Tail index and second order parameters' semi-parametric estimation based on the log-excesses

M. Ivette Gomes
University of Lisbon, D.E.I.O. and C.E.A.U.L.

Lígia Henriques Rodrigues
Polytechnic Institute of Tomar and C.E.A.U.L.

Hugo Pereira
Secondary High School and C.E.A.U.L.

Dinis Pestana
University of Lisbon, D.E.I.O. and C.E.A.U.L.

Abstract. In this paper we are interested in the derivation of the asymptotic and finite-sample distributional properties of a "quasi-maximum likelihood" estimator of a "scale" second order parameter β, directly based on the log-excesses of an available sample. Such estimation is of primordial importance for the adaptive selection of the optimal sample fraction to be used in the classical semi-parametric tail index estimation as well as for the reduced-bias estimation of the tail index, high quantiles and other parameters of extreme or even rare events. An application in the area of survival analysis is provided, on the basis of a data set on males diagnosed with cancer of the tongue.

AMS 2000 subject classification. Primary 62G32; Secondary 65C05.
Keywords and phrases. Statistics of extremes, heavy tails, log-excesses, semi-parametric estimation, maximum-likelihood, survival analysis, Monte Carlo simulation.

1 Introduction

Heavy-tailed models appear often in practice in fields like telecommunication traffic, insurance, finance and biostatistics. Power laws, such as the Pareto income distribution (Pareto, 1965) and the Zipf's law for city-size distribution (Zipf, 1941), have been observed a few decades ago in some important phenomena in economics and biology and have seriously attracted scientists in recent years. A model F is said to be heavy-tailed whenever the right *tail function*, $\overline{F} := 1 - F$, is a regularly varying function with a negative index of regular variation equal to $-1/\gamma$, i.e., for every $x > 0$,

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma}, \quad \gamma > 0.$$

Then we write $F \in \mathcal{D}_{\mathcal{M}}(EV_{\gamma>0})$, meaning that we are in the *domain of attraction for maxima* of an *extreme value* distribution function (d.f.),

$$EV_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}),\ 1+\gamma x > 0. \tag{1.1}$$

The parameter γ is the *tail index*, one of the primary parameters of extreme events, and the basis of other parameters of extreme events like *high quantiles* and *probabilities of exceedance* of *high thresholds* (see de Haan and Ferreira, 2006).

In *statistics of extremes* and whenever dealing with heavy-tailed models, i.e., with a positive tail index γ, the classical estimator of γ is the Hill estimator (Hill, 1975), the average of either the *log-excesses*

$$V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}, \tag{1.2}$$

or the *scaled log-spacings*

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

for $1 \le i \le k < n$, where $X_{i:n}$ denotes, as usual, the i-th ascending order statistic (o.s.), $1 \le i \le n$, associated with a random sample $(X_1, X_2, \cdots, X_n)$. We thus have the following expression for the Hill estimator of γ,

$$H_{k,n} := \frac{1}{k}\sum_{i=1}^{k} V_{ik} = \frac{1}{k}\sum_{i=1}^{k} U_i, \tag{1.4}$$

a consistent estimator of γ in the whole $\mathcal{D}_{\mathcal{M}}(EV_{\gamma>0})$ provided that $k = k_n$ is a sequence of intermediate integers, i.e.,

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

For $j \ge 1$, let us denote

$$L_{k,n}^{(j)} := \frac{1}{k}\sum_{i=1}^{k}\left\{1 - \frac{X_{n-k:n}}{X_{n-i+1;n}}\right\}^j, \quad M_{k,n}^{(j)} := \frac{1}{k}\sum_{i=1}^{k}\left\{\ln X_{n-i+1:n} - \ln X_{n-k:n}\right\}^j. \tag{1.6}$$

Apart from the Hill estimator $H_{k,n}$ in (1.4), valid for $\gamma > 0$, we also consider three other classical extreme value index estimators, valid for all $\gamma \in \mathbb{R}$:

