REVSTAT – Statistical Journal
Volume 4, Number 2, June 2006, 81–109

IMPROVEMENTS IN THE ESTIMATION OF A HEAVY TAIL*

Authors: Orlando Aníbal Oliveira
–

M. Ivette Gomes
– University of Lisbon, FCUL (DEIO) and CEAUL, Portugal
ivette.gomes@fc.ul.pt

M. Isabel Fraga Alves
– University of Lisbon, FCUL (DEIO) and CEAUL, Portugal
isabel.alves@fc.ul.pt

Received: June 2005 Revised: March 2006 Accepted: March 2006

Abstract:

- In this paper, and in a context of regularly varying tails, we suggest new tail index estimators, which provide interesting alternatives to the classical Hill estimator of the tail index γ. They incorporate some extra knowledge on the pattern of scaled top order statistics and seem to work generally pretty well in a semi-parametric context, even for cases where a second order condition does not hold or we are outside Hall's class of models. We shall give particular emphasis to a class of statistics dependent on a *tuning* parameter τ, which is merely a change in the scale of our data, from X to X/τ. Such a statistic is non-invariant both for changes in location and in scale, but compares favourably with the Hill estimator for a class of models where it is not easy to find competitors to this classic tail index estimator. We thus advance with a slight "controversial" argument: it is always possible to take advantage from a non-invariant estimator, playing with particular *tuning* parameters — either a change in the location or in the scale of our data —, improving then the overall performance of the classical estimators of extreme events parameters.

Key-Words:

- *statistics of extremes; semi-parametric estimation; Monte Carlo methods.*

AMS Subject Classification:

- 62G32, 62H12; 65C05.

*Research partially supported by FCT / POCTI and POCI / FEDER. A tribute to Orlando Oliveira, who died prematurely a month after getting his Ph.D. in May 2003.

1. INTRODUCTION AND PRELIMINARIES

Let $X_1, X_2, ..., X_n$ be independent random variables (r.v.'s) with common distribution function (d.f.) F, with a heavy upper tail, i.e., for large x, there exists $\gamma > 0$ such that

$$\overline{F}(x) := 1 - F(x) = x^{-1/\gamma} L_F(x) \ ,$$

where $L_F(x)$ is a slowly varying function, i.e., for every $x > 0$, $L_F(tx)/L_F(t) \to 1$ as $t \to \infty$. F is thus in the max-domain of attraction of an *Extreme Value* (EV) d.f.,

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

with $\gamma > 0$. We shall denote this fact by $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$.

Recall that, for $\gamma > 0$,

$$F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma) \quad \text{iff} \quad \overline{F} \in RV_{-1/\gamma} \quad \text{iff} \quad U \in RV_\gamma \ , \tag{1.1}$$

where $U(t) := F^{\leftarrow}(1-1/t),\ t > 1$ (Gnedenko, 1943; de Haan, 1970). RV_α stands for the class of regularly varying functions at infinity with index of regular variation equal to α, i.e., positive functions g with infinite right endpoint, and such that $\lim_{t\to\infty} g(tx)/g(t) = x^\alpha$, for all $x > 0$, and the notation $F^{\leftarrow}$ is used for the generalized inverse function of F, i.e., $F^{\leftarrow}(t) = \inf\{x \colon F(x) \geq t\}$.

The function $A(t)$ measures the rate of convergence of $\{\ln U(tx) - \ln U(t)\}$ towards $\{\gamma \ln x\}$ in (1.1), and it is a function of constant sign, such that

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho} \ , \tag{1.2}$$

for every $x > 0$, where ρ (≤ 0) is a *second order parameter.* The limit function in (1.2) must be of the stated form, and $|A(t)| \in RV_\rho$ (Geluk and de Haan, 1987).

1.1. The new estimation procedures

Let $X_{i:n}$ denote the i-th ascending order statistic (o.s.), $1 \leq i \leq n$, associated to the sample $\underline{X}_n = (X_1, X_2, ..., X_n)$. Under the validity of the first order framework in (1.1), with $U(t) = t^\gamma\, L_U(t)$, $L_U \in RV_0$, and for *intermediate* k, i.e.,

$$k = k_n \to \infty, \quad k/n \to 0, \qquad \text{as} \quad n \to \infty \ , \tag{1.3}$$

the classic tail index estimator for a positive γ is Hill's estimator (Hill, 1975), with the functional expression

$$\widehat{\gamma}_n^H(k) := \frac{1}{k} \sum_{i=1}^{k} \Big[\ln X_{n-i+1:n} - \ln X_{n-k:n} \Big] . \tag{1.4}$$

For this estimator, and whenever (1.3) holds, we have the validity of the distributional representation,

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

with P_k asymptotically standard normal (de Haan and Peng, 1998).

Also, under the validity of (1.3), it is possible to scale $X_{n-k:n}$ (or $X_{n-k+1:n}$), with $a_n = U(n)$, so that

$$\ln \frac{X_{n-k+1:n}}{a_n} + \gamma\, \psi(k) \underset{n\to\infty}{\overset{p}{\longrightarrow}} 0 . \tag{1.5}$$

And for every fixed i, $1 \le i < n$, there exists a non-degenerate r.v. ϵ_i, such that $\mathbb{E}[\epsilon_i] = 0$, and

$$\ln \frac{X_{n-i+1:n}}{a_n} + \gamma\, \psi(i) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \epsilon_i . \tag{1.6}$$

As usual, ψ denotes the digamma function, i.e. $\psi(t) = d \ln\Gamma(t)/dt = \Gamma'(t)/\Gamma(t)$, being Γ the complete Gamma function, $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\, dx$, $t > 0$. For a justification of these results see Lemma 4.1. For details on the Γ and ψ functions, see Abramowitz and Stegun (1975??).

Let us then think on the least-squares' type estimators of γ and $b := \ln a$, which come from the minimization, jointly in γ and b, of

$$\sum_{i=1}^{k} \Big\{ \ln X_{n-i+1:n} - b + \gamma\, \psi(i) \Big\}^2 .$$

Straightforward computations lead us to

$$\widetilde{b}_n(k) = \widetilde{\ln a}(k) = \frac{1}{k} \sum_{i=1}^{k} \ln X_{n-i+1:n} + \widetilde{\gamma}_n(k) \left(\frac{1}{k} \sum_{i=1}^{k} \psi(i) \right) , \tag{1.7}$$

with

$$\widetilde{\gamma}_n(k) = \frac{\left(\sum\limits_{i=1}^{k} \psi(i)\right) \left(\sum\limits_{i=1}^{k} \ln X_{n-i+1:n}\right) - k \sum\limits_{i=1}^{k} \psi(i) \ln X_{n-i+1:n}}{k \sum\limits_{i=1}^{k} \psi^2(i) - \left(\sum\limits_{i=1}^{k} \psi(i)\right)^2} . \tag{1.8}$$

Remark 1.1. Notice that the replacement of $\psi(i)$ by $\{\ln i\}$ in the γ-estimator in (1.8) leads us to the estimator, based on a QQ-plot, studied in Kratz and Resnick (1996) and independently in Schultze and Steinbach (1996), and given by

$$(1.9) \qquad \widetilde{\gamma}_n^{(\mathcal{K})}(k) := \frac{\left(\sum_{i=1}^{k} \ln i\right)\left(\sum_{i=1}^{k} \ln X_{n-i+1:n}\right) - k \sum_{i=1}^{k} (\ln i) \ln X_{n-i+1:n}}{k \sum_{i=1}^{k} \ln^2 i - \left(\sum_{i=1}^{k} \ln i\right)^2} .$$

Since $\psi(x) = \ln x + O(1/x)$, as $x \to \infty$, the difference between the estimators $\widetilde{\gamma}_n$ and $\widetilde{\gamma}_n^{(\mathcal{K})}$ is asymptotically negligible. However, for finite samples, their performance differs significantly, because the approximation in terms of the digamma function $\psi(i)$ is usually better than the use of $\{\ln i\}$ for all i between 1 and k.

We may easily simplify the expressions of $\widetilde{b}_n(k)$ and of $\widetilde{\gamma}_n(k)$ in (1.7) and (1.8), respectively, through the use of the following relations involving the digamma function,

$$(1.10) \qquad \sum_{j=1}^{k} \psi(j) \;=\; k\,\psi(k) - (k-1) \;=\; k\big(\psi(k+1) - 1\big) ,$$

$$\sum_{j=1}^{k} \psi^2(j) \;=\; k\,\psi^2(k+1) + 2k - (2k+1)\,\psi(k+1) + \psi(1)$$

and

$$k \sum_{j=1}^{j} \psi^2(j) - \left(\sum_{j=1}^{k} \psi(j)\right)^2 \;=\; k\Big\{k - \psi(k+1) + \psi(1)\Big\} \;=\; k \sum_{j=1}^{k} \left(1 - \frac{1}{j}\right) .$$

We then get the following linear combination of the top log-observations,

$$(1.11) \qquad \widetilde{\gamma}_n(k) \;=\; \frac{\sum_{i=1}^{k} \Big(\psi(k+1) - \psi(i) - 1\Big) \ln X_{n-i+1:n}}{k - \psi(k+1) + \psi(1)} ,$$

and we may also write

$$(1.12) \qquad \widetilde{a}_n(k) \;=\; X_{n-k:n} \exp\Big(\widehat{\gamma}_n^H(k) + \widetilde{\gamma}_n(k)\,\big(\psi(k+1) - 1\big)\Big) ,$$

where $\widehat{\gamma}_n^H(k)$ and $\widetilde{\gamma}_n(k)$ are given in (1.4) and (1.11), respectively.

We shall next assume that we are in Hall's class of models (Hall and Welsh, 1985), where

$$(1.13) \qquad U(t) \;=\; C\, t^{\gamma}\left(1 + \frac{A(t)}{\rho}\,\big(1 + o(1)\big)\right), \qquad A(t) = \gamma\,\beta\, t^{\rho}, \quad \text{as } t \to \infty ,$$

or equivalently that the tail function is of the type

$$1 - F(x) \;=\; \left(\frac{x}{C}\right)^{-1/\gamma} \left\{1 + \frac{\beta}{\rho}\left(\frac{x}{C}\right)^{\rho/\gamma} + o\big(x^{\rho/\gamma}\big)\right\}, \qquad \text{as} \quad x \to \infty ,$$

where $\gamma > 0$, $C > 0$, $\rho < 0$ and $\beta \neq 0$.

We may then choose $a = a_n = C\, n^\gamma$, as $n \to \infty$, and, from (1.12), we get a least-squares' estimator of C given by

$$\widetilde{C}_n(k) \;:=\; X_{n-k:n} \exp\left\{\widehat{\gamma}_n^H(k) - \widetilde{\gamma}_n(k)\Big(\ln n - \psi(k{+}1) + 1\Big)\right\} \tag{1.14}$$

$$\sim\; X_{n-k:n}\left(\frac{k}{n}\right)^{\widetilde{\gamma}_n(k)} \exp\left\{\widehat{\gamma}_n^H(k) - \widetilde{\gamma}_n(k)\right\}, \qquad \text{as} \quad k \to \infty , \tag{1.15}$$

again with $\widehat{\gamma}_n^H(k)$ and $\widetilde{\gamma}_n(k)$ given in (1.4) and (1.11), respectively.

