Reduced-bias tail index estimators under a third order framework[*]

Frederico Caeiro
Universidade Nova de Lisboa and C.M.A.

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

Lígia Henriques Rodrigues
Instituto Politécnico de Tomar and C.E.A.U.L.

April 16, 2008

Abstract. In this paper we are interested in the comparison, under a third order framework, of classes of second-order reduced-bias tail index estimators, giving particular emphasis to minimum-variance reduced-bias (MVRB) estimators of the tail index γ. The full asymptotic distributional properties of the proposed classes of γ-estimators are derived under a third order framework and the estimators are compared with other alternative estimators of γ, not only asymptotically, but also for finite samples through Monte Carlo techniques. Applications to the log-exchange rates of the Euro against the USA Dollar and the Nasdaq composite index are also provided.

AMS 2000 subject classification. Primary 62G32; Secondary 65C05.
Keywords and phrases. *Statistics of extremes*, *semi-parametric estimation*, *bias estimation*, *heavy tails*, *log-excesses*, *scaled log-spacings*.

1 The estimators under study and scope of the paper

Let $X_1, X_2, \cdots, X_n$ be independent, identically distributed (i.i.d.) random variables (r.v.'s) with a common distribution function (d.f.) F. Let us denote the associated ascending order statistics (o.s.) by

[*]Research partially supported by FCT / POCTI, POCI and PPCDT / FEDER.

$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ and let us assume that there exist sequences of real constants $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$ such that the maximum, linearly normalized, i.e., $(X_{n:n} - b_n)/a_n$, converges in distribution to a non-degenerate r.v. Then the limit distribution is necessarily an *extreme value* d.f.,

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

The d.f. F is said to belong to the max-domain of attraction of EV_γ, and we write $F \in \mathcal{D}_\mathcal{M}(EV_\gamma)$. The parameter γ is the *extreme value index*, the primary parameter of extreme events, with a low frequency, but a high impact. The tail index measures the heaviness of the right tail function $\overline{F} := 1 - F$, and the heavier the tail, the larger the tail index is. In this paper we shall work with Pareto-type distributions, with a strict positive extreme value index, often called *tail index*.

1.1 First, second and third order conditions for heavy tails

Power laws, such as the Pareto income distribution and the Zipf's law for city-size distribution, have been observed a few decades ago in some important phenomena in economics and biology and have seriously attracted scientists in recent years. In *statistics of extremes*, a model F is said to be *heavy-tailed* whenever the *tail function* $\overline{F}$ is a regularly varying function with a negative index of regular variation equal to $-1/\gamma$, or equivalently, the quantile function $U(t) = F^{\leftarrow}(1-1/t), \ t \geq 1$, with $F^{\leftarrow}(x) = \inf\{y : F(y) \geq x\}$, is of regular variation with index $\gamma > 0$, i.e., for all $x > 0$,

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-1/\gamma} \quad \Longleftrightarrow \quad \lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma \quad \Longleftrightarrow \quad F \in \mathcal{D}_\mathcal{M}(EV_{\gamma>0}). \tag{1.1}$$

The *second order parameter*, ρ (≤ 0), 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 $|A|$ must then be of regular variation with index ρ (Geluk and de Haan, 1987). We shall further assume everywhere in the paper that $\rho < 0$.

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 $\mathcal{D}_\mathcal{M}(EV_{\gamma>0})$. Such a 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 in Gomes* (1984)*, Gomes and de Haan* (1999)*, Kaufmann* (2000)*, Raoult and Worms* (2003)*, Diebolt and Guillou* (2005)*, among others. Very popular tails in insurance and finance are tails of the type,* $\overline{F}(x) = \exp\{-H(x)\}$*,* $H \in RV_{1/\theta}$*,* $\theta > 1$*. In a context of Extreme Value Theory we have* $\gamma = 0$ *and* $\rho = 0$*, i.e., we are working with those tails in* $\mathcal{D}_{\mathcal{M}}(EV_0)$*, 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. For relations among different classes of heavy-tailed distributions see Embrechts et al.* (1997)*, Section* 1.4*.*

To obtain information on the order of a possibly non-null asymptotic bias of the second-order reduced-bias tail index estimators considered in this paper, we shall further assume a third order condition, ruling now the rate of convergence in (1.2), and which guarantees that, for all $x > 0$,

$$\lim_{t\to\infty} \frac{\frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} - \frac{x^\rho - 1}{\rho}}{B(t)} = \frac{x^{\rho+\rho'} - 1}{\rho + \rho'}, \tag{1.3}$$

where $|B|$ must then be of regular variation with index $\rho' \le 0$, which we also assume to be negative. Such a condition has already been used in Gomes *et al.* (2002a) and Fraga Alves *et al.* (2003), for the full derivation of the asymptotic behaviour of ρ-estimators, in Gomes *et al.* (2004), for the study of a specific reduced-bias tail index estimator, and in Caeiro and Gomes (2008) for the study of reduced-bias tail index and quantile estimators.

Remark 1.2. *For Hall-Welsh class of Pareto-type models* (Hall and Welsh, 1985)*, with a tail function* $\overline{F}(x) = Cx^{-1/\gamma}\big(1 + D_1 x^{\rho/\gamma} + o(x^{\rho/\gamma})\big)$*, as* $x \to \infty$*,* $\gamma > 0$*,* $C > 0$*,* $D_1 \ne 0$*,* $\rho < 0$*,* (1.2) *holds and we can choose* $A(t) = \alpha t^\rho$*, for an adequate* α*. If we further specify the term* $o\left(x^{\rho/\gamma}\right)$ *and consider a Pareto-type class of models with a tail function* $\overline{F}(x) = Cx^{-1/\gamma}\big(1 + D_1 x^{\rho/\gamma} + D_2 x^{(\rho+\rho_1)/\gamma} + o\big(x^{(\rho+\rho_1)/\gamma}\big)\big)$*,* $C > 0$*,* $D_1,\ D_2 \ne 0$*,* $\rho,\ \rho_1 < 0$*,* (1.3) *holds, with* $\rho' = \max(\rho, \rho_1) \ge \rho$ *and we may choose* $A(t) = \alpha t^\rho$*,* $B(t) = \alpha' t^{\rho'}$*, for adequate* α *and* α'*.*

Remark 1.3. *Note that for most of the common heavy-tailed models* ($\gamma > 0$)*, the third order parameter* ρ' *in* (1.3) *is equal to the second order parameter* ρ *in* (1.2)*. Among those models we mention: the Fréchet model, with d.f.* $F(x) = \exp(-x^{-1/\gamma})$*,* $x \ge 0$*, for which* $\rho' = \rho = -1$*; the generalized Pareto (GP) model, with d.f.* $F(x) = 1 - (1 + \gamma x)^{-1/\gamma}$*,* $x \ge 0$*, for which* $\rho' = \rho = -\gamma$*; the Burr model, with d.f.* $F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}$*,* $x \ge 0$*,* $\rho' = \rho < 0$*; the Student's* t_ν*-model with* ν *degrees of freedom, for which* $\gamma = 1/\nu$ *and* $\rho' = \rho = -2/\nu$*.*

In this paper, we shall assume that (1.3) holds with $\rho,\ \rho' < 0$ and that we can choose

$$B(t) = \beta'\ t^{\rho'}, \quad A(t) = \alpha\ t^{\rho} =: \gamma\ \beta\ t^{\rho}, \quad \beta,\ \beta' \neq 0, \tag{1.4}$$

with $\beta \neq 0$ and $\beta' \neq 0$ “scale” second and third order parameters, respectively. More generally, we can consider $\beta = \beta(t)$ and $\beta' = \beta'(t)$ as arbitrary slowly varying functions.

1.2 The estimators under study

For heavy-tailed models, the classical tail index estimator is Hill's estimator (Hill, 1975), the average of the log-excesses V_{ik} or of the scaled log-spacings U_i, $1 \le i \le k$. With the notation

$$V_{ik} := \ln \frac{X_{n-i+1:n}}{X_{n-k:n}}, \qquad U_i := i\Big\{ \ln \frac{X_{n-i+1:n}}{X_{n-i:n}} \Big\}, \quad 1 \le i \le k < n, \tag{1.5}$$

Hill's estimator can be written as

$$H_n(k) \equiv H(k) = \frac{1}{k}\sum_{i=1}^{k} V_{ik} = \frac{1}{k}\sum_{i=1}^{k} U_i. \tag{1.6}$$

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

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

it is well known that, for models satisfying (1.1), the log-excesses $V_{ik},\ 1 \le i \le k$, in (1.5), are approximately the k o.s. from an i.i.d. exponential sample of size k, with mean value γ and that the scaled log-spacings U_i, $1 \le i \le k$, also in (1.5), are approximately independent and exponential with mean value γ (see, for instance, Caeiro *et al.* (2005), for the sketch of a proof). The Hill estimator in (1.6) is thus consistent for the estimation of γ whenever (1.1) holds and k is intermediate, i.e., (1.7) holds. If we further assume the second order framework in (1.2), the asymptotic representation

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

holds (de Haan and Peng, 1998), where, with $\{E_i\}$ a sequence of i.i.d. standard exponential r.v.'s, $\overline{Z}_k^{(1)} = \sqrt{k}\big(\sum_{i=1}^{k} E_i/k - 1\big)$ is asymptotically standard normal. Under (1.3), we can write,

$$H_n(k) \stackrel{d}{=} \gamma + \frac{\gamma \overline{Z}_k^{(1)}}{\sqrt{k}} + \frac{A(n/k)}{1-\rho} + A(n/k)\Big(\frac{B(n/k)}{1-\rho-\rho'} + O_p\Big(\frac{1}{\sqrt{k}}\Big)\Big)(1 + o_p(1)). \tag{1.8}$$

The adequate accommodation of the bias of Hill's estimator has been extensively addressed in recent years by several authors, among whom we mention Peng (1998), Beirlant *et al.* (1999), Feuerverger and Hall (1999), Gomes *et al.* (2000; 2002b; 2005a; 2005b; 2007b), Gomes and Martins (2001; 2002), Beirlant *et al.* (2002), Caeiro and Gomes (2002). In all these papers, authors are led to second-order reduced-bias tail index estimators with asymptotic variances always larger or equal to $(\gamma\ (1-\rho)/\rho)^2 > \gamma^2$. Recently, Caeiro *et al.* (2005), Gomes and Pestana (2007a) and Gomes *et al.* (2007*a*; 2008) consider, in different ways and essentially under the second order framework in (1.2), the joint external estimation of both the "scale" and the "shape" parameters, β and ρ, respectively, in the A function 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, at least for values k such that $\sqrt{k}A(n/k) \to \lambda$, finite, as $n \to \infty$. Gomes *et al.* (2008) consider

$$\overline{WH}_{\hat\beta,\hat\rho}(k) := \frac{1}{k}\sum_{i=1}^{k} e^{-\hat\beta\ (n/k)^{\hat\rho}\ \psi_{ik}(\hat\rho)}\ V_{ik}, \qquad \psi_{ik} = \psi_{ik}(\rho) = -\frac{(i/k)^{-\rho}-1}{\rho \ln(i/k)}, \tag{1.9}$$

