Censoring estimators of a positive tail index*

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 analyse some variants of a *maximum likelihood* estimator of a positive tail index γ, under a type II *censoring* scheme. These estimators are compared with the *Hill estimator*, for a Fréchet model and by means of a Monte Carlo simulation. Asymptotic normality of the estimators is derived, and a robustness study of the estimators is undertaken.

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

1 The new semi-parametric estimators and scope of the paper

Under a heavy tail framework, i.e., whenever we assume that the tail of the model $F(\cdot)$, underlying the data, is a regularly varying function with index $-1/\gamma$, $\gamma > 0$, the Pareto behaviour of the top scaled order statistics (o.s.),

$$\frac{X_{n-i+1:n}}{X_{n-k:n}}, \quad 1 \le i \le k,$$

leads us to a maximum likelihood estimator of γ given by

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

which was introduced by Hill (1975). As usual, $X_{i:n}$ denotes the i-th ascending o.s., $1 \le i \le n$, associated to the sample $(X_1, X_2, \cdots, X_n)$ of independent random variables (r.v.'s) with common distribution function (d.f.) $F(\cdot)$. For more details on the asymptotic theory of order statistics see Galambos (1987).

The Hill estimator is a consistent estimator of γ whenever k is intermediate (Mason, 1982), i.e.,

$$k = k_n \to \infty, \; k_n/n \to 0, \; \text{as } n \to \infty. \tag{2}$$

*Research partially supported by FCT / POCTI / FEDER.
†DEIO, Edifício C2, Campo Grande, 1749-016 Lisboa, Portugal.

If we instead consider the Fréchet behaviour of X/a, and estimate jointly γ and a through *Maximum Likelihood* (ML), under a type II censoring scheme, where we have access to the top $k+1$ o.s., $\underline{X}_k = (X_{n-k:n} \le X_{n-k+1:n} \le \cdots \le X_{n:n})$, we get an estimator $\widehat{\gamma}_n(k)$ such that

$$\widehat{\gamma}_n(k) = \frac{1}{k+1}\sum_{i=1}^{k+1} \ln X_{n-i+1:n} - \\ - \frac{\sum_{i=1}^{k} X_{n-i+1:n}^{-1/\widehat{\gamma}_n(k)} \ln X_{n-i+1:n} + (n-k) X_{n-k:n}^{-1/\widehat{\gamma}_n(k)} \ln X_{n-k:n}}{\sum_{i=1}^{k} X_{n-i+1:n}^{-1/\widehat{\gamma}_n(k)} + (n-k) X_{n-k:n}^{-1/\widehat{\gamma}_n(k)}},$$

which may also be written as

$$\widehat{\gamma}_n(k) = \frac{k}{k+1}\widehat{\gamma}_n^H(k) - \frac{\frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{-1/\widehat{\gamma}_n(k)} \ln\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)}{\frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{-1/\widehat{\gamma}_n(k)} + \frac{n}{k} - 1}. \tag{3}$$

In this paper, we have not worked with the estimator in (3), which is easy to get iteratively for one sample, but leads to time-consuming large-scale simulations. We have worked instead with an explicit estimator which is not a long way from the estimator in (3), denoted by $\widehat{\gamma}_n^C(k)$, and given by the expression in the second member of (3), but with $\widehat{\gamma}_n(k)$ replaced by the Hill estimator $\widehat{\gamma}_n^H(k)$ i.e.,

$$\widehat{\gamma}_n^C(k) := \frac{k}{k+1}\widehat{\gamma}_n^H(k) - \frac{\frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{-1/\widehat{\gamma}_n^H(k)} \ln\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)}{\frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{-1/\widehat{\gamma}_n^H(k)} + \frac{n}{k} - 1}. \tag{4}$$

Indeed, the simulation results obtained for the maximum likelihood estimator in (3), although computer time-consuming, reproduce the ones obtained for the estimator in (4), whenever $\rho \ge -1$. However, for a Fréchet model, the estimator in (3) remains consistent till $k = n-1$, and the MSE structure is then decreasing with k, as shown in Figure 1. The performance of the estimator in (3), for $\rho < -1$, is shown in Figure 5, to illustrate its behaviour in this region of ρ-values.

Also, since we may write, for intermediate sequences (as we shall see in detail later on, in section 3 of this paper),

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

we suggest the slightly easier explicit estimator

$$\widehat{\gamma}_n^{C_1}(k) := \widehat{\gamma}_n^H(k) - \frac{1}{n}\sum_{i=1}^{k}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{-1/\widehat{\gamma}_n^H(k)} \ln\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right), \tag{5}$$

or alternatively, in between $\widehat{\gamma}_n^C(k)$ and $\widehat{\gamma}_n^{C_1}(k)$, we may consider the estimator

$$\widehat{\gamma}_n^{C_2}(k) := \frac{k}{k+1}\widehat{\gamma}_n^H(k) - \frac{1}{n}\sum_{i=1}^{k}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{-1/\widehat{\gamma}_n^H(k)} \ln\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right). \quad (6)$$

Indeed they all appear to be able to reduce the asymptotic bias of the Hill estimator for intermediate, but reasonably large values of k, not only for the Fréchet model for which they were explicitly built, but also for other models, as we shall see in the robustness study developed in section 4 of this paper. Previously to such a robustness study we shall discuss in section 2 the finite sample behaviour of the censoring estimators under investigation, for a Fréchet model. The asymptotic behaviour of those estimators is studied in section 3.