Although aware that C is a parameter of the model, which may be estimated for instance through any of the asymptotically equivalent estimators in (1.14) or (1.15), we shall consider $\tau \equiv C$ as a *tuning* parameter. This has been done in a way similar to the one used by Csörgő and Viharos (1998), when they consider a kernel estimator as a function of a tuning parameter $\tau \equiv \rho$, also a model parameter, the second order parameter in (1.2). Notice that if $U_X(t) = C\, t^\gamma (1 + o(1))$, then for $Y = X/C$, $U_Y(t) = t^\gamma\, (1 + o(1))$. This means that a proper scaling of our data enables us to choose $a = n^\gamma$, i.e., $\gamma = \ln a / \ln n$, a particular situation which will merely help us to build a class of statistics, dependent of the control parameter $\tau = C$, which should be regarded as a possible change in the scale of our data. Such a class is got from the least-squares type estimator of $\{\ln a\}$ in (1.7), and is given by

$$\begin{aligned} \widetilde{\widetilde{\gamma}}{}_n^{(\tau)}(k) \;&:=\; \frac{1}{k \ln n}\left\{\sum_{i=1}^{k} \ln \frac{X_{n-i+1:n}}{\tau} + \widetilde{\gamma}_n(k) \sum_{i=1}^{k} \psi(i)\right\} \\ &=\; \frac{1}{\ln n}\left\{\ln \frac{X_{n-k:n}}{\tau} + \big(\psi(k{+}1) - 1\big)\widetilde{\gamma}_n(k) + \widehat{\gamma}_n^H(k)\right\}. \end{aligned} \tag{1.16}$$

As a particular member of the class in (1.16), we shall consider the estimator

$$\widetilde{\widetilde{\gamma}}_n(k) \;\equiv\; \widetilde{\widetilde{\gamma}}{}_n^{(1)}(k) \;=\; \frac{\ln X_{n-k:n} + \big(\psi(k{+}1) - 1\big)\widetilde{\gamma}_n(k) + \widehat{\gamma}_n^H(k)}{\ln n} . \tag{1.17}$$

We shall also consider the estimation of C, and its use in the class of statistics in (1.16), but we are aware that then we are going to get a poorer estimator of the tail index γ, unless the C-estimator is highly efficient. For instance, should we have used $\widetilde{C}_n(k)$, in (1.14), as τ, in (1.16), would we have been led to $\widetilde{\gamma}_n$ in (1.11), i.e., $\widetilde{\widetilde{\gamma}}{}_n^{(\widetilde{C}_n(k))}(k) \equiv \widetilde{\gamma}_n(k)$. We have here decided to follow Hall and Welsh (1985), and to consider the C-estimator

$$\widehat{C}_n(k) := \left(\frac{k}{n}\right)^{\widehat{\gamma}_n^H(k)} X_{n-k:n} . \tag{1.18}$$

Since in Hall's class of models, in (1.13), the mean squared error of both non-degenerate limiting distributions of $\widehat{\gamma}_n^H(k)$ and $\widehat{C}_n(k)$ are minimized by taking

$$k_0 = \left(\frac{(1-\rho)^2}{-2\,\rho\,\beta^2}\; n^{-2\rho}\right)^{1/(1-2\rho)}$$

(Theorem 4.1 in Hall and Welsh, 1985), we shall also consider, in the simulations, and whenever we are in Hall's class of models in (1.13), the estimator of the tail index γ, given by

$$(1.19) \qquad \widetilde{\widetilde{\gamma}}_n^{(\widehat{C})}(k), \qquad \text{where} \quad \widehat{C} = \widehat{C}_n(\widehat{k}_0), \quad \widehat{k}_0 = \left(\frac{(1-\widehat{\rho})^2}{-2\,\widehat{\rho}\,\widehat{\beta}^2}\; n^{-2\widehat{\rho}}\right)^{1/(1-2\widehat{\rho})},$$

with $\widehat{C}_n$ given in (1.18) and $\widehat{\rho}$ and $\widehat{\beta}$ adequate estimators of ρ and β, respectively, already considered in Gomes and Martins (2002). In the simulations of models outside Hall's class, due to the difficulties in the estimation of k_0, we shall exhibit the behaviour of

$$(1.20) \qquad \widetilde{\widetilde{\gamma}}_n^{(\widehat{C}_0)}, \qquad \text{where} \quad \widehat{C}_0 = \widehat{C}_n(k_0), \quad k_0 = \arg\min_k MSE\Big[\widehat{\gamma}_n^H(k)\Big]\;,$$

again with $\widehat{C}_n$ given in (1.18) and k_0 obtained through simulation.

Remark 1.2. In practice, it is sensible to consider τ in (1.16) as a tuning parameter, choosing τ through a data-driven estimation of the mean squared error of $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ as a function of k, for adequately chosen fixed values of τ (Oliveira, 2002). The value of τ may be any value τ^* such that

$$\widehat{MSE}\Big[\widetilde{\widetilde{\gamma}}_n^{(\tau^*)}(k)\Big] \le \widehat{MSE}\Big[\widehat{\gamma}_n^H(k)\Big], \quad \text{for every } k\;.$$

When we consider

$$k_{n0}^{(\tau^*)} := \arg\min_k \widehat{MSE}\Big[\widetilde{\widetilde{\gamma}}_n^{(\tau^*)}(k)\Big]\;,$$

it is then sensible to choose the value τ_0^* providing the minimum $\widehat{MSE}\Big[\widetilde{\widetilde{\gamma}}_n^{(\tau^*)}\big(k_{n0}^{(\tau^*)}\big)\Big]$, i.e.,

$$\tau_0^* := \arg\min_{\tau^*} \widehat{MSE}\Big[\widetilde{\widetilde{\gamma}}_n^{(\tau^*)}\Big(k_{n0}^{(\tau^*)}\Big)\Big]\;.$$

We choose then (see also Remark 5.1)

$$\widehat{k}_{n0} = k_{n0}^{\tau_0^*} \qquad \text{and} \qquad \widetilde{\widetilde{\gamma}}_{n0} := \widetilde{\widetilde{\gamma}}_n^{(\tau_0^*)}\big(\widehat{k}_{n0}\big)\;.$$

This is an open problem, beyond the scope of the present paper, where we intend essentially to present the potentialities of the class of statistics in (1.16) to estimate a positive tail index.

In section 2, we shall briefly review the Peaks Over Threshold (POT) methodology, a classical method of estimation of a tail index, to be also compared with the new estimation procedures considered, as well as the estimation of the second order parameters ρ and β in $A(t) = \gamma\,\beta\,t^\rho$. In section 3 we shall compare asymptotically the estimator in (1.11) (or equivalently, the estimator in (1.9)) with the Hill estimator in (1.4). Section 4 is devoted to the asymptotic behaviour of the class of estimators in (1.16). Sections 5 and 6 are devoted to the illustration of the behaviour of these estimators for finite samples, through the use of Monte Carlo simulation techniques.

2. REVIEW OF WELL-ESTABLISHED ESTIMATION PROCEDURES OF FIRST AND SECOND ORDER PARAMETERS

2.1. The link between the Hill estimator and the POT methodology

Let us think on the excesses over a high random threshold $X_{n-k:n}$,

$$V_{ik} := X_{n-i+1:n} - X_{n-k:n}\ , \qquad 1 \le i \le k\ .$$

Since $X \stackrel{d}{=} U(Y)$, Y a standard unit Pareto r.v. with d.f. $1-1/y$, $y \ge 1$, $Y_{n-i+1:n}/Y_{n-k:n} \stackrel{d}{=} Y_{k-i+1:k}$, $1 \le i \le k$, and, for k intermediate, $Y_{n-k:n} = (n/k)(1+o_p(1))$, we may write, under the validity of the first order condition in (1.1),

$$\begin{aligned}
V_{ik} &= X_{n-i+1:n} - X_{n-k:n} \;=\; X_{n-k:n}\big(X_{n-i+1:n}/X_{n-k:n} - 1\big)\\
&\stackrel{d}{=} X_{n-k:n}\Big(U\big(Y_{n-k:n}\,Y_{k-i+1:k}\big)/U\big(Y_{n-k:n}\big) - 1\Big)\\
&\stackrel{d}{=} X_{n-k:n}\Big(Y^\gamma_{k-i+1:k}\big(1+o_p(1)\big) - 1\Big)\\
&= X_{n-k:n}\Big(\big(Y^\gamma_{k-i+1:k} - 1\big)\,\big(1+o_p(1)\big) + o_p(1)\Big)\ .
\end{aligned}$$

Consequently, we may say that there exists δ such that we have approximately $V_{ik}/\delta \approx \big(Y^\gamma_{k-i+1:k} - 1\big)/\gamma$, i.e., V_{ik}, $1 \le i \le k$, are approximately the k o.s. of a sample of size k from a Generalized Pareto (GP) model,

$$GP_\gamma(x;\delta) \;=\; 1 - \big(1 + \gamma\,x/\delta\big)^{-1/\gamma}\ , \qquad x \ge 0\ \ (\gamma,\delta > 0)\ .$$

The estimation of γ through maximum likelihood (ML) in a GP model has been thoroughly studied in Davison (1984) and Smith (1984a,b). Davison (1984) suggested a re-parameterization of the GP model in $(\gamma,\alpha) = (\gamma,\gamma/\delta)$, which enables us to get only one ML equation to be solved iteratively. Such a re-parameterization has also been used in Gomes and Oliveira (2003a), where a computational study of this methodology has been undertaken. The ML-estimator of γ has, with such a re-parameterization, an explicit expression as a function of the *ML*-estimator $\widehat{\alpha}$ of $\alpha = \gamma/\delta$ and the sample of the excesses. We have

$$\widehat{\gamma}_n^{GP}(k) \;:=\; \frac{1}{k}\sum_{i=1}^{k} \ln\bigl(1 + \widehat{\alpha}\, V_{ik}\bigr) \;, \tag{2.1}$$

and α is such that $\alpha\, V_{ik} \approx Y_{k-i+1:k}^{\gamma} - 1$. Notice that an obvious choice for α is $1/X_{n-k:n}$. Then $1 + \alpha\, V_{ik} = X_{n-i+1:n}/X_{n-k:n}$, and the estimator in (2.1) is the Hill estimator $\widehat{\gamma}_n^H(k)$ in (1.4). Smith (1987) has got the asymptotic behaviour of the estimator in (2.1) for a fixed threshold u. The conclusion of his Theorem 3.2 may be easily rephrased in this set-up (Gomes, 2002; Drees *et al.*, 2004), and, under the second order framework in (1.2), we get the asymptotic distributional representation

$$\widehat{\gamma}_n^{GP}(k) \;\stackrel{d}{=}\; \gamma + \frac{(1+\gamma)}{\sqrt{k}}\, Q_k + \frac{(1+\gamma)\,(\gamma+\rho)\, A(n/k)}{\gamma\,(1-\rho)\,(1-\rho+\gamma)}\,\bigl(1 + o_p(1)\bigr) \;, \tag{2.2}$$

with Q_k asymptotically standard normal.

