Classical maximum likelihood revisited under a semi-parametric context — estimation of the tail index*

M. Ivette Gomes and Orlando Oliveira
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 study computationally the classical *Maximum Likelihood* estimator based on the Paretian behaviour of the excesses over a high threshold, denoted PML-estimator, and a type II *Censoring* estimator based specifically on a Fréchet parent, and denoted $CENS$-estimator. These estimators are considered under a semi-parametric set-up, and compared with the classical *Hill* estimator and a *Generalized Jackknife* (GJ) estimator, which has essentially in mind a reduction of the bias of Hill's estimator.

AMS 2000 subject classification. Primary 62G32, 62E20; Secondary 65C05.
Keywords and phrases. *Statistical Theory of Extremes, Semi-parametric estimation, Maximum Likelihood estimation, Paretian excesses, Censoring.*

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_γ) d.f.

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

with $\gamma > 0$, and we write $F \in \mathcal{D}_M(EV_\gamma)$ (Galambos, 1987).

Then, with $X_{i:n}$ denoting the i-th ascending order statistic (o.s.) associated to the sample $\underline{X}_n = (X_1, X_2, \ldots, X_n)$, the scaled 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 *intermediate* k, i.e. sequences of integers

$$k = k_n \to \infty, \quad k_n = o(n), \text{ as } n \to \infty, \tag{1.4}$$

approximately independent and exponential with mean value γ, and this leads to the consistency of the classical Hill estimator (Hill, 1975),

$$\gamma_n^H(k) := \frac{1}{k}\sum_{i=1}^{k} \left[lnX_{n-i+1:n} - lnX_{n-k:n}\right] \equiv \frac{1}{k}\sum_{i=1}^{k} U_i. \tag{1.5}$$

For $\gamma > 0$,

$$F \in \mathcal{D}_M(EV_\gamma) \quad \text{iff} \quad 1 - F \in RV_{-1/\gamma} \quad \text{iff} \quad U \in RV_\gamma, \tag{1.6}$$

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., positive measurable functions $g(\cdot)$ 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., $F^{\leftarrow}(t) = \inf\{x : F(x) \geq t\}$.

For Hall's class of models (Hall and Welsh (1985)), with a tail

$$1 - F(x) = \alpha\ x^{-1/\gamma}\left\{1 + \beta\ x^{\rho/\gamma}(1 + o(1))\right\},\ \text{as } x \to \infty, \tag{1.7}$$

where $\alpha,\ \gamma \in \mathbb{R}^+$, $\beta \in \mathbb{R} \setminus \{0\}$ and $\rho < 0$, we have $F \in \mathcal{D}_M(EV_\gamma)$ and we may choose $A(t) = Ct^\rho$.

Although aware of the computational time needed to perform Monte Carlo simulations of Maximum Likelihood (ML) estimators in EV_γ and Generalized Pareto (GP_γ) models — $GP_\gamma(x) = 1 + \ln EV_\gamma(x)$, for $x > 0$ —, particularly if under a semi-parametric approach for k ranging from small $k_1 > 1$ till a large $k_2 < n - 1$, we have decided to restrict the scope of the simulation and to consider, in section 2, semi-parametric estimators of the 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. 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 very large sample sizes, we thought it would be of some relevance to study simultaneously 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). Such a derivation is made also in section 2. Finally, in section 3, we shall study through simulation techniques the finite sample properties and robusteness of the estimators for sample sizes up to $n = 2000$. These estimators are compared with the classical

Hill estimator in (1.5) and one of the Generalized Jackknife estimators studied in Gomes et al. (2000), denoted here

$$\gamma_n^{GJ}(k) := \frac{3M_n^{(2)}(k)}{2M_n^{(1)}(k)} - \sqrt{2M_n^{(2)}(k)}, \tag{1.8}$$

where, for $j = 1, 2$,

$$M_n^{(j)}(k) = \frac{1}{k}\sum_{i=1}^{k} [\ln X_{n-i+1:n} - \ln X_{n-k:n}]^j \quad \left[M_n^{(1)}(k) \equiv \gamma_n^H(k)\right]. \tag{1.9}$$

The patterns of mean values and mean squared errors of these estimators, as functions of the level k, are also presented for a few models in $\mathcal{D}_M(EV_\gamma)$, and some comments on that behaviour are put forward.

2 Classical maximum likelihood revisited under a semi-parametric approach

As mentioned before, although aware of the computational time needed to perform Monte Carlo simulations of ML estimators in a GP_γ model with unknown scale δ, under a semi-parametric approach for a wide range of k-values, we have decided to restrict the scope of the simulation and to consider two semi-parametric estimators of a positive tail index γ, described in 2.2 and 2.3, which can be obtained only numerically, through iterative procedures. Our main objective is to quantify how much do we gain with the use of such estimators, relatively to a much simpler estimator like the one in (1.8), which performs better than the Hill estimator, in the sense of minimum mean squared error at optimal levels, for a large variety of models in $\mathcal{D}_M(EV_\gamma)$, $\gamma > 0$ (Gomes et al., 2000).