WH standing here for *weighted Hill* estimator. Caeiro *et al.* (2005) consider an estimator of this same type, now denoted,

$$CH_{\hat\beta,\hat\rho}(k) := H(k)\Big(1 - \frac{\hat\beta}{1-\hat\rho}\left(\frac{n}{k}\right)^{\hat\rho}\Big), \tag{1.10}$$

where the dominant component of the bias of Hill's estimator, given by $A(n/k)/(1-\rho) = \gamma\beta\,(n/k)^{\rho}/(1-\rho)$ if (1.4) holds, is thus estimated through $H(k)\hat\beta\,(n/k)^{\hat\rho}/(1-\hat\rho)$, and directly removed from Hill's classical tail index estimator. The notation CH stands for *corrected Hill.* Another class has been introduced in Gomes *et al.* (2007a), with the functional form

$$ML_{\hat\beta,\hat\rho}(k) := H(k) - \hat\beta\ \left(\frac{n}{k}\right)^{\hat\rho}\ \Big(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\hat\rho}U_i\Big). \tag{1.11}$$

These authors consider also the estimator $\overline{ML}_{\hat\beta,\hat\rho}(k) := \sum_{i=1}^{k}\exp\big(-\hat\beta(n/i)^{\hat\rho}\big)U_i/k$, the estimator directly derived from the likelihood equation for γ, based upon the exponential approximation $U_i \approx \gamma\exp(\beta(n/i)^{\rho})E_i$, $1 \le i \le k$, with (β,ρ) fixed, being claimed a better performance of the ML estimator, comparatively to $\overline{ML}$, for a large class of models. This is the reason why we shall here essentially work not with the $\overline{WH}$-estimator in (1.9), but with the bias-corrected Hill estimator

$$WH_{\hat\beta,\hat\rho}(k) := H(k) - \hat\beta\ \left(\frac{n}{k}\right)^{\hat\rho}\Big(\frac{1}{k}\sum_{i=1}^{k}\psi_{ik}(\hat\rho)\ V_{ik}\Big), \tag{1.12}$$

with ψ_{ik} given in (1.9). In all these classes $\hat{\beta}$ and $\hat{\rho}$ need to be adequate consistent estimators of the second order parameters β and ρ, respectively.

1.3 Scope of the paper

In this paper, we derive asymptotic distributional properties of the three classes of *"unbiased Hill"* (UH) estimators in (1.10), (1.11) and (1.12), under the third order framework in (1.3), obtaining full information on their asymptotic bias. In Section 2, we shall briefly review the estimation of the two second order parameters β and ρ, adding some new results related with the estimation of the second order parameter β, under a third order framework. In Section 3, first assuming that only γ is unknown, we shall state a theorem that provides an obvious technical motivation for the estimators under consideration and full information on the bias. Next, we shall derive the asymptotic behaviour of the UH estimators, estimating first β and ρ at a larger k value than the one used for the tail index estimation. We also do that only with the estimation of ρ, estimating β at the same level k used for the tail index estimation. Finally, generalizing for any of the UH-statistics under study, a result obtained in Gomes and Pestana (2007c) for the statistic CH in (1.10), we refer the estimation of the three parameters at the same level k. In Section 4, and through the use of Monte Carlo simulation techniques, we exhibit the performance of these *UH*-estimators, comparatively to the classical Hill estimator and to the *generalized Jackknife* estimator,

$$GJ_{\hat{\rho}}(k) := \left(\sqrt{2\ M_n^{(2)}(k)} - (2-\hat{\rho})\ M_n^{(2)}(k)/\big(2\ M_n^{(1)}(k)\big)\right)/\hat{\rho}, \tag{1.13}$$

studied in Gomes and Martins (2002), and where, with V_{ik} the log-excesses in (1.5),

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k} V_{ik}^{j}, \quad j \geq 1 \quad \left[M_n^{(1)} \equiv H \text{ in } (1.6)\right]. \tag{1.14}$$

The *generalized Jackknife* estimator in (1.13) was considered as a representative of the classical "asymptotically unbiased" tail index estimators with an asymptotic variance larger than or equal to $\gamma^2\left((1-\rho)/\rho\right)^2$, the minimal asymptotic variance of any "asymptotically unbiased" estimator in Drees' class of functionals (Drees, 1998). In the simulations, we shall consider only an external estimation of ρ at a level k_1 of a larger order than that of the level k on which we base the tail index estimation. Such a decision is related with the discussion initiated in Gomes and Martins (2002) on the advantages of using an external estimation of the second order parameter ρ, or even its misspecification, as in Gomes and Martins (2004), versus an internal estimation at the same k, which leads to high volatile sample paths for small up to “moderate” k. In Section 5 we provide an illustration of the behaviour of these new minimum-variance reduced-bias (MVRB) estimators through the analysis of two data sets in the field of finance, and some overall conclusions are drawn. Finally, in Section 6, we provide the proofs of the theorems in Sections 2 and 3.

2 Third order framework and second order parameters estimation

2.1 The estimation of ρ

We shall consider the class of estimators of the second order parameter ρ proposed by Fraga Alves *et al.* (2003). Under adequate general conditions, they are semi-parametric asymptotically normal estimators of ρ, whenever $\rho < 0$, which show, for a large variety of models, highly stable sample paths as functions of k, the number of top o.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, 2006). It is defined as

$$\hat{\rho}(k;\tau) \equiv \hat{\rho}_\tau(k) := -\left|3(T_n(k;\tau)-1)/(T_n(k;\tau)-3)\right|, \qquad (2.1)$$

where, with $M_n^{(j)}(k)$ given in (1.14) and the notation $a^{b\tau} = b\ln a$, whenever $\tau = 0$,

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

We shall next summarize the results proved in Fraga Alves *et al.* (2003), making explicit both the asymptotic bias and variance of the ρ-estimators in (2.1):

Proposition 2.1 (Fraga Alves *et al.*, 2003)**.** *Under the second order framework in* (1.2)*, with* $\rho < 0$*, if* (1.7) *holds and* $\sqrt{k}\, A(n/k) \to \infty$*, as* $n \to \infty$*, the statistics* $\hat{\rho}(k;\tau)$ *in* (2.1) *converge in probability to* ρ*, as* $n \to \infty$*, for any real* τ*. Moreover, under the third order framework in* (1.3)*, if* $\sqrt{k}\, A^2(n/k) \to \lambda_A$*, finite, and* $\sqrt{k}\, A(n/k)\, B(n/k) \to \lambda_B$*, also finite,* $\sqrt{k}\, A(n/k)\,(\hat{\rho}(k;\tau)-\rho)$ *are asymptotically normal with asymptotic variance*

$$\sigma_R^2 \equiv \sigma_R^2(\gamma;\rho) = \left(\frac{\gamma(1-\rho)^3}{\rho}\right)^2 \left(2\rho^2 - 2\rho + 1\right), \qquad (2.2)$$

the variance of a r.v. $\sigma_R W_k^R$*, with* W_k^R *asymptotically standard normal and given by*

$$W_k^R = \frac{1}{\sqrt{2\rho^2-2\rho+1}}\left((3-\rho)\,\overline{N}_k^{(1)} - (3-2\rho)\,\overline{N}_k^{(2)} + (1-\rho)\,\overline{N}_k^{(3)}\right), \qquad (2.3)$$

where, with $\{E_i\}$ i.i.d. standard exponential r.v.'s, and $\Gamma(t)$ the complete Gamma function,

$$\overline{N}_k^{(\alpha)} = \frac{1}{\Gamma(\alpha+1)\sqrt{k}} \sum_{i=1}^{k} E_i^\alpha - \sqrt{k}, \quad \alpha \geq 1. \tag{2.4}$$

There is moreover a possibly non-null asymptotic bias, given by $b_{_R} := \lambda_{_A} u_{_R} + \lambda_{_B} v_{_R}$, where

$$u_{_R} \equiv u_{_R}(\gamma,\rho;\tau) = \frac{\rho\left(\tau(1-2\rho)^2(3-\rho)(3-2\rho) - 6\rho\left(4\rho^3 - 16\rho^2 + 20\rho - 7\right)\right)}{12\ \gamma\left((1-\rho)(1-2\rho)\right)^2} \tag{2.5}$$

and

$$v_{_R} \equiv v_{_R}(\rho,\rho') = \rho'\Big(1+\frac{\rho'}{\rho}\Big)\Big(\frac{1-\rho}{1-\rho-\rho'}\Big)^3. \tag{2.6}$$

Indeed, we can write the asymptotic distributional representation,

$$\hat\rho(k;\tau) \ \stackrel{d}{=}\ \rho + \frac{\sigma_{_R} W_k^R}{\sqrt{k}\ A(n/k)} + \left(u_{_R} A(n/k) + v_{_R} B(n/k)\right)(1+o_p(1)).$$

Remark 2.1. *Note that if if $\sqrt{k}\ A^2(n/k) \to \infty$ or $\sqrt{k}\ A(n/k)\ B(n/k) \to \infty$, $\hat\rho(k;\tau) - \rho$ is of the order of $A(n/k)$ or of $B(n/k)$, the one of highest order.*

2.1.1 A few remarks on the choice of a level k_1 for the estimation of ρ

We now rephrase for the more general set of models in this paper, the comments made in Caeiro and Gomes (2008) on the choice of the value k_1 that should be used for the estimation of ρ.

(1) The ideal situation would perhaps be the choice of an "optimal" level k_1 for the estimation of ρ, in the sense of a k_1 that enables us to guarantee the asymptotic normality of the ρ-estimators with a non-null asymptotic bias. That level k_1 is then such that $\sqrt{k_1}\ A(n/k_1)B(n/k_1) \ \to\ \lambda_{_{B_1}}$ and $\sqrt{k_1}\ A^2(n/k_1) \ \to\ \lambda_{_{A_1}}$, both finite, with at least one of them non-null, let us say $\lambda_{_{B_1}}$. We should then get $k_1 = O\big(n^{-2(\rho+\rho')/(1-2(\rho+\rho'))}\big)$. Denoting $\hat\rho = \hat\rho(k_1;\tau)$ any of the ρ-estimators in this section computed at such a k_1, $\hat\rho - \rho = O_p\big(1/\left(\sqrt{k_1}A(n/k_1)\right)\big) = O_p\big(n^{\rho'/(1-2(\rho+\rho'))}\big)$, i.e.,

$$\hat\rho - \rho = o_p(1/\ln n), \quad \text{as } n \to \infty, \tag{2.7}$$

a condition needed later on in Section 3. In practice, such a k_1 has only a "limited" interest, at the current state-of-the-art. It is however of high theoretical interest.

(2) It is easy to show that if $\rho = \rho'$, a large subset of the class of models in this paper (see Remark 1.3), there exists τ such that the asymptotic bias $b_{_R} := \lambda_{_A} u_{_R} + \lambda_{_B} v_{_R} \equiv 0$. This partially justifies the assumption of the following condition:

Condition U: There exist $\tau_{_U}$ and k_1, with $\sqrt{k_1}\ A(n/k_1)B(n/k_1) \to \infty$ and/or $\sqrt{k_1}\ A^2(n/k_1) \to \infty$, such that, with $\hat\rho(k;\tau)$ defined in (2.1), $\hat\rho_{_U} - \rho = \hat\rho(k_1;\tau_{_U}) - \rho = O_p\left(1/\left(\sqrt{k_1}A(n/k_1)\right)\right)$.