1. the *moment* estimator (Dekkers *et al.*, 1989), with the functional form

$$M_{k,n} := M_{k,n}^{(1)} + \tfrac{1}{2}\Big\{1 - \Big(\frac{M_{k,n}^{(2)}}{[M_{k,n}^{(1)}]^2} - 1\Big)^{-1}\Big\}, \tag{1.7}$$

with $M_{k,n}^{(j)}$, $j = 1, 2$, defined in (1.6),

2. the *mixed moment* estimator (Fraga Alves *et al.*, 2007), with the functional form

$$MM_{k,n} := \frac{\hat{\varphi}_{k,n} - 1}{1 + 2\min(\hat{\varphi}_{k,n} - 1, 0)}, \qquad \hat{\varphi}_{k,n} := \frac{M_{k,n}^{(1)} - L_{k,n}^{(1)}}{\big(L_{k,n}^{(1)}\big)^2}, \tag{1.8}$$

with $L_{k,n}^{(1)}$ and $M_{k,n}^{(1)}$ defined in (1.6), and

3. the *generalized Hill* estimator, introduced in Beirlant *et al.* (1996) and further studied in Beirlant *et al.* (2005), with the functional form

$$GH_{k,n} = H_{k,n} + \frac{1}{k}\sum_{i=1}^{k}\{\ln H_{n,i} - \ln H_{k,n}\}, \tag{1.9}$$

where $H_{k,n}$ is the Hill estimator in (1.4).

The three estimators in (1.7), (1.8) and (1.9) are consistent in all $\mathcal{D}_{\mathcal{M}}(G_\gamma)$, $\gamma \in \mathbb{R}$, provided that $k = k_n$ is an intermediate sequence, i.e. a sequence of integers such that (1.5) holds.

Slightly more restrictively than the class of models $F \in \mathcal{D}_{\mathcal{M}}(EV_{\gamma>0})$, we shall assume throughout the paper that we are working in Hall-Welsh class of right *tail functions* of the type

$$\overline{F}(x) = 1 - F(x) = \Big(\frac{x}{C}\Big)^{-1/\gamma}\Big(1 + \frac{\beta}{\rho}\Big(\frac{x}{C}\Big)^{\rho/\gamma} + o\Big(x^{\rho/\gamma}\Big)\Big), \tag{1.10}$$

as $x \to \infty$, with $C > 0$, $\beta \neq 0$, $\rho < 0$ (Hall and Welsh, 1985). This is a large class of d.f.'s. Indeed, models like the *extreme value* d.f. in (1.1), the generalized Pareto d.f., $GP_\gamma(x) = 1 + \ln EV_\gamma(x)$, $x \geq 0$ ($\gamma > 0$), and the *Student*'s t_ν, with $\nu(> 0)$ degrees of freedom ($\gamma = 1/\nu$), belong to the class in (1.10). Then, the optimal sample fraction for the estimation of γ through $H_{k,n}$ in (1.4), in the sense of minimal mean squared error, is given by

$$k_0^H \equiv k_0^H(n; \beta, \rho) = \Big(\frac{(1-\rho)n^{-\rho}}{\beta\sqrt{-2\rho}}\Big)^{2/(1-2\rho)}. \tag{1.11}$$

Hence a first need to estimate β and ρ adequately, in order to estimate k_0^H in (1.11). But the Hill estimator has usually a very "peaked" mean squared error structure, as a function of k, and slight changes in the estimation of k_0^H may induce large confidence intervals associated with $H_{\hat{k}_0^H,n}$. Recently, researchers have thus developed bias-corrected Hill estimators, as well as other types of reduced-bias statistics for the estimation of a positive tail index γ. In a recent paper, Gomes *et al.* (2008a) have decided to accommodate bias directly in the log-excesses in (1.2). For heavy tails, they were then led to a weighted log-excesses (WLE) estimator, given by

$$WLE_{k,n,\hat{\beta},\hat{\rho}} := \frac{1}{k}\sum_{i=1}^{k} e^{-\hat{\beta}\ (n/k)^{\hat{\rho}}\ \psi_{\hat{\rho}}(i/k)}\ V_{ik}, \qquad \psi_\rho(u) \equiv \psi_\rho = -\frac{u^{-\rho}-1}{\rho\ \ln u}, \quad 0 < u \le 1, \quad (1.12)$$

