A new class of estimators of a “scale” second order parameter*

Frederico Caeiro

C.M.A. and Departamento de Matemática, F.C.T.U.N.L.

and

M. Ivette Gomes

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

Abstract. For a large class of heavy-tailed distribution functions F in the domain of attraction for maxima of an *Extreme Value* distribution with tail index $\gamma > 0$, the function $A(t)$, controlling the speed of convergence of maximum values, linearly normalized, towards a non-degenerate limiting random variable, may be parameterized as $A(t) = \gamma\ \beta\ t^\rho$, $\rho < 0$, $\beta \in \mathbb{R}$, where β and ρ are second order parameters. The estimation of ρ, the “shape” second order parameter has been extensively addressed in the literature, but practically nothing has been done related to the estimation of the “scale” second order parameter β. In this paper, and motivated by the importance of a reliable β-estimation in recent reduced bias tail index estimators, we shall deal with such a topic. Under a semi-parametric framework, we introduce a class of β-estimators and study their consistency. We deal with the conditions enabling us to get the asymptotic normality of the members of this class, and we illustrate the behaviour of the estimators, through Monte Carlo simulation techniques.

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

Keywords and phrases. *Statistics of extremes*, *semi-parametric estimation*, *third order framework*.

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

1 Introduction and preliminaries

Heavy-tailed models are quite useful in most diversified areas, ranging from telecommunication traffic till financial data analysis. A model F is *heavy-tailed* whenever the *tail function*, $\overline{F} := 1 - F$, is a regularly varying function with a negative index of regular variation equal to $-1/\gamma$, $\gamma > 0$, or equivalently, whenever 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., for every $x > 0$,

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

Then we are in the domain of attraction for maxima of an *Extreme Value* distribution function (d.f.),

$$EV_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}),\ x \geq -1/\gamma,\ \gamma > 0,$$

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

The second order parameter ρ (≤ 0) rules the rate of convergence in the first order condition in (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 $\rho < 0$, together with a third order assumption, required to achieve non-degenerate asymptotic properties of the estimators, and ruling now the rate of convergence in (1.2), with a third order parameter $\eta < 0$. To be more precise, we shall assume that

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

for every $x > 0$, and where $|B(t)|$ must then be of regular variation with index η (Fraga Alves *et al.*, 2003b).

Remark 1.1. *Sometimes condition* (1.3) *also appears written as*

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

for every $x > 0$*, and as* $t \longrightarrow \infty$*.*

For sake of simplicity, we shall assume here that $\eta = \rho < 0$. We shall further assume that the A function in (1.2) may be parameterised as

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

begin $|B(t)|$ also regularly varying and polynomial in ρ. We are thus working in a large sub-class of Hall-Welsh class of models (Hall and Welsh, 1985), with a tail function of the type,

$$1 - F(x) = \left(\frac{x}{C}\right)^{-1/\gamma} \left(1 + \frac{\beta}{\rho} \left(\frac{x}{C}\right)^{\rho/\gamma} + \beta' \left(\frac{x}{C}\right)^{2\rho/\gamma} (1 + o(1)) \right), \tag{1.5}$$

with $\gamma, \ C > 0, \ \rho < 0, \ \beta, \ \beta' \neq 0$. Common heavy-tailed models, like the *Fréchet*, *Burr*, *Generalized Pareto* and *Student's t*, among others, belong to this class. More generally, β and β' may be also regarded not as constants but as any slowly varying functions, i.e., regular varying functions with a null index of regular variation, but then, the second order parameters' estimates have a very erratic behaviour.

The adequate estimation of this parameter (or functional) β has revealed to be of great importance in the reduced bias' tail index estimators recently introduced in the literature (Gomes *et al.*, 2004; Caeiro *et al.*, 2005). In this paper, and to estimate β in (1.4) we shall consider the same type of statistics used in Fraga Alves *et al.* (2003a) for the estimation of ρ. Those statistics are based on the moment statistics

$$M_n^{(\alpha)}(k) := \frac{1}{k} \sum_{i=1}^{k} \left\{ \ln X_{n-i+1:n} - \ln X_{n-k:n} \right\}^\alpha, \quad \alpha > 0, \tag{1.6}$$

where $X_{i:n}$, $1 \leq i \leq n$, denotes the sample of ascending order statistics (o.s.) associated to the available random sample $(X_1, X_2, \cdots, X_n)$.

In section 2 of this paper we shall introduce a new class of β-estimators, and we shall consider their behaviour for intermediate values of k, i.e., sequences of integers $k = k_n$, belonging to $[1, n)$, and such that

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

proving their consistency whenever we go further in the tail, assuming that k is such that, for the A function in (1.2),

$$\lim_{n\to\infty} \sqrt{k}\ A(n/k) = \infty. \tag{1.8}$$

Next, in section 3, we deal with the conditions enabling us to get the asymptotic normality of this new class of estimators. In section 4 we shall illustrate the behaviour of the new class of estimators presented in section 2, through Monte Carlo simulation techniques, for a large variety of heavy-tailed models. A few final remarks are drawn in section 5. Section 6 is devoted to the proofs of the results stated in sections 2 and 3.

2 The class of β-estimators — consistency

In order to estimate β in the function $A(t) = \gamma\ \beta\ t^\rho$, we shall use, and extend for $\tau \in \mathbb{R}$, the same type of statistics considered for the estimation of ρ in Fraga Alves *et al.* (2003a). Before introducing the new class of β-estimators, we shall introduce the following notations: let W be a unit exponential r.v., with d.f. $F_W(x) = 1 - \exp(-x)$, $x > 0$,

$$\mu_\alpha^{(1)} = \mathbb{E}\left[W^\alpha\right] = \Gamma(\alpha+1), \quad \overline{\mu}_\alpha^{(1)} = 1, \tag{2.1}$$

$$\sigma_\alpha^{(1)} = \left(\mathbb{V}ar\left[W^\alpha\right]\right)^{1/2} = \left(\Gamma(2\alpha+1) - \Gamma^2(\alpha+1)\right)^{1/2}, \tag{2.2}$$

$$\overline{\sigma}_\alpha^{(1)}(\rho) = \frac{\sigma_\alpha^{(1)}(\rho)}{\mu_\alpha^{(1)}(\rho)} = \left(\frac{2\Gamma(2\alpha)}{\alpha\Gamma^2(\alpha)} - 1\right)^{1/2} \tag{2.3}$$

and, for $j = 2,\ 3$, let us write

$$\mu_\alpha^{(j)}(\rho) = \mathbb{E}\left[W^{\alpha-j+1}\left(\frac{e^{\rho W} - 1}{\rho}\right)^{j-1}\right], \quad \overline{\mu}_\alpha^{(j)}(\rho) = \frac{\mu_\alpha^{(j)}(\rho)}{\Gamma(\alpha+1)}, \tag{2.4}$$

$$\sigma_\alpha^{(j)}(\rho) = \left(\mathbb{V}ar\left[W^{\alpha-j+1}\left(\frac{e^{\rho W}-1}{\rho}\right)^{j-1}\right]\right)^{1/2}, \quad \overline{\sigma}_\alpha^{(j)}(\rho) = \frac{\sigma_\alpha^{(j)}(\rho)}{\Gamma(\alpha+1)}. \quad (2.5)$$

We shall provide in the following lemma explicit expressions for the functions $\mu_\alpha^{(j)}(\rho)$ in (2.4), $j = 2,\ 3$ and $\sigma_\alpha^{(2)}(\rho)$ in (2.5):

Lemma 2.1. *We may write*

$$\mu_\alpha^{(2)}(\rho) = \frac{\Gamma(\alpha+1)\,(1-(1-\rho)^\alpha)}{\alpha\ \rho\ (1-\rho)^\alpha}, \quad (2.6)$$

$$\begin{aligned} \mu_\alpha^{(3)}(\rho) &= \frac{2}{\rho}\left\{\mu_{\alpha-1}^{(2)}(2\rho) - \mu_{\alpha-1}^{(2)}(\rho)\right\} \\ &= \begin{cases} \frac{1}{\rho^2}\ln\frac{(1-\rho)^2}{1-2\rho} & \textit{if}\quad \alpha = 1 \\ \frac{\Gamma(\alpha+1)}{\rho^2\ \alpha\ (\alpha-1)}\left\{\frac{1}{(1-2\rho)^{\alpha-1}} - \frac{2}{(1-\rho)^{\alpha-1}} + 1\right\} & \textit{if}\quad \alpha \neq 1 \end{cases} \end{aligned} \quad (2.7)$$

and

$$\sigma_\alpha^{(2)}(\rho) = \left(\sqrt{\mu_{2\alpha}^{(3)}(\rho) - \left(\mu_\alpha^{(2)}(\rho)\right)^2}\right)^{1/2}.$$

For any $\alpha > 0$, let us consider $M_n^{(\alpha)}(k)$ in (1.6), and for real *tuning* parameters τ and $\theta_1 \neq \theta_2$, different from zero, let us consider the statistic

$$D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k) := \left(\frac{M_n^{(\alpha\theta_1)}(k)}{\mu_{\alpha\theta_1}^{(1)}}\right)^{\tau/\theta_1} - \left(\frac{M_n^{(\alpha\theta_2)}(k)}{\mu_{\alpha\theta_2}^{(1)}}\right)^{\tau/\theta_2}. \quad (2.8)$$

With the notation

$$d_\alpha^{(\theta_1,\theta_2)}(\rho) := \overline{\mu}_{\alpha\theta_1}^{(2)}(\rho) - \overline{\mu}_{\alpha\theta_2}^{(2)}(\rho), \quad (2.9)$$

we shall consider the class of r.v.'s,

$$\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \rho) := \frac{2d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)}{\alpha\ \tau\left\{d_\alpha^{(\theta_1,\theta_2)}(\rho)\right\}^2}\left(\frac{k}{n}\right)^\rho \frac{\left\{D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k)\right\}^2}{D_n^{(2\alpha,\theta_1,\theta_2,\tau)}(k)}, \quad (2.10)$$

and for any consistent estimator $\widehat{\rho}$ of ρ, the corresponding class of statistics $\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \widehat{\rho})$.