This is obviously a strong assumption, practically equivalent to saying that, for a specific model, there is $(\tau_{_U}, k_1)$ such that $\hat\rho_{_U} = \hat\rho(k_1;\tau_{_U})$ is an "unbiased" estimator of ρ, so that the bias has no influence in the rate of convergence, which is kept at $1/\left(\sqrt{k_1}A(n/k_1)\right)$. Indeed, such a claim is made essentially on the basis of the high stability of sample paths of the ρ-estimates in (2.1) for a specific $\tau = \tau_{_U}$ and large values of k (see the sample path of $\hat\rho_0(k) = \hat\rho(k;0)$ in Figure 5). Then, the use of a value k_1 larger than the "optimal" level for the ρ-estimation considered in item (1), but intermediate, like for instance, the one suggested in Gomes and Martins (2002), $k_1 := \min(n-1, 2n/\ln\ln n)$, enables us to guarantee that $\hat\rho_{_U} - \rho = o_p(1/\ln n)$. Indeed, if we do not assume *Condition U* for this k_1, since $\sqrt{k_1}\ A^2(n/k_1)$ and/or $\sqrt{k_1}\ A(n/k_1)B(n/k_1)$ go both to infinity, $\hat\rho(k_1;\tau) - \rho$, being of the order of $A(n/k_1)$ or of the order of $B(n/k_1)$, is a $O_p\left((\ln\ln n)^\eta\right)$ for some $\eta < 0$. Consequently, $\hat\rho(k_1;\tau) - \rho$ would be of a larger order than $1/\ln n$. If we assume the validity of *Condition U* for $k_1 = O(n/\ln\ln n)$, we get $\hat\rho_{_U} - \rho = O_p\left(1/\left(\sqrt{k_1}A(n/k_1)\right)\right) = O_p\left((\ln\ln n)^{(1-2\rho)/2}/\sqrt{n}\right)$, which is obviously of smaller order than $1/\ln n$, i.e., (2.7) holds. This will be the unique situation under which we can work with the k_1 suggested in Gomes and Martins (2002) and still guarantee the above mentioned property (2.7) on the ρ-estimator, and a possible generalization of Theorem 3.1 for $UH_{\hat\beta_{_U},\hat\rho_{_U}}$, with $\hat\beta_{_U}$ an adequate β-estimator, to be specified in subsection 2.2.

(3) If we consider a level k_1 of the order of $n^{1-\epsilon}$, for some small $\epsilon > 0$, we can also guarantee (2.7) for a large class of models, without the need to assume a condition as strong as *Condition U*. To have consistent estimation of ρ, we need to have $\sqrt{k_1}\ A(n/k_1) \to \infty$, and this holds if and only if $\rho > \frac{1}{2} - \frac{1}{2\epsilon} \to -\infty$, as $\epsilon \to 0$, i.e., we have an almost irrelevant restriction in the model. Then, with further similar irrelevant restrictions, we get either $\sqrt{k_1}\ A(n/k_1)B(n/k_1) \to \infty$ or $\sqrt{k_1}\ A^2(n/k_1) \to \infty$, and $\hat\rho - \rho$, being possibly of the order of $A(n/k_1)$ or of the order of $B(n/k_1)$, is of the order of a negative power of n, i.e., again of smaller order than $1/\ln n$. This is the reason why, such as done in Caeiro *et al.* (2005), Gomes and Pestana (2007*a*; 2007*b*) and Gomes *et al.* (2007*a*; 2008), we advise in practice,

as a compromise between theoretical and practical considerations, the use of any intermediate level like $k_1 = [n^{1-\epsilon}]$ for some $\epsilon > 0$, small.

2.1.2 A remark on the choice of the tuning parameter τ in the estimation of ρ

The theoretical and simulated results in Fraga Alves *et al.* (2003), together with the use of these estimators in the generalized Jackknife statistics in Gomes *et al.* (2000), as done in Gomes and Martins (2002), as well as their use in the estimators in (1.9) (Gomes *et al.*, 2008), in (1.10) (Caeiro *et al.*, 2005; Gomes and Pestana, 2007*a*) and in (1.11) (Gomes *et al.*, 2007a), lead us again to advise in practice the use of $\tau = 0$ for $\rho \in [-1, 0)$ and $\tau = 1$ for $\rho \in (-\infty, -1)$. However, practitioners should not choose blindly the value of τ in (2.1). It is sensible to draw a few sample paths of $\hat\rho_\tau(k) = \hat\rho(k;\tau)$, as functions of k, electing the value of τ which provides the highest stability for large k, by means of any stability criterion, like the one suggested in Gomes *et al.* (2005*a*) or Gomes and Pestana (2007*a*). For not too small n, we are most frequently led to the above mentioned choice, $\hat\rho_0$ if $\rho \geq -1$ and $\hat\rho_1$ if $\rho < -1$, when we consider only the tuning parameters $\tau = 0$ and $\tau = 1$ as possible alternatives. In practice, the adequate choice of τ is more crucial than the choice of k_1.

2.2 The estimation of β

We shall consider the β-estimator first obtained in Gomes and Martins (2002) and based on the scaled log-spacings U_i in (1.5), $1 \leq i \leq k$. On the basis of any consistent estimator $\hat\rho$ of the second order parameter ρ, we shall consider the β-estimator, $\hat\beta(k;\hat\rho)$, where, with $\rho < 0$,

$$\hat\beta(k;\rho) := \frac{\left(\frac{k}{n}\right)^{\rho}\left\{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}\right)\left(\frac{1}{k}\sum_{i=1}^{k}U_i\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}\, U_i\right)\right\}}{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}\right)\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}\, U_i\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\rho}\, U_i\right)}. \tag{2.8}$$

Gomes and Martins (2002), keeping up to the second order framework in (1.2), have got the asymptotic behaviour of $\hat\beta(k;\rho)$ in (2.8). If we go into the third order framework, we then get the following result, a generalization of a result in Gomes *et al.* (2008):

Theorem 2.1. *If the third order framework in* (1.3) *holds, as well as* (1.4), $k = k_n$ *is a sequence of intermediate positive integers, i.e.* (1.7) *holds, and* $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$, *then we get*

$$\hat\beta(k;\rho) \ \stackrel{d}{=}\ \beta + \frac{\gamma\ \beta(1-\rho)\sqrt{(1-2\rho)}}{\rho\ \sqrt{k}\ A(n/k)}\ W_k^{B} + \beta\big(u_B A(n/k) + v_B B(n/k)\big)(1 + o_p(1)), \tag{2.9}$$

where W_k^B *is asymptotically standard normal. More precisely we can write*

$$W_k^B = \frac{(1-\rho)\sqrt{1-2\rho}}{|\rho|}\Big(\frac{\overline{Z}_k^{(1)}}{1-\rho} - \frac{\overline{Z}_k^{(1-\rho)}}{\sqrt{1-2\rho}}\Big), \tag{2.10}$$

where, with $\{E_i\}$ *a sequence of i.i.d. standard exponential r.v.'s,*

$$\overline{Z}_k^{(\alpha)} := \sqrt{(2\alpha-1)\ k}\ \Big(\frac{1}{k}\sum_{i=1}^{k}\Big(\frac{i}{k}\Big)^{\alpha-1}E_i - 1/\alpha\Big), \tag{2.11}$$

for any real $\alpha \geq 1$ $\big[\overline{Z}_k^{(1)} \equiv \overline{N}_k^{(1)}$ *in* (2.4)$\big]$. *Moreover,*

$$u_B \equiv u_B(\gamma,\ \rho) = -\frac{2(1-\rho)}{\gamma(1-3\rho)}, \quad v_B \equiv v_B(\rho,\ \rho') = \frac{(1-\rho)(1-2\rho)(\rho+\rho')}{\rho(1-\rho-\rho')(1-2\rho-\rho')}. \tag{2.12}$$

Consequently, $\hat\beta(k;\rho)$ *converges in probability to* β*, whenever* k *is intermediate and* $\sqrt{k}\ A(n/k) \to \infty$*, as* $n \to \infty$*. Moreover, if* $\sqrt{k}\ A^2(n/k) \to \lambda_A$ *and* $\sqrt{k}\ A(n/k)\ B(n/k) \to \lambda_B$*, both finite,*

$$\sqrt{k}\ A(n/k)\big(\hat\beta(k;\rho)-\beta\big)/\beta \overset{a}{\sim} Normal\big(\lambda_A u_B + \lambda_B v_B,\ \sigma_B^2\big), \tag{2.13}$$

where

$$\sigma_B^2 \equiv \sigma_B^2(\gamma,\ \rho) = \Big(\frac{\gamma\ (1-\rho)}{\rho}\Big)^2(1-2\rho). \tag{2.14}$$

Under the validity of Condition U for a $k_1 = O(n/\ln\ln n)$*, these same results hold for* $\hat\beta(k;\hat\rho_U)$*.*

If we replace ρ *by* $\hat\rho(k;\tau)$ *in* (2.1)*, the rate of convergence of* $\hat\beta(k;\hat\rho(k;\tau))$ *is then of the order of* $\big\{\ln(n/k)/(\sqrt{k}\ A(n/k))\big\}$*, which must converge to zero, so that* $\hat\beta(k;\hat\rho(k;\tau))$ *is consistent for the estimation of* β*, and*

$$\big(\hat\beta(k;\hat\rho(k;\tau))-\beta\big)/\beta \overset{p}{\sim} -\ln(n/k)\big(\hat\rho(k;\tau)-\rho\big). \tag{2.15}$$

If apart from $\sqrt{k}\ A(n/k)/\ln(n/k) \to \infty$*, we further assume that* $\sqrt{k}\ A^2(n/k) \to \lambda_A$ *and* $\sqrt{k}\ A(n/k)\ B(n/k) \to \lambda_B$*, both finite, then* $\sqrt{k}\ A(n/k)\big(\beta-\hat\beta(k;\hat\rho(k;\tau))\big)/(\beta\ln(n/k)) \overset{a}{\sim}$ $Normal\big(\lambda_A u_R + \lambda_B v_R,\ \sigma_R^2\big)$*, with* σ_R^2*,* u_R *and* v_R *given in* (2.2), (2.5) *and* (2.6)*, respectively.*

Remark 2.2. *If* $\rho = \rho'$ *in* (1.3)*, as happens with most common heavy-tailed models* (see Remark 1.3), we get $u_B = -v_B/\gamma$. Since for a Burr and a GP model, we can choose $B(t) = A(t)/\gamma$, we have a null mean value for $\sqrt{k}\big(\hat\beta(k;\rho)-\beta\big)$, even when $\lambda_A, \lambda_B \neq 0$. This justifies the good performance of this β-estimator for Burr and GP models, as detected in Caeiro and Gomes (2006).

