Bias reduction and efficient estimation of the tail index*

M. Ivette Gomes
D.E.I.O. and C.E.A.U.L., Universidade de Lisboa

Abstract. In this paper, and in a context of regularly varying tails, we analyse two particular but interesting cases of the *Maximum Likelihood* (ML) and *Least Squares* (LS) estimators proposed by Feuerverger and Hall (1999). These estimators are alternatives to a well-known estimator of the tail index, the *Hill estimator* (Hill, 1975), and jointly with the *Generalized Jackknife* estimators in Gomes et al.(1998, 2001) have essentially in mind a reduction in bias, preferably without increasing mean square error, leading to more efficient estimators of the *tail index*, provided we may use extreme-value data relatively deep into the sample. Classical ML estimators and quasi-ML estimators are proposed under a semi-parametric set-up, and studied computationally.

AMS 2000 subject classification. Primary 62G32, 62E20; Secondary 65C05.
Keywords and phrases. *Statistical Theory of Extremes, Semi-parametric estimation, Maximum Likelihood and Least Squares 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 $t \to \infty$. The d.f. F is thus in the max-domain of attraction of an *Extreme Value* (EV) d.f.

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

with $\gamma > 0$.

Then, whith $X_{i:n}$ denoting the i-th ascending order statistic (o.s.) associated to the sample $\underline{X}_n = (X_1, X_2, \ldots, X_n)$, the spacings

$$U_i = i\left[\ln X_{n-i+1:n} - \ln X_{n-i:n}\right],\ 1 \le i \le k, \tag{1.3}$$

*Research partially supported by FCT / POCTI / FEDER.

are, for $k = k_n = o(n)$, as $n \to \infty$, approximately independent and exponential with mean value

$$\mu_i = \gamma\, e^{\frac{A(n/k)}{\gamma}\left(\frac{i}{k}\right)^{-\rho}} = \gamma + A(n/k)\left(\frac{i}{k}\right)^{-\rho}(1 + o(1)), \tag{1.4}$$

as $n \to \infty$, where $A(.)$ and $\rho < 0$ are related to the second order behaviour of F (Draisma, 2000, pages 43-59). Indeed, 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 regularly varying functions at infinity with index of regular variation equal to α, i.e., functions $g(\cdot)$ with infinite right endpoint, and such that $\lim_{t\to\infty} g(tx)/g(t) = x^\alpha$, for all $x > 0$, and the notation $F^{\leftarrow}(\cdot)$ is used for the generalized inverse function of F, i.e., $F^{\leftarrow}(t) = \inf\{x : F(x) \geq t\}$. The function $A(t)$ measures the rate of convergence of $U(tx)/U(t)$ towards x^γ, and is a function of constant sign, such that

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}, \tag{1.5}$$

for every $x > 0$, where $\rho\ (\leq 0)$ is a *second order parameter.* The limit function in (1.5) must be of the stated form, and $|A(t)| \in RV_\rho$ (Geluk and de Haan, 1987).

For Hall's class of models (Hall and Welsh (1985)), where we may choose $A(t) = Ct^\rho$, $\rho < 0$, Hall and Feuerverger(1999) work then with the spacings U_i as approximately independent and exponential with mean value

$$\mu_i = \gamma \exp\{D(i/n)^{-\rho}\}, \quad 1 \leq i \leq k, \tag{1.6}$$

where $D = C/\gamma$, i.e., they show that the main effects of bias may be accommodated by modelling the scale of log-spacings of o.s.'s, and suggest the use of either maximum likelihood or least-squares methodologies for the joint estimation of γ, ρ and D. In this paper, and in order to obtain explicit expressions for the estimators of the tail index γ, we assume ρ known and equal to -1 and study, under a general context, the maximum likelihood and the least squares estimators of the tail index γ, comparing them with one of the Generalized Jackknife estimators proposed in Gomes et al. (1998, 2001), the one specifically devised for Fréchet models, denoted here by

$$\gamma_n^{GJ}(k) := \frac{\gamma_n^H(k) - \frac{\ln(1-k/n)}{\ln(1-k/2n)}\gamma_n^H(k/2)}{1 - \frac{\ln(1-k/n)}{\ln(1-k/2n)}},\ 2 \leq k \leq n-1, \tag{1.7}$$

where

$$\gamma_n^H(k) := \frac{1}{k}\sum_{i=1}^{k}\ [lnX_{n-i+1:n} - lnX_{n-k:n}] \equiv \frac{1}{k}\sum_{i=1}^{k} U_i,\ 1 \leq k \leq n-1, \tag{1.8}$$

is the Hill estimator for γ (Hill, 1975).

2 Maximum likelihood and Least Squares estimators

The log-likelihood of the k-spacings U_i, $1 \le i \le k$, is

$$\ln L(\gamma, D, \rho; U_i, 1 \le i \le k) = -k \ln \gamma - D \sum_{i=1}^{k} \left(\frac{i}{n}\right)^{-\rho} - \frac{1}{\gamma} \sum_{i=1}^{k} U_i e^{-D\left(\frac{i}{n}\right)^{-\rho}}. \quad (2.1)$$

Since the maximization of such a function is not an easy computational job, Feuerverger and Hall (1999), in the data analysis performed, suggest to assume ρ known and equal to -1. We do it here from the beginning, and we then explicitly get

$$\widehat{\gamma} = \frac{1}{k} \sum_{i=1}^{k} U_i e^{-\frac{i}{n}\widehat{D}},$$

where $\widehat{D}$ is such that

$$\sum_{i=1}^{k} U_i e^{-\frac{i}{n}\widehat{D}} \left(i - \frac{k+1}{2}\right) = 0.$$

Now, if we take the first order approximation $\{1 - \frac{i}{n}D\}$ for $e^{-\frac{i}{n}D}$, we get the explicit expression

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

which will be studied later on, under a general framework.

Notice that $\frac{1}{k}\sum_{i=1}^{k} U_i$ is the Hill estimator, and consequently

$$\gamma_n^{ML}(k) = \gamma_n^{H}(k) - \left(\frac{1}{k} \sum_{i=1}^{k} iU_i\right) \frac{\sum_{i=1}^{k} (2i-k-1)U_i}{\sum_{i=1}^{k} i(2i-k-1)U_i}. \quad (2.2)$$

Let us consider next the least squares approach, also developed in Feuerverger and Hall (1999). Since we may write, approximately,

$$U_i = \gamma e^{D(n/i)^{\rho}} E_i, \text{ with } E_i \text{ i.i.d. unit exponential,}$$

then

$$\ln U_i = \ln \gamma + D(n/i)^{\rho} + \Gamma'(1) + \epsilon_i, \quad (2.3)$$

where the ϵ_i are i.i.d. r.v's with null mean value, $\Gamma(\cdot)$ is the complete gamma function, $\Gamma'(\cdot)$ its derivative, being thus $\Gamma'(1)$ the symmetric of Euler's constant, i.e. $\Gamma'(1) = -0.577216$. It is thus sensible to proceed to the estimation of the unknown parameters through the solution of the following minimization problem:

$$\underset{D,\gamma,\rho}{Minimize} \sum_{i=1}^{k} \left[\ln U_i - \ln\gamma - \Gamma'(1) - D(n/i)^{\rho}\right]^2. \tag{2.4}$$

Assuming again $\rho = -1$ we get an explicit expression for the least-squares estimator of γ, which is given by

$$\gamma_n^{LS}(k) := \exp\left\{\frac{2(2k+1)}{k(k-1)}\sum_{i=1}^{k}\ln U_i - \Gamma'(1) - \frac{6}{k(k-1)}\sum_{i=1}^{k} i\ln U_i\right\}. \tag{2.5}$$

3 The asymptotic behaviour of the estimators

Whenever the second order condition (1.5) holds and the threshold k is intermediate, i.e.,

$$k = k_n \to \infty, \quad k_n/n \to 0, \text{ as } n \to \infty, \tag{3.1}$$

we have the asymptotic distributional representations

$$\frac{\alpha}{k^\alpha}\sum_{i=1}^{k} i^{\alpha-1}U_i = \gamma + \frac{\gamma\alpha}{\sqrt{(2\alpha-1)k}}\, P_n^{(\alpha)} + \frac{\alpha A(n/k)}{\alpha-\rho} + o_p(A(n/k)), \tag{3.2}$$

and

$$\frac{\alpha}{k^\alpha}\sum_{i=1}^{k} i^{\alpha-1}\ln U_i = \ln\gamma + \Gamma'(1) + \frac{\pi\alpha}{\sqrt{6(2\alpha-1)k}}\, Q_n^{(\alpha)} + \frac{\alpha A(n/k)}{\gamma(\alpha-\rho)} + o_p(A(n/k)), \tag{3.3}$$

where $P_n^{(\alpha)}$ and $Q_n^{(\alpha)}$ are asymptotically standard normal r.v.'s.

