Statistics of Extremes — discrepancy between asymptotic and finite sample behaviour*

M. Ivette Gomes
CEAUL and DEIO (FCUL), Universidade de Lisboa
and
Fernanda Figueiredo
CEAUL and Faculdade de Economia da Universidade do Porto

Abstract. For regularly varying tails, estimation of the index of regular variation or tail index, γ, is often performed through the classical Hill estimator, a statistic strongly dependent on the number k of top order statistics used, and with a high asymptotic bias as k increases. On the basis of the asymptotic structure of Hill's estimator for different k-values, we here propose "asymptotically best linear unbiased" estimators (BLUE) of the tail index. A similar derivation on the basis of the log-excesses and of the scaled log-spacings has also been performed. The asymptotic behaviour of those estimators is derived, and they are compared with other alternative estimators, including the Hill estimator, both asymptotically and for finite samples. Asymptotic equivalent estimators may exhibit indeed very diversified finite sample properties.

AMS 2000 subject classification. Primary 62G32, 62E20; Secondary 65C05.
Keywords and phrases. *Statistics of Extremes, Semi-parametric estimation, BLUE estimation.*

1 Introduction and BLUE estimators

In the general theory of Statistics, whenever we ask the question whether the combination of information may improve the performance of an estimator, we are led to think on *Best Linear Unbiased Estimators* ($BLUE$), i.e., on unbiased linear combinations of an adequate set of statistics, with minimum variance among the class of such linear combinations. The basic theorem underlying this theory is due to Aitken (1935): *If* $\mathbf{X}$ *is a vector of observations with mean values* $\mathbb{E}\mathbf{X} = \mathbf{A}\theta$ *depending linearly on the unknown vector of parameters* θ*, with a known coefficient matrix* $\mathbf{A}$*, and with a covariance matrix* $\delta^2\Sigma$*, known up to a scale factor* δ^2*, the least-squares estimator of* θ *is the*

*Research partially supported by FCT / POCTI / FEDER

vector θ^* *which minimizes the quadratic form* $(\mathbf{X}-\mathbf{A}\ \theta)'\,\Sigma^{-\mathbf{1}}\,(\mathbf{X}-\mathbf{A}\ \theta)$. *Such a vector is thus the vector of solutions of the "normal equations",* $\mathbf{A}'\Sigma^{-\mathbf{1}}\mathbf{A}\ \theta^*=\mathbf{A}'\Sigma^{-\mathbf{1}}\mathbf{X}$. *This solution is explicitly given by*

$$\theta^*=\left(\mathbf{A}'\Sigma^{-\mathbf{1}}\mathbf{A}\right)^{-\mathbf{1}}\mathbf{A}'\Sigma^{-\mathbf{1}}\mathbf{X},$$

and its variance matrix is $\mathbf{Var}(\theta^*)=\delta^{\mathbf{2}}\left(\mathbf{A}'\Sigma^{-\mathbf{1}}\mathbf{A}\right)^{-\mathbf{1}}$.

Given a vector of m statistics directly related to the tail index γ, let us say

$$\underline{\mathbf{T}}\equiv\left(T_{ik},\quad i=k-m+1,\cdots,k\right),\quad 1\le m\le k,$$

where k is intermediate, i.e.,

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

let us assume that, asymptotically, the covariance matrix of $\underline{\mathbf{T}}$ is well approximated by $\gamma^2\Sigma$, i.e., it is known up to the scale factor γ^2, and that its mean value is asymptotically well approximated by

$$\gamma\ \underline{\mathbf{s}}+\varphi(n,k)\ \underline{\mathbf{b}}.$$

It is thus sensible to think on the following question: Is it possible to find a linear combination of our set of statistics with "minimum variance" and "unbiased" in an asymptotic sense? Such a linear combination will be called an "asymptotically best linear unbiased" estimator and will be denoted BL_T.

Our objective is then to find a vector $\underline{\mathbf{a}}'=(a_1,a_2,\cdots,a_m)$ such that $\underline{\mathbf{a}}'\ \Sigma\ \underline{\mathbf{a}}$ is minimum, subject to the conditions $\underline{\mathbf{a}}'\ \mathbf{s}=1$ and $\underline{\mathbf{a}}'\ \underline{\mathbf{b}}=0$. The solution of such a problem is easily obtained if we consider the function,

$$H(\underline{\mathbf{a}};\alpha,\beta)=\underline{\mathbf{a}}'\ \Sigma\ \underline{\mathbf{a}}-\alpha\left(\underline{\mathbf{a}}'\ \mathbf{s}-1\right)-\beta\underline{\mathbf{a}}'\ \underline{\mathbf{b}},$$

and obtain the solution of the stationarity equations:

$$\left\{\begin{array}{l}2\Sigma\underline{\mathbf{a}}-\alpha\underline{\mathbf{s}}-\beta\underline{\mathbf{b}}=0\\ \underline{\mathbf{a}}'\ \mathbf{s}=\mathbf{1}\\ \underline{\mathbf{a}}'\ \underline{\mathbf{b}}=\mathbf{0}\end{array}\right.. \tag{1.2}$$

From the first equation in (1.2) we get $2\ \Sigma\ \underline{\mathbf{a}}=\mathbf{P}\begin{bmatrix}\alpha\\ \beta\end{bmatrix}$, with $\mathbf{P}=[\underline{\mathbf{s}}\quad\underline{\mathbf{b}}]$, and consequently,

$$\underline{\mathbf{a}}=\tfrac{1}{2}\ \Sigma^{-1}\mathbf{P}\begin{bmatrix}\alpha\\ \beta\end{bmatrix}. \tag{1.3}$$

Then $\mathbf{P}'\underline{\mathbf{a}} = \frac{1}{2}\left(\mathbf{P}'\Sigma^{-1}\mathbf{P}\right)\begin{bmatrix}\alpha\\ \beta\end{bmatrix}$, and, if there exists $\left(\mathbf{P}'\Sigma^{-1}\mathbf{P}\right)^{-1}$,

$$\begin{bmatrix}\alpha\\ \beta\end{bmatrix} = 2\ \left(\mathbf{P}'\Sigma^{-1}\mathbf{P}\right)^{-1}\mathbf{P}'\underline{\mathbf{a}}.$$

But from the last two equations in (1.2) we get $\underline{\mathbf{a}}'\mathbf{P} = [1\quad 0]$, i.e., $\mathbf{P}'\underline{\mathbf{a}} = \begin{bmatrix}1\\ 0\end{bmatrix}$, and consequently,

$$\begin{bmatrix}\alpha\\ \beta\end{bmatrix} = 2\ \left(\mathbf{P}'\Sigma^{-1}\mathbf{P}\right)^{-1}\begin{bmatrix}1\\ 0\end{bmatrix}. \tag{1.4}$$

