How can non-invariant statistics work in our benefit in the semi-parametric estimation of parameters of rare events*

M. Ivette Gomes
and
Orlando Oliveira
D.E.I.O. and C.E.A.U.L., Universidade de Lisboa (F.C.U.L.)

Abstract. In this paper, and in a context of regularly varying tails, we suggest a simple generalization of the classical Hill estimator of a positive tail index γ. Such an estimator is merely the Hill estimator associated to artificially shifted data. The shift a imposed to the data is the *tuning* parameter of the statistical procedure of estimation. Such a tuning parameter enables us, in a great diversity of situations, to reduce the main component of the bias of Hill's estimator, getting thus estimates with stable sample paths around the target value γ, and with small mean squared error.

AMS 1991 subject classification. Primary 62G05, 62E25, 62E20; Secondary 62F35.
Keywords and phrases. *Statistical Theory of Extremes, Semi-parametric estimation.*

1 Introduction and preliminaries

Let $X_1, X_2, \ldots, X_n$ be independent random variables (r.v.'s) with common distribution function (d.f.) $F(\cdot)$, with a heavy upper tail, i.e. for large x,

$$1 - F(x) = x^{-1/\gamma} L(x), \tag{1.1}$$

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

$$G_\gamma(x) := exp\left\{-(1+\gamma x)^{-1/\gamma}\right\},\ 1+\gamma x > 0, \tag{1.2}$$

with $\gamma > 0$.

Recall that, for $\gamma > 0$, $F \in D(G_\gamma)$ iff $1 - F \in RV_{-1/\gamma}$ iff $U \in RV_\gamma$, where $U(t) := F^{\leftarrow}(1 - 1/t),\ t > 1$. The notation RV_α stands for the class of

---

[*] Research partially supported by FCT / POCTI / FEDER.

regularly varying functions at infinity with index of regular variation equal to α, i.e., positive measurable functions $g(\cdot)$ such that $\lim\limits_{t\to\infty} g(tx)/g(t) = x^{\alpha}$, for all $x > 0$, and the notation $F^{\leftarrow}(\cdot)$ is used for the generalized inverse function of F, i.e., $F^{\leftarrow}(t) = \inf\{x : F(x) \geq t\}$. We shall assume that there exists a function $A(\cdot)$, which measures the rate of convergence of $U(tx)/U(t)$ towards x^{γ}, and is a function of constant sign near infinity, such that

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

for every $x > 0$, where ρ (≤ 0) is a *second order parameter*. The limit function in (1.3) must be of the stated form, and $|A(t)| \in RV_{\rho}$ (Geluk and de Haan, 1987).

The functional expression of the classical Hill estimator for $\gamma > 0$ (Hill, 1975) is

$$\gamma_n^H(k) := \frac{1}{k} \sum_{i=1}^{k} \left[lnX_{n-i+1:n} - lnX_{n-k:n} \right], \tag{1.4}$$

where $X_{i:n}$ denotes the i-th ascending order statistic (o.s.), $1 \leq i \leq n$, associated to the sample $\underline{X}_n = (X_1, X_2, \ldots, X_n)$, and for which we have, provided k is intermediate, i.e.

$$k = k_n \to \infty, \quad k/n \to 0, \text{ as } n \to \infty, \tag{1.5}$$

the validity of the distributional representation (de Haan and Peng, 1998),

$$\gamma_n^H(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} P_n^H + \frac{1}{1-\rho} A(n/k)(1 + o_p(1)) \tag{1.6}$$

where P_n^H is asymptotically standard normal. Such a result comes from the fact that $\gamma_n^H(k) \overset{d}{=} \frac{1}{k}\sum\limits_{i=1}^{k} \ln \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})}$, where $Y_{n-i+1:n}$, $1 \leq i \leq n$, are the o.s. associated with a sample of n independent, identically distributed (i.i.d.) r.v.'s Y_i, $1 \leq i \leq n$, from a standard Pareto d.f., i.e., $F_Y(y) = 1 - 1/y$, $y \geq 1$. Then, the use of the second order condition in (1.3) with $t = Y_{n-k:n}$ and $x = Y_{n-i+1:n}/Y_{n-k:n} \overset{d}{=} Y_{k-i+1:k}$, $1 \leq i \leq k$, enables us to identify the first two terms of the distributional representation of $\gamma_n^H(k)$, which come from $\frac{1}{k}\sum\limits_{i=1}^{k} \gamma \ln Y_{k-i+1:k} \overset{d}{=} \gamma\, \frac{1}{k}\sum\limits_{i=1}^{k} E_{k-i+1:k} \overset{d}{=} \gamma\, \frac{1}{k}\sum\limits_{i=1}^{k} E_i$, where $\{E_i\}_{i\geq 1}$ denotes a sequence of i.i.d. standard exponential r.v.'s. The third term of the representation (1.6) comes from the weak law of large numbers, i.e., from the fact that $\frac{1}{k}\sum\limits_{i=1}^{k} \frac{Y_{k-i+1:k}^{\rho}-1}{\rho} \overset{d}{=} \frac{1}{k}\sum\limits_{i=1}^{k} \frac{e^{\rho E_i}-1}{\rho}$ converges in probability, as $k \to \infty$, towards $E\left[\frac{e^{\rho E}-1}{\rho}\right] = \frac{1}{1-\rho}$, together with the fact that $(k/n)Y_{n-k:n} \underset{n\to\infty}{\xrightarrow{p}} 1$.

The Hill estimator has a nice explicit expression, and it is appealing from a theoretical point of view. The main problem with this semi-parametric estimator is

- its high variance for small values of k, i.e. high thresholds, and
- its high bias for large k, i.e., low thresholds.

And this *trade-off* between bias and variance leads us to an estimator with very sharp Mean Squared Error (MSE) patterns, and a difficult problem of choice of the best threshold to be considered. Moreover the reflection of this on the estimates' sample paths is often catastrophic!

This is the main reason why for the last few years several authors have been doing research in the reduction of bias of such an estimator, in the hope to get *bath-tube* MSE patterns, as functions of k (Drees (1996), Peng (1998), Feuerverger and Hall (1999), Beirlant et al. (1999), Gomes et al. (2000), Caeiro and Gomes (2002), Gomes (2001), among others). This would make the choice of k almost irrelevant, and would give rise to estimates with stable sample paths close to the target value, whatever such a target is, here the tail index γ. It is then expected to be led to more efficient estimators, provided one might use extreme-value data relatively deep into the sample. So all depends on the amount of data available and/or on the price to pay for it.

But one of the most recent criticisms against most of the semi-parametric estimators of the tail index, among which the Hill estimator is included, is the fact that most of them are non-invariant for location. However, although this may be controversial we would dare to say that such non-invariance may work in our benefit. Maybe even that, instead of moving from *scale invariant/location non-invariant* estimators towards *location/scale invariant* estimators, should we move the other way round, towards *location/scale non-invariant* estimators if our main objective is the reduction of MSE. The development of this paper will provide a background for such a controversial remark.

There is indeed a great diversity of sample paths for $\gamma_n^H(k)$, as a function of k, when we shift the original data. The main objective of this paper is to work with such a diversity in our benefit, considering the following generalization of the Hill estimator,

$$\gamma_n^{(a)}(k) := \frac{1}{k}\sum_{i=1}^{k} \left[ln(X_{n-i+1:n}+a) - ln(X_{n-k:n}+a)\right], \quad a \in \mathbb{R}, \tag{1.7}$$

with a *tuning* parameter a, which we may choose adequately. Notice that the Hill estimator in (1.4) is merely the estimator in (1.7) for $a = 0$, i.e., $\gamma_n^H(k) \equiv \gamma_n^{(0)}(k)$.

We shall derive in section 2 the asymptotic behaviour of this class of estimators, which is quite similar to that of the Hill estimator, together with a discussion on how to play with non-invariance for location in our benefit, particularly whenever $\gamma+\rho=0$. In section 3 we shall obtain through simulation the finite sample distributional behaviour of this class of estimators for models where $\gamma+\rho\neq 0$. In section 4 we shall deal with an interesting special case, that of Generalized Pareto parents, and suggest a few possible estimators of the tail index γ, whose study is beyond the scope of this paper. The estimator in (1.7), but with a replaced by $1-X_{1:n}$ is studied in section 5 for models with a finite left endpoint, and compared (in terms of mean value and mean squared error) with a location invariant estimator recently proposed in the literature by Fraga Alves (2001). Finally, some overall conclusions are drawn.