where $(\hat{\beta}, \hat{\rho})$ is any consistent estimator of the vector of second order parameters (β, ρ) in (1.10). Under additional light restrictions on $\hat{\rho}$, the asymptotic variance of the estimator in (1.12) is kept at γ^2, the asymptotic variance of the Hill estimator in (1.4), i.e. we get, up to the second-order, a *minimum variance reduced-bias* ($MVRB$) estimator of γ. Other papers dealing with $MVRB$-estimators of γ are Caeiro *et al.* (2005), Gomes and Pestana (2007a) and Gomes *et al.* (2007). Here, we shall also consider the simplest $MVRB$ estimator, the bias-corrected Hill estimator introduced in Caeiro *et al.* (2005), and given by

$$\overline{H}_{k,n,\hat{\beta},\hat{\rho}} := H_{k,n}\Big(1 - \frac{\hat{\beta}}{1-\hat{\rho}}\left(\frac{n}{k}\right)^{\hat{\rho}}\Big). \quad (1.13)$$

$MVRB$ quantile estimators have been addressed in Gomes and Pestana (2007b) and Beirlant *et al.* (2008).

However, in all the above mentioned papers, as well as in other papers dealing with reduced-bias tail index estimators based on functions of the log-excesses V_{ik} in (1.2), whereas the estimation of ρ is usually done on the basis of moments of the log-excesses, the estimation of β has been done on the basis of another type of statistics, the scaled log-spacings in (1.3), through an estimator introduced in Gomes and Martins (2002). In this paper we shall introduce, in Section 2, a new β-estimator, based on moments of statistics directly related with the log-excesses V_{ik} in (1.2). The asymptotic behaviour of such an estimator will be derived in Section 3. In Section 4, we describe an algorithm for the estimation of the second order parameters (β, ρ) and we also study the finite-sample properties of the new β-estimator, through Monte-Carlo simulation. Finally, in Section 5, we provide an application in the area of biostatistics and draw some general final comments.

2 An estimator of β based on the log-excesses

Since most of the tail index estimators, like the ones in (1.12) and (1.13), are linear combinations of the log-excesses, we think sensible to consider an approach for the estimation of β, based also on the log-excesses, V_{ik}, $1 \le i \le k$, in (1.2). In Gomes *et al.* (2008a), and with E_i, $1 \le i \le k$ denoting a sample of k standard exponential random variables (r.v.'s), the approximation

$$V_{ik} \;\approx\; \gamma e^{\beta\ (n/k)^\rho\ \psi_\rho(i/k)}\ E_{k-i+1:k},$$

with ψ_ρ given in (1.12), has been considered and justified for any k satisfying (1.5). Then, the log-likelihood associated with the k log-excesses, V_{ik}, $1 \le i \le k$, is, up to an additive constant, well approximated by

$$-k \ln\gamma - \beta\left(\frac{n}{k}\right)^\rho \sum_{i=1}^k \psi_\rho(i/k) - \frac{1}{\gamma}\sum_{i=1}^k e^{-\beta\ (n/k)^\rho\ \psi_\rho(i/k)}\ V_{ik}. \tag{2.1}$$

The derivative of the log-likelihood (2.1) in order to β leads to the maximum likelihood equation

$$\frac{1}{k}\sum_{i=1}^k \psi_\rho(i/k)\ e^{-\beta\ (n/k)^\rho\ \psi_\rho(i/k)}\ V_{ik} - \frac{\gamma}{k}\ \sum_{i=1}^k \psi_\rho(i/k) = 0.$$

