Reduced bias' estimators: an overview*

M. Ivette Gomes

Universidade de Lisboa, FCUL (DEIO) and CEAUL

1 Introduction

The semi-parametric estimators of first order parameters of extreme or even rare events, like the tail index, the extremal index, an high quantile of probability p, a return period of a high level, and so on, are often based on the k top order statistics in the sample, $X_{n-i+1:n},\ 1 \leq i \leq k$. They are weakly consistent for the estimation of an extreme events' parameter, say ξ, in the whole domain of attraction of EV_γ, whenever k is *intermediate*, i.e.,

$$k = k_n \to \infty \quad \text{and} \quad k = o(n), \text{ as } n \to \infty. \tag{1.1}$$

We shall assume here that we are working with heavy tails, i.e., the *tail function*, $\overline{F} := 1 - F$, is a regularly varying function with a negative index of regular variation equal to $\{-1/\gamma\}$, $\gamma > 0$, or equivalently, the quantile function $U(t) = F^{\leftarrow}(1-1/t),\ t \geq 1$, with $F^{\leftarrow}(x) = \inf\{y : F(y) \geq x\}$, is of regular variation with index γ (Gnedenko, 1943; de Haan, 1970). This means that, for every $x > 0$,

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-1/\gamma}, \quad \text{or equivalently,} \quad \lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\gamma}. \tag{1.2}$$

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

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

*Research partially supported by FCT / POCTI / FEDER.

with $\gamma > 0$, and we write, as usual, $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$. We shall here concentrate on these Pareto-type distributions, for which (1.2) holds true. Note that (1.2) is equivalent to say that

$$1 - F(x) = x^{-1/\gamma}\, L_F(x), \quad \text{or equivalently,} \quad U(x) = x^{\gamma}\, L_U(x), \tag{1.3}$$

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

Under the general first order framework in (1.2), the asymptotic normality of this type of estimators is attained whenever we assume a second order condition, i.e., when we assume to know the rate of convergence towards zero of, for instance, $\{\ln U(tx) - \ln U(t) - \gamma \ln x\}$, i.e., the rate of convergence in the first order condition in (1.2). Such a second order condition may be written as,

$$\frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} \underset{t\to\infty}{\longrightarrow} \frac{x^\rho - 1}{\rho}, \tag{1.4}$$

where $\rho \le 0$, and $|A| \in RV_\rho$ (Geluk and de Haan, 1987).

For any classical semi-parametric estimator $\widehat{\xi}_n(k)$, weakly consistent for the estimation of ξ, and under the second order framework in (1.4), there exists a function $\varphi(k)$, converging towards zero as $k \to \infty$, such that we may write the asymptotic distributional representation,

$$\widehat{\xi}_n(k) \overset{d}{=} \xi + \sigma\, \varphi(k)\, P_k + b_\rho A(n/k)\, (1 + o_p(1))\,, \tag{1.5}$$

where P_k is asymptotically standard normal, $\sigma > 0$, $b_\rho \in \mathbb{R}$, $b_\rho \neq 0$, being $A(\cdot)$ the function in (1.4). We may thus provide approximations for the variance and the bias of $\widehat{\xi}_n(k)$ given by $(\sigma\varphi(k))^2$ and $b_\rho A(n/k)$, respectively. Consequently, the pattern of these estimators exhibit the same type of peculiarity:

- high variance for high thresholds $X_{n-k:n}$, i.e., for small values of k;
- high bias for low thresholds, i.e., for large values of k;

- a small region of stability of the sample path (plot of the estimates versus k), as a function of k, making problematic the adaptive choice of the threshold, on the basis of any sample paths' stability criterion;

- a "very peaked" mean squared error, making then difficult the choice of the value k_0 where that mean squared error function, $MSE\left[\widehat{\xi}_n(k)\right]$, attains its minimum value.

This has led researchers to consider the possibility of dealing with the bias term in an appropriate way, building new estimators, $\widehat{\xi}_n^{\,R}(k)$ say, the so-called *reduced bias'* estimators:

Definition 1.1. *Under the second order framework in* (1.4) *and for k intermediate, i.e., whenever* (1.1) *holds, the statistic $\widehat{\xi}_n^{\,R}(k)$, a consistent estimator of a functional of extreme events $\xi = \xi(F)$, based on the k top o.s. in a sample from a heavy-tailed model F, is said to be a reduced bias' semi-parametric estimator of ξ, whenever we may write,*

$$\widehat{\xi}_n^{\,R}(k) \;\stackrel{d}{=}\; \xi + \sigma_R \varphi(k) P_k^{R} + o_p(A(n/k)), \tag{1.6}$$

with P_k^R asymptotically normal, and where $\sigma_R > 0$, $\varphi(k) \to 0$, as $n \to \infty$, being $A(.)$ the function in (1.4)*, controlling the speed of convergence of maximum values, linearly normalized, towards a non-degenerate r.v.*

Remark 1.1. *Note that for the reduced bias' estimators in Definition 1.1, we no longer have a dominant component of bias of the order of $A(n/k)$, as in* (1.5)*. Then, $\sqrt{k}\left(\widehat{\xi}_n^{\,R}(k) - \xi\right)$ is asymptotically normal with null mean value not only when $\sqrt{k}\ A(n/k) \to 0$* (*as for the classical estimators*)*, but also when $\sqrt{k}\ A(n/k) \to \lambda$, finite and non-null.*

Such a bias reduction provides usually a stable sample path for a wide region of k-values, a reduction of the mean squared error at the optimal level and a reduction of the sensitivity to changes in location, one of the "non-properties" of most of the semi-parametric estimators of extreme events' parameters.

Such a bias reduction has been done in the most diversified ways, and from now on we shall restrict ourselves to the tail index estimation, i.e., we shall replace the generic parameter ξ by the tail index γ, in (1.2), being then $\varphi(k) = 1/\sqrt{k}$. The key ideas are either to find ways of getting rid of the dominant component of bias in (1.5), or to go further into the second order behaviour of the basic statistics used for the estimation of γ, like the log-excesses or the scaled log-spacings.

Bias corrected estimators may be dated back to Reiss (1989), Gomes (1994), Drees (1996) and Peng (1998), among others. Gomes (1994) uses the Generalized Jackknife methodology in Gray and Schucany (1972), and Peng (1998) deals with linear combinations of adequate tail index estimators, in a spirit quite close to the one associated to the Generalized Jackknife technique. This technique has also been used in Martins *et al.* (1999), where convex mixtures of two Hill's estimators, computed at two different levels, are considered, and later on, in Gomes *et al.* (2000, 2002b). These authors suggest, just as had already been done in Feuerverger and Hall (1999), the possible misspecification of the second order parameter ρ at -1, a value that corresponds to many commonly used heavy-tailed models, like the Fréchet model, $F(x) = \exp\left(-x^{-1/\gamma}\right),\ x \geq 0\ (\gamma > 0)$. Different approaches to reduced bias' tail index estimation, related to specifically built classes of estimators, have been developed in Gomes and Martins (2001) and Caeiro and Gomes (2002a,b). Those classes depend on a *tuning* or *control* parameter, and a specific choice of that parameter, which depends on the second-order parameter ρ, enables the practitioner to get a reduced bias' estimator.

Under the second set-up, Beirlant *et al.* (1999) and Feuerverger and Hall (1999) consider the accommodation of bias in the scaled log-spacings and derive "Maximum Likelihood" (ML) and "Least Squares" (LS) reduced bias' tail index estimators. The new classes of ρ-estimators, introduced in Gomes *et al.* (2002a) and in Fraga Alves *et al.* (2003), enable Gomes and Martins (2002) to use, successfully, this second class of estimators in the Generalized Jackknife estimators provided by Gomes *et al.* (2000). Gomes and Martins (2002) consider also the accommodation of bias in the scaled log-spacings, in a way similar to the one in Feuerveger and Hall (1999), but together with the use of the first order approximation for $\exp(-x)$, as $x \to 0$. Such an approximation enables them to obtain an explicit tail index estimator, dependent on an "external" estimation of ρ — it is there advised the use of a consistent ρ-estimator, to be computed at a higher level than the one used for the tail index estimation. Comparatively to the joint estimation of the first and second order parameters at the same level k, this procedure enables these authors to reduce the asymptotic variance of the reduced bias' estimator. Gomes *et al.* (2003) develop "asymptotically best linear combinations" of scaled log-spacings, both under a misspecification $\rho = -1$ and for a general ρ, to be estimated from our sample under an adequate methodology. More recently, Gomes *et al.* (2004b) and Caeiro *et al.* (2004) consider the external estimation of both the shape parameter and the scale parameter (or functional) in the function controlling the speed of convergence of maximum values, linearly normalized, towards a non-degenerate limiting r.v., being then able to further decrease the asymptotic variance of the reduced bias's estimators, which attains the value γ^2, the same we get for the Hill estimator.

Reduced bias' quantile estimators have been studied in Figueiredo and Gomes (2003) and Mattys and Beirlant (2003).