Remark 2.1. *The results in Fraga Alves et al.* (2003*a*) *enable us to guarantee that, under the second order framework in* (1.2), *if* (1.7) *and* (1.8) *hold,*

$$\frac{D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k)}{A(n/k)} = \frac{D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k)}{\beta\ \gamma\ (n/k)^\rho} \underset{n\to\infty}{\overset{p}{\longrightarrow}} \alpha\ \tau\ \gamma^{\alpha\tau-1}\ d_\alpha^{(\theta_1,\theta_2)}(\rho).$$

Consequently, we may get rid of γ *if we work with*

$$\left(\frac{k}{n}\right)^\rho \frac{\left(D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k)\right)^2}{D_n^{(2\alpha,\theta_1,\theta_2,\tau)}(k)} \underset{n\to\infty}{\overset{p}{\longrightarrow}} \frac{\beta\alpha\tau\left\{d_\alpha^{(\theta_1,\theta_2)}(\rho)\right\}^2}{2\ d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)}.$$

It is thus sensible to introduce the class of r.v.'s in (2.10), *which converges in probability towards* β *for every positive real* α *and real values* τ *and* $\theta_1 \neq \theta_2$.

Remark 2.2. *We suggest in practice the consideration of* $\alpha = 1$, $\theta_1 = 1$ *and* $\theta_2 = 2$, *in order to get a class of estimators,* $\widehat{\beta}^{(\tau)}(k;\ \widehat{\rho})$, *dependent only on the tuning parameter* $\tau \in \mathbb{R}$, *with the functional expression,*

$$\widehat{\beta}^{(\tau)}(k;\widehat{\rho}) := \begin{cases} -\frac{2(2-\widehat{\rho})^2}{\tau\widehat{\rho}}\left(\frac{k}{n}\right)^{\widehat{\rho}} \frac{\left\{\left(M_n^{(1)}(k)\right)^\tau - \left(M_n^{(2)}(k)/2\right)^{\tau/2}\right\}^2}{\left(M_n^{(2)}(k)/2\right)^\tau - \left(M_n^{(4)}(k)/24\right)^{\tau/2}} & \text{if } \tau \neq 0 \\ \\ -\frac{2(2-\widehat{\rho})^2}{\widehat{\rho}}\left(\frac{k}{n}\right)^{\widehat{\rho}} \frac{\left\{\ln\left(M_n^{(1)}(k)\right) - \frac{1}{2}\ \ln\left(M_n^{(2)}(k)/2\right)\right\}^2}{\ln\left(M_n^{(2)}(k)/2\right) - \frac{1}{2}\ln\left(M_n^{(4)}(k)/24\right)} & \text{if } \tau = 0 \end{cases}. \tag{2.11}$$

This unique tuning parameter provides an adequate flexible class of estimators of β *and has revealed to be suitable for practical purposes.*

We shall further use the notation

$$a_\alpha^{(\theta_1,\theta_2,\tau)}(\rho) := (\alpha\theta_1 - 1)\overline{\mu}_{\alpha\theta_1}^{(3)}(\rho) + \alpha(\tau - \theta_1)\left(\overline{\mu}_{\alpha\theta_1}^{(2)}(\rho)\right)^2 - (\alpha\theta_2 - 1)\overline{\mu}_{\alpha\theta_2}^{(3)}(\rho) - \alpha(\tau - \theta_2)\left(\overline{\mu}_{\alpha\theta_2}^{(2)}(\rho)\right)^2, \tag{2.12}$$

and

$$W_\alpha^{(\theta_1,\theta_2)} := \frac{\overline{\sigma}_{\alpha\theta_1}^{(1)}\ P_n^{(\alpha\theta_1)}}{\theta_1} - \frac{\overline{\sigma}_{\alpha\theta_2}^{(1)}\ P_n^{(\alpha\theta_2)}}{\theta_2}, \tag{2.13}$$

with $\overline{\sigma}_\alpha^{(1)}$, $\mu_\alpha^{(2)}(\rho)$ and $\mu_\alpha^{(3)}(\rho)$ provided explicitly in (2.3), (2.6) and (2.7), respectively, being $\overline{\mu}_\alpha^{(j)}(\rho)$ given in (2.4). Moreover,

$$P_n^{(r)} = \sqrt{k}\left\{\tfrac{1}{k}\sum_{j=1}^{k} W_j^r - \mu_r^{(1)}\right\}/\sigma_r^{(1)}, \tag{2.14}$$

with $\{W_j\}_{j\geq 1}$ a sequence of i.i.d. standard exponential r.v.'s, being $\mu_\alpha(1)$ and $\sigma_\alpha^{(1)}$ given in (2.1) and (2.2), respectively. We may state the following:

Theorem 2.1. *Under the validity of the second order condition in* (1.2)*, with* $A(t) = \gamma\ \beta\ t^\rho$*,* $\rho < 0$*, and for intermediate levels* k*, such that both* (1.7) *and* (1.8) *hold, the estimators* $\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \widehat{\rho})$*, with* $\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \rho)$ *given in* (2.10)*, converge in probability towards* β*, for any consistent estimator* $\widehat{\rho}$ *of* ρ *such that* $(\widehat{\rho}-\rho)\ln(n/k) = o_p(1)$*, and for any* $\alpha \in \mathbb{R}^+$*,* $\tau \in \mathbb{R}$*, and* $\theta_1,\ \theta_2 \in \mathbb{R}^+$*,* $\theta_1 \neq \theta_2$*.*

3 The asymptotic normality of the estimators

We may state the following:

Theorem 3.1. *Apart from the conditions in Theorem 2.1, if we further assume the third order condition in* (1.3)*, with* $\eta = \rho$*, for sake of simplicity,* $\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \rho)$ *in* (2.10) *is asymptotically normal, at a rate of convergence of the order of* $1/\left(\sqrt{k}\ A(n/k)\right)$*. The asymptotic mean value of* $\sqrt{k}\ \ A(n/k)\left(\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \rho)-\beta\right)$ *is null whenever* $\sqrt{k}\ \ A^2(n/k)$ *and* $\sqrt{k}\ \ A(n/k)\ \ B(n/k)$ *converge towards zero. If* $\sqrt{k}\ \ A^2(n/k) \to \lambda_{_A}$ *and* $\sqrt{k}\ A(n/k)\ B(n/k) \to \lambda_{_B}$*, both finite, as* $n \to \infty$*, then such asymptotic mean value is given by*

$$\lambda_{_A} u_\alpha^{(\theta_1,\theta_2,\tau)}(\gamma,\rho,\beta) + \lambda_{_B} v_\alpha^{(\theta_1,\theta_2)}(\rho,\beta),$$

where, with $d_\alpha^{(\theta_1,\theta_2)}(\rho)$ *and* $a_\alpha^{(\theta_1,\theta_2,\tau)}(\rho)$ *given in* (2.9) *and* (2.12)*, respectively,*

$$u_\alpha^{(\theta_1,\theta_2,\tau)}(\gamma,\rho,\beta) = \frac{\beta}{\gamma}\left(\frac{a_\alpha^{(\theta_1,\theta_2,\tau)}(\rho)}{d_\alpha^{(\theta_1,\theta_2)}(\rho)} - \frac{a_{2\alpha}^{(\theta_1,\theta_2,\tau)}(\rho)}{2d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)}\right) \tag{3.1}$$

and

$$v_\alpha^{(\theta_1,\theta_2)}(\rho,\beta) = 2\beta\left(\frac{d_\alpha^{(\theta_1,\theta_2)}(2\rho)}{d_\alpha^{(\theta_1,\theta_2)}(\rho)} - \frac{d_{2\alpha}^{(\theta_1,\theta_2)}(2\rho)}{2d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)}\right). \tag{3.2}$$