From the representations (3.2), and from the fact that we may write (6.5) as

$$\gamma_n^{ML}(k) = \gamma_n^{H}(k) - \left(\frac{1}{k^2}\sum_{i=1}^{k} iU_i\right)\frac{\frac{1}{k^2}\sum_{i=1}^{k}(2i-k-1)U_i}{\frac{1}{k^3}\sum_{i=1}^{k} i(2i-k-1)U_i}$$

where $\frac{1}{k^2}\sum_{i=1}^{k} i\, U_i$ and $\frac{1}{k^3}\sum_{i=1}^{k} i(2i-k-1)U_i$ converge in probability towards $\gamma/2$ and $\gamma/6$, respectively, we have

$$\gamma_n^{ML}(k) = \gamma_n^{H}(k) - \frac{3\rho}{(1-\rho)(2-\rho)}A(n/k)(1+o_p(1)),$$

and consequently,

$$\gamma_n^{ML}(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, P_n^{(1)} + \frac{2(1+\rho)}{(1-\rho)(2-\rho)} A(n/k) + o_p(A(n/k)). \tag{3.4}$$

From the representations (3.3), for $\alpha = 1$ and $\alpha = 2$, and from the fact that the asymptotic covariance between $Q_n^{(1)}$ and $Q_n^{(2)}$ is equal to $\sqrt{3}/2$, it follows that

$$\gamma_n^{LS}(k) \stackrel{d}{=} \gamma + \frac{2\gamma\pi}{\sqrt{6k}}\, Q_n + \frac{2(1+\rho)}{(1-\rho)(2-\rho)} A(n/k) + o_p(A(n/k)). \tag{3.5}$$

The dominant component of bias is in both cases null only for $\rho = -1$, but the finite sample behaviour of $\gamma_n^{ML}(k)$ is quite interesting for all values of $\rho > -1$, as we shall see in section 4.

For comparison, we refer that we have

$$\gamma_n^{H}(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} P_n^{(1)} + \frac{1}{1-\rho} A(n/k) + o_p(A(n/k)) \text{ (de Haan and Peng, 1998)},$$

and

$$\gamma_n^{GJ}(k) \stackrel{d}{=} \gamma + \frac{\gamma\sqrt{5}}{\sqrt{k}} Z_n^{GJ} + \frac{2^{\rho+1}-1}{1-\rho} A(n/k) + o_p(A(n/k)) \text{ (Gomes et al., 1998)},$$

where Z_n^{GJ} is also asymptotically standard normal.

If we now proceed to an asymptotic comparison of the estimators at their optimal levels in the lines of de Haan and Peng (1998), and also Gomes et al. (1998) for a set of Generalized Jackknife statistics, we get, for $\rho \neq -1$, the following asymptotic efficiences relatively to the Hill estimator:

$$AEFF_{ML|H} = \left(\frac{2-\rho}{2|1+\rho|}\right)^{\frac{1}{1-2\rho}}, \tag{3.6}$$

and

$$AEFF_{LS|H} = \left(\left(\frac{4\pi^2}{6}\right)^{\rho} \frac{2-\rho}{2|1+\rho|}\right)^{\frac{1}{1-2\rho}}, \tag{3.7}$$

i.e.

$$AEFF_{ML|H} > 1 \quad \Longleftrightarrow \quad \rho > -4,\ \rho \neq 1,$$

but

$$AEFF_{LS|H} > 1 \quad \Longleftrightarrow \quad -1.1738 < \rho < -0.5425,\ \rho \neq 1.$$

The smallest asymptotic mean square error is achieved by the ML estimator for values of $|\rho| \leq 4$; for $|\rho| > 4$ the Hill estimator overpassees asymptotically all the other estimators. The asymptotic mean square error of the LS-estimator is bigger than that of the Generalized Jackknife statistic herewith considered, for every ρ. Comparatively to the Hill estimator, the LS-estimator performs asymptotically better for values of ρ in the interval $(-1.1738, -0.5425)$, $\rho \neq -1$. Notice that the GJ statistic, which has the same limit distribution as the Generalized Jackknife statistic $\gamma_n^{G_4}$ in Gomes et al. (1998), behaves asymptotically better than the Hill estimator for values of ρ in the interval $(-1.2187, -0.3491)$, $\rho \neq -1$.

In Figure 1 we picture the asymptotic relative efficiencies of the estimators under study together with that of the GJ-statistic.

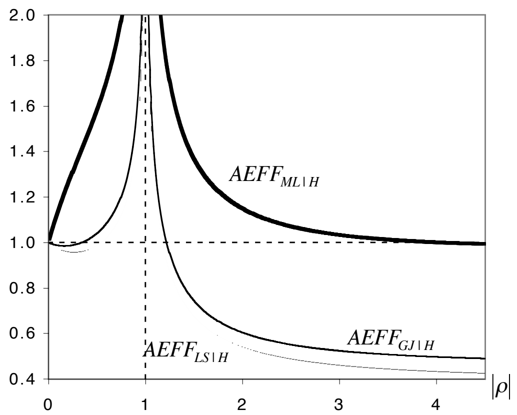

Figure 1: Asymptotic efficiencies of γ_n^{ML}, γ_n^{LS} and γ_n^{GJ}, relatively to the Hill estimator, all computed at their optimal levels.

If we distinguish both components of the mean square error, the squared bias and the variance, we may say that the asymptotic squared bias of the Hill estimator is bigger than that of the GJ statistics for all values of ρ, and is bigger than that of either the ML or the LS statistic only for $|\rho| < .25$. For values of $|\rho| \leq 1$ the asymptotic squared bias of the GJ statistic is smaller than that of either the ML or the LS statistic, and things work the other way round for $|\rho| > 1$. But whereas the asymptotic variance of the ML estimator is equal to that of the Hill estimator, the GJ statistic has an asymptotic variance 5 times greater than that of the Hill estimator, and the asymptotic variance of the LS-estimator is even bigger than that of the GJ-statistic. This gives rise to an overall better asymptotic behaviour for the ML-estimator, which is not generally exhibited for finite samples, as we shall see in the following section.

4 Finite sample properties and robustness of the estimators — a simulation study

We shall consider in this section the finite sample properties of the above mentioned estimators of the tail index, for the following set of models in Hall's class of distributions, with a second order parameter $\rho < 0$,

1. the *Fréchet* model, $F(x) = \exp(-x^{-1/\gamma}),\ x \geq 0$, with $\gamma = 1$, 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 $\gamma = 1$ and for $\rho = -0.25, -0.5, -1, -2$;
3. the *Student-t* model with $\nu = 8, 4, 2, 1$ degrees of freedom, for which $\gamma = .125, .25, .5, 1$ and $\rho = -0.25, -.5, -1, -2$, respectively;

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 $\rho = -1$

and the following models for which the second order condition in (1.5) does not hold:

5. the *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
6. the *sin-Burr* model, with a quantile function $F^{\leftarrow}(1-t) = (t^\rho - \sin(t^\rho))^{-\gamma/\rho}$, $0 < t \leq 1$, with $\gamma = 1$ and for $\rho = -.25, -.5, -1$.

The simulation results were based on a multi-sample simulation of size 5000×10 in order to guarantee small standard errors (not presented in the tables, but available from the author) for the simulated characteristics, the Mean Value ($E_\bullet$), the Mean Squared Error ($MSE_\bullet$), the Optimal Sample Fraction, $k_0^\bullet/n$, with $k_0^\bullet := \arg\min_k MSE_\bullet(k)$, and the Relative Efficiency ($REFF_\bullet$), defined as

$$REFF_\bullet = REFF[\gamma_{n0}^\bullet] = \sqrt{\frac{MSE_s\left[\gamma_n^{(1)}(k_{0s}^H(n))\right]}{MSE_s\left[\gamma_n^\bullet(k_{0s}^\bullet(n))\right]}}, \tag{4.1}$$

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

In Tables 1 and 2, we show some finite sample properties, at their optimal levels, of the estimators γ_n^{LS} and γ_n^{ML}, together with the same properties of γ_n^H (reproduction of results in Gomes and Oliveira (2000)), and γ_n^{GJ} (reproduction of results in Gomes et al. (2001)), for different models with $\rho = -1$, the *Fréchet*, the *Out-Hall* and the *Sin-Fréchet* models. Tables 3 and 4 are equivalent to tables 1 and 2, but for Burr models. Finally tables 5 and 6 refer to the Sin-Burr model. For each model, we place in *italic and underlined* the smallest *BIAS* and *MSE* among all the estimators studied.