2 The Jackknife and related methodologies

The Jackknife methodology (Tukey, 1958) is a non-parametric resampling technique, essentially in the field of exploratory data analysis, whose main objective is the reduction of bias of an estimator, by means of the construction of an auxiliary estimator based on Quenouille's resampling technique (Quenouille, 1956), and the consideration of a suitable combination of the two estimators. The Generalized Jackknife statistics of Gray and Schucany (1972) are, more generally based on two different estimators of the same functional, with similar bias' properties. More precisely, and as a particular case of the Jackknife theory, if we have two different biased consistent estimators $\widehat{\xi}_n^{(1)}$ and $\widehat{\xi}_n^{(2)}$ of the functional $\xi(F)$, such that $\mathbb{E}\left[\widehat{\xi}_n^{(1)}\right] = \xi + \varphi(\xi)\ d_1(n)$ and $\mathbb{E}\left[\widehat{\xi}_n^{(2)}\right] = \xi + \varphi(\xi)\ d_2(n)$, then, denoting by

$$q_n := \frac{BIAS\left[\widehat{\xi}_n^{(1)}\right]}{BIAS\left[\widehat{\xi}_n^{(2)}\right]} = \frac{d_1(n)}{d_2(n)},$$

the Generalized Jackknife statistic associated to $\left(\widehat{\xi}_n^{(1)}, \widehat{\xi}_n^{(2)}\right)$ is

$$\widehat{\xi}_n^G\left(\widehat{\xi}_n^{(1)}, \widehat{\xi}_n^{(2)}\right) = \frac{\widehat{\xi}_n^{(1)} - q_n \widehat{\xi}_n^{(2)}}{1 - q_n},$$

which is an unbiased consistent estimator of $\xi(F)$, provided that $q_n \neq 1$ for all n.

In *Statistics of Extremes*, whenever we are dealing with semi-parametric estimators of the tail index, or even other parameters of extreme events, we have usually information about the asymptotic bias of those estimators. We may thus choose estimators with similar asymptotic properties, and build the associated Generalized Jackknife statistic. This methodology has first been used by Gomes (1994), later developed in Martins *et al.* (1999), where convex mixtures of two Hill's estimators, computed at two different levels, are considered, and still later on, in Gomes *et al.* (2000, 2002b). Gomes and Martins (2000) suggest several *Generalized Jackknife* estimators of the tail index γ. We shall here refer only, the one based on the classical Hill's estimator for γ (Hill, 1975),

$$\widehat{\gamma}_n^{(1)}(k) \equiv H(k) := \frac{1}{k}\sum_{i=1}^{k}\left[\ln X_{n-i+1:n} - \ln X_{n-k:n}\right], \qquad (2.1)$$

and on the alternative estimator

$$\widehat{\gamma}_n^{(2)}(k) := \frac{M_n^{(2)}(k)}{2M_n^{(1)}(k)}, \tag{2.2}$$

where

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k} \left[\ln X_{n-i+1:n} - \ln X_{n-k:n}\right]^j, \ j \geq 1. \tag{2.3}$$

Under the second order framework in (1.4), and with $P_k^{(1)}$ and $P_k^{(2)}$ asymptotically standard normal r.v.'s, we have the validity of the following asymptotic distributional representations:

$$\widehat{\gamma}_n^{(1)}(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, P_k^{(1)} + \frac{A(n/k)}{1-\rho} + o_p(A(n/k)), \tag{2.4}$$

$$\widehat{\gamma}_n^{(2)}(k) \stackrel{d}{=} \gamma + \frac{\sqrt{2}\,\gamma}{\sqrt{k}}\, P_k^{(2)} + \frac{A(n/k)}{(1-\rho)^2} + o_p(A(n/k)). \tag{2.5}$$

Remark 2.1. *Note that we may choose* $P_k^{(1)} = \sqrt{k}\left(\sum_{i=1}^{k} E_i/k - 1\right)$ *and* $P_k^{(2)} = \frac{\sqrt{2}}{2}\left(\frac{\sqrt{k}}{2}\left(\sum_{i=1}^{k} E_i^2/k - 2\right) - P_k^{(1)}\right)$*, where* $E_i,\ i \geq 1$ *are i.i.d. standard exponential r.v.'s. Consequently,* $Cov\left(P_k^{(1)}, P_k^{(2)}\right) = \sqrt{2}/2$.

The quotient between the dominant components of bias of $\widehat{\gamma}_n^{(1)}$ and $\widehat{\gamma}_n^{(2)}$ is equal to $\{1-\rho\}$, and we thus have the *Generalized Jackknife* estimator

$$\widehat{\gamma}_{\widehat{\rho}}^{GJ}(k) := \frac{\widehat{\gamma}_n^{(1)}(k) - (1-\widehat{\rho})\ \widehat{\gamma}_n^{(2)}(k)}{\widehat{\rho}}, \tag{2.6}$$

an affine combination of the tail index estimators $\widehat{\gamma}_n^{(1)}$ and $\widehat{\gamma}_n^{(2)}$ in (2.1) and (2.2), respectively. Note that this estimator is exactly the estimator studied in Peng (1998), who claims that no good estimator for the second order parameter ρ was then available, and considers a new ρ-estimator, alternative to the ones in Hall and Welsh (1985), Beirlant *et al.* (1996) and Drees and Kaufmann (1997). It is then possible to state the following:

Theorem 2.1. *Under the second order framework in* (1.4) *for k intermediate, i.e., such that* (1.1) *holds, and with the obvious notation for $\widehat{\gamma}_\rho^{GJ}(k)$,*

$$\widehat{\gamma}_\rho^{GJ}(k) \stackrel{d}{=} \gamma + \frac{\gamma\sqrt{2\rho^2-2\rho+1}}{|\rho|\,\sqrt{k}}\; P_k^{GJ} + o_p(A(n/k)), \tag{2.7}$$

where P_k^{GJ} is an asymptotically standard normal r.v. The result in (2.7) *remains true for the Generalized Jackknife estimator $\widehat{\gamma}_{\widehat{\rho}}^{GJ}(k)$ in* (2.6)*, provided that $\widehat{\rho}-\rho = o_p(1)$ for all k on which we base the tail index estimation, i.e., whenever $\sqrt{k}\ A(n/k) \to \lambda$, finite, possibly non-null,*

$$\sqrt{k}\left(\widehat{\gamma}_{\widehat{\rho}}^{GJ}(k) - \gamma\right) \underset{n\to\infty}{\overset{w}{\longrightarrow}} Normal\left(0, \sigma^2_{_{GJ}} = \frac{\gamma^2(2\rho^2-2\rho+1)}{\rho^2}\right). \tag{2.8}$$

Proof. The result in (2.7) comes directly from the expression of $\widehat{\gamma}_\rho^{GJ}(k)$, together with the distributional representations in (2.4) and (2.5) and the results in Remark 2.1. The last part of the theorem comes from the fact that

$$\frac{d\widehat{\gamma}_\rho^{GJ}(\rho)}{d\rho} = \frac{\widehat{\gamma}_n^{(2)}(k) - \widehat{\gamma}_n^{(1)}(k)}{\rho^2} = O_p\left(\frac{1}{\sqrt{k}}\right) + O_p\left(A(n/k)\right)$$

and

$$\widehat{\gamma}_{\widehat{\rho}}^{GJ}(k) \stackrel{d}{=} \widehat{\gamma}_\rho^{GJ}(k) + (\widehat{\rho}-\rho)\left(O_p\left(\frac{1}{\sqrt{k}}\right) + O_p\left(A(n/k)\right)\right)(1+o_p(1)). \tag{2.9}$$

□

Remark 2.2. *A closer look at* (2.9) *reveals that it does not seem convenient to compute $\widehat{\rho}$ at the same level k we use for the tail index estimation. Indeed, if we do that, and since we have usually $\widehat{\rho}(k) - \rho = O_p\left(1/\left(\sqrt{k}\ A(n/k)\right)\right)$, $\{\widehat{\rho}(k)-\rho\}$ is not a $o_p(1)$ whenever $\sqrt{k}\ A(n/k) \to \lambda$, finite. We are then going to have a change in the asymptotic variance of the tail index estimator, because $(\widehat{\rho}(k)-\rho)\,A(n/k)$ is a term of the order of $1/\sqrt{k}$.*

Remark 2.3. *Gomes et al.* (2000) *have indeed suggested the misspecification of ρ at $\rho = -1$, and the consideration of the estimator* $\widehat{\gamma}^{GJ}(k) := 2\,\widehat{\gamma}_n^{(1)}(k) - \widehat{\gamma}_n^{(1)}(k)$, *which is a reduced bias' estimator, in the sense herewith defined, i.e., in the sense of* (1.6), *if and only if* $\rho = -1$. *This was essentially due to the high bias and variance of the exhisting estimators of ρ at that time, together with the idea of considering* $\widehat{\rho} = \widehat{\rho}(k)$.

2.1 The estimation of ρ

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

$$\widehat{\rho}_\tau(k) \equiv \widehat{\rho}_n^{(\tau)}(k) := -\left|\frac{3(T_n^{(\tau)}(k) - 1)}{T_n^{(\tau)}(k) - 3}\right|, \tag{2.10}$$

where

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

with $M_n^{(j)}(k)$ given in (2.3).