2.1 The Paretian behaviour of the excesses over a high threshold

The first estimator to be considered in this section is 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.

Since the first order condition (1.6) holds, $U \in RV_\gamma$, and we may write, for every $x > 0$, and as $t \to \infty$,

$$U(tx) - U(t) = U(t)\{x^\gamma - 1\}(1 + o(1)).$$

This means that for an intermediate k the excesses over a high random level $X_{n-k:n}$, i.e., the random sample of size k, $X_{n-i+1:n} - X_{n-k:n}$, $1 \le i \le k$, which may be writen as

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

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

$$X_{n-i+1:n} - X_{n-k:n} = X_{n-k:n}\left\{Y^{\gamma}_{k-i+1:k} - 1\right\}(1 + o_p(1)),\ 1 \leq i \leq k. \quad (2.1)$$

The result in (2.1) follows straightforwardly from the fact that for k intermediate, $t := Y_{n-k:n} = O(n/k) \to \infty$, as $n \to \infty$, and $x := Y_{n-i+1:n}/Y_{n-k:n} \stackrel{d}{=} Y_{k-i+1:k},\ 1 \leq i \leq k$.

Since $\frac{Y^{\gamma}-1}{\gamma}$ is a standard $GP(\gamma)$ 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 $\delta = \delta_k > 0$ such that the excesses $\{X_{n-i+1:n} - X_{n-k:n}\}_{1\leq i\leq k}$ are approximately the k o.s. of a sample of size k from a GP model with scale δ and shape γ, i.e., it is sensible to consider a parametric set-up where

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

come from a GP model, with d.f.

$$GP(z;\delta,\gamma) = 1 - \left(1 + \gamma\,\frac{z}{\delta}\right)^{-1/\gamma},\ 1 + \gamma z/\delta \geq 0,\ z \geq 0. \quad (2.2)$$

Remark. Notice that if we look at (2.1), we see that $\delta = \gamma X_{n-k:n}$ is an obvious choice for δ. Indeed, we have the approximation

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

i.e. the Z_i^*, are a sample of size k from a model with d.f. $F_{Z^*}(z) = 1 - (1+z)^{-1/\gamma}$, $z \geq 0$, and the ML estimation of γ leads us to $\widehat{\gamma} = \frac{1}{k}\sum_{i=1}^{k}\ln(1 + Z_i^*)$, which is exactly the Hill estimator in (1.5).

2.2 Maximum likelihood through the GP model

For a sample of size k, $\underline{z}_k = (z_i,\ 1 \leq i \leq k)$, from the GP model in (2.2), we have the log-likelihood

$$\ln L(\delta,\gamma;\underline{z}_k) = -k\ln\delta - \left(\frac{1}{\gamma} + 1\right)\sum_{i=1}^{k}\ln\left(1 + \gamma\,\frac{z_i}{\delta}\right), \quad (2.3)$$

and the ML-equations,

$$k\delta - (1+\gamma)\sum_{i=1}^{k} \frac{z_i}{1+\gamma\ z_i/\delta} = 0; \quad k\gamma - \sum_{i=1}^{k} \ln(1+\gamma\ z_i/\delta) = 0, \tag{2.4}$$

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:

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

and, denoting $C = \sum_{i=1}^{k} \left(\frac{z_i}{1+\gamma\ z_i/\delta}\right)^2$, 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 $k_1 = 5$ till $k_2 = n-1$ consider as initial estimates of γ and δ, $\widehat{\gamma}_0 \equiv \widehat{\gamma}_0(k) = \gamma_n^H(k)$ 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 (1984a, 1984b).

2.2.1 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. Indeed, 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

$$\frac{1}{k}\sum_{i=1}^{k}\frac{z_i}{1+\alpha\ z_i}+\frac{\sum_{i=1}^{k}\frac{z_i}{1+\alpha\ z_i}}{\sum_{i=1}^{k}\ln(1+\alpha\ z_i)}-\frac{1}{\alpha}\equiv 0, \tag{2.5}$$

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

$$\widehat{\gamma}(k):=\frac{1}{k}\sum_{i=1}^{k}\ln\left(1+\widehat{\alpha}(k)z_i\right). \tag{2.6}$$

We have solved the equation (2.5) 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 $k_1 = 5$ till $k_2 = n-1$, 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, again from $k_1 = 5$ till $k_2 = n-1$ 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 $j \geq 100$.

2.2.2 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 to be 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 first 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.