Table 1: Simulated optimal sample fractions and mean values of γ_n^H, γ_n^{LS}, γ_n^{ML} and γ_n^{GJ} at the simulated optimal levels, for parents with $\rho = -1$, and the *Sin-Fréchet* parent.

n	$\frac{k_{0s}^H}{n}$	$\frac{k_{0s}^{LS}}{n}$	$\frac{k_{0s}^{ML}}{n}$	$\frac{k_{0s}^{GJ}}{n}$	E_s^H	E_s^{LS}	E_s^{ML}	E_s^{GJ}
Fréchet parent: $\rho = -1$, $\gamma = 1$								
100	0.3370	0.8930	0.9480	0.9900	1.1083	0.8893	0.9204	*0.9966*
200	0.2815	0.8205	0.8970	0.9950	1.0850	0.9199	0.9399	*1.0045*
500	0.2208	0.7108	0.8122	0.9938	1.0632	0.9454	0.9585	*1.0106*
1000	0.1762	0.4701	0.7441	0.9901	1.0489	0.9769	0.9690	*1.0092*
2000	0.1418	0.1773	0.6747	0.9837	1.0381	0.9872	0.9766	*1.0059*
5000	0.1110	0.0442	0.5831	0.9757	1.0295	0.9834	0.9845	*1.0032*
10000	0.0880	0.0160	0.5306	0.9713	1.0232	0.9771	0.9880	*1.0021*
20000	0.0706	0.0058	0.4725	0.9671	1.0184	0.9634	0.9910	*1.0012*
Out-Hall parent: $\rho = -1$, $\gamma = 1$								
100	0.0940	0.4550	0.6440	0.3200	0.7178	*0.7516*	0.6211	0.7512
200	0.0685	0.3710	0.5460	0.2535	0.7738	*0.8205*	0.7447	0.8035
500	0.0480	0.2818	0.4548	0.1902	0.8255	0.8770	*0.8789*	0.8541
1000	0.0354	0.2089	0.4128	0.1545	0.8613	0.9120	*0.9333*	0.8825
2000	0.0266	0.1347	0.3831	0.1237	0.8892	0.9413	*0.9643*	0.9071
5000	0.0186	0.0663	0.3575	0.0932	0.9166	0.9661	*0.9858*	0.9308
10000	0.0136	0.0353	0.3474	0.0740	0.9343	0.9787	*0.9922*	0.9456
20000	0.0108	0.0151	0.3412	0.0603	0.9458	0.9902	*0.9955*	0.9559
Sin-Fréchet parent: $\gamma = 1$								
100	0.3130	0.9880	0.8820	0.8590	1.0284	*1.0033*	1.0301	1.0645
200	0.3065	0.9950	0.8580	0.8260	1.0137	0.9803	*1.0099*	1.0325
500	0.3036	0.9962	0.8590	0.8049	1.0067	0.9592	*1.0050*	1.0120
1000	0.3014	0.2536	0.8560	0.7985	1.0025	0.9533	*1.0022*	1.0054
2000	0.3009	0.2402	0.8547	0.7959	1.0011	0.9733	*1.0008*	1.0030
5000	0.3007	0.0620	0.8555	0.7941	1.0007	0.9535	*1.0007*	1.0013
10000	0.3004	0.0335	0.8552	0.7934	*1.0002*	0.9637	1.0003	1.0006
20000	0.3004	0.0143	0.8551	0.7932	*1.0001*	0.9858	1.0002	1.0003

From Figure 2 till Figure 6 we present the simulated mean values and MSE's of $\gamma_n^{ML}(k)$, $\gamma_n^{LS}(k)$, for the sample size $n = 1000$ and for the following parents for which $\rho = -1$, the *Fréchet* ($\gamma = 1$), the *Sin-Fréchet* ($\gamma = 1$), the *Out-Hall* ($\gamma = 1$), the *Burr* ($\gamma = 1$), and *Student-t* with 2 degrees of freedom ($\gamma = 0.5$), respectively. In these Figures, we place also, for easier comparison, the corresponding results for Hill's estimator $\gamma_n^H(k)$ and for the Generalized Jackknife estimator $\gamma_n^{GJ}(k)$ studied in Gomes et al. (2001). Simulations are based on 5000 runs in order to achieve the desired size for confidence intervals.

In Figure 7 we present the patterns of the REFF's of the estimators under study for different sample sizes n (n=100, 200, 500, 1000, 2000, 5000, 10000, 20000), and for different models with $\rho = -1$. Figure 8 is the equivalent of Figure 7, but for models with $\rho = -.25$ and $\rho = -.5$

Table 2: Simulated *MSE*'s and *REFF*'s of γ_n^H, γ_n^{LS}, γ_n^{ML} and γ_n^{GJ} at the simulated optimal levels, and for parents with $\rho = -1$, together with the *Sin-Fréchet* parent.

n	MSE_s^H	MSE_s^{LS}	MSE_s^{ML}	MSE_s^{GJ}	$REFF_s^{LS}$	$REFF_s^{ML}$	$REFF_s^{GJ}$
Fréchet parent: $\rho = -1$, $\gamma = 1$							
100	0.0447	0.0677	*0.0253*	0.0316	0.8126	1.3285	1.1909
200	0.0265	0.0392	*0.0149*	0.0153	0.8223	1.3320	1.3170
500	0.0136	0.0234	0.0075	*0.0062*	0.7637	1.3436	1.4784
1000	0.0083	0.0257	0.0044	*0.0032*	0.5702	1.3759	1.6160
2000	0.0051	0.0378	0.0026	*0.0016*	0.3664	1.4015	1.7704
5000	0.0027	0.0608	0.0013	*0.0007*	0.2115	1.4695	2.0237
10000	0.0017	0.0867	0.0007	*0.0003*	0.1394	1.5249	2.2467
20000	0.0011	0.1219	0.0004	*0.0002*	0.0933	1.5860	2.5054
Out-Hall parent: $\rho = -1$, $\gamma = 1$							
100	0.1574	0.1787	0.2346	*0.1431*	0.9384	0.8192	1.0487
200	0.1080	0.1120	0.1469	*0.0942*	0.9820	0.8575	1.0716
500	0.0653	0.0627	0.0678	*0.0535*	1.0206	0.9814	1.1010
1000	0.0437	0.0444	0.0355	*0.0349*	0.9919	1.1103	1.1175
2000	0.0291	0.0351	0.0185	*0.0227*	0.9101	1.2537	1.1290
5000	0.0169	0.0307	*0.0075*	0.0127	0.7417	1.4964	1.1549
10000	0.0111	0.0306	*0.0038*	0.0082	0.6036	1.7056	1.1662
20000	0.0073	0.0333	*0.0019*	0.0052	0.4678	1.9436	1.1782
Sin-Fréchet parent: $\gamma = 1$							
100	*0.0359*	0.0719	0.0374	0.0536	0.7066	0.9798	0.8178
200	0.0180	0.0354	*0.0176*	0.0277	0.7136	1.0111	0.8071
500	0.0073	0.0221	*0.0069*	0.0114	0.5765	1.0333	0.7985
1000	0.0036	0.0323	*0.0034*	0.0057	0.3366	1.0376	0.7998
2000	0.0018	0.0304	*0.0017*	0.0029	0.2453	1.0397	0.7977
5000	0.0007	0.0325	*0.0007*	0.0011	0.1497	1.0495	0.8006
10000	0.0004	0.0331	*0.0003*	0.0006	0.1048	1.0479	0.8007
20000	0.0002	0.0364	*0.0002*	0.0003	0.0710	1.0481	0.7976

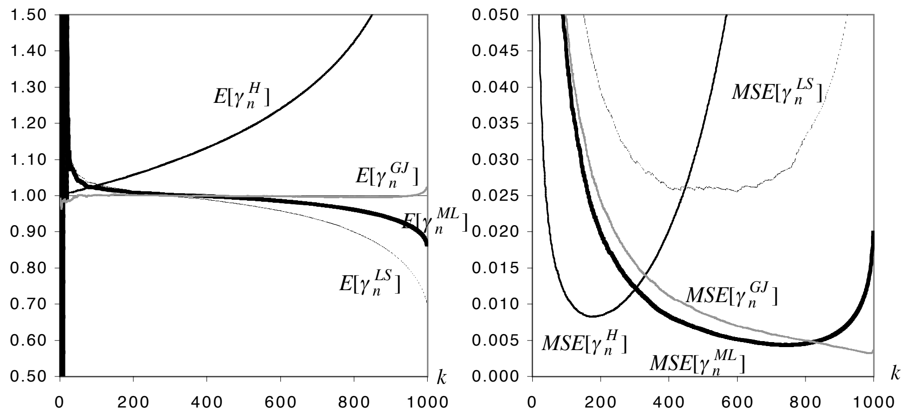

Figure 2: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^H(k)$, $\gamma_n^{ML}(k)$, $\gamma_n^{LS}(k)$ and $\gamma_n^{GJ}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Fréchet* parent with $\gamma = 1$.