The asymptotic variance of $\sqrt{k}\ A(n/k)\ \widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \rho)$ *is given by*

$$\sigma^2_{\alpha,\theta_1,\theta_2,\tau}(\rho) = \left(\frac{\gamma\beta}{\alpha}\right)^2 Var\left[\frac{2W_\alpha^{(\theta_1,\theta_2)}}{d_\alpha^{(\theta_1,\theta_2)}(\rho)} - \frac{W_{2\alpha}^{(\theta_1,\theta_2)}}{2d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)}\right]$$

$$= \left(\frac{\gamma\beta}{\alpha}\right)^2 \left[\frac{4\sigma^2_{W|\alpha,\theta_1,\theta_2}}{\left(d_\alpha^{(\theta_1,\theta_2)}(\rho)\right)^2} + \frac{\sigma^2_{W|2\alpha,\theta_1,\theta_2}}{4\left(d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)\right)^2} - \frac{2\sigma_{\alpha,\theta_1,\theta_2}}{d_\alpha^{(\theta_1,\theta_2)}(\rho)d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)}\right], \quad (3.3)$$

where $W_\alpha^{(\theta_1,\theta_2)}$ *is given in* (2.13),

$$\sigma_{\alpha,\theta_1,\theta_2} = Cov\left(W_\alpha^{(\theta_1,\theta_2)}, W_{2\alpha}^{(\theta_1,\theta_2)}\right)$$
$$= \frac{1}{2\alpha}\left[\frac{3\ \Gamma(3\alpha\theta_1)}{\theta_1^3\ \Gamma(\alpha\theta_1)\Gamma(2\alpha\theta_1)} - \frac{(\theta_1+2\theta_2)\ \Gamma(\alpha(\theta_1+2\theta_2))}{\theta_1^2\ \theta_2^2\ \Gamma(\alpha\theta_1)\Gamma(2\alpha\theta_2)}\right.$$
$$\left. - \frac{(2\theta_1+\theta_2)\ \Gamma(\alpha(2\theta_1+\theta_2))}{\theta_1^2\ \theta_2^2\ \Gamma(2\alpha\theta_1)\Gamma(\alpha\theta_2)} + \frac{3\ \Gamma(3\alpha\theta_2)}{\theta_2^3\ \Gamma(\alpha\theta_2)\Gamma(2\alpha\theta_2)}\right] - \left(\frac{1}{\theta_1} - \frac{1}{\theta_2}\right)^2 \quad (3.4)$$

and

$$\sigma^2_{W|\alpha,\theta_1,\theta_2} = Var\left[W_\alpha^{(\theta_1,\theta_2)}\right]$$
$$= \frac{2}{\alpha}\left[\frac{\Gamma(2\alpha\theta_1)}{\theta_1^3\ \Gamma^2(\alpha\theta_1)} + \frac{\Gamma(2\alpha\theta_2)}{\theta_2^3\ \Gamma^2(\alpha\theta_2)}\frac{(\theta_1+\theta_2)\Gamma(\alpha(\theta_1+\theta_2))}{\theta_1^2\ \theta_2^2\ \Gamma(\alpha\theta_1)\Gamma(\alpha\theta_2)}\right] - \left(\frac{1}{\theta_1} - \frac{1}{\theta_2}\right)^2. \quad (3.5)$$

If $\alpha = \theta_1 = 1$ *and* $\theta_2 = 2$, *i.e., if we consider the particular case* $\widehat{\beta}^{(\tau)}(k;\rho)$, *with* $\widehat{\beta}^{(\tau)}(k;\widehat{\rho})$ *given in* (2.11), *we have*

$$\sigma^2_\tau(\rho) \equiv \sigma^2_{1,1,2,\tau}(\rho) = \left(\frac{\gamma\beta(1-\rho)}{2-\rho}\right)^2 \left(\frac{21\rho^4 - 68\rho^3 + 86\rho^2 - 68\rho + 33}{\rho^2}\right). \quad (3.6)$$

The same results remain true if we replace ρ *by* $\widehat{\rho}$, *a consistent estimator of* ρ *such that* $(\widehat{\rho}-\rho)\ln(n/k) = o_p\left(1/\left(\sqrt{k}\ A(n/k)\right)\right)$.

Remark 3.1. *Since the asymptotic variance in* (3.3), *is independent of* τ, *we next show, in Figure 1, for* $\theta_1 = 1$, $\theta_2 = 2$, *such a variance as a function of* α, *for the values* $\beta = 0.5$ (*Fréchet model*) *and* $\beta = 1$ (*Burr model*). *Note that the minimum asymptotic variance is achieved near* 0.5. *This is the main reason why we have decided to use* $\alpha = 1$ *in the simulations of section 4, the positive integer value where the variance is minimum. In practice, the tuning parameter* $\alpha = 0.5$ *is however a sensible alternative to* $\alpha = 1$.

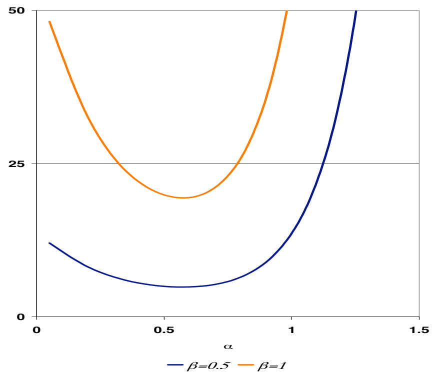

Figure 1: Asymptotic variance of $\sqrt{k}\ A(n/k)\left(\widehat{\beta}^{(\alpha,1,2,\tau)}(k;\ \widehat{\rho})-\beta\right)$ with $\gamma=-\rho=1$.

Remark 3.2. *For the estimation of ρ, we shall consider here the class of estimators introduced in Fraga Alves et al.* (2003a)*. Those estimates depend on a tuning parameter $\theta \ge 0$, that may be straightforwardly generalised to $\theta \in \mathbb{R}$, and have the functional expression,*

$$\widehat{\rho}_\theta(k) = \min\left(0, \frac{3(T_n^{(\theta)}(k)-1)}{T_n^{(\theta)}(k)-3}\right), \quad T_n^{(\theta)}(k) = \frac{D_n^{(1,1,2,\theta)}(k)}{D_n^{(1,2,3,\theta)}(k)}. \tag{3.7}$$

If, by means of any stability criterion for large k, like the ones suggested in Gomes and Pestana (2004)*, we consider a tuning parameter $\theta=\theta^*$ that enables us to guarantee that for levels k_1, like*

$$k_1 = \min(n-1, 2n/\ln\ln n), \tag{3.8}$$

the heuristic level often suggested in the literature, $\widehat{\rho}-\rho := \widehat{\rho}_{\theta^*}(k_1)-\rho = O_p\left(1/\left(\sqrt{k_1}\ A(n/k_1)\right)\right)$, we get $\widehat{\rho}-\rho = O_p\left((\ln\ln n)^{(1-2\rho)/2}/\sqrt{n}\right)$. Consequently, for any level k, $(\widehat{\rho}-\rho)\ln(n/k) = o_p(1)$, a condition needed in Theorem 2.1. If we consider levels k such that $\sqrt{k}\ A(n/k) \to \infty$ and $\sqrt{k}\ A^2(n/k) \to \lambda_{_A}$, finite, two of the conditions in Theorem 3.1, and models in (1.5)*, k is at most of the order of $n^{-4\rho/(1-4\rho)}$, being of a larger order than $n^{-2\rho/(1-2\rho)}$. Then $O\left(n^{-\rho/(1-4\rho)}\right) \le (n/k)^{-\rho} < O\left(n^{-\rho/(1-2\rho)}\right)$, and consequently $(\widehat{\rho}-\rho)\ln(n/k)/A(n/k) = o_p(1)$. Hence, if $\sqrt{k}\ A^2(n/k) \to \lambda_{_A}$, finite, $\sqrt{k}\ A(n/k)\,(\widehat{\rho}-\rho)\ln(n/k) \xrightarrow{p} 0$, and the conditions on $\widehat{\rho}$, assumed in Theorem 3.1, are true for this ρ-estimator. Note that in practice, the adequate choice of* *this tuning parameter θ^* is much more relevant then the choice of the threshold k_1.*

If we compute the β and ρ estimators at the same level k, i.e., if we consider $\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \widehat{\rho}_\theta(k))$, the asymptotic normal behaviour of this statistic, often considered in practice, is directly related with the behaviour of the ρ-estimator and no longer with the behaviour of the r.v. $\beta^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \rho)$. Indeed, directly from the results in Fraga Alves *et al.* (2003a), and from the fact that

$$\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \widehat{\rho}_\theta(k)) = \widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \rho) - \beta\,(\widehat{\rho}_\theta(k)-\rho)\ln(n/k)(1+o_p(1)),$$

we may state, without the need of a proof, the following:

Theorem 3.2. *Under the conditions of Theorem 3.1, the β-estimator $\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \widehat{\rho}_\theta(k))$, with $\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\rho)$ and $\widehat{\rho}_\theta(k)$ given in* (2.10) *and* (3.7)*, respectively, is asymptotically normal, at a rate of convergence equal to $\ln(n/k)/\left(\sqrt{k}\ A(n/k)\right)$. The asymptotic mean value of $\sqrt{k}\ A(n/k)\left(\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \widehat{\rho}(k))-\beta\right)/\ln(n/k)$ is null whenever $\sqrt{k}\ A^2(n/k)$ and $\sqrt{k}\ A(n/k)\ B(n/k)$ converge towards zero. If $\sqrt{k}\ A^2(n/k)\to\lambda_A$ and $\sqrt{k}\ A(n/k)\ B(n/k)\to\lambda_B$, both finite, as $n\to\infty$, such asymptotic mean value is given by $\beta\,(\lambda_A u_\rho+\lambda_B v_\rho)\equiv\beta\,(\lambda_A u_\rho(\gamma;\theta)+\lambda_B v_\rho)$, where*

$$u_\rho = -\frac{\rho\left(\theta(1-2\rho)^2(3-\rho)(3-2\rho)-6\rho\left(4\rho^3-16\rho^2+20\rho-7\right)\right)}{12\ \gamma\left((1-\rho)(1-2\rho)\right)^2}$$

and

$$v_\rho = -2\ \rho\left(\frac{1-\rho}{1-2\rho}\right)^3.$$

The asymptotic variance is given by

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

4 Finite sample behaviour of the estimators

In this section we will present the finite sample properties of some members of the class of β-estimators introduced, for the following set of typical heavy-tailed models:

1. The Fréchet(γ) model: $F(x) = \exp(-x^{-1/\gamma})$, $x > 0$, with $\gamma = 1$ ($\rho =-1$, $\beta = 0.5$);

2. The Burr(γ, ρ) model: $F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho}$, $x > 0$, with $\gamma = 1$ and $\rho = -0.5, -1, -2$ ($\beta = 1$);

The simulation results presented are based on a multi-sample simulation of size 10×1000 to guarantee reasonably small standard errors (not shown here, but available from the authors) for the following simulated characteristics: the mean value, the mean squared error and the optimal level, i.e. the level k_0 where the mean squared error attains its minimum. We shall present in the next figures and tables the simulated output. Notice that small values of k have not been represented in any of the figures. Indeed, as expected, we get a very high volatility of the estimators' characteristics for such values.

We shall use several values for the *tuning* parameter τ in the β estimator, $\widehat{\beta}^{(\tau)}(k;\ \widehat{\rho})$, provided in (2.11). One of our goals is to find, for large values of k, a τ-value which provides high stability around the true β value. Regarding the estimation of ρ: any stability criterion for values of k between $n^{0.900}$ and $n^{0.995}$ led us to the choice $\tau = 0$ whenever $|\rho| \leq 1$ and $\tau = 1$ for $|\rho > 1$. Since educated guesses are much more precise than possibly noisy estimates of τ, we advance with the consideration of the ρ-estimator $\widehat{\rho}_j = \widehat{\rho}_j(k_1)$, $j = 0$ or 1, with $\widehat{\rho}_\theta(k)$ and k_1 given in (3.7) and (3.8), respectively, according as $|\rho \leq 1$ or $|\rho| > 1$. We shall also consider, for comparison with the new β-estimators in (2.11), the estimator of β obtained in Gomes and Martins (2002), given by

$$\widehat{\beta}^{ML}(k;\ \widehat{\rho}) := \frac{1}{n^{\widehat{\rho}}} \frac{\left(\sum_{i=1}^{k} i^{-\widehat{\rho}}\right)\left(\sum_{i=1}^{k} U_i\right) - k\left(\sum_{i=1}^{k} i^{-\widehat{\rho}}\, U_i\right)}{\left(\sum_{i=1}^{k} i^{-\widehat{\rho}}\right)\left(\sum_{i=1}^{k} i^{-\widehat{\rho}}\, U_i\right) - k\left(\sum_{i=1}^{k} i^{-2\widehat{\rho}}\, U_i\right)}, \tag{4.1}$$

where $U_i = i\,[\ln X_{n-i+1:n} - \ln X_{n-i:n}]$, $1 \le i \le k$, are the scaled log-spacings.

Remark 4.1. *Supose that the model F underlying the available data is such that $F^{-1}(t) = (\gamma^j g(t))^\gamma$, for any $j \in \mathbb{R}$, a condition that holds for both Fréchet and Burr models. Then $\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \widehat{\rho})$ is independent of γ. This property comes from the identity,*

$$M_n^{(\alpha)}(k) = \frac{1}{k}\sum_{i=1}^{k}\left(\ln\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{\alpha} \stackrel{d}{=} \frac{\gamma^\alpha}{k}\sum_{i=1}^{k}\left(\ln\frac{g(U_{n-i+1:n})}{g(U_{n-k:n})}\right)^{\alpha},$$

where $U_{i:n}$ denotes the i-th ascending o.s., $1 \le i \le n$, associated with a Uniform$(0,1)$ random sample of size n. Consequently, when we compute $\left(D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k)\right)^2 / D_n^{(2\alpha,\theta_1,\theta_2,\tau)}(k)$, we get rid of γ.

4.1 Mean values and mean squared errors patterns

The simulated mean values and mean squared errors' patterns of the two estimators, $\widehat{\beta}^{(\tau)}(k;\ \widehat{\rho})$ in (2.11) and $\widehat{\beta}^{ML}(k;\ \widehat{\rho})$ in (4.1), as a function of k, are presented in Figures 2, 3, 4 and 5.

$\mathbb{E}[\cdot]$ $\qquad\qquad$ $MSE[\cdot]$

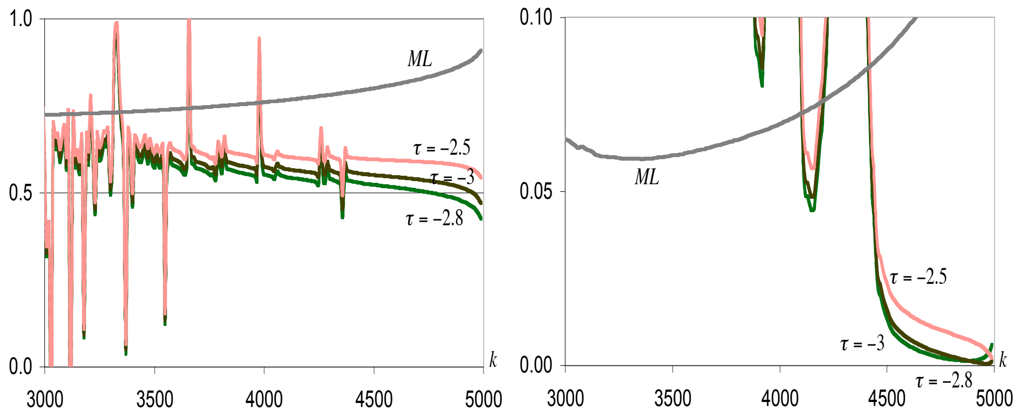

Figure 2: Simulated mean values (left) and MSE (right) of $\widehat{\beta}^{(\tau)}(k;\ \widehat{\rho})$ for $\tau = -3, -2.8, -2.5$ for a sample of size 5000 from a Fréchet(1) model ($\rho = -1,\ \beta = 0.5$).

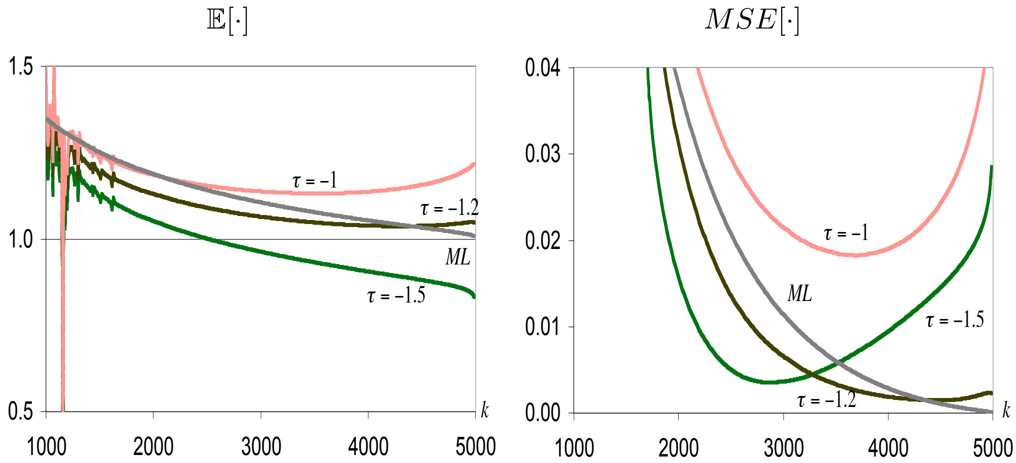

Figure 3: Simulated mean values (*left*) and MSE (*right*) of $\widehat{\beta}^{(\tau)}(k;\ \widehat{\rho})$ with $\tau = -1.5, -1.2, -1$ for a sample of size 5000 from a Burr$(1, -0.5)$ model $(\beta = 1)$.