2 Finite sample behaviour of the estimators for a Fréchet model

We shall here obtain the finite sample properties of the above mentioned estimators of the tail index, for the *Fréchet* model, $F(x) = \exp(-x^{-1/\gamma}),\ x \geq 0$, with $\gamma = 1$. The simulation results were based on a multi-sample simulation of size 5000×10 in order to guarantee small standard errors for the simulated characteristics, the Mean Value ($E^\bullet$), the Mean Squared Error ($MSE^\bullet$), the Optimal Sample Fraction, $OSF^\bullet \equiv k_o^\bullet/n$, with $k_o^\bullet := \arg\min_k MSE^\bullet(k)$, and the Relative Efficiency ($REFF^\bullet$), defined as

$$REFF^\bullet = REFF[\widehat{\gamma}_{n,o}^{(\bullet)}|\widehat{\gamma}_{n,o}^H] = \sqrt{\frac{MSE\left[\widehat{\gamma}_{n,o}^H\right]}{MSE\left[\widehat{\gamma}_{n,o}^\bullet\right]}}, \quad (7)$$

with $\widehat{\gamma}_{n,o}^\bullet = \widehat{\gamma}_n^\bullet(k_o^\bullet(n))$, and where the index o indicates "optimality". With the index s indicating "simulation", the simulator of for instance $k_o^\bullet(n)$, denoted by $k_{o,s}^\bullet(n)$, is $\widehat{E}_{10}[\overline{k}_o^\bullet(n)]$, the average of 10 independent replicates of $\overline{k}_o^\bullet(n) = \arg\min_k \sum_{j=1}^{5000}\left(\widehat{\gamma}_{nj}^\bullet(k) - \gamma\right)^2$. 95% confidence intervals are presented in the tables. They are meaningful only when the simulated characteristics are asymptotically normal, which possibly does not happen to the optimal sample fraction.

In Tables 1 and 2, we show some finite sample properties of $\widehat{\gamma}_{n,o}^H$ (reproduction of results in Gomes and Oliveira, 2000), $\widehat{\gamma}_{n,o}^C$, and $\widehat{\gamma}_{n,o}^{C_j}$, $j = 1, 2$, for a *Fréchet* model. We do not place 95% confidence levels for the *MSE*'s whenever they are 0 up to 4 decimal figures.

Table 1: Simulated mean values and mean squared errors of the estimators under study, at their simulated optimal levels, and for a *Fréchet* parent.

n	E_s^H	E_s^C	E_s^{C1}	E_s^{C2}	MSE_s^H	MSE_s^C	MSE_s^{C1}	MSE_s^{C2}
100	1.1083	0.9524	1.0581	1.0403	0.0447	0.0130	0.0216	0.0187
	±.0041	±.0026	±.0038	±.0032	±.0007	±.0002	±.0002	±.0002
200	1.0850	0.9684	1.0449	1.0354	0.0265	0.0069	0.0119	0.0108
	±.0038	±.0014	±.0026	±.0017	±.0005	±.0001	±.0002	±.0002
500	1.0632	0.9796	1.0316	1.0270	0.0136	0.0030	0.0056	0.0053
	±.0025	±.0009	±.0018	±.0021	±.0002			
1000	1.0489	0.9863	1.0239	1.0226	0.0083	0.0016	0.0032	0.0031
	±.0019	±.0010	±.0021	±.0012	±.0001			
2000	1.0381	0.9908	1.0184	1.0171	0.0051	0.0009	0.0018	0.0018
	±.0013	±.0005	±.0007	±.0009				
5000	1.0295	0.9945	1.0128	1.0123	0.0027	0.0004	0.0009	0.0009
	±.0010	±.0003	±.0007	±.0006				
10000	1.0232	0.9966	1.0100	1.0098	0.0017	0.0002	0.0005	0.0005
	±.0008	±.0002	±.0005	±.0006				
20000	1.0184	0.9978	1.0077	1.0076	0.0011	0.0001	0.0003	0.0003
	±.0007	±.0001	±.0002	±.0004				

Table 2: Simulated relative efficiencies and optimal sample fractions of the estimators under study, at their simulated optimal levels, for a *Fréchet* parent.

n	$REFF_s^C$	$REFF_s^{C1}$	$REFF_s^{C2}$	$k_{o,s}^H/n$	$k_{o,s}^C/n$	$k_{o,s}^{C1}/n$	$k_{o,s}^{C2}/n$
100	1.8547	1.4403	1.5444	0.3370	0.7440	0.5440	0.5570
	±.0134	±.0102	±.0110	±.0101	±.0118	±.0140	±.0101
200	1.9572	1.4917	1.5693	0.2815	0.7010	0.4935	0.5030
	±.0183	±.0097	±.0107	±.0089	±.0087	±.0101	±.0059
500	2.1136	1.5553	1.6070	0.2208	0.6588	0.4260	0.4298
	±.0168	±.0125	±.0135	±.0079	±.0068	±.0100	±.0106
1000	2.2694	1.6095	1.6491	0.1762	0.6199	0.3725	0.3836
	±.0126	±.0063	±.0074	±.0057	±.0102	±.0120	±.0090
2000	2.4411	1.6638	1.6936	0.1418	0.5855	0.3281	0.3313
	±.0186	±.0131	±.0134	±.0038	±.0076	±.0068	±.0074
5000	2.7262	1.7545	1.7757	0.1110	0.5474	0.2718	0.2745
	±.0253	±.0092	±.0096	±.0032	±.0057	±.0054	±.0047
10000	2.9745	1.8197	1.8362	0.0880	0.5171	0.2396	0.2421
	±.0386	±.0159	±.0163	±.0023	±.0058	±.0057	±.0049
20000	3.3127	1.9032	1.9163	0.0706	0.4974	0.2085	0.2102
	±.0439	±.0134	±.0136	±.0024	±.0045	±.0048	±.0051

In Figure 1 we present the simulated mean values and MSE's of $\widehat{\gamma}_n^H(k)$, $\widehat{\gamma}_n^C(k)$ and $\widehat{\gamma}_n^{C_j}(k)$, $j = 1, 2$, for a sample size $n = 1000$ and for an underlying Fréchet parent with $\gamma = 1$. We present also the same results for the maximum likelihood censoring estimator $\widehat{\gamma}_n(k)$ in (3), which exhibits no bias.