Table 3: Simulated optimal sample fractions and mean values of γ_n^H, γ_n^{LS}, γ_n^{ML} and γ_n^{GJ} at the simulated optimal levels, and for *Burr* parents with $\rho = -.25, -.5, -1, -2$.

n	$\frac{k_{0s}^H}{n}$	$\frac{k_{0s}^{LS}}{n}$	$\frac{k_{0s}^{ML}}{n}$	$\frac{k_{0s}^{GJ}}{n}$	E_s^H	E_s^{LS}	E_s^{ML}	E_s^{GJ}
Burr parent: $\rho = -0.25$, $\gamma = 1$								
100	0.0500	0.9890	0.3280	0.9520	1.6649	*1.2904*	1.8078	1.3042
200	0.0335	0.9940	0.2020	0.9750	1.5547	1.2465	1.6496	*1.2372*
500	0.0212	0.9970	0.1166	0.9876	1.4564	1.1937	1.5103	*1.1869*
1000	0.0138	0.9984	0.0743	0.9928	1.3904	*1.0912*	1.4297	1.1546
2000	0.0093	0.0726	0.0479	0.9954	1.3404	1.4003	1.3643	*1.1317*
5000	0.0054	0.0366	0.0258	0.9975	1.2833	1.3188	1.2954	*1.1024*
10000	0.0034	0.0217	0.0168	0.9984	1.2436	1.2671	1.2546	*1.0835*
20000	0.0022	0.0125	0.0105	0.9990	1.2112	1.2208	1.2178	*1.0665*
Burr parent: $\rho = -0.5$, $\gamma = 1$								
100	0.1010	0.9600	0.5410	0.9055	1.2920	0.9277	1.2959	*1.0369*
200	0.0780	0.9370	0.4045	0.9275	1.2423	0.9617	1.2384	*1.0261*
500	0.0520	0.9120	0.2580	0.9411	1.1851	*0.9817*	1.1776	1.0184
1000	0.0384	0.8700	0.1933	0.9500	1.1545	*0.9969*	1.1475	1.0095
2000	0.0300	0.1726	0.1433	0.9537	1.1329	1.1159	1.1222	*1.0056*
5000	0.0187	0.0693	0.0929	0.9562	1.1021	1.0741	1.0948	*1.0027*
10000	0.0133	0.0375	0.0646	0.9574	1.0840	1.0518	1.0769	*1.0012*
20000	0.0094	0.0172	0.0464	0.9578	1.0694	1.0320	1.0640	*1.0007*
Burr parent: $\rho = -1$, $\gamma = 1$								
100	0.2230	0.8210	0.9900	0.8440	1.1362	0.8820	*1.0015*	0.9201
200	0.1885	0.7430	0.9950	0.7990	1.1097	0.9129	*1.0007*	0.9360
500	0.1404	0.6224	0.9976	0.7326	1.0782	0.9456	*1.0000*	0.9522
1000	0.1168	0.4644	0.9988	0.6647	1.0640	0.9723	*1.0001*	0.9644
2000	0.0930	0.2449	0.9993	0.6103	1.0498	0.9918	*1.0001*	0.9724
5000	0.0709	0.0735	0.9997	0.5291	1.0373	0.9997	*1.0001*	0.9812
10000	0.0561	0.0384	0.9998	0.4694	1.0293	0.9986	*1.0000*	0.9860
20000	0.0452	0.0170	0.9999	0.4196	1.0233	0.9971	*1.0000*	0.9895
Burr parent: $\rho = -2$, $\gamma = 1$								
100	0.4200	0.7920	0.8400	0.8580	*1.0693*	0.8719	0.8826	0.8946
200	0.3775	0.7020	0.7380	0.7900	*1.0531*	0.9069	0.9120	0.9152
500	0.3258	0.5840	0.6200	0.6740	*1.0376*	0.9380	0.9385	0.9404
1000	0.2877	0.4684	0.5422	0.5956	*1.0290*	0.9584	0.9531	0.9536
2000	0.2511	0.2432	0.4605	0.5104	1.0218	*0.9874*	0.9659	0.9658
5000	0.2083	0.0731	0.3844	0.4307	1.0149	*0.9998*	0.9763	0.9753
10000	0.1815	0.0385	0.3290	0.3720	1.0112	*0.9987*	0.9826	0.9812
20000	0.1608	0.0164	0.2867	0.3198	1.0087	*0.9980*	0.9866	0.9859

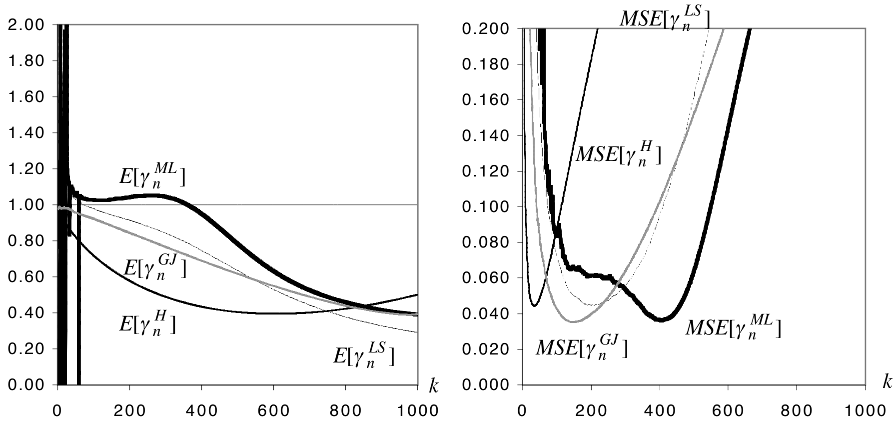

Figure 3: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^H(k)$, $\gamma_n^{ML}(k)$, $\gamma_n^{LS}(k)$ and $\gamma_n^{GJ}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Out-Hall* parent with $\gamma = 1$.

Table 4: Simulated *MSE*'s and *REFF*'s of γ_n^H, γ_n^{LS}, γ_n^{ML} and γ_n^{GJ} at the simulated optimal levels, and for *Burr* parents with $\rho = -.25, -.5, -1, -2$.

n	MSE_s^H	MSE_s^{LS}	MSE_s^{ML}	MSE_s^{GJ}	$REFF_s^{LS}$	$REFF_s^{ML}$	$REFF_s^{GJ}$
Burr parent: $\rho = -0.25$, $\gamma = 1$							
100	0.8976	*0.1749*	0.8919	0.6068	2.2705	1.0033	1.2181
200	0.6105	*0.1087*	0.6038	0.3447	2.3754	1.0056	1.3330
500	0.3815	*0.0939*	0.3726	0.1748	2.0191	1.0118	1.4844
1000	0.2758	0.1503	0.2683	*0.1073*	1.3548	1.0140	1.6045
2000	0.2014	0.2473	0.1929	*0.0677*	0.9026	1.0218	1.7267
5000	0.1350	0.1660	0.1302	*0.0373*	0.9018	1.0184	1.9001
10000	0.1004	0.1246	0.0960	*0.0240*	0.8978	1.0228	2.0414
20000	0.0750	0.0969	0.0715	*0.0155*	0.8798	1.0243	2.2044
Burr parent: $\rho = -0.5$, $\gamma = 1$							
100	0.2286	*0.0552*	0.1519	0.1448	2.0353	1.2268	1.2552
200	0.1450	*0.0305*	0.1017	0.0719	2.1800	1.1940	1.4227
500	0.0834	*0.0237*	0.0607	0.0289	1.8799	1.1716	1.6998
1000	0.0557	0.0452	0.0415	*0.0145*	1.1112	1.1588	1.9569
2000	0.0374	0.0441	0.0284	*0.0073*	0.9211	1.1489	2.2597
5000	0.0228	0.0355	0.0173	*0.0029*	0.8008	1.1460	2.7810
10000	0.0154	0.0330	0.0119	*0.0015*	0.6832	1.1375	3.2523
20000	0.0106	0.0349	0.0083	*0.0007*	0.5504	1.1320	3.8118
Burr parent: $\rho = -1$, $\gamma = 1$							
100	0.0706	0.0714	*0.0124*	0.0683	0.9940	2.3843	1.0161
200	0.0420	0.0423	*0.0059*	0.0364	0.9970	2.6598	1.0738
500	0.0216	0.0236	*0.0023*	0.0162	0.9559	3.0519	1.1522
1000	0.0132	0.0199	*0.0011*	0.0089	0.8147	3.4116	1.2181
2000	0.0082	0.0227	*0.0006*	0.0049	0.6009	3.8153	1.2881
5000	0.0043	0.0270	*0.0002*	0.0023	0.3996	4.4456	1.3799
10000	0.0027	0.0273	*0.0001*	0.0013	0.3143	4.9826	1.4527
20000	0.0017	0.0318	*0.0001*	0.0007	0.2294	5.5549	1.5314
Burr parent: $\rho = -2$, $\gamma = 1$							
100	*0.0299*	0.0784	0.0378	0.0510	0.6180	0.8900	0.7667
200	*0.0166*	0.0469	0.0241	0.0297	0.5944	0.8283	0.7487
500	*0.0078*	0.0259	0.0130	0.0147	0.5466	0.7727	0.7250
1000	*0.0044*	0.0211	0.0079	0.0087	0.4567	0.7460	0.7094
2000	*0.0025*	0.0233	0.0048	0.0051	0.3278	0.7238	0.6969
5000	*0.0012*	0.0269	0.0024	0.0026	0.2100	0.7048	0.6808
10000	*0.0007*	0.0270	0.0014	0.0015	0.1586	0.6949	0.6710
20000	*0.0004*	0.0319	0.0008	0.0009	0.1106	0.6875	0.6631