But this equation does not lead to a consistent estimation of β, because its first member, denoted $A_k^{(1)} - \gamma\ s_k$, where

$$s_k := \frac{1}{k}\sum_{i=1}^k \psi_\rho(i/k),$$

and, for any $j \ge 0$,

$$A_k^{(j)} := \frac{1}{k}\sum_{i=1}^k \psi_\rho^j(i/k)\ \ e^{-\beta\ (n/k)^\rho \psi_\rho(i/k)}\ V_{ik},$$

converges to $\gamma\,(1/(1-\rho) + \ln(1-\rho)/\rho) \ne 0$, as $k \to \infty$. In order to get convergence to 0 we shall replace, in the second sum, ψ_ρ by

$$\psi_\rho^*(u) \equiv \psi^* = -\psi_\rho(u)\ln u = \frac{u^{-\rho}-1}{\rho},\quad 0 < u \le 1,$$

together with the notation

$$s_k^* := \frac{1}{k}\sum_{i=1}^k \psi_\rho^*(i/k). \tag{2.2}$$

The "quasi-maximum likelihood" β-estimator is thus solution of the implicit equation,

$$\frac{1}{k}\sum_{i=1}^{k}\psi_{\hat\rho}(i/k)e^{-\hat\beta\ (n/k)^{\hat\rho}\ \psi_{\hat\rho}(i/k)}V_{ik}-\Big(\frac{1}{k}\sum_{i=1}^{k}\psi^*_{\hat\rho}(i/k)\Big)\times\Big(\frac{1}{k}\sum_{i=1}^{k}\ e^{-\hat\beta\ (n/k)^{\hat\rho}\ \psi_{\hat\rho}(i/k)}V_{ik}\Big)$$
$$=:\hat A_k^{(1)}-\hat s_k^*\ \hat A_k^{(0)}=0. \quad (2.3)$$

Let us further introduce the notation:

$$B_k^{(j)}:=\frac{1}{k}\sum_{i=1}^{k}\psi_\rho^j(i/k)V_{ik},\quad j\ge 0. \quad (2.4)$$

With the obvious notation for $\hat B_k^{(j)}$, $j\ge 0$, with s_k^* and $B_k^{(j)}$ given in (2.2) and (2.4), respectively, if we use in (2.3) the first order approximation $e^x=1+x$, as $x\to 0$, we come to the explicit β-estimator:

$$\hat\beta_{k;\hat\rho|V}=\left(\frac{k}{n}\right)^{\hat\rho}\ \frac{\hat s_k^*\ \hat B_k^{(0)}-\hat B_k^{(1)}}{\hat s_k^*\ \hat B_k^{(1)}-\hat B_k^{(2)}}, \quad (2.5)$$

with a functional form not a long way from the one of the β-estimator in Gomes and Martins (2002), denoted by $\hat\beta_{k;\hat\rho|U}$, where, generally speaking, $\psi_\rho(i/k)$ and V_{ik} are replaced in (2.5) by $(i/k)^{-\rho}$ and the scaled log-spacing U_i, respectively, for all $1\le i\le k$, i.e.

$$\hat\beta_{k;\hat\rho|U}:=\left(\frac{k}{n}\right)^{\hat\rho}\ \frac{\left(\frac{1}{k}\sum\limits_{i=1}^{k}\left(\frac{i}{k}\right)^{-\hat\rho}\right)\left(\frac{1}{k}\sum\limits_{i=1}^{k}U_i\right)-\frac{1}{k}\sum\limits_{i=1}^{k}\left(\frac{i}{k}\right)^{-\hat\rho}U_i}{\left(\frac{1}{k}\sum\limits_{i=1}^{k}\left(\frac{i}{k}\right)^{-\hat\rho}\right)\left(\frac{1}{k}\sum\limits_{i=1}^{k}\left(\frac{i}{k}\right)^{-\hat\rho}U_i\right)-\frac{1}{k}\sum\limits_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\hat\rho}U_i}, \quad (2.6)$$

with U_i given in (1.3).