If we incorporate (1.4) in (1.3), we get

$$\underline{\mathbf{a}} = \Sigma^{-1}\mathbf{P}\left(\mathbf{P}'\Sigma^{-1}\mathbf{P}\right)^{-1}\begin{bmatrix}1\\ 0\end{bmatrix} = \frac{1}{\Delta}\mathbf{b}'\Sigma^{-1}\left(\underline{\mathbf{s}}\ \underline{\mathbf{b}}' - \underline{\mathbf{b}}\ \underline{\mathbf{s}}'\right)\Sigma^{-1}, \tag{1.5}$$

where $\Delta = \left|\left|\mathbf{P}'\Sigma^{-1}\mathbf{P}\right|\right|$. Since we have denoted $\underline{\mathbf{T}}$ the vector of the m statistics on which we are going to base our estimation, we get the final random variable

$$BL_T^{(\rho)}(k;m) := \underline{\mathbf{a}}'\ \underline{\mathbf{T}}, \quad \underline{\mathbf{a}} \text{ given in (1.5).} \tag{1.6}$$

Provided the results were not asymptotic, could we derive that

$$\mathbb{V}ar\left(BL_T^{(\rho)}(k;m)\right) = \gamma^2\ \frac{\underline{\mathbf{b}}'\ \Sigma^{-1}\ \underline{\mathbf{b}}}{\Delta}.$$

We shall here assume to be working in a context of heavy tail models, i.e., for all $x > 0$, and with $U(t) = F^{\leftarrow}(1 - 1/t),\ t \geq 1,\ F^{\leftarrow}$ the generalized inverse of the underlying model F, one of the following equivalent conditions holds true:

$$\lim_{t\to\infty}\frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma} \quad \Longleftrightarrow \quad \lim_{t\to\infty}\frac{U(tx)}{U(t)} = x^{\gamma}, \tag{1.7}$$

with γ the above mentioned tail index. For heavy tails, the classical tail index estimator is then Hill's estimator (Hill, 1975):

$$H(k) := \frac{1}{k}\sum_{i=1}^{k}\left\{\ln X_{n-i+1:n} - \ln X_{n-k:n}\right\},$$

where $X_{i:n},\ 1 \leq i \leq n$, are the ascending order statistics (o.s.) associated to our random sample $(X_1, X_2, \cdots, X_n)$. Hill's estimator is consistent under the first order framework in (1.7) and for intermediate k, i.e., levels k such that (1.1) holds. To achieve asymptotic normality we need to assume a second order condition. We shall here assume that

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho} \tag{1.8}$$

for all $x > 0$, where A is a suitably chosen function of constant sign near infinity (positive or negative), and $\rho \le 0$ is the second order parameter. The limit function in (1.8) is necessarily of this given form, and $|A| \in RV_\rho$ (Geluk and de Haan, 1987). The notation RV_β stands for the class of *regularly varying* functions at infinity with *index of regular variation* equal to β, i.e., positive measurable functions g such that $\lim_{t\to\infty} g(tx)/g(t) = x^\beta$, for all $x > 0$.

Under the validity of (1.8), and for intermediate k, the following asymptotic distributional representation

$$H(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, P_k^H + \frac{1}{1-\rho}\, A(n/k)(1 + o_p(1)) \tag{1.9}$$

holds, where P_k^H is asymptotically a standard normal random variable.

In section 2 of this paper we shall consider "asymptotically best linear combinations" of Hill's estimators in (1.6), both under a mispecification $\rho = -1$ and for a general ρ, to be estimated under an adequate methodology. For a general ρ, we have obtained a computer time consuming estimator of the tail index γ. And since the computation time has increased for "asymptotically best linear combinations" of log-excesses, we have decided to consider in section 3 the same kind of derivation, but now based on the scaled log-spacings. Such a derivation led us to much simpler linear combinations, with almost the same exact behaviour and equivalent asymptotic properties. In section 4 we exhibit the finite sample behaviour of the "asymptotically best linear unbiased" estimators, comparatively to the Hill estimator and to the "asymptotically unbiased" estimator with smallest asymptotic variance, among the ones considered in Gomes and Martins (2002b). Finally, section 5 is devoted to the finite distributional behaviour of the estimators under consideration, through the use of Monte Carlo techniques, and to the drawing of some overall conclusions.

2 "Asymptotically unbiased" linear combinations of Hill's estimators

Let us consider Hill's estimators computed at different intermediate levels $k - m + 1, k - m + 2, \cdots, k$, i.e., let us think on the vector

$$\underline{\mathbf{H}} \equiv (H_n(k-m+1),\ H_n(k-m+2),\ \cdots,\ H_n(k)).$$

We are thus working with the top $k+1$ order statistics, down to $X_{n-k:n}$. We know that, asymptotically, the covariance matrix of $\underline{\mathbf{H}}$ is well approximated by

$$\gamma^2\ \Sigma = \gamma^2\left[\sigma_{i,j}\right], \quad \sigma_{i,j} = \sigma_{j,i} = \frac{1}{k-m+j},\ 1 \le i \le j \le m.$$

On the other side, its mean value is asymptotically well approximated by

$$\gamma \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + b_\rho A(n/k) \begin{bmatrix} \left(\frac{k}{k-m+1}\right)^\rho \\ \left(\frac{k}{k-m+2}\right)^\rho \\ \vdots \\ 1 \end{bmatrix} =: \gamma\underline{\mathbf{1}} + b_\rho\ A(n/k)\underline{\mathbf{b}}, \tag{2.1}$$

where $\rho \le 0$ is the second order parameter formally defined in (1.8), and related to the rate of convergence of normalized maximum values towards a non-degenerate limit law. Here, and taking into account (1.9), we have $b_\rho = 1/(1-\rho)$.

We get straightforwardly:

Proposition 2.1. *The inverse matrix,* $\Sigma^{-\mathbf{1}}$, *of*

$$\Sigma = \left[\sigma_{ij}\right], \quad \sigma_{i,j} = \sigma_{j,i} = \frac{1}{k-m+j},\ 1 \le i \le j \le m,$$

has entries $\sigma^{i,j}$, $1 \le i,\ j \le m$, *given by*

$$\sigma^{i,i} = \begin{cases} (k-m+1)(k-m+2) & \textit{if} \quad i=1 \\ 2(k-m+i)^2 & \textit{if} \quad i = 2,3,\cdots,m-1, \\ k^2 & \textit{if} \quad i=m \end{cases}$$

$$\sigma^{i-1,i} = \sigma^{i,i-1} = -(k-m+i-1)(k-m+i),\ i = 2,3,\cdots,m,$$

and

$$\sigma^{i,j} = 0,\ |i-j| > 1.$$

2.1 Mispecification of ρ ($\rho = -1$)

Proposition 2.2. *Under a mispecification of* ρ *at* -1,

$$\mathbf{P}'\Sigma^{-\mathbf{1}}\mathbf{P} = \begin{bmatrix} k & k \\ k & P_{mk} \end{bmatrix}$$

where

$$3\ k^2\ P_{mk} = 3\ k^3 + 3(m-1)(k-m+1)^2 \\ + 3(m-1)^2(k-m+1) + (m-1)^3 - (m-1).$$