A few general remarks:

1. The estimator $\widehat{\gamma}_n^C(k)$ reduces more drastically than expected the bias of the Hill estimator, for large values of k; at the optimal level, which is attained for large k, provides high efficiencies relatively to the Hill estimator, not only for the Fréchet model, but also for other simulated models, as shown in section 4.

2. The simplified estimators $\widehat{\gamma}_n^{C_1}(k)$ and $\widehat{\gamma}_n^{C_2}(k)$, although with a smaller rela-

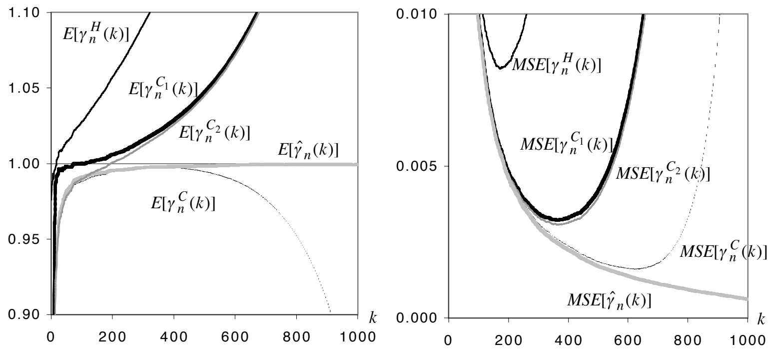

Figure 1: Simulated mean values (*left*) and *MSE's* (*right*) of $\widehat{\gamma}_n^H(k)$, $\widehat{\gamma}_n^C(k)$, $\widehat{\gamma}_n^{C_j}(k)$, $j = 1, 2$, and the maximum likelihood censoring estimator $\widehat{\gamma}_n(k)$ in (3), based on 5000 runs, for a sample size $n = 1000$ from a *Fréchet* parent with $\gamma = 1$.

tive efficiency than $\widehat{\gamma}_n^C(k)$ at their optimal levels, provide mean squared errors, $MSE[\widehat{\gamma}_n^{C_j}(k)]$, $j = 1,\ 2$, with an interesting bath-tube pattern, reasonably flat for a wide range of k-values, making less relevant the choice of the threshold.

3. The mean squared error of any of the censoring estimators is smaller than that of the Hill estimator at its optimal level, for a wide region of k-values.

3 The asymptotic normality of the estimators

As may be easily noticed from the explicit expressions of the estimators in (4), (5) and (6), their dominant component is always the Hill estimator, $\widehat{\gamma}_n^H(k)$, in (1), with a remaining term converging to 0 in probability. Apart from that main component, we have the r.v.'s $\frac{1}{k}\sum_{i=1}^{k}\left(X_{n-i+1:n}/X_{n-k:n}\right)^{-1/\gamma}$, and $\frac{1}{k}\sum_{i=1}^{k}\left(X_{n-i+1:n}/X_{n-k:n}\right)^{-1/\gamma}\ln\left(X_{n-i+1:n}X_{n-k:n}\right)$, to be studied here.

The study of these r.v.'s may be done in the lines of de Haan and Peng's (1998) derivation of asymptotic normality of the Hill estimator, under the second order framework, usual in *Extreme Value Theory* due to the large variety of models for which such a framework holds. Here, with $U(t) := F^{\leftarrow}(1 - 1/t)$, $t \geq 1$, $F(\cdot)$ the underlying model and $F^{\leftarrow}(\cdot)$ the generalized inverse function of F, we shall also assume that there exists a function $A(t)$ of constant sign, going to 0 as $t \to \infty$, such that

$$\lim_{t\to\infty}\frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}, \qquad (8)$$

for every $x > 0$, where ρ (≤ 0) is a *second order parameter*. The limit function in (8) must be of the stated form, and $|A(t)| \in RV_\rho$ (Geluk and de Haan, 1987).

We may then write, under the validity of (2) and (8),

$$\ln\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right) = \ln\left(\frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})}\right)$$
$$= \gamma \ln Y_{k-i+1:k} + A(n/k)\frac{Y_{k-i+1:k}^{\rho}-1}{\rho}(1+o_p(1)),$$

where $Y_{k-i+1:k},\ 1 \le i \le k$ are the o.s.'s associated to a standard Pareto(1) random sample of size k, with d.f $F_Y(y) = 1 - 1/y,\ y \ge 1$. Then, since $E\left[\frac{Y^{\rho}-1}{\rho}\right] = \frac{1}{1-\rho}$, we have for intermediate k, i.e., k such that (2) holds, the following distributional representation for the Hill estimator,

$$\widehat{\gamma}_n^H(k) = \gamma + \frac{\gamma}{\sqrt{k}} Z_n^H + \frac{1}{1-\rho} A(n/k)(1+o_p(1)), \tag{9}$$

where Z_n^H is an asymptotically standard normal r.v. Then, provided $\sqrt{k}A(n/k) \to \lambda$, finite, the Hill estimator is asymptotically normal. More precisely, $\sqrt{k}\left(\widehat{\gamma}_n^H(k) - \gamma\right)$ is asymptotically normal with a possibly non-null asymptotic bias, $\lambda/(1-\rho)$, and an asymptotic variance equal to γ^2.

