Direct reduction of bias of the classical Hill estimator*

Frederico Caeiro

Universidade Nova de Lisboa

M. Ivette Gomes and Dinis Pestana

Universidade de Lisboa, FCUL (DEIO) and CEAUL

Abstract. In this paper we are interested in the estimation of the dominant component of the bias of Hill's estimator of a positive tail index γ, in order to remove it from the classical Hill estimator in different asymptotically equivalent ways. If the scale and shape second order parameters in the bias are both computed at a fixed level k_1 of a larger order than that of the level k at which the Hill estimator is computed, there is no change in the asymptotic variances of these reduced bias tail index estimators, which are kept equal to the asymptotic variance of the Hill estimator, i.e., equal to γ^2. To enhance the good performance of this type of estimators, we also consider the estimation of the shape second order parameter only, at the same level k used for the tail index estimation, as well the estimation of all unknown parameters at the same level k. The asymptotic distributional properties of the proposed estimators of γ are derived and the estimators are compared not only asymptotically, but also for finite samples through Monte Carlo techniques. An application to the log-exchange rates of the Euro against the USA Dollar is also provided.

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

*Research partially supported by FCT / POCTI / FEDER.

Keywords and phrases. *Statistics of Extremes*, *Semi-parametric estimation*, *Bias estimation*, *Heavy tails.*

1 Introduction and motivation for the new tail index estimators

In *Statistics of Extremes*, the tail index γ is the basic parameter of extreme events. Such a parameter plays a relevant role in other extreme events' parameters, like the high quantiles and the return periods of high levels, among others. The tail index is a real-valued parameter and the heavier the tail, the larger the tail index γ is. Heavy-tailed models have revealed to be quite useful in most diversified fields, like computer science, telecommunication networks, insurance and finance. In the field of Extremes, we say that a model F is *heavy-tailed* whenever the *tail function*, $\overline{F} := 1 - F$, is a regularly varying function with a negative index of regular variation equal to $\{-1/\gamma\}$, $\gamma > 0$, or equivalently, the quantile function $U(t) = F^{\leftarrow}(1 - 1/t),\ t \geq 1$, with $F^{\leftarrow}(x) = \inf\{y : F(y) \geq x\}$, is of regular variation with index γ. This means that, for every $x > 0$,

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

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

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

with $\gamma > 0$, and we denote this fact by $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$. We shall here concentrate on these Pareto-type distributions, for which (1.1) holds true. Note that (1.1) is equivalent to say that

$$1 - F(x) = x^{-1/\gamma}\, L_F(x) \quad \Longleftrightarrow \quad U(x) = x^{\gamma}\, L_U(x), \tag{1.2}$$

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

The *second order parameter* ρ (≤ 0), rules the rate of convergence in the first order condition (1.1) (or equivalently, (1.2)), and is the non-positive parameter appearing in the limiting relation

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

which we assume to hold for every $x > 0$, and where $|A(t)|$ must then be of regular variation with index ρ (Geluk and de Haan, 1987). We shall assume everywhere that $\rho < 0$.

Remark 1.1. *For the strict Pareto model,* $F(x) = 1 - C\ x^{-1/\gamma}$, $x \geq C^\gamma$, *and indeed only for this model, the numerator of the fraction in the left hand-side of* (1.3) *is null, i.e.,* $\ln U(tx) - \ln U(t) - \gamma \ln x \equiv 0$. *We then consider that* $A(t) \equiv 0$ *in* (1.3).

Remark 1.2. *For Hall's class of Pareto-type models* (*Hall, 1982; Hall and Welsh, 1985*), *with a tail function*

$$1 - F(x) = Cx^{-1/\gamma}\left(1 + Dx^{\rho/\gamma} + o\left(x^{\rho/\gamma}\right)\right), \quad \textit{as} \quad x \to \infty, \quad (1.4)$$

$C > 0$, $D \in \mathbb{R}$, $\rho < 0$, (1.3) *holds and we may choose* $A(t) = \gamma\rho D C^\rho t^\rho$. *This is a class of models where* (1.2) (*or equivalently,* (1.1)) *holds true, with an asymptotically constant slowly varying function* L_F (*or equivalently,* L_U).

For intermediate k, i.e., a sequence of integers $k = k_n$, $1 \leq k < n$, such that

$$k = k_n \to \infty, \quad k_n = o(n), \text{ as } n \to \infty, \quad (1.5)$$

we shall consider, as basic statistics, both the log-excesses over the random high level $\{\ln X_{n-k:n}\}$, i.e.,

$$V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}, \quad 1 \leq i \leq k < n, \quad (1.6)$$

and the scaled log-spacings,

$$U_i := i \{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}, \quad 1 \le i \le k < n, \tag{1.7}$$

where $X_{i:n}$ denotes, as usual, the i-th ascending order statistic (o.s.), $1 \le i \le n$, associated to a random sample $(X_1, X_2, \cdots, X_n)$.

Since $X_{i:n} \stackrel{d}{=} U(Y_{i:n})$ where $\{Y_i\}$ denotes a sequence of unit Pareto random variables (r.v.'s), i.e., $P(Y \le y) = 1 - 1/y,\ y \ge 1$, and since for $j > i$, $Y_{j:n}/Y_{i:n} \stackrel{d}{=} Y_{j-i:n-i}$, $\ln Y_{i:n} \stackrel{d}{=} E_{i:n}$ where $\{E_i\}$ denotes a sequence of independent standard exponential r.v.'s and for intermediate k, $Y_{n-k:n} \sim n/k \ \to \ \infty$, as $n \to \infty$, we may write, whenever we are under the first order framework in (1.1),

$$V_{ik} \ \stackrel{d}{=} \ \ln \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})} \ = \ \ln \frac{U(Y_{n-k:n}\ Y_{k-i+1:k})}{U(Y_{n-k:n})} \ \sim \ \gamma\ E_{k-i+1:k},\ 1 \le i \le k,$$

i.e., $V_{ik},\ 1 \le i \le k$, are approximately the k o.s.'s from an exponential sample of size k and mean value γ. Also, since $\ln Y_{1:i} \stackrel{d}{=} E_{1:i} \stackrel{d}{=} E_i/i$,

$$U_i \ \stackrel{d}{=} \ i \left(\ln \frac{U(Y_{n-i+1:n})}{U(Y_{n-i:n})} \right) \ = \ i \left(\ln \frac{U(Y_{n-i:n}\ Y_{1:i})}{U(Y_{n-i:n})} \right) \ \sim \ \gamma\ E_i,\ \ 1 \ \le \ i \ \le \ k,$$

i.e., the U_i's, $1 \le i \le k$, are approximately independent and exponential with mean value γ. Then the Hill estimator of γ (Hill, 1975),

$$H(k) \equiv \hat{\gamma}_n^H(k) := \frac{1}{k} \sum_{i=1}^{k} V_{ik} = \frac{1}{k} \sum_{i=1}^{k} U_i, \tag{1.8}$$

is consistent for the estimation of γ whenever (1.1) holds and k is intermediate, i.e., (1.5) holds.

Under the second order framework in (1.3) the asymptotic distributional representation

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

holds true (de Haan and Peng, 1998), where $Z_k^{(1)} = \sqrt{k}\left(\sum_{i=1}^{k} E_i/k - 1\right)$ is an asymptotically standard normal r.v.

Remark 1.3. *If the underlying model is the strict Pareto model in Remark 1.1,* $\ln X_{i:n} = \gamma\ E_{i:n} + \gamma \ln C$*, and the use of Rényi's representation of exponential order statistics, as a linear combination of independent unit exponential r.v.'s (Rényi, 1953),* $E_{i:n} = \sum_{j=1}^{i} E_j/(n-j+1),\ 1 \le i \le n$*, leads us to*

$$H(k) \ \stackrel{d}{=}\ \frac{\gamma}{k}\sum_{i=1}^{k}\{E_{n-i+1:n} - E_{n-k:n}\} \ \stackrel{d}{=}\ \frac{\gamma}{k}\sum_{i=1}^{k}\sum_{j=i}^{k}\frac{E_j}{j} \ \stackrel{d}{=}\ \frac{\gamma}{k}\sum_{j=1}^{k}E_j \ \stackrel{d}{=}\ \frac{\gamma}{k}\,Ga(k),$$

where $Ga(k)$ *denotes a Gamma r.v. with a shape parameter equal to* k*, i.e., a r.v. with probability density function (p.d.f.)* $x^{k-1}\exp(-x)/\Gamma(k),\ x \ge 0$*, with* $\Gamma(t)$ *denoting the complete gamma function,* $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$*. Then for every* k*, the Hill estimator in* (1.8) *is unbiased for the estimation of* γ*, i.e.,* $\mathbb{E}\left(H(k)\right) = \gamma$ *for any* k*, and* $\sqrt{k}\left(H(k) - \gamma\right)/\gamma$ *is asymptotically standard normal, as* $k \to \infty$*.*

We shall assume that we are in Hall's class of models in (1.4). Consequently, we may choose

$$A(t) = \alpha\ t^\rho =: \gamma\ \beta\ t^\rho, \quad \beta \ne 0, \quad \rho < 0. \tag{1.10}$$

