“Asymptotically unbiased” estimators of the tail index based on external estimation of the second order parameter*

M. Ivette Gomes
D.E.I.O. and C.E.A.U.L., Universidade de Lisboa
e-mail: ivette.gomes@fc.ul.pt
and
M. João Martins
Departamento de Matemática, Instituto Superior de Agronomia
Universidade Técnica de Lisboa
e-mail: mjmartins@isa.utl.pt

Abstract. In this paper we shall deal with the asymptotic and finite sample properties of “asymptotically unbiased” estimators of the tail index γ, based on “external” adequate estimators of the second order parameter ρ. The behaviour of the ρ-estimator considered has indeed a high impact on the distributional properties of the final estimator of γ, and must be carefully chosen. As a by-product of the final study we present also the finite sample properties of a few ρ-estimators available in the literature.

AMS 2000 subject classification. Primary 62G32; Secondary 65C05.
Keywords and phrases. *Statistics of Extremes*, *Semi-parametric estimation*, *Generalized Jackknife statistics.*

1 The estimators under study and scope of the paper

In a context of heavy tails, Gomes *et al.* (2000a) have worked with a set of estimators of a positive tail index γ associated to an underlying model F with a regularly varying tail function, $1-F$, with an index of regular variation equal to $-1/\gamma$. Those estimators appeared as competitors to the well-known Hill

*Research partially supported by FCT / POCTI / FEDER.

estimator for γ (Hill, 1975),

$$\gamma_n^{(1)}(k) \equiv \gamma_n^H(k) := \frac{1}{k}\sum_{i=1}^{k}\left[\ln \frac{X_{n-i+1:n}}{X_{n-k:n}}\right], \tag{1.1}$$

where, as usual, $X_{i:n}$ denotes the i-th ascending order statistic (o.s.) associated to the sample $\underline{X}_n = (X_1, X_2, \ldots, X_n)$. Based on the moment statistics,

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left[\ln \frac{X_{n-i+1:n}}{X_{n-k:n}}\right]^j, \quad j \geq 1, \quad [\gamma_n^{(1)} \equiv M_n^{(1)}], \tag{1.2}$$

those estimators were

$$\gamma_n^{(2)}(k) := M_n^{(2)}(k)/(2M_n^{(1)}(k))$$

and

$$\gamma_n^{(3)}(k) := \sqrt{M_n^{(2)}(k)/2},$$

together with four Generalized Jackknife (GJ) statistics based on the pivot estimators $\gamma_n^{(i)}$, $i = 1, 2, 3$. The first three GJ statistics were asymptotically equivalent, although with a distinct pattern for finite n, which led us to choose the third one, denoted here

$$\gamma_n^{GJ}(k) := 3\gamma_n^{(2)}(k) - 2\gamma_n^{(3)}(k). \tag{1.3}$$

The fourth GJ statistic therewith introduced is also going to be used here; such a statistic may be considered as the Natural Generalized Jackknife (NGJ) estimator associated to the Hill estimator, and we shall denote it

$$\gamma_n^{NGJ}(k) := 2\gamma_n^{(1)}(k/2) - \gamma_n^{(1)}(k). \tag{1.4}$$

The estimators in (1.3) and (1.4) were obtained, assuming a known value $\widehat{\rho} = -1$, eventually mispecified, in the Generalized Jackknife estimators to be studied here,

$$\gamma_n^{GJ(\widehat{\rho})}(k) := \frac{-(2-\widehat{\rho})\gamma_n^{(2)}(k) + 2\gamma_n^{(3)}(k)}{\widehat{\rho}}, \tag{1.5}$$

and

$$\gamma_n^{NGJ(\widehat{\rho})}(k) := \frac{\gamma_n^{(1)}(k) - 2^{-\widehat{\rho}}\gamma_n^{(1)}(k/2)}{1 - 2^{-\widehat{\rho}}}, \tag{1.6}$$

respectively, where $\widehat{\rho}$ may be any adequate consistent estimator of the second order parameter ρ.

We also recall that the first GJ estimator considered in Gomes *et al.* (2000a), now denoted $\gamma_n^P(k) := 2\gamma_n^{(2)}(k) - \gamma_n^{(1)}(k)$, could also be obtained, assuming again $\widehat{\rho} = -1$, in

$$\gamma_n^{P(\widehat{\rho})}(k) := \frac{\gamma_n^{(1)}(k) - (1-\widehat{\rho})\gamma_n^{(2)}(k)}{\widehat{\rho}}, \tag{1.7}$$

which is exactly Peng's "asymptotically unbiased" estimator (Peng, 1998).

In Gomes *et al.* (2000a) the study of the estimators (1.5), (1.6) and (1.7) has been postponed due to practical reasons, and it has been there claimed that the known estimators of ρ, like the ones suggested, among other authors, by Hall (1982), Beirlant *et al.* (1996a,b), Drees and Kaufmann (1998) and Peng (1998) have a very high mean squared error for most of the common parent distributions, and that erratic behaviour destroys drastically the theoretical nice properties of the r.v.'s $\gamma_n^{\bullet(\rho)}$, should ρ be known. However it has also been argued there that these estimators would potentially be the "optimal" Generalized Jackknife estimators, provided we were able to get a suitable way of estimating the second order parameter ρ, as we think has now been partially achieved by Fraga Alves *et al.* (2001), for heavy tails.

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), because it corresponds to a Moments' estimation versus a Maximum Likelihood estimation, is much more simple in practice, and may work better asymptotically, leading to a much smaller asymptotic variance, as seen in section 2 of this paper. More than that, the implicit estimation of Feuerverger and Hall leads to serious convergence problems for a great diversity of samples, and in a single sample for several values of k. Thus a study of the distributional properties of such an estimator as a function of k through simulation techniques is computationally unfeasible. Gomes and Martins (2002) have proposed a particular but interesting case of the maximum likelihood estimator of Feuerverger and Hall, with a mispecification of ρ in -1, together with the first order approximation $1 + x$ for e^x, as $x \to 0$. They were led to

$$\gamma_n^{ML}(k) := \frac{1}{k}\sum_{i=1}^{k} U_i - \left(\frac{1}{k}\sum_{i=1}^{k} iU_i\right) \frac{\sum_{i=1}^{k}(2i-k-1)U_i}{\sum_{i=1}^{k} i(2i-k-1)U_i},$$

where

$$U_i = i\left[\ln \frac{X_{n-i+1:n}}{X_{n-i:n}}\right], \quad 1 \le i \le k,$$

are the scaled log-spacings. Such an estimator was then studied under a general second order framework.

In this paper, we rely on the fact that in Hall's class of models (Hall, 1982; Hall and Welsh, 1985), the U_i's are approximately exponential with mean value $\mu_i = \gamma \; e^{D(i/n)^{-\rho}}$, $1 \le i \le k$, $D \in \mathbb{R}$. The joint maximization, in γ, D and ρ, of the log-likelihood of the scaled log-spacings,

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

leads us to an explicit expression for $\widehat{\gamma}$ as a function of $\widehat{D}$ and $\widehat{\rho}$ given by

$$\widehat{\gamma} = \frac{1}{k} \sum_{i=1}^{k} \; U_i e^{-\widehat{D}\left(\frac{i}{n}\right)^{-\widehat{\rho}}}.$$

One of the maximum likelihood equations is then

$$\sum_{i=1}^{k} i^{-\widehat{\rho}} e^{-\widehat{D}(n/i)^{-\widehat{\rho}}} U_i = \widehat{\gamma} \; \sum_{i=1}^{k} i^{-\widehat{\rho}},$$

and if we use a first order approximation for $e^x = 1 + x$, as $x \to 0$, we get

$$\widehat{D} \equiv \widehat{D}_{\widehat{\rho}}(k) = \frac{1}{n^{\widehat{\rho}}} \; \frac{\left(\sum\limits_{i=1}^{k} i^{-\widehat{\rho}}\right) \left(\sum\limits_{i=1}^{k} U_i\right) - k \left(\sum\limits_{i=1}^{k} i^{-\widehat{\rho}} U_i\right)}{\left(\sum\limits_{i=1}^{k} i^{-\widehat{\rho}}\right) \left(\sum\limits_{i=1}^{k} i^{-\widehat{\rho}} U_i\right) - k \left(\sum\limits_{i=1}^{k} i^{-2\widehat{\rho}} U_i\right)}, \tag{1.8}$$

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

$$\begin{aligned} \gamma_n^{ML(\widehat{\rho})}(k) := {} & \frac{1}{k} \sum_{i=1}^{k} U_i \\ & - \left(\frac{1}{k} \sum_{i=1}^{k} i^{-\widehat{\rho}} \; U_i\right) \; \frac{\left(\sum\limits_{i=1}^{k} i^{-\widehat{\rho}}\right) \left(\sum\limits_{i=1}^{k} U_i\right) - k \left(\sum\limits_{i=1}^{k} i^{-\widehat{\rho}} U_i\right)}{\left(\sum\limits_{i=1}^{k} i^{-\widehat{\rho}}\right) \left(\sum\limits_{i=1}^{k} i^{-\widehat{\rho}} U_i\right) - k \left(\sum\limits_{i=1}^{k} i^{-2\widehat{\rho}} U_i\right)}, \end{aligned} \tag{1.9}$$

which will be compared with the other estimators of γ herewith considered, all dependent on an external estimation of the second order parameter ρ.

The Monte Carlo simulations were based on one-replicate of size 5000, and we have considered the following set of models in Hall's class of distributions,

1. the *Fréchet* model, $F(x) = \exp(-x^{-1/\gamma})$, $x \ge 0$, with $\gamma = 1$ $(\rho = -1)$,

2. the *Burr* model, $F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho},\ x \geq 0,\ \gamma > 0,\ \rho < 0$, with $\gamma = 1$ and for $\rho = -0.5, -1, -2$,

and

3. the *Student-t* model with $\nu = 4, 2, 1$ degrees of freedom, for which $\gamma = 0.25, 0.5, 1$ and $\rho = -0.5, -1, -2$, respectively.