Similarly, and with $U_i,\ 1 \le i \le k$, denoting i.i.d. standard uniform r.v.'s, and P_n and Q_n asymptotically standard normal r.v.'s,

$$\frac{1}{k}\sum_{i=1}^{k} Y_{k-i+1:k}^{-1} = \frac{1}{k}\sum_{i=1}^{k} U_i = \frac{1}{2} + \frac{1}{\sqrt{12k}}\ P_n(1+o_p(1)),$$

and

$$\frac{1}{k}\sum_{i=1}^{k} Y_{k-i+1:k}^{-1} \ln Y_{k-i+1:k} = -\frac{1}{k}\sum_{i=1}^{k} U_i \ln U_i = \frac{1}{4} + \frac{1}{12}\sqrt{\frac{5}{3k}}\ Q_n(1+o_p(1)).$$

Also, $E\left[Y^{-1}\frac{Y^{\rho}-1}{\rho}\right] = \frac{1}{2(2-\rho)}$ and $E\left[Y^{-1}\frac{Y^{\rho}-1}{\rho}(1-\ln Y)\right] = -\frac{\rho}{4(2-\rho)^2}$.

We thus get, for intermediate k,

$$\Phi(\gamma) = \Phi(\gamma, \underline{X}_k) \equiv \frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{-1/\gamma}$$
$$\stackrel{d}{=} \frac{1}{2} + \frac{1}{\sqrt{12k}}\ P_n - \frac{1}{2\gamma(2-\rho)} A(n/k)(1+o_p(1)), \tag{10}$$

and

$$\Psi(\gamma) = \Psi(\gamma, \underline{X}_k) \equiv \frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{-1/\gamma} \ln\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)$$
$$\stackrel{d}{=} \frac{\gamma}{4} + \frac{\gamma}{12}\sqrt{\frac{5}{3k}}\ Q_n - \frac{\rho}{4\gamma(2-\rho)^2} A(n/k)(1+o_p(1)). \tag{11}$$

It thus follows that we have for the estimators under study the same type of asymptotically normal behaviour of the Hill estimator, eventually with a non-null asymptotic bias, but the value of that asymptotic bias is going to be slightly different from that of the Hill estimator. Noticing that, for intermediate k, $\Phi(\widehat{\gamma}_n^H(k)) = \Phi(\gamma) + O_p(1/\sqrt{k})$ and $\Psi(\widehat{\gamma}_n^H(k)) = \Psi(\gamma) + O_p(1/\sqrt{k})$ converge in probability towards $1/2$ and $\gamma/4$, respectively, as $n \to \infty$, the second term in either (4) or (5) or (6) is of the order of k/n, and the distributional representation

$$\widehat{\gamma}_n^\bullet(k) = \gamma + \left(\frac{\gamma}{\sqrt{k}} Z_n^H - \frac{\gamma}{4}\,\frac{k}{n} + \frac{1}{1-\rho} A(n/k)\right)(1 + o_p(1)) \qquad (12)$$

holds for any of the estimators in (4), (5) and (6), here denoted by $\widehat{\gamma}_n^\bullet(k)$.

If we take into account only the dominant component of asymptotic bias, we have for every fixed n and k an asymptotic approximation for the mean squared error of any of the censoring estimators $\widehat{\gamma}_n^\bullet(k)$ given by

$$AMSE_\rho^\bullet(n,k) = \begin{cases} \frac{\gamma^2}{k} + \frac{\gamma^2}{16}\,\left(\frac{k}{n}\right)^2 & \text{if} \quad \rho < -1 \\ \frac{\gamma^2}{k} + \left(\frac{1}{1-\rho} A(n/k) - \frac{\gamma}{4}\,\frac{k}{n}\right)^2 & \text{if} \quad \rho = -1 \\ \frac{\gamma^2}{k} + \left(\frac{1}{1-\rho} A(n/k)\right)^2 & \text{if} \quad \rho > -1 \end{cases}$$

Notice however that the approximation given for $\rho = -1$ is more accurate than any of the other two, for any value of ρ. Indeed, had we decided to go into a higher order development of $\ln U(t)$, would we get better and better approximations. For results related to the role of the third order behaviour in Statistics of Extremes see for instance Gomes and de Haan (1999).

A few additional remarks:

4. Whenever $\rho < -1$, the dominant term of the asymptotic bias is thus negative, and given by $-\frac{\gamma}{4}\,\frac{k}{n}$ (a term of the order of k/n dominates then $A(n/k)$, which is of the order of $(k/n)^{-\rho}$). Asymptotically, the estimator is then worse than the Hill estimator.

5. If $\rho = -1$, $A(n/k) = O(k/n)$, and we have a dominant term of asymptotic bias equal to $\left\{\frac{1}{1-\rho} A(n/k) - \frac{\gamma}{4}\,\frac{k}{n}\right\}$; then we are for sure going to have a decreasing in bias relatively to the Hill estimator, possibly in a wrong direction if the function $A(\cdot)$ is a negative function; but we may even have for very specific models, like for instance the Fréchet model where $\rho = -1$ and $A(t)$ may be chosen equal to $\gamma/(2t)$, a null dominant component for the asymptotic bias.

6. If $\rho > -1$, the dominant component of asymptotic bias is equal to the one we get for the Hill estimator, $\frac{1}{1-\rho} A(n/k)$.