The adequate accommodation of the bias of Hill's estimator has been extensively addressed in recent years by several authors. The idea is to go further into the second order framework in (1.3). Then,

$$V_{ik} \ \stackrel{d}{=}\ \ln\frac{U(Y_{n-k:n}Y_{k-i+1:k})}{U(Y_{n-k:n})} \ \stackrel{d}{\approx}\ \gamma\ E_{k-i+1:k} + A(n/k)\frac{Y_{k-i+1:k}^\rho - 1}{\rho},\ \ 1 \le i \le k,$$

and

$$U_i \ \stackrel{d}{=}\ i\ln\frac{U(Y_{n-i:n}Y_{1:i})}{U(Y_{n-i:n})} \ \stackrel{d}{\approx}\ \gamma\left(1 + \frac{A(n/k)}{\gamma}\left(\frac{k}{i}\right)^\rho\right)E_i,\ \ 1 \le i \le k.$$

Beirlant *et al.* (1999) and Feuerverger and Hall (1999) work with the scaled log-spacings U_i, $1 \le i \le k$, and consider the joint estimation of the first order parameter γ and the second order parameters at the same level k; in a similar set-up, Gomes and Martins (2002) advance with the external estimation of the second order parameter ρ at a lower level (larger k) than the one used for the tail index estimation, being then able to reduce the asymptotic variance of the proposed tail index estimator, but they pay no special attention to the extra "scale" parameter $\beta \neq 0$ in the A function in (1.10). More recently, Gomes *et al.* (2004b) deal with a joint external estimation of both the "scale" and the "shape" parameters in the A function, being able to reduce the bias without increasing the asymptotic variance, which is kept at the value γ^2, the asymptotic variance of Hill's estimator. Such an estimator, also considered here for comparison with the two new estimators herewith proposed, is based on a linear combination of the excesses V_{ik} in (1.6), and is given by

$$WH_{\widehat{\beta},\,\widehat{\rho}}(k) := \frac{1}{k}\sum_{i=1}^{k} e^{\widehat{\beta}\ (n/k)^{\widehat{\rho}}\ \left((i/k)^{-\widehat{\rho}}-1\right)/(\widehat{\rho}\ln(i/k))}\ V_{ik}, \tag{1.11}$$

for adequate consistent estimators $\widehat{\beta}$ and $\widehat{\rho}$ of the second order parameters β and ρ, respectively. In the same spirit, Gomes and Pestana (2004) study, mainly computationally, the estimator

$$\widetilde{H}_{(\widehat{\beta},\,\widehat{\rho})}(k) \ := \ H(k) \ - \ \frac{\widehat{\beta}}{1-\widehat{\rho}}\left(\frac{n}{k}\right)^{\widehat{\rho}} \ H\left(\left(\frac{(1-\widehat{\rho})^2 n^{-2\widehat{\rho}}}{(-2\widehat{\rho})\ \widehat{\beta}^2}\right)^{1/(1-2\widehat{\rho})}\right), \tag{1.12}$$

with H the Hill estimator in (1.8).

We shall herewith consider the estimator

$$\overline{H}_{\widehat{\beta},\,\widehat{\rho}}(k) := H(k)\left(1-\frac{\widehat{\beta}}{1-\widehat{\rho}}\left(\frac{n}{k}\right)^{\widehat{\rho}}\right), \tag{1.13}$$

together with the asymptotically equivalent variant,

$$\overline{\overline{H}}_{\widehat{\beta},\,\widehat{\rho}}(k) := H(k)\exp\left(-\frac{\widehat{\beta}}{1-\widehat{\rho}}\left(\frac{n}{k}\right)^{\widehat{\rho}}\right). \tag{1.14}$$

The dominant component of the bias of Hill's estimator, $A(n/k)/(1-\rho) = \gamma\,\beta\,(n/k)^{\rho}\,/(1-\rho)$, is thus estimated through $H(k)\ \widehat{\beta}\ (n/k)^{\widehat{\rho}}/(1-\widehat{\rho})$ and directly removed from Hill's classical tail index estimator, through two asymptotically equivalent expressions, provided that k is intermediate, i.e., provided that (1.5) holds true.

Remark 1.4. *Note that the estimator* $\widetilde{H}$ *in* (1.12) *has been built in a way similar to the one herewith considered. The difference is that the bias* $\gamma\ \beta\,(n/k)^{\rho}\,/(1-\rho)$ *is estimated through* $H(\widehat{k}_0)\ \widehat{\beta}\,(n/k)^{\widehat{\rho}}\,/(1-\widehat{\rho})$*, being* $\widehat{k}_0$ *an estimate of the "optimal" level for the Hill estimator, in the sense of minimum mean squared error in Hall's class of models.*

Remark 1.5. *The reason to consider the two asymptotically equivalent estimators in* (1.13) *and* (1.14) *— also symptotically equivalent to the estimator in* (1.12) *— lies on the fact that we have the clear experience that asymptotically equivalent estimators may exhibit quite different sample paths' properties. In practice, one never knows the peculiarities of the underlying models, and it is thus sensible to work with a set of a few estimators of the primary parameter of rare events, in order to estimate it in the most possible reliable way.*

1.1 A graphical motivation

For the second order parameters' estimators, discussed later on, in section 2, we exhibit here, in Figure 1, the differences between the sample paths of the estimators $\overline{H}_{\bullet}(k)$ in (1.13), for a sample of size $n = 10,000$ from a Fréchet model, with d.f. $F(x) = \exp\left(-x^{-1/\gamma}\right),\ x \geq 0$, with $\gamma = 1$, when both $\widehat{\rho}$ and $\widehat{\beta}$ are computed at that high level k_1 (*left*), when we compute $\widehat{\beta}$ and $\widehat{\rho}$ at the same level k used for the estimation of the tail index γ (*right*) and when we compute only $\widehat{\beta}$ at that same level k, being $\widehat{\rho}$ computed at a larger k-value, let us say an intermediate level k_1 such that $\sqrt{k_1}\ A(n/k_1) \to \infty$, as $n \to \infty$ (*center*). It thus seems sensible to investigate the behaviour of the first type of estimation procedure, the one which seems to be the most adequate to be used in practice, due to the higher stability of the sample path in Figure 1 (*left*), around the target value $\gamma = 1$. It is also sensible to compare asymptotically the three estimation procedures, in order to detect the reasons for the differences in behaviour.

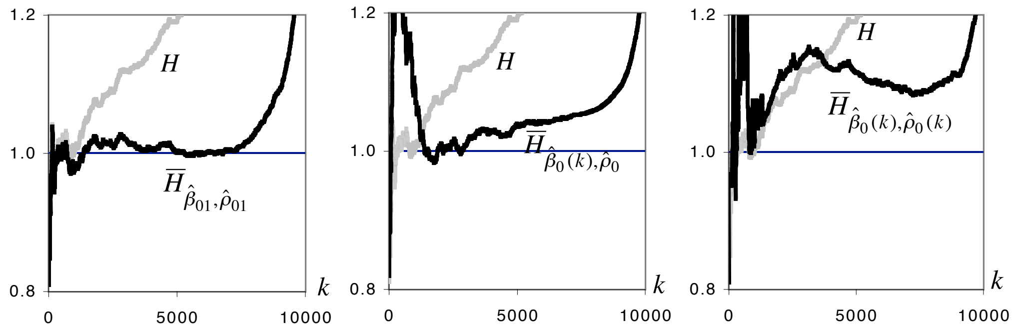


Figure 1: Sample paths of the Hill and the tail index estimates $\overline{H}$ in (1.13), obtained through the estimation of (β, ρ) at the level $k_1 := \min(n-1, 2n/\ln\ln n)$ (*left*) versus the estimation at the same level both for β and ρ (*right*) and only for β (*center*).

1.2 A technical motivation

In the lines of Gomes and Martins (2004):

Lemma 1.1. *Under the second order framework in* (1.3)*, and for levels k such that* (1.5) *holds, the distributional representation*

$$\frac{\alpha}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1} U_i \stackrel{d}{=} \gamma + \frac{\gamma\ \alpha}{\sqrt{(2\alpha-1)\ k}}\ Z_k^{(\alpha)} + \frac{\alpha\ A(n/k)}{\alpha-\rho}(1+o_p(1)) \quad (1.15)$$

holds true for any $\alpha \geq 1$, where

$$Z_k^{(\alpha)} = \sqrt{(2\alpha-1)\ k}\ \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1} E_i - \frac{1}{\alpha}\right) \quad (1.16)$$

are asymptotically standard normal r.v.'s.

If we assume that only the tail index parameter γ in unknown, and similarly to the result in Gomes and Pestana (2004), we shall now state and prove a theorem that provides an obvious technical motivation for the estimator in

(1.13) (or in (1.14)):

Theorem 1.1. *Under the second order framework in* (1.3)*, further assuming that* $A(t)$ *may be chosen as in* (1.10)*, and for levels* k *such that* (1.5) *holds, we get, for* $\overline{H}_{\beta,\ \rho}(k)$ *in* (1.13) (*or for* $\overline{\overline{H}}_{\beta,\ \rho}(k)$ *in* (1.14))*, an asymptotic distributional representation of the type*

$$\gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + R_k, \quad \textit{with} \quad R_k = o_p(A(n/k)), \tag{1.17}$$

where $Z_k^{(1)}$ *is the asymptotically standard normal r.v. in* (1.16) *for* $\alpha = 1$*. Consequently, both* $\sqrt{k}\ \left(\ \overline{H}_{\beta,\ \rho}(k) - \gamma\right)$ *and* $\sqrt{k}\ \left(\overline{\overline{H}}_{\beta,\ \rho}(k) - \gamma\right)$ *are asymptotically normal with variance equal to* γ^2*, and with a null mean value not only when* $\sqrt{k}\ A(n/k) \longrightarrow 0$*, but also when* $\sqrt{k}\ A(n/k) \longrightarrow \lambda \neq 0$*, finite, as* $n \to \infty$*.*

Proof. The results related to the estimator in (1.13) come straighforwardly from the fact that if all parameters are known, apart from the tail index γ, we get from (1.9),

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

i.e., (1.17) holds. Since, for intermediate k,

$$\exp\left(-\frac{A(n/k)}{1-\rho}\right) = 1 - \frac{A(n/k)}{1-\rho} + o_p(A(n/k)),$$

the same distributional representation in (1.17) holds true for the tail index estimator in (1.14). The remaining of the theorem follows then straightforwardly.