We have simulated the Mean Value ($E_s^\bullet$), the Mean Squared Error ($MSE_s^\bullet$), the Optimal Sample Fraction, $k_{0s}^\bullet/n$, with $k_0^\bullet := \arg\min_k MSE[\gamma_n^\bullet(k)]$, and two indicators at the optimal level, the Relative Efficiency ($REFF^\bullet$), defined as

$$REFF^\bullet = \sqrt{\frac{MSE_s\left[\gamma_{n0}^{(1)}\right]}{MSE_s\left[\gamma_{n0}^\bullet\right]}},$$

and a Bias Reduction Indicator ($BRI^\bullet$), defined as

$$BRI^\bullet = \left|\frac{BIAS_s\left[\gamma_{n0}^{(1)}\right]}{BIAS_s\left[\gamma_{n0}^\bullet\right]}\right|,$$

with $\gamma_{n0}^\bullet = \gamma_n^\bullet(k_{0s}^\bullet(n))$. Also, to illustrate the loss of sensitivity of the new estimators to the choice of the level k, we have simulated two other indicators, related to the stability with k of the mean squared error and of mean value, respectively, and denoted

$$STI_1^\bullet = \frac{\max\left(\sum_{k=1}^{k_{0s}^\bullet} I_{\left\{\left|\frac{MSE_s(\gamma_n^\bullet(k))}{MSE_s(\gamma_n^\bullet(k_{0s}^\bullet))}-1\right|\leq 0.2\right\}},\ \sum_{k=k_{0s}^\bullet}^{n-1} I_{\left\{\left|\frac{MSE_s(\gamma_n^\bullet(k))}{MSE_s(\gamma_n^\bullet(k_{0s}^\bullet))}-1\right|\leq 0.2\right\}}\right)}{\max\left(\sum_{k=1}^{k_{0s}^{(1)}} I_{\left\{\left|\frac{MSE_s(\gamma_n^{(1)}(k))}{MSE_s(\gamma_n^{(1)}(k_{0s}^{(1)}))}-1\right|\leq 0.2\right\}},\ \sum_{k=k_{0s}^{(1)}}^{n-1} I_{\left\{\left|\frac{MSE_s(\gamma_n^{(1)}(k))}{MSE_s(\gamma_n^{(1)}(k_{0s}^{(1)}))}-1\right|\leq 0.2\right\}}\right)}.$$

and

$$STI_2^\bullet = \frac{\sum_{k=1}^{n-1} I_{\left\{\left|\frac{E_s(\gamma_n^\bullet(k))}{\gamma}-1\right|\leq 0.2\right\}}}{\sum_{k=1}^{n-1} I_{\left\{\left|\frac{E_s(\gamma_n^{(1)}(k))}{\gamma}-1\right|\leq 0.2\right\}}}.$$

The subscript s denotes "simulated" and I_A denotes, as usual, the indicator function of the event A. The higher the indicators, the better the estimator is. Indicator values higher than one mean improvement relatively to the Hill estimator.

As an overall description of the results in this paper, we may say that after the derivation, in section 2, of the asymptotic distributional properties of the estimators under study, we shall make, in section 3, a derivation of the asymptotic behaviour of the herewith proposed estimator of D in (1.8) and a presentation of the simulated distributional behaviour of four estimators of the second order parameter ρ, drawing some conclusions about the one which appears to be the most adequate, at the present state of the art. In section 4, and for the selected estimators of the second order parameter ρ, we present the patterns of mean values and mean squared errors of the tail index estimators under study, as functions of the level k. The consideration of the estimator (1.9) does not enable us to perform a large scale simulation like the one in Gomes and Martins (2001) for the GJ estimators, but only a one-replicate simulation of size 5000. Finally, in section 5 we shall deal with the capacity of these estimators to reduce bias and with their insensitivity to the choice of the level k; we shall also present their efficiency relatively to the Hill estimator, and draw some overall conclusions.

2 The asymptotic behaviour of the tail index estimators

We here just recall that, in a context of heavy tails, and with the notation $U(t) = F^{\leftarrow}(1 - 1/t),\ t \geq 1,\ F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$ the generalized inverse function of the underlying model F, the first order parameter (or tail index) γ (> 0) appears, for every $x > 0$, as the limiting value, as $t \to \infty$, of

$$\frac{\ln U(tx) - \ln U(t)}{\ln x}.$$

The second order parameter ρ is the non-positive value which appears in the limiting relation

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}, \tag{2.1}$$

which holds for every $x > 0$, and where $|A(t)|$ is of regular variation with index ρ (Geluk and de Haan, 1987).

The r.v.'s $\gamma_n^{GJ(\rho)}(k)$, $\gamma_n^{NGJ(\rho)}(k)$, $\gamma_n^{P(\rho)}(k)$ and $\gamma_n^{ML(\rho)}(k)$ (replace $\widehat{\rho}$ by ρ in (1.5), (1.6), (1.7) and (1.9), respectively) were built to remove the main component of bias of the original estimators $\gamma_n^{(i)}(k),\ i = 1, 2, 3$, which is of the order of $A(n/k)$. For these r.v.'s, and with the notation Normal(μ, σ^2) for a normal random variable with mean value μ and variance σ^2, we have the validity of the following distributional representations,

Theorem 2.1. *If the second order condition (2.1) holds, if $k = k_n$ is a sequence of intermediate positive integers, i.e.,*

$$k = k_n \to \infty, \quad k_n = o(n), \ \text{as } n \to \infty, \tag{2.2}$$

and if $\sqrt{k}A(n/k) \underset{n\to\infty}{\longrightarrow} \lambda$, finite, non necessarily null, then

$$\sqrt{k}\left(\gamma_n^{\bullet(\rho)}(k) - \gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \ Normal\left(0, \ \frac{\gamma^2(2\rho^2 - 2\rho + 1)}{\rho^2}\right), \tag{2.3}$$

both for $\gamma_n^{GJ(\rho)}$ and $\gamma_n^{P(\rho)}$.

We further have

$$\sqrt{k}\left(\gamma_n^{NGJ(\rho)}(k) - \gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \ Normal\left(0, \ \frac{\gamma^2(1 + 2^{1-2\rho} - 2^{1-\rho})}{(1-2^{-\rho})^2}\right) \tag{2.4}$$

and

$$\sqrt{k}\left(\gamma_n^{ML(\rho)}(k) - \gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \ Normal\left(0, \ \frac{\gamma^2(1-\rho)^2}{\rho^2}\right). \tag{2.5}$$

Proof. Notice that, with $P_k^{(r)} = \sqrt{k}\left\{\frac{1}{k}\sum_{j=1}^{k} E_j^r - r\right\}$, $r = 1, 2$, $\{E_j\}_{j\geq 1}$ a sequence of i.i.d. standard exponential r.v.'s, we may write

$$\sqrt{k}\left\{\gamma_n^{(i)}(k) - \gamma\right\} \overset{d}{=} \gamma \ V_k^{(i)} + O_p\left(\sqrt{k}A(n/k)\right),$$

where $V_k^{(1)} = P_k^{(1)}$, $V_k^{(2)} = P_k^{(2)}/2 - P_k^{(1)}$ and $V_k^{(3)} = P_k^{(2)}/4$. Then, for $\gamma_n^{GJ(\rho)}(k)$ and $\gamma_n^{P(\rho)}(k)$, straighforward computations lead us to the distributional representation,

$$\sqrt{k}\left\{\gamma_n^{\bullet(\rho)}(k) - \gamma\right\} \overset{d}{=} \gamma \ V_k^{\bullet(\rho)} + o_p\left(\sqrt{k} \ A(n/k)\right), \tag{2.6}$$

where

$$V_k^{GJ(\rho)} = \frac{2 \ V_k^{(3)} - (2-\rho) \ V_k^{(2)}}{\rho} = \frac{(2-\rho)P_k^{(1)} - (1-\rho)P_k^{(2)}/2}{\rho}$$

and

$$V_k^{P(\rho)} = \frac{2 \ V_k^{(1)} - (1-\rho) \ V_k^{(2)}}{\rho} = \frac{(2-\rho)P_k^{(1)} - (1-\rho)P_k^{(2)}/2}{\rho} \equiv V_k^{GJ(\rho)}.$$

Since $Var[P_k^{(1)}] = 1$, $Var[P_k^{(2)}] = 20$ and $Cov[P_k^{(1)}, \ P_k^{(2)}] = 4$ (Dekkers *et al.*, 1989), (2.3) follows immediately.

For the r.v. $\gamma_n^{NGJ(\rho)}$ we have a distributional representation of the type (2.6), where

$$V_k^{NGJ(\rho)} = \frac{\sqrt{2}\ 2^{-\rho}\ P_{k,1}^{(1)} - P_{k,2}^{(1)}}{1-2^{-\rho}}, \quad P_{k,r}^{(1)} = \sqrt{\frac{rk}{2}} \left\{ \frac{2}{rk} \sum_{j=1}^{rk/2} E_j - 1 \right\},\ r=1,2.$$

Then $Var[P_{k,1}^{(1)}] = Var[P_{k,2}^{(1)}] = 1$, $Cov[P_{k,1}^{(1)}, P_{k,2}^{(1)}] = \sqrt{2}/2$, and (2.4) follows.

The r.v. $\gamma_n^{ML(\rho)}(k)$ may be written as

$$\gamma_n^{ML(\rho)}(k) := \frac{\sum_{i=1}^{k} U_i}{k} - \frac{\sum_{i=1}^{k} i^{-\rho}\ U_i}{k^{1-\rho}}\ \frac{\frac{\sum_{i=1}^{k} i^{-\rho}}{k^{1-\rho}}\frac{\sum_{i=1}^{k} U_i}{k} - \frac{\sum_{i=1}^{k} i^{-\rho} U_i}{k^{1-\rho}}}{\frac{\sum_{i=1}^{k} i^{-\rho}}{k^{1-\rho}}\frac{\sum_{i=1}^{k} i^{-\rho} U_i}{k^{1-\rho}} - \frac{\sum_{i=1}^{k} i^{-2\rho} U_i}{k^{1-2\rho}}}.$$

It has been proved in Gomes and Martins (2002) that

$$\frac{\alpha}{k^\alpha} \sum_{i=1}^{k} i^{\alpha-1} U_i \stackrel{d}{=} \gamma + \frac{\gamma\ \alpha}{\sqrt{(2\alpha-1)k}}\ Z_k^{(\alpha)} + \frac{\alpha\ A(n/k)}{\alpha-\rho}(1+o_p(1)), \quad \alpha \geq 1, \quad (2.7)$$

where $Z_k^{(\alpha)} = \sqrt{(2\alpha-1)k}\ \left(\frac{1}{k^\alpha} \sum_{i=1}^{k} i^{\alpha-1} E_i - \frac{1}{\alpha} \right)$ is asymptotically standard normal. Consequently,

$$\frac{1}{k^{1-\rho}} \sum_{i=1}^{k} i^{-\rho} U_i \xrightarrow[n\to\infty]{p} \frac{\gamma}{1-\rho},$$

and

$$\frac{1}{k^{1-2\rho}} \left(\sum_{i=1}^{k} i^{-2\rho}\ U_i - \frac{1}{k} \sum_{i=1}^{k}\ i^{-\rho} \sum_{i=1}^{k}\ i^{-\rho} U_i \right) \xrightarrow[n\to\infty]{p} \frac{\gamma\ \rho^2}{(1-\rho)^2(1-2\rho)}.$$

Using (2.7), and the approximation $\sum_{i=1}^{k} i^{-\rho}/k^{1-\rho} = \frac{1}{1-\rho}(1+o(1))$,

$$\begin{aligned} &\frac{1}{k^{1-\rho}} \sum_{i=1}^{k} i^{-\rho} U_i - \left(\frac{1}{k^{1-\rho}} \sum_{i=1}^{k} i^{-\rho} \right) \left(\frac{1}{k} \sum_{i=1}^{k} U_i \right) \\ &\quad \stackrel{d}{=} \frac{\gamma}{\sqrt{k}} \left(\frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}} - \frac{Z_k^{(1)}}{1-\rho} \right) + \frac{\rho^2}{(1-\rho)^2(1-2\rho)}\ A(n/k)(1+o_p(1)). \end{aligned}$$

Consequently

$$\gamma_n^{ML(\rho)} \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} \left(\left(\frac{1-\rho}{\rho} \right)^2\ Z_k^{(1)} - \frac{(1-\rho)\sqrt{1-2\rho}}{\rho^2}\ Z_k^{(1-\rho)} \right) + o_p(A(n/k)),$$

and from the fact that the asymptotic covariance between $Z_k^{(1)}$ and $Z_k^{(1-\rho)}$ is $\sqrt{1-2\rho}/(1-\rho)$, (2.5) follows immediately. □

Corolary 2.1. *Under the conditions of Theorem 2.1, the same distributional results hold true if we consider the tail index estimators* $\gamma_n^{\bullet(\widehat{\rho})}$, *for any second order parameter estimator* $\widehat{\rho}$ *such that* $\widehat{\rho}-\rho=o_p(1)$ *independently of* k.