3 Asymptotic behaviour of the new β-estimator

Before stating a theorem related with the asymptotic behaviour of $\hat\beta_{k;\hat\rho|V}$ in (2.5), we state, without proof, the following lemma, a generalization of Lemma 2 in Gomes *et al.* (2008a):

Lemma 3.1. *For integer values j, let us consider*

$$P_k^{(j)}:=\frac{1}{k}\sum_{i=1}^{k}\psi_\rho^j(i/k)\ E_{k-i+1:k},\ j\ge 0, \quad (3.1)$$

and

$$Q_k^{(j)}:=\frac{1}{k}\sum_{i=1}^{k}\psi_\rho^{j-1}(i/k)\frac{Y^\rho_{k-i+1:k}-1}{\rho},\ j\ge 1, \quad (3.2)$$

with $\psi_\rho(\cdot)$ given in (1.12), being $\{E_i\}$ and $\{Y_i\}$ sequences of i.i.d. standard exponential and unit Pareto r.v.'s, respectively. Denoting $\mathbb{E}$ the mean value operator and

$$a_j \equiv a_j(\rho) := \frac{(-1)^{j-1}}{\rho^j} \int_0^1 \frac{(v^{-\rho}-1)^j}{\ln^{j-1} v}\, dv = -\int_0^1 \psi_\rho^j(v)\; lnv\; dv < \infty, \quad j \geq 0 \quad (a_0 = 1), \tag{3.3}$$

both $\mathbb{E}\big(P_k^{(j)}\big)$, $j \geq 0$, and $\mathbb{E}\big(Q_k^{(j)}\big)$ $j \geq 1$, with $P_k^{(j)}$ and $Q_k^{(j)}$ given in (3.1) and (3.2), respectively, converge towards a_j in (3.3). For the particular cases $j = 1, 2$, a_1 and a_2 are explicitly given by the formulas

$$a_1 = \frac{1}{1-\rho}, \qquad a_2 = -\frac{\ln(1-2\rho) - 2\ln(1-\rho)}{\rho^2}. \tag{3.4}$$

Moreover, for $j \geq 0$, and with $\mathbb{V}ar$ denoting the variance operator,

$$\begin{aligned} \sigma_j^2 &= \lim_{n\to\infty} k\ \mathbb{V}ar\left(P_k^{(j)}\right) \\ &= \frac{2}{\rho^{2j}} \iint\limits_{0\le u<v\le 1} \left(\frac{u^{-\rho}-1}{\ln u}\ \frac{v^{-\rho}-1}{\ln v}\right)^j \frac{1-v}{v} du\ dv < \infty \quad [\sigma_0 = 1]. \end{aligned} \tag{3.5}$$

Consequently, for $j \geq 0$, $P_k^{(j)}$ in (3.1) converges in probability towards a_j, as $k \to \infty$, with a_j, $j \geq 0$, given in (3.3). Also $k\ \mathbb{C}ov\big(P_k^{(0)}, P_k^{(1)}\big) \sim a_1 = 1/(1-\rho)$.

Similarly to Theorem 3 in Gomes *et al.* (2008a), related with the estimator $\hat\beta_{k;\hat\rho|U}$ in (2.6), we shall now state, the following result for the β-estimator $\hat\beta_{k;\hat\rho|V}$ in (2.5):

Theorem 3.1. *For models in (1.10), if $\hat\rho$ is any consistent estimator of ρ, if $k = k_n$ is a sequence of intermediate integers, i.e., (1.5) holds, and if, with $A(t) = t^\rho$, we further have $\lim_{n\to\infty} \sqrt{k}\ A(n/k) = \infty$, then $\hat\beta_{k;\hat\rho|V}$ in (2.5) converges in probability towards β, as $n \to \infty$. With an extra condition on $\hat\rho$, of the type $\hat\rho - \rho = o_p(1/\ln n)$, and an extra assumption on the term $o(x^{\rho/\gamma})$ in (1.10), which we assume to be of the order of $x^{2\rho/\gamma}$, we can further guarantee the asymptotic normality of $\hat\beta_{k;\hat\rho|V}$ whenever $\sqrt{k}\ A^2(n/k) \to \lambda$, finite, as $n \to \infty$. The asymptotic standard deviation of $\hat\beta_{k;\hat\rho|V}$ is then ruled by*

$$\sigma_{\hat\beta_V} = \frac{\gamma\ |\beta|\ \sqrt{\sigma_1^2 - a_1^2}}{a_1^2 - a_2}, \tag{3.6}$$

with (a_1, a_2) and σ_j^2 given in (3.4) and (3.5), respectively. Again, the rate of convergence of $\hat\beta_{k;\hat\rho(k)|V}$ towards β is of the order of $\ln(n/k)/\big(\sqrt{k}\ A(n/k)\big)$ for any $\hat\rho(k) = O_p(1/(\sqrt{k}A(n/k)))$.