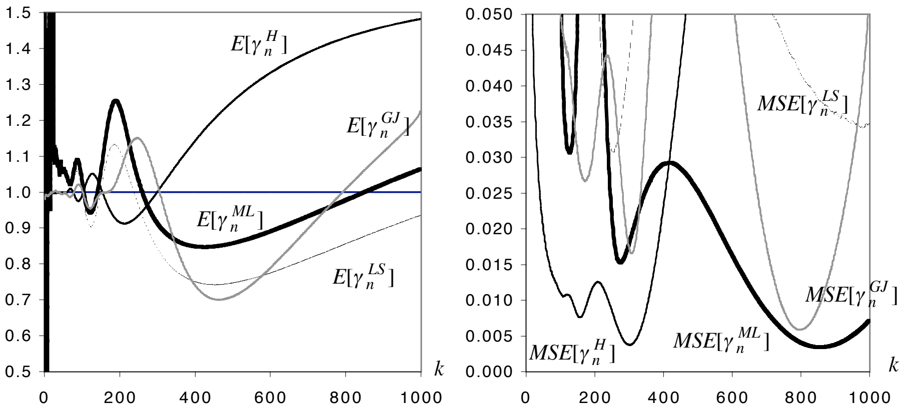

Figure 4: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^H(k)$, $\gamma_n^{ML}(k)$, $\gamma_n^{LS}(k)$ and $\gamma_n^{GJ}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Sin-Fréchet* parent with $\gamma = 1$.

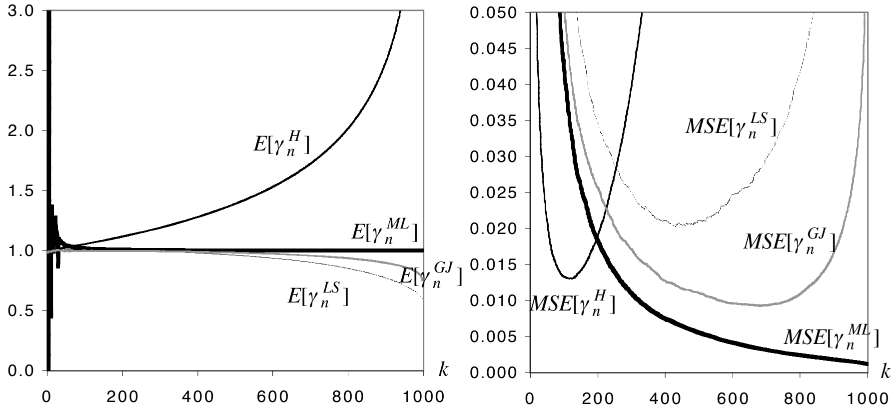

Figure 5: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^H(k)$, $\gamma_n^{ML}(k)$, $\gamma_n^{LS}(k)$ and $\gamma_n^{GJ}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -1$.

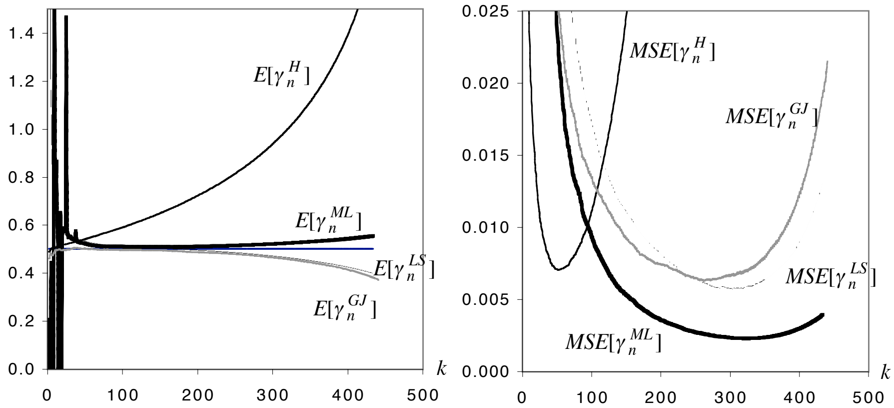

Figure 6: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^H(k)$, $\gamma_n^{ML}(k)$, $\gamma_n^{LS}(k)$ and $\gamma_n^{GJ}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Student*(2) parent with $\gamma = .5$ ($\rho = -1$).

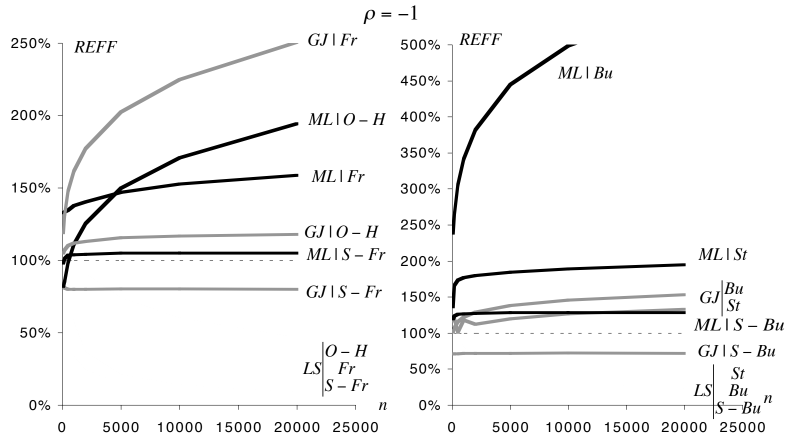

Figure 7: Simulated Relative Efiiciencies for different models with $\rho = -1$.

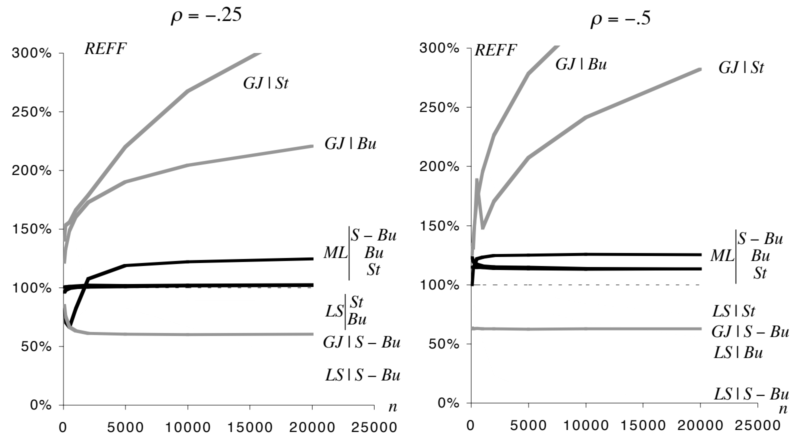

Figure 8: Simulated Relative Efiiciencies for different models with $\rho > -1$.

A few remarks regarding the behaviour of these statistics:

1. The GJ statistic revealed the best behaviour, among the estimators considered, not only for the Fréchet model, for which was specifically devised, but also for large sample sizes from Burr and the Student models with $\rho = -.25$ and $\rho = -.5$, which illustrate values of $\rho > -1$.

2. For the above mentioned models with $\rho > -1$, and for small sample sizes ($n \leq 500$) the LS estimator exhibits nice properties.