The models simulated are the following models in Hall's class of distributions (Hall, 1982; Hall and Welsh, 1985),

1. the *Fréchet* model, $F(x)=\exp(-x^{-1/\gamma}),\ x\geq 0$, for which $\rho=-1$;
2. the *Burr* model, $F(x)=1-(1+x^{-\rho/\gamma})^{1/\rho},\ x\geq 0,\ \gamma>0,\ \rho<0$, with $\gamma=1$;
3. the *GP* model, $F(x)=1-(1+\gamma x)^{-1/\gamma},\ x\geq 0,\ \gamma>0$ (for which $\rho=-\gamma$);

and a model outside Hall's class,

4. the *Out-Hall* model, with a quantile function $F^{\leftarrow}(1-t)=t^{-1}e^{-2t(\ln t-1)}$, for all $0<t\leq 1$, for which we have $\rho=-1$.

The overall simulation results were based on a multi-sample simulation of size 5000 × 10 in order to guarantee small standard errors (not presented in the table provided, but available from the authors) for the simulated characteristics of an estimator $\gamma_n^\bullet(k)$, the Mean Value ($E^\bullet$), the Mean Squared Error ($MSE^\bullet$), the Optimal Sample Fraction ($OSF^\bullet$), defined as

$$OSF^\bullet := k_0^\bullet/n, \quad k_0^\bullet := \arg\min_k MSE^\bullet(k) \tag{1.8}$$

and the Relative Efficiency ($REFF^\bullet$), defined as

$$REFF^\bullet = REFF[\gamma_{n0}^\bullet] = \sqrt{\frac{MSE_s^{(1)}\left[\gamma_n^{(1)}(k_{0s}^{(1)}(n))\right]}{MSE_s^\bullet\left[\gamma_n^\bullet(k_{0s}^\bullet(n))\right]}}, \tag{1.9}$$

with $\gamma_{n0}^\bullet=\gamma_n^\bullet(k_{0s}^\bullet(n))$, and where $MSE_s^\bullet$ denotes the simulated $MSE^\bullet$ of the estimator at its simulated optimal level. The simulator of for instance $k_0^\bullet(n)$, denoted by $k_{0s}^\bullet(n)$, is $\widehat{E}_{10}[\overline{k}_0^\bullet(n)]$, the average of 10 independent replicates of $\overline{k}_0^\bullet(n)=\arg\min_k \sum_{j=1}^{5000}\left(\gamma_{nj}^\bullet(k)-\gamma\right)^2$. All Figures with mean values and MSE's are based on the first 5000 runs.

2 The asymptotic behaviour of the estimators — how far is location invariance essential?

From the expression of $\gamma_n^{(a)}(k)$ in (1.7) it is obvious that, just like for the Hill estimator, we have consistency for the estimation of γ, for every fixed a, whenever the classical Hill estimator is consistent. We shall thus assume that (1.5) holds, so that $\gamma_n^{(a)}(k)$ is consistent for the estimation of γ, for every $a \in \mathbb{R}$. Consequently, the class of estimators in (1.7) is asymptotically location/scale invariant for the estimation of γ.

Let X be our original parent, and put $Y = X + a$. Then $U_Y(t) = a + U_X(t)$, and consequently

$$\begin{aligned}
\frac{U_Y(tx)}{U_Y(t)} &= \frac{a + U_X(tx)}{a + U_X(t)} = \frac{U_X(tx)}{U_X(t)}\ \frac{1 + a/U_X(tx)}{1 + a/U_X(t)} \\
&= \frac{U_X(tx)}{U_X(t)} \left\{ 1 + \frac{a}{U(t)} \left(\frac{U_X(t)}{U_X(tx)} - 1 \right) (1 + o(1)) \right\} \\
&= x^{\gamma} \left\{ 1 + \frac{x^{\rho} - 1}{\rho} A(t)(1 + o(1)) \right\} \left\{ 1 - \frac{x^{-\gamma} - 1}{-\gamma} \frac{a\gamma}{U(t)} (1 + o(1)) \right\} \\
&\qquad = x^{\gamma} \left\{ 1 + \frac{x^{\rho} - 1}{\rho} A(t) - \frac{x^{-\gamma} - 1}{-\gamma} \frac{a\gamma}{U(t)} + o(A(t)) + o(1/U(t)) \right\}.
\end{aligned}$$

Looking at the previous formula, and thinking on the fact that $A \in RV_{\rho}$ and $U \in RV_{\gamma}$, the bias component of the new estimator $\gamma_n^{(a)}(k)$ depends on the sign of $\rho + \gamma$. If $\rho + \gamma > 0$, $A(t)$ dominates and the dominant component of the bias remains the same. If $\rho + \gamma \le 0$ there is a change in the bias behaviour of the estimator in (1.7). More precisely, whereas we may define, for intermediate k,

$$Bias_{\infty}\left(\gamma_n^H(k)\right) := \frac{1}{1 - \rho}\ A(n/k), \tag{2.1}$$

we shall have, whenever $a \neq 0$, and for intermediate k,

$$Bias_{\infty}\left(\gamma_n^{(a)}(k)\right) := \begin{cases} \frac{1}{1-\rho}\ A(n/k) & \text{if} \quad \rho > -\gamma \\ \frac{1}{1-\rho} \left(A(n/k) + \frac{\rho a}{U(n/k)} \right) & \text{if} \quad \rho = -\gamma \\ -\frac{1}{1+\gamma} \frac{\gamma a}{U(n/k)} & \text{if} \quad \rho < -\gamma \end{cases}. \tag{2.2}$$

If we put $A^*(k, n; \gamma, \rho) \equiv Bias_{\infty}\left(\gamma_n^{(a)}(k)\right)$, we have for intermediate k the validity of the distributional representation

$$\gamma_n^{(a)}(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} P_n^{(a)} + A^*(k, n; \gamma, \rho)(1 + o_p(1)), \tag{2.3}$$

where $P_n^{(a)}$ is asymptotically standard normal.

In previous papers we have been trying to find alternatives to the Hill estimator, which perform better than this classical semi-parametric estimator of a heavy tail index, either asymptotically or for finite samples, for a wide range of k-values or at least at the optimal k. We are going to show here that we may partially play with the non-invariance for location of the Hill estimator in our benefit. Indeed, if we think on the class of estimators in (1.7), where the tuning parameter a is merely a shift in the data, the plot of $(k,\ \gamma_n^{(a)}(k))$, $k = 1, 2, \cdots, n-1$, for a few values of a, enables us often, to identify easily (by means of any adequate stability criterion) the estimate of γ.

In Figure 1 we present the simulated mean values (which provide a smooth version of the sample paths) of $\gamma_n^{(a)}(k)$ for $a = 0,\ 0.5,\ 1,$ and 2, and for samples of size $n = 1000$ from a Fréchet model with $\gamma = 1$. The location $a = 0.5$ seems to be here playing a "magic" role.

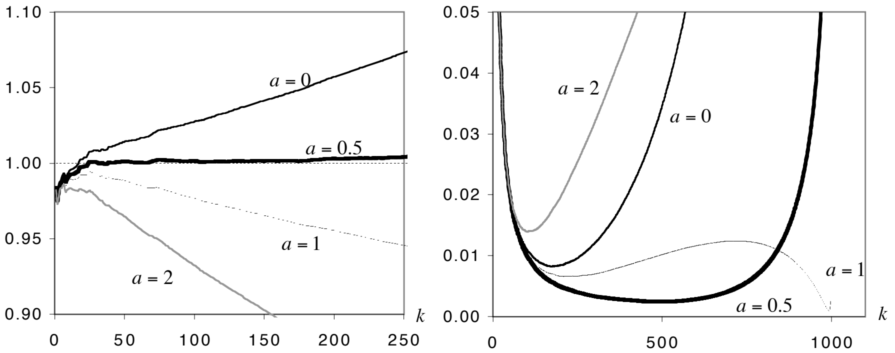

Figure 1: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^{(a)}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Fréchet* parent with $\gamma = 1$ ($\rho = -1$).

We next consider Figure 2, equivalent to Figure 1, but for a Burr model with $\rho = -1$, $\gamma = 1$. Then, the magic number seems to be $a = 1$. And, the"magic" value $a = -9$ appears for *Out-Hall* parents, as illustrated in Figure 3.