The statistics $\rho_n^{(\tau)}(k)$ in (2.10) converge towards ρ, for every τ, whenever the second order condition (1.4) holds and k is such that $k \to \infty$, $k = o(n)$ and

$\sqrt{k}\ A(n/k) \to \infty$, as $n \to \infty$. We shall here summarize a particular case of the results proved in Fraga Alves *et al.* (2003), now related to the asymptotic behaviour of the ρ-estimator in (2.10), under the second order framework in (1.4):

Proposition 2.1. *Under the second order framework in* (1.4)*, with* $\rho < 0$*, if* (1.1) *holds, and if* $\sqrt{k}\ A(n/k) \to \infty$*, as* $n \to \infty$*, the statistic* $\widehat{\rho}_n^{(\tau)}(k)$ *in* (2.10) *converges in probability towards* ρ*, as* $n \to \infty$*, for any real* $\tau \geq 0$*, i.e., we have* $\widehat{\rho}_n^{(\tau)}(k) - \rho = O_p\left(1/\left(\sqrt{k}\ A(n/k)\right)\right)$*. More specifically,*

$$\widehat{\rho}_n^{(\tau)}(k) - \rho \stackrel{d}{=} \frac{\sigma_{\widehat{\rho}} W_k^{*(\rho)}}{\sqrt{k}\ A(n/k)} + R_k^{(\rho)}, \quad \text{with} \quad R_k^{(\rho)} = o_p(1) \tag{2.11}$$

and

$$\sigma_{\widehat{\rho}}^2 \equiv \sigma_\rho^2(\gamma) = \left(\frac{\gamma(1-\rho)^3}{\rho}\right)^2 \left(2\rho^2 - 2\rho + 1\right). \tag{2.12}$$

Consequently, if $\sqrt{k}\ A(n/k)\ R_k^{(\rho)} \to 0$*,* $\sqrt{k}\ A(n/k)\left(\widehat{\rho}_n^{(\tau)}(k) - \rho\right)$ *is asymptotically normal with null mean value and variance equal to* σ_ρ^2*.*

Remark 2.4. *The theoretical and simulated results in Fraga Alves et al.* (2003)*, together with the use of these estimators in the Generalized Jackknife statistics of Gomes et al.* (2000)*, as done in Gomes and Martins* (2002)*, as well as their use in the estimators in Gomes et al.* (2004*b*) *and Caeiro et al.* (2004)*, lead us to advise in practice the consideration of the level*

$$k_1 = \min\left(n-1, [2n/\ln\ln n]\right) \tag{2.13}$$

(*not chosen in any optimal way*)*, and of the tuning parameters* $\tau = 0$ *for the region* $\rho \in [-1, 0)$ *and* $\tau = 1$ *for the region* $\rho \in (-\infty, -1)$*. As done before, we however advise practitioners not to choose blindly the value of* τ *in* (2.10)*. It is sensible to draw a few sample paths of* $\widehat{\rho}_n^{(\tau)}(k)$ *in* (2.10)*, as functions of* k*, electing the value of* τ *which provides higher stability for large* k*, by means of* *any stability criterion. Anyway, in Monte Carlo simulations we have considered the level* k_1 *in* (2.13) *and the* ρ*-estimators*

$$\widehat{\rho}_0 := -\left|\frac{3\left(T_n^{(0)}(k_1)-1\right)}{T_n^{(0)}(k_1)-3}\right| \quad \textit{if} \quad \rho \geq -1, \tag{2.14}$$

and

$$\widehat{\rho}_1 := -\left|\frac{3\left(T_n^{(1)}(k_1)-1\right)}{T_n^{(1)}(k_1)-3}\right| \quad \textit{if} \quad \rho < -1. \tag{2.15}$$

Remark 2.5. *When we consider the level* k_1 *in* (2.13)*, together with any of the* ρ*-estimators in this section, computed at the level,* k_1*,* $\{\widehat{\rho}-\rho\}$ *is of the order of* $1/\left(\sqrt{k_1}A(n/k_1)\right) = O\left((\ln_2 n)^{(1-2\rho)/2}/\sqrt{n}\right)$*, with the obvious notation* $\ln_2 n = \ln\ln n$*.*

Remark 2.6. *Moreover, for any level* k*,* $(\widehat{\rho}-\rho)\ln(n/k) = o_p(1)$*, and consequently* $\sqrt{k}\ A(n/k)\,(\widehat{\rho}-\rho)\ln(n/k) = o_p(1)$ *whenever* $\sqrt{k}\ A(n/k) \to \lambda$*, finite.*

3 Specifically built reduced bias classes of estimators

Gomes and Martins (2001) generalize the statistic in (2.2), and consider the class of estimators

$$\widehat{\gamma}_n^{(\alpha)}(k) := \frac{M_n^{(\alpha)}(k)}{\Gamma(\alpha+1)\left[M_n^{(1)}(k)\right]^{\alpha-1}}, \tag{3.1}$$

dependent on the *tuning* or *control* parameter α.

In a quite similar way, Caeiro and Gomes (2002a) generalize the possible tail index estimator $\sqrt{M_n^{(2)}(k)/2}$, studied in Gomes *et al.* (2000), considering the class of estimators

$$\widetilde{\gamma}_n^{(\alpha)}(k) := \frac{\Gamma(\alpha)}{M_n^{(\alpha-1)}(k)}\left(\frac{M_n^{(2\alpha)}(k)}{\Gamma(2\alpha+1)}\right)^{1/2}. \tag{3.2}$$

We have:

Theorem 3.1. *Under the validity of the first order condition in* (1.2) *and for k intermediate, i.e., a k-value such that* (1.1) *holds, the two classes of statistics in* (3.1) *and in* (3.2) *converge in probability towards γ. If we further assume the general second order framework in* (1.4)*, the asymptotic distributional representations*

$$\widehat{\gamma}_n^{(\alpha)}(k) \stackrel{d}{=} \gamma + +\frac{\gamma\sqrt{\widehat{v}(\alpha)}}{\sqrt{k}}\ \widehat{Z}_k^{(\alpha)} + \widehat{b}(\rho,\alpha)A(n/k) + o_p(A(n/k))$$

and

$$\widetilde{\gamma}_n^{(\alpha)}(k) \stackrel{d}{=} \gamma + \frac{\gamma\sqrt{\widetilde{v}(\alpha)}}{\sqrt{k}}\ \widetilde{Z}_k^{(\alpha)} + \widetilde{b}(\rho,\alpha)A(n/k) + o_p(A(n/k)),$$

hold true, where $\widehat{Z}_k^{(\alpha)}$ and $\widetilde{Z}_k^{(\alpha)}$ are asymptotically standard normal,

$$\widehat{v}(\alpha) = \frac{\Gamma(2\alpha+1)}{\Gamma^2(\alpha+1)} - \alpha^2,$$

$$\widetilde{v}(\alpha) = \frac{1}{4}\left\{\frac{\Gamma(4\alpha)}{\alpha\Gamma^2(2\alpha)} + \frac{4\Gamma(2\alpha-1)}{\Gamma^2(\alpha)} - \frac{2\Gamma(3\alpha)}{\alpha\Gamma(\alpha)\Gamma(2\alpha)} - 1\right\},$$

and, for $\rho < 0$,

$$\widehat{b}(\rho,\alpha) = \frac{1-(1-\rho)^\alpha}{\rho(1-\rho)^\alpha} - \frac{\alpha-1}{1-\rho}, \tag{3.3}$$

$$\widetilde{b}(\rho,\alpha) = \frac{1}{2\rho}\left\{(1-\rho)^{-2\alpha} - 2(1-\rho)^{-\alpha+1} + 1\right\}, \tag{3.4}$$

being $\widehat{b}(\rho,\alpha) = \widetilde{b}(\rho,\alpha) = 1$ if $\rho = 0$.

Remark 3.1. *Note that the bias function $\widehat{b}(\rho,\alpha)$ in* (3.3) *has a unique zero, at the value α_0 such that $(1-\rho)^{\alpha_0-1}\left[1+\rho(\alpha_0-2)\right] = 1$, i.e., for an adequate choice of α, $\widehat{\gamma}_n^{(\alpha)}$ in* (3.1) *is a reduced bias' tail index estimator. Also, for every $\rho < 0$, there exists a value α_0, now explicitly given by $\alpha_0 = -\frac{\ln\left[1-\rho-\sqrt{(1-\rho)^2-1}\right]}{\ln(1-\rho)}$, such that $\widetilde{b}(\rho,\alpha_0) = 0$, i.e., such that $\widetilde{\gamma}_n^{(\alpha)}$ in* (3.2) *is a reduced bias' tail index estimator.*

Let us denote $\widehat{\rho}$ any consistent estimator of ρ. It is then sensible to consider the reduced bias' tail index estimators,

$$\widehat{\gamma}_{\widehat{\rho}}^{(\widehat{\alpha})}(k) := \frac{M_n^{(\widehat{\alpha})}(k)}{\Gamma(\widehat{\alpha}+1)\left[M_n^{(1)}(k)\right]^{\widehat{\alpha}-1}}, \quad \widehat{\alpha}: \ (1-\widehat{\rho})^{\widehat{\alpha}-1}[1+\widehat{\rho}(\widehat{\alpha}-2)] = 1. \tag{3.5}$$

and

$$\widetilde{\gamma}_{\widehat{\rho}}^{(\widehat{\alpha})}(k) := \frac{\Gamma(\widehat{\alpha})}{M_n^{(\widehat{\alpha}-1)}(k)}\left(\frac{M_n^{(2\widehat{\alpha})}(k)}{\Gamma(2\widehat{\alpha}+1)}\right)^{1/2}, \quad \widehat{\alpha} = -\frac{\ln\left[1-\widehat{\rho}-\sqrt{(1-\widehat{\rho})^2-1}\right]}{\ln(1-\widehat{\rho})}. \tag{3.6}$$

Remark 3.2. *The estimator in* (3.6) *has been studied under a third order framework in Caeiro et al.* (2004).