□

1.3 Scope of the paper

When we look at the expression of the estimators from (1.11) till (1.14), we see that one of the topics to deal with is the adequate estimation of (β, ρ) in order to get the tail index estimators $\overline{H}_{\widehat{\beta},\, \widehat{\rho}}(k)$ and $\overline{\overline{H}}_{\widehat{\beta},\, \widehat{\rho}}(k)$. In section 2 of this paper, we shall thus briefly review the estimation of the two second order parameters β and ρ. Section 3 is devoted to the derivation of the asymptotic behaviour of the estimator $\overline{H}_{\widehat{\beta},\, \widehat{\rho}}(k)$ in (1.13) (equivalently, $\overline{\overline{H}}_{\widehat{\beta},\, \widehat{\rho}}(k)$ in (1.14)), estimating β and ρ at a larger k value than the one used for the tail index estimation. We also do that only with the estimation of ρ, estimating β at the same level k used for the tail index estimation. We shall consider as well the estimation of γ, β and ρ at the same level k. In section 4, and through the use of simulation techniques, we shall exhibit the performance of the new estimators in (1.13) and (1.14), comparatively to the WH estimator in (1.11), to the classical Hill estimator and to $\overline{H}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)$, for the β-estimator, $\widehat{\beta}_{\widehat{\rho}}(k)$, in Gomes and Martins (2002). Computationally, we have considered only an external estimation of the second order parameter ρ. Such a decision is related to the discussion in Gomes and Martins (2002) on the advantages of an external estimation of the second order parameter ρ (or even their misspecification, as in Gomes and Martins (2004)) versus an internal estimation at the same level k. Indeed, the estimation of γ, β and ρ at the same level k leads to very volatile mean values and mean squared error patterns. Finally, in section 5, an illustration of the behaviour of these new reduced bias' estimators is provided, through the analysis of the exchange rate of Euro against USA Dollar and some overall conclusions are drawn.

2 Second order parameter estimation

2.1 The estimation of ρ

We shall first address the estimation of ρ. We have nowadays two general classes of ρ-estimators, which work well in practice, the ones introduced in Gomes *et al.* (2002) and Fraga Alves *et al.* (2003). We shall consider here particular members of the class of estimators of the second order parameter ρ proposed by Fraga Alves *et al.* (2003). Under adequate general conditions, they are semi-parametric asymptotically normal estimators of ρ, whenever $\rho < 0$, which show, for a large diversity of models, highly stable sample paths as functions of k, the number of top o.s.'s used, for a wide range of large k-values. Such a class of estimators has first been parameterised in a tuning parameter $\tau > 0$, which may be straightforwardly generalized to the case $\tau \in \mathbb{R}$, as done in Caeiro and Gomes (2004). Such a class may be defined as

$$\widehat{\rho}_\tau(k) \equiv \widehat{\rho}_n^{(\tau)}(k) := -\left|\frac{3(T_n^{(\tau)}(k)-1)}{T_n^{(\tau)}(k)-3}\right|, \tag{2.1}$$

where

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

with

$$M_n^{(j)}(k) := \sum_{i=1}^{k}\left\{\ln\frac{X_{n-i+1:n}}{X_{n-k:n}}\right\}^j, \; j \geq 1 \quad \left[M_n^{(1)} \equiv H \text{ in } (1.8)\right].$$

We shall here summarize a particular case of the results proved in Fraga Alves *et al.* (2003), now related to the asymptotic behaviour of the ρ-estimator in (2.1), under the second order framework in (1.3):

Proposition 2.1. *Under the second order framework in* (1.3), *with* $\rho < 0$, *if* (1.5) *holds, and if* $\sqrt{k}\ A(n/k) \to \infty$, *as* $n \to \infty$, *the statistic* $\widehat{\rho}_n^{(\tau)}(k)$ *in* (2.1) *converges in probability towards* ρ, *as* $n \to \infty$, *for any* $\tau \in \mathbb{R}$. *More than that: we have* $\widehat{\rho}_n^{(\tau)}(k) - \rho = O_p\left(1/\left(\sqrt{k}\ A(n/k)\right)\right)$, *and more specifically,*

$$\widehat{\rho}_n^{(\tau)}(k) - \rho \overset{d}{=} \frac{\sigma_\rho W_k^*}{\sqrt{k}\ A(n/k)} + R_k^*, \quad \textit{with} \quad R_k^* = o_p(1) \tag{2.2}$$

and

$$\sigma_\rho^2 \equiv \sigma_\rho^2(\gamma) = \left(\frac{\gamma(1-\rho)^3}{\rho}\right)^2 \left(2\rho^2 - 2\rho + 1\right). \tag{2.3}$$

The r.v. W_k^* *is a standard normal r.v., which may be written as*

$$W_k^* = \frac{1}{\sqrt{2\rho^2 - 2\rho + 1}} \left((3-\rho)\overline{P}_k^{(1)} - (3-2\rho)\overline{P}_k^{(2)} + (1-\rho)\overline{P}_k^{(3)}\right), \tag{2.4}$$

where, for $\alpha \geq 1$,

$$\overline{P}_k^{(\alpha)} = \frac{1}{\Gamma(\alpha+1)\sqrt{k}} \sum_{i=1}^{k} E_i^\alpha - \sqrt{k}, \qquad \left[\overline{P}_k^{(1)} \equiv Z_k^{(1)}, \ \textit{in (1.16)}\right],$$

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

Remark 2.1. *The theoretical and simulated results in Fraga Alves et al.* (2003), *together with the use of these estimators in the Generalized Jackknife statistics of Gomes et al.* (2000), *as done in Gomes and Martins* (2002), *as well as their use in the estimator in* (1.11) (*Gomes et al.,* 2004*b*) *and in the estimator in* (1.12) (*Gomes and Pestana, 2004*), *lead us to advise in practice the consideration of the tuning parameters* $\tau = 0$ *for the region* $\rho \in [-1, 0)$ *and* $\tau = 1$ *for the region* $\rho \in (-\infty, -1)$, *together with the level*

$$k_1 = \min\left(n-1, [2n/\ln_2 n]\right) \tag{2.5}$$

(*not chosen in any optimal way*), *with the obvious notation* $\ln_2 n = \ln\ln n$. *As done before, we however advise practitioners not to choose blindly the value of*

τ. It is sensible to draw a few sample paths of $\widehat{\rho}_n^{(\tau)}(k)$ in (2.1)*, as functions of k, electing the value of τ which provides higher stability for large k, by means of any stability criterion. For details, see Gomes and Figueiredo* (2003)*. Anyway, in all the Monte Carlo simulations we have considered the level k_1 in* (2.5) *and the following ρ-estimators in* (2.1)*: $\widehat{\rho}_0$ if $\rho \geq -1$ and $\widehat{\rho}_1$ if $\rho < -1$.*

Remark 2.2. *When we consider the level k_1 in* (2.5)*, together with any of the ρ-estimators in this section, computed at the level k_1, $\{\widehat{\rho} - \rho\}$ is of the order of $1/\left(\sqrt{k_1}A(n/k_1)\right) = O\left((\ln_2 n)^{(1-2\rho)/2}/\sqrt{n}\right)$.*

Remark 2.3. *Moreover, for any level k, $(\widehat{\rho} - \rho)\ln(n/k) = o_p(1)$, and consequently $\sqrt{k}\ A(n/k)\,(\widehat{\rho} - \rho)\ln(n/k) = o_p(1)$ whenever $\sqrt{k}\ A(n/k) \to \lambda$, finite.*

2.2 Estimation of the second order parameter β

We have here considered the estimator of β obtained in Gomes and Martins (2002) and based on the scaled log-spacings $U_i = i\,\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}$ in (1.7), $1 \leq i \leq k$. Let us denote $\widehat{\rho}$ any of the estimators in (2.1) computed at the level k_1 in (2.5). The β-estimator is given by

$$\widehat{\beta}_{\widehat{\rho}}(k) := \left(\frac{k}{n}\right)^{\widehat{\rho}} \frac{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^{k}U_i\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\ U_i\right)}{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\ U_i\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\widehat{\rho}}\ U_i\right)}. \tag{2.6}$$

In Gomes and Martins (2002) and later in Gomes *et al.* (2004b), the following result has been proved:

Proposition 2.2. *If the second order condition* (1.3) *holds, with $A(t) = \gamma\ \beta\ t^{\rho}$, $\rho < 0$, if $k = k_n$ is a sequence of intermediate positive integers, i.e.* (1.5) *holds,* *and if* $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$*, then* $\widehat{\beta}_{\widehat{\rho}}(k)$ *in* (2.6) *is consistent for the estimation of* β*, whenever* $\widehat{\rho}$ *is consistent for the estimation of* ρ*. Moreover, if* ρ *is known,*

$$\widehat{\beta}_{\rho}(k) \stackrel{d}{=} \beta + \frac{\gamma\ \beta\ (1-\rho)\ \sqrt{1-2\rho}}{\rho\ \sqrt{k}\ A(n/k)}\ B_k\ + R_k^{**},\ \textit{with } R_k^{**} = o_p(1) \tag{2.7}$$