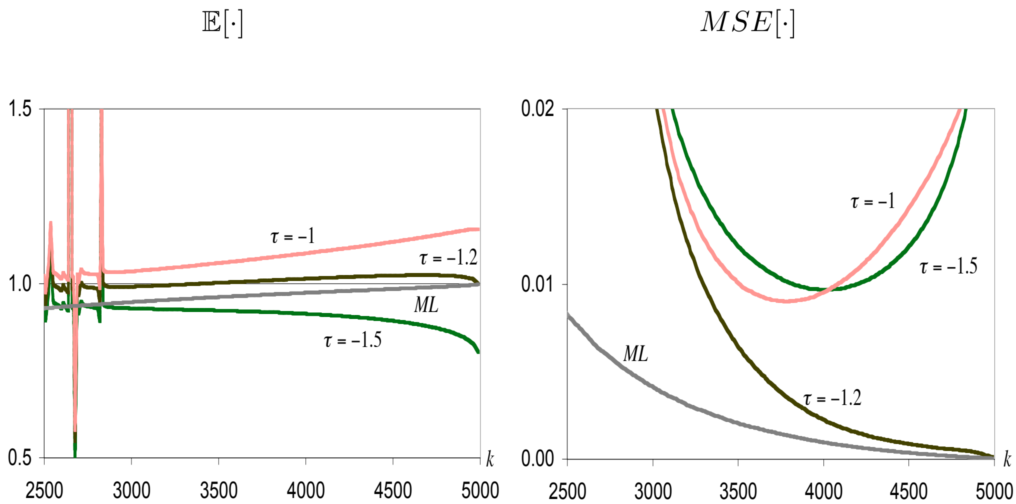

Figure 4: Simulated mean values (left) and MSE (right) of $\widehat{\beta}^{(\tau)}(k;\ \widehat{\rho})$ with $\tau = -1.5, -1.2, -1$ for a sample of size 5000 from a Burr$(1, -1)$ model $(\beta = 1)$.

In Table 1, for values of $n = 100,\ 500,\ 1000,\ 2000,\ 5000$ and 10000, we present the simulated mean values ($\mathbb{E}_0$) and mean squared errors (MSE_0) of the estimators at their optimal levels, i.e., at the levels k_0 where the mean squared error attains their minimum. Although irrelevant from a practical point of view, this information shows us the potentialities of the different estimation

$\mathbb{E}[\cdot]$ $MSE[\cdot]$

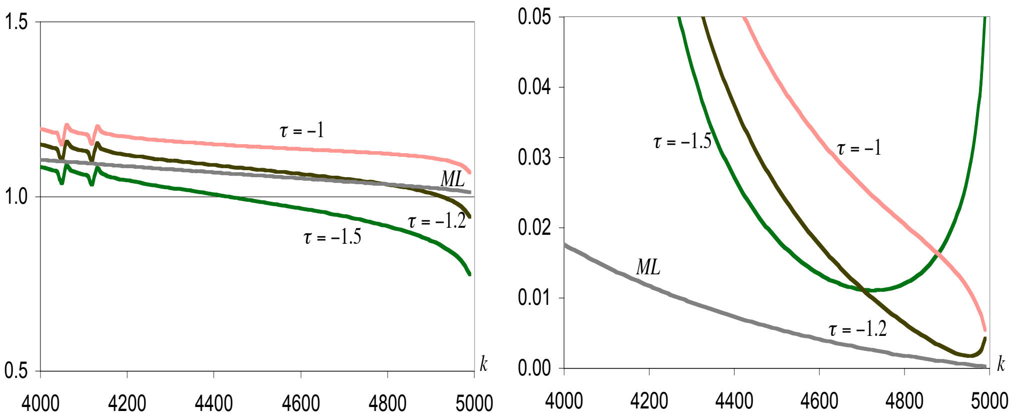

Figure 5: Simulated mean values (left) and MSE (right) of $\widehat{\beta}^{(\tau)}(k;\ \widehat{\rho})$ with $\tau = -1.5, -1.2, -1$ for a sample of size 5000 from a Burr$(1, -2)$ model ($\beta = 1$).

procedures. For each underlying model and each value of n, the smallest squared bias and mean squared error are underlined.

Remark 4.2. *The results presented allow us to make the following specific comments:*

- *The sample path of $\widehat{\beta}^{ML}(k;\ \widehat{\rho})$ has revealed to be, for all simulated models, less volatile than that of $\widehat{\beta}^{(\tau)}(k;\ \widehat{\rho})$. However, for a suitably chosen value of τ, $\widehat{\beta}^{(\tau)}(k;\ \widehat{\rho})$ exhibits for large values of k, a quite stable sample path around the true value of β.*

- *The choice of the tuning parameter τ depends heavily on the model, and contrarily to what happens with the choice of the tuning parameter θ in the estimator of ρ in (3.7), the best choice of τ in (2.11) is provided by negative values of this control parameter. The choice of τ for the estimation of β through (2.11), is much less resistant to changes in the model than the choice of θ in the estimation of ρ through (3.7).*

- *For values of $\rho \neq -1$ and in a Burr model (Figures 3 and 5), $\widehat{\beta}^{ML}(k;\ \widehat{\rho})$ provides for a wider region of k-values, and comparatively to $\widehat{\beta}^{(\tau)}(k;\ \widehat{\rho})$,*

Table 1: Simulated mean value ($\mathbb{E}_0$) and MSE (MSE_0) of the estimators at their optimal levels.

Fréchet($\gamma = 1$) parent: $\beta = 0.5$								
	$\mathbb{E}_0$				MSE_0			
	$\tau=-3$	$\tau=-2.8$	$\tau=-2.5$	ML	$\tau=-3$	$\tau=-2.8$	$\tau=-2.5$	ML
100	0.5028	0.5450	0.6155	0.8577	1.0466	1.1054	1.2082	0.1456
500	0.4386	0.4653	0.5397	0.7250	0.0137	0.0101	0.0112	0.0665
1000	0.4542	0.4753	0.5295	0.7035	0.0067	0.0041	0.0040	0.0536
2000	0.4681	0.4853	0.5155	0.7231	0.0033	0.0015	0.0010	0.0589
5000	0.4832	0.4946	0.5061	0.7254	0.0012	0.0004	0.0003	0.0576
10000	0.4917	0.4982	0.5012	0.7120	0.0004	0.0001	0.0002	0.0494
Burr($\gamma = 1, \rho = -0.5$) parent: $\beta = 1$								
	$\tau=-1.5$	$\tau=-1.2$	$\tau=-1$	ML	$\tau=-1.5$	$\tau=-1.2$	$\tau=-1$	ML
100	0.8635	1.0348	1.1458	1.0088	0.0226	0.0013	0.0300	0.0002
500	0.9102	1.0350	1.1285	1.0070	0.0135	0.0012	0.0210	0.0001
1000	0.9280	1.0353	1.1288	1.0061	0.0096	0.0012	0.0193	0.0000
2000	0.9506	1.0287	1.1252	1.0056	0.0066	0.0009	0.0173	0.0000
5000	0.9715	1.0362	1.1309	1.0045	0.0035	0.0014	0.0181	0.0000
10000	0.9816	1.0417	1.1346	1.0039	0.0019	0.0018	0.0187	0.0000
Burr($\gamma = 1, \rho = -1$) parent: $\beta = 1$								
	$\tau=-1.5$	$\tau=-1.2$	$\tau=-1$	ML	$\tau=-1.5$	$\tau=-1.2$	$\tau=-1$	ML
100	0.8707	1.0355	1.1689	0.9893	0.0249	0.0065	0.0391	0.0004
500	0.9083	1.0377	1.1482	0.9934	0.0116	0.0014	0.0313	0.0001
1000	0.9206	1.0380	1.1300	0.9941	0.0088	0.0014	0.0265	0.0000
2000	0.9111	1.0033	1.0985	0.9961	0.0101	0.0001	0.0158	0.0000
5000	0.9117	1.0020	1.0712	0.9976	0.0096	0.0001	0.0089	0.0000
10000	0.9153	1.0017	1.0574	0.9981	0.0087	0.0001	0.0063	0.0000
Burr($\gamma = 1, \rho = -2$) parent: $\beta = 1$								
	$\tau=-1.5$	$\tau=-1.2$	$\tau=-1$	ML	$\tau=-1.5$	$\tau=-1.2$	$\tau=-1$	ML
100	0.8415	0.9955	1.1132	1.0141	0.5151	0.5424	0.5939	0.0024
500	0.8630	0.9584	1.0598	1.0116	0.0312	0.0061	0.0063	0.0002
1000	0.9003	0.9704	1.0494	1.0101	0.0203	0.0034	0.0035	0.0001
2000	0.9199	0.9776	1.0407	1.0092	0.0150	0.0024	0.0023	0.0001
5000	0.9359	0.9843	1.0282	1.0065	0.0110	0.0017	0.0013	0.0001
10000	0.9469	0.9888	1.0193	1.0053	0.0081	0.0011	0.0007	0.0000

a sample path closer to the target value β. Also, the minimum mean squared error of $\widehat{\beta}^{ML}(k;\ \widehat{\rho})$ is smaller than the minimum mean squared error of $\widehat{\beta}^{(\tau)}(k;\ \widehat{\rho})$ for any of the τ-values considered. However, for $\rho = -1$ (Figures 2 and 4), if we choose properly the parameter τ, the reverse may happen.