The asymptotic variance of the estimator in (3.5), i.e., the square of the value $\sigma_{_R}$, in (1.6), associated to this estimator, is thus

$$\sigma^2_{_{GM_1}} \equiv \sigma^2_{_{GM_1}}(\rho) = \gamma^2 \left(\frac{\Gamma(2\widehat{\alpha}+1)}{\Gamma^2(\widehat{\alpha}+1)} - \widehat{\alpha}^2\right), \tag{3.7}$$

with

$$\widehat{\alpha}: \quad (1-\rho)^{\widehat{\alpha}-1}[1+\rho(\widehat{\alpha}-2)] = 1.$$

The asymptotic variance of the estimator in (3.6) is ruled by

$$\sigma^2_{_{CG}} \equiv \sigma^2_{_{CG}}(\rho) = \frac{\gamma^2}{4}\left\{\frac{\Gamma(4\widetilde{\alpha})}{\widetilde{\alpha}\Gamma^2(2\widetilde{\alpha})} + \frac{4\,\Gamma(2\widetilde{\alpha}-1)}{\Gamma^2(\widetilde{\alpha})} - \frac{2\,\Gamma(3\widetilde{\alpha})}{\widetilde{\alpha}\Gamma(\widetilde{\alpha})\Gamma(2\widetilde{\alpha})} - 1\right\}, \tag{3.8}$$

with

$$\widetilde{\alpha} = \widetilde{\alpha}(\rho) = -\frac{\ln\left[1-\rho-\sqrt{(1-\rho)^2-1}\right]}{\ln(1-\rho)}.$$

4 Accommodation of bias in the scaled log-spacings

Let us consider the scaled log-spacings

$$U_i := i\left(\ln X_{n-i+1:n} - \ln X_{n-i:n}\right), \quad 1 \le i \le k. \tag{4.1}$$

Under the second order framework in (1.4), and for $\rho < 0$, Beirlant *et al.* (1999) motivated the following approximation for the scaled log-spacings:

$$U_i \sim \left(\gamma + A(n/k)\left(\frac{i}{k}\right)^{-\rho}\right) E_i, \quad 1 \le i \le k, \tag{4.2}$$

where $\{E_i\}$, $i \ge 1$, denotes a sequence of i.i.d., standard exponential r.v.'s. In the same context, Feuerverger and Hall (1999), consider the approximation,

$$U_i \sim \gamma \exp\left(\frac{A(n/k)}{\gamma}\left(\frac{i}{k}\right)^{-\rho}\right) E_i = \gamma \exp\left(\frac{A(n/i)}{\gamma}\right) E_i, \quad 1 \le i \le k. \tag{4.3}$$

The representation (4.2), or equivalently (4.3), has been made more precise, in the asymptotic sense, in Beirlant *et al.* (2002), in a way quite close in spirit to the approximations established by Kaufmann and Reiss (1998) and Drees *et al.* (2000).

We shall here further assume, just as in Feuerverger and Hall (1999), that we are in Hall's class of Pareto-type models (Hall, 1982; Hall and Welsh, 1985), with a tail function

$$1 - F(x) = Cx^{-1/\gamma}\left(1 + Dx^{\rho/\gamma} + o\left(x^{\rho/\gamma}\right)\right), \quad \text{as} \quad x \to \infty, \tag{4.4}$$

$C > 0$, $D \in \mathbb{R}$, $\rho < 0$. Then, (1.4) holds and we may choose

$$A(t) = \alpha\, t^\rho =: \gamma\, \beta\, t^\rho, \quad \beta \ne 0, \quad \rho < 0. \tag{4.5}$$

Notice however that β may be regarded as a slowly varying function, and not as a constant.

4.1 The ML estimation based on the scaled log-spacings

The use of the approximation in (4.3) and the joint maximization, in γ, β and ρ, of the approximate log-likelihood of the scaled log-spacings, i.e., of

$$\ln L(\gamma, \beta, \rho; U_i, 1 \le i \le k) = -k \ln \gamma - \beta \sum_{i=1}^{k} \left(\frac{i}{n}\right)^{-\rho} - \frac{1}{\gamma} \sum_{i=1}^{k} e^{-\beta\left(\frac{i}{n}\right)^{-\rho}}\ U_i,$$

led Feuerverger and Hall (1999) to an explicit expression for $\widehat{\gamma}$, as a function of $\widehat{\beta}$ and $\widehat{\rho}$, given by

$$\widehat{\gamma}(k) \equiv \widehat{\gamma}^{FH}_{\widehat{\beta},\widehat{\rho}}(k) := \frac{1}{k} \sum_{i=1}^{k}\ e^{-\widehat{\beta}\left(\frac{i}{n}\right)^{-\widehat{\rho}}}\ U_i. \tag{4.6}$$

Then, $\widehat{\beta} \equiv \widehat{\beta}_{\widehat{\rho}(k)} = \widehat{\beta}^{FH}_{\widehat{\rho}(k)}(k)$ and $\widehat{\rho} \equiv \widehat{\rho}_{\widehat{\beta}(k)} = \widehat{\rho}^{FH}_{\widehat{\beta}(k)}(k)$ are numerically obtained, through

$$(\widehat{\beta}, \widehat{\rho}) := \underset{(\beta,\rho)}{\arg\min} \left\{ \ln \left(\frac{1}{k} \sum_{i=1}^{k} e^{-\beta\left(\frac{i}{n}\right)^{-\rho}}\ U_i \right) + \beta \left(\frac{1}{k} \sum_{i=1}^{k} \left(\frac{i}{n}\right)^{-\rho} \right) \right\}. \tag{4.7}$$

It is possible to state the following:

Theorem 4.1. *If we assume ρ to be known, if the second order condition* (1.4) *holds, if* (1.1) *holds and if $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$, then $\widehat{\beta} \equiv \widehat{\beta}^{FH}_{\rho} = \widehat{\beta}^{FH}_{\rho}(k)$, obtained through the minimization in* (4.7)*, is consistent for the estimation of β at a rate of convergence of the order of $1/\left(\sqrt{k}\ A(n/k)\right)$. Moreover, if $\sqrt{k}\ A^2(n/k) \underset{n\to\infty}{\longrightarrow} 0$ we have,*

$$\sqrt{k}\ A(n/k) \left\{ \frac{\widehat{\beta}^{FH}_{\rho}(k) - \beta}{\beta} \right\} \underset{n\to\infty}{\overset{w}{\longrightarrow}} Normal\left(0, \frac{\gamma^2(1-\rho)^2(1-2\rho)}{\rho^2}\right). \tag{4.8}$$

For the tail index estimator, we have then

$$\widehat{\gamma}^{FH}_{\widehat{\beta},\rho}(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} \left(\frac{1-\rho}{\rho}\right)\ \Gamma_k + o_p(A(n/k)), \tag{4.9}$$

where Γ_k is asymptotically standard normal.

4.2 Joint estimation of first and second order parameters

However, if ρ is unknown as well as β, as usually happens, and they are both estimated through (4.7), as suggested in Feuerverger and Hall (1999), we get:

Theorem 4.2. *Under the validity of the second order framework in* (1.4)*, and for k intermediate, such that* $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$, $\widehat{\rho} = \widehat{\rho}^{FH}_{\widehat{\beta}(k)}(k)$ *is consistent for the estimation of* ρ. *If we further have* $\sqrt{k}\ A(n/k)/\ln(n/k) \underset{n\to\infty}{\longrightarrow} \infty$, $\widehat{\beta} = \widehat{\beta}^{FH}_{\widehat{\rho}(k)}(k)$ *is consistent for the estimation of* β. *For any intermediate k, and also under the second order framework in* (1.4)*, the distributional representation*

$$\widehat{\gamma}^{FH}_{\widehat{\beta},\widehat{\rho}}(k) \ \stackrel{d}{=}\ \gamma + \gamma\left(\frac{1-\rho}{\rho}\right)^2 \frac{\Gamma^*_k}{\sqrt{k}} + o_p(A(n/k)) \tag{4.10}$$

holds true, where Γ^*_k *is an asymptotically standard normal r.v.*

Remark 4.1. *Note that, even when* $\sqrt{k}\ A(n/k) \to \lambda$*, non-null, we have an asymptotic normal behaviour for* $\widehat{\gamma}^{FH}_{\widehat{\beta},\widehat{\rho}}(k)$*, with a null asymptotic bias, but at the expenses of an asymptotic variance ruled by*

$$\sigma^2_{FH} = \gamma^2\left((1-\rho)/\rho\right)^4, \tag{4.11}$$

which is greater than the value $\gamma^2\left((1-\rho)/\rho\right)^2$ *in Theorem 4.1, which is on its turn greater than the value* γ^2*, for all* $\rho < 0$.