Is there any reason for this? Intuitively we would say that the bias of the Hill estimator in (1.4) begins increasing drastically whenever $-\ln(X_{n-k:n})$ becomes positive, i.e., whenever $X_{n-k:n} < 1$. This means that we should plot the Hill estimator only for $k \le k_1 = \arg\max_k\{k : X_{n-k:n} < 1\}$, and zoom that region. In a Fréchet model or in any other model attracted to an EV_γ, $\gamma > 0$, if we shift adequately the data from X to $X + a$, we may widden the stability region of the estimate plot, around the target unknown value γ, in the sense that the value k_1 increases. In fact, if we have all data $X + a \ge 1$, i.e., $a \ge 1 - x_{1:n}$, then $k_1 = n - 1$. In practice, the best choice seems then to be a

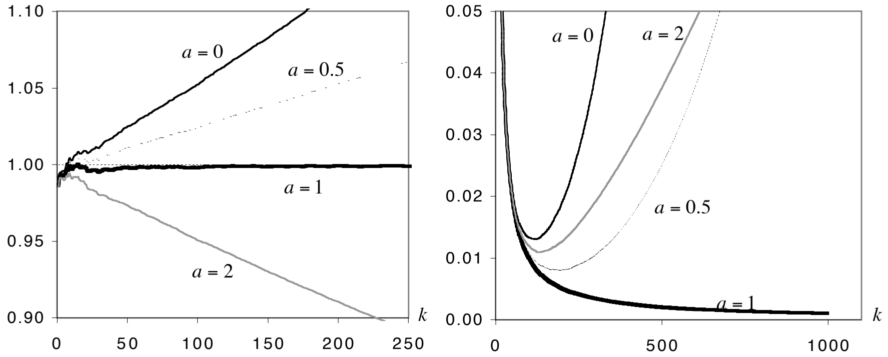

Figure 2: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^{(a)}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -1$.

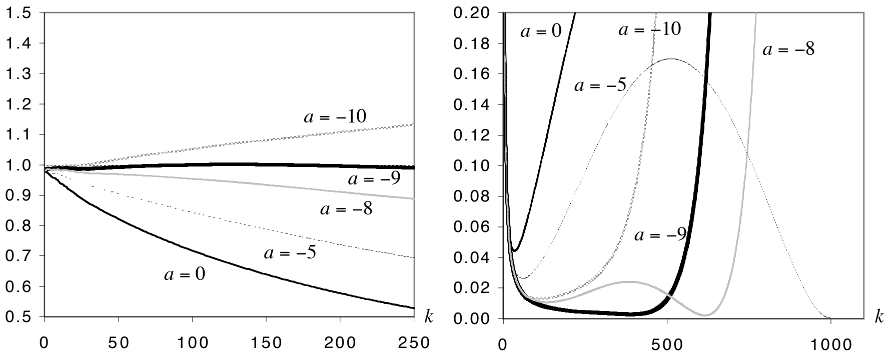

Figure 3: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^{(a)}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Out-Hall* parent with $\gamma = 1$ ($\rho = -1$).

value close to $a = 1 - x_{1:n}$. This does not mean that we may in general replace a by $1 - X_{1:n}$ in (1.7), and consider $\gamma_n^{(1-X_{1:n})}(k)$, which may eventually not to be even consistent for the estimation of γ, in the general domain of attraction of an EV_γ. However $\gamma_n^{(1-X_{1:n})}(k)$ is obviously consistent whenever we work with a class of heavy tail models with a finite left endpoint, and an illustration of its simulated properties is performed in section 5 of this paper, as stated before.

We shall here try to find a theoretical justification for these magic numbers $a = 0.5$ and $a = 1$ which appear for *Fréchet* and *Burr* parents with $\rho = -1$ (and $\gamma = 1$). From a theoretical point of view, let us assume we are working in Hall's class of distributions, where

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

Then, regular variation theory enables us to obtain the asymptotic inverse of F, and we have

$$U(t) \sim (Ct)^{\gamma} \left(1 + \gamma D (Ct)^{\rho} (1 + o(1))\right), \text{ as } t \to \infty,$$

and

$$A(t) \sim \gamma \rho D (Ct)^{\rho}, \text{ as } t \to \infty.$$

Whenever $\rho = -\gamma$,

$$\frac{1}{U(t)} \sim -\frac{A(t)}{\rho^2 D}.$$

If we look at (2.2) we see that the dominant component of asymptotic bias is null whenever $a = \rho D$. For the Fréchet model, which belongs to Hall's class, we have $C = 1$, $D = -1/2$ and $\rho = -1$. Then, for $\gamma = 1$, $a = 0.5$ enables us to remove the main component of asymptotic bias. For the Burr models, also in Hall's class, $C = 1$ and $D = 1/\rho$. Consequently, for every $\rho < 0$ and $\gamma = -\rho$, the main component of asymptotic bias of $\gamma_n^{(a)}(k)$ is equal to 0 whenever $a = 1$.

3 Simulation studies for models whith $\gamma + \rho \neq 0$

3.1 The case $\gamma + \rho > 0$

Let us first consider the situation of a *Burr* model with $\gamma = 1$ and $\rho = -0.5$ (which illustrates the region $\gamma + \rho > 0$), presented in Figure 4, for $n = 1000$. Theoretically the dominant component of asymptotic bias is similar to that of the Hill estimator for every a but what happens for finite samples suggests us the following comments, on the basis of a search through $a = 0.1, 2.0(0.1)$: as a increases from $a = 0.1$ till $a = 2$ the MSE of $\gamma_n^{(a)}(k)$ is *always below* the MSE of $\gamma_n^{H}(k)$, but with 2 local minima for $a \leq 1$. The global mimimum is achieved at the largest value of k for $a \geq 0.3$. For values of $a > 1$ there is only one local minimum at a value k which decreases as a increases from $a = 1$ till $a = 2$. The MSE at that optimal level is increasing slightly with a as we increase a from $a = 0.3$ till $a = 2$. Anyway, all values of a are preferable to the value $a = 0$.

The situation for a *Fréchet* model in this same region, here illustrated in Figure 5 with a Fréchet parent with $\gamma = 2$ ($\gamma + \rho = 1 > 0$) is not quite different, and the final message is the same. As a increases from $a = 0.1$ till $a = 2$ the MSE of $\gamma_n^{(a)}(k)$ is also always below the MSE of $\gamma_n^{H}(k)$. There is a local minimum at $k = n - 1$ for $a \geq 0.8$, which is never the global minimum. For $a \leq 0.4$ there exists a unique minimum, which decreases as a increases and is attained at a value k which increases with a. For $a \geq 0.9$ the global minimum is increasing slightly with a, and is attained at a value k which decreases with a. For the intermediate values $0.5 \leq a \leq 0.8$ there are 2 local minima attained at values k strictly between 1 and $n - 1$. For $a = 0.5$ the global minimum is attained at the smallest k, being attained at the largest k for $a \neq 0.5$.

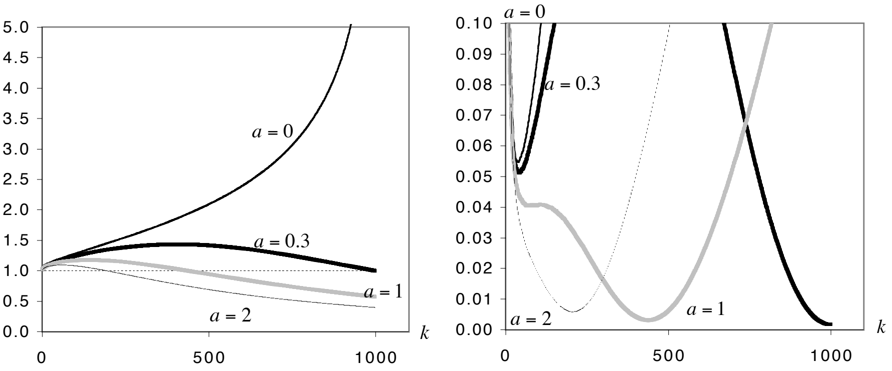

Figure 4: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^{(a)}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -0.5$.

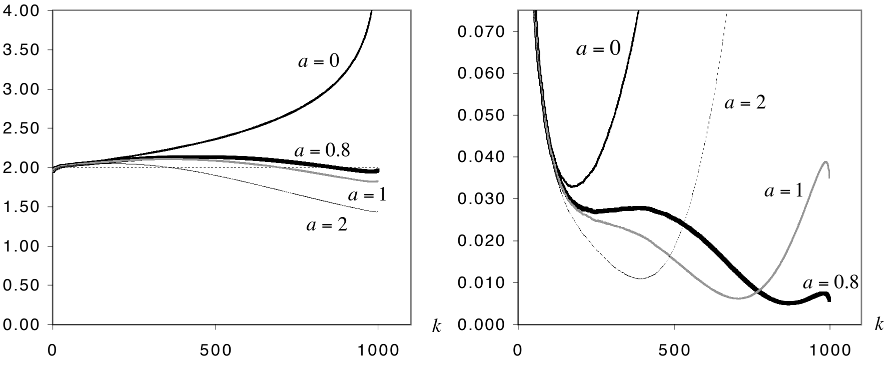

Figure 5: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^{(a)}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Fréchet* parent with $\gamma = 2$ ($\rho = -1$).