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

$$B_k = \frac{(1-\rho)\sqrt{1-2\rho}}{|\rho|}\left(\frac{Z_k^{(1)}}{1-\rho} - \frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}}\right), \tag{2.8}$$

with $Z_k^{(\alpha)}$*,* $\alpha \geq 1$*, given in* (1.16)*.*

The asymptotic distributional representation (2.7) *holds true as well for* $\widehat{\beta}_{\widehat{\rho}}(k)$*, with* $\widehat{\rho}$ *any of the estimators in* (2.1) *computed at the level* k_1 *in* (2.5)*. If* $\sqrt{k}\ A(n/k)\ R_k^{**} \to \lambda_R^{**}$*, finite, we may further guarantee the asymptotic normality of* $\widehat{\beta}_{\widehat{\rho}}(k)$*. If we consider* $\widehat{\beta}_{\widehat{\rho}(k)}(k)$*, then*

$$\widehat{\beta}_{\widehat{\rho}(k)}(k) - \beta \ \stackrel{p}{\sim}\ -\beta\ \ln(n/k)\ (\widehat{\rho}(k) - \rho). \tag{2.9}$$

Remark 2.4. *Note that when we consider the level* k_1 *in* (2.5)*, and* $\widehat{\beta} \equiv \widehat{\beta}_{\widehat{\rho}}(k_1)$*, with* $\widehat{\rho}$ *any of the estimator in* (2.1)*, computed also at the same level* k_1*,* $\widehat{\beta} - \beta$ *is of the order of* $\ln(n/k_1)/\left(\sqrt{k_1}A(n/k_1)\right) = O\left(\ln_3 n\ (\ln_2 n)^{(1-2\rho)/2}/\sqrt{n}\right)$*.*

3 Asymptotic behaviour of the estimators

Let us assume first that we estimate both β and ρ externally at the level k_1 in (2.5). We may state the following:

Theorem 3.1. *Under the conditions of Theorem 1.1, let us consider the tail index estimators* $\overline{H}_{\widehat{\beta},\ \widehat{\rho}}(k)$ *and* $\overline{\overline{H}}_{\widehat{\beta},\ \widehat{\rho}}(k)$ *in* (1.13) *and* (1.14)*, respectively, for any of the estimators* $\widehat{\beta}$ *and* $\widehat{\rho}$ *in* (2.1) *and in* (2.6)*, respectively,* *both computed at the level k_1 in (2.5). Then, $\sqrt{k}\left\{\overline{H}_{\widehat{\beta},\ \widehat{\rho}}(k)-\gamma\right\}$ as well as $\sqrt{k}\left\{\overline{\overline{H}}_{\widehat{\beta},\ \widehat{\rho}}(k)-\gamma\right\}$ are asymptotically normal with null mean value, not only when $\sqrt{k}\ A(n/k)\to 0$, but also whenever $\sqrt{k}\ A(n/k)\to\lambda$, finite or infinite, provided that $\sqrt{k}\ A^2(n/k)\to\lambda^*$, finite, as $n\to\infty$.*

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

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

Since $\widehat{\beta}$ and $\widehat{\rho}$ are consistent for the estimation of β and ρ, respectively, and $(\widehat{\rho}-\rho)\ln(n/k)=o_p(1)$ (see Remarks 2.2 and 2.3), the summands related to $\left(\widehat{\beta}-\beta\right)$ and $(\widehat{\rho}-\rho)$ are both $o_p(A(n/k))$, and the result in the theorem follows immediately, provided that $\sqrt{k}\ A(n/k)\to\lambda$, finite. Moreover, since from (2.9), $\widehat{\beta}-\beta\ \stackrel{p}{\sim}\ -\beta\ \ln(n/k_1)\ (\widehat{\rho}-\rho)$,

$$\begin{aligned}\overline{H}_{\widehat{\beta},\ \widehat{\rho}}(k)-\overline{H}_{\beta,\ \rho}(k)\ &\stackrel{p}{\sim}\ -\frac{A(n/k)}{1-\rho}\left(\frac{\widehat{\beta}-\beta}{\beta}+(\widehat{\rho}-\rho)\ \ln(n/k)\right)\\ &\stackrel{p}{\sim}\ \frac{A(n/k)}{1-\rho}\ (\widehat{\rho}-\rho)\ln(k/k_1).\end{aligned}\tag{3.1}$$

If $\sqrt{k}\ A(n/k)\to\infty$ but $\sqrt{k}\ A^2(n/k)\to\lambda^*$, finite, k is at most of the order of $n^{-4\rho/(1-4\rho)}$. Then, k/k_1 is at most of the order of $n^{-1/(1-4\rho)}\ln_2 n$, and consequently,

$$\begin{aligned}0\le\sqrt{k}\ \left|(\widehat{\rho}-\rho)\ln(k/k_1)\right|\ &\le\ O_p\left(\frac{n^{\frac{-2\rho}{1-4\rho}}(\ln_2 n)^{\frac{1-2\rho}{2}}\ln n}{n^{\frac{1}{2}}}\right)\\ &=\ O_p\left(\frac{(\ln_2 n)^{\frac{1-2\rho}{2}}\ln n}{n^{\frac{1}{2(1-4\rho)}}}\right).\end{aligned}$$

Hence, $\sqrt{k}\ (\widehat{\rho}-\rho)\ A(n/k)\ \ln(k/k_1)$ converges towards zero, as $n \to \infty$, and the results in the theorem, related to the $\overline{H}$-estimator, follow. The reasoning is exactly the same for the $\overline{\overline{H}}$-estimator. $\square$

Remark 3.1. *Note that the levels k such that $\sqrt{k}\ A(n/k) \to \lambda$, finite, are sub-optimal for this type of estimators. To go further to the optimal level, we should go into a third order framework, like the one considered in Gomes and de Haan* (1999)*, Gomes et al.* (2002)*, Fraga Alves et al.* (2003) *and Gomes et al.* (2004*a*)*, considering levels k such that $\sqrt{k}\ A(n/k) \to \infty$, as $n \to \infty$.*

If we consider γ and β estimated at the same level, we are going to have an increase in the variance of our final tail index estimator $\overline{H}_{\widehat{\beta}_{\widehat{\rho}}(k),\ \widehat{\rho}}(k)$ (or equivalently, $\overline{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\ \widehat{\rho}}(k)$). Similarly to Corollary 2.1 of Theorem 2.1 in Gomes and Martins (2002), there in connection with a ML-tail index estimator, as well as in Theorem 3.2 in Gomes *et al.* (2004b), in connection with the tail index estimator in (1.11), we may also get:

Theorem 3.2. *If the second order condition* (1.3) *holds, if $k = k_n$ is a sequence of intermediate integers, i.e.,* (1.5) *holds, and if $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \lambda$, finite, non necessarily null, then*

$$\sqrt{k}\left(\overline{H}_{\widehat{\beta}_{\widehat{\rho}}(k),\ \widehat{\rho}}\ (k)-\gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}}\ Normal\left(0,\ \sigma^2_{H_2} := \gamma^2\ \left(\frac{1-\rho}{\rho}\right)^2\right), \tag{3.2}$$

i.e., the asymptotic variance of $\overline{H}_{\widehat{\beta}_{\widehat{\rho}}(k),\ \widehat{\rho}}(k)$ increases of a factor $((1-\rho)/\rho)^2$, greater than one, for every $\rho \le 0$. The same result holds obviously true for $\overline{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k)}$.

Proof. If we consider

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

we now get

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

Since $\left(\widehat{\beta}_{\widehat{\rho}}(k)-\beta\right)/\beta$ is now of the order of $1/\left(\sqrt{k}\ A(n/k)\right)$, the term of the order of $1/\sqrt{k}$ is going to be, from (2.7) and (2.8),

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

which may be written as

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

with $Z_k^{(\alpha)}$ the asymptotically standard normal r.v. in (1.16). Taking again into account the fact that the asymptotic variance between $Z_k^{(1)}$ and $Z_k^{(1-\rho)}$ is given by $\sqrt{1-2\rho}/(1-\rho)$, together with the fact that $\sqrt{k}\ A(n/k)\,(\widehat{\rho}-\rho)\ln(n/k) \to 0$ (see Remark 2.3), (3.2) follows.

□

Furthermore, if we estimate the three parameters γ, β and ρ at the same level k, we are going to have again some changes in the asymptotic behaviour of our final tail index estimators $\overline{H}_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\ \widehat{\rho}(k)}(k)$ $\left(\text{or } \overline{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\ \widehat{\rho}(k)}(k)\right)$, with an increase in their asymptotic variances. Indeed,

Theorem 3.3. *Let us assume that the second order condition* (1.3) *holds, that $k = k_n$ is a sequence of intermediate integers, i.e.,* (1.5) *holds, and that $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$. If we further have $\sqrt{k}\ R_k \to 0$, with R_k given in* (1.17), *and* (2.2) *holds with $\sqrt{k}\ A(n/k)\ R_k^* = o_p(1)$, then*

$$\sqrt{k}\left(\overline{H}_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\ \widehat{\rho}(k)}(k) - \gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal\left(0,\ \sigma^2_{H_3}\right),$$

where

$$\sigma^2_{H_3} = \gamma^2\left(1 + \frac{(1-\rho)^2(1-2\rho(1-\rho))}{\rho^2}\right), \tag{3.3}$$

i.e., we get the same rate of convergence, of the order of $1/\sqrt{k}$, *for* $\overline{H}_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\ \widehat{\rho}(k)}(k)$, *but the asymptotic variance increases. This same asymptotic result remains true if we consider the estimator* $\overline{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\ \widehat{\rho}(k)}(k)$ *instead of* $\overline{H}_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\ \widehat{\rho}(k)}(k)$.