4.3 A simplified ML tail index estimator and the external estimation of ρ

The ML estimators of β and ρ in (4.7) are the solution of the ML system of equations:

$$\begin{cases} \widehat{b}_{10}\widehat{B}_{00} - \widehat{B}_{10} = 0 \\ \widehat{B}_{11} - \widehat{b}_{11}\widehat{B}_{00} = 0 \end{cases}, \tag{4.12}$$

where, for non-negative integers j and l,

$$\widehat{b}_{jl} \equiv \widehat{b}_{jl}(\widehat{\rho}) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{n}\right)^{-j\widehat{\rho}}\left(\ln\frac{i}{n}\right)^l$$

and

$$\widehat{B}_{jl} \equiv \widehat{B}_{jl}(\widehat{\rho},\widehat{\beta}) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{n}\right)^{-j\widehat{\rho}}\left(\ln\frac{i}{n}\right)^{l} e^{-\widehat{\beta}(i/n)^{-\widehat{\rho}}}\; U_i.$$

The first ML equation may then be written as,

$$\sum_{i=1}^{k} i^{-\widehat{\rho}} \exp\left(-\widehat{\beta}\,(i/n)^{\widehat{\rho}}\right)\; U_i = \widehat{\gamma}\left(\sum_{i=1}^{k} i^{-\widehat{\rho}}\right),$$

with $\widehat{\gamma}$ given in (4.6). The use the first order approximation, $e^x = 1+x$, as $x\to 0$, led Gomes and Martins (2002) to an explicit estimator for β, given by

$$\widehat{\beta}_{\widehat{\rho}}^{\,GM}(k) := \left(\frac{k}{n}\right)^{\widehat{\rho}} \frac{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^{k}U_i\right)-\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\;U_i\right)}{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\;U_i\right)-\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\widehat{\rho}}\;U_i\right)}, \tag{4.13}$$

and the following maximum likelihood estimator for the tail index γ,

$$\begin{aligned}\widehat{\gamma}_{\widehat{\rho}}^{\,GM}(k) := &\frac{1}{k}\sum_{i=1}^{k}U_i \\ &-\left(\frac{1}{k}\sum_{i=1}^{k} i^{-\widehat{\rho}}\;U_i\right)\frac{\left(\sum_{i=1}^{k} i^{-\widehat{\rho}}\right)\left(\sum_{i=1}^{k}U_i\right)-k\left(\sum_{i=1}^{k} i^{-\widehat{\rho}}U_i\right)}{\left(\sum_{i=1}^{k} i^{-\widehat{\rho}}\right)\left(\sum_{i=1}^{k} i^{-\widehat{\rho}}U_i\right)-k\left(\sum_{i=1}^{k} i^{-2\widehat{\rho}}U_i\right)},\end{aligned} \tag{4.14}$$

based on an adequate, consistent estimator for ρ.

Gomes and Martins (2002) and later, Gomes *et al.* (2004a), keeping up to the second order framework in (1.4), have proved the following result:

Theorem 4.3. *If the second order condition in* (1.4) *holds true, if $k = k_n$ is a sequence of intermediate positive integers, i.e.,* (1.1) *holds, and $\sqrt{k}\,A(n/k)\to\lambda$, finite, as $n\to\infty$, with $\widehat{\rho}$ any of the estimators in* (2.10) *computed at the level k_1 in* (2.13),

$$\sqrt{k}\left(\widehat{\gamma}_{\widehat{\rho}}^{\,GM}(k)-\gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal\left(0,\; \sigma^2_{GM_2} = \frac{\gamma^2(1-\rho)^2}{\rho^2}\right). \tag{4.15}$$

The limiting behaviour in (4.15) *holds more generally whenever* $\widehat{\rho} - \rho = o_p(1)$ *for any* k *on which we base the tail index estimation.*

If, with $A(t) = \gamma\ \beta\ t^\rho$*, as in* (4.5)*, we assume that* $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$*, then* $\widehat{\beta}_{\widehat{\rho}}^{\ GM}(k)$ *in* (4.13) *is consistent for the estimation of* β*, whenever* $\widehat{\rho}$ *is consistent for the estimation of* ρ*. Moreover, if* ρ *is known,*

$$\widehat{\beta}_{\rho}^{\ GM}(k) \stackrel{d}{=} \beta + \frac{\gamma\ \beta\ (1-\rho)\ \sqrt{1-2\rho}}{\rho\ \sqrt{k}\ A(n/k)}\ B_k\ + R_k^{(\beta)},\ \ with\ R_k^{(\beta)} = o_p(1), \tag{4.16}$$

where B_k *is asymptotically standard normal. More precisely we may write*

$$B_k = \frac{(1-\rho)\sqrt{1-2\rho}}{|\rho|}\left(\frac{Z_k^{(1)}}{1-\rho} - \frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}}\right), \tag{4.17}$$

with $Z_k^{(\alpha)},\ \alpha \geq 1$*, given by,*

$$Z_k^{(\alpha)} = \sqrt{(2\alpha-1)\ k}\ \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1} E_i - \frac{1}{\alpha}\right). \tag{4.18}$$

The asymptotic distributional representation (4.16) *holds true as well for* $\widehat{\beta}_{\widehat{\rho}}^{\ GM}(k)$*, with* $\widehat{\rho}$ *any of the estimators in* (2.10) *computed at the level* k_1 *in* (2.13)*. If* $\sqrt{k}\ A(n/k)\ R_k^{(\beta)} \to \lambda_R^{(\beta)}$*, finite, we may further guarantee the asymptotic normality of* $\widehat{\beta}_{\widehat{\rho}}^{\ GM}(k)$*, and* (4.8) *holds true, with* $\widehat{\beta}_{\rho}^{FH}(k)$ *replaced by* $\widehat{\beta}_{\widehat{\rho}}^{\ GM}(k)$*. If we consider* $\widehat{\beta}_{\widehat{\rho}(k)}^{\ GM}(k)$*, then*

$$\widehat{\beta}_{\widehat{\rho}(k)}^{\ GM}(k) - \beta\ \sim\ -\beta\ \ln(n/k)\ (\widehat{\rho}(k) - \rho). \tag{4.19}$$

Remark 4.2. *Note that when we consider the level* k_1 *in* (2.13)*, and* $\widehat{\beta} \equiv \widehat{\beta}_{\widehat{\rho}}^{\ GM}(k_1)$*, with* $\widehat{\rho}$ *any of the estimator in* (2.10)*, computed also at the same level* k_1*,* $\widehat{\beta} - \beta$ *is thus of the order of* $\ln(n/k_1)/\left(\sqrt{k_1}A(n/k_1)\right) = O\left(\ln_3 n\ (\ln_2 n)^{(1-2\rho)/2}/\sqrt{n}\right)$*.*

Remark 4.3. *Notice that the implicit maximum likelihood estimation of* γ *and* ρ *made at the same level* k*, as performed by Feuerverger and Hall* (1999)*, leads* *them to an asymptotic variance, ruled by* $\gamma^2\left((1-\rho)/\rho\right)^4$*, which is the square of the value exhibited in* (4.15). *Such an increase in variance is due to the simultaneous estimation of* γ *and* ρ *at the same level* k.

Remark 4.4. *Notice also that, as it is well-known from the literature, we may have a non-null asymptotic bias for any classical estimator* $\gamma_n(k)$*, like the Hill estimator in* (2.1)*, if the threshold* k *is such that* $\sqrt{k}\ A(n/k) \to \lambda \neq 0$. *It is indeed a sequence* $k_0 = k_0(n)$ *such that* $\sqrt{k_0}\ A(n/k_0) \to \varphi(\rho) \neq 0$*, the one which provides a minimum mean squared error of such a classical estimator. From Theorem 4.3, even when* $\sqrt{k}\ A(n/k) \to \lambda \neq 0$*, finite and non-null, do we have a null mean value for the limit of* $\sqrt{k}\left(\widehat{\gamma}_{\widehat{\rho}}^{GM}(k) - \gamma\right)$*, whenever we consider the semi-parametric* ρ*-estimators in Gomes et al.* (2002*a*) *or in Fraga Alves et al.* (2003)*, computed at an adequate fixed level* k_1 *such that* $\sqrt{k_1}\ A(n/k_1) \underset{n\to\infty}{\longrightarrow} \infty$*, so that we have* $\widehat{\rho} - \rho = o_p(1)$*, for every* k *on which we are going to base the estimation of the tail index* γ.