Remark 2.3. *If we consider* $\hat\beta \equiv \hat\beta(k_1;\hat\rho)$*, with* $\hat\rho$ *any of the estimators in* (2.1)*, computed also at the level* k_1*,* $\hat\beta-\beta$ *is thus, from* (2.15)*, of the order of* $(\hat\rho-\rho)\ln(n/k_1)$*. Consequently, the validity of* (2.7)*, enables us to guarantee the consistency of* $\hat\beta \equiv \hat\beta(k_1;\hat\rho)$*.*

3 Asymptotic behaviour of the "*Unbiased Hill*" estimators

3.1 Known β and ρ

Under the second order framework in (1.2), further assuming that $A(\cdot)$ can be chosen as in (1.4), and for levels k such that (1.7) holds, Gomes *et al.* (2008), Caeiro *et al.* (2005) and Gomes *et al.* (2007a) got, for the r.v.'s $\overline{WH}_{\beta,\rho}(k)$, $CH_{\beta,\rho}(k)$ and $ML_{\beta,\rho}$, respectively, generically denoted $UH_{\beta,\rho}$, an asymptotic distributional representation of the type

$$UH_{\beta,\rho}(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, \overline{Z}_k^{(1)} + R_{UH}(k), \quad R_{UH}(k) = o_p(A(n/k)), \tag{3.1}$$

where $\overline{Z}_k^{(\alpha)}$ is the asymptotically standard normal r.v. in (2.11). It is trivial that (3.1) also holds for the r.v. $WH_{\beta,\rho}$, with $WH_{\hat\beta,\hat\rho}$ given in (1.12). Assuming now the third order framework in (1.3):

Theorem 3.1. *Under* (1.3)*, further assuming that $A(\cdot)$ and $B(\cdot)$ can be chosen as in* (1.4)*, and for levels k such that* (1.7) *holds, we can specify the term $R_{UH}(k)$ in* (3.1)*, writing,*

$$R_{UH}(k) = A(n/k)\Big(u_{UH}A(n/k) + v_{UH}B(n/k) + O_p\Big(\frac{1}{\sqrt{k}}\Big)\Big)(1 + o_p(1)), \tag{3.2}$$

where, with

$$a_2(\rho) = -\left(\ln(1-2\rho) - 2\ln(1-\rho)\right)/\rho^2, \tag{3.3}$$

$$u_{WH} = -\frac{a_2(\rho)}{\gamma}, \quad u_{CH} = -\frac{1}{\gamma(1-\rho)^2}, \quad u_{ML} = -\frac{1}{\gamma(1-2\rho)}, \quad v_{UH} = \frac{1}{1-\rho-\rho'}. \tag{3.4}$$

Consequently, even if $\sqrt{k}\ A(n/k) \to \infty$, with $\sqrt{k}\ A^2(n/k) \to \lambda_A$ and $\sqrt{k}\ A(n/k)\ B(n/k) \to \lambda_B$, λ_A and λ_B finite, $\sqrt{k}\ \ (UH_{\beta,\rho}(k) - \gamma) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal\big(b_{UH} := \lambda_A u_{UH} + \lambda_B v_{UH}, \gamma^2\big)$. If $\sqrt{k}\ A^2(n/k) \to \infty$ or $\sqrt{k}\ A(n/k)B(n/k) \to \infty$, $(UH_{\beta,\rho}(k) - \gamma)/A(n/k)$ is either $O_p(A(n/k))$ or $O_p(B(n/k))$, the one of highest order.

Remark 3.1. *Note that since $\lambda_A \ge 0$ and $(1-\rho)^2 > 1/a_2(\rho) > 1-2\rho$ for any $\rho < 0$, $b_{CH} \ge b_{WH} \ge b_{ML}$. All depends then on the sign of the bias, but we expect the sample paths of CH to be always above the sample paths of WH, these ones above the ones of ML. The ML-statistics are then preferable to the other ones whenever the bias are all positive.*

Remark 3.2. *If $\rho = \rho'$, we get $b_{ML} = (\lambda_B - \lambda_A/\gamma)/(1-2\rho)$. On the basis of Remark 2.2, $\lambda_B = \lambda_A/\gamma$ and $b_{ML} = 0$ for Burr and GP models, a point in favour of the ML-statistic, as mentioned in* Gomes et al. (2007a)*. For the Student's t_ν model with $\nu = 1$ degree of freedom, $|b_{ML}| \ge b_{CH} \ge |b_{WH}|$, and $|b_{ML}| \ge |b_{WH}| \ge |b_{CH}|$ if $\nu \ge 2$.*

Remark 3.3. *A possible adaptive choice of k based on the minimum MSE of this kind of "asymptotically unbiased" estimators is now closer, but still slightly problematic at the current state-of-the-art, because we also need to estimate β' and ρ', the parameters in the B function in* (1.3)*. It seems worth investing in the use of the bootstrap methodology to estimate the optimal sample fraction in these classes of UH estimators, but that is beyond the scope of the present paper.*

3.2 Estimation of both second order parameters at the same level k_1

We may state the following:

Theorem 3.2. *Under the third order framework in* (1.3)*, let us consider the tail index estimators in* (1.10)*,* (1.11) *and* (1.12)*, generically denoted $UH_{\hat\beta,\hat\rho}(k)$, for any of the estimators $\hat\beta$ and $\hat\rho$ in* (2.1) *and in* (2.8)*, respectively, both computed at a level k_1 such that* (2.7) *holds. Then, $\sqrt{k}\{UH_{\hat\beta,\hat\rho}(k)-\gamma\}$ are asymptotically normal with null mean value and variance*

$$\sigma_1^2 = \gamma^2, \tag{3.5}$$

not only when $\sqrt{k}\ A(n/k)\to 0$, but also whenever $\sqrt{k}\ A(n/k)\to\lambda$, finite.

We can still get this same limiting result for levels k such that $\sqrt{k}\ A(n/k)\to\infty$, provided that $k = o(k_1)$, as $n\to\infty$, and we choose k_1 optimal for the estimation of ρ, i.e., such that $\sqrt{k_1}\ A^2(n/k_1)\to\lambda_{A_1}$ and $\sqrt{k_1}\ A(n/k_1)B(n/k_1)\to\lambda_{B_1}$, finite.

If for k such that $\sqrt{k}\ A^2(n/k)\to\lambda_A$ and $\sqrt{k}\ A(n/k)B(n/k)\to\lambda_B$, both finite, we assume the validity of Condition U for a level $k_1 = O(n/\ln\ln n)$, and consider $\hat\beta_U = \hat\beta(k_1;\hat\rho_U)$, the tail index estimators $UH_{\hat\beta_U,\hat\rho_U}(k)$ have an asymptotic variance still equal to γ^2 and a non-null asymptotic bias, still given by $b_{UH} := \lambda_A u_{UH} + \lambda_B v_{UH}$, with u_{UH} and v_{UH} given in (3.4)*. This same result holds for any $UH_{\hat\beta,\hat\rho}(k)$ provided that $\hat\rho-\rho = o_p(\ln(n/k)/(\sqrt{k}A(n/k))$.*

3.3 Estimation of γ and β at the same level k

If we consider γ and β estimated at the same k, keeping the estimation of ρ at a larger k_1, we are going to have an increase in the variance of our final tail index estimators. Indeed, in the above mentioned papers on the UH-statistics, an asymptotic normal behaviour has already been obtained for any of the tail index estimators $UH^*_{\hat\rho} := UH_{\hat\beta(k;\hat\rho),\hat\rho}$, being the rate of convergence of the order of $1/\sqrt{k}$ and the asympotic variance $\gamma^2\left((1-\rho)/\rho\right)^2$. Under conditions similar to the ones in Theorem 3.2, we may also obtain information on the asymptotic bias of $UH^*_{\hat\rho}(k)$, if we assume a third order framework:

Theorem 3.3. *If the third order condition* (1.3) *holds,* $k = k_n$ *is a sequence of intermediate integers, i.e.,* (1.7) *holds, and* $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$*, with* $\sqrt{k}\ A^2(n/k) \to 0$ *and* $\sqrt{k}\ A(n/k)\ B(n/k) \to 0$*, as* $n \to \infty$*,* $\sqrt{k}\big(UH^*_{\hat\rho}(k) - \gamma\big)$ *is asymptotically normal, with null mean value and variance*

$$\sigma_2^2 := \gamma^2\Big(\frac{1-\rho}{\rho}\Big)^2, \tag{3.6}$$

provided that we consider the estimator $\hat\rho(k;\tau)$ *in* (2.1)*, computed at a level* k_1 *such that* $\sqrt{k_1}\ A^2(n/k_1) \to \lambda_{A_1}$ *and* $\sqrt{k_1}\ A(n/k_1)B(n/k_1) \to \lambda_{B_1}$ *or, more generally, such that* $\hat\rho - \rho = O_p\left(1/\left(\sqrt{k_1}\ A(n/k_1)\right)\right)$ *and* $k = o(k_1)$*, as* $n \to \infty$*.*

If we assume that Condition U holds for a level $k_1 = O(n/\ln\ln n)$*, then, the asymptotic behaviour of* $UH^*_{\hat\rho_U}(k)$ *is equivalent to the one of* $UH^*_{\rho}(k)$*. More specifically, the asymptotic variances of* $UH^*_{\rho}(k)$ *are kept equal to* $(\gamma(1-\rho)/\rho)^2$*, and whenever* $\sqrt{k}\ A^2(n/k) \to \lambda_A$ *and* $\sqrt{k}\ A(n/k)B(n/k) \to \lambda_B$*, both finite, the asymptotic bias of* $UH^*_{\rho}(k)$ *is equal to*

$$b^*_{UH} = \lambda_A\left(u_{UH} - \frac{u_B}{1-\rho}\right) + \lambda_B\left(v_{UH} - \frac{v_B}{1-\rho}\right) =: \lambda_A u^*_{UH} + \lambda_B v^*_{UH}, \tag{3.7}$$

with (u_B, v_B) *and* (u_{UH}, v_{UH}) *given in* (2.12) *and* (3.4)*, respectively.*

Remark 3.4. *Results similar to the ones mentioned in Remarks 3.1 and 3.2 hold also for the bias in Theorem 3.3. Indeed,* $b^*_{ML} = 0$ *for Burr and GP models, whenever* $\rho = \rho'$*, and* $b^*_{CH} \ge b^*_{WH} \ge b^*_{ML}$*.*

Remark 3.5. *As noticed in* Caeiro *et al.* (2005)*, if we compare Theorems 3.2 and 3.3, we see that the estimation of* 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* $((1-\rho)/\rho)^2 > 1$*. As noticed before, like for instance in* Gomes and Martins (2002)*, the asymptotic variance of the estimator in* Feuerverger and Hall (1999) (*where the three parameters are computed at the same level* k) *is given by* $\sigma^2_{FH} := \gamma^2\left((1-\rho)/\rho\right)^4$*. Note that* Peng and Qi (2004) *have derived this same asymptotic variance for an approximate second-order maximum-likelihood tail index estimator. The asymptotic variance of the MVRB estimators is* $\sigma_1^2 = \gamma^2$*, given in* (3.5)*, and we have* $\sigma_1 < \sigma_2 < \sigma_{FH}$ *for all* $\rho \le 0$*.*

3.4 Estimation of γ, β and ρ at the same level k

If we consider γ, β and ρ estimated at the same k, we are further going to have an extra increase in the variance of the tail index estimators $UH^{**}(k) := UH_{\hat\beta(k;\hat\rho(k)),\hat\rho(k)}(k)$. Indeed, we are able to derive an asymptotic normal behaviour, at a rate of convergence still of the order of $1/\sqrt{k}$, but the asympotic variance increases. Indeed, as a particular case of Theorem 3.1 in Gomes and Pestana (2007c):