We picture here as an illustration of the behaviour of the three branches of $AMSE_\rho^\bullet(n,k)$, now expressed in terms of the sample fraction $s = k/n$, and for $n = 1000$ and $n = 5000$, the three functions $\varphi_1(s)/\gamma^2 = \frac{1}{ns} + \frac{s^2}{16}$, $\varphi_2(s)/\gamma^2 = \frac{1}{ns} + \left(\frac{s^{-\rho}}{1-\rho} - \frac{s}{4}\right)^2$, and $\varphi_3(s)/\gamma^2 = \frac{1}{ns} + \left(\frac{s^{-\rho}}{1-\rho}\right)^2$, for $0 < s < 1$, $\rho = -2$. The function A was here chosen equal to $A(t) = \gamma\, t^\rho$, and is the second order function associated to a Burr model, $F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho},\ x \ge 0,\ \gamma > 0,\ \rho < 0$.

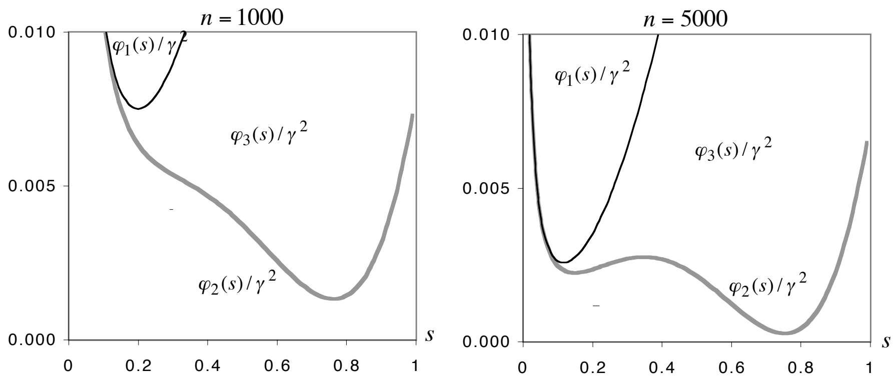

Figure 2: Asymptotic mean squared error patterns of the estimators under study, for a *Burr* parent with $\rho = -2$.

4 Robustness of the estimators — a simulation study

Figures 3, 4 and 5 are analogue to Figure 1, and also based on 5000 runs, but for *Burr* models, with $\rho = -0.5, -1$ and -2, respectively.

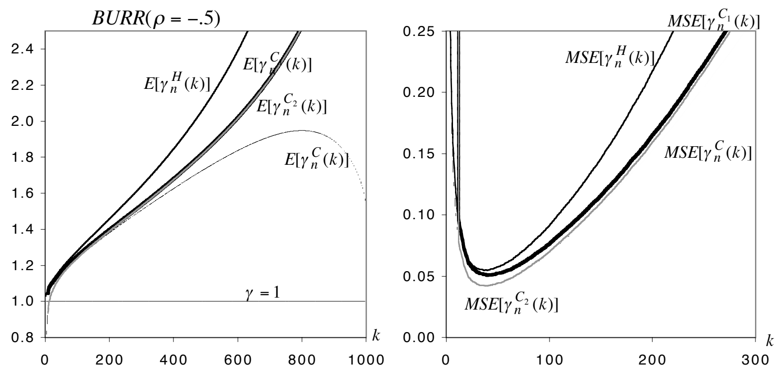

Figure 3: Simulated mean values (*left*) and *MSE's* (*right*) of $\widehat{\gamma}_n^H(k)$, $\widehat{\gamma}_n^C(k)$ and $\widehat{\gamma}_n^{C_j}(k)$, $j = 1, 2$, based on 5000 runs, for a sample size $n = 1000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -.5$.

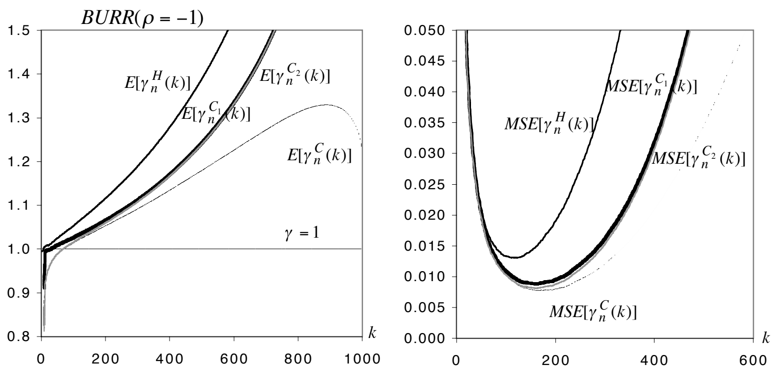

Figure 4: Simulated mean values (*left*) and *MSE's* (*right*) of $\widehat{\gamma}_n^H(k)$, $\widehat{\gamma}_n^C(k)$ and $\widehat{\gamma}_n^{C_j}(k)$, $j = 1, 2$, based on 5000 runs, for a sample size $n = 1000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -1$.