Proof. Note that

$$\frac{\hat{\beta}_{k;\rho|V}}{\beta} = \frac{\gamma}{A(n/k)}\ \frac{s_k^*\ B_k^{(0)} - B_k^{(1)}}{s_k^*\ B_k^{(1)} - B_k^{(2)}},$$

with $B_k^{(j)}$ given in (2.4). Noticing that $a_1 = 1/(1-\rho)$ is the limiting value of s_k^* in (2.2), as $n \to \infty$, and assuming ρ known, we have

$$\frac{A(n/k)\ \hat{\beta}_{k;\rho|V}}{\gamma\ \beta} = \frac{a_1\ B_k^{(0)}(1+o(1)) - B_k^{(1)}}{a_1\ B_k^{(1)}(1+o(1)) - B_k^{(2)}}.$$

On the other side, for $j \geq 0$, with $a_0 = 1$, a_j given in (3.3) for $j \geq 1$, there exist σ_j^2 ($\sigma_0^2 = 1$), given in (3.5), and $\overline{P}_k^{(j)} := \sqrt{k}\big(P_k^{(j)} - \mathbb{E}\big(P_k^{(j)}\big)\big)/\sigma_j$ asymptotically standard normal, with $P_k^{(j)}$ given in (3.1), such that

$$B_k^{(j)} \stackrel{d}{=} \gamma\ a_j + \frac{\gamma\ \sigma_j}{\sqrt{k}}\ \overline{P}_k^{(j)} + a_{j+1}A(n/k)(1+o_p(1)).$$

Then,

$$\frac{A(n/k)\ \hat{\beta}_{k;\rho|V}}{\gamma\ \beta} \stackrel{d}{=} \frac{\frac{\gamma}{\sqrt{k}}\left(a_1\overline{P}_k^{(0)} - \sigma_1\overline{P}_k^{(1)}\right) + (a_1^2 - a_2)\ A(n/k) + o_p(A(n/k))}{\gamma(a_1^2 - a_2)(1+o_p(1))},$$

and consequently,

$$\frac{\hat{\beta}_{k;\rho|V}}{\beta} \stackrel{d}{=} 1 + \frac{\gamma}{\sqrt{k}\ A(n/k)}\left(\frac{a_1\overline{P}_k^{(0)} - \sigma_1\overline{P}_k^{(1)}}{a_1^2 - a_2}\right) + R_k^{V}, \quad \text{with} \quad R_k^{V} = o_p(1),$$

i.e., $\hat{\beta}_{k;\rho|V}$ converges in probability towards β, provided that $\sqrt{k}\ A(n/k) \to \infty$, as $n \to \infty$. The same result is true if we replace ρ by $\hat{\rho}$, any consistent estimator of ρ.

Next, note that

$$\frac{d}{d\rho}\hat{\beta}_{k;\rho|V} = -\hat{\beta}_{k;\rho|V}\ln(n/k)(1+o_p(1)).$$

Under the conditions on $\hat{\rho}$, the use of Cramer's delta-method enables us to write