Also

$$\Delta = \left|\left|\mathbf{P}'\Sigma^{-\mathbf{1}}\mathbf{P}\right|\right| = k\left(P_{mk} - k\right),$$

and

$$\underline{\mathbf{b}}'\Sigma^{-\mathbf{1}}\underline{\mathbf{b}} = P_{mk}.$$

If we put $m = \theta k$, $0 < \theta \leq 1$,

$$k\ \frac{\underline{\mathbf{b}}'\Sigma^{-\mathbf{1}}\underline{\mathbf{b}}}{\Delta} \underset{n\to\infty}{\longrightarrow} \frac{\theta^3 - 3\theta^2 + 3\theta + 3}{\theta\left(\theta^2 - 3\theta + 3\right)},$$

which is a decreasing function of θ, converging towards 4 as $\theta \to 1$.

It is thus sensible to put $m = k$, and we then obtain

$$P'\Sigma^{-1} = \frac{1}{\gamma^2}\left[\begin{array}{cccccc} 0 & 0 & 0 & \cdots & 0 & k \\ -\frac{2}{k} & -\frac{4}{k} & -\frac{6}{k} & \cdots & -\frac{2(k-1)}{k} & 2k-1 \end{array}\right],$$

$$\left(P'\Sigma^{-1}P\right)^{-1} = \tfrac{3}{k^2-1}\left[\begin{array}{cc} \frac{4k^2-1}{3k} & -k \\ -k & k \end{array}\right],$$

and consequently

$$a' = \tfrac{3}{k^2-1}\left[\begin{array}{cccccc} 2 & 4 & 6 & \cdots & 2(k-1) & -\frac{(k-1)(2k-1)}{3} \end{array}\right].$$

We may thus state the following result.

Proposition 2.3. *If we consider $m = k$, we get*

$$P_{kk} = \frac{4k^2-1}{3k}, \quad \Delta = \frac{k^2-1}{3}, \quad \gamma^2\ \frac{b'\Sigma^{-1}b}{\Delta} = \frac{\gamma^2}{k}\ \frac{4k^2-1}{k^2-1}.$$

The weights a_i, $i = 1, 2, \cdots, k$, in $BL_H^{(-1)}(k) = \sum_{i=1}^{k} a_i\ H(i)$, are given by

$$a_i = \frac{6i}{k^2-1},\ i = 1, 2, \cdots, k-1, \qquad a_k = -\frac{2k-1}{k+1}.$$

We are thus interested in the estimator

$$BL_{H}^{(-1)}(k) := \frac{6}{k^2-1}\sum_{i=1}^{k-1} i\ H(i) - \frac{2k-1}{k+1}\ H(k). \tag{2.2}$$

Since we have misspecified ρ, we no longer have a null asymptotic bias whenever $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \lambda$, finite and non-null, unless $\rho = -1$. We may state the following result:

Theorem 2.1. *Under the first order framework in* (1.7)*, and for intermediate k, the linear combination in* (2.2) *is consistent for the estimation of the tail index γ. If we further assume a second order framework, i.e. if we assume that* (1.8) *holds, our estimator is asymptotically normal. More specifically, the asymptotic distributional representation*

$$BL_{H}^{(-1)}(k) \stackrel{d}{=} \gamma + \frac{2\gamma}{\sqrt{k}}\ P_k^{BL_H} + \frac{2(1+\rho)}{(1-\rho)(2-\rho)} A(n/k)(1+o_p(1)) \tag{2.3}$$

holds true, where $P_k^{BL_H}$ is an asymptotically standard normal random variable.

Proof. The proof comes straightforwardly from Hill's asymptotic distributional representation in (1.9), together with the results in Proposition 2.3. The term in (2.3), related to the asymptotic variance, comes from the fact that

$$k\ \frac{\mathbf{b}'\Sigma^{-1}\mathbf{b}}{\Delta} \underset{k\to\infty}{\longrightarrow} 4$$

and the bias term comes from the fact that

$$\underline{\mathbf{a}}'\underline{\mathbf{b}} = \frac{6}{k^2-1}\left(k^{\rho}\sum_{i=1}^{k-1} i^{1-\rho} - \frac{(k-1)(2k-1)}{6}\right)$$

converges towards $2(1+\rho)/(2-\rho)$ as $k \to \infty$.

□

Remark 2.1. *Note that we are able to get a null bias for $\rho = -1$, but at the expenses of an increase in the variance, which is 4 times the asymptotic variance of Hill's estimator — the old trade-off betwen variance and bias.*

Remark 2.2. *The asymptotic behaviour of this new estimator is the same than that of the ML-estimator in Gomes and Martins (2002a), given by*

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

based on the scaled log-spacings

$$U_i := i \ \left[\ln X_{n-i+1:n} - \ln X_{n-i:n}\right]. \tag{2.4}$$

Despite of that, their finite sample behaviour is quite different, as we shall see in section 4.

2.2 "Asymptotically unbiased" linear combination of Hill's estimators for a general ρ

The equivalent of Proposition 2.3 is:

Proposition 2.4. *For a general ρ, and whenever we consider the k levels $m = 1, 2, \cdots, k$, we have again $\mathbf{P}'\Sigma^{-1}\mathbf{P} = \begin{bmatrix} k & k \\ k & P_{kk} \end{bmatrix}$, but where*

$$k^{-2\rho} \ P_{kk} = k^{2(1-\rho)} - (k(k-1))^{1-\rho} - \sum_{i=1}^{k-1} i^{1-\rho} \left((i-1)^{1-\rho} - 2 \ i^{1-\rho} + (i+1)^{1-\rho}\right).$$

As well as in Proposition 2.2, we have $\Delta = \left|\left|\mathbf{P}'\Sigma^{-1}\mathbf{P}\right|\right| = k\left(P_{kk} - k\right)$, and $\underline{\mathbf{b}}'\Sigma^{-1}\underline{\mathbf{b}} = P_{kk}$.