2.3 A ML censoring estimator for a Fréchet model

Since we are working with heavy tails, and we were prepared to have a time-consuming simulation algorithm, we thought that, as we have said before, 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). We thus assume to be dealing with the d.f.

$$\Phi_\gamma(x;\delta) = e^{-(x/\delta)^{-1/\gamma}}, \quad x \geq 0. \tag{2.7}$$

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

$$\ln L(\gamma, \delta; \underline{y}_{k+1}) = -(k+1)\ln\gamma - (k+1)\ln\delta - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{k+1}\ln\left(\frac{y_i}{\delta}\right) - \sum_{i=1}^{k+1}\left(\frac{y_i}{\delta}\right)^{-1/\gamma} - (n-k-1)\left(\frac{y_{k+1}}{\delta}\right)^{-1/\gamma}. \tag{2.8}$$

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

$$\widehat{\delta} = \widehat{\delta}(k) := \left\{\frac{1}{k+1}\left(\sum_{i=1}^{k+1} y_i^{-1/\widehat{\gamma}(k)} + (n-k-1)y_{k+1}^{-1/\widehat{\gamma}(k)}\right)\right\}^{-\widehat{\gamma}(k)}. \tag{2.9}$$

The other ML equation may be written as

$$\widehat{\gamma}(k) = \frac{1}{k+1}\left\{\sum_{i=1}^{k+1}\ln\left(\frac{y_i}{\widehat{\delta}}\right) - \sum_{i=1}^{k+1}\left(\frac{y_i}{\widehat{\delta}}\right)^{-1/\widehat{\gamma}(k)}\ln\left(\frac{y_i}{\widehat{\delta}}\right) - (n-k-1)\left(\frac{y_{k+1}}{\widehat{\delta}}\right)^{-1/\widehat{\gamma}(k)}\ln\left(\frac{y_{k+1}}{\widehat{\delta}}\right)\right\}, \tag{2.10}$$

which may be used as an iterative formula for the estimation of γ. The iterative procedure (**Algorithm** 4) is then obvious and the stopping rules are the same as before. Due to non-convergence problems for large k and $\rho \neq -1$ we have considered $k_2 = .995(n-1)$. 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 and 4. The corresponding estimators of the tail index will be denoted by γ_n^{PML} and γ_n^{CENS}, respectively.

3 Finite sample properties and robustness of the PML and CENS estimators

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 in (1.7), 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 *Generalized Pareto* model, $F(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, $1 + \gamma x > 0$, for which $\rho = -\gamma$;

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

and a model outside Hall's class,

5. 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 $\rho = -1$.

The simulation results were based on 1000 runs, without replicates. For any of the estimators, denoted generically by $\gamma_n^{\bullet}(k)$, we have simulated the Mean Value ($E^{\bullet}$), the Mean Squared Error ($MSE^{\bullet}$), the Optimal Sample Fraction, $k_o^{\bullet}/n$, with $k_o^{\bullet} := \arg\min_k MSE^{\bullet}(k)$, and the Relative Efficiency ($REFF^{\bullet}$), defined as

$$REFF^{\bullet} = REFF[\gamma_{no}^{\bullet}] = \sqrt{\frac{MSE\left[\gamma_{n,o}^{H}\right]}{MSE\left[\gamma_{n,o}^{\bullet}\right]}}, \tag{3.1}$$

with $\gamma_{n,o}^{\bullet} = \gamma_n^{\bullet}(k_o^{\bullet})$. The subscript "$o$" denotes "*o*ptimality" and the subscript"s", used later, denotes "*s*imulated" characteristics.

In Table 1, we show some finite sample properties, at their optimal levels, of the estimators γ_n^{PML} and γ_n^{CENS}, for four models with $\rho = -1$, the *Fréchet*, the *Out-Hall*, the *GP* and the Burr. Table 2 is equivalent to table 1, but for *GP* and *Burr* models with $\rho \neq -1$. The same properties of γ_n^{H} and γ_n^{GJ} may be seen in Gomes et al. (2000).

Table 1: Simulated optimal sample fractions, mean values and mean squared errors of γ_n^{PML} and γ_n^{CENS} at the simulated optimal levels, for selected parents with $\rho = -1$.

n	k_{0s}^{PML}/n	k_{0s}^{CENS}/n	E_s^{PML}	E_s^{CENS}	MSE_s^{PML}	MSE_s^{CENS}
Fréchet parent: $\rho = -1$, $\gamma = 1$						
100	0.8500	0.9900	0.9010	0.9923	0.0566	0.0064
200	0.8150	0.9950	0.9286	0.9957	0.0291	0.0031
500	0.7480	0.9980	0.9529	0.9982	0.0126	0.0013
1000	0.6950	0.9990	0.9644	0.9992	0.0068	0.0006
2000	0.6270	0.9990	0.9782	0.9995	0.0034	0.0003
Out-Hall parent: $\rho = -1$, $\gamma = 1$						
100	0.4400	0.1000	0.8675	0.6327	0.1075	0.1972
200	0.4350	0.0800	0.9127	0.6990	0.0507	0.1336
500	0.3960	0.0480	0.9586	0.7819	0.0207	0.0809
1000	0.3910	0.0340	0.9701	0.8336	0.0108	0.0526
2000	0.3730	0.0230	0.9807	0.8795	0.0055	0.0331
GP and Burr parents: $\rho = -1$, $\gamma = 1$						
100	0.9900	0.3300	0.9737	1.0631	0.0427	0.0330
200	0.9950	0.2850	0.9870	1.0629	0.0195	0.0204
500	0.9960	0.2420	0.9954	1.0609	0.0081	0.0118
1000	0.9980	0.1650	0.9970	1.0417	0.0038	0.0078
2000	0.9980	0.1450	0.9984	1.0362	0.0020	0.0047