Proof. If we now consider

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

the use of the delta-method enables us to guarantee that the statistic $\widehat{\beta}_{\widehat{\rho}(k)}(k)\left(\frac{n}{k}\right)^{\widehat{\rho}(k)}/(1-\widehat{\rho}(k))$ is asymptotically equivalent to

$$\frac{A(n/k)}{\gamma\ (1-\rho)}\left(1 + \frac{\widehat{\beta}_{\widehat{\rho}(k)}(k) - \beta}{\beta} + \left(\frac{1}{1-\rho} + \ln(n/k)\right)(\widehat{\rho}(k) - \rho)\right),$$

which is on its turn equivalent to

$$\frac{A(n/k)}{\gamma(1-\rho)}\left(1 + \frac{\widehat{\rho}(k)-\rho}{1-\rho}\right),$$

due to the validity of (2.9). We thus have, under the second order framework in (1.3) and the extra condition in the theorem,

$$\begin{aligned}\widetilde{\overline{H}}(k) &= H(k)\left(1 - \frac{A(n/k)}{\gamma\ (1-\rho)} - \frac{A(n/k)}{\gamma\ (1-\rho)^2}(\widehat{\rho}(k)-\rho)\right)(1+o_p(1)) \\ &= \overline{H}_{\beta,\ \rho}(k) - \frac{A(n/k)}{(1-\rho)^2}(\widehat{\rho}(k)-\rho)(1+o_p(1)).\end{aligned}$$

Now, since as $n \to \infty$, $\widehat{\rho}(k) - \rho = O_p\left(1/(\sqrt{k}\ A(n/k))\right)$, $\sqrt{k}\ R_k \to 0$ and $\sqrt{k}\ A(n/k)R_k^* \to 0$,

$$\sqrt{k}\left(\widetilde{\overline{H}}(k) - \gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \gamma\ Z_k^{(1)} - \frac{\sigma_\rho}{(1-\rho)^2}\ W_k^*,$$

with σ_ρ and W_k^* given in (2.3) and (2.4), respectively. In order to obtain the asymptotic variance of $\sqrt{k}\left(\widetilde{\overline{H}}(k) - \gamma\right)$ we need to obtain the variance of $\left\{\gamma\ Z_k^{(1)} - \sigma_\rho\ W_k^*/(1-\rho)^2\right\}$. Since $Z_k^{(1)} = \overline{P}_k^{(1)}$, and

$$\mathbb{C}ov\left(\overline{P}_k^{(\alpha)},\ \overline{P}_k^{(\beta)}\right) = \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta+1)} - 1,$$

we get $\mathbb{C}ov\left(\overline{P}_k^{(1)},\ \sigma_\rho\ W_k^*/(1-\rho)^2\right) = 0$. From (2.3),

$$\mathbb{V}ar\left(\gamma\ Z_k^{(1)} - \frac{\sigma_\rho}{(1-\rho)^2}\ W_k^*\right) = \gamma^2\left(1 + \frac{(1-\rho)^2(2\rho^2-2\rho+1)}{\rho^2}\right),$$

and the results in the theorem related to the $\overline{H}$-estimator follow. The equivalent result for the $\overline{\overline{H}}$-estimator follows in a similar way. □

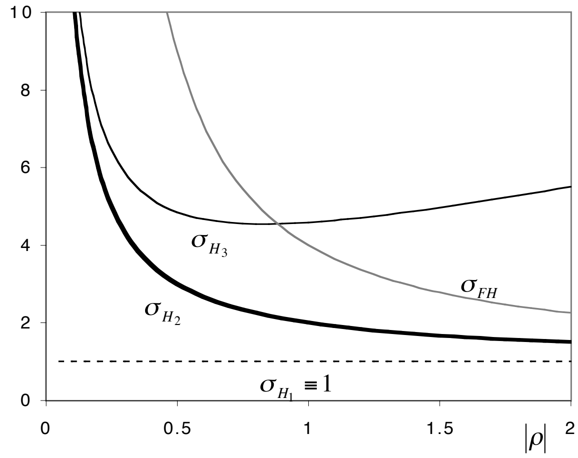
Remark 3.2. *If we compare Theorems 3.1, 3.2 and 3.3 we see that the estimation of the two parameters γ and β at the same level k induces an increase in the asymptotic variance of the final γ-estimator of a factor given by $((1-\rho)/\rho)^2$, greater than 1, whereas the estimation of γ, β and ρ at the same level k induces an extra increase in the asymptotic variance of the final γ-estimator. As may be seen in Gomes and Martins* (2002) *the asymptotic variance of the estimator in Feuerverger and Hall* (1999) (*where also the three parameters are computed at the same level k*) *is given by*

$$\sigma^2_{_{FH}} := \gamma^2\left(\frac{1-\rho}{\rho}\right)^4.$$

In Figure 2 we provide both a picture and some values of $\sigma_{_{H_1}}/\gamma \equiv 1$, $\sigma_{_{H_2}}/\gamma$, $\sigma_{_{H_3}}/\gamma$ and $\sigma_{_{FH}}/\gamma$, as functions of $|\rho|$.

It is obvious from Figure 2 that, whenever possible, it seems convenient to estimate both β and ρ externally, at a k-value higher than the one used for the estimation of the tail index γ.

Remark 3.3. *More generally, to obtain information on the asymptotic bias of $\overline{H}_{\widehat{\beta},\ \widehat{\rho}}(k)$, $\overline{H}_{\widehat{\beta}_{\widehat{\rho}}(k),\ \widehat{\rho}}(k)$ and $\overline{H}_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\ \widehat{\rho}(k)}(k)$ — or equivalently, of $\overline{\overline{H}}_{\widehat{\beta},\ \widehat{\rho}}(k)$, $\overline{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\ \widehat{\rho}}(k)$ and $\overline{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\ \widehat{\rho}(k)}(k)$ — we need to go further into a third order framework, specifying the rate of convergence in the second order condition in* (1.3). *This is however beyond the scope of this paper.*



$\lvert\rho\rvert$	σ_{H_1}	σ_{H_2}	σ_{H_3}	σ_{FH}
0.1	1.00	11.00	12.19	121.00
0.2	1.00	6.00	7.37	36.00
0.3	1.00	4.33	5.87	18.78
0.4	1.00	3.50	5.19	12.25
0.5	1.00	3.00	4.85	9.00
1.0	1.00	2.00	4.58	4.00
1.5	1.00	1.67	4.96	2.78
2.0	1.00	1.50	5.50	2.25
2.5	1.00	1.40	6.10	1.96
3.0	1.00	1.33	6.74	1.78

Figure 2: "Rulers" of the asymptotic standard deviations, σ_{H_1}, σ_{H_2} and σ_{H_3}, of the estimators under study, together with σ_{FH}, for $\gamma = 1$

4 Finite sample behaviour of the estimators

4.1 Underlying models

In this section we shall consider the following models in Hall's class (1.4):

- the *Fréchet* model, with distribution function (d.f.) $F(x) = \exp(-x^{-1/\gamma})$, $x \geq 0$, $\gamma > 0$, for which $\rho = -1$;
- the *Generalized Pareto* (*GP*) model, with d.f. $F(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, $x \geq 0$, $\gamma > 0$, for which $\rho' = \rho = -\gamma$;
- the *Burr* model, with d.f. $F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}$, $x \geq 0$, $\gamma > 0$, $\rho < 0$;
- the *Student*'s t_ν-model with ν degrees of freedom, with a probability density function (*p.d.f.*)

$$f_{t_\nu}(t) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\;\Gamma(\nu/2))} \left[1 + \frac{t^2}{\nu}\right]^{-(\nu+1)/2}, \; t \in \mathbb{R} \quad (\nu > 0),$$

for which $\gamma = 1/\nu$ and $\rho = -2/\nu$.