Remark 4.5. *If we further work with a third order expansion, like the one used in Gomes and de Haan* (1999)*, Gomes et al.* (2002*a*)*, Fraga Alves et al.* (2003) *and Caeiro et al.* (2004)*, we may see that for a large class of models in Hall's class, more precisely, for models with a tail function*

$$1 - F(x) = Cx^{-1/\gamma}\left(1 + D_1 x^{\rho/\gamma} + D_2 x^{2\rho/\gamma} + o(x^{2\rho/\gamma})\right), \ \text{as } x \to \infty, \qquad (4.20)$$

the minimum mean squared error of any of the reduced bias' tail index estimators is attained further in the tail. Indeed, it is then attained whenever $\sqrt{k}\ A^2(n/k) \underset{n\to\infty}{\longrightarrow} \lambda \neq 0$*, finite, such as happens for the* ρ*-estimators in Gomes et al.* (2002*a*) *and in Fraga Alves et al.* (2003). *Since the remainder* $o_p\left(\sqrt{k}\ A(n/k)\right)$ *in the distributional representation of these* ρ*-estimators, is then of the order of* $\sqrt{k}\ A^2(n/k)$*, Theorem 4.3 holds true even when* $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$*, provided that* $\sqrt{k}\ A^2(n/k) \underset{n\to\infty}{\longrightarrow} 0$.

Remark 4.6. *Notice that should we have gone further into a third order set-up, assuming for instance that we were working in the class* (4.20)*, would we be able to guarantee asymptotic normality for* $\widehat{\beta}_{\widehat{\rho}}^{GM}(k)$ *in* (4.13)*, with a null asymptotic bias provided that* $\sqrt{k}\ A^2(n/k) \underset{n\to\infty}{\longrightarrow} 0$*. For a* ρ*-estimator,* $\widehat{\rho}$*, in the conditions of Theorem 4.3, the asymptotic variance of* $\sqrt{k}\ A(n/k)\left(\widehat{\beta}_{\widehat{\rho}}^{GM}(k) - \beta\right)/\beta$ *would then be given by* $\gamma^2(1-\rho)^2(1-2\rho)/\rho^2$*, like in* (4.8).

Remark 4.7. *Notice also that with the joint estimation of Feuerverger and Hall* (1999)*, the rate of convergence of their* β*-estimator is slower than the one achieved here: more precisely we have there a rate of convergence of the order of* $\ln(n/k)/\left(\sqrt{k}\ A(n/k)\right)$ *— see Theorem 4.2 — whereas we get here a rate of convergence of the order of* $1/\left(\sqrt{k}\ A(n/k)\right)$*, as in Theorem 4.1, provided we do not estimate* ρ *at the same level used for the* β*-estimation.*

The external estimation of ρ, although may appear at a first sight less appealing, from a theoretical point of view, than the implicit estimation proposed by Feuerverger and Hall (1999) and Beirlant *et al.* (1999), is much more simple in practice, and may work better asymptotically, leading to a much smaller asymptotic variance, as seen before.

5 External estimation of both scale and shape second order parameters

More recently, Gomes *et al.* (2004b) deal with a joint external estimation of both the "scale", β, and the "shape" parameter, ρ, in the A function in (4.5), both estimated at a higher level than the one used for the tail index estimation. Doing so, they have been 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. Such an estimator is based on a linear combination of the log-excesses

$$V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}, \quad 1 \le i \le k, \tag{5.1}$$

and is given by

$$WH_{\widehat{\beta},\ \widehat{\rho}}(k) := \frac{1}{k}\sum_{i=1}^{k} e^{\widehat{\beta}\ (n/k)^{\widehat{\rho}}\ \left((i/k)^{-\widehat{\rho}}-1\right)/(\widehat{\rho}\ln(i/k))}\ V_{ik}, \tag{5.2}$$

for adequate consistent estimators $\widehat{\beta}$ and $\widehat{\rho}$ of the second order parameters β and ρ, respectively.

With the same objectives, but with a simpler analytic expression, we shall here consider in more detail, for an illustration, the estimator

$$\overline{H}_{\widehat{\beta},\ \widehat{\rho}}(k) := H(k)\left(1 - \frac{\widehat{\beta}}{1-\widehat{\rho}}\left(\frac{n}{k}\right)^{\widehat{\rho}}\right), \tag{5.3}$$

studied in Caeiro *et al.* (2004), where $H(k)$ denotes the Hill estimator in (2.1). Note that the dominant component of the bias of Hill's estimator, $A(n/k)/(1-\rho) = \gamma\ \beta\ (n/k)^{\rho}/(1-\rho)$, is thus estimated through $H(k)\ \widehat{\beta}\ (n/k)^{\widehat{\rho}}/(1-\widehat{\rho})$ and directly removed from Hill's classical tail index estimator.

We exhibit here, in Figure 1, the differences between the sample paths of the estimators $\overline{H}_{\bullet}(k)$ in (5.3), for a sample of size $n = 10,000$ from a Fréchet model, with d.f. $F(x) = \exp\left(x^{-1/\gamma}\right),\ x \ge 0$, with $\gamma = 1$, when we compute $\widehat{\beta}$ and $\widehat{\rho}$ at the same level k used for the estimation of the tail index γ (*left*), when we compute only $\widehat{\beta}$ at that same level k, being $\widehat{\rho}$ computed at a larger k-value, let us say an intermediate level k_1 such that $\sqrt{k_1}\ A(n/k_1) \to \infty$, as $n \to \infty$ (*center*) and when both $\widehat{\rho}$ and $\widehat{\beta}$ are computed at that high level k_1 (*right*). We have estimated β through $\widehat{\beta}_{\widehat{\rho}_0}^{GM}(k)$ in (4.13), computed at the level k used for the estimation of the tail index, as well as computed at the level $k_1 = \min(n-1, [2n/\ln\ln n])$ in (2.13), the one used for the estimator

$\widehat{\rho}_0$ in (2.14), and again not chosen in any optimal way. We use the notation $\widehat{\beta}_{01} = \widehat{\beta}_{\widehat{\rho}_0}^{GM}(k_1)$. The estimates of β and ρ have been incorporated in the $\overline{H}$-estimator, leading to $\overline{H}_{\widehat{\beta}_{\widehat{\rho}_0}(k),\ \widehat{\rho}_0}(k) \equiv \overline{H}_{\widehat{\beta}_{\widehat{\rho}_0}^{GM}(k),\ \widehat{\rho}_0}(k)$ and $\overline{H}_{\widehat{\beta}_{01},\ \widehat{\rho}_0}(k)$. It thus seems sensible to investigate the behaviour of this last type of estimation procedure, the one which seems to be the most adequate to be used in practice, due to the higher stability of its sample path around the target value $\gamma = 1$.

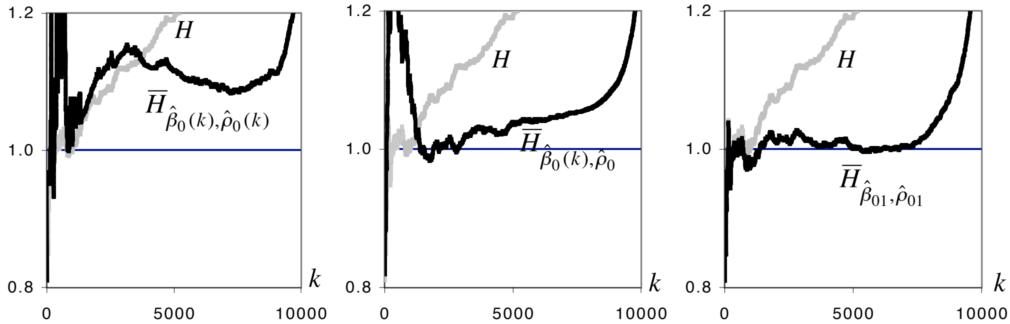

Figure 1: External estimation of (β, ρ) at a larger fixed level k_1 (*right*) versus estimation at the same level both for β and ρ (*left*) and only for β (*center*).

- We are thus interested in the direct estimation of the dominant component of the bias of Hill's estimator of a positive tail index γ.
- Such an estimated bias is then directly removed from the classical Hill estimator.
- The second order parameters in the bias are computed at a fixed level k_1 of a larger order than that of the level k at which we compute the Hill estimator. Doing this, we are able to keep the asymptotic variance of the new reduced bias' tail index estimator equal to γ^2, the asymptotic variance of the Hill estimator.

5.1 Asymptotic behaviour of the new reduced bias' tail index estimator

In the lines of Gomes and Martins (2004) and Gomes *et al.* (2003), we may state:

Lemma 5.1. *Under the second order framework in* (1.4)*, and for levels* k *such that* (1.1) *holds, the distributional representations*

$$\frac{\alpha}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1} U_i \;\stackrel{d}{=}\; \gamma + \frac{\gamma\ \alpha}{\sqrt{(2\alpha-1)\ k}}\ Z_k^{(\alpha)} + \frac{\alpha\ A(n/k)}{\alpha-\rho}(1+o_p(1)) \quad (5.4)$$

and

$$-\frac{\alpha^2}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1}\ln\left(\frac{i}{k}\right) U_i \;\stackrel{d}{=}\; \gamma + \frac{\gamma\ \alpha^2}{(2\alpha-1)\sqrt{(2\alpha-1)\ k/2}}\ W_k^{(\alpha)} + \frac{\alpha^2\ A(n/k)}{(\alpha-\rho)^2}(1+o_p(1)) \quad (5.5)$$

hold true for any $\alpha \geq 1$*, where* $Z_k^{(\alpha)}$*, given in* (4.18)*, and*

$$W_k^{(\alpha)} = (2\alpha-1)\sqrt{(2\alpha-1)\ k/2}\ \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1}\ln\left(\frac{i}{k}\right) E_i + \frac{1}{\alpha^2}\right) \quad (5.6)$$

are asymptotically standard normal r.v.'s.