Remark 4.3. *We again advise practitioners not to choose blindly the value of τ. It is sensible to draw a few sample paths of $\widehat{\beta}^{(\tau)}(k;\ \widehat{\rho})$ in (2.11), for instance* *for $\tau = 0, \ -0.5, \ -1, \ -1.5, \ -2, \ -3$, as functions of k, electing the value of τ which provides higher stability for $k \geq k_{01}$, large, by means of any stability criterion for $\widehat{\beta}^{(\tau)}(k; \ \widehat{\rho})$, $k_{01} \leq k < n$, around the median or the average of the sample estimates associated to a suitable region of k-values. Let us define*

$$\chi(\tau, \ i_n, \ j_n) := \widehat{\chi}_{1/2}\left(\widehat{\beta}^{(\tau)}(k; \ \widehat{\rho}), \ i_n \leq k \leq j_n\right)$$

and

$$\Sigma(\tau, \ i_n, \ j_n) := \sum_{k=i_n}^{j_n} \left\{\widehat{\beta}^{(\tau)}(k; \ \widehat{\rho}) - \chi(\tau, \ i_n, \ j_n)\right\}^2,$$

where $\widehat{\chi}_{1/2}\,(u_k, \ i_n \leq k \leq j_n)$ denotes the sample median of the sample $(u_k, \ i_n \leq k \leq j_n)$. We may then take

$$\tau_0 := \arg\min_{\tau} \ \Sigma(\tau, \ i_n, \ j_n), \qquad \widehat{\beta}_0 := \chi(\tau_0, \ i_n, \ j_n),$$

for suitably chosen values of i_n and j_n, like for instance $i_n = n^{0.95}$ and $j_n = n^{0.99}$. The choice of these values, provided they are large enough, is not quite important due to the robustness of the method regarding such a choice.

4.2 Simulation study at the fixed level k_1

In Table 2 we present for the heavy-tailed models considered before, the mean values and mean squared errors of the estimator $\widehat{\beta}^{(\tau)}(k; \ \widehat{\rho})$ computed also at the heuristic level k_1 in (3.8), for the τ-value leading to the minimum mean squared error, among the ones simulated. Again, the smallest squared bias and mean squared error are underlined.

As may be seen from Table 2, with the heuristic choice of k_1 in (3.8), the new estimator $\widehat{\beta}^{(\tau)}\,(k_1; \ \widehat{\rho})$ overpasses the existing ML estimator in (4.1), provided we choose carefully the *tuning* parameter τ in (2.11), which is unfortunately strongly stuck to the underlying model. For large values of n ($n > 10000$), and unless $|\rho| < 1$, the new estimator works always better than the existing one.

Table 2: Simulated mean value and mean squared errors of the β-estimators computed at the level k_1 in (3.8).

n	100	500	1000	2000	5000	10000	25000	50000
	Fréchet parent($\gamma = 1$): $\beta = 0.5,\ \rho = -1$							
$\mathbb{E}\left[\widehat{\beta}^{(-2.8)}_{\widehat{\rho}}(k)\right]$	0.5702	0.4611	0.4531	0.5055	0.5400	0.5474	0.5529	0.5555
$\mathbb{E}\left[\widehat{\beta}^{ML}_{\widehat{\rho}}(k)\right]$	0.8914	0.9106	0.9166	0.8750	0.8200	0.7975	0.7765	0.7644
$MSE[\widehat{\beta}^{(-2.8)}_{\widehat{\rho}}(k)]$	6.5223	0.0498	0.0051	0.0028	0.0054	0.0038	0.0037	0.0036
$MSE[\widehat{\beta}^{ML}_{\widehat{\rho}}(k)]$	0.1585	0.1689	0.1738	0.1408	0.1029	0.0889	0.0767	0.0701
	Burr($\gamma = 1,\ \rho = -0.5$) parent: $\beta = 1$							
$\mathbb{E}\left[\widehat{\beta}^{(-1.2)}_{\widehat{\rho}}(k)\right]$	1.0348	1.0350	1.0353	1.0342	1.0388	1.0431	1.0480	1.0512
$\mathbb{E}\left[\widehat{\beta}^{ML}_{\widehat{\rho}}(k)\right]$	1.0088	1.0070	1.0061	1.0135	1.0249	1.0300	1.0350	1.0380
$MSE[\widehat{\beta}^{(-1.2)}_{\widehat{\rho}}(k)]$	0.0013	0.0012	0.0012	0.0012	0.0015	0.0019	0.0023	0.0026
$MSE[\widehat{\beta}^{ML}_{\widehat{\rho}}(k)]$	0.0002	0.0001	0.0000	0.0002	0.0006	0.0009	0.0012	0.0014
	Burr($\gamma = 1,\ \rho = -1$) parent: $\beta = 1$							
$\mathbb{E}\left[\widehat{\beta}^{(-1.2)}_{\widehat{\rho}}(k)\right]$	1.0355	1.0377	1.0380	1.0341	1.0226	1.0175	1.0124	1.0094
$\mathbb{E}\left[\widehat{\beta}^{ML}_{\widehat{\rho}}(k)\right]$	0.9893	0.9934	0.9941	0.9908	0.9866	0.9850	0.9834	0.9824
$MSE[\widehat{\beta}^{(-1.2)}_{\widehat{\rho}}(k)]$	0.0017	0.0014	0.0014	0.0012	0.0006	0.0004	0.0002	0.0002
$MSE[\widehat{\beta}^{ML}_{\widehat{\rho}}(k)]$	0.0004	0.0001	0.0000	0.0001	0.0002	0.0003	0.0003	0.0003
	Burr($\gamma = 1,\ \rho = -2$) parent: $\beta = 1$							
$\mathbb{E}\left[\widehat{\beta}^{(-1.2)}_{\widehat{\rho}}(k)\right]$	1.0480	1.0536	1.0530	1.0491	1.0141	0.9966	0.9840	0.9708
$\mathbb{E}\left[\widehat{\beta}^{ML}_{\widehat{\rho}}(k)\right]$	1.0141	1.0116	1.0101	1.0244	1.0449	1.0526	1.0579	1.0596
$MSE[\widehat{\beta}^{(-1.2)}_{\widehat{\rho}}(k)]$	0.0139	0.0036	0.0030	0.0033	0.0038	0.0043	0.0035	0.0034
$MSE[\widehat{\beta}^{ML}_{\widehat{\rho}}(k)]$	0.0024	0.0002	0.0001	0.0007	0.0032	0.0046	0.0053	0.0052

5 Final remarks

- The class of β-estimators investigated may provide an interesting alternative to the available estimator in Gomes and Martins (2002).

- In practice, the choice of the tuning parameters may be done adaptively, but the choice $(\alpha, \theta_1, \theta_2) = (1, 1, 2)$ is with no doubt the most convenient for simulation studies, since the algorithms are much less time consuming. And the unique *tuning* parameter τ has revealed to be sufficient in practice.

- It is however possible to reduce the asymptotic variance, and hopefully get better results, if we work with $(\alpha, \theta_1, \theta_2) = (0.5, 1, 2)$ and $\widehat{\beta}^{(0.5,1,2,\tau)}(k;\ \widehat{\rho})$.

- For all the simulated models and large k values, the sample paths of $\widehat{\beta}^{(1,1,2,\tau)}(k;\ \widehat{\rho})$ are quite stable if we chose correctly the parameter τ. Stability is achieved only for very large values of k.