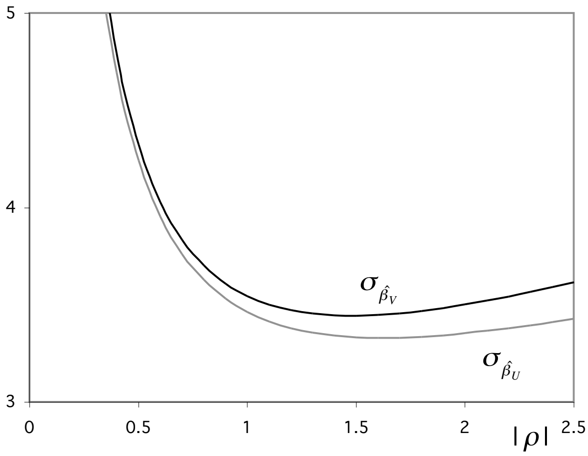
$$\hat{\beta}_{k;\hat{\rho}|V} = \hat{\beta}_{k;\rho|V} - \hat{\beta}_{k;\rho|V}(\hat{\rho} - \rho)\ln(n/k)(1+o_p(1)),$$

and the remaining of the theorem follows. □

Under the same conditions as before in Theorem 3.1, the β-estimator $\hat{\beta}_{k;\hat{\rho}|U}$, in Gomes and Martins (2002), has an asymptotic standard deviation ruled by

$$\sigma_{\hat{\beta}_U} = \frac{\gamma\ |\beta|(1-\rho)\sqrt{1-2\rho}}{|\rho|}. \tag{3.7}$$

Another β-estimator can be found in Caeiro and Gomes (2006).



$\|\rho\|$	$\sigma_{\hat\beta_U}$	$\sigma_{\hat\beta_V}$
0.1	12.05	13.20
0.2	7.10	7.40
0.3	5.48	5.63
0.4	4.70	4.80
0.5	4.24	4.32
1.0	3.46	3.55
1.5	3.33	3.44
2.0	3.35	3.50
2.5	3.43	3.62

Figure 1: "Rulers" of the asymptotic standard deviations of $\hat\beta_{k;\hat\rho|U}$ and $\hat\beta_{k;\hat\rho|V}$, for $\gamma = \beta = 1$.

Remark 3.1. *For the difference between* $\sigma_{\hat\beta_V}$ *in* (3.6) *and* $\sigma_{\hat\beta_U}$ *in* (3.7), *see* Figure 1. *Notice that there is only a very slight difference between the asymptotic variances of* $\hat\beta_{k;\hat\rho|U}$, *based on the scaled log-spacings, and* $\hat\beta_{k;\hat\rho|V}$, *based on the log-excesses. Such a difference is not quite relevant in practice, and the two estimators provide practically the same results when incorporated in the estimators in* (1.12) *or in* (1.13). *But the estimator in* (2.5) *is more natural, in the sense that it is based exactly on the same type of statistics, the log-excesses in* (1.2). *Moreover* $\hat\beta_{k;\hat\rho|V}$ *in* (2.5) *exhibits often a much smaller bias than the estimator* $\hat\beta_{k;\hat\rho|U}$ *in Gomes and Martins* (2002), *also reproduced in* (2.6), *as can be seen in Section 4.*

4 An algorithm for the estimation of the second order parameter (ρ, β) and a Monte Carlo simulation

4.1 The algorithm

We propose the following **Algorithm** (similar to the ones in Gomes and Pestana, 2007b and Gomes *et al.*, 2008a) for an estimation of ρ, prior to the estimation of β through the estimator in (2.5). The estimation of ρ is done on the basis of the class of estimators in Fraga Alves *et al.* (2003).

1. Given a sample $(X_1, X_2, \cdots, X_n)$, plot, for $\tau = 0$ and $\tau = 1$, the estimates

$$\hat{\rho}_{k,\tau} := \min\Big\{0,\ \frac{3(T_{k,n}^{(\tau)} - 1)}{T_{k,n}^{(\tau)} - 3}\Big\}, \tag{4.1}$$

where, with $M_{k,n}^{(j)}$ given in (1.6), $j = 1, 2, 3$, and the notation $a^{b\tau} = b \ln a$ whenever we consider $\tau = 0$,

$$T_{k,n}^{(\tau)} := \frac{\left(M_{k,n}^{(1)}\right)^{\tau} - \left(M_{k,n}^{(2)}/2\right)^{\tau/2}}{\left(M_{k,n}^{(2)}/2\right)^{\tau/2} - \left(M_{k,n}^{(3)}/6\right)^{\tau/3}},$$

for any $\tau \in \mathbb{R}$.