Table 2: Simulated optimal sample fractions, mean values and mean squared errors of γ_n^{PML} and γ_n^{CENS} at the simulated optimal levels, and for *GP* and *Burr* parents with $\rho = -0.5$ and -2.

n	k_{0s}^{PML}/n	k_{0s}^{CENS}/n	E_s^{PML}	E_s^{CENS}	MSE_s^{PML}	MSE_s^{CENS}
GP parent: $\rho = -0.5$, $\gamma = 0.5$						
100	0.9900	0.1000	0.4737	0.5574	0.0252	0.0323
200	0.9950	0.0700	0.4877	0.5516	0.0114	0.0232
500	0.9960	0.0600	0.4955	0.5669	0.0047	0.0146
1000	0.9980	0.0410	0.4973	0.5576	0.0022	0.0104
2000	0.9980	0.0275	0.4987	0.5484	0.0011	0.0074
Burr parent: $\rho = -0.5$, $\gamma = 1$						
100	0.3600	0.1000	1.2075	1.1149	0.1846	0.1294
200	0.2900	0.0700	1.1957	1.1031	0.1225	0.0929
500	0.2020	0.0600	1.1506	1.1337	0.0700	0.0584
1000	0.1380	0.0440	1.1108	1.1166	0.0458	0.0411
2000	0.1280	0.9970	1.1180	1.0321	0.0316	0.0132
GP parent: $\rho = -2$, $\gamma = 2$						
100	0.9900	0.9300	1.9699	2.0004	0.0938	0.0345
200	0.9950	0.9100	1.9836	1.9847	0.0429	0.0169
500	0.9960	0.9140	1.9942	1.9951	0.0178	0.0074
1000	0.9990	0.9130	1.9961	1.9948	0.0084	0.0034
2000	0.9980	0.9150	1.9977	1.9988	0.0045	0.0018
BURR parent: $\rho = -2$, $\gamma = 1$						
100	0.6700	0.9300	0.8667	1.0002	0.0741	0.0086
200	0.6300	0.9100	0.9022	0.9923	0.0376	0.0042
500	0.5520	0.9140	0.9362	0.9975	0.0176	0.0019
1000	0.4970	0.9130	0.9535	0.9974	0.0101	0.0008
2000	0.4070	0.9150	0.9697	0.9994	0.0057	0.0004

From Figure 1 till Figure 7 we present the simulated mean values and MSE's of $\gamma_n^{PML}(k)$ and $\gamma_n^{CENS}(k)$, for the sample size $n = 1000$ and for the following parents: the *Fréchet* ($\gamma = -\rho = 1$), the *Out-Hall* ($\gamma = -\rho = 1$), the GP ($\gamma = -\rho = 1$) — which are the same as the ones obtained for a *Burr* parent with $\gamma = 1$ and $\rho = -1$, since denoting by P and B a $GP(-\rho)$ ($\gamma = -\rho$) and a Burr(γ, ρ) r.v.'s, respectively, we have $P = -B^{-\rho/\gamma}/\rho$ — the GP ($\gamma = -\rho = 0.5$), the *Burr* ($\gamma = 1,\ \rho = -0.5$), the GP ($\gamma = -\rho = 2$) and the *Burr* ($\gamma = 1,\ \rho = -2$), respectively. In these Figures, we place also, for easier comparison, the corresponding results for Hill's estimator $\gamma_n^H(k)$ in (1.5) and for the Generalized Jackknife estimator $\gamma_n^{GJ}(k)$ in (1.8).

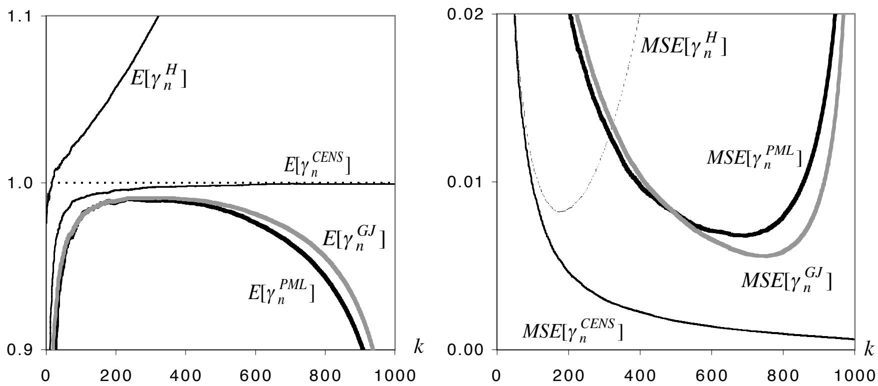

Figure 1: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^H(k)$, $\gamma_n^{GJ}(k)$, $\gamma_n^{PML}(k)$ and $\gamma_n^{CENS}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Fréchet* parent with $\gamma = 1$.