6 Proofs

6.1 Weak consistency (Proof of Theorem 2.1)

Proof. Under the second order condition in (1.2), assuming that (1.7) holds, and using the same arguments as in lemma 2 of Draisma *et al.* (1999) and more recently in Gomes *et al.* (2002) and Fraga Alves *et al.* (2003a), we may write for any $\alpha,\ \theta_1, \theta_2 > 0$ and $\tau \in \mathbb{R}$ the distributional representation

$$\left(\left(\frac{M_n^{(\alpha\theta_1)}(k)}{\mu_{\alpha\theta_1}^{(1)}}\right)^{\tau/\theta_1}, \left(\frac{M_n^{(\alpha\theta_2)}(k)}{\mu_{\alpha\theta_2}^{(1)}}\right)^{\tau/\theta_2}\right)$$
$$\stackrel{d}{=} \gamma^{\alpha\tau}\left(1 + \frac{\tau\ \overline{\sigma}_{\alpha\theta_1}^{(1)}\ P_n^{(\alpha\theta_1)}}{\theta_1\sqrt{k}} + \frac{\tau\ \alpha\ \overline{\mu}_{\alpha\theta_1}^{(2)}(\rho)A(n/k)}{\gamma} + o_p\left(A(n/k)\right),\right.$$
$$\left. 1 + \frac{\tau\ \overline{\sigma}_{\alpha\theta_2}^{(1)}\ P_n^{(\alpha\theta_2)}}{\theta_2\sqrt{k}} + \frac{\tau\ \alpha\ \overline{\mu}_{\alpha\theta_2}^{(2)}(\rho)A(n/k)}{\gamma} + o_p\left(A(n/k)\right)\right),$$

where $P_n^{(\alpha\theta_1)}$ and $P_n^{(\alpha\theta_2)}$ are the dependent sequences of asymptotically standard Normal r.v.'s in (2.14).

Consequently, as $(X_1, X_2) \stackrel{d}{=} (Y_1, Y_2)$ implies $X_1 - X_2 \stackrel{d}{=} Y_1 - Y_2$, we may write, with $D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k)$ and $W_\alpha^{(\theta_1,\theta_2)}$ given in (2.8) and (2.13), respectively,

$$D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k) =\alpha\ \tau\ \gamma^{\tau\alpha-1}\left(\frac{\gamma\ W_\alpha^{(\theta_1,\theta_2)}}{\alpha\sqrt{k}}\ +\ A(n/k)\ d_\alpha^{(\theta_1,\theta_2)}(\rho) + o_p\left(A(n/k)\right)\right).$$

For intermediate k, if $\sqrt{k}\ A(n/k) \to \infty$, as $n \to \infty$, i.e., $1/\sqrt{k} = o(A(n/k))$, the second term in the previous distributional representation is the dominant one, the other terms are $o_p(A(n/k))$ and may be discarded, and

$$\frac{D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k)}{A(n/k)} \underset{n\to\infty}{\overset{p}{\longrightarrow}} \alpha\ \tau\ \gamma^{\alpha\tau-1}\ d_\alpha^{(\theta_1,\theta_2)}(\rho).$$

Consequently, in Hall-Welsh sub-class of models in (1.5), and writing $A(t)$ as in (1.4), we get

$$\left(\frac{k}{n}\right)^{\rho} \ D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k) \ \underset{n\to\infty}{\overset{p}{\longrightarrow}} \ \beta \ \alpha \ \tau \ \gamma^{\alpha\tau} \ d_\alpha^{(\theta_1,\theta_2)}(\rho),$$

i.e.

$$\left(\frac{k}{n}\right)^{2\rho} \ \left(D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k)\right)^2 \ \underset{n\to\infty}{\overset{p}{\longrightarrow}} (\beta \ \alpha \ \tau)^2 \ \gamma^{2\alpha\tau} \left\{d_\alpha^{(\theta_1,\theta_2)}(\rho)\right\}^2. \qquad (6.1)$$

To get rid of the unknown γ, it is enough to consider that

$$\left(\frac{k}{n}\right)^{\rho} \ D_n^{(2\alpha,\theta_1,\theta_2,\tau)}(k) \ \underset{n\to\infty}{\overset{p}{\longrightarrow}} \ 2\beta \ \alpha \ \tau \ \gamma^{2\alpha\tau} \ d_{2\alpha}^{(\theta_1,\theta_2)}(\rho). \qquad (6.2)$$

The quotient between (6.1) and (6.2), enables us to say that

$$\left(\frac{k}{n}\right)^{\rho} \ \frac{\left(D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k)\right)^2}{D_n^{(2\alpha,\theta_1,\theta_2,\tau)}(k)} \ \underset{n\to\infty}{\overset{p}{\longrightarrow}} \ \tfrac{1}{2} \ \beta \ \alpha \ \tau \ \frac{\left\{d_\alpha^{(\theta_1,\theta_2)}(\rho)\right\}^2}{d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)},$$

and consequently, $\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\rho)$ in (2.10) converge in probability towards β, as $n \to \infty$. The replacement of ρ by $\widehat{\rho}$ in the scale factor $2 \ d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)/\left(\alpha\tau\left\{d_\alpha^{(\theta_1,\theta_2)}(\rho)\right\}^2\right)$ poses no problem because of continuity, provided that $\widehat{\rho}$ is consistent for the estimation of ρ. But the replacement of ρ by $\widehat{\rho}$ in $(k/n)^\rho$ requires that $(k/n)^{\widehat{\rho}}/(k/n)^\rho \underset{n\to\infty}{\overset{p}{\longrightarrow}} 1$, and hence the need to impose the condition $(\widehat{\rho}-\rho)\ln(n/k) = o_p(1)$.

6.2 Normality (Proof of Theorem 3.1)

Let us now assume that $\sqrt{k} \ A^2(n/k) \ \to \ \lambda_{_A}$, $\sqrt{k} \ A(n/k)B(n/k) \to \lambda_{_B}$, as $n \to \infty$, both finite and perhaps non-null. If $\lambda_{_A} = \lambda_{_B} = 0$, then both $A^2(n/k)$ and $A(n/k)B(n/k)$ are of smaller order than $1/\sqrt{k}$.

Going further into the third order framework in (1.3), and with the same notation as before, we may write

$$\left(\frac{M_n^{(\alpha\theta)}(k)}{\mu_{\alpha\theta}^{(1)}}\right)^{\tau/\theta} \stackrel{d}{=} \gamma^{\alpha\tau}\left(1+\frac{\tau\overline{\sigma}_{\alpha\theta}^{(1)}\ P_n^{(\alpha\theta)}}{\theta\sqrt{k}}+\frac{\tau\alpha\overline{\mu}_{\alpha\theta}^{(2)}(\rho)A(n/k)}{\gamma}+O_p\left(\frac{A(n/k)}{\sqrt{k}}\right)\right.$$
$$+\frac{\tau\ \alpha\ A^2(n/k)}{2\gamma^2}\left((\alpha\theta-1)\ \overline{\mu}_{\alpha\theta}^{(3)}(\rho)+\alpha(\tau-\theta)(\overline{\mu}_{\alpha\theta}^{(2)}(\rho))^2\right)(1+o_p(1))$$
$$\left.+\frac{\tau\ \alpha\ \overline{\mu}_{\alpha\theta}^{(2)}(2\rho)\ A(n/k)\ B(n/k)}{\gamma}(1+o_p(1))\right) \quad (6.3)$$

and

$$D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k) = \alpha\ \tau\ \gamma^{\tau\alpha-1}\left(\frac{\gamma\ W_\alpha^{(\theta_1,\theta_2)}}{\alpha\sqrt{k}}\ +\ A(n/k)\ d_\alpha^{(\theta_1,\theta_2)}(\rho)+O_p\left(\frac{A(n/k)}{\sqrt{k}}\right)\right.$$
$$\left.+A(n/k)\left(\frac{1}{2\gamma}\ A(n/k)\ a_\alpha^{(\theta_1,\theta_2,\tau)}(\rho)+B(n/k)\ d_\alpha^{(\theta_1,\theta_2)}(2\rho)\right)(1+o_p(1))\right).$$

Then,

$$\frac{D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k)}{A(n/k)} = \alpha\ \tau\ \gamma^{\tau\alpha-1}d_\alpha^{(\theta_1,\theta_2)}(\rho)\left(1+\frac{\gamma\ W_\alpha^{(\theta_1,\theta_2)}}{\alpha\ d_\alpha^{(\theta_1,\theta_2)}(\rho)\sqrt{k}\ A(n/k)}\right.$$
$$\left.+\ \frac{1}{d_\alpha^{(\theta_1,\theta_2)}(\rho)}\left(\frac{1}{2\gamma}\ A(n/k)\ a_\alpha^{(\theta_1,\theta_2,\tau)}(\rho)+B(n/k)\ d_\alpha^{(\theta_1,\theta_2)}(2\rho)\right)\right)\ (1+o_p(1)).$$