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

Theorem 5.1. *Under the second order framework in* (1.4)*, further assuming that* $A(t)$ *may be chosen as in* (4.5)*, and for levels* k *such that* (1.1) *holds, we get, for* $\overline{H}_{\beta,\ \rho}(k)$ *in* (5.3)*, an asymptotic distributional representation of the type*

$$\gamma + \frac{\gamma}{\sqrt{k}}\ Z_k^{(1)} + R_k^{(\gamma)}, \quad \text{with} \quad R_k^{(\gamma)} = o_p(A(n/k)), \quad (5.7)$$

where $Z_k^{(1)}$ *is the asymptotically standard normal r.v. in* (4.18) *for* $\alpha = 1$*. Consequently,* $\sqrt{k}\ \left(\overline{H}_{\beta,\ \rho}(k) - \gamma\right)$ *is asymptotically normal with variance*

$$\sigma^2_{CGP_1} = \gamma^2 \quad (5.8)$$

and a null mean value, not only when $\sqrt{k}\ A(n/k) \longrightarrow 0$*, but also when* $\sqrt{k}\ A(n/k) \longrightarrow \lambda \neq 0$*, finite, as* $n \to \infty$*.*

Proof. The result in (5.7) comes straighforwardly from the fact that if all parameters are known, apart from the tail index γ, we get from (5.4), with $\alpha = 1$,

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

i.e., (5.7) holds. The remaining of the theorem follows then straightforwardly.

□

Let us denote $\widehat{\rho}$ any of the estimators in (2.10) computed at the level k_1 in (2.13). Gomes *et al.* (2004b) and Caeiro *et al.* (2004) suggest the consideration of the estimator $\widehat{\beta} \equiv \widehat{\beta}_{\widehat{\rho}}^{\,GM}(k_1)$, with $\widehat{\beta}_{\widehat{\rho}}^{\,GM}(k)$ given in (4.13) and k_1 given in (2.13). Let us assume first that we estimate both β and ρ externally at the level k_1 in (2.13). We may state the following:

Theorem 5.2. *Under the conditions of Theorem 5.1, let us consider the tail index estimators* $\overline{H}_{\widehat{\beta},\,\widehat{\rho}}(k)$ *in* (5.3), *for any of the estimators* $\widehat{\beta}$ *and* $\widehat{\rho}$ *in* (2.10) *and in* (4.13), *respectively, both computed at the level* k_1 *in* (2.13). *Then,* $\sqrt{k}\left\{\overline{H}_{\widehat{\beta},\,\widehat{\rho}}(k) - \gamma\right\}$ *is asymptotically normal with null mean value, not only when* $\sqrt{k}\ A(n/k) \to 0$, *but also whenever* $\sqrt{k}\ A(n/k) \to \lambda$, *finite or infinite, provided that* $\sqrt{k}\ R_k^{(\gamma)} \to 0$, *as* $n \to \infty$.

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

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

Since $\widehat{\beta}$ and $\widehat{\rho}$ are consistent for the estimation of β and ρ, respectively, and $(\widehat{\rho}-\rho)\ln(n/k) = o_p(1)$ (see Remarks 2.5 and 2.6), the summands related to $\left(\widehat{\beta}-\beta\right)$ and $(\widehat{\rho}-\rho)$ are both $o_p(A(n/k))$. Moreover, since from (4.19), $\widehat{\beta}-\beta\ \sim$

$-\beta\ \ln(n/k_1)\ (\widehat\rho-\rho)$,

$$\overline{H}_{\widehat\beta,\ \widehat\rho}(k)-\overline{H}_{\beta,\ \rho}(k)\ \sim\ -\frac{A(n/k)}{1-\rho}\left(\frac{\widehat\beta-\beta}{\beta}+(\widehat\rho-\rho)\ \ln(n/k)\right) \quad (5.9)$$
$$\sim\ \frac{A(n/k)}{1-\rho}\ (\widehat\rho-\rho)\ln(k/k_1).$$

If $\sqrt{k}\ A^2(n/k)\to\lambda$, finite, k is at most of the order of $n^{-4\rho/(1-4\rho)}$. Then, k/k_1 is at most of the order of $n^{-1/(1-4\rho)}\ln\ln n$. Consequently,

$$0\le\sqrt{k}\ \left|(\widehat\rho-\rho)\ln(k/k_1)\right|\ \le\ O_p\left(\frac{n^{\frac{-2\rho}{1-4\rho}}(\ln\ln n)^{\frac{1}{2}-\rho}\ln n}{n^{\frac{1}{2}}}\right)$$
$$=\ O_p\left(\frac{(\ln\ln n)^{\frac{1}{2}-\rho}\ln n}{n^{\frac{1}{2(1-4\rho)}}}\right).$$

Hence, $\sqrt{k}\ (\widehat\rho-\rho)\ A(n/k)\ \ln(k/k_1)$ converges towards zero, as $n\to\infty$, and the results in the theorem follow. □

Remark 5.1. *Note that the levels k such that $\sqrt{k}\ A(n/k)\to\lambda$, finite, are sub-optimal for this type of estimators. To go further to the optimal level, we should go into a third order framework, like the one considered in Gomes and de Haan* (1999)*, Gomes et al.* (2002*a*)*, Fraga Alves et al.* (2003) *and Caeiro et al.* (2004)*, considering levels k such that $\sqrt{k}\ A(n/k)\to\infty$, as $n\to\infty$.*

If we consider γ and β estimated at the same level, we are going to have an increase in the variance of our final tail index estimator $\overline{H}_{\widehat\beta_{\widehat\rho}(k),\ \widehat\rho}(k)$. Similarly to the result in Theorem 4.1, there in connection with the r.v. $\widehat\gamma^{FH}_{\widehat\beta,\rho}$, as well as in Theorem 4.3, in connection with the tail index estimator $\widehat\gamma^{GM}_{\widehat\rho}$, we may also get:

Theorem 5.3. *If the second order condition* (1.4) *holds, if $k=k_n$ is a sequence of intermediate integers, i.e.,* (1.1) *holds, and if $\sqrt{k}\ A(n/k)\underset{n\to\infty}{\longrightarrow}\lambda$, finite, non* *necessarily null, then*

$$\sqrt{k}\left(\overline{H}_{\widehat{\beta}_{\widehat{\rho}}(k),\ \widehat{\rho}}\ (k)-\gamma\right)\underset{n\to\infty}{\overset{d}{\longrightarrow}}\ Normal\left(0,\ \gamma^2\ \left(\frac{1-\rho}{\rho}\right)^2\right), \tag{5.10}$$

i.e., the asymptotic variance of $\overline{H}_{\widehat{\beta}_{\widehat{\rho}}(k),\ \widehat{\rho}}(k)$ *increases of a factor* $((1-\rho)/\rho)^2$, *greater than one, for every* $\rho<0$.

Proof. If we consider

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

we now get

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

Since $\left(\widehat{\beta}_{\widehat{\rho}}(k)-\beta\right)/\beta$ is now of the order of $1/\left(\sqrt{k}\ A(n/k)\right)$, the term of the order of $1/\sqrt{k}$ is going to be

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

which may be written as

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

with $Z_k^{(\alpha)}$ the asymptotically standard normal r.v. in (4.18). Taking again into account the fact that the asymptotic variance between $Z_k^{(1)}$ and $Z_k^{(1-\rho)}$ is given by $\sqrt{1-2\rho}/(1-\rho)$, together with the fact that $\sqrt{k}\ A(n/k)\,(\widehat{\rho}-\rho)\ln(n/k)\to 0$ (see Remark 2.6), (5.10) follows.

□

Remark 5.2. *Caeiro et al.* (2004) *consider also the estimation of the three parameters* γ, β *and* ρ *at the same level* k. *Then, there occur some changes in the asymptotic behaviour of the final tail index estimators* $\overline{H}_{\beta^{GM}_{\widehat{\rho}(k)}(k),\ \widehat{\rho}(k)}(k)$, *with* *an increase in their asymptotic variance. Such an asymptotic variance is then ruled by*

$$\sigma^2_{CGP_2} = \gamma^2 \left(1 + \frac{(1-\rho)^2(2\rho^2 - 2\rho + 1)}{\rho^2}\right), \tag{5.11}$$

5.2 A simulation experiment

We have here implemented a simulation experiment, with 1000 runs, for an underlying Burr parent, $F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho}$, $x \geq 0$, with $\rho = -0.5$ and $\gamma = 1$. For these Burr models, $\beta = \gamma$ for any ρ. We have again estimated β through $\widehat{\beta}^{GM}_{\widehat{\rho}_0}(k)$, computed at the level k used for the estimation of the tail index, as well as computed at the level $k_1 = \min(n-1, [2n/\ln\ln n])$ in (2.13), the one used for the estimator $\widehat{\rho}_0$ in (2.14), and again not chosen in any optimal way. We use the notation $\widehat{\beta}_{01} = \widehat{\beta}^{GM}_{\widehat{\rho}_0}(k_1)$. The estimates of β and ρ have been incorporated in the $\overline{H}$-estimator, leading to $\overline{H}_{\widehat{\beta}_{\widehat{\rho}_0}(k),\ \widehat{\rho}_0}(k) \equiv \overline{H}_{\widehat{\beta}^{GM}_{\widehat{\rho}_0}(k),\ \widehat{\rho}_0}(k)$ and $\overline{H}_{\widehat{\beta}_{01},\ \widehat{\rho}_0}(k)$. The simulations show that the tail index estimator $\overline{H}_{\widehat{\beta}_{01},\ \widehat{\rho}_0}$ seems to work reasonably well, as illustrated in Figure 2.