4.2 The simulation design

We have here implemented simulation experiments of size 1000×10, essentially to obtain, for the above mentioned models, the distributional behaviour of the new estimators $\overline{H}_{\widehat{\beta},\,\widehat{\rho}}$ and $\overline{\overline{H}}_{\widehat{\beta},\,\widehat{\rho}}$ in (1.13) and (1.14), respectively, based on the estimation of β at the same level $k_1 = \min\left(n-1, [2n/\ln_2 n]\right)$ we have used for the estimation of ρ, again not chosen in any optimal way. We use the notation $\widehat{\beta}_{j1} = \beta_{\widehat{\rho}_j}(k_1)$, $j = 0, 1$, with $\widehat{\rho}_0$, $\widehat{\rho}_1$ and $\beta_{\widehat{\rho}}(k)$ given in (2.1), (2.1) and (2.6), respectively. Similarly to what has been done in Gomes *et al.* (2004b) for the WH-estimator in (1.11), these estimators of ρ and β have been incorporated in the estimators under study, leading to $\overline{H}_{\widehat{\beta}_{01},\,\widehat{\rho}_0}(k)$ and $\overline{\overline{H}}_{\widehat{\beta}_{01},\,\widehat{\rho}_0}(k)$ or to $\overline{H}_{\widehat{\beta}_{11},\,\widehat{\rho}_1}(k)$ and $\overline{\overline{H}}_{\widehat{\beta}_{11},\,\widehat{\rho}_1}(k)$. The simulations show that the tail index estimators $\overline{H}_{\widehat{\beta}_{j1},\,\widehat{\rho}_j}(k)$ and $\overline{\overline{H}}_{\widehat{\beta}_{j1},\,\widehat{\rho}_j}(k)$, j equal to either 0 or 1, according as $|\rho| \le 1$ or $|\rho| > 1$ seem to work reasonably well, as illustrated in the sequel. In the simulation we have also included the Hill estimator in (1.8), the "Weighted Hill" estimator in (1.11) and $\overline{H}_{\widehat{\beta}_{\widehat{\rho}}(k),\,\widehat{\rho}}$. The estimator in (1.12) exhibits a behaviour quite similar to that of $\overline{\overline{H}}$ in (1.14), and was not pictured, for sake of simplicity.

We have also simulated four different indicators. Let us denote generally $\widetilde{H}_n$ any of the estimators in (1.11), (1.13) and (1.14), and let

$$k_{0s}^{H}(n) := \arg\min_{k} MSE_s\left[H_n(k)\right]$$

be the simulated optimal k (in the sense of minimum mean squared error) for the Hill estimator $H_n(k) \equiv H(k)$ in (1.8). The two first indicators are related to the behaviour of the new estimators at Hill's optimal simulated level, i.e.,

$$REFF_n^{\widetilde{H}|H_0} := \sqrt{\frac{MSE_s\left[H_n\left(k_{0s}^{H}(n)\right)\right]}{MSE_s\left[\widetilde{H}_n\left(k_{0s}^{H}(n)\right)\right]}} \tag{4.1}$$

and

$$BRI_n^{\widetilde{H}|H_0} := \left|\frac{E_s\left[H_n\left(k_{0s}^{H}(n)\right) - \gamma\right]}{E_s\left[\widetilde{H}_n\left(k_{0s}^{H}(n)\right) - \gamma\right]}\right|. \tag{4.2}$$

We have also simulated two additional indicators, now related to the comparison of mean squared errors and bias of the new estimators with those of the Hill's estimators, when all the estimators are considered at their optimal levels. Denoting

$$H_{n0} := H_n\left(k_0^H(n)\right) \quad \text{and} \quad \widetilde{H}_{n0} := \widetilde{H}_n\left(k_0^{\widetilde{H}}(n)\right),$$

with the obvious meanig for $k_0^{\bullet}(n)$, the two extra simulated indicators are

$$REFF_{n0}^{\widetilde{H}|H} := \sqrt{\frac{MSE_s\left[H_{n0}\right]}{MSE_s\left[\widetilde{H}_{n0}\right]}}. \tag{4.3}$$

and

$$BRI_{n0}^{\widetilde{H}|H} := \left|\frac{E_s\left[H_{n0}-\gamma\right]}{E_s\left[\widetilde{H}_{n0}-\gamma\right]}\right|. \tag{4.4}$$

Remark 4.1. *Note that an indicator higher than one means a better performance than the Hill estimator. Consequently, the higher these indicators are, the better the new estimators perform, comparatively to the Hill estimator.*

Remark 4.2. *Note also that whereas we have appropriate techniques to deal with the estimation of the optimal level for Hill's estimator, in the sense of minimum mean squared error, we do not have yet equivalent techniques for the reduced bias' estimators. Consequently, the indicators in* (4.3) *and* (4.4) *are not useful in practice, but they give us an indication of the potentialities of this type of estimators.*

4.3 Mean values and mean squared error patterns

In Figures from 3 till 9, we picture for the different underlying models herewith considered, and a sample of size $n = 1000$, the mean values ($E[\bullet]$) and the mean squared errors ($MSE[\bullet]$) of the Hill estimator H, together with $\overline{H}_{\widehat{\beta}_{j1},\, \widehat{\rho}_j}$,

$\overline{\overline{H}}_{\widehat{\beta}_{j1},\ \widehat{\rho}_j}$, $WH_{\widehat{\beta}_{j1},\ \widehat{\rho}_j}$ and $\overline{H}_{\widehat{\beta}_j(k),\ \widehat{\rho}_j}$, $j = 0$ or $j = 1$, according as $|\rho| \leq 1$ or $|\rho| > 1$. For comparison, we also picture the analogue behaviour of the r.v. $\overline{H}_{\beta,\ \rho}$ for the models where there is a big discrepancy between the behaviour of the estimators and of the r.v.'s. Such a discrepancy suggests that some improvement in the estimation of second order parameters β and ρ is still welcome.

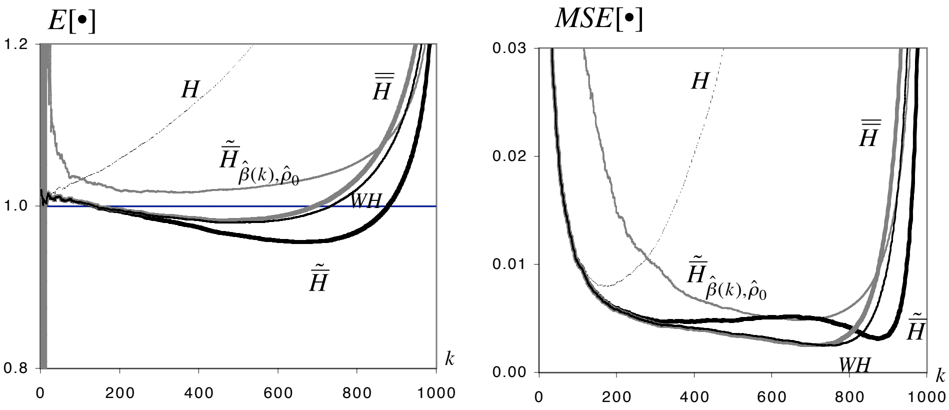


Figure 3: Underlying *Fréchet* parent with $\gamma = 1$ ($\rho = -1$).

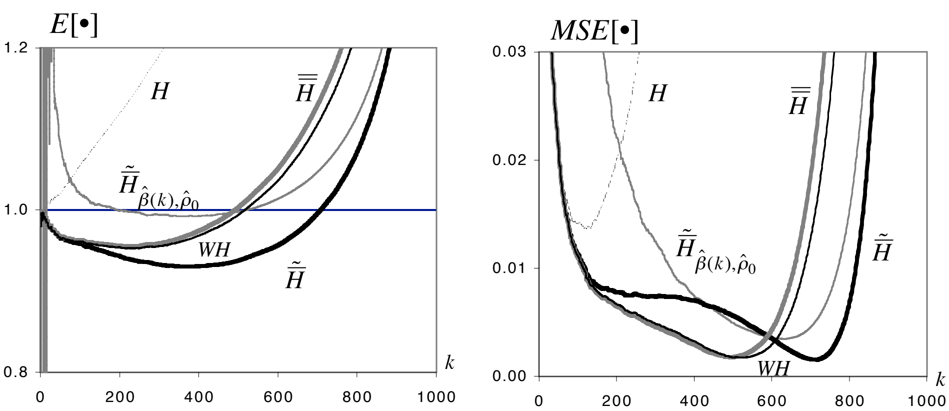


Figure 4: Underlying *Burr* parent with $\gamma = 1$ and $\rho = -1$.

Remark 4.3. *For a Burr model and for any of the estimators considered, $BIAS/\gamma$ and MSE/γ^2 are independent of γ, for every ρ. And we have not pictured the mean value and mean squared error patterns of the estimators for*

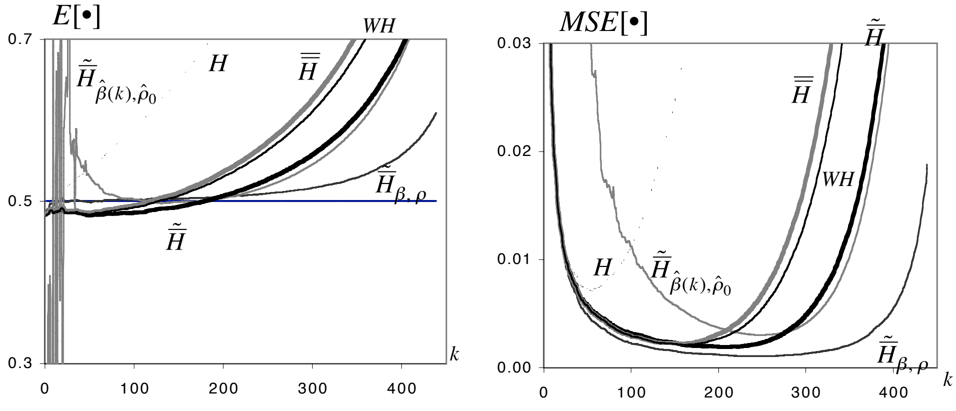


Figure 5: Underlying *Student* parent with $\nu = 2$ degrees of freedom ($\gamma = 0.5$ and $\rho = -1$).