Proof. The result comes essentially from the assumption that $\widehat{\rho}-\rho=o_p(1)$ independently of k, and from the fact that we have the distributional representation $\gamma_n^{\bullet(\widehat{\rho})}(k)\overset{d}{=}\gamma_n^{\bullet(\rho)}(k)+(\widehat{\rho}-\rho)\,\xi_k^{\bullet}(\rho)\,(1+o_p(1))$, with $\xi_k^{\bullet}=O_p(1/\sqrt{k})$. Consequently, $\sqrt{k}\left(\gamma_n^{\bullet(\widehat{\rho})}(k)-\gamma\right)\overset{d}{=}\sqrt{k}\left(\gamma_n^{\bullet(\rho)}(k)-\gamma\right)+o_p(1)$, whenever $\sqrt{k}A(n/k)\underset{n\to\infty}{\longrightarrow}\lambda$ finite, non necessarily null. □

Remark 2.1. *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* $\gamma^2\sigma_1^2$, *with* σ_1^2 *given in their Remark 4.1. The complex expression therewith presented and denoted* σ_1^2 *may be written as*

$$\left(\sigma^{FH}\right)^2\equiv\sigma_1^2=\left(\frac{1-\rho}{\rho}\right)^4$$

which is the square of the value exhibited in (2.5). Such an increase in variance is due to the simultaneous estimation of γ *and* ρ *at the same level* k.

Remark 2.2. *Notice also that, as it is well-known from the literature, we may have a non-null asymptotic bias for the original estimators* $\gamma_n^{(i)}(k)$, $i=1,2,3$, *if the threshold* k *is such that* $\sqrt{k}\;A(n/k)\to\lambda\neq0$. *It is indeed a sequence* $k_0=k_0(n)$ *such that* $\sqrt{k_0}\;A(n/k_0)\to\varphi(\rho,\gamma)\neq0$, *the one which provides a minimum mean squared error of* $\gamma_n^{(i)}(k)$.
From Theorem 2.1, even when $\sqrt{k}\;A(n/k)\to\lambda\neq0$, *finite, do we have a null dominant component of asymptotic bias for any of the r.v.'s,* $\gamma_n^{\bullet(\rho)}$. *The same happens for* γ*-estimators,* $\gamma_n^{\bullet(\widehat{\rho})}$, *whenever we consider the semi-parametric* ρ*-estimators in Gomes et al. (2000b) or in Fraga Alves et al. (2001), 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 2.3. *If we further work with a third order expansion, like the one used in Gomes and de Haan (1999), Gomes et al. (2000b) and Fraga Alves et al. (2001) we may see that for a large class of models in Hall's class, more precisely, for models with a tail function*

$$1-F(x)=Cx^{-1/\gamma}\left(1+D_1x^{\rho/\gamma}+D_2x^{2\rho/\gamma}+o(x^{2\rho/\gamma})\right),\;\; as\; x\to\infty, \tag{2.8}$$

the minimum mean squared error of any of the four estimators $\gamma_n^{\bullet(\widehat{\rho})}$ 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. (2000b) and in Fraga Alves et al. (2001). Since the remainder $o_p\left(\sqrt{k}\ A(n/k)\right)$ in (2.6) is then of the order of $\sqrt{k}\ A^2(n/k)$, Theorem 2.1 holds true even when $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$, provided that $\sqrt{k}\ A^2(n/k) \underset{n\to\infty}{\longrightarrow} 0$.

As mentioned in Remark 2.3, the asymptotic study of the proposed estimators at their optimal levels would need the use of a third order theory, and that overpasses the aims of the present investigation. We are however going to compare the estimators asymptotically through their asymptotic variances, given in terms of

$$\sigma^{GJ} \equiv \sigma^{P} = \sqrt{\frac{2\rho^2 - 2\rho + 1}{\rho^2}},\ \sigma^{NGJ} = \sqrt{\frac{1 + 2^{1-2\rho} - 2^{1-\rho}}{(1 - 2^{-\rho})^2}},\ \sigma^{ML} = \frac{1-\rho}{\rho},$$

and pictured in Figure 1, together with σ^{FH}, as functions of $|\rho|$.

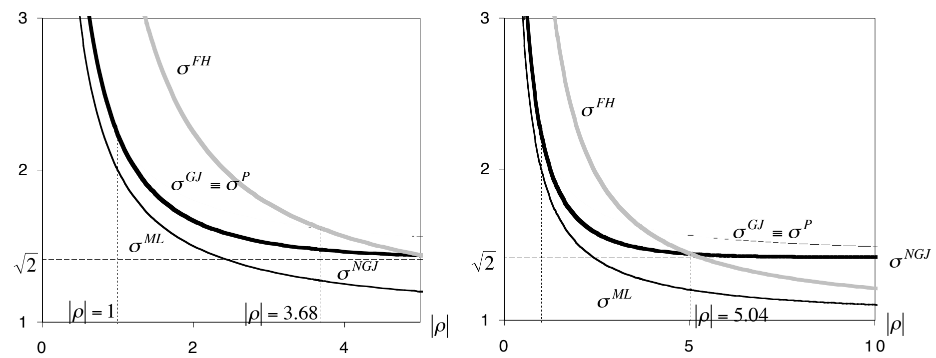

Figure 1: Asymptotic standard deviations of the estimators under study (with two different scales in the horizontal axis).

For all these estimators we have $\lim_{\rho\to 0} \sigma^\bullet = +\infty$ and

$$\lim_{\rho\to-\infty} \sigma^{GJ} = \lim_{\rho\to-\infty} \sigma^{NGJ} = \lim_{\rho\to-\infty} \sigma^{P} = \sqrt{2}, \quad \lim_{\rho\to-\infty} \sigma^{ML} = \lim_{\rho\to-\infty} \sigma^{FH} = 1.$$

It is clear from the picture that asymptotically, and regarding variances, the ML estimator herewith proposed has an overall better performance. The comparison between the other estimators enable us to say that, asymptotically, the GJ (or the P) estimator performs better than the NGJ estimator for $\rho \geq -1$, and things happen the other way round for $\rho < -1$. The implicit estimation of

Feuerverger and Hall (1999) leads to an asymptotic variance greater than that of the estimators either in (1.5) or (1.7) for values of $|\rho| \le 3.08$, and greater than that of the estimator in (1.6) for $|\rho| \le 5.04$.

As a final conclusion, and from a theoretical point of view, regarding the behaviour of asymptotic variances, the estimator in (1.9) has an overall better performance for every $\gamma > 0$ and for every value of the second order parameter $\rho < 0$. We shall see that this behaviour also holds often for finite samples, but not generally.

3 Some comments on the estimation of the second order structure

We shall first briefly refer the rate of convergence of the estimator of D in (1.8):

Theorem 3.1. *If the second order condition (2.1) holds with $A(t) = D\gamma\ t^\rho$, if $k = k_n$ is a sequence of intermediate positive integers, i.e., (2.2) holds and if $\sqrt{k}A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$, then $\widehat{D}$ in (1.8) is consistent for the estimation of D at a rate of convergence of the order of $\frac{1}{\sqrt{k}\ A(n/k)}$.*

Proof. We shall consider the r.v.

$$\begin{aligned} D_\rho(k) &:= \frac{1}{n^\rho} \frac{\left(\sum\limits_{i=1}^k i^{-\rho}\right)\left(\sum\limits_{i=1}^k U_i\right) - k\left(\sum\limits_{i=1}^k i^{-\rho}U_i\right)}{\left(\sum\limits_{i=1}^k i^{-\rho}\right)\left(\sum\limits_{i=1}^k i^{-\rho}U_i\right) - k\left(\sum\limits_{i=1}^k i^{-2\rho}U_i\right)} \\ &= \left(\frac{k}{n}\right)^\rho \frac{\left(\frac{1}{k}\sum\limits_{i=1}^k \left(\frac{i}{k}\right)^{-\rho}\right)\left(\frac{1}{k}\sum\limits_{i=1}^k U_i\right) - \left(\frac{1}{k}\sum\limits_{i=1}^k \left(\frac{i}{k}\right)^{-\rho}U_i\right)}{\left(\frac{1}{k}\sum\limits_{i=1}^k \left(\frac{i}{k}\right)^{-\rho}\right)\left(\frac{1}{k}\sum\limits_{i=1}^k \left(\frac{i}{k}\right)^{-\rho}U_i\right) - \left(\frac{1}{k}\sum\limits_{i=1}^k \left(\frac{i}{k}\right)^{-2\rho}U_i\right)}. \end{aligned}$$

Recall that under the context herewith considered $A(t) = D\gamma\ t^\rho$, and since the denominator in the expression converges, as $n \to \infty$, towards

$$\frac{1}{1-\rho}\ \frac{\gamma}{1-\rho} - \frac{\gamma}{1-2\rho} = -\frac{\gamma\ \rho^2}{(1-\rho^2(1-2\rho)},$$

we may say the $D_\rho(k)$ is asymptotically equivalent to

$$-\frac{D(1-\rho)^2(1-2\rho)}{\rho^2 A(n/k)}\left\{\left(\frac{1}{k}\sum_{i=1}^k\left(\frac{i}{k}\right)^{-\rho}\right)\left(\frac{1}{k}\sum_{i=1}^k U_i\right) - \left(\frac{1}{k}\sum_{i=1}^k\left(\frac{i}{k}\right)^{-\rho}U_i\right)\right\}.$$