Theorem 3.4. *If the third order condition* (1.3) *holds,* $k = k_n$ *is a sequence of intermediate integers, i.e.,* (1.7) *holds, and* $\sqrt{k}\ A(n/k) \to \infty$, *with* $\sqrt{k}\ A^2(n/k) \to 0$ *and* $\sqrt{k}\ A(n/k)B(n/k) \to 0$, $\sqrt{k}\{UH^{**}(k) - \gamma\} \underset{n\to\infty}{\overset{d}{\longrightarrow}}$ *Normal* $\left(0, \sigma_3^2\right)$, *with*

$$\sigma_3^2 := \gamma^2\Big(1 + \Big(\frac{1-\rho}{\rho}\Big)^2 - \frac{2(1-\rho)^3}{\rho}\Big) > \gamma^2\Big(\frac{1-\rho}{\rho}\Big)^2, \tag{3.8}$$

i.e., we get the same rate of convergence of the order of $1/\sqrt{k}$ *for* $UH^{**}(k)$, *but the asymptotic variance increases. If* $\sqrt{k}\ A^2(n/k) \to \lambda_{_A}$ *and* $\sqrt{k}\ A(n/k)B(n/k) \to \lambda_{_B}$, *both finite, then, with* $u_{_R}$, $v_{_R}$ *and* $(u_{_{UH}}, v_{_{UH}})$ *given in* (2.5), (2.6) *and* (3.4), *respectively, the asymptotic bias of* $UH^{**}(k)$ *is*

$$b^{**}_{_{UH}} := \lambda_{_A}\Big(u_{_{UH}} - \frac{u_{_R}}{(1-\rho)^2}\Big) + \lambda_{_B}\Big(v_{_{UH}} - \frac{v_{_R}}{(1-\rho)^2}\Big) =: \lambda_{_A} u^{**}_{_{UH}} + \lambda_{_B} v^{**}_{_{UH}}. \tag{3.9}$$

4 Finite sample behaviour of the estimators

We have implemented simulation experiments, with 5000 runs, based on the estimation of β at the same level k_1 we have used for the estimation of ρ, the level $k_1 = n^{0.999}$. We use the notation $\hat\beta_{j1} = \hat\beta(k_1;\hat\rho_{_j})$, $\hat\rho_j = \hat\rho(k_1;j)$, $j = 0, 1$, with $\hat\rho(k;\tau)$ and $\hat\beta(k;\rho)$ given in (2.1) and (2.8), respectively. These estimators of ρ and β have been incorporated in the *"unbiased Hill"* estimators, leading to $UH_{\hat\beta_{j1},\hat\rho_j}(k)$, $j = 0,\ 1$. The simulations show that the tail index estimators $UH_{\hat\beta_{j1},\hat\rho_j}(k)$, j equal to either 0 or 1, according as $|\rho| \le 1$ or $|\rho| > 1$, seem to work quite well, as illustrated in Figures from 1 until 4. In these figures we picture for Fréchet and generalized Pareto (GP) 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 and the generalized Jacknife estimator $GJ_j = GJ_{\hat\rho_j}$ in (1.13), together with $\overline{WH}_j = \overline{WH}_{\hat\beta_{j1},\hat\rho_j}$, $CH_j = CH_{\hat\beta_{j1},\hat\rho_j}$ $ML_j = ML_{\hat\beta_{j1},\hat\rho_j}$ and $WH_j = WH_{\hat\beta_{j1},\hat\rho_j}$, $j = 0$ or $j = 1$, according as $|\rho| \le 1$ or $|\rho| > 1$. The r.v.'s $ML = ML_{\beta,\rho}$ are also pictured, so that we may see that, for some models, there exists still a significant difference between the behaviour of the statistics under study and that of the r.v.'s. Such a discrepancy suggests that some improvement in the estimation of second order parameters β and ρ is still welcome.

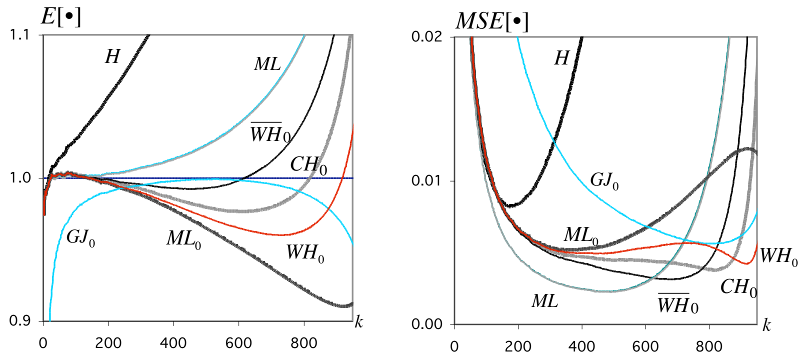

Figure 1: Underlying *Fréchet* parent with $\gamma = 1$ $(\rho = -1)$.

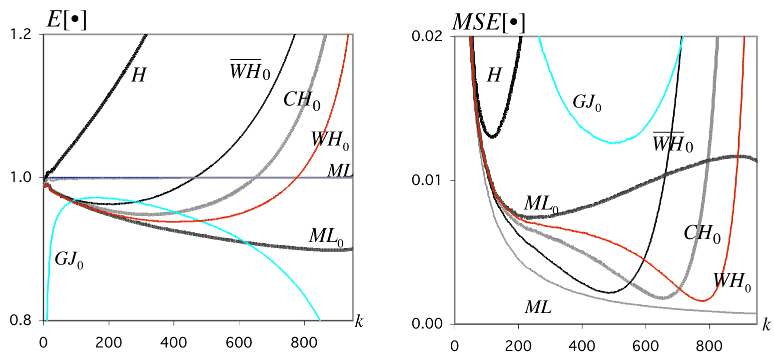

Figure 2: Underlying *GP* parent with $\gamma = 1$ $(\rho = -1)$.

Remark 4.1. *Note that the comment made in Remark 3.1 is coherent with the pictures of the mean values of* CH_0*,* ML_0 *and* WH_0*.*

Remark 4.2. *For the Fréchet model (Figure 1), and among the* UH*-estimators simulated, the* $\overline{WH}$*-estimator exhibits the best performance.*

Remark 4.3. *For a generalized Pareto (GP) model, we may further draw the following comments:*

- *For any* ρ*, the* ML *r.v. seems to be unbiased for the estimation of* γ*, for all* k*.*
- *For* $\rho = -1$ *(Figure 2), the* WH *statistic is the best one regarding* MSE *at the optimal level,*

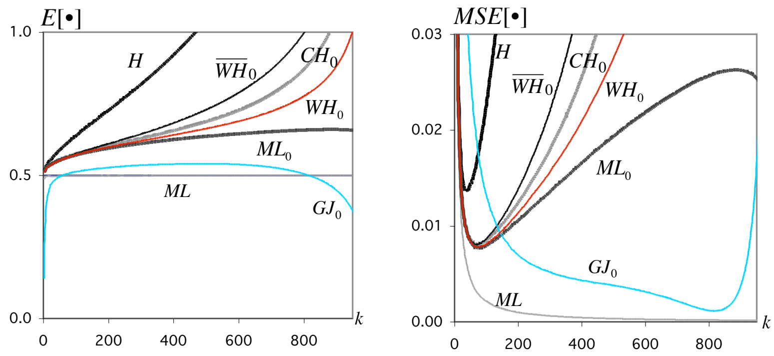

Figure 3: Underlying GP parent with $\gamma = 0.5$ $(\rho = -0.5)$.

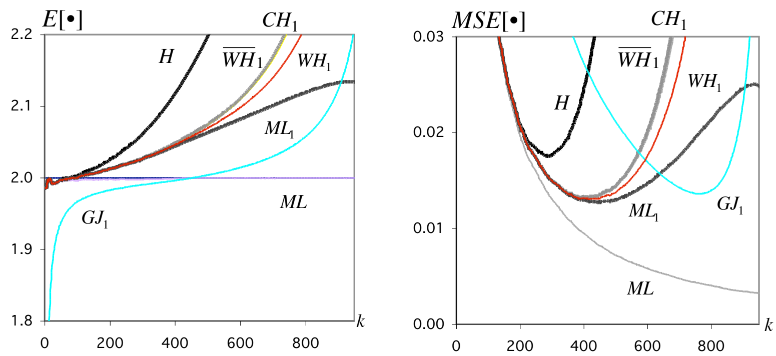

Figure 4: Underlying GP parent with $\gamma = 2$ $(\rho = -2)$.

but merely because the associated sample paths cross the target at a lower threshold. The $\overline{WH}$-statistic is however the one with the smallest bias for not too large values of k.

- *For values of $|\rho| < 1$ (Figure 3), the WH-statistic behaves slightly better than the CH-statistic, which behaves slightly better than the $\overline{WH}$-statistic, but the ML-statistic outperforms all UH-estimators. The behaviour of the GJ-statistic is quite appealing for these ρ-values.*
- *For $|\rho| > 1$ (Figure 4), we need to use $\hat{\rho}_1$ (instead of $\hat{\rho}_0$) or any of the hybrid estimators suggested in* Gomes and Pestana (2007*a*). *In all the simulated cases, the CH and the $\overline{WH}$-statistics almost overlap and the ML-statistics again outperform all other statistics. The GJ statistic, although exhibiting a better performance than the Hill estimator, when both statistics are compared at their optimal levels, is not able to overpass the UH-statistics.*

Further simulation comparisons can be found in Caeiro *et al.* (2005) and Gomes *et al.* (2007a; 2008).

5 Two case-studies and overall conclusions

When analysing heavy-tailed data, quite common in financial time series, one never knows how much the underlying model differs from a strict Pareto model. And this is the unique situation where the Hill estimator is “perfect”. Otherwise, all depends on the specificity of the underlying heavy-tailed model and on the practitioner’s objectives. If we want to use (or have only access to) a very small number of top o.s. from a heavy-tailed model, the Hill estimator has been considered the most adequate one. This is no longer true: the new estimators are similar to Hill’s estimator from small up to moderate values of k, being much better than the Hill estimator, when we consider larger values of k, although intermediate. After taking a decision on the estimate of ρ, and assuming that $|\rho| \leq 1$, a situation which seems to appear often in practice, we should simultaneously picture the sample path of a few tail index estimators, with different specificities. On the basis of those sample paths we may then get, in a more appropriate way, an accurate estimate of the tail index γ, like we shall see later on, in the applications provided. And the new estimators under study in this paper should be taken as the adequate substitutes of Hill’s estimator.

5.1 Log-exchange rates of Euro against USA Dollar

We shall first consider an illustration of the performance of the above mentioned estimators, through the analysis of the positive log-returns $P_i = \ln(S_{i+1}/S_i) = -L_i$, $1 \leq i \leq n-1$, with S_i, $1 \leq i \leq n$, the Euro-USA Dollar daily exchange rates from January 4, 1999 until December 15, 2004. In Figure 5, working with the $n_0 = 748$ positive log-returns, we present the sample path of the $\hat{\rho}_\tau(k) = \hat{\rho}(k;\tau)$ estimates in (2.1) (*left*), as function of k, for $\tau = 0$ and $\tau = 1$. Note that 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 k, the estimate associated to $\tau = 0$. From previous experience with this type of estimates, we conclude that the underlying ρ-value is larger than or equal to -1, and the consideration of $\tau = 0$ is then suitable, although negative values of τ might be even better, as illustrated in Caeiro and Gomes (2008) for a GP underlying model. The estimate of ρ is in this case $\hat{\rho}_0 = -0.69$. In this same figure (*right*), we also present the sample paths of the classical Hill estimator, H, of $UH_0 = UH_{\hat{\beta}_{01},\hat{\rho}_0}$, with $UH = CH, ML$ and WH, as well as of $GJ_0 = GJ_{\hat{\rho}_0}$.

Regarding the tail index estimation, note that the Hill estimator exhibits here a relevant bias.