In Table 3 we provide the measure of efficiency in (3.1) for these two estimators and for the estimators introduced in (1.5) and (1.8), for a few models in $\mathcal{D}_M(EV_\gamma)$.

For GP and $Burr$ parents with $\rho = -2$, and as n increases, the mean squared error of the $CENS$ estimator exhibits two minima (although still difficult to detect for $n = 1000$, as may be seen in Figures 6 and 7). The results presented in Tables 2 and 3 are related to the minimum attained at the largest k, which is the global minimum. For $n = 1000$ and GP parents we have a first local minimum for the sample fraction 0.3090, where the mean value and the mean squared error of the $CENS$ estimator are 1.8840 and 0.0248, respectively. For $n = 2000$ those same values are 0.2020, 1.9158 and 0.0158. This would lead to lower efficiencies in Table 3, which would be 0.8365 and 0.7963 for $n = 1000$ and $n = 2000$, respectively. Notice that, with * denoting either the $CENS$ or the GJ estimator, $OSF^*|Burr(1,-\gamma) = OSF^*|GP(\gamma)\ E^*|Burr(1,-\gamma) = \frac{1}{\gamma}\ E^*|GP(\gamma)$, and $MSE^*|Burr(1,-\gamma) = \frac{1}{\gamma^2}\ MSE^*|GP(\gamma)$.

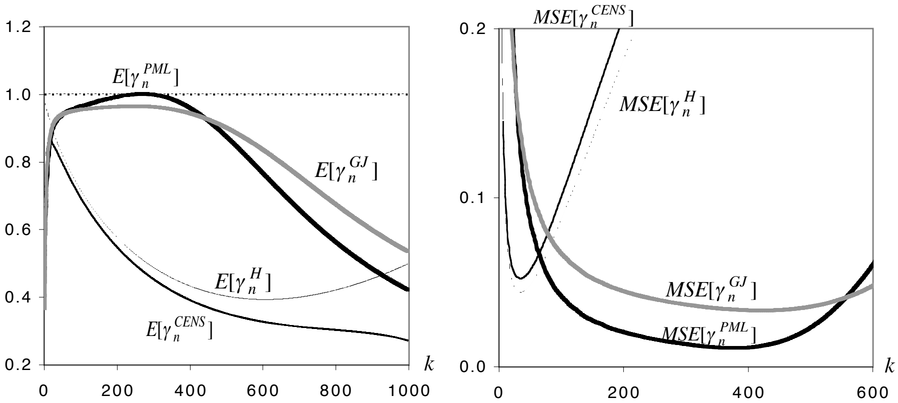

Figure 2: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^H(k)$, $\gamma_n^{GJ}(k)$, $\gamma_n^{PML}(k)$ and $\gamma_n^{CENS}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Out-Hall* parent with $\gamma = 1$.

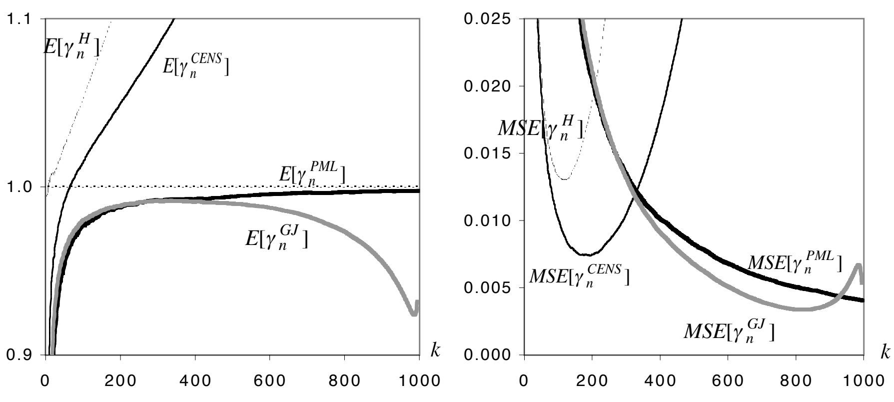

Figure 3: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^H(k)$, $\gamma_n^{GJ}(k)$, $\gamma_n^{PML}(k)$ and $\gamma_n^{CENS}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *GP* parent with $\gamma = 1$ ($\rho = -1$) or a *Burr* parent with $\gamma = 1$ and $\rho = -1$.