Then, computations in the lines of the proof of Theorem 2.1, and with the same notation as before, enable us to write

$$D_\rho(k) = -\frac{D(1-\rho)^2(1-2\rho)}{\rho^2 A(n/k)} \left\{ \frac{\gamma}{\sqrt{k}} \left(\frac{Z_k^{(1)}}{1-\rho} - \frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}} \right) \right.$$
$$\left. - \frac{\rho^2\ A(n/k)}{(1-\rho)^2(1-2\rho)} + o_p(A(n/k)) \right\}.$$

We then need to have $\frac{1}{\sqrt{k}} = o(A(n/k))$, i.e. $\sqrt{k}A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$, so that the second term above is the dominant one, and then

$$D_\rho(k) = D + \frac{D\ \gamma\ (1-\rho)\sqrt{1-2\rho}}{\rho\sqrt{k}\ A(n/k)}\ U_k + o_p(1),$$

where U_k is asymptotically standard normal, and the result follows. □

Remark 3.1. *Notice that should we have gone further in a third order set-up, assuming for instance that we were working in the class (2.8), would we be able to guarantee asymptotic normality for* $\widehat{D}$ *in (1.8), with a null asymptotic bias provided we assumed that* $\sqrt{k}A^2(n/k) \underset{n\to\infty}{\longrightarrow} 0$. *The asymptotic variance of* $\sqrt{k}\ A(n/k)\left\{\frac{D_{\widehat{\rho}}(k)-D}{D}\right\}$ *would then be given by* $\frac{\gamma^2(1-\rho)^2(1-2\rho)}{\rho^2}$, *for a* $\widehat{\rho}$ *in the conditions of Corolary 2.1.*

Remark 3.2. *Notice also that with the joint estimation of Feuerverger and Hall (1999) the rate of convergence of their D-estimator is slower than the one achieved here: more preceisely we have there a rate of convergence of the order of* $\frac{\ln(n/k)}{\sqrt{k}\ A(n/k)}$.

We shall next deal with the simulated distributional properties of estimators of the second order parameter ρ. As mentioned before, we shall here work with four estimators of ρ, elected among a large set of such estimators, all built upon the statistics $M_n^{(j)}(k)$ in (1.2).

The first estimator is a ρ-estimator of the type of the one provided by Hall and Welsh (1985), given by

$$\widehat{\rho}_1 := -\left| \log \left| \frac{1/M_n^{(1)}([n^{0.9}]) - 1/M_n^{(1)}([n^{0.5}])}{1/M_n^{(1)}([n^{0.95}]) - 1/M_n^{(1)}([n^{0.5}])} \right| / \log \frac{[n^{0.9}]}{[n^{0.95}]} \right|. \tag{3.1}$$

The second estimator is the one suggested in Peng (1998),

$$\widehat{\rho}_2 := -\frac{1}{\ln 2} \left| \log \left| \frac{M_n^{(2)}\left(\left[\frac{n}{2\ln n}\right]\right) - 2\left(M_n^{(1)}\left(\left[\frac{n}{2\ln n}\right]\right)\right)^2}{M_n^{(2)}\left(\left[\frac{n}{\ln n}\right]\right) - 2\left(M_n^{(1)}\left(\left[\frac{n}{\ln n}\right]\right)\right)^2} \right| \right|. \tag{3.2}$$

The other two estimators are particular members of the class of estimators proposed by Fraga Alves *et al.* (2001). Under adequate general conditions, they are semi-parametric asymptotically normal estimators of ρ, which show highly stable sample paths as functions of k, the number of top order statistics used, for a wide range of large k-values. Such a class of estimators is parametrized in a tuning parameter τ and depends on the statistics

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

which converge towards $3(1-\rho)/(3-\rho)$, independently of τ, whenever the second order condition (2.1) holds and k is such that $k = o(n)$ and $\sqrt{k}\ A(n/k) \to \infty$, as $n \to \infty$. The estimators are thus given by

$$\widehat{\rho}_n^{(\tau)}(k) := -\left|\frac{3(T_n^{(\tau)}(k) - 1)}{T_n^{(\tau)}(k) - 3}\right|, \quad \tau \geq 0. \tag{3.3}$$

The theoretical and simulated results in Fraga Alves *et al.* (2001) led us to consider the following two estimators of ρ, associated to the high level $k_1 = \min(n-1, [2n/\ln\ln n])$ (not chosen in an optimal way), and to the tuning parameters $\tau = 0$ and $\tau = 1$,

$$\widehat{\rho}_3 \equiv \widehat{\rho}_n^{(0)}(k_1) := -\left|3(T_n^{(0)}(k_1) - 1)/(T_n^{(0)}(k_1) - 3)\right| \tag{3.4}$$

and

$$\widehat{\rho}_4 \equiv \widehat{\rho}_n^{(1)}(k_1) := -\left|3(T_n^{(1)}(k_1) - 1)/(T_n^{(1)}(k_1) - 3)\right|. \tag{3.5}$$

In Tables 1 and 2 we present the simulated mean values and root mean squared errors, respectively, of the four estimators of ρ in (3.1), (3.2), (3.4) and (3.5), for *Fréchet*, *Burr* and *Student* models. For each n and for each model the smallest bias and the smallest root mean squared error is *underlined and in italic*. All the simulated results in this section are based on 5000 runs.

Among the models simulated, with $\rho \geq -1$, Hall and Welsh and Pengs's ρ-estimators, $\widehat{\rho}_1$ and $\widehat{\rho}_2$, respectively, have generally a high bias comparatively to $\widehat{\rho}_3$ ($\tau = 0$) unless n is very high, say $n \geq 5000$; then $\widehat{\rho}_1$ has the smallest bias among the estimators considered. Hall and Welsh's estimator has generally small mean squared errors for large n, but is often overpassed by $\widehat{\rho}_3$, the estimator we think to be the most advisable to be used in practice, whenever we expect to be in the region $|\rho| \leq 1$. For small values of ρ, here illustrated with $\rho = -2$,

Table 1: Mean values of the estimators $\widehat{\rho}_j$, $j = 1, 2, 3$.

n	100	200	500	1000	2000	5000	10000	20000
				$\rho = \mathbf{-0.5}$				
$E[\widehat{\rho}_1 \| BU]$	-1.2746	-1.0796	-0.9174	-0.8375	-0.7823	*-0.7276*	*-0.6962*	*-0.6713*
$E[\widehat{\rho}_2 \| BU]$	-1.6034	-1.4813	1.3438	-1.2738	-1.1944	-1.1063	-1.0519	-0.9747
$E[\widehat{\rho}_3 \| BU]$	*-0.7534*	*-0.7539*	*-0.7518*	*-0.7494*	*-0.7566*	-0.7370	-0.7255	-0.7163
$E[\widehat{\rho}_4 \| BU]$	-2.0766	-2.1377	-2.1904	-2.2172	-2.0036	-1.7006	-1.5895	-1.5123
$E[\widehat{\rho}_1 \| STU]$	-1.6954	-1.1341	-0.8410	-0.7489	-0.6922	*-0.6419*	*-0.6152*	*-0.5939*
$E[\widehat{\rho}_2 \| STU]$	-1.8815	-1.6496	-1.4429	-1.3377	-1.2471	-1.1578	-1.1022	-1.0192
$E[\widehat{\rho}_3 \| STU]$	*-0.6317*	*-0.6408*	*-0.6600*	*-0.6750*	*-0.6857*	-0.6875	-0.6771	-0.6694
$E[\widehat{\rho}_4 \| STU]$	-1.3694	-1.3581	-1.5232	-1.6309	-1.7130	-1.7256	-1.6418	-1.5812
				$\rho = \mathbf{-1}$				
$E[\widehat{\rho}_1 \| FRE]$	-2.4995	-2.3022	-2.1450	-2.0413	-1.9406	-1.8013	-1.7147	-1.6472
$E[\widehat{\rho}_2 \| FRE]$	-1.6351	-1.5449	-1.4900	-1.4588	-1.4568	-1.4346	-1.4313	-1.4043
$E[\widehat{\rho}_3 \| FRE]$	*-1.4660*	*-1.3023*	*-1.2213*	*-1.1989*	*-1.2683*	*-1.2612*	*-1.2450*	*-1.2298*
$E[\widehat{\rho}_4 \| FRE]$	-2.89194	-3.19677	-2.48229	-2.49452	-2.25796	-1.93791	-1.8218	-1.7413
$E[\widehat{\rho}_1 \| BU]$	-2.1679	-1.9505	-1.7544	-1.6301	-1.5336	-1.4141	-1.3438	-1.2869
$E[\widehat{\rho}_2 \| BU]$	-1.6225	-1.5340	-1.4647	-1.4245	-1.4023	-1.3760	-1.3512	-1.3368
$E[\widehat{\rho}_3 \| BU]$	*-0.8204*	*-0.8089*	*-0.7956*	*-0.7867*	*-0.8416*	*-0.8851*	*-0.8982*	*-0.9068*
$E[\widehat{\rho}_4 \| BU]$	-2.1317	-2.1822	-2.2258	-2.2473	-2.0739	-1.8198	-1.7268	-1.6625
$E[\widehat{\rho}_1 \| STU]$	-2.3729	-1.9518	-1.5785	-1.4072	-1.2841	*-1.1805*	*-1.1290*	*-1.0893*
$E[\widehat{\rho}_2 \| STU]$	-1.8768	-1.6614	-1.5088	-1.4534	-1.4168	-1.3572	-1.2987	-1.2342
$E[\widehat{\rho}_3 \| STU]$	*-0.9684*	*-0.8531*	*-0.7635*	*-0.7577*	*-0.7543*	-0.7536	-0.7567	-0.7601
$E[\widehat{\rho}_4 \| STU]$	-1.8424	-1.5797	-1.5990	-1.6853	-1.7530	-1.7684	-1.6996	-1.6476
				$\rho = \mathbf{-2}$				
$E[\widehat{\rho}_1 \| BU]$	-3.0571	-2.8547	-2.6956	-2.6001	-2.5204	-2.3965	-2.3299	*-2.2702*
$E[\widehat{\rho}_2 \| BU]$	-1.6454	-1.5612	-1.5055	-1.4804	-1.4872	-1.4695	-1.4599	-1.4618
$E[\widehat{\rho}_3 \| BU]$	-1.1084	-1.0503	-0.9538	-0.9184	-1.1487	-1.4003	-1.4853	-1.5413
$E[\widehat{\rho}_4 \| BU]$	*-2.6431*	*-2.4320*	*-2.4092*	*-2.4025*	*-2.4162*	*-2.3498*	*-2.3093*	-2.2767
$E[\widehat{\rho}_1 \| STU]$	-2.7493	-2.5948	-2.6762	-2.6684	-2.6118	-2.5236	-2.4603	-2.4258
$E[\widehat{\rho}_2 \| STU]$	-1.8614	-1.6711	-1.5531	-1.5075	-1.4807	-1.4662	-1.4537	-1.4527
$E[\widehat{\rho}_3 \| STU]$	*-2.0470*	*-2.4821*	*-1.6067*	-1.1698	-1.0501	-1.0263	-1.0848	-1.1359
$E[\widehat{\rho}_4 \| STU]$	-2.2139	-3.2579	-4.1449	*-2.1319*	*-2.0622*	*-2.0570*	*-2.0350*	*-2.0208*

the *MSE* of $\widehat{\rho}_1$ is quite high for all n. The mean squared error of $\widehat{\rho}_2$ is, for the simulated models, smaller than that of $\widehat{\rho}_1$ for all n, being practically independent on n, and around 2.25, and consequently quite high for large n. For small values of n (say $n < 500$) neither $\widehat{\rho}_3$ nor $\widehat{\rho}_4$ show an interesting performance, and $\widehat{\rho}_2$ performs better; anyway we think that the alternative to be considered here is $\widehat{\rho}_4$, which exhibits a nice behaviour for large n.