Remark 2.1. Note that the result in (2.2), although appearing to produce a different bias term, agrees with the one in Drees *et al.* (2004). Indeed, whereas we here assume (1.2), the most common second order condition for heavy-tailed models, Drees *et al.* (2004) consider the general case $\gamma \in \mathbb{R}$, and assume that there exists $a^*(\cdot)$ and $A^*(\cdot)$ such that

$$\frac{\frac{U(tx)-U(t)}{a^*(t)} - \frac{x^\gamma - 1}{\gamma}}{A^*(t)} \;\underset{t\to\infty}{\longrightarrow}\; \frac{1}{\rho^*}\left(\frac{x^{\gamma+\rho^*}-1}{\gamma+\rho^*} - \frac{x^\gamma - 1}{\gamma}\right).$$

If we consider $\rho^* < 0$, we may then guarantee that, with $A_0(t) = A^*(t)/\rho^*$ and $a_0(t) = a^*(t)\,(1 - A^*(t)/\rho^*)$, we get,

$$\frac{\frac{U(tx)-U(t)}{a_0(t)} - \frac{x^\gamma - 1}{\gamma}}{A_0(t)} \;\underset{t\to\infty}{\longrightarrow}\; \frac{x^{\gamma+\rho^*}-1}{\gamma+\rho^*} \;. \tag{2.3}$$

For $\gamma > 0$ (and $\rho^* < 0$), condition (2.3) is equivalent to saying that, as $t \to \infty$,

$$U(t) \;=\; C\, t^\gamma \Bigl(1 + A\, t^{\rho*} + o\bigl(t^{\rho*}\bigr)\Bigr) \;. \tag{2.4}$$

Then

$$U(tx) - U(t) \;=\; C\,\gamma\, t^\gamma \left(\frac{x^\gamma - 1}{\gamma} + \frac{A(\gamma+\rho^*)\, t^{\rho^*}}{\gamma}\left(\frac{x^{\gamma+\rho^*}-1}{\gamma+\rho^*}\right) + o\bigl(t^{\rho*}\bigr)\right).$$

If $\gamma+\rho^* \neq 0$, we then need to choose $a_0(t) = C\,\gamma\, t^\gamma$, $A_0(t) = A(\gamma+\rho^*)\, t^{\rho^*}/\gamma$. Then

$$\begin{aligned}
\frac{U(tx)}{U(t)} \;&=\; 1 + \frac{U(tx)-U(t)}{C\, t^\gamma}\Bigl(1 - A\, t^{\rho^*} + o\bigl(t^{\rho^*}\bigr)\Bigr) \\
&=\; x^\gamma \Bigl(1 + A\, t^{\rho^*}\bigl(x^{\rho^*} - 1\bigr) + o\bigl(t^{\rho^*}\bigr)\Bigr) \\
&=\; x^\gamma \left(1 + A\,\rho^*\, t^{\rho^*}\left(\frac{x^{\rho^*}-1}{\rho^*}\right) + o\bigl(t^{\rho^*}\bigr)\right) \;,
\end{aligned}$$

and consequently,

$$\ln U(tx) - \ln U(t) \;=\; \gamma \,\ln x + A\,\rho^*\, t^{\rho^*}\biggl(\frac{x^{\rho^*}-1}{\rho^*}\biggr) + o\bigl(t^{\rho^*}\bigr)\;,$$

i.e., provided that $\gamma+\rho^* \neq 0$, and with $A(t) = \frac{\gamma\,\rho^* A_0(t)}{\gamma+\rho^*}$,

$$\text{(2.5)}\qquad \frac{\ln U(tx) - \ln U(t) - \gamma\,\ln x}{A(t)} \;\underset{t\to\infty}{\longrightarrow}\; \frac{x^{\rho^*}-1}{\rho^*} = \frac{x^{\rho}-1}{\rho}\;,$$

i.e., ρ^* in (2.3) is equal to ρ in (1.2). Consequently, if $\sqrt{k}\,A(n/k) \to \lambda$, $\sqrt{k}\,A_0(n/k) \to \lambda(\gamma{+}\rho)/(\gamma\,\rho)$. The bias provided in Drees *et al.* (2004) for the *POT-ML* tail index estimator is then

$$\frac{\lambda(\gamma+\rho)}{\gamma\,\rho}\left(\frac{\rho(1+\gamma)}{(1-\rho)\,(1-\rho+\gamma)}\right) \;=\; \frac{\lambda(1+\gamma)\,(\gamma+\rho)}{\gamma(1-\rho)\,(1-\rho+\gamma)}\;,$$

the values provided in both Smith (1987) and Gomes (2002).

We shall now make explicit the term $o(t^{\rho^*})$ in (2.4), assuming that

$$U(t) \;=\; C\,t^{\gamma}\Bigl(1 + A\,t^{\rho*} + B\,t^{\rho*+\rho'} + o\bigl(t^{\rho*+\rho'}\bigr)\Bigr)\;, \qquad \rho' < 0\;.$$

If $\gamma+\rho^* = 0$, i.e., $\rho^* = -\gamma$,

$$U(tx) - U(t) \;=\; C\,\gamma\,t^{\gamma}\left(\frac{x^{\gamma}-1}{\gamma} + \frac{2B\,\rho^*\,t^{\rho^*+\rho'}}{\gamma}\biggl(\frac{x^{\rho^*+\rho'}-1}{\rho^*+\rho'}\biggr) + o\bigl(t^{\rho^*+\rho'}\bigr)\right)\;,$$

and

$$a_0(t) = C\,\gamma\,t^{\gamma}\;, \qquad A_0(t) = \frac{2B\,\rho^*\,t^{\rho^*+\rho'}}{\gamma}\;.$$

But for the model in (2.4), we may choose for any $\rho<0$, $A(t) = \rho\,A\,t^{\rho}$, and we get

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

If $\sqrt{k}\,A(n/k) \to \lambda$, $\sqrt{k}\,A_0(n/k) \to 0$. So, both from Smith (1987) and from Drees *et al.* (2004), we get a null dominant component for the bias term of the *POT-ML* tail index estimator, whenever $\gamma+\rho^* = 0$, as expected.

2.2. Estimators of the second order parameters ρ and β

The estimation of the second order parameter ρ, in $A(t) = \gamma\,\beta\,t^\rho$, is going to be done through particular members of the class of ρ-estimators in Fraga Alves *et al.* (2003). Those estimators are given by

$$\widehat{\rho}_n^{(i)}(k) := \min\left(0\,,\ \frac{3\big(T_n^{(i)}(k)-1\big)}{T_n^{(i)}(k)-3}\right)\,, \qquad i=0,1\,, \tag{2.6}$$

where

$$T_n^{(i)}(k) := \begin{cases} \dfrac{M_n^{(1)}(k) - \big(M_n^{(2)}(k)/2\big)^{1/2}}{\big(M_n^{(2)}(k)/2\big)^{1/2} - \big(M_n^{(3)}(k)/6\big)^{1/3}} & \text{if} \quad i=1 \\[2ex] \dfrac{\ln\big(M_n^{(1)}(k)\big) - \frac12 \ln\big(M_n^{(2)}(k)/2\big)}{\frac12 \ln\big(M_n^{(2)}(k)/2\big) - \frac13 \ln\big(M_n^{(3)}(k)/6\big)} & \text{if} \quad i=0 \end{cases}\ .$$

The statistics is (2.6) are consistent for the estimation of ρ whenever the second order condition (1.2) holds and k is such that $k \to \infty$, $k = o(n)$ and $\sqrt{k}\,A(n/k) \to \infty$, as $n \to \infty$.

Remark 2.2. The theoretical and simulated results in Fraga Alves *et al.* (2003), together with the use of these estimators in the Generalized Jackknife statistics of Gomes *et al.* (2000), as done in Gomes and Martins (2002), has led these authors to advise the consideration of the level

$$k_1 = \min\Big(n-1,\ \big[2n/\ln\ln n\big]\Big) \tag{2.7}$$

and of the ρ-estimators

$$\widehat{\rho}_0 := \min\Big(0,\ 3\big(T_n^{(0)}(k_1)-1\big)\big/\big(T_n^{(0)}(k_1)-3\big)\Big) \qquad \text{if} \quad \rho \ge -1\,, \tag{2.8}$$

and

$$\widehat{\rho}_1 := \min\Big(0,\ 3\big(T_n^{(1)}(k_1)-1\big)\big/\big(T_n^{(1)}(k_1)-3\big)\Big) \qquad \text{if} \quad \rho < -1\,. \tag{2.9}$$

For the estimation of β we have here considered the estimator of β in Gomes and Martins (2002) and based on the scaled log-spacings $U_i = i\,\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}$, $1 \le i \le k$. Let us denote $\widehat{\rho}$ any of the estimators either in (2.8) or in (2.9) (or even in (2.6)). The β-estimator is given by

$$\widehat{\beta}(k) := \frac{1}{n^{\widehat{\rho}}}\, \frac{\left(\sum\limits_{i=1}^k i^{-\widehat{\rho}}\right)\left(\sum\limits_{i=1}^k U_i\right) - k\left(\sum\limits_{i=1}^k i^{-\widehat{\rho}}\,U_i\right)}{\left(\sum\limits_{i=1}^k i^{-\widehat{\rho}}\right)\left(\sum\limits_{i=1}^k i^{-\widehat{\rho}}\,U_i\right) - k\left(\sum\limits_{i=1}^k i^{-2\widehat{\rho}}\,U_i\right)}\ . \tag{2.10}$$

We have then considered $\widehat{\beta} = \widehat{\beta}(k_1)$, k_1 given in (2.7).

3. ASYMPTOTIC PROPERTIES OF $\widetilde{\gamma}_n$

3.1. The estimator $\widetilde{\gamma}_n$ as a linear combination of Hill's estimators

We first state the following:

Lemma 3.1. *A semi-parametric estimator of the tail index γ which is a linear combination of the k top log-observations, i.e.,*

$$\gamma_n(k) \;=\; \sum_{i=1}^{k} a_i \,\ln X_{n-i+1:n} \tag{3.1}$$

is scale invariant if and only if $\sum\limits_{i=1}^{k} a_i = 0$.

Proof: If we consider a change in scale, moving from X to X/C, $C>0$, $C\neq 1$, the estimator in (3.1) changes to $\sum_{i=1}^k a_i \ln X_{n-i+1:n} - \ln C \sum_{i=1}^k a_i$, which equals $\gamma_n(k) = \sum_{i=1}^k a_i \ln X_{n-i+1:n}$ if and only if $\sum_{i=1}^k a_i = 0$. □

Lemma 3.2. *A semi-parametric estimator of the type (3.1) may be expressed as a linear combination of Hill's estimators, i.e.,*

$$\gamma_n(k) \;=\; \sum_{i=1}^{k} a_i \,\ln X_{n-i+1:n} \;=\; \sum_{j=1}^{k-1} b_j \;\widehat{\gamma}_n^H(j)\;, \tag{3.2}$$

where

$$b_j \;=\; -a_{j+1} - \frac{1}{j+1} \sum_{i=j+2}^{k} a_i, \quad j=1,...,k-2\;, \qquad b_{k-1} = -a_k\;, \tag{3.3}$$

if and only if it is scale invariant, i.e., if and only if $\sum_{i=1}^k a_i = 0$.

Proof: We may write

$$\begin{aligned}
\sum_{j=1}^{k-1} b_j\;\widehat{\gamma}_n^H(j) \;&=\; \sum_{j=1}^{k-1} b_j \left\{\frac{1}{j}\sum_{i=1}^{j} \ln X_{n-i+1:n} - \ln X_{n-j:n}\right\}\\
&=\; \sum_{i=1}^{k-1}\left(\sum_{j=i}^{k-1}\frac{b_j}{j}\right)\ln X_{n-i+1:n} \;-\; \sum_{i=1}^{k} b_{i-1}\,\ln X_{n-i+1:n} \quad (b_0\equiv 0)\\
&=\; \sum_{i=1}^{k-1}\left(\sum_{j=i}^{k-1}\frac{b_j}{j} - b_{i-1}\right)\ln X_{n-i+1:n} -\; b_{k-1}\,\ln X_{n-k+1:n}\;,
\end{aligned}$$

i.e., $a_i = \sum_{j=i}^{k-1} b_j/j - b_{i-1}$, $1 \le i \le k-1$, and $a_k = -b_{k-1}$. This linear system has a unique and possible solution if and only if $\sum_{i=1}^{k} a_i = 0$. Then we just need to solve the linear system of equations:

$$\begin{bmatrix} a_2 \\ a_3 \\ \cdots \\ a_k \end{bmatrix} = \begin{bmatrix} -1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{k-1} \\ 0 & -1 & \frac{1}{3} & \cdots & \frac{1}{k-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_{k-1} \end{bmatrix} =: \mathbf{A}\ \mathbf{b}\ .$$

Since the inverse matrix of $\mathbf{A}$ is

$$\mathbf{A}^{-1} = \begin{bmatrix} -1 & -\frac{1}{2} & -\frac{1}{2} & \cdots & -\frac{1}{2} \\ 0 & -1 & -\frac{1}{3} & \cdots & -\frac{1}{3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix},$$

the result follows. □

Then, from the relation (1.10) and from Lemma 3.2, it follows straightforwardly that:

Proposition 3.1. *The estimator in* (1.11), *which may be written as*

$$\widetilde{\gamma}_n(k) = \sum_{i=1}^{k} a_i \ln X_{n-i+1:n}\ , \qquad a_i = \frac{\psi(k+1) - \psi(i) - 1}{k - \psi(k+1) + \psi(1)},\quad 1 \le i \le k\ , \tag{3.4}$$

is scale invariant, i.e. $\sum_{i=1}^{k} a_i = 0$, *and we may write it as the following linear combination of Hill's estimator,*

$$\widetilde{\gamma}_n(k) = \sum_{j=1}^{k-1} b_j\ \widehat{\gamma}_n^H(j)\ , \qquad b_j = \frac{j}{(j+1)\Big(k - \psi(k+1) + \psi(1)\Big)}\ . \tag{3.5}$$

3.2. The asymptotic behaviour of $\widetilde{\gamma}_n(k)$

Theorem 3.1. *Under the first order framework* (1.1) *and for k such that* (1.3) *holds, the estimator in* (1.11) *is a consistent estimator of γ. Moreover, under the second order framework in* (1.2), *we have the validity of the following distributional representation,*

$$\widetilde{\gamma}_n(k) \ \stackrel{d}{=}\ \gamma + \frac{\gamma\sqrt{2}}{\sqrt{k}}\, P_k + \frac{1}{(1-\rho)^2}\ A(n/k)\,\big(1 + o_p(1)\big)\ , \tag{3.6}$$

where P_k is asymptotically standard normal.