Consequently,

$$\frac{\left\{D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k)\right\}^2}{A^2(n/k)} = (\alpha\tau)^2\gamma^{2\tau\alpha-2}\left(\left\{d_\alpha^{(\theta_1,\theta_2)}(\rho)\right\}^2+\frac{2\gamma\ d_\alpha^{(\theta_1,\theta_2)}(\rho)W_\alpha^{(\theta_1,\theta_2)}}{\alpha\sqrt{k}\ A(n/k)}\right.$$
$$\left.+2\left(\frac{1}{2\gamma}\ A(n/k)\ a_\alpha^{(\theta_1,\theta_2,\tau)}(\rho)+B(n/k)\ d_\alpha^{(\theta_1,\theta_2)}(2\rho)\right)d_\alpha^{(\theta_1,\theta_2)}(\rho)\right)\ (1+o_p(1)).$$

and since $1/(1+x)=1-x+o(x)$, as $x\to 0$, we get

$$\frac{A(n/k)}{D_n^{(2\alpha,\theta_1,\theta_2,\tau)}(k)} = \frac{1}{2\alpha\tau\gamma^{2\tau\alpha-1}d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)}\left(1-\frac{\gamma\ W_{2\alpha}^{(\theta_1,\theta_2)}}{2\alpha\ d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)\sqrt{k}\ A(n/k)}\right.$$
$$\left.-\frac{1}{d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)}\left(\frac{1}{2\gamma}\ A(n/k)\ a_{2\alpha}^{(\theta_1,\theta_2,\tau)}(\rho)+B(n/k)\ d_{2\alpha}^{(\theta_1,\theta_2)}(2\rho)\right)\ (1+o_p(1))\right).$$

Now,

$$\frac{\left\{D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k)\right\}^2}{A(n/k)\ D_n^{(2\alpha,\theta_1,\theta_2,\tau)}(k)} = \frac{\alpha\tau}{2\gamma}\ \frac{\left\{d_\alpha^{(\theta_1,\theta_2)}(\rho)\right\}^2}{d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)} \left(1 + \frac{2\gamma\ W_\alpha^{(\theta_1,\theta_2)}}{\alpha\ d_\alpha^{(\theta_1,\theta_2)}(\rho)\sqrt{k}\ A(n/k)}\right.$$

$$\left. + \frac{2}{d_\alpha^{(\theta_1,\theta_2)}(\rho)} \left(\frac{1}{2\gamma}\ A(n/k)\ a_\alpha^{(\theta_1,\theta_2,\tau)}(\rho) + B(n/k)\ d_\alpha^{(\theta_1,\theta_2)}(2\rho)\right)\ (1+o_p(1))\right)$$

$$\times \left(1 - \frac{\gamma\ W_{2\alpha}^{(\theta_1,\theta_2)}}{2\alpha\ d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)\sqrt{k}\ A(n/k)}\right.$$

$$\left. - \frac{1}{d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)} \left(\frac{1}{2\gamma}\ A(n/k)\ a_{2\alpha}^{(\theta_1,\theta_2,\tau)}(\rho) + B(n/k)\ d_{2\alpha}^{(\theta_1,\theta_2)}(2\rho)\right)\ (1+o_p(1))\right).$$

Since $A(t) = \gamma\ \beta\ t^\rho$, we may write

$$\widehat{\beta}^{(\alpha,\theta_1,\theta_2)}(k;\ \rho) = \frac{2d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)}{\alpha\tau\left\{d_\alpha^{(\theta_1,\theta_2)}(\rho)\right\}^2}\left(\frac{k}{n}\right)^\rho \frac{\left\{D_n^{(\alpha,\theta_1,\theta_2,\tau)}(k)\right\}^2}{D_n^{(2\alpha,\theta_1,\theta_2,\tau)}(k)}$$

$$= \beta\left(1 + \frac{2\gamma\ W_\alpha^{(\theta_1,\theta_2)}}{\alpha\ d_\alpha^{(\theta_1,\theta_2)}(\rho)\sqrt{k}\ A(n/k)} - \frac{\gamma\ W_{2\alpha}^{(\theta_1,\theta_2)}}{2\alpha\ d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)\sqrt{k}\ A(n/k)}\right.$$

$$+ \frac{2}{d_\alpha^{(\theta_1,\theta_2)}(\rho)} \left(\frac{1}{2\gamma}\ A(n/k)\ a_\alpha^{(\theta_1,\theta_2,\tau)}(\rho) + B(n/k)\ d_\alpha^{(\theta_1,\theta_2)}(2\rho)\right)\ (1+o_p(1))$$

$$\left. - \frac{1}{d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)} \left(\frac{1}{2\gamma}\ A(n/k)\ a_{2\alpha}^{(\theta_1,\theta_2,\tau)}(\rho) + B(n/k)\ d_{2\alpha}^{(\theta_1,\theta_2)}(2\rho)\right)\ (1+o_p(1))\right),$$

and consequently,

$$\sqrt{k}\ A(n/k)\left(\widehat{\beta}^{(\alpha,\theta_1,\theta_2)}(k;\ \rho) - \beta\right) = \beta\left(\frac{2\gamma\ W_\alpha^{(\theta_1,\theta_2)}}{\alpha\ d_\alpha^{(\theta_1,\theta_2)}(\rho)} - \frac{\gamma\ W_{2\alpha}^{(\theta_1,\theta_2)}}{2\alpha\ d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)}\right.$$

$$+ \frac{2}{d_\alpha^{(\theta_1,\theta_2)}(\rho)} \left(\frac{1}{2\gamma}\ \lambda_A\ a_\alpha^{(\theta_1,\theta_2,\tau)}(\rho) + \lambda_B\ d_\alpha^{(\theta_1,\theta_2)}(2\rho)\right)\ (1+o_p(1))$$

$$\left. - \frac{1}{d_{2\alpha}^{(\theta_1,\theta_2)}(\rho)} \left(\frac{1}{2\gamma}\ \lambda_A\ a_{2\alpha}^{(\theta_1,\theta_2,\tau)}(\rho) + \lambda_B\ d_{2\alpha}^{(\theta_1,\theta_2)}(2\rho)\right)\ (1+o_p(1))\right).$$

We may thus say that $\sqrt{k}\ A(n/k)\left(\widehat{\beta}^{(\alpha,\theta_1,\theta_2)}(k;\ \rho) - \beta\right)$ is asymptotically normal, with a bias given by

$$\lambda_A u_\alpha^{(\theta_1,\theta_2,\tau)}(\gamma,\rho,\beta) + \lambda_B v_\alpha^{(\theta_1,\theta_2)}(\rho,\beta),$$

where $u_\alpha^{(\theta_1,\theta_2,\tau)}(\gamma,\rho,\beta)$ and $v_\alpha^{(\theta_1,\theta_2)}(\rho,\beta)$ are given in (3.1) and (3.2), respectively.

From the asymptotic covariance between $\overline{\sigma}^{(1)}_{\alpha\theta_1} P_n^{(\alpha\theta_1)}$ and $\overline{\sigma}^{(1)}_{\alpha\theta_2} P_n^{(\alpha\theta_2)}$ (see Gomes and Martins, 2002), given by

$$\frac{\alpha(\theta_1+\theta_2)\Gamma(\alpha(\theta_1+\theta_2))}{\alpha^2\theta_1\theta_2\Gamma(\alpha\theta_1)\Gamma(\alpha\theta_2)} - 1,$$

we easily derive the asymptotic covariance between $W_\alpha^{(\theta_1,\theta_2)}$ and $W_{2\alpha}^{(\theta_1,\theta_2)}$, which is given by the expression $\sigma_{\alpha,\theta_1,\theta_2}$ in (3.4). The asymptotic variance of $W_\alpha^{(\theta_1,\theta_2)}$ is the function $\sigma^2_{W|\alpha,\theta_1,\theta_2}$ given in (3.5). Consequently, (3.3) holds, and after a few simplications we arrive at (3.6) as well.

If we estimate ρ by means of $\widehat{\rho}$, we may write,

$$\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \widehat{\rho}) \ = \ \beta^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \rho) - \beta\,(\widehat{\rho}-\rho)\ln(n/k)(1+o_p(1)).$$

Consequently, under the conditions in the theorem, i.e., if $(\widehat{\rho}-\rho)\ln(n/k)$ $= o_p\left(1/\left(\sqrt{k}\ A(n/k)\right)\right)$, $\sqrt{k}\ \ A(n/k)\left(\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \widehat{\rho})-\beta\right)$ is also, whenever $\sqrt{k}\ \ A^2(n/k)$ and $\sqrt{k}\ \ A(n/k)B(n/k)$ converge towards finite limiting values, asymptotic normal with the same asymptotic variance and bias of $\sqrt{k}\ A(n/k)\left(\widehat{\beta}^{(\alpha,\theta_1,\theta_2,\tau)}(k;\ \rho)-\beta\right)$. □

References

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

[2] Draisma, G., Haan, L. de, Peng, L. and T. T. Pereira (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes* **2**:4, 367-404.

[3] Fraga Alves, M. I., Gomes, M. I. and L. de Haan (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* **60**:2, 193-214.

[4] Fraga Alves, M. I., de Haan, L. and Lin, T. (2003b). Estimation of the parameter controlling the speed of convergence in Extreme Value Theory. *Mathem. Methods of Statist.* **12**:2, 155-176.

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

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

[7] 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.

[8] 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.

[9] Gomes, M. I. and Pestana (2004). A simple second order reduced bias' tail index estimator. Accepted at *J. Satist. Comp. and Simul.*

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