Remark 3.3. *The first two estimators induce in general high variances whenever incorporated in any of the γ-estimators (1.5), (1.6), (1.7) and (1.9). Consequently we do not advise the use, in practice, of Peng's "asymptotically unbiased" estimator (Peng, 1998). More than that, we do not advise the incorporation neither of $\widehat{\rho}_1$ nor of $\widehat{\rho}_2$ in any of the estimators herewith proposed.*

Remark 3.4. *As may be seen from both tables, with the ρ-estimators in (3.3) we obtain in general nice distributional properties. For values of $\rho \geq -1$ the estimator $\widehat{\rho}_3$ in (3.4) exhibits a nice performance for values of $n > 200$ and should be the one to be used. For values of $\rho < -1$, here illustrated with $\rho = -2$, the best performance for large n is achieved by the estimator $\widehat{\rho}_4$ in (3.5), the one we think should be used in practice in this region of ρ-values.*

Table 2: Root Mean Squared Errors (RMSE) of the estimators $\widehat{\rho}_j$, $j = 1, 2, 3$.

n	100	200	500	1000	2000	5000	10000	20000
				$\rho = -\mathbf{0.5}$				
$RMSE[\widehat{\rho}_1\|BU]$	1.5364	1.0314	0.6550	0.4850	0.3829	0.2946	0.2471	*0.2107*
$RMSE[\widehat{\rho}_2\|BU]$	1.8208	1.7061	1.5771	1.5129	1.4307	1.3150	1.2119	1.0372
$RMSE[\widehat{\rho}_3\|BU]$	*0.2559*	*0.2546*	*0.2520*	*0.2495*	*0.2568*	*0.2374*	*0.2258*	0.2165
$RMSE[\widehat{\rho}_4\|BU]$	1.5834	1.6406	1.6915	1.7177	1.5041	1.2010	1.0898	1.0125
$RMSE[\widehat{\rho}_1\|STU]$	3.1414	1.6520	0.6790	0.4199	0.3044	0.2168	*0.1733*	*0.1408*
$RMSE[\widehat{\rho}_2\|STU]$	2.1210	1.8668	1.6568	1.56 22	1.4585	1.3497	1.2415	1.0321
$RMSE[\widehat{\rho}_3\|STU]$	*1.6047*	1.8313	*0.1718*	*0.1785*	*0.1870*	*0.1880*	0.1775	0.1696
$RMSE[\widehat{\rho}_4\|STU]$	14.6614	*0.9646*	1.0377	1.1382	1.2173	1.2273	1.1425	1.0815
				$\rho = -\mathbf{1}$				
$RMSE[\widehat{\rho}_1\|FRE]$	3.6317	3.1025	2.6860	2.4324	2.1957	1.8904	1.6960	1.5403
$RMSE[\widehat{\rho}_2\|FRE]$	*1.6067*	*1.5410*	1.5038	1.4777	1.4825	1.4665	1.4682	1.4449
$RMSE[\widehat{\rho}_3\|FRE]$	9.0992	2.9106	*0.4006*	*0.2438*	*0.3198*	*0.3201*	*0.2895*	*0.2608*
$RMSE[\widehat{\rho}_4\|FRE]$	28.2725	144.5569	1.7250	1.5147	1.2882	0.9725	0.8467	0.7580
$RMSE[\widehat{\rho}_1\|BU]$	2.8999	2.3449	1.9016	1.5914	1.3472	1.0400	0.8541	0.6991
$RMSE[\widehat{\rho}_2\|BU]$	1.5953	1.5344	1.4740	1.4548	1.4289	1.4277	1.4254	1.4375
$RMSE[\widehat{\rho}_3\|BU]$	*0.2073*	*0.1968*	*0.2060*	*0.2141*	*0.1599*	*0.1199*	*0.1070*	*0.0976*
$RMSE[\widehat{\rho}_4\|BU]$	1.1844	1.1870	1.2272	1.2478	1.0749	0.8218	0.7286	0.6639
$RMSE[\widehat{\rho}_1\|STU]$	3.9900	2.9526	1.8545	1.3505	0.9580	0.6184	0.4535	0.3439
$RMSE[\widehat{\rho}_2\|STU]$	*1.8435*	*1.6368*	1.5040	1.4606	1.4422	1.4043	1.3607	1.3356
$RMSE[\widehat{\rho}_3\|STU]$	15.0276	3.2851	*0.2684*	*0.2500*	*0.2480*	*0.2471*	*0.2439*	*0.2403*
$RMSE[\widehat{\rho}_4\|STU]$	52.0324	7.3018	0.6418	0.7002	0.7603	0.7710	0.7009	0.6483
				$\rho = -\mathbf{2}$				
$RMSE[\widehat{\rho}_1\|BU]$	3.7920	3.2483	2.8056	2.5754	2.3802	2.1339	1.9889	1.8553
$RMSE[\widehat{\rho}_2\|BU]$	*1.5288*	*1.5152*	1.5176	1.5072	1.5074	1.5076	1.5003	1.5091
$RMSE[\widehat{\rho}_3\|BU]$	2.2149	7.4335	1.0508	1.0836	0.8566	0.6257	0.5426	0.4831
$RMSE[\widehat{\rho}_4\|BU]$	13.0856	1.6843	*0.4323*	*0.4093*	*0.4463*	*0.5079*	*0.3976*	*0.3473*
$RMSE[\widehat{\rho}_1\|STU]$	4.1229	3.5374	3.0790	2.8066	2.5536	2.2527	2.0552	1.9229
$RMSE[\widehat{\rho}_2\|STU]$	*1.6125*	*1.5462*	*1.5250*	1.5139	1.5065	1.5111	1.5084	1.5208
$RMSE[\widehat{\rho}_3\|STU]$	69.2614	89.7699	13.7532	*1.4958*	0.9696	0.9779	0.9189	0.8672
$RMSE[\widehat{\rho}_4\|STU]$	27.8365	84.3638	183.4609	2.7067	*0.3722*	*0.1455*	*0.1276*	*0.1096*

Remark 3.5. *Notice that the "best" choice of τ in (3.3) is easily achieved in practice due to the high stability of the sample path of the adequate ρ-estimator, for a wide range of large k values. This means that, given a sample, the best thing is not to work automatically, but to draw a few sample paths of the ρ-estimator based on (3.3) for different values of τ, say $\tau = 0, 0.5, 1$. It is then obvious to choose τ, on the basis of any stability criterion for large values of k. At the moment, this class of estimators of the second order parameter is really the one we strongly advise to the practioner.*

Remark 3.6. *We would like to add that when we choose the fixed level $k = k_1$ in $T_n^{(\tau)}(k)$, and consider the associated ρ-estimators $\widehat{\rho}_n^{(\tau)}(k_1)$, we are obviously loosing efficiency relatively to the best estimator $\widehat{\rho}_n^{(\tau)}(k_0^{(\tau)})$, with $k_0^{(\tau)} := \arg\min_k MSE[\widehat{\rho}_n^{(\tau)}(k)]$. In Table 3 we present, for $\tau = 0$ and $\tau = 1$ the quotient between the Root Mean Squared Error (RMSE) of such an optimal estimator of ρ (and ideal, because we do not have yet an available technique for the estimation of $k_0^{(\tau)}$) and the estimator herewith considered, i.e.,*

$$Q_n^{(\tau)}|model := \frac{RMSE[\widehat{\rho}_n^{(\tau)}(k_0^{(\tau)})|model]}{RMSE[\widehat{\rho}_n^{(\tau)}(k_1)|model]},$$

for $\tau = 0$ and $\tau = 1$, and for the same models simulated before. The values of simulated $Q_n^{(\tau)}$'s larger than 50% are written in **bold**.

Table 3: Simulated $Q_n^{(\tau)}|model$ for $\tau = 0$, $\tau = 1$ and Fréchet, Burr and Student models.

n	100	200	500	1000	2000	5000	10000	20000
				$\tau = \mathbf{0}$				
$BU(-.5)$	**97.77%**	**93.03%**	**85.57%**	**79.76%**	**70.95%**	**68.18%**	**64.94%**	**61.10%**
$STU(-.5)$	13.07%	10.33%	**94.86%**	**82.50%**	**71.62%**	**62.84%**	**60.23%**	**56.96%**
$FRECHET$	46.57%	**51.34%**	**88.15%**	**99.97%**	**63.24%**	**52.92%**	**52.39%**	**52.29%**
$BU(-1)$	**98.13%**	**89.74%**	**74.66%**	**64.98%**	**78.89%**	**91.54%**	**92.30%**	**90.76%**
$STU(-1)$	5.14%	8.94%	**94.87%**	**99.81%**	**96.64%**	**90.11%**	**86.20%**	**82.44%**
$BU(-2)$	**82.37%**	13.43%	**80.47%**	**69.03%**	**77.32%**	**89.64%**	**91.07%**	**89.06%**
$STU(-2)$	3.96%	2.49%	7.93%	**66.75%**	**93.15%**	**80.56%**	**76.42%**	**71.35%**
				$\tau = \mathbf{1}$				
$BU(-.5)$	**61.35%**	**51.98%**	42.13%	36.18%	36.12%	37.82%	36.69%	34.38%
$STU(-.5)$	6.69%	**87.96%**	**68.82%**	**55.29%**	45.41%	38.17%	35.95%	33.67%
$FRECHET$	18.08%	2.42%	**89.45%**	**87.53%**	**82.28%**	**84.30%**	**80.78%**	**75.96%**
$BU(-1)$	**86.76%**	**76.52%**	**62.54%**	**53.44%**	**53.61%**	**58.05%**	**56.57%**	**53.04%**
$STU(-1)$	2.52%	10.63%	**91.52%**	**71.02%**	**56.30%**	44.48%	41.34%	36.92%
$BU(-2)$	23.22%	44.36%	**100.00%**	**99.99%**	**88.92%**	**75.62%**	**94.45%**	**98.84%**
$STU(-2)$	11.72%	5.14%	0.36%	8.89%	49.24%	**90.31%**	**82.03%**	**76.27%**

As may be seen there is a loss, although not drastic. But the choice of the fixed level to be considered in the semi-parametric estimators of ρ studied in Fraga Alves et al. (2001) still needs to be improved. Particularly with the statistic $\widehat{\rho}_4 \equiv \widehat{\rho}_n^{(1)}(k_1)$ and for the Student model we loose drastically efficiency when we fix the level at k_1.