Proof: Since in the linear combination in (3.5), $\sum_{j=1}^{k-1} b_j = 1$, $\widetilde{\gamma}_n(k)$ is, under the conditions of the theorem, a consistent estimator of γ. The linear combination of Hill's estimators, $\sum_{j=1}^{k-1} b_j\, \widehat{\gamma}_n^H(j)$ may be written as

$$\begin{aligned}\sum_{j=1}^{k-1} b_j\, \widehat{\gamma}_n^H(j) \;&=\; \sum_{j=1}^{k-1} \frac{b_j}{j} \sum_{i=1}^{j} i\Big[\ln X_{n-i+1:n} - \ln X_{n-i:n}\Big] \\ &=\; \sum_{i=1}^{k-1} i \left(\sum_{j=i}^{k-1} \frac{b_j}{j}\right) \Big[\ln X_{n-i+1:n} - \ln X_{n-i:n}\Big] ,\end{aligned}$$

and consequently, with $\{E_i\}_{i\geq 1}$ i.i.d. standard exponential r.v.'s, we may write

$$\begin{aligned}\text{(3.7)}\qquad \sum_{j=1}^{k-1} b_j\, \widehat{\gamma}_n^H(j) \;&\stackrel{d}{=}\; \gamma \sum_{i=1}^{k-1}\left(\sum_{j=i}^{k-1} \frac{b_j}{j}\right) E_i \\ &\quad + A(n/k)\, k^{\rho} \sum_{i=1}^{k-1} i^{1-\rho} \left(\sum_{j=i}^{k-1} \frac{b_j}{j}\right) \frac{e^{\rho E_i/i} - 1}{\rho} \big(1 + o_p(1)\big) .\end{aligned}$$

For the particular linear combination under study we have

$$\sum_{i=1}^{k-1}\left(\sum_{j=i}^{k-1} \frac{b_j}{j}\right) = \sum_{j=1}^{k-1} b_j = 1 , \qquad \sum_{j=i}^{k-1} \frac{b_j}{j} = \frac{\psi(k+1) - \psi(i+1)}{k - \psi(k+1) + \psi(1)} ,$$

and

$$\begin{aligned}\sum_{i=1}^{k-1}\left(\sum_{j=i}^{k-1} \frac{b_j}{j}\right)^2 \;&=\; \frac{2k - \psi^2(k+1) + \big(2\psi(1) - 1\big)\, \psi(k+1) + \psi(1) - \psi^2(1)}{\Big(k - \psi(k+1) + \psi(1)\Big)^2} \\ &=\; \frac{2}{k} \big(1 + o(1)\big) .\end{aligned}$$

Since $\mathbb{E}\big\{\big(e^{\rho E_i/i} - 1\big)/\rho\big\} = 1/(i-\rho)$, and $\sum_{i=1}^{k-1} i^{-\rho}\big\{\psi(k+1) - \psi(i+1)\big\} = O(k^{-\rho+1})/(1-\rho)^2$, we finally get (3.6). □

Remark 3.1. The result in Theorem 3.1 has already been obtained for the estimator in (1.9) by Csörgő and Viharos (1997), who have shown that for intermediate sequences k, and with $\mu_n(k) = -\frac{n}{k} \int_0^{k/n} \big(1 + \ln(ns/k)\big) \ln U(1/s)\, ds$,

$$\sqrt{k}\; \Big\{\widetilde{\gamma}_n^{(\mathcal{K})}(k) - \mu_n(k)\Big\} \;\underset{n\to\infty}{\overset{d}{\longrightarrow}}\; \mathit{Normal}\,(0, 2\gamma^2) .$$

But under the second order framework in (1.2), $\mu_n(k)$ may be written as

$$\mu_n(k) \;=\; \gamma + \frac{A(n/k)}{(1-\rho)^2} \big(1 + o(1)\big) ,$$

which agrees with the result in (3.6).

Remark 3.2. Notice that, relatively to the Hill estimator, the asymptotic variance of $\widetilde{\gamma}_n(k)$ duplicates, but the bias decreases by a factor $1/(1-\rho)$.

3.3. Asymptotic comparison at optimal levels

Now we proceed to an asymptotic comparison of the estimators $\widetilde{\gamma}_n$, $\widehat{\gamma}_n^H$ and $\widehat{\gamma}_n^{GP}$ at their optimal levels in the lines of de Haan and Peng (1998), Gomes *et al.* (2000, 2002) for sets of Generalized Jackknife statistics, Gomes and Martins (2001) and also Caeiro and Gomes (2002), for specifically built "asymptotically unbiased" estimators of the tail index. Suppose $\gamma_n(k)$ is a general semi-parametric estimator of the tail index, for which the distributional representation

$$\gamma_n(k) \;=\; \gamma + \frac{\sigma}{\sqrt{k}}\, Z_k + b\, A(n/k) + o_p\big(A(n/k)\big) \tag{3.8}$$

holds for any intermediate k, and where Z_k is an asymptotically standard normal r.v.; then we have

$$\sqrt{k}\big[\gamma_n(k) - \gamma\big] \overset{d}{\to} N\big(\lambda b,\, \sigma^2\big), \qquad \text{as} \;\; n \to \infty \;,$$

provided k is such that $\sqrt{k}\, A(n/k) \to \lambda$, finite, as $n \to \infty$. In this situation we write $\mathit{Bias}_\infty[\gamma_n(k)] := b\, A(n/k)$ and $\mathit{Var}_\infty[\gamma_n(k)] := \sigma^2/k$. The so-called Asymptotic Mean Squared Error (AMSE) is then given by

$$\mathit{AMSE}\big[\gamma_n(k)\big] \;:=\; \frac{\sigma^2}{k} + b^2 A^2(n/k) \;.$$

Using regular variation theory it may be proved that, whenever $b \neq 0$, there exists a function $\varphi(n)$, dependent only on the underlying model, and not on the estimator, such that

$$\lim_{n\to\infty} \varphi(n)\, \mathit{AMSE}[\gamma_{n0}] \;=\; \frac{2\rho - 1}{2\rho}\, \big(\sigma^2\big)^{-\frac{2\rho}{1-2\rho}}\, \big(b^2\big)^{\frac{1}{1-2\rho}} \;:=\; \mathit{LMSE}[\gamma_{n0}] \;,$$

where $\gamma_{n0} := \gamma_n(k_0(n))$, $k_0(n) := \arg\min_k \mathit{AMSE}[\gamma_n(k)]$.

It is then sensible to consider the following measure of efficiency, defined in a way that the larger such a measure is the better is the estimator.

Definition 3.1. Given two biased estimators $\gamma_n^{(1)}(k)$ and $\gamma_n^{(2)}(k)$, both computed at their optimal levels, and for which distributional representations of the type (3.8) hold, with constants (σ_1, b_1) and (σ_2, b_2), respectively, $b_1, b_2 \neq 0$, the Asymptotic Root Efficiency (AREFF) of $\gamma_{n0}^{(2)}$ relatively to $\gamma_{n0}^{(1)}$ is

$$\begin{aligned} \mathit{AREFF}_{2|1} \;\equiv\; \mathit{AREFF}_{\gamma_{n0}^{(2)}|\gamma_{n0}^{(1)}} \;&:=\; \sqrt{\mathit{LMSE}\big[\gamma_{n0}^{(1)}\big] \,/\, \mathit{LMSE}\big[\gamma_{n0}^{(2)}\big]} \\ &=\; \left(\left(\frac{\sigma_1}{\sigma_2}\right) \left|\frac{b_1}{b_2}\right|\right)^{\frac{1}{1-2\rho}} . \end{aligned}$$

The comparison of the estimator $\widetilde{\gamma}_n$ with the Hill estimator $\widehat{\gamma}_n^H$, both computed at their optimal levels, leads us to the following result:

Proposition 3.2. *The asymptotic root efficiency of* $\widetilde{\gamma}_n$ *relatively to the Hill estimator* $\widehat{\gamma}_n^H$*, both computed at their optimal levels, is given by*

$$AREFF_{\widetilde{\gamma}_n|\widehat{\gamma}_n^H} = \left(2^\rho(1-\rho)\right)^{\frac{1}{1-2\rho}} , \tag{3.9}$$

being thus greater than 1 iff $\rho > -1$*, and equal to 1 at* $\rho = 0$ *and* $\rho = -1$.

The comparison of the three estimators $\widetilde{\gamma}_n$, $\widehat{\gamma}_n^H$ and $\widehat{\gamma}_n^{GP}$ is done graphically in Figure 1, where the "best" estimator, in terms of minimum *LMSE* at the optimal level, is exhibited. As expected, all depends on the region (γ, ρ), but for values of ρ close to 0, say $\rho > -1$, a region where Hill's estimator exhibits "disturbing" sample paths, the new estimator $\widetilde{\gamma}_n$, at its optimal level, not only overpasses the Hill estimator for all γ, as stated in Proposition 3.2, but also overpasses the *GP*-estimator, at their respective optimal levels, for a wide region of (γ, ρ)-values.

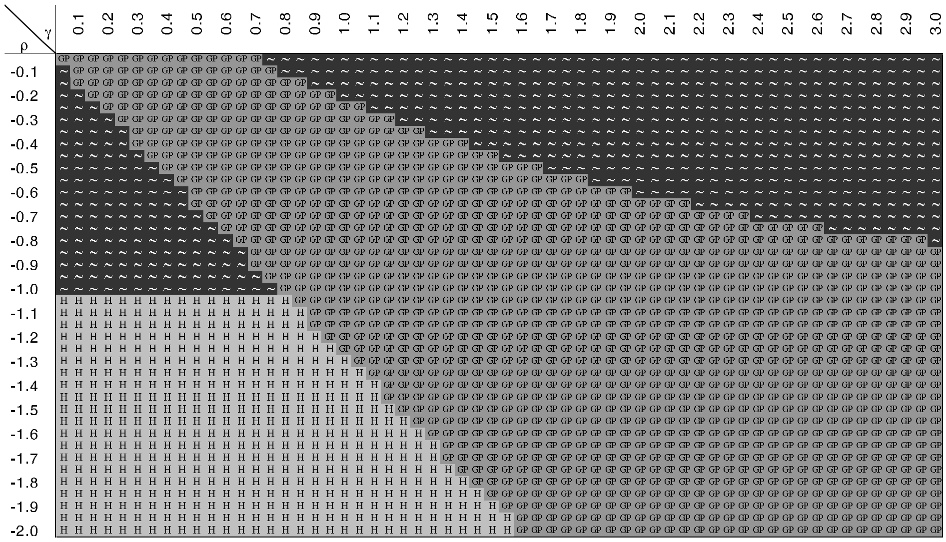

Figure 1: Mimum *LMSE* among the estimators $\widehat{\gamma}^H$, $\widetilde{\gamma}$ and $\widehat{\gamma}^{GP}$ in (1.4), (1.11) and (2.1), respectively.