The weights $a_i^H = a_i^H(\rho)$, $i = 1, 2, \cdots, k$, in

$$BL_H^{(\rho)}(k) = \sum_{i=1}^{k} a_i^H(\rho) \ H(i),$$

are given by

$$a_i^H(\rho) = \frac{k^{2\rho}}{k(P_{kk}-k)} \left(-i(i-1)S_{i-1} + 2i^2 S_i - i(i+1)S_{i+1}\right), \quad 1 \le i \le k-1, \tag{2.5}$$

$$a_k^H(\rho) = \frac{k^{2\rho}}{k(P_{kk}-k)} \left(-k(k-1)S_{k-1} + k^2 S_k\right), \tag{2.6}$$

where

$$S_i = \sum_{j=1}^{k-1} j \left(j^{-\rho} - i^{-\rho}\right) \left(2 \ j^{1-\rho} - (j-1)^{1-\rho} - (j+1)^{1-\rho}\right) + k\left(k^{-\rho} - i^{-\rho}\right)\left(k^{1-\rho} - (k-1)^{1-\rho}\right), \quad 1 \le i \le k.$$

Moreover,

$$\lim_{k\to\infty} \frac{k\ \mathbf{b}'\Sigma^{-1}\mathbf{b}}{\Delta} = \lim_{k\to\infty} \frac{P_{kk}}{P_{kk}-k} = \left(\frac{1-\rho}{\rho}\right)^2.$$

We may thus state the following general result.

Theorem 2.2. *If the second order condition* (1.8) *holds and if* $k = k_n$ *is a sequence of intermediate positive integers, i.e.,* (1.1) *holds, then, with*

$$BL_H^{(\rho)}(k) := \sum_{i=1}^{k} a_i^H(\rho)\ H(i),$$

$a_i^H(\rho),\ 1 \le i \le k-1$ *and* $a_k^H(\rho)$ *given in* (2.5) *and* (2.6)*, respectively,*

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

with $P_k^{BL_H^{(\rho)}}$ *asymptotically standard normal.*
Consequently, if $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \lambda$*, finite, non necessarily null, then*

$$\sqrt{k}\left(BL_H^{(\rho)}(k)-\gamma\right) \underset{n\to\infty}{\stackrel{d}{\longrightarrow}} Normal\left(0,\ \frac{\gamma^2(1-\rho)^2}{\rho^2}\right).$$

Moreover, the same distributional result holds true if we consider the tail index estimators $BL_H^{(\widehat{\rho})}$*, for any second order parameter estimator* $\widehat{\rho}$ *such that* $\widehat{\rho}-\rho = o_p(1)$ *for the levels* k *on which we are going to base the estimation of the tail index* γ*.*

Proof. The first part of the theorem comes straighforwardly from the previous results. The last result in the theorem comes essentially from the assumption that $\widehat{\rho}-\rho = o_p(1)$, and from the fact that we have the distributional representation $BL_H^{(\widehat{\rho})}(k) \stackrel{d}{=} BL_H^{(\rho)}(k) + (\widehat{\rho}-\rho)\ \xi_k(\rho)\ (1+o_p(1))$, with $\xi_k = O_p\left(1/\sqrt{k}\right)$. Consequently, we may write $\sqrt{k}\left(BL_H^{(\widehat{\rho})}(k)-\gamma\right) \stackrel{d}{=} \sqrt{k}\left(BL_H^{(\rho)}(k)-\gamma\right)+o_p(1)$, whenever $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \lambda$ finite, non necessarily null. □

Remark 2.3. *For $\widehat{\rho}$ we may choose, such as in Gomes and Martins (2002b), one of the estimators in Fraga Alves et al. (2003). More specifically, we shall here consider particular members of the class of estimators*

$$\widehat{\rho}_\tau := \min\left(0, \frac{3(T_n^{(\tau)}(k_1)-1)}{(T_n^{(\tau)}(k_1)-3)}\right), \quad k_1 = \min\left(n-1, \left[\frac{2n}{\ln\ln n}\right]\right) \quad (2.7)$$

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

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\ln\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^j, \quad j = 1, 2, 3.$$

We have $\widehat{\rho} - \rho = O_p\left(1/\left(\sqrt{k_1}\ A(n/k_1)\right)\right) = o_p(1)$ for any level k such that $\sqrt{k}\ A(n/k) \to \lambda$, finite.

We advise practitioners not to choose blindly the value of τ in (2.7). It is sensible to draw a few sample paths of $\widehat{\rho}_n^{(\tau)}(k)$ in (2.7), as functions of k, electing the value of τ which provides higher stability for large k, by means of any stability criterion. Anyway, in all the Monte Carlo simulations we have considered the level k_1 in (2.7) and the ρ-estimators

$$\widehat{\rho}_0 := \min\left(0, \frac{3\left(T_n^{(0)}(k_1)-1\right)}{T_n^{(0)}(k_1)-3}\right), \quad \textit{advisable for} \quad \rho \geq -1, \qquad (2.8)$$

and

$$\widehat{\rho}_1 := \min\left(0, \frac{3\left(T_n^{(1)}(k_1)-1\right)}{T_n^{(1)}(k_1)-3}\right), \quad \textit{advisable for} \quad \rho < -1. \qquad (2.9)$$

Remark 2.4. *The results in Theorem 2.2 are now equivalent to the ones got in Gomes and Martins (2002b) for the "Maximum Likelihood" estimator,* *based on an external estimation $\widehat{\rho}$ of ρ, and given by*

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

based on the scaled log-spacings U_i in (3.1).

3 Linear combinations of scaled log-spacings

Since the linear combination obtained in section 2 is a bit intrincate, and things become still a bit more intricate if we directly approach linear combinations of log-excesses, we are now going to think on linear combinations of the vector of scaled log-spacings

$$\underline{\mathbf{U}} \equiv \left(U_i = i\left[\ln X_{n-i+1:n} - \ln X_{n-i:n}\right],\ 1 \le i \le k\right). \tag{3.1}$$

For heavy tails, and whenever $k = k_n \to \infty$ and $k = k_n = o(n)$, as $n \to \infty$, the scaled log-spacings U_i, $1 \le i \le k$, are approximately independent and exponential with mean value

$$\mu_i = \gamma + A(n/k)\left(\frac{i}{k}\right)^{-\rho}(1 + o(1)), \tag{3.2}$$

as $n \to \infty$ (Draisma, 2000, pages 43-59), where $A(.)$ and $\rho < 0$ are related to the second order behaviour of F. We may thus consider approximately that $\Sigma = \mathbf{I}$, the identity matrix, and the "best linear unbiased" combinations of the scaled log-spacings are for sure much easier to derive. We my state the following:

Proposition 3.1. *We now have*

$$\mathbf{P}'\Sigma^{-\mathbf{1}}\mathbf{P} = \left[\begin{array}{cc} k & \sum_{j=1}^{k}\left(\frac{j}{k}\right)^{-\rho} \\ \sum_{j=1}^{k}\left(\frac{j}{k}\right)^{-\rho} & \sum_{j=1}^{k}\left(\frac{j}{k}\right)^{-2\rho} \end{array}\right],$$

and consequently.