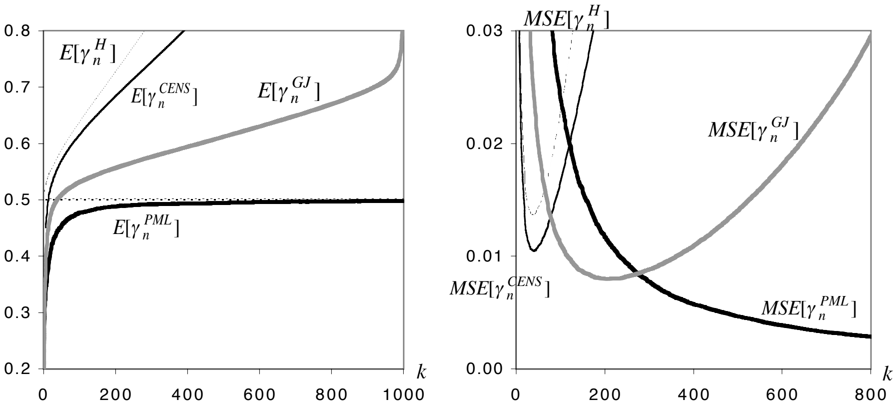

Figure 4: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^H(k)$, $\gamma_n^{GJ}(k)$, $\gamma_n^{PML}(k)$ and $\gamma_n^{CENS}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *GP* parent with $\gamma = 0.5$ ($\rho = -0.5$).

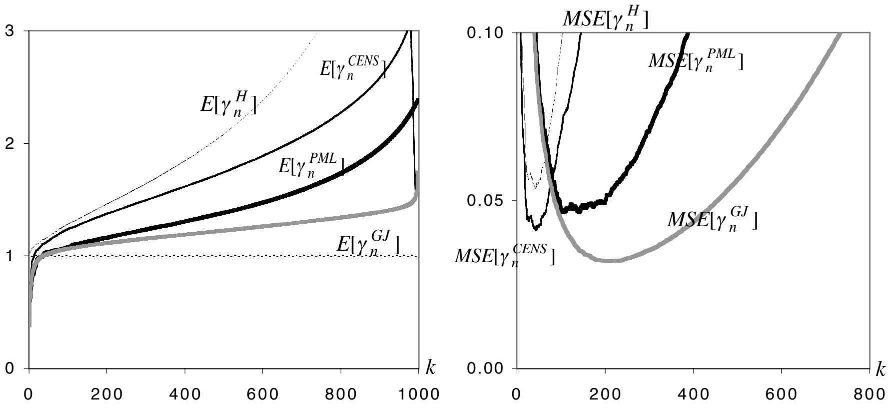

Figure 5: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^H(k)$, $\gamma_n^{GJ}(k)$, $\gamma_n^{PML}(k)$ and $\gamma_n^{CENS}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -0.5$.

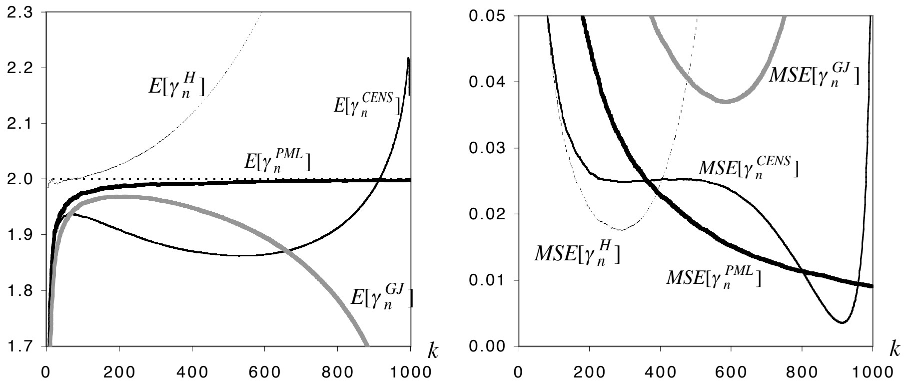

Figure 6: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^H(k)$, $\gamma_n^{GJ}(k)$, $\gamma_n^{PML}(k)$ and $\gamma_n^{CENS}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *GP* parent with $\gamma = 2$ ($\rho = -2$).