4 The patterns of Mean Values and Mean Squared Errors of the tail index estimators

The estimators $\gamma_n^{ML(\widehat{\rho}_i)}$, $i = 3$ or 4, are not a long way from what would happen should we consider ρ known and the estimator $\gamma_n^{ML(\rho)}$. Their best performance is exhibited for large values of n, and is particularly relevant for values of $\rho < -1$. We shall not go too far with their simulated distributional properties, mainly due to the computational time associated to the expression (1.9), but we shall illustrate the Mean Values and Mean Squared Error structures of the estimators for samples of size $n = 5000$ from Fréchet, Burr and Student models. As said before, the underlying simulation is based on 5000 runs.

We present in Figures 2 and 3 the simulated mean values and MSE's of $\gamma_n^H(k)$, $\gamma_n^{GJ(\widehat{\rho}_3)}(k)$, $\gamma_n^{NGJ(\widehat{\rho}_3)}(k)$ and $\gamma_n^{ML(\widehat{\rho}_3)}(k)$, for a sample size $n = 5000$ from *Fréchet* and *Burr* parents, with $\rho = -1$ and $\gamma = 1$. In Figure 3 we also picture $\gamma_n^{ML(\rho)}$, which really behaves quite well, clearly overpassing $\gamma_n^{ML(\widehat{\rho}_3)}$. For the Fréchet model, $\gamma_n^{ML(\rho)}$ has a behaviour quite similar to that of $\gamma_n^{ML(\widehat{\rho}_3)}$ and so was not pictured in Figure 2. For all simulated models with $\rho = -1$, $\gamma_n^{ML(\widehat{\rho}_3)}$ exhibits the best bevaviour among the estimators considered, not a long way from $\gamma_n^{GJ(\widehat{\rho}_3)}$. The worst behaviour is exhibited by $\gamma_n^{NGJ(\widehat{\rho}_3)}$, but with smooth sample paths. Unexpetedly, $\gamma_n^{NGJ(\widehat{\rho}_4)}$ exhibits an interesting behaviour for all models with $\rho = -1$, as illustrated, later on, in Figures 10 and 11.

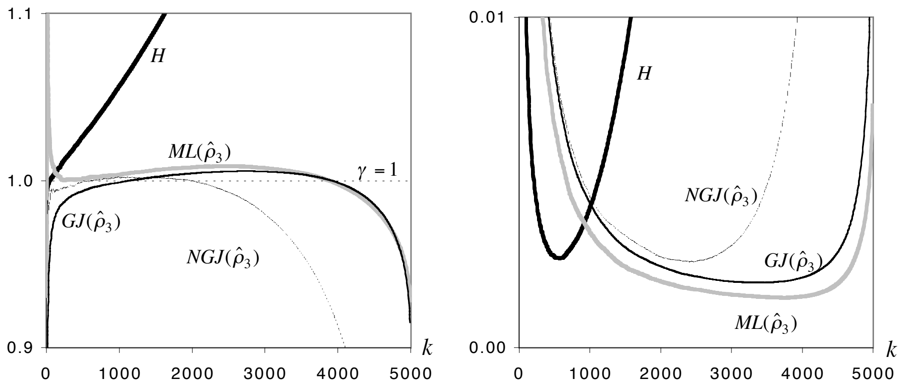

Figure 2: Simulated mean values and *MSE's* of the Hill estimator and of the estimators under study, for a sample size $n = 5000$, from a *Fréchet* parent with $\gamma = 1$ ($\rho = -1$).

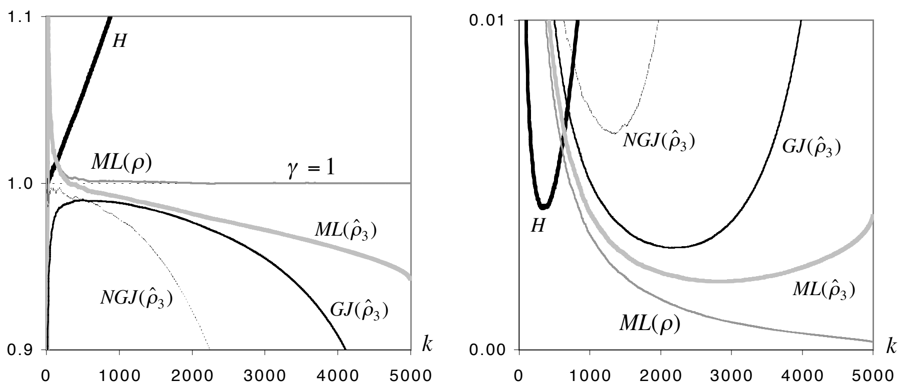

Figure 3: Simulated mean values and *MSE's* of the Hill estimator and of the estimators under study, for a sample size $n = 5000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -1$.

Next, in Figures 4 and 5, we present the patterns of the mean values and mean squared errors of the estimators under study, but for parents with $\rho > -1$, illustrated with the value ρ equal to -0.5, in Burr and Student models, respectively. We also picture again $\gamma_n^{ML(\rho)}$. In this region there is a high discrepancy between $\gamma_n^{ML(\widehat{\rho}_3)}$ and $\gamma_n^{ML(\rho)}$, and $\gamma_n^{GJ(\widehat{\rho}_3)}$ exhibits a better behaviour than $\gamma_n^{ML(\widehat{\rho}_3)}$. All the estimators herewith proposed clearly overpass the Hill estimator, provided we may use a slightly larger sample fraction, usually available in practice.

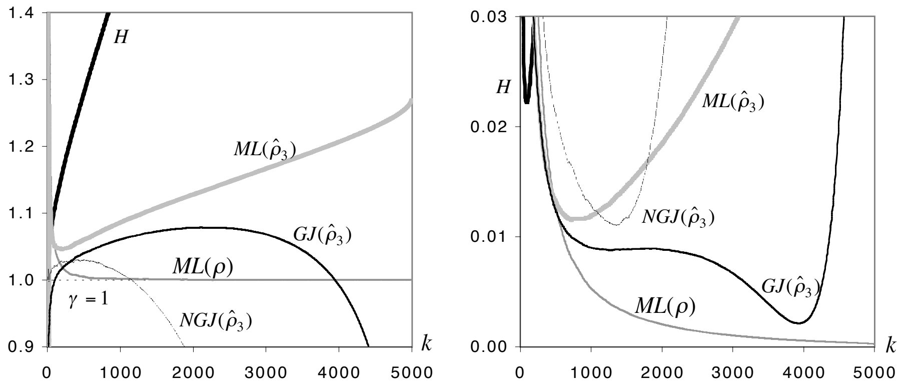

Figure 4: Simulated mean values and *MSE's* of the Hill estimator and of the estimators under study, for a sample size $n = 5000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -0.5$.

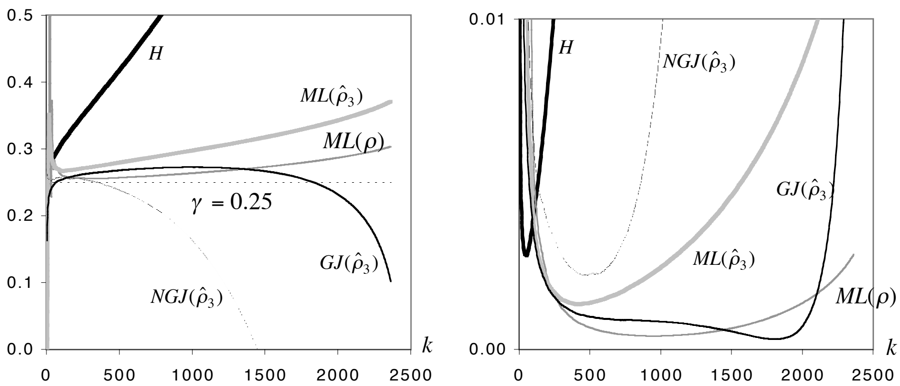

Figure 5: Simulated mean values and *MSE's* of the Hill estimator and of the estimators under study, for a sample size $n = 5000$, from a *Student* parent with $\nu = 4$ degrees of freedom ($\gamma = 0.25$ and $\rho = -0.5$).

In Figures 6 and 7, and with samples from a Burr parent with $\rho = -2$ and a Student parent with $\nu = 1$ degree of freedom, respectively, we illustrate what happens in the region $\rho < -1$, a region where has been difficult to find competitors for the Hill estimator, which already exhibits a nice performance. In these figures we picture $\gamma_n^{GJ(\widehat{\rho}_3)}$, $\gamma_n^{NGJ(\widehat{\rho}_3)}$ and $\gamma_n^{ML(\widehat{\rho}_3)}$, which do not overpass the Hill estimator at the optimal level, together with $\gamma_n^{ML(\rho)}$, where we replace the unknown ρ by its true value -2; we are obviously aware that this is not at all sensible from a practical point of view, but the results obtained suggest that we need to work with a better estimator of ρ. Indeed if we had worked with either $\widehat{\rho}_1$ or $\widehat{\rho}_2$ the tail index estimators under study would have a much worse performance, particularly in what concerns the mean squared error structure, which would lie totally outside the sketched region, and with a terribly erratic behaviour.

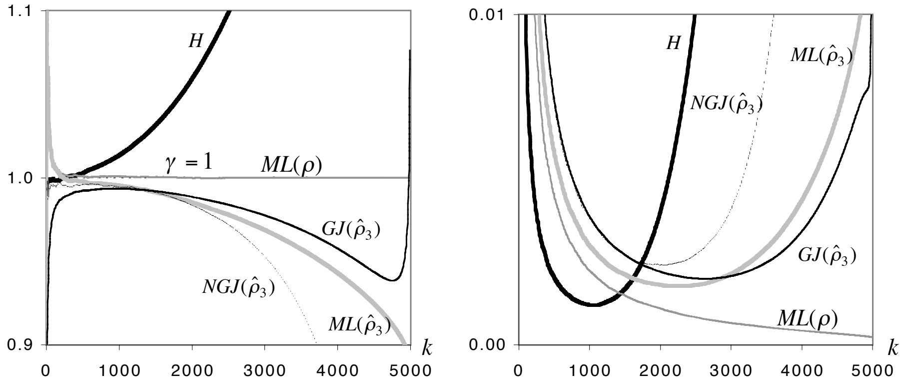

Figure 6: Simulated mean values and *MSE's* of the Hill estimator, of the estimators $\gamma_n^{\bullet(\widehat{\rho}_3)}$ and of the r.v. $\gamma_n^{ML(\rho)}$, for a sample size $n = 5000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -2$.