3. The ML estimator is the best one, among the estimators considered, for Burr and Student models with $\rho = -1$.

4. For the model outside Hall's class the GJ estimator performs better than

the *ML* estimator for small sample sizes, but the reverse happens for large sample sizes ($n \geq 2000$).

5. For the models for which the second order condition does not hold the *ML* estimator reveals a reasonably high efficiency relatively to the Hill estimator, the *LS* is quite poor, and the *GJ* statistic, although working worse than the Hill estimator, has a relative efficiency greater than 60%.

6. The sample paths of both the *ML* and the *GJ* estimators are quite close to the target value of the tail index γ for a wide range of k-values.

7. For all these alternatives to the Hill estimator, the optimal sample fraction is higher than the one needed for the Hill estimator — so, all depends on the available information or on the price of getting that information.

8. For models with $\rho = 0$, also simulated but not presented in the paper, all estimators have a terribly bad behaviour comparatively to the Hill estimator, with relative efficiencies smaller than 10% for large n.

References

[1] Balkema, A.A. and L. de Haan (1978). Limit distributions for order statistics I (II). *Th. Probab. Appl.* (I), 77-92 ((II), 341-358).

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

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

[4] Draisma, G. (2000). *Parametric and Semi-parametric methods in extreme value theory.* Tinbergen Institute research Series 239.

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

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

[7] Fraga Alves, M.I., Gomes, M.I. and L. de Haan (2001). *A new class of semi-parametric estimators of the second order parameter.* Notas e Comunicações CEAUL 4/01. Submitted.

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

[9] Gomes, M.I., Martins, M.J. and M. Neves (1998). *Alternatives to a semi-parametric estimator of parameters of rare events — the jackknife methodology.* Notas e Comunicações CEAUL 16/98. Accepted at *Extremes.*

[10] Gomes, M. I. and O. Oliveira (2000). *The bootstrap methodology in Statistical Extremes — the choice of the optimal sample fraction.* Notas e Comunicações CEAUL 4/00. Submitted.

[11] Gomes, M.I., Martins, M.J. and M. Neves (2001). *Generalized Jackknife semi-parametric estimators of the tail index.* Technical Report IISA 5/01. Submitted.

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

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

[16] Pickands III, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119-131.

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

[18] Smith, R.L. (1984b). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**, 67.90.

5 Comments on the estimation of D

The "estimators" of D used to build the ML and LS estimators of the tail index γ are not consistent for the estimation of D, unless $\rho = -1$. This is not strange due the way estimation was developed. Let us write explicitely the statistics considered,

$$\widehat{D}_{ML}(k) := \frac{n \sum_{i=1}^{k} (2i-k-1)U_i}{\sum_{i=1}^{k} i(2i-k-1)U_i}, \tag{5.1}$$

and

$$\widehat{D}_{LS}(k) := \frac{12n}{k(k^2-1)} \left\{ \sum_{i=1}^{k} i \ln U_i - \frac{k+1}{2} \sum_{i=1}^{k} \ln U_i \right\}. \tag{5.2}$$

We may write

$$\begin{aligned}
\widehat{D}_{ML}(k) &:= \frac{n}{k} \frac{\frac{1}{k^2} \sum_{i=1}^{k} (2i-k-1)U_i}{\frac{1}{k^3} \sum_{i=1}^{k} i(2i-k-1)U_i} \\
&= \left(\frac{6}{\gamma} \left(\frac{n}{k}\right) \frac{-\rho}{(1-\rho)(2-\rho)} D\gamma \left(\frac{n}{k}\right)^{\rho} \right) (1+o_p(1)) \\
&= -\frac{6\rho}{(1-\rho)(2-\rho)} \left(\frac{n}{k}\right)^{1+\rho} D(1+o_p(1)).
\end{aligned}$$

Also,

$$\begin{aligned}
\widehat{D}_{LS}(k) &= 12 \left(\frac{n}{k}\right) \left(\frac{1}{\gamma(2-\rho)} - \frac{1}{2\gamma(1-\rho)} \right) A(n/k)(1+o_p(1)) \\
&= -\frac{6\rho}{(1-\rho)(2-\rho)} \left(\frac{n}{k}\right)^{1+\rho} D(1+o_p(1)).
\end{aligned}$$

This means that

$$\widehat{D}_{\bullet} \underset{n\to\infty}{\overset{p}{\longrightarrow}} \begin{cases} \infty & \text{if} \quad \rho > -1 \\ D & \text{if} \quad \rho = -1 \\ 0 & \text{if} \quad \rho < -1 \end{cases}, \tag{5.3}$$

i.e., we have consistency for the estimation of D if and only if $\rho = -1$.

Consistency is achieved if we consider the r.v.

$$\widehat{\widehat{D}}_{\bullet} := -\frac{(1-\rho)(2-\rho)}{6\rho} \left(\frac{k}{n}\right)^{1+\rho} \widehat{D}_{\bullet}, \tag{5.4}$$

which converges in probability towards D, as $n \to \infty$.

This may suggest the following alternative procedure of estimation of the unknown parameters:

1. Estimate first ρ by means of a good estimator of the second order parameter, like the ones developed by Fraga Alves et al. (2001);

2. Consider next the following estimator of D:

$$\widehat{D}(k) := -\frac{(1-\widehat{\rho})(2-\widehat{\rho})}{6\widehat{\rho}} \left(\frac{k}{n}\right)^{1+\widehat{\rho}} \frac{n \sum_{i=1}^{k}(2i-k-1)U_i}{\sum_{i=1}^{k} i(2i-k-1)U_i}, \tag{5.5}$$

   which is consistent for the estimation of D, for every $\rho < 0$;

3. Get finally the estimator of the tail index,

$$\widehat{\gamma}(k) := \frac{1}{k}\sum_{i=1}^{k} U_i e^{-\widehat{D}(i)\left(\frac{i}{n}\right)^{-\widehat{\rho}}}. \tag{5.6}$$

It is also possible to built an estimation procedure on the basis of the least squares technique, replacing 2. and 3. by

2'. Estimate D by means of

$$D^*(k) := \frac{\sum_{i=1}^{k}\left(\frac{i}{n}\right)^{-\widehat{\rho}} \ln U_i - \frac{1}{k}\sum_{i=1}^{k}\ln U_i \sum_{i=1}^{k}\left(\frac{i}{n}\right)^{-\widehat{\rho}}}{\sum_{i=1}^{k}\left(\frac{i}{n}\right)^{-2\widehat{\rho}} - \frac{1}{k}\left(\sum_{i=1}^{k}\left(\frac{i}{n}\right)^{-\widehat{\rho}}\right)^2}; \tag{5.7}$$

3'. Obtain

$$\gamma^*(k) := exp\left(\frac{1}{k}\sum_{i=1}^{k}\left\{\ln U_i - D^*(i)\left(\frac{i}{n}\right)^{-\widehat{\rho}}\right\} - \Gamma'(1)\right). \tag{5.8}$$

The estimator of D in (5.5) depends strongly on the level k, and the overall method in 1-2-3 did not work. An *apriori* choice of the threshold k for estimating D has to be made. Here we have chose the value $k = 2n/\ln\ln n$ in (5.5), and the semi-parametric estimator of γ given by

$$\widehat{\gamma}(k) := \frac{1}{k}\sum_{i=1}^{k} U_i e^{-\widehat{D}(n/\ln\ln n)\left(\frac{i}{n}\right)^{-\widehat{\rho}}}. \tag{5.9}$$

6 Classical maximum likelihood revisited under a semi-parametric approach

Although aware of the computational time needed to perform Monte Carlo simulations of ML estimators in EV and Generalized pareto (GP) models, particularly under a semi-parametric approach, we have decided to restrict the scope of the simulation and to consider a few additional semi-parametric estimators of a positive tail index γ, which cannot be obtained explicitly, and which are based essentially on the Paretian behaviour of the excesses over a high level (Pickands, 1975; Balkema and de Hann, 1978) and the maximum likelihood technique.