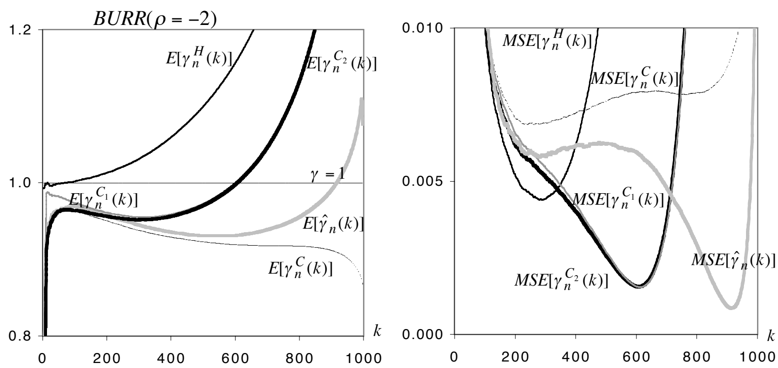

Figure 5: Simulated mean values (*left*) and *MSE's* (*right*) of $\widehat{\gamma}_n^H(k)$, $\widehat{\gamma}_n^C(k)$, $\widehat{\gamma}_n^{C_j}(k)$, $j = 1, 2$, and $\widehat{\gamma}_n(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -2$.

Only for $\rho = -2$ do we picture the behaviour of the estimator in (3), which is then quite different from that of $\widehat{\gamma}_n^C(k)$, contrarily to what happens for models with $\rho \geq -1$, different from the Fréchet model. Notice that the exact behaviour of the estimators, which are asymptotically similar, is quite diversified. Indeed, the rate of convergence of any of the estimators under study is of the order of $1/\sqrt{k}$, and for not too large k (and n), the exact behaviour of the estimators may be a long way from the asymptotic one. Notice however the similarities between the MSE of $\widehat{\gamma}_n(k)$ in Figure 5, and the asymptotic behaviour pictured in Figure 2.

We next present simulation results, based again on 10 replicates of 5000 runs each, to evaluate the robustness of the estimators at their optimal levels, i.e., the estimators $\widehat{\gamma}_{n,o}^C = \widehat{\gamma}_n^C(k_o^C(n))$, and $\widehat{\gamma}_{n,o}^{(C_j)} = \widehat{\gamma}_n^{C_j}(k_o^{C_j}(n))$, $j = 1, 2$. The measure of comparison is the relative efficiency explicited in (7). In Table 3 we present these relative efficiencies for different models with a second order parameter $\rho = -1$, different from the *Fréchet*: the *Student-t* with $\nu = 2$ degrees of freedom, a *Burr* model, with $\gamma = -\rho = 1$, a model, denoted *Out-Hall*, also considered in Drees and Kaufmann (1998), with a quantile function $F^{\leftarrow}(1 - t) = t^{-1} e^{-2t(\ln t - 1)}$, for all $0 < t \leq 1$, which does not belong to Hall's class (Hall and Welsh (1985)), and finally two models, also considered by Gomes *et al.* (2000), one denoted *Sin-Fréchet*, with a quantile function $F^{\leftarrow}(1 - t) = \left(-\frac{1}{\sin(1/t)} \ln(1 - t\sin(1/t))\right)^{-1}$, $0 < t \leq 1$, close to the *Fréchet* parent with $\gamma = 1$, and another related to the *Burr* model, denoted *Sin-Burr*, with a quantile function $F^{\leftarrow}(1 - t) = (t^{\rho} - \sin(t^{\rho}))^{-\gamma/\rho}$, $0 < t \leq 1$. For these two models the second order condition (8) no longer holds.

Table 3: Simulated efficiencies of $\widehat{\gamma}^{\bullet}_{n,o}$ relatively to $\widehat{\gamma}^{H}_{n,o}$ for models with $\rho = -1$.

n	100	200	500	1000	2000	5000	10000	20000
Burr parent: $\rho = -1$, $\gamma = 1$								
C	1.3903	1.3538	1.3191	1.3025	1.2914	1.2784	1.2716	1.2660
	±.0132	±.0105	±.0064	±.0087	±.0073	±.0039	±.0049	±.0091
C_1	1.1939	1.2066	1.2110	1.2192	1.2268	1.2315	1.2343	1.2361
	±.0091	±.0068	±.0049	±.0088	±.0070	±.0027	±.0041	±.0090
C_2	1.3295	1.3057	1.2784	1.2706	1.2662	1.2595	1.2558	1.2529
	±.0115	±.0080	±.0047	±.0094	±.0072	±.0028	±.0042	±.0092
Student ($\nu = 2$) parent: $\rho = -1$, $\gamma = .5$								
C	1.2784	1.2164	1.1650	1.1379	1.1221	1.1004	1.0912	1.0833
	±.0061	±.0060	±.0036	±.0036	±.0024	±.0020	±.0026	±.0023
C_1	1.0561	1.0565	1.0584	1.0584	1.0615	1.0582	1.0588	1.0582
	±.0035	±.0041	±.0026	±.0029	±.0019	±.0017	±.0025	±.0020
C_2	1.2743	1.2129	1.1624	1.1357	1.1204	1.0991	1.0902	1.0825
	±.0056	±.0057	±.0035	±.0035	±.0024	±.0019	±.0026	±.0023
Out-Hall parent: $\rho = -1$, $\gamma = 1$								
C	0.7675	0.7981	0.8433	0.8708	0.8944	0.9176	0.9321	0.9421
	±.0044	±.0037	±.0015	±.0022	±.0010	±.0012	±.0009	±.0014
C_1	0.4397	0.7620	0.8927	0.9249	0.9446	0.9590	0.9678	0.9719
	±.0057	±.0263	±.0033	±.0029	±.0012	±.0011	±.0008	±.0017
C_2	0.4364	0.7152	0.8328	0.8682	0.8938	0.9176	0.9322	0.9422
	±.0040	±.0221	±.0027	±.0025	±.0010	±.0012	±.0009	±.0015
Sin-Fréchet parent: $\gamma = 1$								
C	1.9223	1.9126	1.9196	1.9161	1.9238	1.9252	1.9260	1.9283
	±.0303	±.0250	±.0265	±.0206	±.0162	±.0147	±.0181	±.0125
C_1	1.1639	1.1571	1.1555	1.1519	1.1522	1.1529	1.1529	1.1530
	±.0056	±.0035	±.0068	±.0083	±.0095	±.0066	±.0071	±.0055
C_2	1.2346	1.1945	1.1706	1.1594	1.1562	1.1543	1.1539	1.1534
	±.0081	±.0053	±.0071	±.0080	±.0096	±.0068	±.0072	±.0056
Sin-Burr parent: $\rho = -1$, $\gamma = 1$								
C	1.0975	1.0753	1.0728	1.0663	1.0681	1.0659	1.0641	1.0657
	±.0078	±.0035	±.0057	±.0053	±.0078	±.0046	±.0052	±.0039
C_1	1.0265	1.0288	1.0407	1.0398	1.0463	1.0455	1.0436	1.0451
	±.0056	±.0044	±.0052	±.0051	±.0073	±.0045	±.0054	±.0035
C_2	1.0688	1.0510	1.0501	1.0438	1.0487	1.0465	1.0441	1.0452
	±.0059	±.0033	±.0049	±.0051	±.0070	±.0045	±.0054	±.0036