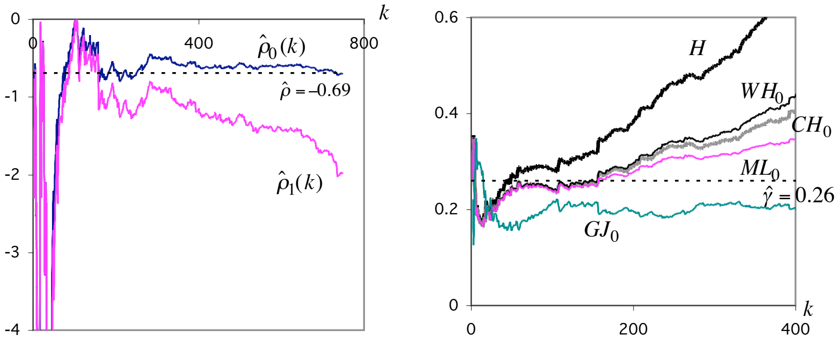

Figure 5: Estimates of the second order parameter ρ (*left*) and of the tail index γ (*right*) for the Daily Log-Returns of the Euro-USA Dollar.

We are thus a long way from a strict Pareto underlying model. The other estimators, which are "asymptotically unbiased" up to the second order, reveal a smaller bias, and enable us to take a decision upon the estimate of γ to be used, with the help of any stability criterion or any heuristic procedure, like a largest run method, of the type of the one described in the sequel, and already suggested in Gomes *et al.* (2005a): let us consider a set of reduced-bias tail index estimates $\hat{\gamma}_i(k)$, $1 \le k < n/2$, $i \in \mathcal{I}$, with a small number r of decimal figures. Let us denote them $\hat{\gamma}_{i|r}(k)$. Then, for any value $i \in \mathcal{I}$ and for any possible value a in the domain of $\hat{\gamma}_{i|r}(k)$, consider the largest run associated with a, i.e., $R_i(a)$, the maximum number of consecutive k values such that $\hat{\gamma}_{i|r}(k) = a$. Next, compute $a_i^M := \arg\max_a R_i(a)$, and consider as a data-driven estimate of the tail index γ, $\hat{\gamma} = a_{i_0}^M$ with $i_0 := \arg\max_i a_i^M$. Here, if we consider $r = 1$, the largest run is achieved by the sample path of the ML-estimator in (1.11) and the WH-estimator in (1.12). Such a largest run has a size equal to 320, for k between 60 and 379, and is associated to $\hat{\gamma} = 0.3$. Working with the CH-estimator and this criterion, would we also be led to an estimate equal to 0.3, with a run of size 266 ($68 \le k \le 333$). For the $\overline{WH}$-estimate, we would also get a tail index estimate equal to 0.3, with a run of size 235 ($60 \le k \le 294$). With this same criterion, the Hill estimator would provide an estimate also equal to 0.3, with a run of size 126 ($43 \le k \le 168$). According to the previous heuristic procedure we would thus be led to the choice of the ML or WH estimators, computed at any level from $k = 60$ until $k = 379$, all providing the same estimate $\hat{\gamma} = 0.3$. Should we consider this same criterion, but the estimates with two decimal figures, would we be led to the estimate $\hat{\gamma} = 0.26$ for any of the reduced-bias estimators. This is the value pictured in Figure 5.

5.2 Nasdaq Composite index

We next proceed to the data analysis of positive log-returns $P_i = \ln(S_{i+1}/S_i)$, $1 \le i \le n-1$, with S_i, $1 \le i \le n$, the close prices of Nasdaq Composite Index. We have used the daily log-returns from 1997 until 2000, which correspond to a sample of size $n = 1037$, with $n_0 = 750$ positive values.

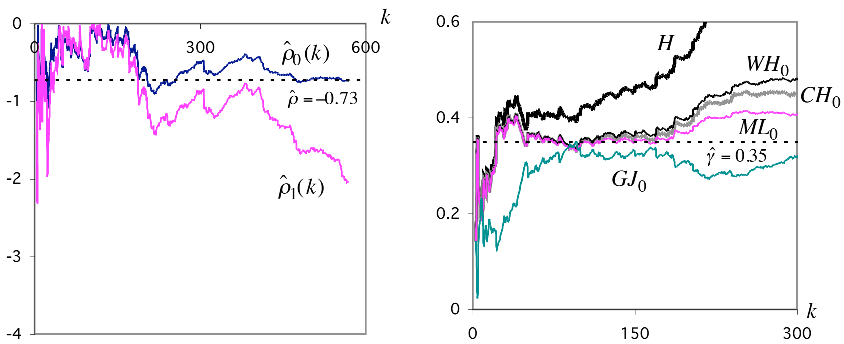

Figure 6: Estimates of the second order parameter ρ (*left*) and of the tail index γ (*right*) for the Daily Log-Returns of the Nasdaq Composite Index.

5.3 Some overall conclusions

- Generally, we may say that there is not any significant difference between the different UH-estimators. Anyway, whenever confronted with real data, the drawing of sample paths of a few alternative estimates can help us in the choice of the most accurate tail index estimate.
- The generalized Jacknife statistic in (1.13) exhibits, for some of the models, sample paths more stable around the target value γ for a wider region of k-values, comparatively to the statistics here studied, but at the expenses of mean squared errors much higher than those of the Hill and the new MVRB statistics, for small up to moderate values of k.
- Indeed, the main advantage of the so-called MVRB estimators $\overline{WH}$, CH, ML and WH in (1.9), (1.10), (1.11) and (1.12), respectively, 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 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 previous alternatives available in the literature, like the

generalized Jackknife estimator in (1.13).

- Despite of this, it is sensible to understand the comparative behaviour at optimal levels of these MVRB estimators and the other reduced-bias estimators, not only for finite samples, but also asymptotically. It is thus crucial to have information on the order of the dominant component(s) of their asymptotic bias, the main contribution in this paper, for the MVRB tail index estimators in (1.10), (1.11) and (1.12). The adaptive choice of the threshold is now potentially feasible, but out of the scope of this paper.

6 Proofs

Let us further introduce the notations,

$$P_k^{(\alpha)} \equiv P_k^{(\alpha)}(\rho) := \frac{1}{k}\sum_{i=1}^{k} \psi_{ik}^{\alpha}(\rho)\ E_{k-i+1:k},\ \alpha \ge 0, \tag{6.1}$$

$$Q_k^{(\alpha)} \equiv Q_k^{(\alpha)}(\rho) := \frac{1}{k}\sum_{i=1}^{k} \psi_{ik}^{\alpha-1}(\rho)\ \big(\exp(\rho E_{k-i+1:k}) - 1\big)/\rho,\ \alpha \ge 1, \tag{6.2}$$

$$R_k^{(\alpha)} \equiv R_k^{(\alpha)}(\rho) := \frac{1}{k}\sum_{i=1}^{k} \psi_{ik}^{\alpha}(\rho)\ \ln(i/k)\ \ E_{k-i+1:k},\ \alpha \ge 0, \tag{6.3}$$

and

$$\overline{W}_k^{(\alpha)} := \frac{(2\alpha-1)\sqrt{(2\alpha-1)\ k}}{\sqrt{2}}\ \ \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1}\ln\left(\frac{i}{k}\right)E_i + \frac{1}{\alpha^2}\right),\ \alpha \ge 1, \tag{6.4}$$

with ψ given in (1.9), being $\{E_i\}$, again, a sequence of i.i.d. standard exponential r.v.'s and $E_{i:n}$, $1 \le i \le n$, the associated ascending o.s. As usual, let us denote $\mathbb{E}$ the mean value operator. More generally than in Lemma 2 of Gomes *et al.* (2008), but with the same type of proof, we may state:

Lemma 6.1. *For intermediate k, i.e., whenever* (1.7) *holds, and with $P_k^{(\alpha)}$ and $Q_k^{(\alpha)}$ in* (6.1) *and* (6.2)*, respectively, for any real $\alpha \ge 1$, both $\mathbb{E}\big(P_k^{(\alpha)}\big)$ and $\mathbb{E}\big(Q_k^{(\alpha)}\big)$ converge to*

$$a_\alpha \equiv a_\alpha(\rho) = -\int_0^1 \psi_\rho^\alpha(v)\ lnv\ dv < \infty, \tag{6.5}$$

with $a_1 = 1/(1-\rho)$. For $0 \le \alpha < 1$, $\mathbb{E}\big(P_k^{(\alpha)}\big)$ also converges to a_α in (6.5)*, with $a_0 \equiv 1$. Moreover, $\sigma_0^2 := k\mathbb{V}ar\big(P_k^{(0)}\big) = 1$, and for any $\alpha > 0$*

$$\sigma_\alpha^2 \equiv \sigma_\alpha^2(\rho) = \lim_{n\to\infty} k\ \mathbb{V}ar\big(P_k^{(\alpha)}\big) = \frac{2}{\rho^{2\alpha}} \iint\limits_{0\le u<v\le 1} \left(\frac{u^{-\rho}-1}{\ln u}\ \frac{v^{-\rho}-1}{\ln v}\right)^{\alpha} \frac{1-v}{v}\ du\ dv < \infty. \tag{6.6}$$

Consequently, for $\alpha \geq 0$, $P_k^{(\alpha)}$ *in* (6.1) *converges in probability to* a_α, *as* $k \to \infty$, *with* $a_0 = 1$ *and* a_α, $\alpha > 0$ *given in* (6.5). *The sequence* $R_k^{(1)}$, *with* $R_k^{(\alpha)}$ *given in* (6.3), *converges in probability to* $b_1 = -\int_0^1 \psi(v)\ ln^2 v\ dv = -(2-\rho)/(1-\rho)^2$.

We can also easily derive the results in the following lemma:

Lemma 6.2. *Under the third order framework in* (1.3), *for levels* k *such that* (1.7) *holds, with* U_i *given in* (1.5), *and for any* $\alpha \geq 1$, *the distributional representations*

$$\frac{1}{k}\sum_{i=1}^{k}\Big(\frac{i}{k}\Big)^{\alpha-1} U_i \ \stackrel{d}{=}\ \frac{\gamma}{\alpha} + \frac{\gamma\ \overline{Z}_k^{(\alpha)}}{\sqrt{(2\alpha-1)\ k}} + \frac{A(n/k)}{\alpha-\rho} + O_p\Big(\frac{A(n/k)}{\sqrt{k}}\Big) + \frac{A(n/k)\ B(n/k)}{\alpha-\rho-\rho'}\ (1+o_p(1)) \tag{6.7}$$

and

$$\frac{1}{k}\sum_{i=1}^{k}\Big(\frac{i}{k}\Big)^{\alpha-1} \ln\Big(\frac{i}{k}\Big) U_i \ \stackrel{d}{=}\ -\frac{\gamma}{\alpha^2} - \frac{\gamma\ \overline{W}_k^{(\alpha)}}{(2\alpha-1)\sqrt{(2\alpha-1)\ k/2}} - \frac{A(n/k)}{(\alpha-\rho)^2} + O_p\Big(\frac{A(n/k)}{\sqrt{k}}\Big) - \frac{A(n/k)\ B(n/k)}{(\alpha-\rho-\rho')^2}\ (1+o_p(1)) \tag{6.8}$$

hold, where $\overline{Z}_k^{(\alpha)}$ *and* $\overline{W}_k^{(\alpha)}$ *in* (2.11) *and* (6.4), *respectively, are asymptotically standard normal.*