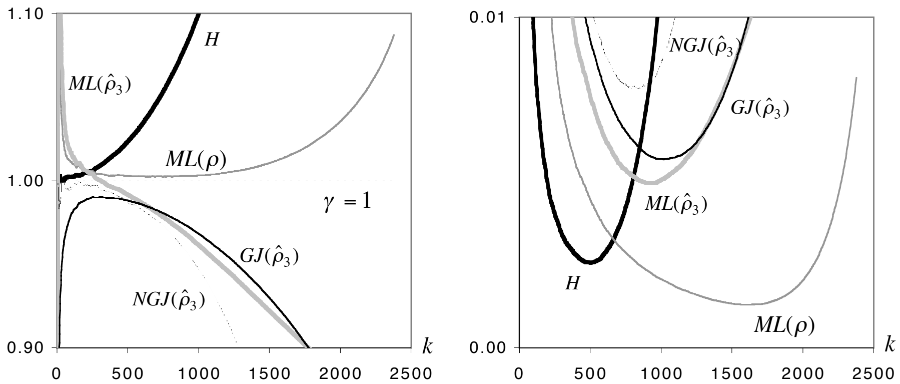

Figure 7: Simulated mean values and *MSE's* of the Hill estimator, of the estimators $\gamma_n^{\bullet(\widehat{\rho}_3)}$, and of $\gamma_n^{ML(\rho)}$, for a sample size $n = 5000$, from a *Student* parent with $\nu = 1$ degree of freedom ($\gamma = 1$ and $\rho = -2$).

Figures 8 and 9 are equivalent to Figures 6 and 7, but now with $\widehat{\rho}_3$ replaced by $\widehat{\rho}_4$. We now get much better results, being able to overpass the Hill estimator at the optimal level. Maybe it is even worth to invest in the adaptive choice of the high level to be considered in the semi-parametric estimation of ρ, eventually with the use of the bootstrap methodology, like has been done for the optimal choice of k in the semi-parametric estimation of the tail index γ (see for instance Draisma *et al.* (1999) and Danielsson *et al.* (2001)). In this region of ρ-values $\gamma_n^{ML(\widehat{\rho}_4)}$ exhibits the best behaviour among all estimators herewith considered.

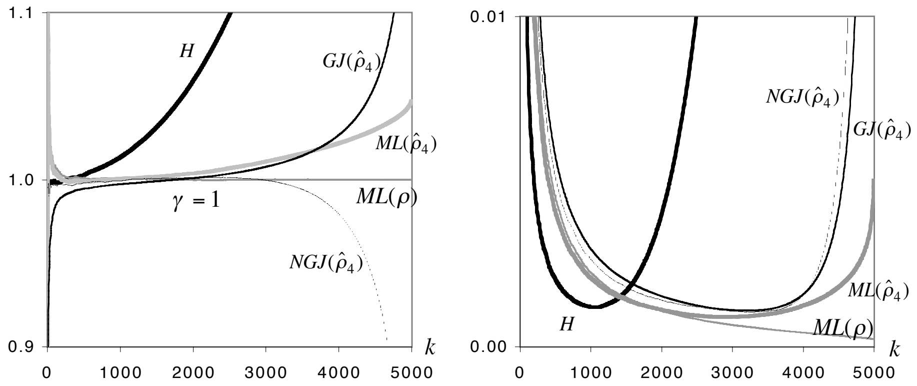

Figure 8: Simulated mean values and *MSE's* of the Hill estimator, of the estimators $\gamma_n^{\bullet(\hat{\rho}_4)}$ and of the r.v. $\gamma_n^{ML(\rho)}$, for a sample size $n = 5000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -2$.

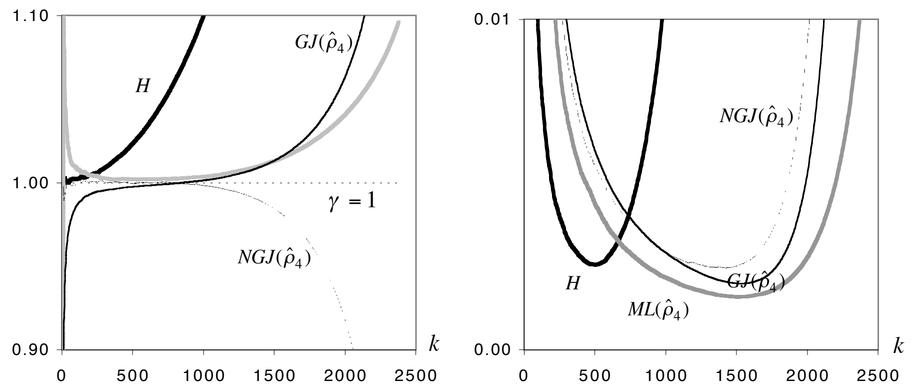

Figure 9: Simulated mean values and *MSE's* of the Hill estimator, of the estimators $\gamma_n^{\bullet(\hat{\rho}_4)}$, and of $\gamma_n^{ML(\rho)}$, for a sample size $n = 5000$, from a *Student* parent with $\nu = 1$ degree of freedom ($\gamma = 1$ and $\rho = -2$).

The Generalized Jackknife estimators studied in Gomes *et al.* (2000a), i.e., $\gamma_n^{GJ(-1)}$ and $\gamma_n^{NGJ(-1)}$, as well as the estimator $\gamma_n^{ML(-1)}$ studied in Gomes and Martins (2002), although devised for $\rho = -1$, had already some general nice properties. We put again in evidence some of the overall conclusions drawn in the two above mentioned papers, adding some extra comments, mainly related to the estimators studied here.

Remarks:

1. The estimators $\gamma_n^{\bullet(-1)}$ were already highly efficient in Hall's class of parent

distributions for values of $\rho \geq -1$, the most problematic region for the Hill estimator (see Gomes *et al.* (2000a) and Gomes and Martins (2002)). However, that was sometimes due to a global minimum of the *MSE*, for a very large value of k, difficult to justify as intermediate. In this same region of ρ-values, the estimators $\gamma_n^{\bullet(\widehat{\rho}_3)}$ have even a higher efficiency, without exhibiting a *MSE* with such a peculiar behaviour.

2. For large values of $|\rho| \geq 1$ and models in Hall's class, here illustrated in Figures 6, 7, 8 and 9 with $\rho = -2$, the Hill estimator compared favourably to the alternatives considered in Gomes *et al.* (2000a) and Gomes and Martins (2002). We had however noticed there that should we replace ρ by its true value (an irrelevant methodology from a practical point of view, since ρ is not known), the r.v.'s $\gamma_{n0}^{\bullet(\rho)}$ would clearly overpass γ_{n0}^{H}. The same happens when we adequately estimate ρ, through $\widehat{\rho}_4$ (see Figures 8 and 9).

3. The most attractive features of the Generalized Jackknife estimators are on one side their stable sample paths (for a wide region of k values), close to the target value γ, (look at all Figures shown in this section, where the mean values provide a hint of what happens to the sample paths), and on the other side the "bath-tube" patterns of their $MSE(k)$ function, which make less relevant the choice of the optimal sample fraction k_0/n.

5 Relative efficiencies, bias reduction at the optimal level, and sensitiveness to the choice of the threshold

In the following Figures we have placed together the four indicators, $REFF$ and STI_1 at the top, and BRI and STI_2 at the bottom. Recall that the first two indicators are related to the mean squared error structure of the estimators, and the last two indicators are related to the mean value structure of the estimators. Regarding mean squared error the best estimator, among the ones considered, would be the one with highest $REFF$ and STI_1 indicators — an association of a very high $REFF$ together with a very low STI_1 means that the MSE of the estimator at the optimal level is much smaller than that of the Hill estimator, also at its optimal level, but with a sharp shape. A trade off between the $REFF$ and the STI_1 indicators would be preferable than this latter situation. Similarly, and regarding bias reduction, the best estimator would be the one with highest BRI and STI_2 indicators — again, a very high BRI together with a very low STI_2 means that the mean value of the estimator at its optimal level is quite close to the target value γ, but it does not stay close to γ for a long time. Again, if necessary, it is preferable to elect an estimator on the basis of a trade-off between the BRI and the STI_2 indicators. Among all the indicators considered, we think sensible to base our selection preferencially on high values of $REFF$, associated with high values of STI_2.

The message obtained for the different models simulated with $\rho = -1$, with $|\rho| < 1$ and with $|\rho| > 1$ is similar within each of these regions. This led us to present in Figure 10 the indicators for a Fréchet model, the classical heavy tail model, with $\rho = -1$. Next, in Figures 11, 12 and 13 we picture the above mentioned indicators only for *Burr* parents with $\rho = -1$, $\rho = -0.5$ and $\rho = -2$, respectively.

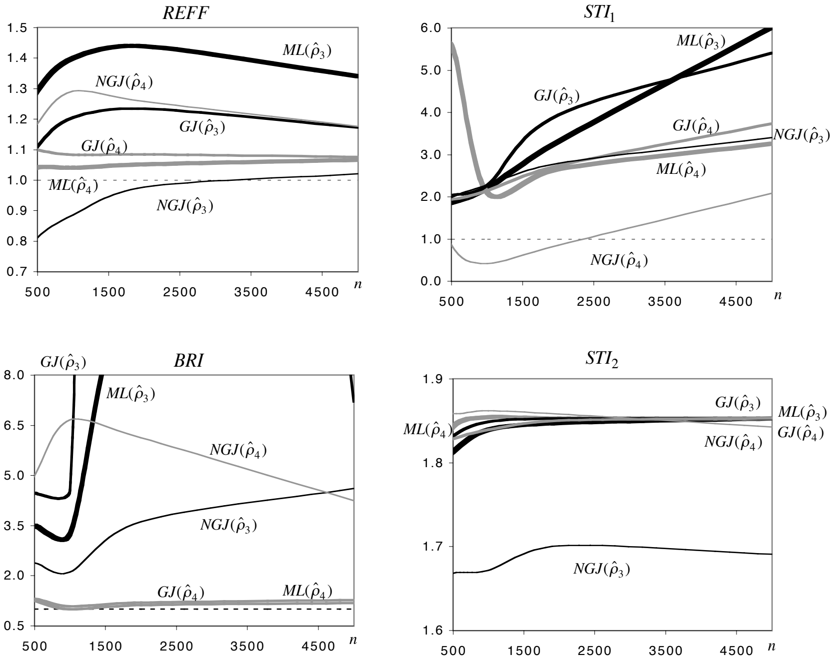

Figure 10: Simulated indicators associated to the estimators under study, for sample sizes till $n = 5000$, from a *Fréchet* parent with $\gamma = 1$ ($\rho = -1$).

As reported in Gomes and Martins (2001), for $\rho = -1$ even the consideration of $\gamma_{n0}^{GJ(\widehat{\rho}_1)}$ leads to a very high reduction in bias, which is totally compensated by a drastic increase in variance, leading, for instance, for Fréchet models, to efficiencies around 50% relatively to the Hill estimator at its optimal level. Such a loss of efficiency does not occur when we consider either $\widehat{\rho}_3$ or $\widehat{\rho}_4$.