4. THE ASYMPTOTIC BEHAVIOUR OF $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$

Notice first of all that we no longer have linear combinations of the top log-observations, unless $\tau = 1$, and then:

Proposition 4.1. *If we consider* $\tau = 1$ *in* (1.16)*, the statistic* $\widetilde{\widetilde{\gamma}}(k) \equiv \widetilde{\widetilde{\gamma}}_n^{(1)}(k)$*, in* (1.17)*, may be written as*

$$\widetilde{\widetilde{\gamma}}_n(k) = \sum_{i=1}^{k} a_i^* \ln X_{n-i+1:n} , \tag{4.1}$$

where

$$(4.2) \qquad a_i^* = \frac{1}{k\,\ln n}\left\{1+\frac{k\big(\psi(k+1)-1\big)\,\big(\psi(k+1)-\psi(i)-1\big)}{k-\psi(k+1)+\psi(1)}\right\} .$$

The statistic $\widetilde{\widetilde{\gamma}}_n(k)$ *is only asymptotically scale invariant, and consequently cannot be expressed as a linear combination of Hill's estimators.*

Proof: To get the coefficients of the linear combination in (4.1) we just need to use again Lemma 3.2. Since $\sum_{i=1}^k a_i^* = \frac{1}{\ln n} \neq 0$, but converging towards 0, as $n\to\infty$, $\widetilde{\widetilde{\gamma}}_n(k)$ is not scale invariant, but it is asymptotically scale invariant. □

The asymptotic behaviour of $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ in (1.16) is not directly related to that of the Hill estimator. Indeed the dominant term of $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ is $\{\ln X_{n-k:n}\}$, and we shall base the proof of the asymptotic behaviour of this estimator on the following:

Lemma 4.1. *If* $i \geq 1$ *is fixed, and under the first order condition* (1.1),

$$(4.3) \qquad \ln\frac{X_{n-i+1:n}}{U(n)} \underset{n\to\infty}{\overset{\mathrm{d}}{\longrightarrow}} \gamma\, W_i\ ,$$

where W_i *is a non-degenerate r.v. with a probability density function* (*p.d.f.*) $g_i(w) = \Lambda(w)\,(-\ln\Lambda(w))^i/\Gamma(i)$, $\Lambda(w) = e^{-e^{-w}}$, $w\in\mathbb{R}$. *For* k *intermediate, and under the validity of the second order condition* (1.2), *the distributional representation*

$$(4.4) \qquad \ln\frac{X_{n-k:n}}{U(n/k)} = \frac{\gamma}{\sqrt{k}}\,B_k + o_p\big(A(n/k)\big)$$

holds, with B_k *an asymptotically standard normal r.v.*

Proof: The result in (4.3) is well-known from the field of Extreme Value Theory (see, for instance, Galambos, 1987). Indeed, since $Y_{n-i+1:n}/n$ converges towards a non-degenerate r.v. $Z_i = \exp(W_i)$, and

$$\ln\frac{X_{n-i+1:n}}{U(n)} = \ln\frac{U\big(n(Y_{n-i+1:n}/n)\big)}{U(n)} = \gamma\,\ln Z_i + o_p(1)\ ,$$

(4.3) follows.
For k intermediate (Ferreira *et al.*, 2003),

$$\frac{X_{n-k:n}}{U(n/k)} = \frac{U(Y_{n-k:n})}{U(n/k)} = 1+\frac{\gamma}{\sqrt{k}}\,B_k + o_p\big(A(n/k)\big)\ ,$$

with B_k asymptotically standard normal r.v., and consequently (4.4) holds true. □

Remark 4.1. Notice that $W_i \stackrel{d}{=} -\ln Gama(i)$, where $Gama(i)$ denotes a gamma r.v., with p.d.f. $f(w) = w^{i-1}\exp(-w)/\Gamma(i)$, $w \geq 0$. Consequently $\mathbb{E}(W_i) = -\psi(i)$, and hence (1.6). The relation (1.5) is also a direct consequence of (4.4), together with the fact that $\psi(k) = \ln k + O(1/k)$, as $k \to \infty$.

We thus have, for every $\tau > 0$, consistency of $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ for the estimation of the tail index γ, but we cannot guarantee asymptotic normality. We may however state the following:

Theorem 4.1. *In Hall's class of models, where* (1.13) *holds, and both for fixed and intermediate* k, $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ *is consistent for the estimation of* γ, *for every* $\tau > 0$. *For intermediate* k *we have*

$$\ln n \left\{\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k) - \gamma\right\} \underset{n\to\infty}{\overset{p}{\longrightarrow}} \ln\bigl\{C/\tau\bigr\} , \tag{4.5}$$

i.e., $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ *exhibits a degenerate behaviour. For models where* $C=1$ (*or if we scale our data, dividing them by the appropriate scale* $C\neq 1$, *so that we have a unit scale*), *we get*

$$\frac{\sqrt{k}\,\ln n}{\ln k}\Bigl(\widetilde{\widetilde{\gamma}}_n^{(1)}(k) - \gamma\Bigr) \stackrel{d}{=} \gamma\sqrt{2}\,P_k + \frac{\sqrt{k}\,A(n/k)}{(1-\rho)^2}\bigl(1 + o_p(1)\bigr) , \tag{4.6}$$

i.e., $\widetilde{\widetilde{\gamma}}_n^{(1)}(k)$ *is asymptotically normal, at a rate of convergence of the order of* $\ln k/(\sqrt{k}\,\ln n)$, *with an asymptotic variance equal to* $2\gamma^2$ *and an asymptotic bias equal to* $\lambda/(1-\rho)^2$, *whenever* $\sqrt{k}\,A(n/k) \underset{n\to\infty}{\longrightarrow} \lambda$, *finite.*

Proof: The expression of $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ in (1.16) enables us to get, for fixed k,

$$\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k) \stackrel{d}{=} \gamma + \frac{\ln C - \ln\tau + \gamma\bigl(W_k + \widetilde{H}_k + H_k\bigr) + o_p(1)}{\ln n} ,$$

with W_k, $\widetilde{H}_k$ and H_k non-degenerate r.v.'s. Hence, consistency follows.
For intermediate k,

$$\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k) \stackrel{d}{=} \gamma + \frac{\ln C - \ln\tau}{\ln n} + \gamma\sqrt{2}\,\frac{\ln k}{\sqrt{k}\,\ln n}\,P_k + \frac{\ln k\, A(n/k)}{(1-\rho)^2\,\ln n}\bigl(1 + o_p(1)\bigr) .$$

Consequently (4.5) follows and, for $C = 1$, (4.6) follows, as well as the remaining of the theorem. □

Remark 4.2. Note again that the value $C = 1$ may be achieved through a change in the scale of our data. Indeed, as said from the beginning, if for the original r.v. X we have a quantile function $U_X(t) = C\,t^\gamma(1 + o(1))$, for $Y = X/C$, $U_Y(t) = U_X(t)/C = t^\gamma(1 + o(1))$, and (4.6) holds.

Remark 4.3. Note also that the rate of convergence in (4.6) is of the order of $\ln k/(\sqrt{k}\,\ln n)$, which is a $o(1/\sqrt{k})$, for k intermediate and such that $\ln k = o(\ln n)$. The rate of convergence $1/\sqrt{k}$ is the usual rate of convergence for the most common tail index estimators. The rate of convergence here is also the usual one, whenever $k = O(n^\epsilon)$.

5. PATTERNS OF MEAN VALUES AND MEAN SQUARE ERRORS OF THE ESTIMATORS

From Figure 2 till Figure 7, and with the obvious notation H, GP, $\widetilde{H}$ and $\widetilde{\widetilde{H}}$ instead of $\widehat{\gamma}_n^H$, $\widehat{\gamma}_n^{GP}$, $\widetilde{\gamma}_n$ and $\widetilde{\widetilde{\gamma}}_n$, respectively, we present, in the top, the simulated mean values and MSE's of $\widetilde{\gamma}_n^{(\mathcal{K})}$, $\widetilde{\gamma}_n$ and $\widehat{\gamma}_n^{GP}$ in (1.9), (1.11) and (2.1), respectively. In the bottom part of each figure we picture the same characteristics of $\widetilde{\widetilde{\gamma}}_n^{(1)}|C=0.5$, $\widetilde{\widetilde{\gamma}}_n^{(C)} \equiv \widetilde{\widetilde{\gamma}}_n^{(1)}|C=1$ and $\widetilde{\widetilde{\gamma}}_n^{(1)}|C=2$, as well as of $\widetilde{\widetilde{\gamma}}_n^{(\widehat{C})}$, in (1.19), for models in Hall's class and $\widetilde{\widetilde{\gamma}}_n^{(\widehat{C}_0)}$, in (1.20), for models outside Hall's class. We place in all Figures the same characteristics of the Hill estimator $\widehat{\gamma}_n^H$, for an easier comparison. Simulations related to these Figures are based on 2000 runs, due to the computational time associated to the Peaks Over Threshold methodology. For some of the models, and due the erratic behaviour of the GP estimator for small values of k, we picture its mean value only for $k \geq 100$. The sample size is $n = 1000$ and we have considered the following set of models:

1. *the Fréchet model*, $F(x;C) = \exp\big(-(x/C)^{-1/\gamma}\big)$, $x \geq 0$, with $\gamma = 1$ and $C = 0.5$, 1 and 2, for which $\rho = -1$ (Figure 2);

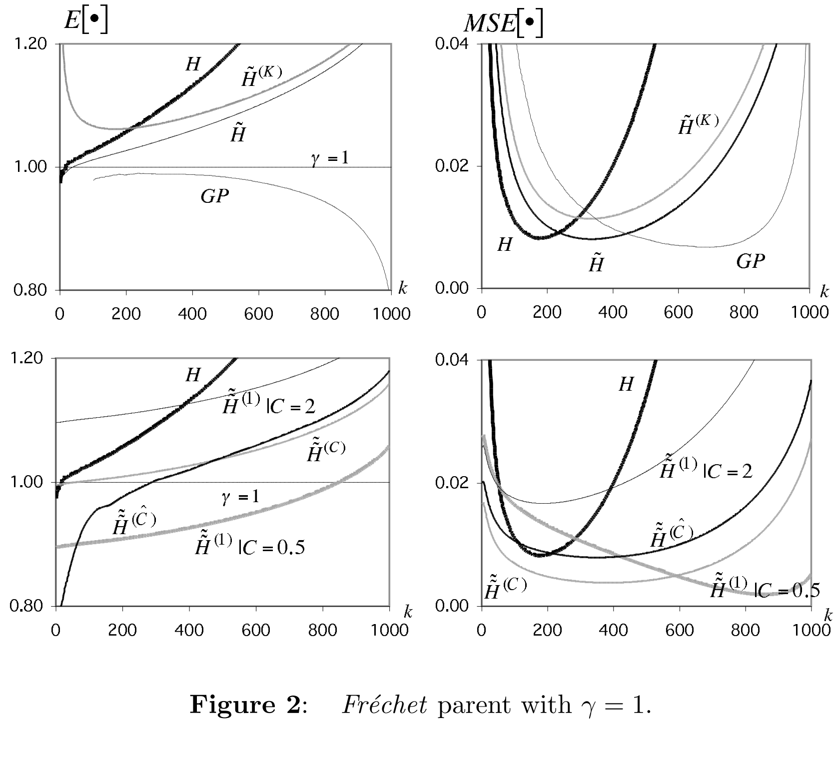

Figure 2: *Fréchet* parent with $\gamma = 1$.