Similarly, now with V_{ik} *and* ψ_{ik} *given in* (1.5) *and* (1.9), *respectively, the representations*

$$\frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{\alpha}(\rho)\ V_{ik} \ \stackrel{d}{=}\ \gamma\ a_\alpha(\rho) + \frac{\gamma\ \sigma_\alpha(\rho)\ \overline{P}_k^{(\alpha)}(\rho)}{\sqrt{k}} + a_{\alpha+1}(\rho)\ A(n/k) + O_p\Big(\frac{A(n/k)}{\sqrt{k}}\Big) + b_{\alpha+1}(\rho,\rho')\ A(n/k)B(n/k)(1+o_p(1)) \tag{6.9}$$

hold for any $\alpha \geq 0$, *with* a_α *given in* (6.5), σ_α^2 *given in* (6.6) *for* $\alpha > 0$, $\sigma_0^2 = 1$, *and*

$$b_\alpha(\rho,\rho') = \lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{\alpha-1}(\rho)\ \mathbb{E}\Big[\frac{Y_{k-i+1:k}^{\rho+\rho'}-1}{\rho+\rho'}\Big] = -\int_0^1 \psi_\rho^{\alpha-1}(v)\ \psi_{\rho+\rho'}(v)\ln v\ dv < \infty. \tag{6.10}$$

Consequently, and with $P_k^{(\alpha)}(\rho)$ *defined in* (6.1), $\overline{P}_k^{(\alpha)}(\rho) := \sqrt{k}\big(P_k^{(\alpha)}(\rho) - a_\alpha(\rho)\big)/\sigma_\alpha(\rho)$ *are asymptotically standard normal r.v.'s.*

Proof. The summand $O_p\big(A(n/k)/\sqrt{k}\big)$ is due to the replacement of $A(Y_{n-k:n})$ by $A(n/k)$, together with the fact that $Y_{n-k:n} - n/k = O_p\big(n/\big(k\sqrt{k}\big)\big)$ and $A'(t) = O\left(A(t)/t\right)$, as $t \to \infty$. The remaining of the lemma comes from the third order set-up in (1.3) together with the same kind of reasoning as in Gomes and Martins (2004), Gomes *et al.* (2005a; 2007a) and Caeiro *et al.* (2005). ■

Before the proof of Theorem 2.1, we prove the following lemma, related with *Condition U*:

Lemma 6.3. *Under the third order framework in* (1.3)*, if we further assume the validity of Condition U for a level* $k_1 = O(n/\ln\ln n)$ *and consider values* k *such that* $\sqrt{k}\ A^2(n/k) \to \lambda_{_A}$*, finite,*

$$\sqrt{k}\ A(n/k)\ln(n/k)\,(\hat\rho_{_U} - \rho) = o_p(1), \quad \text{as } n \to \infty. \tag{6.11}$$

Proof. Under the validity of *Condition U*, for a level $k_1 = O(n/\ln\ln n)$, we may guarantee that $\hat\rho_{_U} - \rho = O_p\left(1/\left(\sqrt{k_1}\ A(n/k_1)\right)\right) = O_p\left((\ln\ln n)^{(1-2\rho)/2}/\sqrt{n}\right)$. Consequently, condition (2.7) holds with $\hat\rho$ replaced by $\hat\rho_{_U}$. *A fortiori* we have $(\hat\rho_{_U} - \rho)\ln(n/k) = o_p(1)$, as $n \to \infty$, and (6.11) holds whenever $\sqrt{k}\ A(n/k) \to \lambda$, finite. Next, if $\sqrt{k}\ A(n/k) \to \infty$, k is of a larger order than $n^{-2\rho/(1-2\rho)}$, and n/k is at most of the order of $n^{1/(1-2\rho)}$. Consequently, $\ln(n/k)/A(n/k)$ is at most of the order of $(\ln n)n^{-\rho/(1-2\rho)}$, and

$$0 \le |(\hat\rho - \rho)\ln(n/k)/A(n/k)| < O\Big((\ln\ln n)^{(1-2\rho)/2}\,\frac{\ln n}{n^{1/(2(1-2\rho))}}\Big) \underset{n\to\infty}{\longrightarrow} 0.$$

If $\sqrt{k}\ A^2(n/k) \to \lambda_{_A}$, finite, (6.11) follows for this type of levels. ∎

Proof. (**Theorem 2.1**). The distributional representation in (2.9) comes directly from (6.7). Indeed, if we write $\hat\beta(k;\rho) =: (k/n)^\rho \times (\varphi(\rho)/\psi(\rho)) = (\gamma\ \beta/A(n/k)) \times (\varphi(\rho)/\psi(\rho))$, we get

$$\varphi(\rho) \overset{d}{=} \frac{\gamma}{\sqrt{k}}\Big(\frac{\overline{Z}_k^{(1)}}{1-\rho} - \frac{\overline{Z}_k^{(1-\rho)}}{\sqrt{1-2\rho}}\Big) - \frac{\rho^2 A(n/k)}{(1-\rho)^2(1-2\rho)} + O_p\big(A(n/k)/\sqrt{k}\big)$$
$$- \frac{\rho(\rho+\rho')A(n/k)\ B(n/k)}{(1-\rho)(1-\rho-\rho')(1-2\rho-\rho')}(1+o_p(1)) + O_p(1/k),$$

and

$$\psi(\rho) \overset{d}{=} -\frac{\gamma\ \rho^2}{(1-\rho)^2(1-2\rho)} + \frac{\gamma}{\sqrt{k}}\Big(\frac{\overline{Z}_k^{(1-\rho)}}{(1-\rho)\sqrt{1-2\rho}} - \frac{\overline{Z}_k^{(1-2\rho)}}{\sqrt{1-4\rho}}\Big) + O_p\big(A(n/k)/\sqrt{k}\big)$$
$$- \frac{2\rho^2 A(n/k)}{(1-\rho)(1-2\rho)(1-3\rho)} - \frac{\rho(2\rho+\rho')A(n/k)B(n/k)}{(1-\rho)(1-2\rho-\rho')(1-3\rho-\rho')}(1+o_p(1)) + O_p(1/k).$$

Consequently, if $\sqrt{k}\ A(n/k) \to \infty$, i.e., $1/\sqrt{k} = o(A(n/k))$,

$$\frac{1}{\psi(\rho)} = -\frac{(1-\rho)^2(1-2\rho)}{\gamma\ \rho^2}\Big(1 - \frac{2(1-\rho)A(n/k)}{\gamma(1-3\rho)} + o_p(A(n/k))\Big),$$

and with W_k^B and $(u_{_B}, v_{_B})$ given in (2.10) and (2.12), respectively,

$$\frac{\varphi(\rho)}{\psi(\rho)} \overset{p}{\sim} \frac{A(n/k)}{\gamma} + \frac{(1-\rho)\sqrt{1-2\rho}}{\rho\sqrt{k}} W_k^B + \frac{A(n/k)}{\gamma}\Big(u_{_B} A(n/k) + v_{_B} B(n/k)\Big),$$

and

$$\hat\beta(k;\rho) \overset{p}{\sim} \beta\Big(1 + \frac{\gamma\,(1-\rho)\sqrt{1-2\rho}}{\rho\sqrt{k}\;A(n/k)} W_k^B + u_{_B} A(n/k) + v_{_B} B(n/k)\Big).$$

Since the asymptotic covariance between $\overline{Z}_k^{(1)}$ and $\overline{Z}_k^{(1-\rho)}$ is given by $\sqrt{1-2\rho}/(1-\rho)$, W_k^B is asymptotically standard normal and (2.9), (2.12), (2.13) and (2.14) follow.

The remaining of the theorem comes from the fact that, since $d\hat\beta(k;\rho)/d\rho = -\ln(n/k)\hat\beta(k;\rho)(1 + o_p(1))$, we can write

$$\hat\beta(k;\hat\rho) = \hat\beta(k;\rho) - \hat\beta(k;\rho)\,(\hat\rho-\rho)\ln(n/k)(1+o_p(1)) = \beta + \frac{\gamma\,\beta(1-\rho)\sqrt{1-2\rho}}{\rho\sqrt{k}\;A(n/k)} W_k^B$$
$$+ \beta\; u_{_B} A(n/k) + \beta\; v_{_B} B(n/k) - \beta\;\,(\hat\rho-\rho)\ln(n/k)(1+o_p(1)), \quad (6.12)$$

provided that $(\hat\rho - \rho)\ln(n/k) = o_p(1)$. In order to keep the results in the theorem, we need to work with a level k such that $\sqrt{k}\;A(n/k) \to \infty$, $\sqrt{k}\;A^2(n/k) \to \lambda_{_A}$, $\sqrt{k}\;A(n/k)B(n/k) \to \lambda_{_B}$, both finite, and $\sqrt{k}\;A(n/k)\ln(n/k)\,(\hat\rho-\rho) = o_p(1)$, as $n \to \infty$. This holds if we assume the validity of *Condition U* for a level $k_1 = O(n/\ln\ln n)$ and work with $\hat\rho_{_U}$, as stated and proved in Lemma 6.3.

If we estimate ρ through $\hat\rho(k;\tau)$, in (2.1), $(\hat\rho(k;\tau)-\rho)\ln(n/k)$ is the dominant term among the terms in (6.12), dependent on k. Then the behaviour of $\hat\beta(k;\hat\rho(k;\tau))$ is related with the behaviour of $\hat\rho(k;\tau)-\rho$, as stated in (2.15). If we estimate ρ through $\hat\rho(k;\tau)$, we get, from Proposition 2.1,

$$\frac{\hat\beta(k;\hat\rho(k;\tau)) - \beta}{-\beta\ln\left(\frac{n}{k}\right)} \overset{d}{=} \frac{\sigma_{_R} W_k^R}{\sqrt{k}\;A(n/k)} + (u_{_R} A(n/k) + v_{_R} B(n/k))\,(1+o_p(1)).$$

Consequently, we need to have $\ln(n/k)/\big(\sqrt{k}\;\,A(n/k)\big) \to 0$, in order to be able to guarantee that $\hat\beta(k;\hat\rho(k;\tau))$ is consistent for the estimation of β. Then, if $\sqrt{k}\;A^2(n/k) \to \lambda_{_A}$ and $\sqrt{k}\;A(n/k)B(n/k) \to \lambda_{_B}$, both finite, $\sqrt{k}\;A(n/k)\big(\hat\beta(k;\hat\rho(k;\tau)) - \beta\big)/(\beta\ln(n/k))$ is asymptotically normal, with variance $\sigma_{_R}^2$, given in (2.2), and a possibly non-null asymptotic bias given by $\{-\,(\lambda_{_A} u_{_R} + \lambda_{_B} v_{_R})\}$, with $u_{_R}$ and $v_{_R}$ given in (2.5) and (2.6), respectively. ■

We shall now proceed with the proofs of the theorems in section 3.