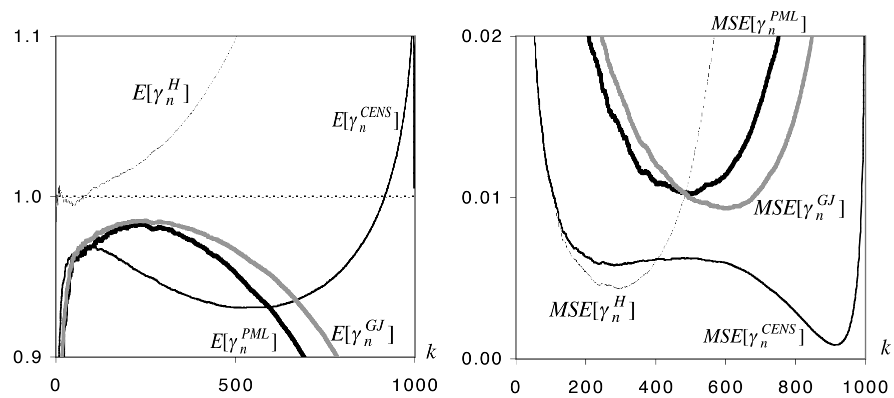

Figure 7: Simulated mean values (*left*) and *MSE's* (*right*) of $\gamma_n^H(k)$, $\gamma_n^{GJ}(k)$, $\gamma_n^{PML}(k)$ and $\gamma_n^{CENS}(k)$, based on 5000 runs, for a sample size $n = 1000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -2$.

Table 3: Relative efficiencies of γ_n^{GJ}, γ_n^{PML} and γ_n^{CENS} for a selected set of parents.

n	100	200	500	1000	2000
Fréchet parent: $\rho = -1$, $\gamma = 1$					
$REFF_s^{GJ}$	1.1021	1.1223	1.1818	1.2110	1.3099
$REFF_s^{PML}$	0.8978	0.9606	1.0484	1.0997	1.2362
$REFF_s^{CENS}$	2.6732	2.9392	3.3205	3.6808	4.0045
Out-Hall parent: $\gamma = 1$, $\rho = -1$					
$REFF_s^{GJ}$	0.8659	0.9336	1.0664	1.1469	1.2667
$REFF_s^{PML}$	1.2172	1.4627	1.8063	2.0374	2.2805
$REFF_s^{CENS}$	0.8986	0.9010	0.9131	0.9235	0.9303
GP and Burr parents: $\gamma = 1$, $\rho = -1$					
$REFF_s^{GJ}$	2.2777	2.3411	2.0476	2.1819	2.2707
$REFF_s^{PML}$	1.2512	1.4479	1.6139	1.8957	2.0116
$REFF_s^{CENS}$	1.4242	1.4180	1.3365	1.3223	1.3169
GP parent: $\gamma = 0.5$, $\rho = -0.5$					
$REFF_s^{GJ}$	1.5277	1.4209	1.3378	1.3076	1.2578
$REFF_s^{PML}$	1.4905	1.7694	2.0827	2.5141	2.8748
$REFF_s^{CENS}$	1.3150	1.2383	1.1788	1.1541	1.1165
Burr parent: $\gamma = 1$, $\rho = -0.5$					
$REFF_s^{GJ}$	1.5277	1.4209	1.3378	1.3076	1.2579
$REFF_s^{PML}$	1.1008	1.0783	1.0775	1.1023	1.0783
$REFF_s^{CENS}$	1.3150	1.2383	1.1788	1.1541	1.11651
GP parent: $\gamma = 2$, $\rho = -2$					
$REFF_s^{GJ}$	0.7557	0.7075	0.7019	0.6820	0.6927
$REFF_s^{PML}$	1.1169	1.1899	1.3001	1.4393	1.4855
$REFF_s^{CENS}$	1.8428	1.8937	2.0133	2.2613	2.3609
Burr parent: $\gamma = 1$, $\rho = -2$					
$REFF_s^{GJ}$	0.7557	0.7075	0.7019	0.6820	0.6927
$REFF_s^{PML}$	0.6282	0.6355	0.6537	0.6539	0.6648
$REFF_s^{CENS}$	1.8428	1.8937	2.0133	2.2613	2.3609

In Figures 8 and 9 we exhibit graphically the data in Table 3.

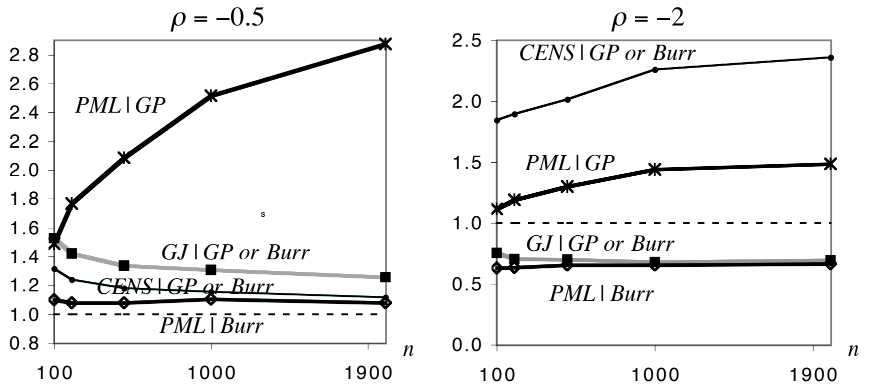

Figure 8: Simulated Relative Efficiencies for different models with $\rho = -1$.