$$\Delta = \left|\left|\mathbf{P}'\Sigma^{-\mathbf{1}}\mathbf{P}\right|\right| = k\ \sum_{j=1}^{k}\left(\frac{j}{k}\right)^{-2\rho} - \left(\sum_{j=1}^{k}\left(\frac{j}{k}\right)^{-\rho}\right)^{2}.$$

The weights $a_i^U = a_i^U(\rho)$, $i = 1, 2, \cdots, k$, *are given by*

$$a_i^U = \frac{1}{\Delta}\left(\sum_{j=1}^{k}\left(\frac{j}{k}\right)^{-2\rho} - \left(\frac{i}{k}\right)^{-\rho}\sum_{j=1}^{k}\left(\frac{j}{k}\right)^{-\rho}\right), \; 1 \le i \le k. \qquad (3.3)$$

Moreover,

$$\lim_{k\to\infty} \frac{k\ \mathbf{b}'\Sigma^{-1}\mathbf{b}}{\Delta} = \lim_{k\to\infty} \frac{\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\rho}}{\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\rho} - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}\right)^2} = \left(\frac{1-\rho}{\rho}\right)^2.$$

We may thus state the following general result:

Theorem 3.1. *The results of Theorem 2.2 hold true for the r.v.*

$$\begin{aligned} BL_U^{(\rho)}(k) &:= \sum_{i=1}^{k} a_i^U(\rho)\ U_i \\ &= \frac{\left(\frac{1}{k}\sum_{i=1}^{k} U_i\right)\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\rho}\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho} U_i\right)\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}\right)}{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\rho}\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}\right)^2}, \end{aligned}$$

with U_i *and* $a_i^U(\rho)$, $1 \le i \le k$ *given in* (3.1) *and in* (3.3), *respectively, as well as for the tail index estimators* $BL_U^{(\widehat{\rho})}$, *with* $\widehat{\rho}$ *in the conditions of Theorem 2.2.*

Proof. The proof of Theorem 2.2 applies here as well. The result comes also easily from the fact that $\sum_{i=1}^{k} i^{\alpha-1}/k^{\alpha} = 1/\alpha + O(1/k)$, and that, as has been proved in Gomes and Martins (2002a),

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

where $Z_k^{(\alpha)} = \sqrt{(2\alpha-1)k}\left(\sum_{i=1}^{k} i^{\alpha-1}E_i/k^{\alpha} - 1/\alpha\right)$ is asymptotically standard normal. The asymptotic variance comes from the random term

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

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

□

As expected, the exact behaviour of "best linear unbiased" combinations based on the Hill estimators, on the log-excesses or on the scaled log-scacings, do not differ significantly, as it is shown in Figure 1, where we picture, for a Fréchet underlying parent and for a sample size $n = 1000$, the sample paths of $BL_H^{(-1)}$, $BL_U^{(-1)}$ and $BL_V^{(-1)}$, where V denotes the vector of the log-excesses

$$\underline{\mathbf{V}} \equiv \left(V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n},\ 1 \le i \le k\right),$$

and which "almost" overlap. We picture also the behaviour of $BL_H^{(\widehat{\rho}_i)}$, $BL_U^{(\widehat{\rho}_i)}$ and $BL_V^{(\widehat{\rho}_i)}$, $i = 0, 1$, also "overlapping". The sample paths of $ML^{(-1)}$ and $ML^{(\widehat{\rho}_i)}$, $i = 0, 1$ are also pictured, being possible to detect a significantly different behaviour between the BL and the ML statistics.

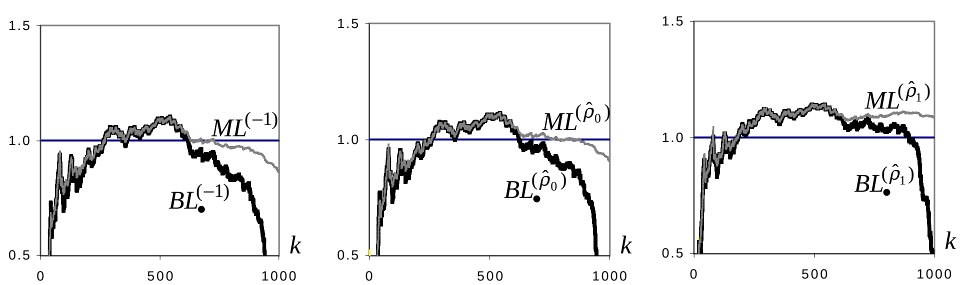

Figure 1: Simulated sample paths of the BL and ML-estimators for a Fréchet(1) parent.

In Figure 2 we picture the differences $BL_U^{(\bullet)}(k) - BL_H^{(\bullet)}(k)$ in black and $BL_U^{(\bullet)}(k) - BL_V^{(\bullet)}(k)$ in grey, also when we consider ρ mispecified at -1 (*left*), estimated through $\widehat{\rho}_0$ (*center*) and estimated through $\widehat{\rho}_1$ (*right*). Only for very small values of k, as well as, but not so significantly, for large values of k, do appear significative differences, which have no special influence in the final properties of the estimators. This justifies the use, in practice, and also in the simulations performed, of $BL_U^{(\widehat{\rho})}$, instead of either $BL_H^{(\widehat{\rho})}$ or $BL_V^{(\widehat{\rho})}$.

4 The finite sample behaviour of the estimators

To enhance the fact that despite their asymptotic equivalence $BL^{(-1)}$ and $ML^{(-1)}$, here denoted BL and ML, respectively, have a reasonably different behaviour as the underlying model changes, we present, in Figure 3, the

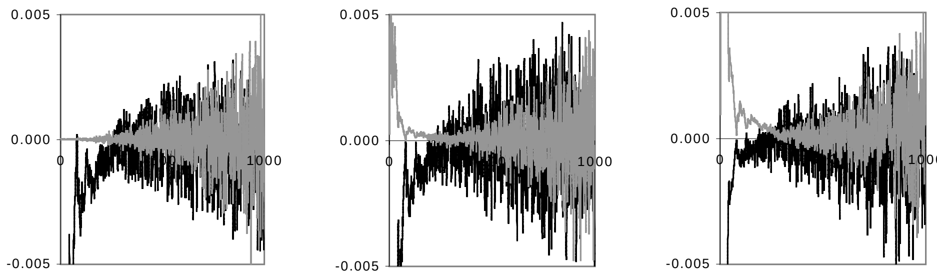

Figure 2: Differences $BL_U^{(\bullet)} - BL_H^{(\bullet)}$, in black, and $BL_U^{(\bullet)} - BL_V^{(\bullet)}$, in grey, for a Fréchet(1) parent and for $\rho = -1$ (*left*), $\widehat{\rho}_0$ (*center*) and $\widehat{\rho}_1$ (*right*).

relative efficiencies of $ML|BL$ at their optimal levels. Such a measure is given by

$$REFF_{ML|BL} = \sqrt{\frac{MSE\,[BL_0]}{MSE\,[ML_0]}},$$

with both estimators considered at their optimal levels, i.e., the levels where the mean squared error is minimum: for any estimator G of the tail index γ, we denote $G_0 = G(k_0^G(n))$, with $k_0^G(n) := \arg\min_k MSE\,[G(k)]$. Notice that high relative efficiencies correspond to better performances of the ML-estimator relatively to the BL-estimator, and the other way round for low relative efficiencies. Notice also that in Figure 3 (*right*) we have pictured, in a different scale the three central $REFF$ measures presented in Figure 3 (*left*).