When we look at Figure 2, we immediately notice that the expected changes have obviously occurred, despite the asymptotic scale invariance of $\widetilde{\widetilde{\gamma}}_n(k)$. Indeed, the changes in C induce a "shift" in the sample paths of our estimator. Note that for a scale C, we should get a dominant term of bias given by $\ln C/\ln n = -0.10, 0$ and $+0.10$ for $C = 0.5$, 1 and 2, respectively, which really agrees with the simulated mean values' patterns presented in Figure 2 (*bottom, left*).

Remark 5.1. Looking at the mean squared error patterns, presented also at the Figure 2 (*bottom, right*), we think that we may play with the *tuning* parameter τ in our benefit, in the lines of the work developed in Gomes and Oliveira (2003*b*), where, the use of a control parameter $\{a\}$, which is merely a shift, artificially imposed to the data, and the choice of the adequate value of $\{a\}$ improves greatly the performance of our original estimator. The criterion used there for the choice of $\{a\}$ is a stability criterion of sample paths. Here the methodology must be different, and further research is under development, but we have already the adequate methodologies to deal with this problem. As said before, in Remark 1.2, we think that the best way to proceed (Oliveira, 2002) is to estimate the mean squared errors of our estimators as functions of k, merely on the basis of the available sample, proceeding next to the adaptive choice of the k and τ-values providing the minimum mean squared error: we already have access to suitable procedures of estimation of *MSE*(k), either through the regression diagnostic methodology of Beirlant *et al.* (1996*a*, 1996*b*) or through the use of the bootstrap methodology in Draisma *et al.* (1999), Danielsson *et al.* (2001) and Gomes and Oliveira (2001). Such a computer intensive study is however beyond the scope of this paper.

2. *the Burr model*, $F(x;C) = 1 - \left(1+(x/C)^{-\rho/\gamma}\right)^{1/\rho}$, $x \ge 0$, $\gamma > 0$, $\rho < 0$, with $\gamma = 1$, $C = 0.5$, 1 and 2 and for $\rho = -0.5$, -1 and -2 (Figures 3, 4 and 5);

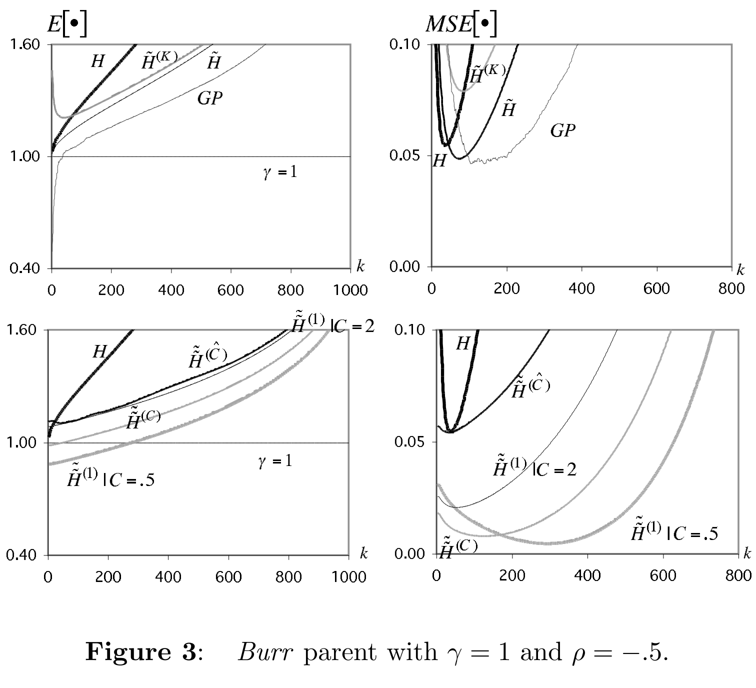

Figure 3: *Burr* parent with $\gamma = 1$ and $\rho = -.5$.

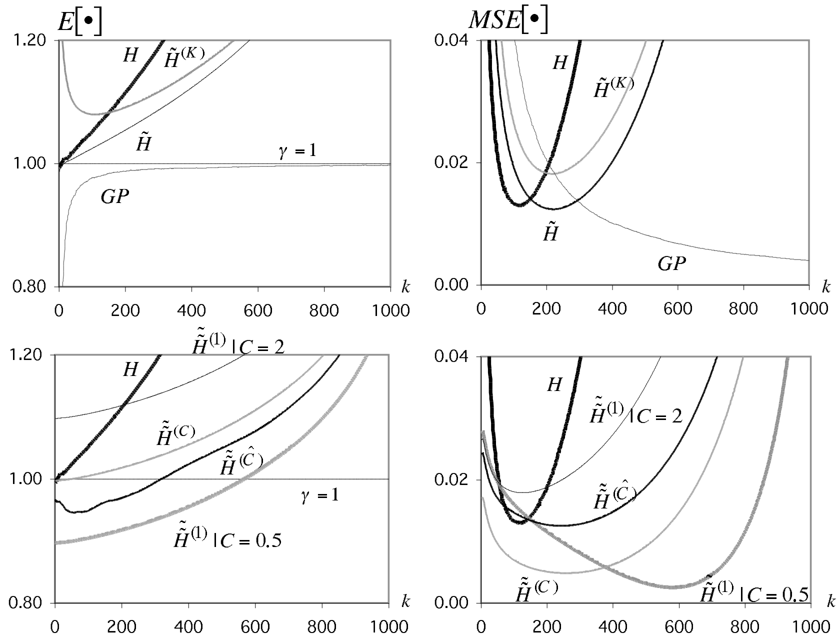

Figure 4: *Burr* parent with $\gamma = 1$ and $\rho = -1$.

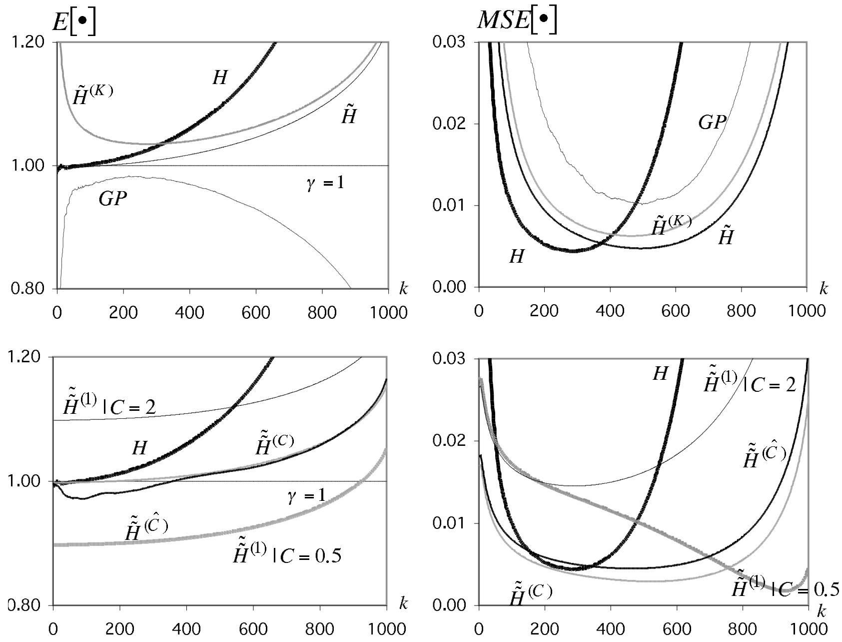

Figure 5: *Burr* parent with $\gamma = 1$ and $\rho = -2$.

a model outside Hall's class,

3. *the Out-Hall model*, with a quantile function $F^{\leftarrow}(1-t) = C\, t^{-1}\, e^{-2t(\ln t - 1)}$, for all $0 < t \leq 1$, $C = 0.5$, 1 and 2 (Figure 6);

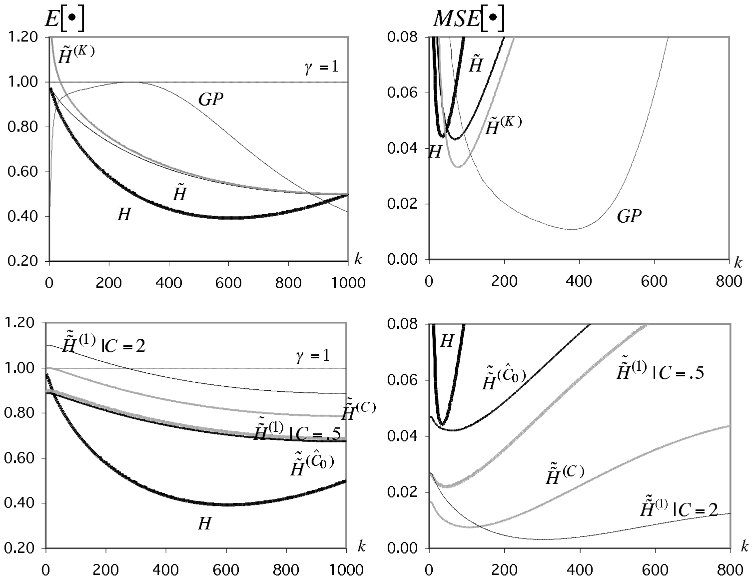

Figure 6: *Out-Hall* parent with $\gamma = 1$.

and the following model for which the second order condition in (1.2) does not hold:

4. *the sin-Burr model*, with a quantile function given by $F^{\leftarrow}(1-t) = C\big(t^\rho - \sin(t^\rho)\big)^{-\gamma/\rho}$, $0 < t \le 1$, with $\gamma = 1$, $C = 0.5$, 1 and 2 and for $\rho = -0.5$ (Figure 7).

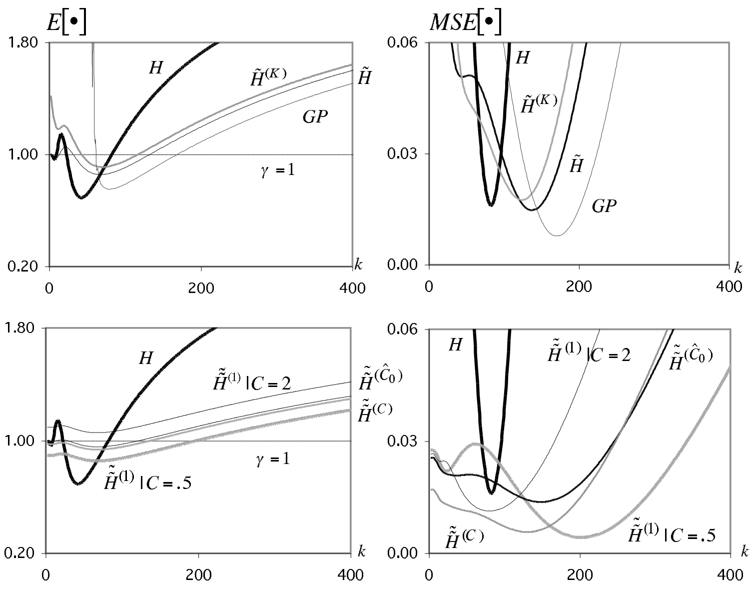

Figure 7: *Sin-Burr* parent with $\gamma = 1$ and $\rho = -.5$.