Apart the *Out-Hall* model, where $\widehat{\gamma}^{C_1}_{n,o}$ overpasses $\widehat{\gamma}^{C}_{n,o}$ for $n \geq 500$, the highest efficiency is always attained by $\widehat{\gamma}^{C}_{n,o}$, as expected, but not a long way from the other estimators. In Table 4 we present similar results for *Burr*, *Student* and *Sin-Burr* models, but now with values of $\rho > -1$, more specifically, for $\rho = -0.25$ and -0.5.

Table 4: Simulated efficiencies of $\widehat{\gamma}^{\bullet}_{n,o}$ relatively to $\widehat{\gamma}^{H}_{n,o}$ for models with $\rho = -0.25$ and -0.5

n	100	200	500	1000	2000	5000	10000	20000
Burr parent: $\rho = \mathbf{0.25}$, $\gamma = \mathbf{1}$								
C	1.5011	1.3832	1.2815	1.2245	1.1859	1.1410	1.1205	1.0978
	±.0098	±.0093	±.0076	±.0065	±.0068	±.0043	±.0032	±.0023
C_1	0.6906	1.0152	1.0113	1.0085	1.0067	1.0047	1.0035	1.0027
	±.3408	±.0010	±.0009	±.0005	±.0002	±.0003	±.0003	±.0003
C_2	0.9715	1.3768	1.2795	1.2241	1.1859	1.1410	1.1205	1.0978
	±.4899	±.0075	±.0086	±.0067	±.0068	±.0042	±.0032	±.0023
Student ($\nu = 8$) parent: $\rho = \mathbf{-0.25}$, $\gamma = \mathbf{0.125}$								
C	1.5201	1.4363	1.3277	1.2408	1.1895	1.1343	1.1153	1.0932
	±.0508	±.0210	±.0046	±.0099	±.0056	±.0054	±.0074	±.0054
C_1	0.9138	0.9422	0.9559	0.9581	0.9751	0.9709	0.9823	0.9819
	±.0213	±.0169	±.0109	±.0215	±.0069	±.0061	±.0083	±.0043
C_2	1.1929	1.1904	1.1501	1.1229	1.1328	1.0902	1.0938	1.0771
	±.0503	±.0436	±.0220	±.0416	±.0161	±.0127	±.0147	±.0076
Sin-Burr parent: $\rho = \mathbf{-0.25}$, $\gamma = \mathbf{1}$								
C	1.5784	1.4683	1.3102	1.1859	1.0878	1.0347	1.0184	1.0084
	±.0354	±.0711	±.0343	±.0139	±.0087	±.0065	±.0024	±.0016
C_1	0.9582	0.9056	0.8287	0.8642	0.8832	0.9638	0.9877	0.9947
	±.2247	±.0155	±.0169	±.0410	±.0463	±.0202	±.0015	±.0010
C_2	1.3220	1.2733	1.1043	1.0903	1.0358	1.0286	1.0181	1.0084
	±.3204	±.0436	±.0381	±.0548	±.0508	±.0137	±.0025	±.0016
Burr parent: $\rho = \mathbf{0.5}$, $\gamma = \mathbf{1}$								
C	1.3071	1.2405	1.1765	1.1426	1.1170	1.0895	1.0738	1.0609
	±.0031	±.0027	±.0017	±.0010	±.0007	±.0007	±.0009	±.0007
C_1	1.0597	1.0521	1.0454	1.0389	1.0357	1.0289	1.0235	1.0197
	±.0029	±.0029	±.0016	±.0020	±.0019	±.0013	±.0014	±.0009
C_2	1.3027	1.2379	1.1751	1.1418	1.1164	1.0892	1.0737	1.0608
	±.0027	±.0026	±.0016	±.0009	±.0006	±.0007	±.0008	±.0006
Student ($\nu = 4$) parent: $\rho = \mathbf{-0.5}$, $\gamma = \mathbf{0.25}$								
C	1.4178	1.3107	1.2205	1.1733	1.1375	1.1033	1.0836	1.0683
	±.0053	±.0040	±.0034	±.0019	±.0021	±.0017	±.0011	±.0010
C_1	0.9981	1.0170	1.0157	1.0151	1.0136	1.0122	1.0110	1.0092
	±.0134	±.0026	±.0017	±.0017	±.0008	±.0010	±.0008	±.0003
C_2	1.3315	1.2995	1.2188	1.1726	1.1372	1.1032	1.0836	1.0683
	±.0387	±.0087	±.0030	±.0018	±.0020	±.0017	±.0011	±.0010
Sin-Burr parent: $\rho = \mathbf{-0.5}$, $\gamma = \mathbf{1}$								
C	1.1044	1.0442	1.0256	1.0165	1.0120	1.0100	1.0097	1.0106
	±.0044	±.0081	±.0049	±.0028	±.0020	±.0019	±.0025	±.0020
C_1	0.8189	0.9103	0.9904	0.9994	1.0030	1.0061	1.0076	1.0090
	±.0355	±.0227	±.0045	±.0023	±.0022	±.0024	±.0024	±.0022
C_2	0.9787	1.0075	1.0236	1.0157	1.0115	1.0093	1.0092	1.0100
	±.0477	±.0213	±.0049	±.0028	±.0020	±.0021	±.0025	±.0020