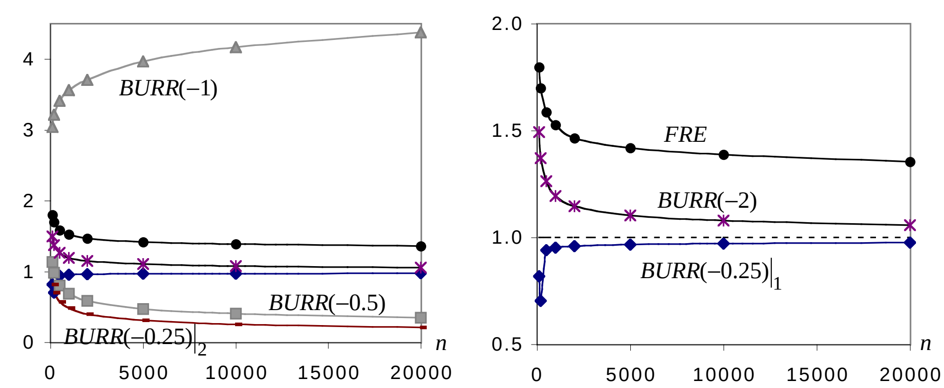

Figure 3: Relative efficiencies of $ML|BL$ for different parents.

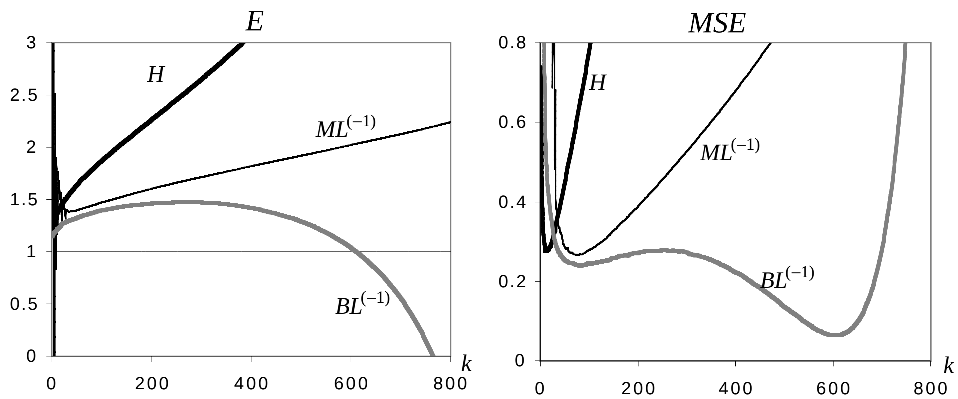

Figure 4: Simulated distributional behaviour of the estimators under study for a Burr(1, −0.25).

Some general comments:

1. As ρ approaches 0, the BL mean squared error exhibits two minima for large n — look at Figure 4, where we picture the mean value and the mean squared error of the BL and ML estimators, together with the Hill estimator, for a Burr(1,−0.25) parent, with d.f. $F_{\gamma,\rho}(x) = 1-(1+x^{-\rho/\gamma})^{1/\rho},\ x \geq 0,\ \gamma > 0,\ \rho < 0$. The global minimum is always achieved at the largest k-value, and the comparison has been done for both minimum values of MSE. The $REFF$ measure associated to the global minimum is pictured in Figure 3 (*left*), with the subscript 2, and the one associated to the first minimum is pictured in the same figure (*left* and *right*), with the subscript 1.

2. For $\rho = -0.5$ the two minima are undistinguishable for the sample sizes considered (see Figure 6, for instance) — that's the reason why we picture the relative efficiency only in Figure 3 (*left*). For $\rho = -1$ the ML estimator reveals no bias, and that's the reason for the high relative efficiency of $ML|BL$, exhibited in Figure 3 (*left*).

3. The results we think sensible to consider are the ones in Figure 3 (*right*), also pictured in the central part of Figure 3 (*left*). And for those parents, the relative efficiency will ultimately achieve the value one, although still a long way from one for the Fréchet model and for a sample size $n = 20000$. All the other $REFF$ measures are related to "peculiarities" of the estimators.

4. The main message seems to be the following: asymptotically equivalent estimators may reveal quite distinct finite sample behaviour, and even if asymptotically equivalent, estimators should be compared for finite sample through Monte Carlo simulation techniques.

In Figures 5, 6 and 7 we present the patterns of mean values (E), mean squared errors (MSE) and of $\sqrt{k\,Var}$ of the Hill (H) and of the $BL \equiv BL^{(-1)}$ and the $ML \equiv ML^{(-1)}$ for the Fréchet(1), the Burr(1,−0.5) and the Burr(1,−2) parents, respectively. The Fréchet(γ) model is a basic heavy tail model, with d.f. $F_\gamma(x) = \exp\left(-x^{-1/\gamma}\right),\ x \geq 0$. The simulation has been based on 1000 runs. In these figures we also picture the mean value and mean squared error of the estimator $BL_U^{(\widehat{\rho}_i)}$, denoted $BL^{\widehat{\rho}_i}$, as well as of $ML^{(\widehat{\rho}_i)}$, for the adequate value of i, either 0 or 1.

We also present in Tables 1, 2 and 3 the simulated optimal sample fractions, bias and mean square errors of the different estimators under play. Denoting generically $G(k)$ any estimator of γ, we shall denote $OSF^G := k_0^G(n)/n$, $k_0^G(n) := \arg\min_k MSE\,[G(k)]$, $E_0^G := E\left[G\left(k_0^G(n)\right)\right]$ and $MSE_0^G := MSE\left[G\left(k_0^G(n)\right)\right]$. The extra subscript $_s$ in the tables denotes "simulated". The Monte Carlo simulation is amulti-sample simulation of size 1000×10 for $n = 100, 200, 500$ and 1000, and of size 1000×5 for n= 2000 and 5000. Standard errors are not presented, but are available from the authors. The smallest bias and mean square errors are double underlined, being only underlined the second smallest.