Some overall comments.

1. For models with $\rho = -1$, we elect with no doubt $\gamma_n^{ML(\widehat{\rho}_3)}$ among all the estimators considered, but not a long way from $\gamma_n^{GJ(\widehat{\rho}_3)}$. It is convenient to say that the relative efficiencies of the estimators under study are greater

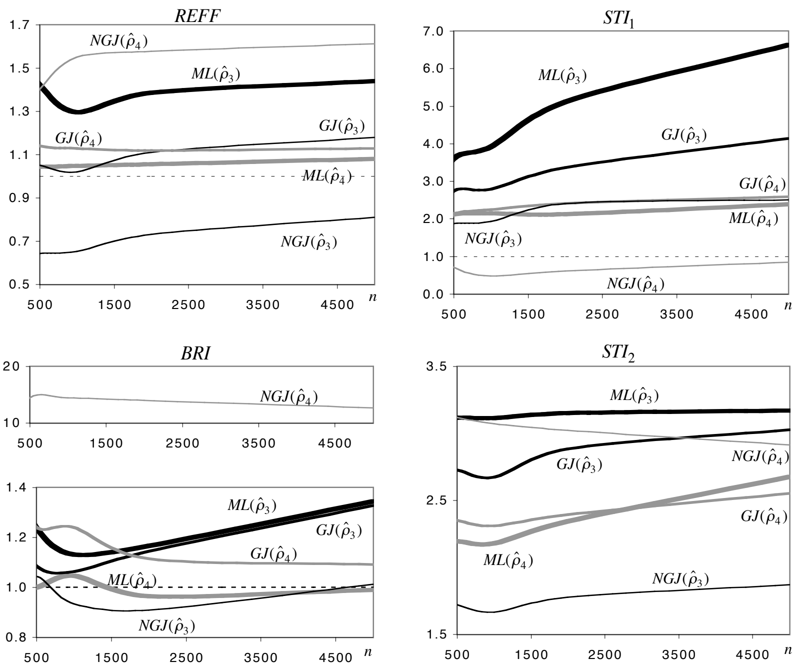

Figure 11: Simulated indicators associated to the estimators under study, for sample sizes till $n = 5000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -1$.

than 1, for large n, only when we use either $\widehat{\rho}_3$ or $\widehat{\rho}_4$ (and not $\widehat{\rho}_1$ or $\widehat{\rho}_2$). Some additional considerations: the relative efficiency of $\gamma_n^{NGJ(\widehat{\rho}_4)}$ is very high at this same region, mainly due to a high bias reduction indicator, but at the expenses of very low STI_1 indicators; for Burr and Student models the relative efficiency of $\gamma_n^{NGJ(\widehat{\rho}_3)}$ is smaller than 1 even for $n = 5000$.

2. For values of ρ close to 0, illustrated with a *Burr* parent with $\rho = -0.5$ (Figure 12) we may conclude that the relative efficiencies are increasing with n, for reasonably large n; this is mainly due to high reductions in bias for large n. A global analysis of all the indicators would lead us to elect, in this region of ρ-values and among the estimators considered, the Generalized Jackknife estimator $\gamma_n^{GJ(\widehat{\rho}_3)}$.

3. Finally, in the region $|\rho| > 1$, illustrated with a *Burr* parent with $\rho = -2$ (Figure 13), the situation is not so clear. With the use of $\widehat{\rho}_3$, there is no decrease in the bias of $\gamma_n^{\bullet(\widehat{\rho}_3)}$ for the simulated sample sizes, and consequently the relative efficiency is always smaller than 1 for the simulated values of n. But with the use of $\widehat{\rho}_4$ we are able to get relative efficiencies higher than one, due to the reduction of bias achieved. Again the global analysis of all indicators would lead us to elect, among the estimators considered, $\gamma_n^{ML(\widehat{\rho}_4)}$, not a long way from $\gamma_n^{GJ(\widehat{\rho}_4)}$ and $\gamma_n^{NGJ(\widehat{\rho}_4)}$.

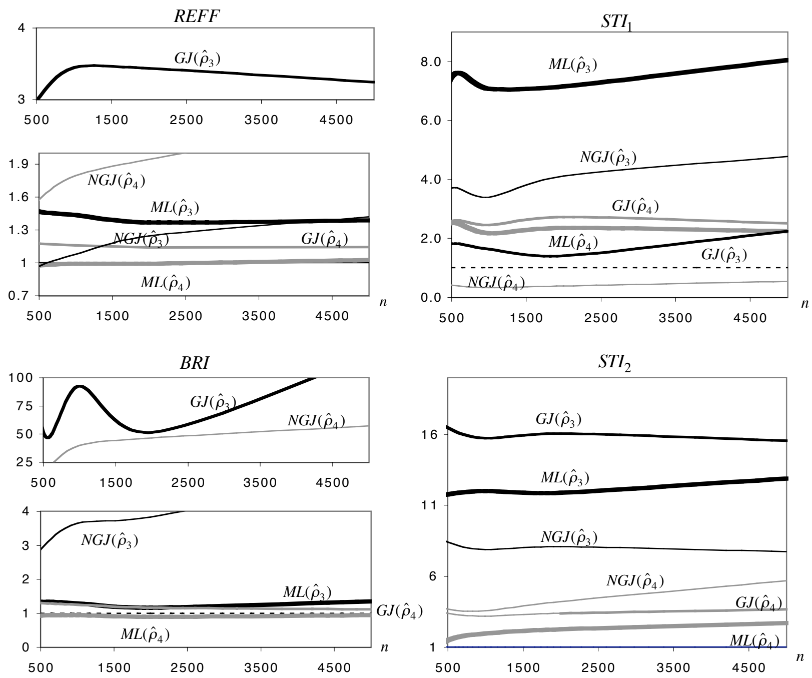

Figure 12: Simulated indicators associated to the estimators under study, for sample sizes till $n = 5000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -0.5$.

4. In general, the most smooth sample paths were attained by either $\gamma_n^{NGJ(\widehat{\rho}_3)}$ or $\gamma_n^{NGJ(\widehat{\rho}_4)}$. Among the estimators using $\widehat{\rho}_4$, the best estimator is often $\gamma_n^{NGJ(\widehat{\rho}_4)}$; but among the estimators using $\widehat{\rho}_3$, $\gamma_n^{NGJ(\widehat{\rho}_3)}$ is the worst one.

5. The adaptive choice of k in the semi-parametric estimators of ρ herewith considered is under development, and seems to be a promising way of improving the ρ-estimator, and consequently of improving the estimators of the tail index γ with a null dominant component of asymptotic bias.

Two final comments.

- The removal of the bias term of the order of $A(n/k)$ improves greatly the estimation of the tail index γ, both from the point of view of the sample paths of the new estimators, which exhibit a much larger stability region around the target value, and from the point of view of the MSE structure, as a function of k, which has now a "bath-tube" shape, making then less relevant the choice of an optimal sample fraction.

- Difficulties with the estimation of the second order parameter ρ, which from our point of view seem to have been now overcome, for positive γ, with the methods of estimation proposed in Fraga Alves et al. (2001),

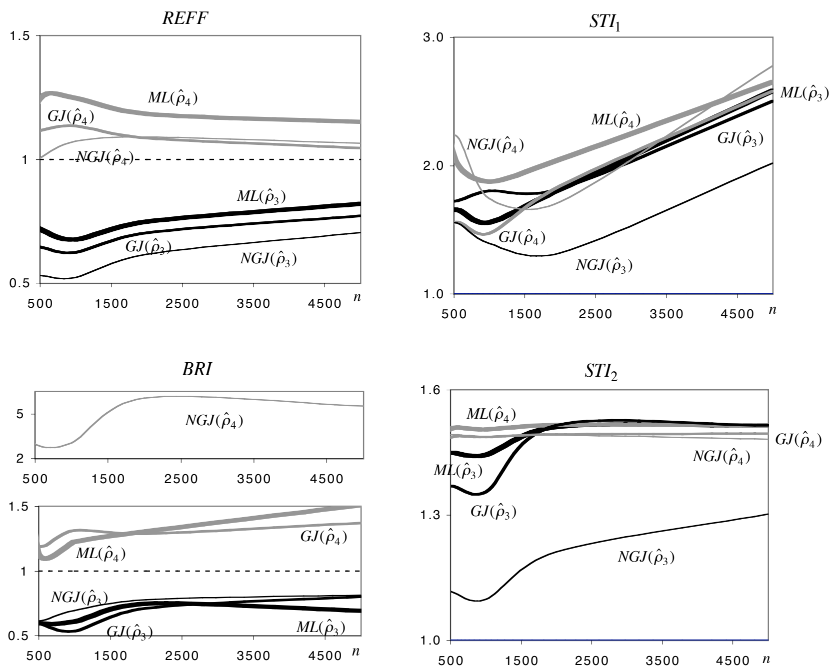

Figure 13: Simulated indicators associated to the estimators under study, for sample sizes till $n = 5000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -2$.

may lead to a mispecification of ρ, in -1, and the consideration of $\gamma_n^{ML} \equiv \gamma_n^{ML(-1)}$. This estimator already exhibits an interesting behaviour for a large diversity of models with a second order parameter in the region $|\rho| \leq 1$, as may be seen in Gomes and Martins (2002). An external estimation of ρ leads to an improvement also for models with $|\rho| > 1$. The use of such an external estimator has, over the joint estimation of ρ, γ and D by maximum likelihood, like has been done in Feuerverger and Hall (1999), the great advantage of not requiring the numerical resolution of a system of non-linear equations, with serious convergence problems, leading also to a smaller asymptotic variance.

References

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

[2] Beirlant, J., Vynckier, P. and J.L. Teugels (1996b). Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.* **91**, 1659-1667.

[3] Danielsson, J., Haan, L. de, Peng, L. and C.G. de Vries (2001). *Using a bootstrap method to choose the sample fraction in the tail index estimation. J. Multivariate Analysis* **76**, 226-248.

[4] Dekkers, A.L.M., J.H.J. Einmahl and L. de Haan (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17**, 1833-1855.

[5] 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.

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

[7] Fraga Alves, M.I., Gomes, M.I. and L. de Haan (2001). *A new class of semi-parametric estimators of the second order parameter.* Notas e Comunicações CEAUL 4/01. Submitted.

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

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

[10] Gomes, M. I. and Martins, M. J. (2001). *Generalized Jackknife estimators of the tail index based on the estimation of the second order parameter.* Notas e Comunicações C.E.A.U.L. 9/2001.

[11] Gomes, M.I. and Martins, M.J. (2002). *Bias reduction and explicit efficient estimation of the tail index.* Notas e Comunicações C.E.A.U.L. 6/2002.

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

[13] Gomes, M.I., de Haan, L. and Peng, L. (2000b). *Semi-parametric estimation of the second order parameter — asymptotic and finite sample behaviour.* Notas e Comunicações CEAUL 8/2000. Submitted.

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

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

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

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