Table 5 is the equivelent of Table 4, but for *Burr* and *Student* models with $\rho = -2$. Here, and for the Burr parent, both $\widehat{\gamma}^{C_1}_{n,o}$ and $\widehat{\gamma}^{C_2}_{n,o}$ behave better than $\widehat{\gamma}^{C}_{n,o}$.

Table 5: Simulated efficiencies of $\widehat{\gamma}^{\bullet}_{n,o}$ relatively to $\widehat{\gamma}^{H}_{n,o}$ for models with $\rho = -2$.

n	100	200	500	1000	2000	5000	10000	20000
Burr parent: $\rho = -\mathbf{2}$, $\gamma = \mathbf{1}$								
C	1.2710	1.1375	0.9161	0.8129	0.7569	0.7018	0.6636	0.6294
	±.0184	±.0146	±.0091	±.0070	±.0038	±.0083	±.0031	±.0046
C_1	1.4163	1.4777	1.5952	1.6905	1.8077	1.9642	2.1054	2.2572
	±.0155	±.0106	±.0146	±.0156	±.0157	±.0134	±.0130	±.0227
C_2	1.4604	1.5061	1.6096	1.6983	1.8126	1.9663	2.1062	2.2579
	±.0156	±.0116	±.0145	±.0154	±.0154	±.0134	±.0129	±.0226
Student ($\nu = 1$) **parent:** $\rho = -\mathbf{2}$, $\gamma = \mathbf{1}$								
C	1.2529	1.2427	1.2522	1.2788	1.2985	1.3521	1.4017	1.4652
	±.0078	±.0060	±.0087	±.0066	±.0074	±.0073	±.0133	±.0084
C_1	1.1200	1.1418	1.1781	1.2194	1.2455	1.3066	1.3565	1.4215
	±.0044	±.0047	±.0058	±.0051	±.0065	±.0073	±.0102	±.0064
C_2	1.2348	1.2222	1.2289	1.2547	1.2706	1.3215	1.3672	1.4284
	±.0075	±.0055	±.0071	±.0058	±.0067	±.0074	±.0108	±.0064

A few additional and final remarks:

7. The estimators seem to be appropriate to reduce bias whenever $\rho = -1$.

8. For values of $\rho > -1$, here illustrated with $\rho = -0.25$ and $\rho = -0.5$, and just like the developed theory suggests, the reduction in bias is not so significant. Even so, there is a slight decrease in mean squared error.

9. For values of $\rho < -1$, here illustrated with $\rho = -2$, it is not possible from the theoretical results in section 3 to predict what is really going to happen in practice. There is for sure a high reduction in bias, which may be in a wrong direction, increasing the mean squared error of the estimators. However, and as it is illustrated in Table 5 for Burr and Student models, apart from $\gamma^{C}_{n,o}$ in a Burr model, these estimators are highly efficient in this region of ρ-values, a region where has been difficult to find competitors to the Hill estimator. This is mainly due to a second local minumum of the mean squared error pattern, which is the global minimum. For values of $n \leq 1000$ it is possible to detect only a unique minimum (see Figures 2 (*left*) and 5). However, even for $n = 1000$, the mean squared error of $\widehat{\gamma}_n(k)$ has already two local minima (see again Figure 5).

10. The results obtained for *Sin-Fréchet* and *Sin-Burr* parents, models for which the second order condition (8) no longer holds, give us some hope that the developed estimators are robust against a non-second order framework, and this is an interesting result, since to date there is no statistical procedure to assess the validity of second order regular variation submodels. Anyway, the *Mean values*' and *MSE*'s patterns of the estimators (not shown here) are, like for Hill's estimator, still sinusoidal, with several humps and bumps.

11. Only for models outside Hall's class, not shown here graphically as well, i.e., models whose tail is not of the type $1 - F(x) =$

$Ax^{-1/\gamma}\left(1 + Bx^{\rho/\gamma}(1 + o(1))\right)$, do the estimators not behave better than the Hill estimator, due to a decreasing in bias, which was already negative for the Hill estimator and for the models simulated.

References

[1] Drees, H. and E. Kaufmann, 1998. Selecting the optimal sample fraction in univariate extreme value estimation. Stoch. Proc. and Appl. 75, 149-172.

[2] Galambos, J., 1987. The Asymptotic Theory of Extreme Order Statistics, 2nd ed. Krieger.

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

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

[5] Gomes, M. I. and O. Oliveira, 2000. The bootstrap methodology in Statistical Extremes — choice of the optimal sample fraction. Notas e Comunicações CEAUL 04/00.

[6] 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.

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

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

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

[10] Mason, D.M., 1982. Laws of large numbers for sums of extreme values. Ann. Probab. 10, 754-774.