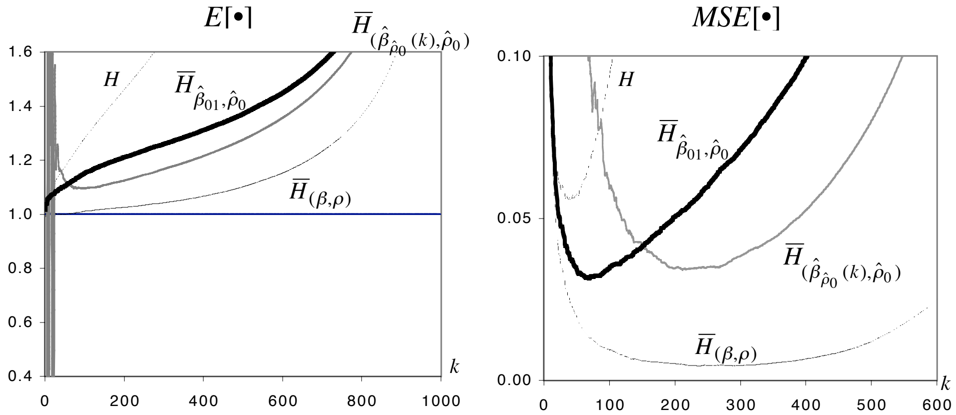

Figure 2: Mean values and Mean Squared errors of the estimators under study for samples of size $n = 1000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -0.5$ ($\beta = 1$).

The discrepancy between the behaviour of the estimator $\overline{H}_{\widehat{\beta}_{01},\ \widehat{\rho}_0}(k)$ and the r.v. $\overline{H}_{\beta,\ \rho}(k)$ suggests that some improvement in the estimation of second order parameters may be still welcome, but the behaviour of the mean squared error of the $\overline{H}-$estimator is rather interesting: the new estimator, $\overline{H}_{\widehat{\beta}_{01},\ \widehat{\rho}_0}(k)$, is better than the Hill estimator not only when both are considered at their optimal levels, but also for every sub-optimal level k, and this contrarily to what happens with $\overline{H}_{\widehat{\beta}_{\widehat{\rho}_0}(k),\ \widehat{\rho}_0}(k)$, as we may also see in this same figure.

5.3 Comparison of asymptotic variances

If we compare Theorems 5.2, 5.3 and the result in (5.11), we see that the estimation of the two parameters γ and β at the same level k induces an increase in the asymptotic variance of the final γ-estimator of a factor given by $((1-\rho)/\rho)^2$, greater than 1, whereas the estimation of γ, β and ρ at the same level k induces an extra increase in the asymptotic variance of the final γ-estimator. In Figure 3, we provide both a picture and some values of $\sigma_{CGP_1}/\gamma \equiv 1$ and of σ_{GJ}/γ, σ_{GM_1}/γ, σ_{FH}/γ, σ_{GM_2}/γ and σ_{CGP_2}/γ, in (2.8), (3.7), (4.11), (4.15) and (5.11), respectively, as functions of $|\rho|$. It is obvious from Figure 3, as well as from

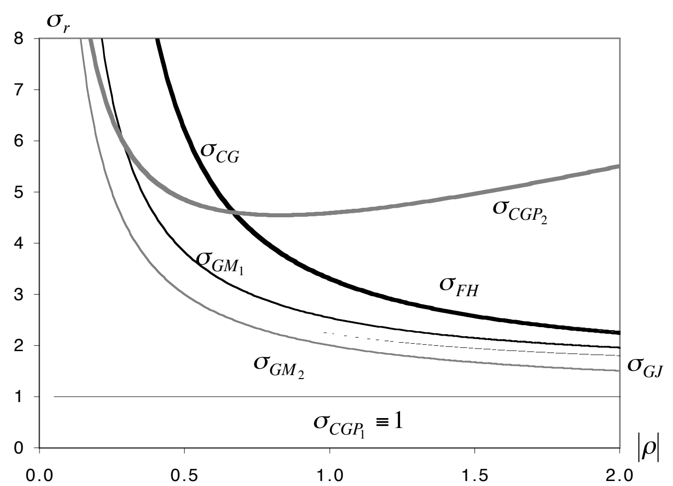

Figure 3: "Rulers" of the asymptotic standard deviations of the reduced bias' tail index estimators, for $\gamma = 1$

Figures 1 and 2, that, whenever possible, it seems convenient to estimate both

β and ρ externally, at a higher level than the one used for the estimation of the tail index γ.

5.4 Some overall conclusions

Remark 5.3. *The main advantage of these estimators lies on the fact that we may estimate* β *and* ρ *adequately through* $\widehat{\beta}$ *and* $\widehat{\rho}$ *so that the* MSE *of the new estimator is smaller than the* MSE *of Hill's estimator for all* k*, even when* $|\rho| > 1$*, a region where has been difficult to find alternatives for the Hill estimator. And this happens together with a higher stability of the sample paths around the target value* γ*.*

Remark 5.4. *More generally, to obtain information on the asymptotic bias of* $\overline{H}_{\widehat{\beta},\ \widehat{\rho}}(k)$*,* $\overline{H}_{\widehat{\beta}_{\widehat{\rho}}(k),\ \widehat{\rho}}(k)$ *and* $\overline{H}_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\ \widehat{\rho}(k)}(k)$ *we should have gone further into a third order framework, specifying the rate of convergence in the second order condition in* (1.4)*. This is however beyond the scope of this overview.*

Remark 5.5. *If we estimate the first order parameter at a level* k*, and use that same level* k *for the estimation of the second order parameter* ρ *and* β *in* $A(t) = \gamma\ \beta\ t^{\rho}$*, we get a much higher asymptotic variance than when we compute* ρ *at a larger level* k_1*, computing* β *at the same level* k*. And we may still decrease the asymptotic variance of the first order parameter' s estimator, if we estimate both second order parameters,* β *and* ρ*, at a larger level* k_1 *than the one used for the estimation of the first order parameter.*

References

[1] Beirlant, J., Vynckier, P. and Teugels, J. L. (1996). Excess function and estimation of the extreme-value index. *Bernoulli* **2**, 293-318.

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

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

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

[5] Caeiro, F. and Gomes, M. I. (2002b) Bias reduction in the estimation of parameters of rare events. *Theory of Stochastic Processes* **8**(24), 67-76.

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

[7] Drees, H. (1996). Refined Pickands estimators with bias correction. *Comm. Statist. Theory and Meth.* **25**, 837-851.

[8] Drees, H., de Haan, L. and Resnick, S. (2000). How to make a Hill plot. *Ann. of Statist.* **28**, 254-274.

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

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

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

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

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

[14] Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423-453.

[15] Gomes, M. I. (1994). Metodologias Jackknife e Bootstrap em Estatística de Extremos. In Mendes-Lopes *et al.* (eds.), *Actas II Congresso S.P.E.*, 31-46.

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

[17] Gomes, M. I. Figueiredo, F. and Mendonça, S. (2003). Asymptotically best linear unbiased tail estimators under a second order regular variation condition. Accepted at *J. Statist. Planning and Inference*.

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

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

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

[21] Gomes, M. I. and M. J. Martins (2001). Alternatives to Hill's estimator — asymptotic versus finite sample behaviour. *J. Statist. Planning and Inference* **93**, 161-180.

[22] 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.

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

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

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

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

[27] Gray, H. L. and Schucany, W. R. (1972). *The Generalized Jackknife Statistic.* Marcel Dekker.

[28] Haan, L. de (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes.* Mathematical Centre Tract 32, Amsterdam.

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

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

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

[32] Kaufmann, E. and Reiss, R. D. (1998). Approximation of the Hill estimator process. *Statistics and Probability Letters* **39**, 347-354.

[33] Martins, M. J., Gomes, M. I. and Neves, M. (1999). Some results on the behaviour of Hill's estimator. *J. Statist. Comput. and Simulation* **63**, 283-297.

[34] Matthys, G. and Beirlant, J. (2003). Estimating the extreme value index and high quantiles with exponential regression models. *Statistica Sinica* **13**, 853-880.

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

[36] Quenouille, B. (1956). Notes on bias in estimation. *Biometrika* **43**, 353-360.

[37] Reiss, R.-D. (1989). Extreme value models and adaptive estimation of the tail index. In Hüsler, J. and Reiss, R.-D. (eds.), *Extreme Value Theory*, Springer-Verlag, 156-165.

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

[39] Tukey, J. (1958). Bias and confidence in not quite large samples. *Ann. Math. Statist.* **29**, 614.