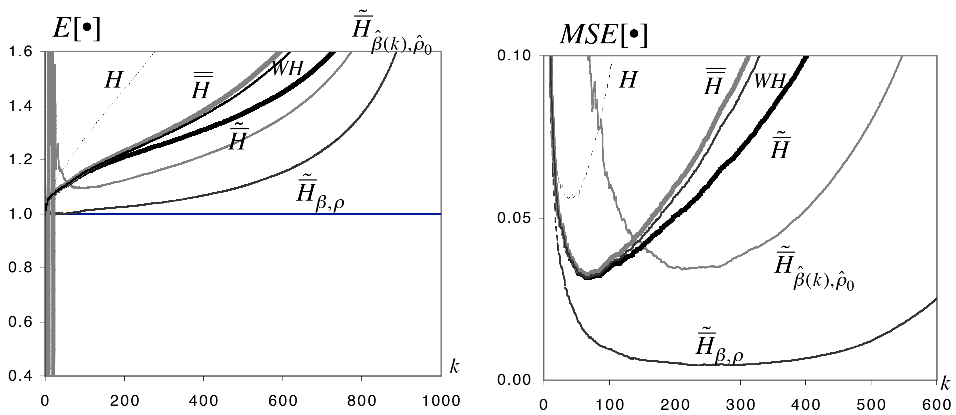


Figure 6: Underlying *Burr* parent with $\gamma = 1$ and $\rho = -0.5$.

a GP underlying model because for all the estimators considered,

$$\mathbb{E}|GP(\gamma) = \gamma \times \mathbb{E}|BURR(\gamma = 1, \rho = -\gamma)$$

and

$$MSE|GP(\gamma) = \gamma^2 \times MSE|BURR(\gamma, \rho = -\gamma).$$

Remark 4.4. *We may further draw the following specific comments:*

- *For a Fréchet model (with $\rho = -1$) (Figure 3) the bias of $\overline{\overline{H}}$ is the smallest one, being the one of $\overline{H}$ the largest one, for small to moderate values of*

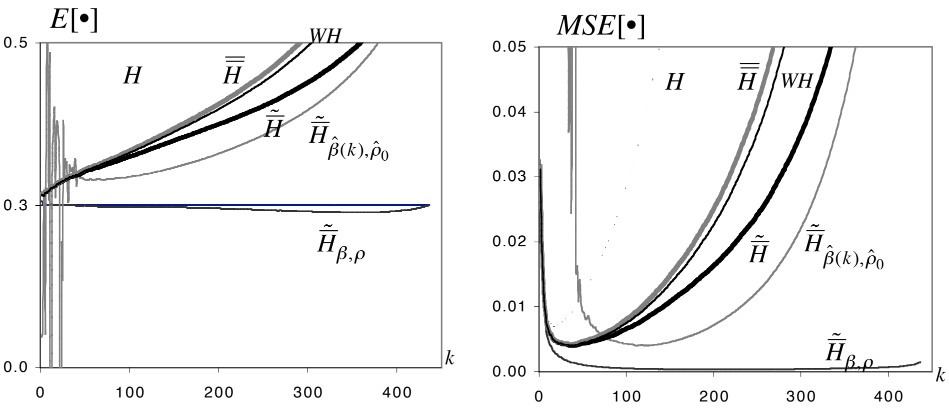


Figure 7: Underlying *Student* parent with $\nu = 4$ degrees of freedom ($\gamma = 0.25$ and $\rho = -0.5$).

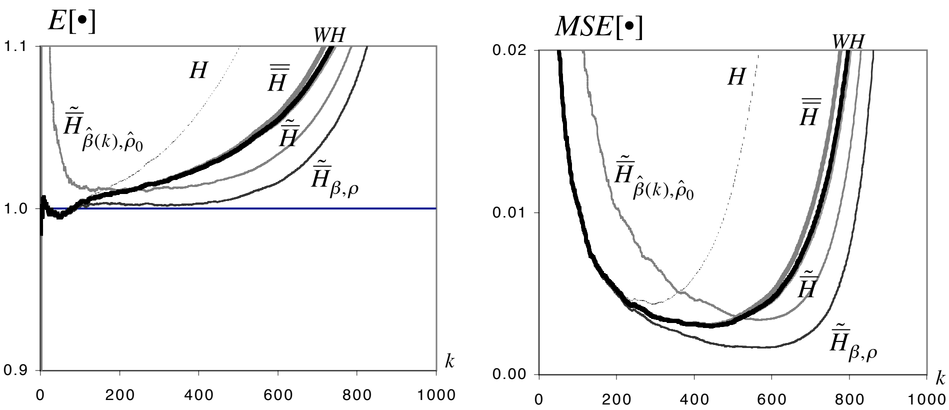


Figure 8: Underlying *Burr* parent with $\gamma = 1$ and $\rho = -2$.

k. All the three reduced bias' statistics overestimate the true value of γ for very small as well as very large values of k, whereas, for moderate values of k, they all underestimate γ.

- *For an underlying Burr parent with $\rho = -1$ (Figure 4), the three reduced bias' statistics are negatively biased for small values of k — again, as for a Fréchet underlying parent, the $\overline{\overline{H}}$-statistic exhibits then the smallest bias and $\overline{H}$ the largest one. The $\overline{H}$ statistic is the best one regarding MSE at the optimal level, but the WH-statistic is the one with the smallest mean squared error for not too large values of k, followed by $\overline{\overline{H}}$. Quite similar*

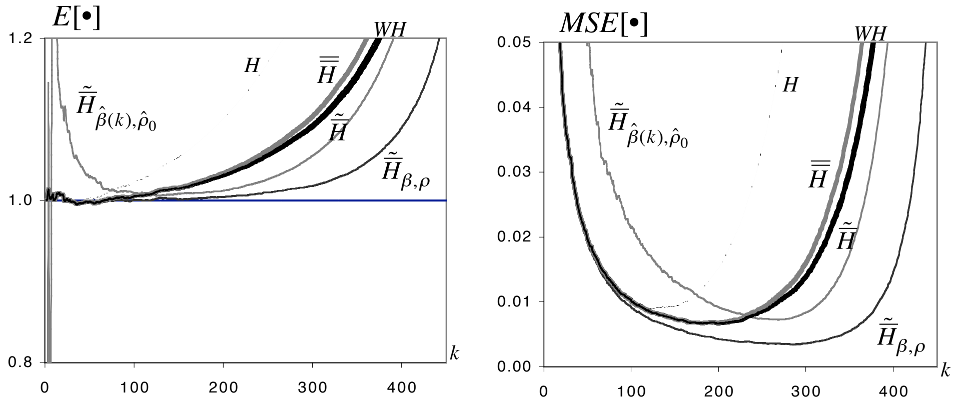


Figure 9: Underlying *Student* parent with with $\nu = 1$ degrees of freedom ($\gamma = 1$ and $\rho = -2$).

results may be drawn for a Student model with $\rho = -1$ (Figure 5), but then the mean squared error of $\overline{H}$ is smaller than that of WH, which on its turn is smaller than that of $\overline{\overline{H}}$, for all values of k.

- *For values of $\rho > -1$ (Figures 6 and 7), the three reduced bias' statistics are positively biased for all k. The $\overline{H}$-statistic is better than the WH-statistic, which on its turn behaves slightly better than the $\overline{\overline{H}}$-statistic, both regarding bias and mean squared error.*
- *For $\rho < -1$ (Figures 8 and 9), we need to use $\widehat{\rho}_1$ (instead of $\widehat{\rho}_0$). In all the simulated cases the $\overline{H}$ and the WH-statistics are the best ones and exhibit quite similar properties, but they are not a long way from the $\overline{\overline{H}}$-statistic.*

Remark 4.5. *For a Student model with ν degrees of freedom (Figures 5, 7 and 9), and whenever we assume β and ρ known, the most stable sample path around the target value γ is achieved by the statistic $\overline{H}_{(\beta,\ \rho)}$, presented in the figures. And such a fact leads this statistic to have the smallest mean squared error, followed by the $\overline{\overline{H}}$ and next the WH statistics, for all values of ν. If we need to estimate β and ρ, the $\overline{H}$-statistic is the one with the smallest mean squared error at the optimal level, also for every ν. Next comes the WH-statistic, quite close to the $\overline{\overline{H}}$-statistic when $\nu < 2$, i.e., when $\rho < -1$.*

4.4 Relative efficiency and bias reduction indicators

In Table 1 we present the $REFF$ indicators, in (4.1) and (4.3), and the BRI indicators in (4.2) and (4.4). For each model, the first, second and third rows are related to the $\overline{H}$-estimator in (1.13), the $\overline{\overline{H}}$-estimator in (1.14) and the WH-estimator in (1.11), respectively. Each entry has two numbers: the first one is either the indicator in (4.1) or in (4.2) and the second one is either the indicator in (4.3) or in (4.4).