The value $a = 1$ seems to be a sensible choice in all the semi-plane $\gamma + \rho > 0$.

3.2 The case $\gamma + \rho < 0$

We next illustrate in Figure 6, the case $\gamma + \rho < 0$, with the simulation of *Burr* samples with $\rho = -2$ and $\gamma = 1$. In a search, running also from $a = 0$ till $a = 2$ with step 0.1, we have got MSE's with a unique minimum for $a \leq 0.6$. For larger values of a there appear two local minima, being the global minimum achieved at the larger of the two k-values whenever $a \leq 1.8$. The value of k where the global minimum is achieved increases as a increases, being equal to 999 for $n = 1000$ and $1.5 \leq a \leq 1.8$. The corresponding MSE decreases till $a = 1.2$ and then increases. It is interesting to notice that till $a = 0.3$ the MSE of $\gamma_n^{(a)}(k)$ is, for every k, smaller than the optimal MSE for the Hill estimator. This suggests a choice of $a = 0.3$. However, such a choice suggested by the MSE behaviour, which is difficult to estimate in practice, has not a strong reflection upon the mean value function, which is already quite stable for the Hill estimator. Indeed, we have already the experience that this region of ρ-values is a region where it is not easy to overpass the Hill estimator.

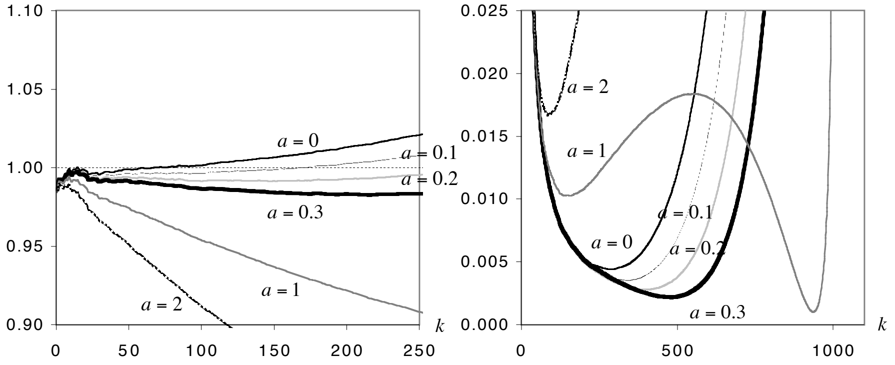

Figure 6: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^{(a)}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -2$.

We finally present in Figure 7 the behaviour of the class of estimators in (1.7) for *Fréchet* parents with $\gamma = 0.5$, again in the region $\gamma + \rho < 0$, which is quite similar to the behaviour we have obtained for the Burr parent in this same region.

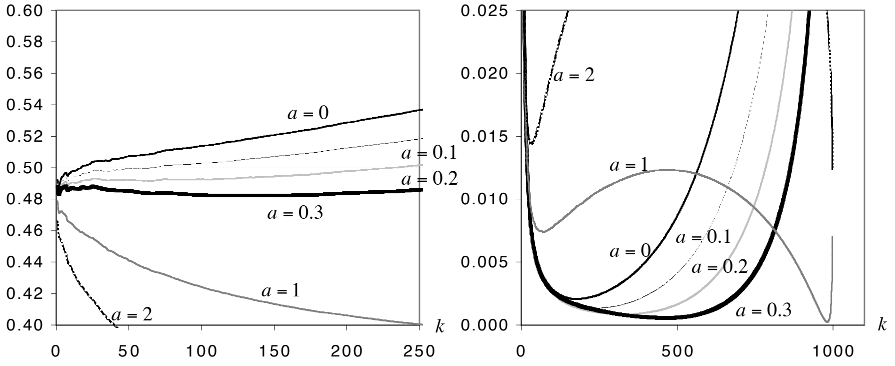

Figure 7: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^{(a)}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Fréchet* parent with $\gamma = 0.5$ ($\rho = -1$).

Here again, in a search running from $a = 0$ till $a = 2$ with step 1, we have now got MSE's with a unique minimum for $a \leq 0.5$. For larger values of a there appear two local minima, being the global minimum always achieved for the larger of the two k-values. That value of k where the global minimum is achieved increases as a increases, being equal to 999 for $n = 1000$ and

$a \geq 1.4$. The corresponding MSE decreases till $a = 0.9$ and then increases. It is interesting to notice that also till $a = 0.3$ the MSE of $\gamma_n^{(a)}(k)$ is, for every k, smaller than the optimal MSE for the Hill estimator. This suggests again a choice of $a = 0.3$. The value $a = 0.3$ provides really slightly better results than $a = 0$, in terms of stability of sample paths, but with a slight underestimation of the tail index γ. The minimum value for the MSE is attained for $a = 0.9$, quite close to the one obtained for $a = 1$, pictured in Figure 7, but at the second local minima, which corresponds to a k value which is difficult to justify as intermediate.

The choice $a = 0.3$ seems thus to be a sensible choice in the semi-plane $\gamma + \rho < 0$, but we stress again that the choice of the *tuning* parameter a must be suggested by means of any adequate stability criterion.

4 The special case of the Generalized Pareto model

Let us next think on a Generalized Pareto (GP) model, with tail function $1 - F(x) = (1 + \gamma x)^{-1/\gamma}$, $x > 0$ ($\gamma > 0$). We are again in Hall's class of models with $C = \gamma^{-1/\gamma}$, $D = -1/\gamma^2$ and $\rho = -\gamma$. Then, for every γ if we induce in the data a shift $a = \rho D = 1/\gamma = -1/\rho$ we are able to remove the dominant component of asymptotic bias. We illustrate this behaviour in Figures 8 and 9, where we picture the mean values' and the mean squared errors' patterns, respectively, for GP parents with $\gamma = 0.5$, $\gamma = 1$ and $\gamma = 2$.

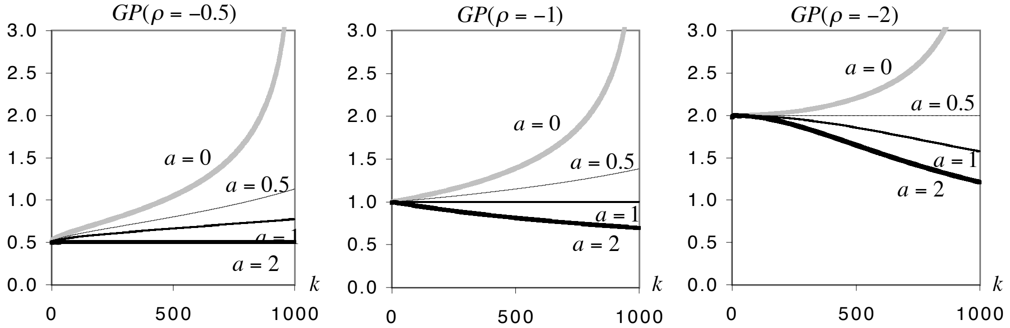

Figure 8: Simulated mean values of $\gamma_n^{(a)}(k)$, $a = 0, 0.5, 1, 2$, based on 5000 runs, for a sample size $n = 1000$, from *Generalized Pareto* parents with $\gamma = 0.5, 1, 2$ ($\rho = -0.5, -1, -2$).

Finally, Figure 10 is a zoom of Figure 8, in a region of size 0.1 around the target value γ.