Figures 8 and 9 are equivalent to the previous figures, but for standard models with $C \neq 1$, in (1.13) — the Student models with $\nu = 4$, and 2 degrees of freedom. Notice that for the Student model with ν degrees of freedom,

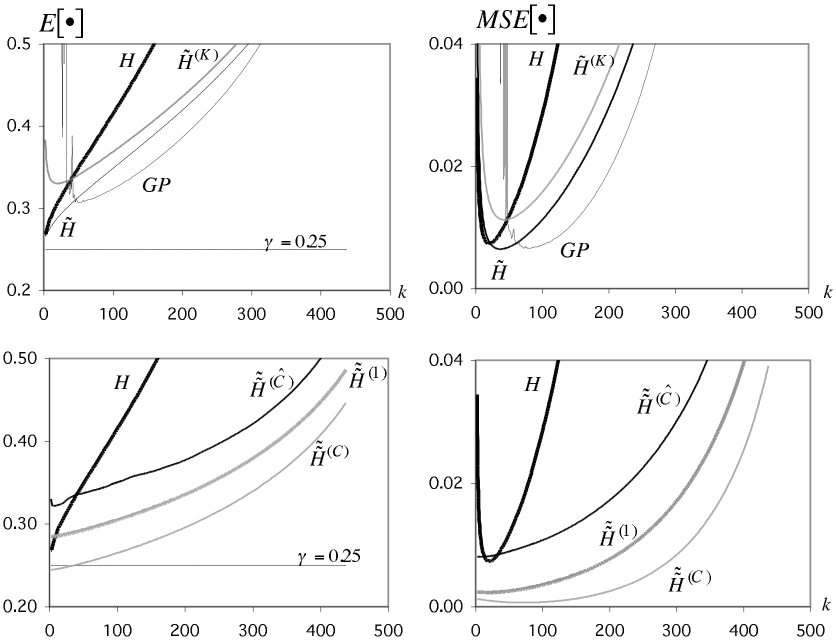

Figure 8: *Student*(4) parent with $\gamma = .25$ and $\rho = -.5$ ($C = 1.32$).

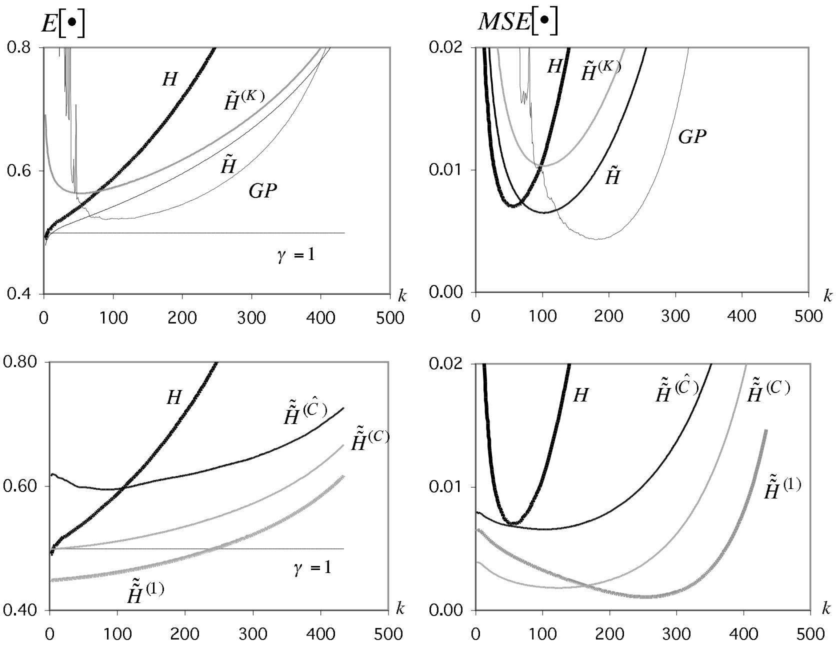

Figure 9: *Student*(2) parent with $\gamma = .5$ and $\rho = -1$ ($C = 0.71$).

$C=(-c_\nu\,\nu^{\nu/2})^{1/\nu}$, where c_ν is given for instance in Martins (2000). For $\nu=4$, and 2 we have $c_4=-3/16$, and $-1/4$, respectively. Consequently, for these models $C=1.32$ and 0.71, respectively. Such as in Figure 2, it is clear the existence of a bias close now to $\ln C/\ln n = 0.04$, and -0.05 for $\nu=4$ (Figure 8) and $\nu=2$ (Figure 9), respectively. For a more exhaustive simulation study see Oliveira (2002).

6. FINITE SAMPLE BAHAVIOUR AND ROBUSTNESS OF THE ESTIMATORS

The estimators $\widehat{\gamma}_n^H$, $\widetilde{\gamma}_n$, $\widetilde{\gamma}_n^{(\mathcal{K})}$ and $\widetilde{\widetilde{\gamma}}_n^{(\widehat{C})}$ (or $\widetilde{\widetilde{\gamma}}_n^{(\widehat{C}_0)}$, whenever we are outside Hall's class) will be also denoted $\gamma_n^{(1)}$, $\gamma_n^{(2)}$, $\gamma_n^{(3)}$ and $\gamma_n^{(4)}$, respectively. The r.v. $\widetilde{\widetilde{\gamma}}^{(C)}$ will be denoted $\gamma_n^{(5)}$. For the comparison of $\gamma_n^{(j)}$, $j=1,2,3,4$ and 5, at their optimal levels, we have implemented a multi-sample simulation of size 5000×10 in order to guarantee small standard errors (not presented in the tables, but available from the authors) for the simulated characteristics, the Mean Value ($E_\bullet$), the Mean Squared Error ($MSE_\bullet$), the Optimal Sample Fraction, $k_0^\bullet/n$, with $k_0^\bullet := \arg\min_k MSE_\bullet(k)$, and the Relative Efficiency ($REFF_\bullet$), defined as

$$(6.1)\qquad REFF_\bullet \;=\; REFF\big[\gamma_{n0}^\bullet\big] \;=\; \sqrt{MSE_s\big[\gamma_n^{(1)}\big(k_{0s}^{(1)}(n)\big)\big]\big/MSE_s\big[\gamma_n^\bullet\big(k_{0s}^\bullet(n)\big)\big]}\;,$$

with $\gamma_{n0}^\bullet=\gamma_n^\bullet(k_{0s}^\bullet(n))$, and where MSE_s denotes the simulated MSE of the estimator at its simulated optimal level. The simulator of for instance $k_0^\bullet(n)$, denoted by $k_{0s}^\bullet(n)$, is $\widehat{E}_{10}\big[\,\overline{k}_0^\bullet(n)\big]$, the average of the 10 independent replicates of $\overline{k}_0^\bullet(n)=\arg\min_k \sum_{j=1}^{5000}\big(\gamma_{nj}^\bullet(k)-\gamma\big)^2$. The simulated mean values of these five estimators, at their optimal levels, are presented in Table 1 (Fréchet and Burr parents), Table 3 (Student parents) and Table 5 (Out-Hall, Sin-Fréchet and Sin-Burr parents). Tables 2, 4 and 6 are equivalent to tables 1, 3 and 5, respectively, with simulated mean values replaced by simulated mean squared errors.

Finally in Table 4 we present the *REFF*'s of the estimator $\widetilde{\widetilde{\gamma}}^{(1)}$ at its optimal level, for models with a scale $C\neq 1$. Note that the lost in efficiency is very high for the Fréchet, Sin-Fréchet and Burr models with a large scale, as it is the scale $C=2$, used here for illustration.

Table 1: Simulated mean values and mean squared errors of $\widehat{\gamma}_n^H$, $\widetilde{\gamma}_n$, $\widetilde{\gamma}_n^{(K)}$, $\widetilde{\widetilde{\gamma}}_n^{(\widehat{C})}$, and $\widetilde{\widetilde{\gamma}}_n^{(C)}$ at the simulated optimal levels, for Fréchet and Burr parents.

n	$E_{(1)}$	$E_{(2)}$	$E_{(3)}$	$E_{(4)}$	$E_{(5)}$	$MSE_{(1)}$	$MSE_{(2)}$	$MSE_{(3)}$	$MSE_{(4)}$	$MSE_{(5)}$
Fréchet parent: $\rho = -1$, $\gamma = 1$										
100	1.0987	1.0863	1.1682	1.0877	1.0347	0.0423	0.0365	0.0624	0.0393	0.0188
500	1.0628	1.0597	1.0958	1.1062	1.0297	0.0135	0.0130	0.0192	0.0131	0.0062
1000	1.0490	1.0487	1.0739	1.0501	1.0252	0.0083	0.0082	0.0115	0.0081	0.0039
2000	1.0380	1.0388	1.0565	1.0453	1.0207	0.0051	0.0050	0.0068	0.0050	0.0024
5000	1.0294	1.0286	1.0397	1.0375	1.0159	0.0027	0.0027	0.0035	0.0027	0.0013
Burr parent: $\rho = -0.5$, $\gamma = 1$										
100	1.2920	1.2825	1.4944	1.5005	1.0308	0.2286	0.1966	0.3745	0.2342	0.0329
500	1.1851	1.1785	1.2876	1.2718	1.0342	0.0834	0.0736	0.1245	0.0826	0.0121
1000	1.1545	1.1488	1.2321	1.2147	1.0330	0.0557	0.0500	0.0808	0.0555	0.0083
2000	1.1329	1.1236	1.1872	1.1861	1.0303	0.0374	0.0339	0.0527	0.0379	0.0057
5000	1.1021	1.0980	1.1424	1.0858	1.0266	0.0228	0.0207	0.0306	0.0228	0.0035
Burr parent: $\rho = -1$, $\gamma = 1$										
100	1.1361	1.1331	1.2424	1.2034	1.0452	0.0705	0.0648	0.1138	0.0664	0.0246
500	1.0782	1.0776	1.1266	1.1207	1.0336	0.0216	0.0206	0.0316	0.0207	0.0080
1000	1.0640	1.0625	1.0968	1.0699	1.0286	0.0132	0.0128	0.0188	0.0129	0.0050
2000	1.0498	1.0494	1.0741	1.0664	1.0238	0.0082	0.0080	0.0112	0.0079	0.0032
5000	1.0373	1.0368	1.0523	1.0330	1.0185	0.0043	0.0043	0.0057	0.0042	0.0017
Burr parent: $\rho = -2$, $\gamma = 1$										
100	1.0660	1.0648	1.1301	1.0179	1.0333	0.0294	0.0295	0.0469	0.0319	0.0181
500	1.0376	1.0385	1.0639	1.0655	1.0232	0.0077	0.0085	0.0118	0.0098	0.0052
1000	1.0290	1.0299	1.0465	1.0534	1.0186	0.0044	0.0049	0.0065	0.0059	0.0030
2000	1.0218	1.0228	1.0338	1.0290	1.0146	0.0025	0.0028	0.0036	0.0030	0.0017
5000	1.0148	1.0160	1.0222	0.9970	1.0103	0.0012	0.0013	0.0016	0.0013	0.0008

Table 2: Simulated mean values and mean squared errors of $\widehat{\gamma}_n^H$, $\widetilde{\gamma}_n$, $\widetilde{\gamma}_n^{(K)}$, $\widetilde{\widetilde{\gamma}}_n^{(\widehat{C})}$, and $\widetilde{\widetilde{\gamma}}_n^{(C)}$ at the simulated optimal levels, for Student parents.