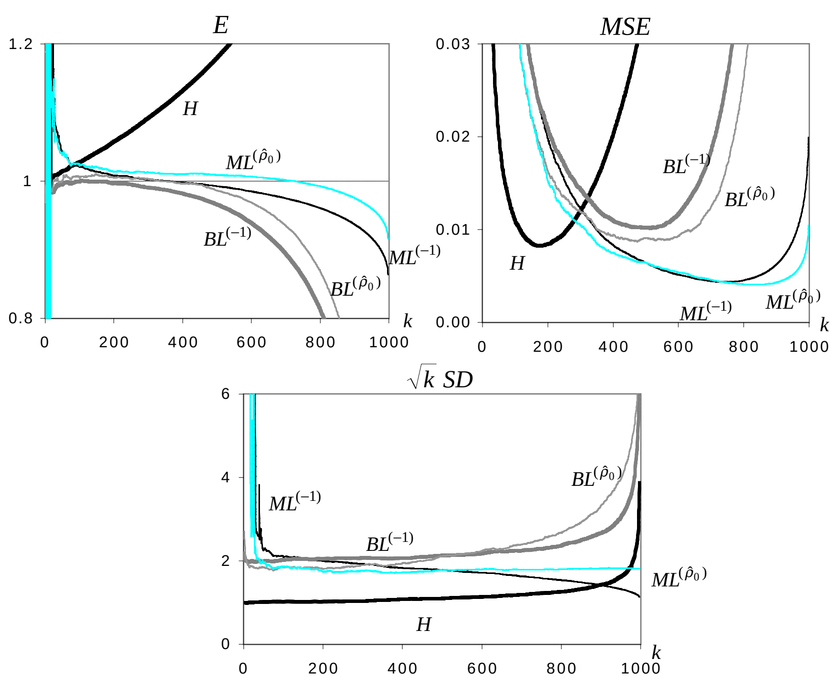

Figure 5: Simulated distributional behaviour of the estimators under study for a Fréchet(1) parent ($\rho = -1$).

We now advance with some extra comments:

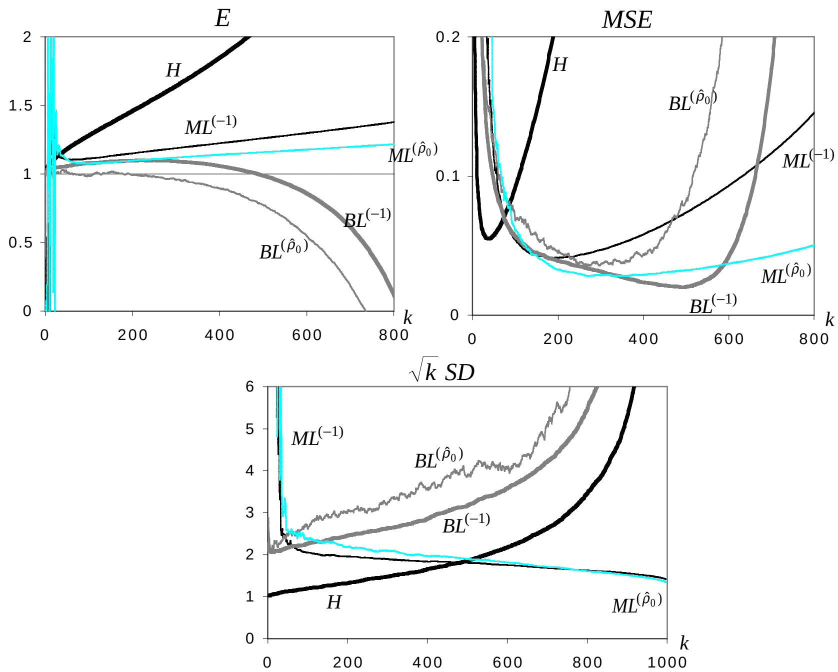

Figure 6: Simulated distributional behaviour of the estimators under study for a Burr$(1, -0.5)$.

Table 1: Simulated Optimal Sample Fractions (OSF) of the tail index estimators

n	H	$BL^{(-1)}$	$ML^{(-1)}$	$BL^{(\widehat{\rho}_0)}$	$ML^{(\widehat{\rho}_0)}$	$BL^{(\widehat{\rho}_1)}$	$ML^{(\widehat{\rho}_1)}$
STUDENT$(\nu = 4)$							
200	0.0405	0.1990	0.1280	0.1180	0.2965	0.1470	0.1450
500	0.0278	0.2024	0.0968	0.1222	0.2096	0.2354	0.0846
1000	0.0195	0.2070	0.0697	0.1181	0.1594	0.3534	0.0577
2000	0.0141	0.2139	0.0419	0.1112	0.1230	0.3697	0.0390
5000	0.0095	0.2126	0.0415	0.1012	0.0853	0.3721	0.0241
STUDENT$(\nu = 2)$							
200	0.0850	0.2115	0.3850	0.1310	0.3855	0.2005	0.2640
500	0.0672	0.1818	0.3536	0.1410	0.4252	0.2468	0.1908
1000	0.0563	0.1678	0.3214	0.1339	0.4439	0.2767	0.1607
2000	0.0458	0.1571	0.2829	0.1143	0.4633	0.3176	0.1139
5000	0.0315	0.1371	0.2465	0.0947	0.4762	0.3194	0.0894
STUDENT$(\nu = 1)$							
200	0.1665	0.2500	0.3780	0.1665	0.3685	0.2445	0.3740
500	0.1448	0.2174	0.3356	0.1950	0.2978	0.2560	0.3752
1000	0.1339	0.1985	0.2816	0.1901	0.2638	0.2175	0.2922
2000	0.1191	0.1815	0.2386	0.1737	0.2199	0.2560	0.3311
5000	0.0971	0.1553	0.1919	0.1496	0.1876	0.2409	0.3055

Table 2: Simulated Bias of the tail index estimators

n	H	$BL^{(-1)}$	$ML^{(-1)}$	$BL^{(\widehat{\rho}_0)}$	$ML^{(\widehat{\rho}_0)}$	$BL^{(\widehat{\rho}_1)}$	$ML^{(\widehat{\rho}_1)}$
STUDENT($\nu = 4$)							
200	0.0572	−0.0541	0.0754	−0.0594	0.0105	−0.0018	0.0760
500	0.0536	−0.0436	0.0658	−0.0434	0.0349	0.0295	0.0780
1000	0.0546	−0.0045	0.0628	−0.0081	0.0349	0.0172	0.0570
2000	0.0856	−0.0022	0.0284	0.0072	0.0540	0.0252	0.0760
5000	0.0259	−0.0072	0.0405	−0.0118	0.0220	0.0027	0.0314
STUDENT($\nu = 2$)							
200	0.1305	0.0498	0.0397	−0.0848	−0.0083	0.0653	0.0986
500	0.0078	−0.0869	−0.0191	−0.1176	−0.0748	−0.0567	−0.0127
1000	0.0155	−0.0545	0.0013	−0.0812	−0.0290	−0.0075	0.0120
2000	0.0154	−0.0180	0.0105	−0.0432	−0.0127	−0.0061	0.0219
5000	0.0111	−0.0252	−0.0006	−0.0413	−0.0100	−0.0047	0.0107
STUDENT($\nu = 1$)							
200	0.0089	−0.2525	−0.1782	−0.2567	−0.2137	−0.2756	−0.0979
500	0.0171	−0.1250	−0.1128	−0.1866	−0.1178	−0.0341	−0.0278
1000	0.0334	−0.0413	−0.0576	−0.0289	−0.0455	−0.0257	0.0074
2000	0.0278	−0.0253	−0.0502	−0.0325	−0.0406	−0.0081	0.0187
5000	0.0206	−0.0402	−0.0419	−0.0413	−0.0443	0.0031	0.0228