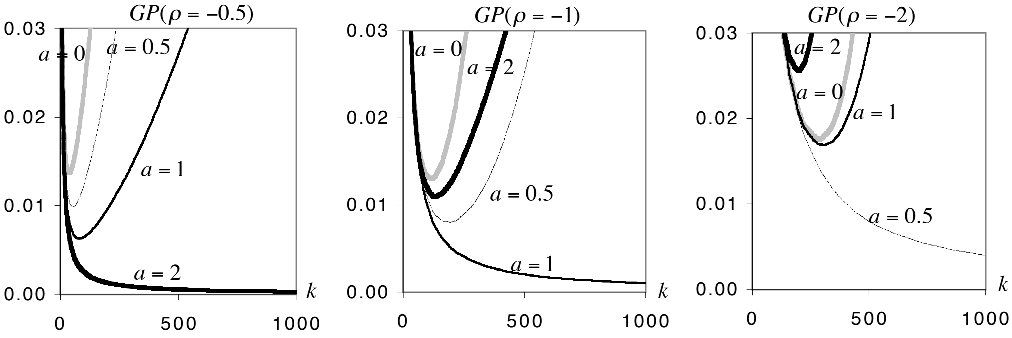

Figure 9: Simulated *MSE's* of $\gamma_n^{(a)}(k)$, $a = 0, 0.5, 1, 2$, based on 5000 runs, for a sample size $n = 1000$, from *Generalized Pareto* parents with $\gamma = 0.5, 1, 2$ ($\rho = -0.5, -1, -2$).

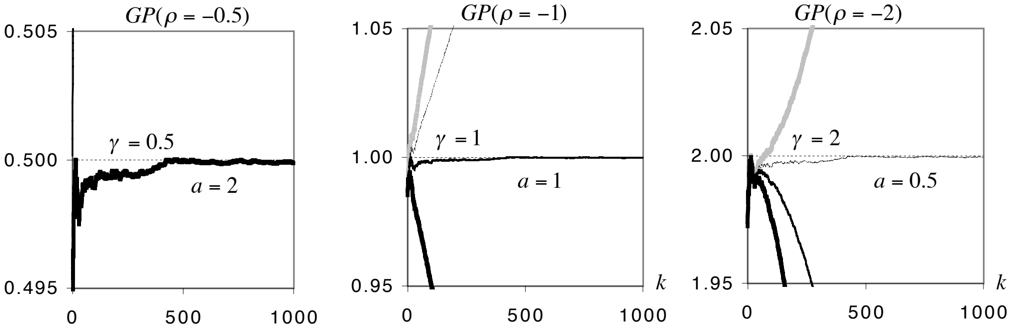

Figure 10: Zoom of Figure 8

This result leads us to make a few statements about future possible research:

1. If we generally work with the excesses over a high level, say $X_{n-k-1:n}$ for instance, i.e., with

$$Z_{i,k} = X_{n-i+1:n} - X_{n-k-1:n}, \quad 1 \le i \le k+1, \quad k = 1, \cdots, n-2, \tag{4.1}$$

and for each k consider the class of shifted Hill estimators

$$\widetilde{\gamma}_k^{(a)}(j) := \frac{1}{j}\sum_{i=1}^{j} \ln \frac{Z_{i,k} + a}{Z_{j+1,k} + a}, \ j = 1, 2, \cdots, k, \tag{4.2}$$

we expect to get interesting stable sample paths for every model $F \in \mathcal{D}(G_\gamma)$, $\gamma > 0$, whenever the value k is such that a Paretian behaviour holds for the excesses in (4.1), i.e., whenever (1.5) holds. Theoretically, the optimal value of $a = a(k)$, which is the value of a associated to a $GP(\delta_k, \gamma)$, with scale δ_k and shape γ, is $a = \delta_k/\gamma$. For each k we thus suggest the choice of the value $a_0 = a_0(k)$ which provides the most stable sample path, accordingly to an adequate stability criterion, for instance a least squares technique,

$$a_0(k) := \arg\min_a \sum_{m=[k/4]+1}^{[k/2]} \left\{ \widetilde{\gamma}_k^{(a)}(m) - \overline{\widetilde{\gamma}_k^{(a)}} \right\}^2 ,$$

where $\overline{\widetilde{\gamma}_k^{(a)}} = \frac{1}{[k/2]-[k/4]} \sum\limits_{j=[k/4]+1}^{[k/2]} \widetilde{\gamma}_k^{(a)}(j)$.

This minimization problem is equivalent to the search of the solution of the equation

$$\varphi(a_0) = \overline{\widetilde{\gamma}_k^{(a_0)} \widetilde{\gamma}_k^{\prime(a_0)}} - \overline{\widetilde{\gamma}_k^{(a_0)}}\; \overline{\widetilde{\gamma}_k^{\prime(a_0)}} = 0,$$

where $\widetilde{\gamma}_k^{\prime(a)}(j) = \frac{1}{j} \sum\limits_{i=1}^{j} (Z_{i,k}+a)^{-1} - (Z_{j+1,k}+a)^{-1}$, being all averages performed over the same set of integer values of j, say from $[k/4]+1$ till $[k/2]$.

2. The methodology proposed in 1. may even give rise to an heuristic choice of the threshold to be considered in the *Peaks Over Threshold* (*POT*) approach. Such a choice is given by

$$k_0 := \arg\min_k \left\{ \frac{1}{[k/2]-[k/4]} \sum_{j=[k/4]+1}^{[k/2]} \left(\widetilde{\gamma}_k^{(a_0(k))}(k) - \overline{\gamma_k^{(a_0(k))}} \right)^2 \right\}.$$

The final estimate of the tail index γ is then, for instance,

$$\widehat{\gamma}_0 := \overline{\gamma_{k_0}^{(a_0(k_0))}}. \tag{4.3}$$

3. Alternatively, since a suitable choice for δ_k based on the extremal limiting process is $\delta_k = \{1 + \gamma X_{n-k-1:n}\}$ (Gomes and Alpuim, 1986), we could also consider the estimator, obtained through the iterative formula,

$$\widetilde{\gamma}_{j+1}^* \equiv \widetilde{\gamma}_{j+1}^*(k) := \frac{1}{k} \sum_{i=1}^{k} \ln \frac{X_{n-i+1:n} + 1/\widetilde{\gamma}_j^*}{X_{n-k:n} + 1/\widetilde{\gamma}_j^*}, \quad j \geq 0. \tag{4.4}$$

The distributional properties of this estimator were indeed simulated, and they are quite similar to those of the estimator proposed in the following section for models with a finite left endpoint, the estimator $\gamma_n^{(1-X_{1:n})}(k)$. Notice also that the possible choice $\delta_k = \gamma X_{n-k-1:n}$ would lead to $a = X_{n-k-1:n}$, and the estimator in (4.2) would be again the Hill estimator.

5 The estimator $\gamma_n^*(k) \equiv \gamma_n^{(1-X_{1:n})}(k)$

We shall now work with the estimator

$$\gamma_n^*(k) = \gamma_n^{(1-X_{1:n})}(k) := \frac{1}{k}\sum_{i=1}^{k} \ln \frac{X_{n-i+1:n} - X_{1:n} + 1}{X_{n-k:n} - X_{1:n} + 1}, \tag{5.1}$$

from which we were expecting a better behaviour than that of the Hill estimator and not a long way from the estimator in (1.7), for the optimal value of a. However, this last statement does not really holds true in general.

First of all it is important to mention that we can guarantee that such an estimator is consistent for the estimation of γ, with asymptotic distributional properties similar to those of the Hill estimator, only if the underlying model in $\mathcal{D}(G_\gamma)$ has a finite left endpoint. Fortunately this happens often in practice, particularly when we are dealing with heavy tail data. We have simulated the distributional properties of the estimator in (5.1) for Fréchet models with $\gamma = 0.5,\ 1$ and 2, Burr models with $\gamma = 1$ and $\rho = -0.5, -1$ and -2, and GP models with $\gamma = 0.5, 1$ and 2.