We recall again that if we assume that

$$U(t) = C\ t^{\gamma}\ e^{A(t)/\rho}, \quad |A| \in RV_{\rho},\ \rho < 0, \tag{6.1}$$

we have the validity of (1.5), and we may write, for every $x > 0$, and as $t \to \infty$,

$$U(tx) - U(t) = U(t) \left\{ x^{\gamma} - 1 + x^{\gamma}\ \frac{x^{\rho} - 1}{\rho} A(t)(1 + o(1)) \right\}.$$

Consequently the excesses over a high level, i.e., $X_{n-i+1:n} - X_{n-k:n}$, $1 \le i \le k$, which may be writen as

$$\begin{aligned} X_{n-i+1:n} - X_{n-k:n} &= U(Y_{n-i+1:n}) - U(Y_{n-k:n}) \\ &= U\left(Y_{n-k:n}\ \tfrac{Y_{n-i+1:n}}{Y_{n-k:n}}\right) - U(Y_{n-k:n}), \end{aligned}$$

where Y denotes a standard Pareto(1) r.v. with d.f. $F_Y(y) = 1 - 1/y,\ y \ge 1$, may be expressed, as $k \to \infty$, as

$$X_{n-i+1:n} - X_{n-k:n} = X_{n-k:n} \left\{ Y^{\gamma}_{k-i+1:k} - 1 \right\} (1 + o_p(1)),\ 1 \le i \le k. \tag{6.2}$$

Since $\frac{Y^{\gamma}-1}{\gamma}$ is a standard Generalized Pareto, GP(γ), r.v., and we are working with all the k o.s. in a sample of size k, we may say that there exists a $\delta > 0$ such that the excesses $\{X_{n-i+1:n} - X_{n-k:n}\}_{1 \le i \le k}$ are approximately GP r.v.'s with scale δ and shape γ, i.e. we may consider a parametric set-up where

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

come from a GP model, with d.f.

$$GP(z; \delta, \gamma) = 1 - \left(1 + \gamma\ \frac{z}{\delta}\right)^{-1/\gamma},\ 1 + \gamma z/\delta \ge 0,\ z \ge 0. \tag{6.3}$$

Remark. Notice that if we look at (6.2) we see that $\delta = \gamma X_{n-k:n}$ is an obvious choice for δ. Indeed, we have the approximation

$$Z_i^* = \frac{X_{n-i+1:n} - X_{n-k:n}}{X_{n-k:n}} \approx Y_{k-i+1:k}^{\gamma} - 1,\ 1 \le i \le k,$$

i.e. the Z_i^*, are a sample of size k from a model with d.f. $F_{Z^*}(z) = 1-(1+z)^{-1/\gamma}$, $z \ge 0$, and the ML estimation of γ leads us to $\widehat{\gamma} = \frac{1}{k}\sum\limits_{i=1}^{k} \ln(1+Z_i^*)$, which is exactly the Hill estimator in (1.8).

6.1 Maximum likelihhod through the GP model

For a random sample of size k, $\underline{Z} = (Z_i,\ 1 \le i \le k)$, from the GP model in (6.3), we have the log-likelihood

$$\ln L(\delta, \gamma; \underline{z}) = -k \ln \delta - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{k} \ln\left(1 + \gamma\ \frac{z_i}{\delta}\right), \tag{6.4}$$

and the ML-equations,

$$k\delta - (1+\gamma)\sum_{i=1}^{k} \frac{z_i}{1+\gamma\ z_i/\delta} = 0; \quad k\gamma - \sum_{i=1}^{k} \ln\left(1+\gamma\ z_i/\delta\right) = 0, \tag{6.5}$$

which may be solved iteratively, for instance by means of the Newton-Raphson procedure. This was exactly one of the algorithms we have used initially, before the choice of the algorithm to be inserted in the overall simulations. Notice that we need to solve the following system of two equations

$$\begin{cases} \varphi_1(\gamma, \delta) = \delta - (1+\gamma)\frac{1}{k}\sum_{i=1}^{k} \frac{z_i}{1+\gamma\ z_i/\delta} =: \delta - \frac{1+\gamma}{k} A \equiv 0 \\ \varphi_2(\gamma, \delta) = \gamma - \frac{1}{k}\sum_{i=1}^{k} \ln\left(1+\gamma\ z_i/\delta\right) =: \gamma - \frac{1}{k} B \equiv 0, \end{cases}$$

and, denoting $C = \sum\limits_{i=1}^{k} \left(\frac{z_i}{1+\gamma\ z_i/\delta}\right)^2$, we have

$$\frac{\partial \varphi_1}{\partial \gamma} = -\frac{A}{k} + \frac{(1+\gamma)C}{k\delta}; \qquad \frac{\partial \varphi_1}{\partial \delta} = 1 - \frac{\gamma(1+\gamma)C}{k\delta^2}$$
$$\frac{\partial \varphi_2}{\partial \gamma} = 1 - \frac{A}{k\delta}; \qquad \frac{\partial \varphi_2}{\partial \delta} = \frac{\gamma A}{k\delta^2}$$

The iterative procedure is then the following:

Algorithm 1.

1. For every k from 2 till $n-1$ consider as initial estimates of γ and δ, $\widehat{\gamma}_0(k) = \gamma_n^H(k)$ and $\widehat{\delta}_0(k) = \widehat{\gamma}_0\ x_{n-k:n}$;

2. Iteratively, compute for $j \geq 1$,

$$\begin{cases} \widehat{\gamma}_j \equiv \widehat{\gamma}_j(k) = \widehat{\gamma}_{j-1}(k) + h_{1j} \\ \widehat{\delta}_j \equiv \widehat{\delta}_j(k) = \widehat{\delta}_{j-1}(k) + h_{2j} \end{cases},$$

where, with $(a,b) = (\widehat{\gamma}_{j-1}(k), \widehat{\delta}_{j-1}(k))$, and $J = \frac{\partial \varphi_1}{\partial \gamma}\frac{\partial \varphi_2}{\partial \delta} - \frac{\partial \varphi_1}{\partial \delta}\frac{\partial \varphi_2}{\partial \gamma}$,

$$h_{1j} = \left.\frac{\varphi_2 \frac{\partial \varphi_1}{\partial \delta} - \varphi_1 \frac{\partial \varphi_2}{\partial \delta}}{J}\right]_{(a,b)}, \quad h_{2j} = \left.\frac{\varphi_1 \frac{\partial \varphi_2}{\partial \gamma} - \varphi_2 \frac{\partial \varphi_1}{\partial \gamma}}{J}\right]_{(a,b)};$$

3. Stop the procedure whenever $\left|[\widehat{\gamma}_j - \widehat{\gamma}_{j-1}]^2 + [\widehat{\delta}_j - \widehat{\delta}_{j-1}]^2\right| \leq 10^{-6}$ or $j \geq 100$.

The estimation of γ through ML in a GP model has been thoroughly studied in Davison (1984) and Smith (1984, 1985).

6.2 A more simple approach to the ML estimation in a GP model

Davison (1984) suggested a re-parametrization of the GP model, in $(\gamma, \alpha) = (\gamma, \gamma/\delta)$, which enables us to get only one ML equation to be solved iteratively, since the ML-estimator of γ has an explicit expression as a function of the ML-estimator of α and the sample. We need first to solve the equation

$$\frac{1}{k}\sum_{i=1}^{k} \frac{z_i}{1+\alpha\, z_i} + \frac{\sum_{i=1}^{k} \frac{z_i}{1+\alpha\, z_i}}{\sum_{i=1}^{k} \ln(1+\alpha\, z_i)} - \frac{1}{\alpha} \equiv 0, \tag{6.6}$$

in order to obtain $\widehat{\alpha}(k)$; then

$$\widehat{\gamma}(k) := \frac{1}{k}\sum_{i=1}^{k} \ln\left(1 + \widehat{\alpha}(k) Z_i\right). \tag{6.7}$$

We have solved the equation (6.6) through two different methodologies: the first one uses the Newton-Raphson's method, and the second one uses the fixed point method.

With the notation

$$\varphi(\alpha) = \frac{1}{k}\sum_{i=1}^{k} \frac{z_i}{1+\alpha\, z_i} + \frac{\sum_{i=1}^{k} \frac{z_i}{1+\alpha\, z_i}}{\sum_{i=1}^{k} \ln(1+\alpha\, z_i)} - \frac{1}{\alpha} =: \frac{A}{k} + \frac{A}{B} - \frac{1}{\alpha} \equiv 0,$$

and noticing that $dB/d\alpha = A$ and $dA/d\alpha = -\sum_{i=1}^{k}\left(\frac{z_i}{1+\alpha\, z_i}\right)^2 =: -C$, we have

$$\varphi'(\alpha) = \frac{1}{\alpha^2} - \left(\frac{A}{B}\right)^2 - C\left(\frac{1}{k} + \frac{1}{B}\right).$$

The algorithms associated to this re-parametrization are then:

Algorithm 2.

1. For every value of k from 2 till $n-1$ consider an initial estimate for α given by $\widehat{\alpha}_0(k) = 1/x_{n-k:n}$;
2. For $j \geq 1$ compute $\widehat{\alpha}_j(k) = \widehat{\alpha}_{j-1}(k) - \frac{\varphi(\widehat{\alpha}_{j-1}(k))}{\varphi'(\widehat{\alpha}_{j-1}(k))}$;
3. Compute $\widehat{\gamma}_j(k) := \frac{1}{k}\sum_{i=1}^{k} \ln\left(1 + \widehat{\alpha}_j(k) z_i\right)$;
4. Stop the procedure whenever $\left|\frac{\widehat{\gamma}_j - \widehat{\gamma}_{j-1}}{\widehat{\gamma}_{j-1}}\right| \leq 10^{-6}$, or $j \geq 100$.