n	$E_{(1)}$	$E_{(2)}$	$E_{(3)}$	$E_{(4)}$	$E_{(5)}$	$MSE_{(1)}$	$MSE_{(2)}$	$MSE_{(3)}$	$MSE_{(4)}$	$MSE_{(5)}$
Student(4) parent: $\rho = -0.5$, $\gamma = 0.25$										
100	0.3559	0.3548	0.4400	0.4435	0.2539	0.0316	0.0269	0.0557	0.0395	0.0029
500	0.3171	0.3137	0.3580	0.3287	0.2575	0.0109	0.0095	0.0174	0.0123	0.0010
1000	0.3037	0.3027	0.3365	0.3355	0.2576	0.0072	0.0063	0.0111	0.0081	0.0007
2000	0.2956	0.2939	0.3197	0.3228	0.2574	0.0048	0.0043	0.0072	0.0054	0.0005
5000	0.2860	0.2844	0.3025	0.2947	0.2569	0.0029	0.0026	0.0041	0.0031	0.0003
Student(2) parent: $\rho = -1$, $\gamma = 0.5$										
100	0.5988	0.5937	0.6832	0.6047	0.5235	0.0374	0.0329	0.0640	0.0381	0.0081
500	0.5556	0.5545	0.5959	0.5295	0.5187	0.0114	0.0104	0.0174	0.0107	0.0029
1000	0.5448	0.5425	0.5721	0.5426	0.5155	0.0069	0.0063	0.0101	0.0064	0.0018
2000	0.5356	0.5335	0.5546	0.5238	0.5127	0.0042	0.0039	0.0059	0.0040	0.0011
5000	0.5253	0.5235	0.5370	0.5240	0.5096	0.0022	0.0020	0.0029	0.0021	0.0006
Student(1) parent: $\rho = -2$, $\gamma = 1$										
100	1.0929	1.0642	1.1866	0.7481	1.0334	0.0608	0.0601	0.1019		0.0261
500	1.0557	1.0551	1.0985	1.0848	1.0305	0.0163	0.0174	0.0258	0.0187	0.0081
1000	1.0410	1.0431	1.0722	1.0340	1.0239	0.0093	0.0101	0.0143	0.0112	0.0048
2000	1.0314	1.0332	1.0526	1.0656	1.0184	0.0053	0.0059	0.0079	0.0066	0.0028
5000	1.0217	1.0231	1.0342	1.0151	1.0132	0.0025	0.0028	0.0036	0.0031	0.0014

Table 3: Simulated mean values and mean squared errors of $\widehat{\gamma}_n^H$, $\widetilde{\gamma}_n$, $\widetilde{\gamma}_n^{(K)}$, $\widetilde{\widetilde{\gamma}}_n^{(\widehat{C}_0)}$, and $\widetilde{\widetilde{\gamma}}_n^{(C)}$ at the simulated optimal levels, for Out-Hall, Sin-Fréchet and Sin-Burr parents.

n	$E_{(1)}$	$E_{(2)}$	$E_{(3)}$	$E_{(4)}$	$E_{(5)}$	$MSE_{(1)}$	$MSE_{(2)}$	$MSE_{(3)}$	$MSE_{(4)}$	$MSE_{(5)}$
Out-Hall parent: $\rho=-1$, $\gamma=1$										
100	0.7178	0.7228	0.7881	0.7320	0.9248	0.1574	0.1568	0.1289	0.1517	0.0146
500	0.8255	0.8325	0.8808	0.8368	0.9569	0.0653	0.0644	0.0502	0.0625	0.0103
1000	0.8613	0.8657	0.9062	0.8704	0.9618	0.0437	0.0431	0.0333	0.0418	0.0076
2000	0.8892	0.8922	0.9252	0.8961	0.9669	0.0291	0.0286	0.0220	0.0278	0.0055
5000	0.9166	0.9189	0.9436	0.9220	0.9728	0.0169	0.0166	0.0128	0.0161	0.0034
Sin-Fréchet parent: $\gamma=1$										
100	1.0284	1.0680	1.1263	1.0629	1.0391	0.0359	0.0380	0.0593	0.0347	0.0209
500	1.0067	1.0223	1.0372	1.0184	1.0155	0.0073	0.0091	0.0114	0.0078	0.0056
1000	1.0025	1.0133	1.0204	1.0107	1.0098	0.0036	0.0047	0.0055	0.0040	0.0031
2000	1.0011	1.0075	1.0104	1.0060	1.0058	0.0018	0.0024	0.0027	0.0021	0.0016
5000	1.0007	1.0033	1.0041	1.0027	1.0027	0.0007	0.0010	0.0010	0.0009	0.0007
Sin-Burr parent: $\rho=-0.5$, $\gamma=1$										
100	1.0459	1.1088	1.2267	1.0873	1.0382	0.1601	0.1314	0.2380	0.1498	0.0296
500	1.0116	1.0301	1.0480	1.0286	1.0184	0.0323	0.0290	0.0378	0.0277	0.0100
1000	1.0042	1.0168	1.0235	1.0150	1.0115	0.0163	0.0151	0.0179	0.0140	0.0059
2000	1.0030	1.0091	1.0110	1.0078	1.0066	0.0081	0.0077	0.0085	0.0070	0.0033
5000	1.0011	1.0037	1.0041	1.0033	1.0028	0.0032	0.0031	0.0033	0.0028	0.0015

Table 4: Relative efficiencies of $\widetilde{\widetilde{\gamma}}_n^{(1)}$ relatively to $\widehat{\gamma}_n^H$, at their optimal levels, for models with $C \neq 1$.

n	100	500	1000	2000	5000
Fréchet ($\gamma=1$, $\rho=-1$, $C=.5$)	1.5713	1.9344	2.0609	2.1933	2.4099
Fréchet ($\gamma=1$, $\rho=-1$, $C=2$)	0.9726	0.7781	0.7039	0.6304	0.5422
Burr ($\gamma=1$, $\rho=-.5$, $C=.5$)	2.8678	3.2438	3.4208	3.5979	3.9156
Burr ($\gamma=1$, $\rho=-.5$, $C=2$)	2.0830	1.7470	1.6127	1.4894	1.3433
Burr ($\gamma=1$, $\rho=-1$, $C=.5$)	1.9624	2.1431	2.2474	2.3686	2.5641
Burr ($\gamma=1$, $\rho=-1$, $C=2$)	1.1404	0.9347	0.8472	0.7660	0.6569
Burr ($\gamma=1$, $\rho=-2$, $C=.5$)	1.3344	1.4941	1.5549	1.6196	1.7241
Burr ($\gamma=1$, $\rho=-2$, $C=2$)	0.8069	0.6189	0.5436	0.4719	0.3808
Student(4) ($\gamma=0.25$, $\rho=-0.5$, $C=1.32$)	2.5488	1.9430	1.7554	1.5875	1.3802
Student(2) ($\gamma=0.5$, $\rho=-1$, $C=.71$)	2.0645	2.4684	2.5236	2.6085	2.7371
Student(1) ($\gamma=1$, $\rho=-2$, $C=.32$)	0.9243	1.0394	1.3273	1.8245	1.9260
Out-Hall ($\gamma=1$, $C=.5$)	1.9073	1.5357	1.4122	1.2991	1.1562
Out-Hall ($\gamma=1$, $C=2$)	3.2856	3.8498	3.6741	3.6192	3.6691
Sin-Fréchet ($\gamma=1$, $C=.5$)	1.4649	1.4346	1.3653	1.3087	1.2586
Sin-Fréchet ($\gamma=1$, $C=2$)	0.8766	0.6537	0.5592	0.4720	0.3658
Sin-Burr ($\gamma=1$, $\rho=-.5$, $C=.5$)	2.5707	2.0930	1.9252	1.7945	1.6664
Sin-Burr ($\gamma=1$, $\rho=-.5$, $C=2$)	1.6669	1.2783	1.1763	1.1088	1.0713

A few final remarks:

1. The class of statistics $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ revealed a surprisingly good behaviour among the estimators considered both for small and for large sample sizes, and for all the simulated models (most of them in the class of models where (1.13) holds, with $C = 1$). Indeed, for every k, we have got $MSE\big[\widetilde{\widetilde{\gamma}}_n(k)\big]$ smaller than $MSE\big[\gamma_n^H(k)\big]$ and also smaller than $MSE\big[\widetilde{\gamma}_n(k)\big]$ for all models simulated. This class of statistics also enables us to find an estimator of the tail index γ, which behaves better than the maximum lilelihood estimator based on the Generalized Pareto excesses, for most of the models simulated.

2. Particularly astonishing is the behaviour of $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ for small values of ρ, like the value $\rho = -2$ used herein for illustration, a region where has been claimed to be difficult to find good competitors for the Hill estimator. Also, the results obtained for models for which the second order condition does not hold deserve further investigation, and are interesting from a point of view of a more general application.

3. It may be claimed that such a good behaviour is due to the fact that $\widetilde{\widetilde{\gamma}}_n(k)$ is not only non-invariant for location, like the Hill statistic, but also non-invariant for scale. The adequate estimation of the parameter C is a possible way out, but that induces an increase in the variance of our final tail index estimator, and the nice features of this estimator will disappear. Alternatively, the best way to proceed is to estimate the mean squared errors of our estimators as functions of k, merely on the basis of the available sample, proceeding next to the adaptive choice of the k and a τ-value providing a mean squared error smaller than that of the Hill estimator for every k. It is perhaps also sensible to use an extra *tuning* parameter a, a shift in the location of our data, like in Gomes and Oliveira (2003b). All this work is essentially computational, and as said before, overpasses the scope of this paper.

REFERENCES

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

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

[3] Caeiro, F. and Gomes, M.I. (2002). A new class of asymptotically unbiased estimators of the index of regular variation, *Test*, **11**(2), 345–364.

[4] Csörgő, S. and Viharos, L. (1997). Asymptotic normality of least squares estimators of tail indices, *Bernoulli*, **3**, 351–370.

[5] Csörgő, S. and Viharos, L. (1998). *Estimating the tail index.* In "Asymptotic Methods in Probability and Statistics" (Szyszkowicz, Ed.), North-Holland, Amsterdam, 833–881.

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

[7] Davison, A. (1984). *Modelling excesses over high threshold with an application.* In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), D. Reidel, 461–482.

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

[9] Drees, H.; Ferreira, A. and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index, *Ann. Appl. Prob.*, **14**, 1179–1201.

[10] Ferreira, A.; de Haan, L. and Peng, L. (2003). On optimising the estimation of high quantiles of a probability distribution, *Statistics*, **37**(5), 401–434.

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

[12] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics* (2nd edition), Krieger.

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

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

[15] Gomes, M.I. (2002). *A note on the excesses over a high threshold*, Notas e Comunicações, CEAUL 10/2002.

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

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

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

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

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

[21] Gomes, M.I. and Oliveira, O. (2001). The bootstrap methodology in Statistics of Extremes — choice of the optimal sample fraction, *Extremes*, **4**(4), 331–358.

[22] Gomes, M.I. and Oliveira, O. (2003a). Maximum likelihood revisited under a semi-parametric context — estimation of the tail index, *J. Statist. Comp. and Simul.*, **73**(4), 285–301.

[23] Gomes, M.I. and Oliveira, O. (2003b). How can non-invariant statistics work in our benefit in the semi-parametric estimation of parameters of rare events, *Commun. in Statist. — Simulation and Computation*, **32**(4), 1005–1028.

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

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

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

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

[28] Kratz, M. and Resnick, S. (1996). The *QQ*-estimator of the index of regular variation, *Commun. in Statist. — Stochastic Models*, **12**, 699–724.

[29] Martins, M.J. (2000). *Estimação de Caudas Pesadas — Variantes ao Estimador de Hill*, Ph. D. Thesis, DEIO, FCUL.

[30] Oliveira, O. (2002). *In Extremis*, Ph. D. Thesis, DEIO, FCUL.

[31] Schultze, J. and Steinbach, J. (1996). On least squares estimates of an exponential tail coefficient, *Statistics and Decisions*, **14**, 353–372.

[32] Smith, R.L. (1984a). *Threshold methods for sample extremes.* In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), D. Reidel, 621–638.

[33] Smith, R.L. (1984b). Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, **72**, 67–90.

[34] Smith, R.L. (1987). Estimating tails of probability distributions, *The Annals of Statistics*, **15**(3), 1174–1207.