We obviously agree that, both from a theoretical and from a practical point of view, the best thing would be to have a *location/scale invariant* estimator with small mean squared error. But, if we are prepared to play also with estimators which may have this property only asymptotically, it is obviously difficult (if not impossible) to achieve minimum MSE with a *location/scale invariant* estimator. We shall illustrate this comment, performing a comparison of the estimators presented in this paper with one of the most recent and interesting location/scale invariant estimators of the tail index γ, available in the literature, the one introduced by Fraga Alves (2001). Such estimator is a Pickands-type estimator, but based upon the excesses

$$V_{ik} = X_{n-i+1:n} - X_{n-k:n}, \quad 1 \le i \le k, \tag{5.2}$$

which are, for intermediate k, i.e., for sequences of integers $k = k_n$ such that (1.5) holds, and after a scaling by for instance $\gamma X_{n-k:n}$, approximately the o.s. of a sample of size k from a Generalized Pareto model, $GP(\gamma)$. Based then on the fact that for the $GP(\gamma)$, the second order parameter $\rho = -\gamma$, Fraga Alves (2001) is able to remove easily the main dominant component of bias, which now depends only on γ (and not on ρ, as usual). The location/invariant estimator (still with bias) has the functional form:

$$\gamma_n(k; k_0) := \frac{1}{k_0}\sum_{i=1}^{k_0} \ln \frac{X_{n-i+1:n} - X_{n-k:n}}{X_{n-k_0:n} - X_{n-k:n}}, \ k_0 = 1, 2, \cdots, k-1. \tag{5.3}$$

Then, under the second order framework in (1.3), for k such that (1.5) holds, and for k_0 intermediate relatively to k, i.e., such that $k_0 \to \infty$ and $k_0/k \to 0$, as

$n \to \infty$, asymptotic normality is achieved by $\gamma_n(k;k_0)$, with an asymptotic bias at the optimal level $k_0 = \left(\frac{1+\gamma}{\sqrt{2}\gamma}\right)^{2/(1+2\gamma)} k^{2\gamma/(1+2\gamma)}$ given by $\sqrt{\gamma/(2k_0)}$. Fraga Alves (2001) proposes then the following algorithm for the estimation of the tail index γ:

1. for each $k = 1, 2, \cdots, n-1$, consider $k_0^* = 2k^{2/3}$, and obtain $\gamma_0^* = \gamma_n(k;k_0^*)$;
2. compute
$$\widehat{k}_0 = \left(\frac{(1+\gamma_0^*)^2}{2\gamma_0^*}\right)^{1/(1+2\gamma_0^*)} k^{2\gamma_0^*/(1+2\gamma_0^*)}; \tag{5.4}$$
3. the final semi-parametric estimator is then
$$\gamma_n^{FA}(k) = \gamma_n(k;\widehat{k}_0) - \sqrt{\gamma_n(k;\widehat{k}_0)/(2\widehat{k}_0)}, \tag{5.5}$$
with $\gamma_n(k;k_0)$ and $\widehat{k}_0$ given in (5.3) and (5.4), respectively.

From Figures 11 till 15 we illustrate the region of the (γ,ρ)-plane for which $\gamma+\rho=0$. In all Figures we picture mean values and mean squared errors of $\gamma_n^H(k)$ in (1.4), of the "optimal" $\gamma_n^{(a)}(k)$ in (1.7), with a_0 obtained from the asymptotic results given before or by simulation, of $\gamma_n^*(k)$ in (5.1) and of $\gamma_n^{FA}(k)$ in (5.5).

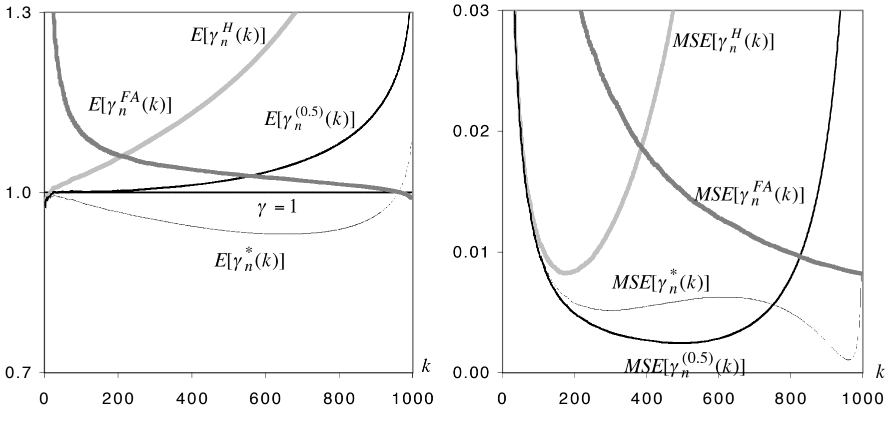

Figure 11: Simulated mean values (*left*) and *MSE's* (*right*) of the estimators under study, based on 5000 runs, for a sample size $n = 1000$, from a *Fréchet* parent with $\gamma = 1$ ($\rho = -1$).

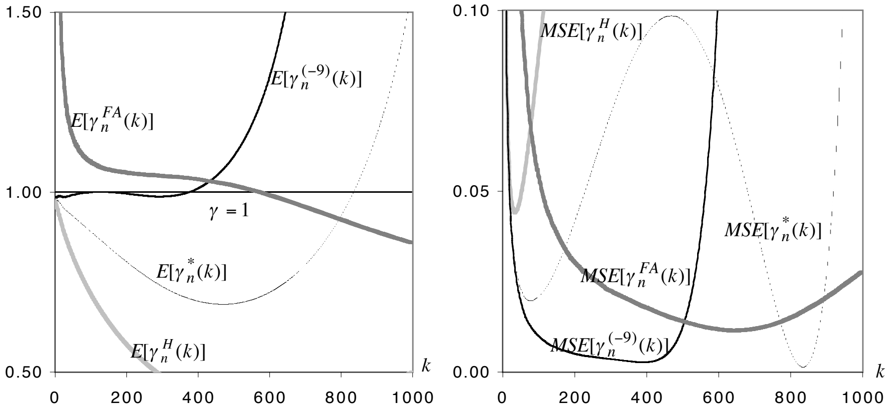

Figure 12: Simulated mean values (*left*) and *MSE's* (*right*) of the estimators under study, based on 5000 runs, for a sample size $n = 1000$, from a *Out-Hall* parent with $\gamma = 1$ ($\rho = -1$).

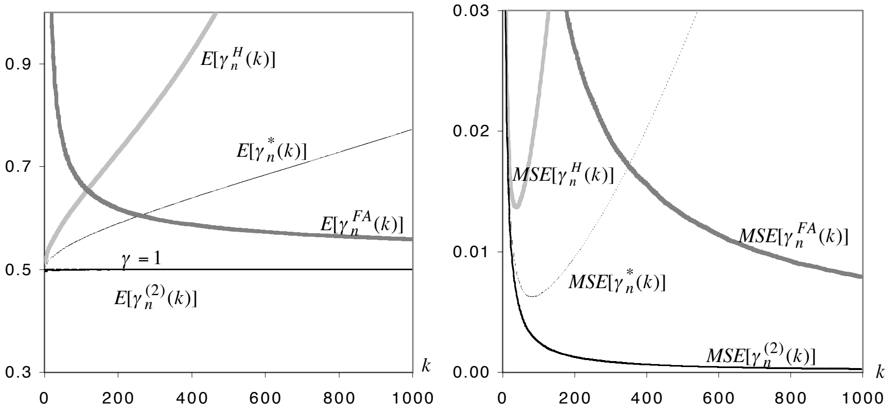

Figure 13: Simulated mean values (*left*) and *MSE's* (*right*) of the estimators under study, based on 5000 runs, for a sample size $n = 1000$, from a *GP* parent with $\gamma = 0.5$ (and $\rho = -0.5$).

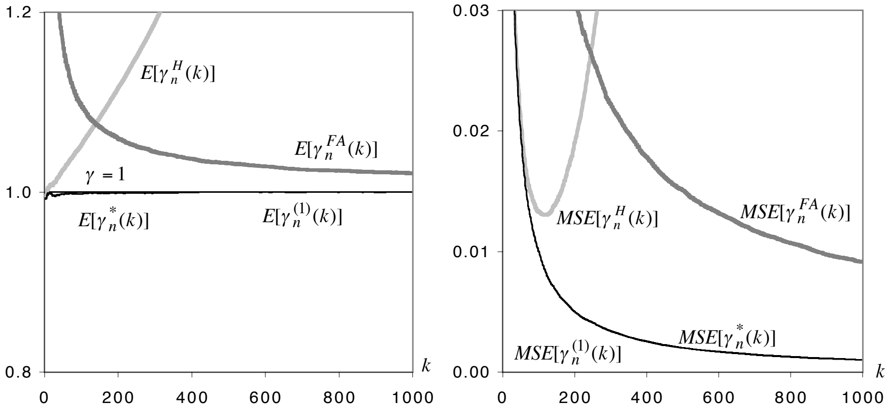

Figure 14: Simulated mean values (*left*) and *MSE's* (*right*) of the estimators under study, based on 5000 runs, for a sample size $n = 1000$, from a *GP* parent with $\gamma = 1$ (and $\rho = -1$) (the same structure than the one obtained for a Burr model with $\gamma = 1$ and $\rho = -1$).