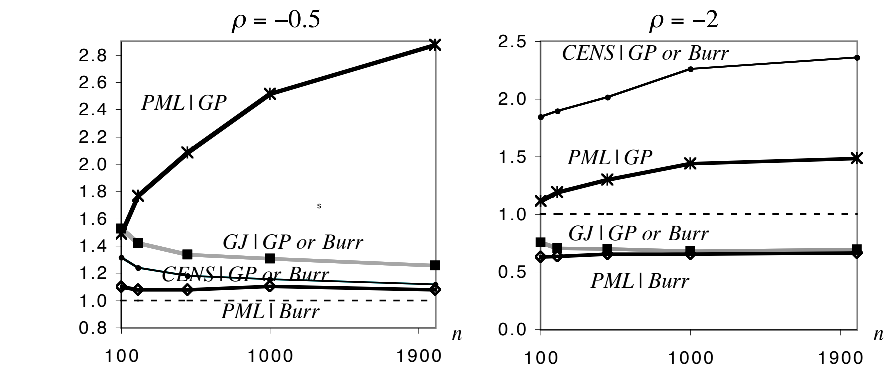

Figure 9: Simulated Relative Efficiencies for GP and $Burr$ models with $\rho \neq -1$.

A few remarks regarding the behaviour of these statistics:

1. The censoring estimator performs badly for the model outside Hall's class, but performs quite well not only for the *Fréchet* model, for which was specifically devised, but also for *Burr* and *GP* models with $\rho = -2$, although due to a second minimum for a low threshold $X_{n-k:n}$, i.e., a high value of k.

2. For models with $\rho < -1$, here illustrated with $\rho = -2$, whereas the $CENS$ estimator overpasses all the other estimators for both simulated models, the PML estimator at their optimal levels overpass the Hill estimator at its optimal level only for the GP model. The GJ estimator exhibits a bad performance in this region of ρ-values.

3. The PML estimator exhibits an interesting performance for the model herewith considered outside of Hall's class, but for which the second order condition does hold. Contrarily, the $CENS$ estimator performs very badly for this model.

4. The PML exhibits again an interesting performance for models in the region $\rho > -1$, being however sometimes overpassed by the GJ estimator.

5. The worst performance of the PML estimator holds for most of the models with a second order parameter $\rho = -1$, and in Hall's class of distributions. Then, it is clearly overpassed by the GJ estimator.

6. The PML estimator has a great advantage relatively to the *Hill* or the $CENS$ estimator — it is invariant to changes in the location of the original data; the GJ estimator due to the way it is build, in order to reduce bias, is not sensitive to changes in location as the Hill estimator, and in this aspect is a good competitor to the PML estimator. In Figure 10, we draw sample paths of the estimators under study for a sample of size $n = 1000$ from *Fréchet* models with location at 0 and at 1.

Some overall conclusions:

- Better not to work only with one estimator; draw sample paths associated to a set of different semi-parametric estimators (i.e., functions of k, the number of order statistics involved in the estimation procedures);

- Estimators which are explicitly expressed in the observations are much easier to obtain and may have high efficiency provided we proceed to some kind of bias reduction, like the ones suggested by Drees (1996), Peng (1998), Beirlant et al. (1999), Feuerverger and Hall (1999), Gomes et al.(2000, 2001), Gomes and Martins (2001), Caeiro and Gomes (2001), among others.

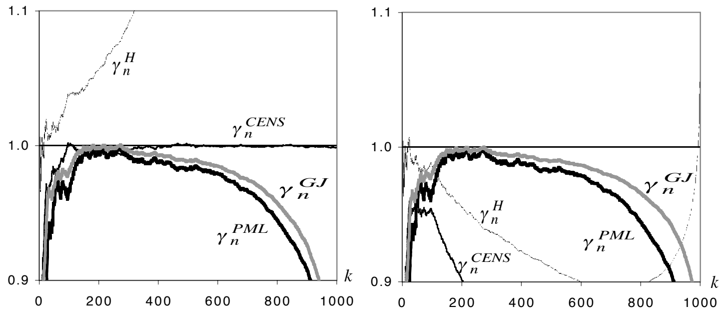

Figure 10: Sample paths of the different estimators under study for a Fréchet model with location 0 (*left*) and with location 1 (*right*).

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] Caeiro, F. and M.I. Gomes (2001). *A new class of asymptotically unbiased estimators of the index of regular variation.* Notas e Comunicações C.E.A.U.L. 8/2001. Submitted.

[4] 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.

[5] Drees, H. (1996). Refined Pickands estimator with bias correction. *Comm. Statist. — Theory and Methods* **25**, 837-851.

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

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

[8] 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.

[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., Martins, M.J. and M. Neves (2001). *Generalized Jackknife semi-parametric estimators of the tail index.* Technical Report IISA 5/01. Accepted at *Portugaliae Mathematica.*

[11] Gomes, M. I. and O. Oliveira (2001). *Censoring estimators of the tail index.* Notas e Comunicações CEAUL 6/01. Submitted.

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

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

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

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

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

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