Table 1: $REFF$ and BRI indicators

$REFF$ indicators			BRI indicators		
n			n		
200	500	1000	200	500	1000
Fréchet parent: $\rho = -\mathbf{1}$, $\gamma = \mathbf{1}$					
1.06/1.39	1.10/1.53	1.12/1.67	4.18/45.93	6.33/31.87	9.99/69.15
1.08/1.52	1.12/1.67	1.12/1.85	5.85/5.56	8.92/47.70	14.28/22.89
1.07/1.53	1.11/1.69	1.12/1.86	4.56/25.40	7.16/13.12	11.50/10.22
Burr parent: $\rho = -\mathbf{0.5}$, $\gamma = \mathbf{1}$					
1.19/1.39	1.26/1.35	1.22/1.33	1.73/1.25	1.73/1.22	1.64/1.24
1.18/1.35	1.25/1.32	1.22/1.31	1.69/1.30	1.69/1.20	1.62/1.22
1.20/1.39	1.26/1.35	1.23/1.33	1.83/1.32	1.76/1.23	1.67/1.24
Burr parent: $\rho = -\mathbf{1}$, $\gamma = \mathbf{1}$					
1.18/2.27	1.22/2.63	1.21/2.94	2.14/306.29	1.74/22.55	1.55/28.80
1.20/2.02	1.24/2.34	1.24/2.61	2.67/9.35	2.07/24.51	1.79/13.63
1.19/2.09	1.23/2.42	1.23/2.69	2.25/11.26	1.91/10.31	1.70/30.21
Burr parent: $\rho = -\mathbf{2}$, $\gamma = \mathbf{1}$					
1.05/1.17	1.08/1.18	1.10/1.21	2.60/1.09	2.25/1.10	2.07/1.09
1.05/1.16	1.08/1.17	1.10/1.20	2.54/1.10	2.22/1.10	2.05/1.06
1.05/1.18	1.08/1.18	1.11/1.21	2.75/1.08	2.29/1.12	2.09/1.10
Student parent: $\rho = -\mathbf{0.5}$, $\gamma = \mathbf{0.25}$					
1.32/1.41	1.24/1.34	1.21/1.32	2.17/1.49	1.87/1.39	1.68/1.26
1.30/1.36	1.23/1.31	1.20/1.30	1.99/1.41	1.80/1.42	1.64/1.26
1.35/1.42	1.25/1.35	1.22/1.32	2.27/1.54	1.90/1.47	1.73/1.30
Student parent: $\rho = -\mathbf{1}$, $\gamma = \mathbf{0.5}$					
1.03/1.51	1.02/1.75	1.14/1.97	2.77/7.88	2.03/7.15	1.25/8.71
1.08/1.51	1.02/1.65	1.16/1.82	4.58/2.84	2.26/3.31	1.24/4.65
1.07/1.56	1.02/1.70	1.15/1.86	3.17/3.40	1.68/3.94	1.25/4.85
Student parent: $\rho = -\mathbf{2}$, $\gamma = \mathbf{1}$					
0.91/1.12	1.06/1.16	1.04/1.15	1.71/15.65	11.52/1.71	3.64/1.47
0.88/1.18	1.07/1.15	1.04/1.15	2.43/8.52	8.83/1.51	3.50/1.45
0.87/1.18	1.07/1.16	1.04/1.16	2.00/7.00	13.80/1.67	3.81/1.48

5 A case-study and overall conclusions

5.1 Log-exchange rates of Euro against USA Dollar

When analysing heavy-tailed data, quite common in financial time series, one never knows how much the underlying model differs from a strict Pareto model. And this is the unique situation where the Hill estimator is "perfect". Otherwise, all depends on the specificity of the underlying heavy-tailed model and on the practitioner's objectives. If we want to use (or have only access to) a very small number of top o.s., the Hill estimator has been considered the most adequate one, among the tail index estimators so far available in the literature. This is no longer true now: these new estimators are similar to Hill's estimator from small to moderate values of k, being much better than the Hill estimator, when we consider larger values of k, although intermediate.

After taking a decision on the estimate of ρ, and assuming that $|\rho| \le 1$, a situation which seems to appear often in practice, we should simultaneously picture the sample path of a few tail index estimators, with different specificities. On the basis of those sample paths we may then get, in a more appropriate way, a sensible estimate of the tail index γ, like we shall see later on, in the application provided. And the new estimators herewith presented should be taken as the adequate substitutes of Hill's estimator.

We shall now consider an illustration of the performance of the above mentioned estimators, through the analysis of the Euro-USA Dollar daily exchange rates from January 4, 1999 till December 15, 2004. In Figure 10, working with the $n^+ = 748$ positive log-returns, we present the sample path of the $\widehat{\rho}_\tau$ estimates in (2.1) (*left*), as function of k, for $\tau = 0$ and $\tau = 1$. Note that the sample paths of the ρ-estimates associated to $\tau = 0$ and $\tau = 1$ lead us to choose, on the basis of any stability criterion for large k, the estimate associated to $\tau = 0$. From previous experience with this type of estimates, we conclude that the underlying ρ-value is larger or equal to -1, and the consideration of $\tau = 0$ is then advisable. The estimate of ρ is in this case $\widehat{\rho}_0 = -0.7$. In this same figure (*right*), we also present the sample paths of the classical Hill estimator, H, of $\overline{H}_{\widehat{\beta}_{01},\, \widehat{\rho}_0}$ and of $WH_{\widehat{\beta}_{01},\, \widehat{\rho}_0}$.

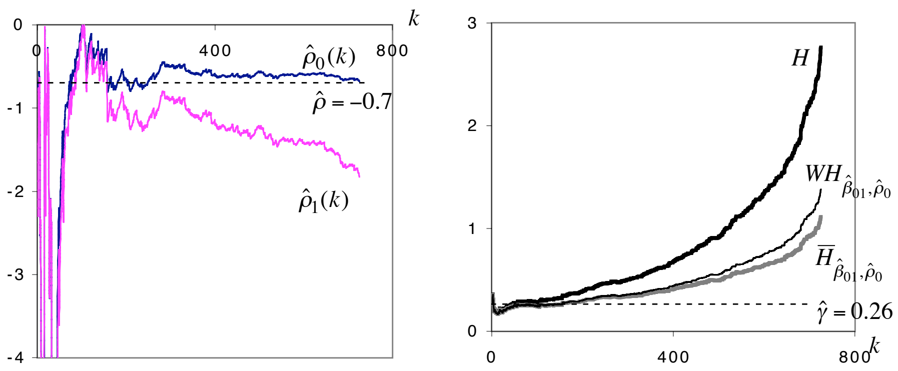


Figure 10: Estimates of the second order parameter ρ (*left*) and of the tail index γ (*right*) for the Daily Log-Returns of the Euro-UK Pound.

Regarding the tail index estimation, note that the Hill estimator exhibits here a relevant bias. We are thus a long way from a strict Pareto underlying model. The other estimators, which are "asymptotically unbiased" up to the second order, reveal a smaller bias, and enable us to take a decision upon the estimate of γ to be used, with the help of any stability criterion or any heuristic procedure, like a largest run method as the one described in the sequel, and already suggested in Gomes *et al.* (2003).

Let us consider a set of reduced bias' tail index estimates $\widehat{\gamma}_i(k)$, $1 \le k < n$, $i \in \mathcal{I}$, based on the observed sample of size n. Consider those estimates with a small number r of decimal figures, and denote them $\widehat{\gamma}_{i|r}(k)$.

1. For any value $i \in \mathcal{I}$ and for any possible value a in the domain of $\widehat{\gamma}_{i|r}(k)$, consider the largest run associated with a, i.e., $R_i(a)$, the maximum num-

ber of consecutive k values such that $\widehat{\gamma}_{i|r}(k) = a$.

2. Compute $a_i^{M} := \arg\max_a R_i(a)$.
3. Consider as a data-driven estimate of the tail index γ, $\widehat{\gamma} = a_{i_0}^{M}$ with $i_0 := \arg\max_i a_i^{M}$.

Here, if we consider the tail index estimates with one decimal figure, the largest run is achieved by the sample path of the $\overline{H}$-estimator in (1.13). Such a largest run has a size equal to 266, for k between 68 and 333, and is associated to $\overline{H} = 0.3$. For the WH-estimate, and also with one decimal figure, we would also get a tail index estimate equal to 0.3, but with a run of size 235 ($60 \leq k \leq 294$). With this same criterion, the Hill estimator would provide an estimate also equal to 0.3, with a run of size 126 ($43 \leq k \leq 168$). According to the previous heuristic procedure we would thus be led to the choice of the $\overline{H}$ estimator, computed at any of the levels from $k = 68$ till $k = 333$, all providing the same estimate $\widehat{\gamma} = 0.3$. Should we consider this same criterion, but the estimates with two decimal figures, would we be led to an estimate equal to 0.26, provided by any of the reduced bias' estimators, the value $\widehat{\gamma}$ pictured in Figure 10.

5.2 Some overall conclusions

- Generally, we may say that there is practically no difference between the estimators, WH, $\widetilde{H}$, $\overline{H}$ and $\overline{\overline{H}}$ in (1.11), (1.12) (1.13) and (1.14), respectively. Anyway, whenever confronted with real data, the drawing of a few sample paths may help us in the choice of the most adequate estimate of the tail index γ.

- The $\overline{H}_{\widehat{\beta}_{\widehat{\rho}}(k),\, \widehat{\rho}}$ statistic may perhaps help us in the choice of the optimal sample fraction of Hill's estimator, and for some of the models exhibits sample paths more stable around the target value γ for a wider region of k-values. This is however a topic which deserves further investigation, being outside the scope of the present paper.

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

References

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

[2] Caeiro, F. and Gomes, M. I. (2004). *A new class of estimators of a "scale" second order parameter*. Notas e Comunicações CEAUL 20/2004. Submitted.

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

[4] 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.

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

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

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

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

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

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

[11] 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.

[12] 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.

[13] 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.

[14] Gomes, M. I. and Pestana (2004). *A simple second order reduced bias' tail index estimator.* Submitted.

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

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

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

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

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