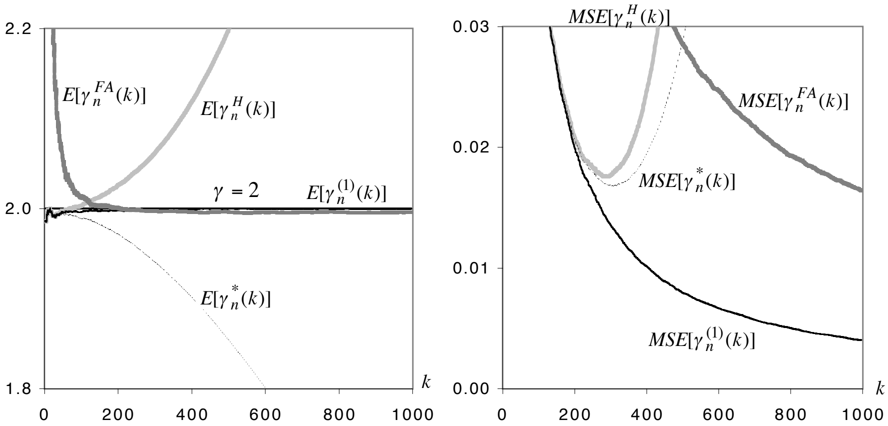

Figure 15: Simulated mean values (*left*) and *MSE's* (*right*) of the estimators under study, based on 5000 runs, for a sample size $n = 1000$, from a *GP* parent with $\gamma = 2$ (and $\rho = -2$).

Notice that for the *Fréchet* and the *Out-Hall* models, and whenever we consider both estimators at their optimal level, $\gamma_n^*(k)$ behaves better than $\gamma_n^{(a_0)}(k)$ only because its MSE pattern has two local minima, with the global minimum achieved at a very large value of k, which is difficult to justify as intermediate. For the first local minimum, γ_n^* behaves better than γ_n^H, but worser than $\gamma_n^{(a_0)}$.

In Figure 16 we illustrate the region $\gamma + \rho < 0$, where things become worse for the first local minimum of the MSE (perhaps the one achieved at the intermediate k!). Indeed, we would expect such a behaviour since the bias is higher asymptotically: for the Fréchet model the bias it is now $O(\sqrt{k/n})$, whereas for the Hill estimator was $O(k/n)$.

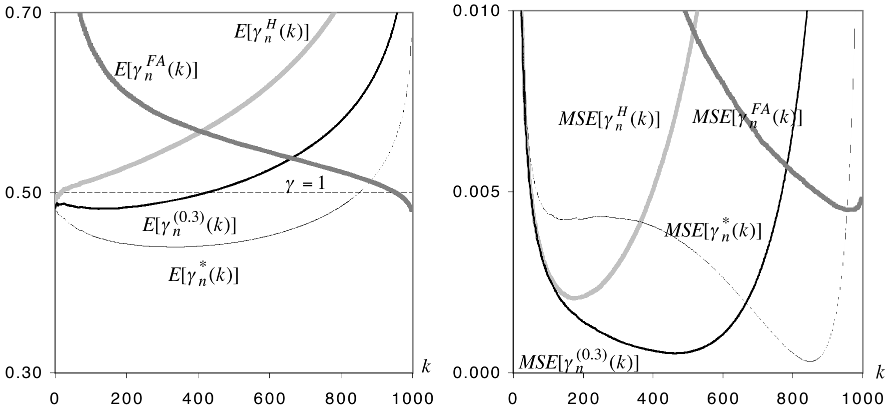

Figure 16: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^{(1-X_{1:n})}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Fréchet* parent with $\gamma = 0.5$ ($\rho = -1$).

Finally, in Figure 17 we illustrate what happens in the region $\gamma + \rho > 0$, now with a Fréchet parent with $\gamma = 2$.

In Table 1 we merely present the main simulated distributional properties of γ_{n0}^*, at the global minimum.

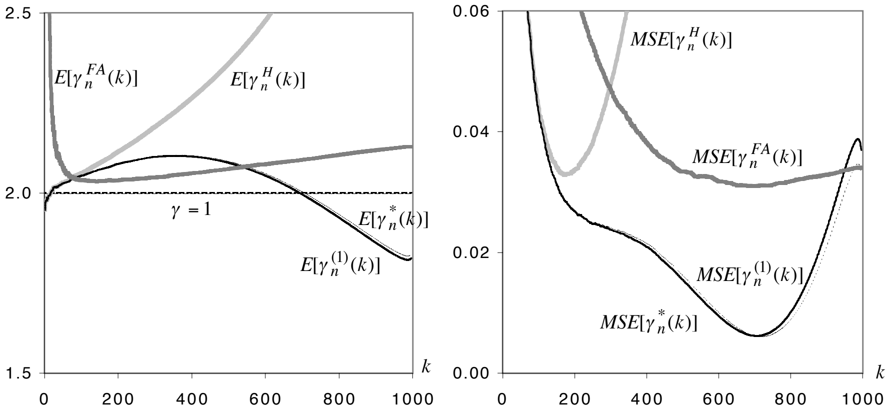

Figure 17: Simulated mean values (*left*) and *MSE's* (*right*) of the estimators under study, based on 5000 runs, for a sample size $n = 1000$, from a *Fréchet* parent with $\gamma = 2$ ($\rho = -1$).

Table 1: Simulated optimal sample fractions, mean values, mean squared errors and relative efficiencies of γ^*_{n0} for heavy tail models with finite left endpoint.

n	100	200	500	1000	2000	5000	10000	20000
Fréchet parent: $\rho = -\mathbf{1}$, $\gamma = \mathbf{1}$								
k_{0s}/n	0.9260	0.9425	0.9542	0.9612	0.9659	0.9707	0.9734	0.9757
E	1.0047	1.0027	1.0003	1.0002	0.9997	0.9998	0.9999	1.0000
MSE	0.0105	0.0052	0.0021	0.0011	0.0005	0.0002	0.0001	0.0001
$REFF$	2.0624	2.2592	2.5503	2.8137	3.0551	3.4106	3.6397	3.8684
Burr parent: $\rho = -\mathbf{1}$, $\gamma = \mathbf{1}$								
k_{0s}/n	0.9700	0.9850	0.9936	0.9963	0.9979	0.9989	0.9995	0.9996
E	1.0047	1.0025	1.0006	1.0003	1.0002	1.0000	1.0000	1.0000
MSE	0.0104	0.0051	0.0020	0.0010	0.0005	0.0002	0.0001	0.0001
$REFF$	2.6091	2.8603	3.2611	3.6284	4.0308	4.6475	5.2048	5.7850
OUT-HALL parent: $\rho = -\mathbf{1}$, $\gamma = \mathbf{1}$								
k_{0s}/n	0.8200	0.8280	0.8328	0.8344	0.8350	0.8357	0.8359	0.8360
E	0.9959	0.9977	0.9987	0.9994	0.9993	0.9999	0.9999	0.9999
MSE	0.0133	0.0067	0.0027	0.0013	0.0007	0.0003	0.0001	0.0001
$REFF$	3.4422	4.0205	4.9233	5.6986	6.5585	7.9025	9.0698	10.3853
GP parent: $\rho = -\mathbf{0.5}$, $\gamma = \mathbf{0.5}$								
k_{0s}/n	0.2300	0.1720	0.1112	0.0820	0.0598	0.0383	0.0276	0.0200
E	0.5996	0.5826	0.5637	0.5536	0.5448	0.5352	0.5295	0.5248
MSE	0.0239	0.0159	0.0093	0.0064	0.0044	0.0027	0.0018	0.0013
$REFF$	1.5448	1.5113	1.4945	1.4792	1.4618	1.4595	1.4452	1.4365
GP parent: $\rho = -\mathbf{1}$, $\gamma = \mathbf{1}$								
k_{0s}/n	0.9700	0.9850	0.9936	0.9963	0.9979	0.9989	0.9995	0.9996
E	1.0047	1.0025	1.0006	1.0003	1.0002	1.0000	1.0000	1.0000
MSE	0.0104	0.0051	0.0020	0.0010	0.0005	0.0002	0.0001	0.0001
$REFF$	2.6091	2.8603	3.2611	3.6284	4.0308	4.6475	5.2048	5.7850
GP parent: $\rho = -\mathbf{2}$, $\gamma = \mathbf{2}$								
k_{0s}/n	0.5070	0.4345	0.3518	0.3069	0.2611	0.2172	0.1903	0.1649
E	1.8499	1.8860	1.9217	1.9400	1.9559	1.9695	1.9765	1.9822
MSE	0.0974	0.0575	0.0284	0.0164	0.0095	0.0045	0.0026	0.0015
$REFF$	1.1090	1.0740	1.0446	1.0339	1.0248	1.0197	1.0162	1.0111
Fréchet parent: $\rho = -\mathbf{1}$, $\gamma = \mathbf{0.5}$								
k_{0s}/n	0.7870	0.8140	0.8362	0.8509	0.8618	0.8743	0.8820	0.8884
E	0.5025	0.5017	0.5001	0.5001	0.4998	0.4999	0.4999	0.4999
MSE	0.0030	0.0015	0.0006	0.0003	0.0002	0.0001	0.0000	0.0000
$REFF$	1.9423	2.1270	2.3984	2.5984	2.7584	2.9157	2.9258	2.8776
BURR parent: $\rho = -\mathbf{2}$, $\gamma = \mathbf{1}$								
k_{0s}/n	0.8910	0.9050	0.9194	0.9247	0.9290	0.9325	0.9341	0.9352
E	1.0027	0.9995	1.0000	0.9993	0.9998	1.0000	1.0000	1.0000
MSE	0.0107	0.0053	0.0021	0.0011	0.0005	0.0002	0.0001	0.0001
$REFF$	1.6732	1.7739	1.9124	2.0434	2.1908	2.3854	2.5573	2.7277
Fréchet parent: $\rho = -\mathbf{1}$, $\gamma = \mathbf{2}$								
k_{0s}/n	0.8110	0.7640	0.7330	0.7172	0.7090	0.7037	0.7011	0.6991
E	1.9433	1.9691	1.9845	1.9926	1.9963	1.9982	1.9991	1.9996
MSE	0.0563	0.0291	0.0120	0.0061	0.0031	0.0012	0.0006	0.0003
$REFF$	1.7826	1.9090	2.1263	2.3395	2.5688	2.9603	3.2893	3.7076
BURR parent: $\rho = -\mathbf{0.5}$, $\gamma = \mathbf{1}$								
k_{0s}/n	0.4730	0.4535	0.4418	0.4386	0.4367	0.4352	0.4348	0.4346
E	0.9609	0.9805	0.9919	0.9955	0.9977	0.9991	0.9995	0.9997
MSE	0.0291	0.0150	0.0061	0.0031	0.0015	0.0006	0.0003	0.0002
$REFF$	2.8027	3.1042	3.6892	4.2599	4.9304	6.1168	7.1132	8.3086