Table 3: Simulated mean square errors of the tail index estimators

n	H	$BL^{(-1)}$	$ML^{(-1)}$	$BL^{(\widehat{\rho}_0)}$	$ML^{(\widehat{\rho}_0)}$	$BL^{(\widehat{\rho}_1)}$	$ML^{(\widehat{\rho}_1)}$
STUDENT($\nu = 4$)							
200	0.0206	0.0179	0.0084	0.0516	0.0084	0.0209	0.0196
500	0.0109	0.0071	0.0055	0.0150	0.0049	0.0103	0.0111
1000	0.0073	0.0035	0.0037	0.0070	0.0032	0.0055	0.0073
2000	0.0047	0.0018	0.0022	0.0034	0.0023	0.0031	0.0048
5000	0.0028	0.0007	0.0022	0.0014	0.0014	0.0012	0.0028
STUDENT($\nu = 2$)							
200	0.0238	0.0370	0.0084	0.1650	0.0129	0.0365	0.0177
500	0.0115	0.0156	0.0038	0.0304	0.0039	0.0100	0.0092
1000	0.0069	0.0083	0.0022	0.0154	0.0017	0.0050	0.0057
2000	0.0041	0.0044	0.0013	0.0082	0.0007	0.0025	0.0035
5000	0.0022	0.0020	0.0007	0.0039	0.0003	0.0010	0.0019
STUDENT($\nu = 1$)							
200	0.0366	0.1025	0.0396	1.2807	0.0809	0.2114	0.0378
500	0.0164	0.0437	0.0217	0.0555	0.0293	0.0249	0.0131
1000	0.0093	0.0241	0.0141	0.0274	0.0169	0.0126	0.0080
2000	0.0054	0.0135	0.0091	0.0151	0.0105	0.0057	0.0036
5000	0.0025	0.0062	0.0046	0.0066	0.0050	0.0023	0.0016

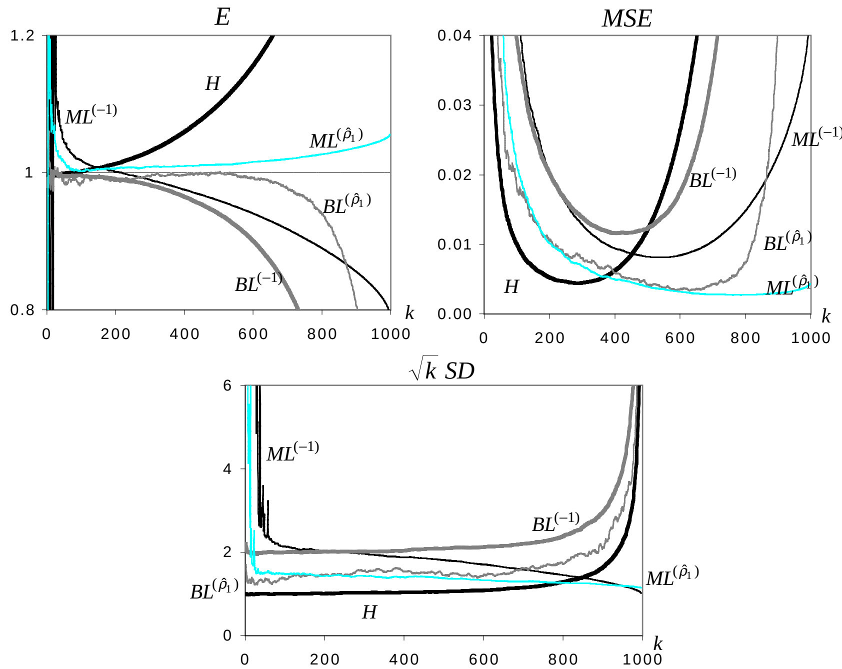

Figure 7: Simulated distributional behaviour of the estimators under study for a Burr$(1, -2)$.

5. For not very large values of n (say $n \leq 1000$) there exists only a slight improvement in terms of smaller minimum mean squared error, when we use $\widehat{\rho}$ instead of a misspecification of ρ at -1, unless ρ is reasonably small (see Figures 5 and 7, for parents with $\rho = -1$ and $\rho = -2$, respectively).

6. As ρ approaches 0 (see Figure 6, for a parent with $\rho = -0.5$), there is a significant difference between the mean value pattern of the estimators $BL^{(-1)}$ and $BL^{(\widehat{\rho}_0)}$. Indeed, the $BL^{(\widehat{\rho}_0)}$ estimator exhibits sample paths highly stable around the target value γ, but with a reasonably high volatility. Such a volatity gives rise to similar mean squared errors, as functions of k, both when we misspecify or estimate ρ.

7. In general, we may say that, whenever $\rho \neq -1$, the replacement of $\rho = -1$ by $\widehat{\rho}$ enables us to achieve sample paths with a reasonable high volatility, but around the target value γ for a wide region of k-values. Indeed, the sample paths of the BL-estimators are much more stable (around the target value γ) than those of the corresponding ML-estimators. However, the trouble with the BL-estimators are related to the fact that the variance of $\sqrt{k}\ (BL(k) - \gamma)$ is, for finite n, an increasing function of k, contrarily to what happens to the corresponding ML-estimators. This gives rise to a general better perfor-

mance of of the ML comparatively to the equivalent BL, when both are considered at their optimal levels.

References

[1] Aitken, A. C. (1935). On least squares and linear combinations of observations. *Proc. Roy. Soc. Edin.* **55**, 42-48.

[2] 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.

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

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

[5] Gomes, M.I. and Martins, M.J. (2002a). *Bias reduction and explicit efficient estimation of the tail index.* Notas e Comunicações C.E.A.U.L. 5/2002. To appear in *J. Statistical Planning and Inference.*

[6] Gomes, M. I. and M. J. Martins (2002b). "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes* **5**:1, 5-31.

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