Proof. (**Theorem 3.1**). Regarding the estimator in (1.10), since (6.7) holds, with $\alpha = 1$, jointly with (1.8), and $CH_{\beta,\rho}(k) = H_n(k) \times \big(1 - A(n/k)/(\gamma(1-\rho))\big)$, we get

$$CH_{\beta,\rho}(k) \;\stackrel{d}{=}\; \gamma + \frac{\gamma}{\sqrt{k}}\, \overline{Z}_k^{(1)} + A(n/k)\Big(\frac{B(n/k)}{1-\rho-\rho'} - \frac{A(n/k)}{\gamma(1-\rho)^2} + O_p\Big(\frac{1}{\sqrt{k}}\Big)\Big)(1+o_p(1)).$$

For the estimator in (1.11), the use of (6.7), for $\alpha = 1$ and $\alpha = 1-\rho$, enables us to get,

$$ML_{\beta,\rho}(k) \;\stackrel{d}{=}\; \gamma + \frac{\gamma}{\sqrt{k}}\, \overline{Z}_k^{(1)} + A(n/k)\Big(\frac{B(n/k)}{1-\rho-\rho'} - \frac{A(n/k)}{\gamma(1-2\rho)}\Big)(1+o_p(1)) \\ + O_p\big(A(n/k)/\sqrt{k}\big)(1+o(1)).$$

Finally, for the estimator in (1.12), the use of (1.8) and (6.9) for $\alpha = 1$, enables us to write,

$$WH_{\beta,\rho}(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, \overline{Z}_k^{(1)} - A(n/k)\Big(\frac{a_2(\rho)A(n/k)}{\gamma} - \frac{B(n/k)}{1-\rho-\rho'} + O_p\Big(\frac{1}{\sqrt{k}}\Big)\Big)(1+o_p(1)).$$

The integral $a_2(\rho)$, with $a_\alpha(\rho)$ defined in (6.5), can be easily computed if we make the substitution $-\ln v = t$ and use (5.1.32) from Abramowitz and Stegun's handbook, providing the value in (3.3). Consequently, (3.2) holds for any $UH_{\beta,\rho}(k)$, with u_{UH} and v_{UH} given in (3.4). Note that since $\sqrt{k}\; O_p\big(A(n/k)/\sqrt{k}\big) = O_p(A(n/k)) \to 0$, as $n \to \infty$, the summands $O_p\big(A(n/k)/\sqrt{k}\big)$ are totally irrelevant for the asymptotic bias, that follow straightforwardly from representations above. ■

Proof. (**Theorem 3.2**). If we estimate consistently β and ρ through the estimators $\hat\beta$ and $\hat\rho$, we may use Cramer's delta-method, and obtain for any of the estimators in (1.10), (1.11) and (1.12), generically denoted UH, and with $a_{UH} = -1/(1-\rho)$, independent of UH,

$$UH_{\hat\beta,\hat\rho}(k) - UH_{\beta,\rho}(k) \;\stackrel{p}{\sim}\; a_{UH} A(n/k)\Big\{\Big(\frac{\hat\beta-\beta}{\beta}\Big) + (\hat\rho-\rho)\,[\ln(n/k) \; - a_{UH}]\Big\}. \tag{6.13}$$

Indeed, directly for the $\overline{H}$-estimator, on the basis of Lemma 6.1 for the WH-estimator, and on the basis of (6.7) and (6.8) for the ML-estimator, we can write,

$$\frac{\partial UH_{\beta,\rho}}{\partial\beta} \;\stackrel{p}{\sim}\; -\frac{A(n/k)}{\beta(1-\rho)}, \qquad \frac{\partial UH_{\beta,\rho}}{\partial\rho} \;\stackrel{p}{\sim}\; -A(n/k)\Big(\frac{1}{1-\rho}\ln(n/k) + \frac{1}{(1-\rho)^2}\Big).$$

The first part of the theorem, related with levels k such that $\sqrt{k}\; A(n/k) \to \lambda$, finite, follows thus straightforwardly from (6.13).

Next, since (2.15) holds, i.e., $(\hat\beta-\beta)/\beta \;\stackrel{p}{\sim}\; -\ln(n/k_1)\,(\hat\rho-\rho)$, we have

$$UH_{\hat\beta,\hat\rho}(k) - UH_{\beta,\rho}(k) \;\stackrel{p}{\sim}\; a_{UH}\,(\hat\rho-\rho)\,A(n/k)\,(-\ln(k/k_1) - a_{UH}) =: W_{k,k_1}. \tag{6.14}$$

Under the conditions in the theorem, i.e., with k_1 optimal for the ρ-estimation, $\hat\rho - \rho = O_p\big(1/\big(\sqrt{k_1}\ A(n/k_1)\big)\big)$,

$$\sqrt{k}\ W_{k,k_1} = O_p\Big(\frac{\sqrt{k}\ A(n/k)}{\sqrt{k_1}\ A(n/k_1)} \ln\Big(\frac{k}{k_1}\Big)\Big) = O_p\Big(\Big(\frac{k}{k_1}\Big)^{\frac{1}{2}-\rho} \ln\Big(\frac{k}{k_1}\Big)\Big) = o_p(1) \quad \text{if} \quad k/k_1 \to 0,$$

and the second part of the theorem follows.

If we further assume the validity of *Condition U* for a level $k_1 = O(n/\ln\ln n)$ and consider $UH_{\hat\beta_U,\hat\rho_U}$, we guarantee, on the basis of Lemma 6.3, that $\sqrt{k}\,(\hat\rho_U - \rho)\,A(n/k)\ln(k/k_1) = o_p(1)$. Consequently, the use of (6.14), with $(\hat\beta,\hat\rho)$ replaced by $(\hat\beta_U,\hat\rho_U)$ or such that $\hat\rho-\rho = o_p(\ln(n/k)/(\sqrt{k}A(n/k))$, enables us to prove the results in the theorem. ■

Proof. (**Theorem 3.3**). If we consider any of the reduced-bias estimators $UH^*_\rho(k) = UH_{\beta(k;\rho),\rho}(k)$, we get, under the validity of (1.3), with (u_B, v_B) and (u_{UH}, v_{UH}) given in (2.12) and (3.4), respectively,

$$\begin{aligned} UH^*_\rho(k) = \gamma + \frac{\gamma}{\sqrt{k}}\Big(\overline{Z}_k^{(1)} + \frac{(1-\rho)(1-2\rho)}{\rho^2}\Big(\frac{\overline{Z}_k^{(1)}}{1-\rho} - \frac{\overline{Z}_k^{(1-\rho)}}{\sqrt{1-2\rho}}\Big)\Big) \\ + A(n/k)\Big(A(n/k)\big(u_{UH} - u_B/(1-\rho)\big) + B(n/k)\big(v_{UH} - v_B/(1-\rho)\big)\Big)(1+o_p(1)). \end{aligned}$$

Under the conditions in the theorem, we thus get, for any $UH^*_\rho(k)$, the asymptotic variance in (3.6) and the asymptotic bias $\lambda_A u^*_{UH} + \lambda_B v^*_{UH}$, in (3.7).

Let us think now on $UH^*_{\hat\rho}(k) = UH_{\hat\beta(k;\hat\rho),\hat\rho}(k) =: H(k) - Bias_{UH}(\hat\rho)$. Since $\partial Bias_{UH}(\rho)/\partial\rho = O_p(A(n/k))$, whenever $\sqrt{k}\ A(n/k) \to \infty$, the use of the delta-method enables us to guarantee that

$$UH^*_{\hat\rho}(k) - UH^*_\rho(k) = O_p\big(A(n/k)\,(\hat\rho - \rho)\,\big).$$

Consequently, if we choose an optimal k_1 for the estimation of ρ,

$$\begin{aligned} \sqrt{k}\,\big(UH^*_{\hat\rho}(k) - UH^*_\rho(k)\big) &= O_p\Big(\sqrt{k}\ A(n/k)\,(\hat\rho-\rho)\,\Big) = O_p\Big(\frac{\sqrt{k}\ A(n/k)}{\sqrt{k_1}\ A(n/k_1)}\Big) \\ &= o_p(1), \quad \text{provided that } k = o(k_1). \end{aligned}$$

Finally, from Lemma 6.3, we may guarantee that if we assume the validity of *Condition U* for a level $k_1 = O(n/\ln\ln n)$, the results in the theorem hold. ■

Proof. (**Theorem 3.4**). If we consider $k = k_1$ in (6.14), we may write

$$\begin{aligned} \sqrt{k}\Big(UH^{**}(k) - \gamma\Big) = \gamma\ \overline{Z}_k^{(1)} + \sqrt{k}A(n/k)\,(u_{UH}A(n/k) + v_{UH}\ B(n/k))\,(1+o_p(1)) \\ - a^2_{UH}\Big(\sigma_R W_k^R + \sqrt{k}\ A(n/k)\,\Big(u_R A(n/k) + v_R B(n/k)\Big)(1+o_p(1))\Big), \end{aligned}$$

with σ_R, W_k^R, u_R, v_R and (u_{UH}, v_{UH}) given in (2.2), (2.3), (2.5), (2.6) and (3.4), respectively. Then

$$\sqrt{k}\Big(UH^{**}(k)-\gamma\Big) = \gamma\ \overline{Z}_k^{(1)} - \sigma_R\ a_{UH}^2 W_k^R + \sqrt{k}\ A^2(n/k)\left(u_{UH} - u_R\ a_{UH}^2\right)(1+o_p(1))$$
$$+\sqrt{k}\ A(n/k)B(n/k)\left(v_{UH} - v_R\ a_{UH}^2\right)(1+o_p(1))$$

We now need to compute the variance of $\{\gamma\ \overline{Z}_k^{(1)} - \sigma_R\ a_{UH}^2 W_k^R\}$. Since $Cov\big(\overline{Z}_k^{(1)},\ W_k^R\big) = 0$, we get an asymptotic variance equal to $\gamma^2 + \sigma_R^2 a_{UH}^4$, the value σ_3^2 in (3.8). If $\sqrt{k}\ A^2(n/k) \to \lambda_A$ and $\sqrt{k}\ A(n/k)\ B(n/k) \to \lambda_B$, both finite, the asymptotic bias in (3.9) follows.

■

References

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

[2] Beirlant, J.; Dierckx, G.; Guillou, A.; Starica, C. (2002). On exponential representations of log-spacings of extreme order statistics. *Extremes* **5**: 2, 157-180.

[3] Caeiro, F.; Gomes, M.I. (2002). A class of asymptotically unbiased semi-parametric estimators of the tail index. *Test* **11**:2, 345-364.

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

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

[6] Caeiro, F.; Gomes, M.I.; Pestana, D.D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat* **3** (2), 111-136.

[7] Diebolt, J.; Guillou, A. (2005). Asymptotic behaviour of regular estimators. *Revstat* **3** (1), 19-44.

[8] Drees, H. (1998). A general class of estimators of the tail index. *J. Statist.Plann.Inference* **98**, 95-112.

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

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

[11] Fisher, R.A.; 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.

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

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

[14] Gomes, M.I. (1984). Penultimate limiting forms in extreme value theory. *Ann. Inst. Statist. Math.* **A36**, 71-85.

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

[16] Gomes, M.I.; Figueiredo, F.; 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.

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

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

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

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

[21] Gomes, M.I.; 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.

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

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

[24] Gomes, M.I.; Martins, M.J.; Neves, M.M. (2002b). Generalized Jackknife semi-parametric estimators of the tail index. *Portugaliae Mathematica* **59**:4, 393-408.

[25] Gomes, M.I.; Martins, M.J.; Neves, M.M. (2007a). Improving second order reduced-bias tail index estimation. *Revstat* **5**(2), 177-207.

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

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

[28] Gomes, M.I.; Pestana, D.D. (2007a). A simple second order reduced-bias tail index estimator. *J. Statist. Comp. and Simul.* **77**, nº 6, 487-504.

[29] Gomes, M.I.; Pestana, D.D. (2007b). A sturdy reduced bias extreme quantile (VaR) estimator. *J. Amer. Statist. Assoc.* **102**, No. 477, 280-292.

[30] Gomes, M.I.; Pestana, D.D. (2007c). A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator. *Bull. of the Internat. Stat. Institute* **LXII** (electronic version).

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

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

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

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

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

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

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