Algorithm 3.

1. For every value of k from 2 till $n-1$ consider the initial estimate for α given by $\widehat{\alpha}_0(k) = 1/x_{n-k:n}$;
2. For $j \geq 1$ compute
$$\widehat{\alpha}_j(k) = \frac{k}{\sum_{i=1}^{k} \frac{z_i}{1+\widehat{\alpha}_{j-1}(k)\ z_i}\left(1 + \frac{k}{\sum_{i=1}^{k}\ln(1+\widehat{\alpha}_{j-1}(k)\ z_i)}\right)};$$
3. Compute $\widehat{\gamma}_j(k) := \frac{1}{k}\sum_{i=1}^{k} \ln\left(1 + \widehat{\alpha}_j(k) z_i\right)$;
4. Stop the procedure whenever $\left|\frac{\widehat{\gamma}_j - \widehat{\gamma}_{j-1}}{\widehat{\gamma}_{j-1}}\right| \leq 10^{-6}$, or $j \geq 100$.

6.3 Comparison of the algorithms

The implementation and running of the algorithms for data from a Fréchet model suggest the following comments:

1. Algorithm 1, i.e., the joint estimation of γ and δ, under a GP model, works generally reasonably well, with a fast convergence, whenever it converges. But there are a few situations spread over different values of k where we cannot get convergence.
2. Algorithm 2 converges only for a few values of k. When it converges, it is much faster than Algorithm 3, but simulation methods are not compatible with non-convergent algorithms, and so, algorithm 2 was the first one discarded.
3. Algorithm 3 converges quite slowly for very small values of k — but even when we get the 100 iterates without stopping the procedure, we have the feeling from the iterates that sooner or later we would get convergence on the basis of the second stopping rule.

Upon the above mentioned considerations we were led to the choice of Algorithm 3, increasing to 500 the number of iterates; then, if we reach for a certain k the 500 iterates without stopping we place the value of γ at the last iterate (this happens only for a few small values of k, which are not important for the development of our work). Indeed, Algorithm 3 turns out to be the most adequate method to go in a simulation, even at the expenses of a higher computational time.

6.4 A "quasi-ML" estimator

At the beginning of section 6 we have mentioned the fact that we would obtain the Hill estimator whenever we considered for the spacings a scale given by $\gamma X_{n-k:n}$. But, whenever we work with the extremal process, i.e. with a vector $V_1 > V_2 > \cdots V_{k+1}$ whose d.f. is

$$f(v_1, v_2, \cdots, v_{k+1}) = \prod_{j=1}^{k} \frac{g_\gamma(v_j)}{G_\gamma(v_j)}\, g_\gamma(v_{k+1}) \quad \text{if} \quad v_1 > v_2 > \cdots > v_{k+1}, \tag{6.8}$$

where G_γ is the EV d.f. in (1.2) and g_γ its derivative, we have that

$$\frac{V_i - V_{k+1}}{1 + \gamma V_{k+1}} \stackrel{d}{=\!=} \frac{Y_{k-i+1:k}^{\gamma} - 1}{\gamma}, \quad 1 \le i \le k \tag{6.9}$$

(Alpuim and Gomes, 1986). Notice that the d.f. in (6.8) is the limiting d.f. of the $k+1$ top o.s., $\{X_{n:n}, X_{n-1:n}, \cdots X_{n-k:n}\}$, for a fixed k.

This result suggested to us the consideration of what we call a "quasi-ML" algorithm, where we consider the second ML-equation in (6.5), putting $\widehat{\delta} = 1 + \widehat{\gamma} X_{n-k:n}$, and solving afterwards, iteratively, the equation

$$\widehat{\gamma} = \frac{1}{k} \sum_{i=1}^{k} \ln\left(1 + \widehat{\gamma} \frac{X_{n-i+1:n} - X_{n-k:n}}{1 + \widehat{\gamma}\; X_{n-k:n}}\right), \tag{6.10}$$

using the fixed point method. We then have

Algorithm 4.

1. For every value of k from 2 till $n-1$ consider the initial estimate for γ given by $\widehat{\gamma}_0(k) = \gamma_n^H(k)$, the Hill estimator in (1.8);

2. For $j \ge 1$ compute

$$\widehat{\gamma}_j(k) = \frac{1}{k} \sum_{i=1}^{k} \ln\left(\frac{1 + \widehat{\gamma}_{j-1}(k) X_{n-i+1:n}}{1 + \widehat{\gamma}_{j-1}(k)\; X_{n-k:n}}\right);$$

3. Stop the procedure whenever $\left|\frac{\widehat{\gamma}_j - \widehat{\gamma}_{j-1}}{\widehat{\gamma}_{j-1}}\right| \leq 10^{-6}$, or $j \geq 100$.

Indeed we have got convergence of Algorithm 4 in a few steps, and the initial simulated sample paths suggested a reduction in bias relatively to the Hill estimator. We have thus also considered this estimation procedure in the overall simulation.

6.5 A ML censoring estimator for a Fréchet model

Since we are working with heavy tails, and we are prepared to have a time-consuming simulation algorithm, which for sure is not going to enable us to deal with the very large sample sizes we have used before, we thought it would be of some relevance to study the behaviour of an estimator of γ under a type II censoring scheme and for a Fréchet model with unknown scale δ (and obviously unknown shape γ), i.e., we assume to be dealing with the d.f.

$$\Phi_\gamma(x;\delta) = e^{-(x/\delta)^{-1/\gamma}}, \quad x \geq 0. \tag{6.11}$$

Under a tipe II censoring scheme, where we have acess to the $k+1$ largest observations $\underline{X} = (X_{n:n}, X_{n-1:n}, \cdots, X_{n-k:n})$ we have, with $y_i = x_{n-i+1:n}$ and $\underline{y} = (y_1, \cdots, y_{k+1})$, a log-likelihood given by

$$\begin{aligned} \ln L(\gamma, \delta; \underline{y}) = -(k+1)\ln\gamma - (k+1)\ln\delta - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{k+1}\ln\left(\frac{y_i}{\delta}\right) \\ - \sum_{i=1}^{k+1}\left(\frac{y_i}{\delta}\right)^{-1/\gamma} - (n-k-1)\left(\frac{y_{k+1}}{\delta}\right)^{-1/\gamma}. \end{aligned} \tag{6.12}$$

The ML equations lead then to an explicit expression for $\widehat{\delta}$,

$$\widehat{\delta} = \widehat{\delta}(k) := \left\{\frac{1}{k+1}\left(\sum_{i=1}^{k+1} y_i^{-1/\widehat{\gamma}(k)} + (n-k-1)y_{k+1}^{-1/\widehat{\gamma}(k)}\right)\right\}^{-\widehat{\gamma}(k)}. \tag{6.13}$$

The other ML equation may be written as

$$\begin{aligned} \widehat{\gamma}(k) = \frac{1}{k+1}\left\{\sum_{i=1}^{k+1}\ln\left(\frac{y_i}{\widehat{\delta}}\right) - \sum_{i=1}^{k+1}\left(\frac{y_i}{\widehat{\delta}}\right)^{-1/\widehat{\gamma}(k)}\ln\left(\frac{y_i}{\widehat{\delta}}\right)\right. \\ \left. -(n-k-1)\left(\frac{y_{k+1}}{\widehat{\delta}}\right)^{-1/\widehat{\gamma}(k)}\ln\left(\frac{y_{k+1}}{\widehat{\delta}}\right)\right\}, \end{aligned} \tag{6.14}$$

which may be used as an iterative formula for the estimation of γ. The iterative procedure (**Algorithm** 5) is then obvious and the stopping rules are the same as before. For other estimators of the tail index related to censoring schemes see Gomes and Oliveira (2001).

In the simulations presented in the following section we have thus implemented algorithms 3, 4 and 5. The corresponding estimators of the tail index will be denoted by γ_n^{PML}, γ_n^{QML} and γ_n^{C}, respectively.

7 Some overall conclusions

- Better not to work only with one estimator;
- Avoid estimators which are not explicitly expressed in the observations;
- Choose adequately a set of simple semi-parametric estimators of a parameter of rare events, or a class of estimators parametrized on a tuning parameter, which one may control at our ease, and picturea few sample pathsassociated to different values of that tuning parameter;
- Do not be afraid of estimators which are not invariant for location. Play with them in your benefit;
- Simple is beautiful! Avoid terribly computer time-consuming methodologies!