We next present in Figures 18, 19 and 20 the simulated relative efficiencies of $\gamma_{n0}^{(a_0)}$ (for the optimal a_0 obtained either asymptotically or heuristically trough Monte Carlo techniques), γ_{n0}^{*} and γ_{n0}^{FA}, also at the global minimum, for n = 100, 200, 500, 1000, 2000 and 5000, and for the models presented before with $\gamma + \rho = 0$, $\gamma + \rho < 0$ and $\gamma + \rho > 0$, respectively.

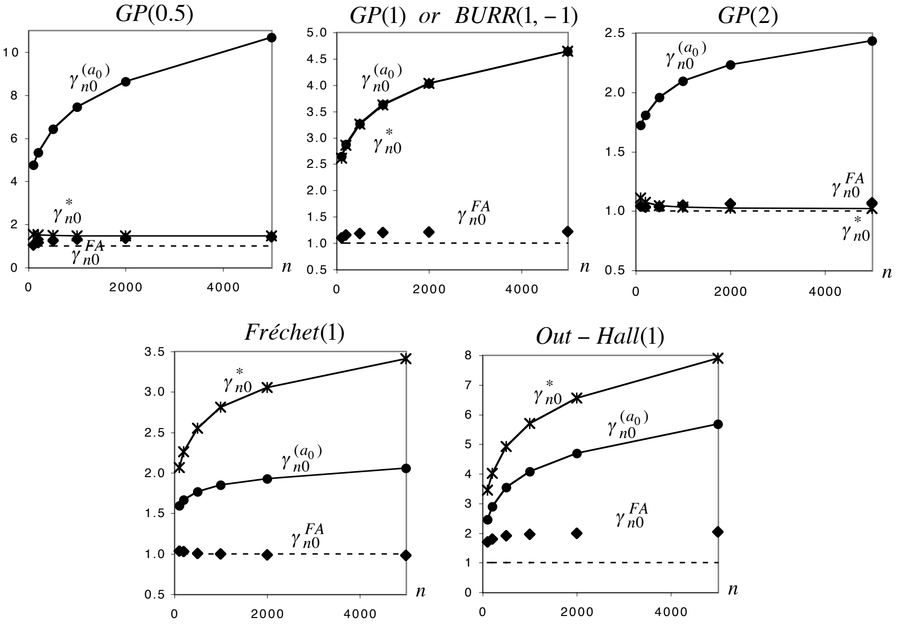

Figure 18: Simulated Efficiencies of the estimators under study, for models in the region $\gamma + \rho = 0$

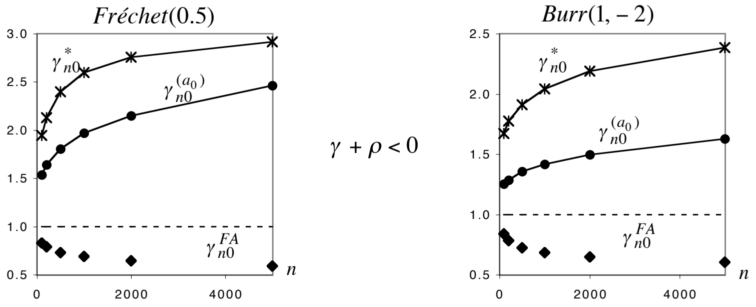

Figure 19: Simulated Efficiencies of the estimators under study, for models in the region $\gamma + \rho < 0$

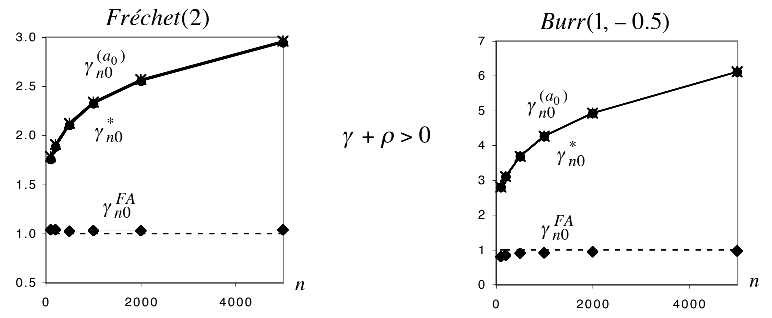

Figure 20: Simulated Efficiencies of the estimators under study, for models in the region $\gamma + \rho > 0$

Two final remarks:

1. The previous pictures enable us to see that we pay a lot in mean squared error whenever we use a *location/scale invariant* estimator, even when we use an extra procedure to remove the main component of bias, like the one used in (5.5). This is only due to the variance of the estimator in (5.5); indeed, the sample paths are usually quite stable around the target value, here the tail index γ.

2. The introduction of a *tuning* parameter a, which is merely a shift in the original data, in the Hill estimator, may help us to increase highly the efficiency of the new estimator, although at the lost of invariance regarding scale transformations.

An overall comment:

As a general final comment we would dare to say that whenever we are dealing with semi-parameteric estimation of a parameter of rare events we should take into account the following facts:

- it is better not to work only with one semi-parametric estimator; choose adequately a set of simple semi-parametric estimators of a parameter of rare events, or, preferably, a class of estimators parametrized on a tuning parameter, which one may control at our ease, and picture a few sample paths associated to different values of that tuning parameter; such a class is provided in this paper, the class (1.7), but there are alternative classes, like the ones considered in Gomes and Martins (2001) and Caeiro and Gomes (2002);

- Do not be afraid of estimators which are not invariant for location. Play with them in your benefit, in order to reduce mean squared error and obtain smooth sample paths;

- Explicit estimators are beautiful and may be highly efficient! There seems to be no reason to go into terribly computer time-consuming methodologies, particularly if we want to study, by means of large scale simulations, the distributional behaviour of the developed estimators for moderate sample sizes.

References

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

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

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

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

[5] Fraga Alves, M. I. (2001). A location invariant Hill-type estimator. *Extremes* **4**:3, 199-217.

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

[7] Gomes, M.I. and M.T. Alpuim (1986). Inference in multivariate generalized extreme value models. *Scandinavian J. Statist.* **13**, 291-300.

[8] Gomes, M.I. (2001). *Bias reduction and efficient estimation of the tail index.* Notas e Comunicações 10/2001.

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

[10] Gomes, M.I. and M.J. Martins (2001). Generalizations of the Hill estimator — asymptotic versus finite sample behaviour. *J. Statist. Planning and Inference* **93**, 161-180.

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

[12] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Royal Statist. Soc.* **44**, 37-42.

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

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

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

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