2. Consider $\{\hat{\rho}_{k,\tau}\}_{k\in\mathcal{K}}$ in (4.1) for large k, say $k \in \mathcal{K} = \left(\left[n^{0.990}\right],\ \left[n^{0.999}\right]\right)$, and compute their median, denoted ρ_τ. Next choose the *tuning parameter* $\tau_0 := \arg\min_\tau \sum_{k\in\mathcal{K}} \left(\hat{\rho}_{k,\tau} - \rho_\tau\right)^2$, and work with $\hat{\rho} \equiv \hat{\rho}_{\tau_0} := \hat{\rho}_{k_1,\tau_0}$, with

$$k_1 = \left[n^{0.995}\right] \tag{4.2}$$

and $\hat{\rho}_{k,\tau}$ given in (4.1).

3. Consider the β-estimator $\hat{\beta} \equiv \hat{\beta}_{\tau_0} = \hat{\beta}_{k_1;\hat{\rho}_{\tau_0}|V}$, with $\hat{\beta}_{k;\rho|V}$ given in (2.5).

The choice of the level k_1 in (4.2) is not crucial. We merely should consider any reasonably large value of k of the order of $n^{1-\epsilon}$ for small ϵ, due to the high stability of $\hat{\rho}_{k,\tau_0}$ around the target ρ for large k-values and for a large class of models. Also, this algorithm leads in almost all situations to $\tau = 0$ whenever $|\rho| \leq 1$ and $\tau = 1$, otherwise. Such an educated guess usually provides slightly better results than a "noisy" estimation of τ, it has been recommended in practice and will be used in the simulations of this Section, as well as in the case study in Section 5.

4.2 Monte Carlo simulations

We have performed large-scale Monte Carlo simulations for the following parents:

- the *Fréchet* model, with d.f. $F(x) = \exp(-x^{-1/\gamma})$, $x \geq 0$, $\gamma > 0$, for which $\rho = -1$ and $\beta = 1/2$;
- the *Generalized Pareto* (*GP*) model, with d.f. $F(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, $x \geq 0$, $\gamma > 0$, for which $\rho = -\gamma$ and $\beta = 1$;

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

$$f_{t_\nu}(t) = \frac{\Gamma((\nu+1)/2)\left[1 + t^2/\nu\right]^{-(\nu+1)/2}}{\sqrt{\pi\nu}\,\Gamma(\nu/2)},$$

  defined for $t \in \mathbb{R}$ $(\nu > 0)$, with $\gamma = 1/\nu$, $\rho = -2/\nu$ and $\beta = (\nu+1)c_\nu^2/(\nu+2)$, with $c_\nu = (\nu\mathcal{B}(\nu/2, 1/2))^{1/\nu}$ ($c_1 = \pi$ leading to the usually called Cauchy d.f.), and where $\mathcal{B}$ is the complete Beta function.

For each value of $n = 100$, 200, 500, 1000, 2000, 5000 and 10000, for each model and for the estimate $\hat\rho = \hat\rho_{\tau_0}$ of ρ, suggested in the algorithm of Section 4.1, we have simulated the mean values and the mean squared errors of both β-estimators, the new estimator $\hat\beta_{k;\hat\rho|V}$ in (2.5) and the old estimator $\hat\beta_{k;\hat\rho|U}$ in (2.6). We have next computed the mean values and the mean squared errors at optimal levels, i.e., at the levels k_0 that minimize the mean squared error of the statistics under play. We have also computed these same characteristics at the level k_1 in (4.2). As an illustration, we picture in Figure 2, for underlying Burr models with $\gamma = 1$ and $\rho = -0.5, -1$ and -2, respectively (all with a second order parameter $\beta = 1$), the mean values of $\hat\beta|U := \hat\beta_{k_1;\hat\rho|U}$ and $\hat\beta|V := \hat\beta_{k_1;\hat\rho|V}$, for the above mentioned values of n. These results were based on multi-sample simulations of size 1000×10. Similar results have been obtained for the other simulated models.

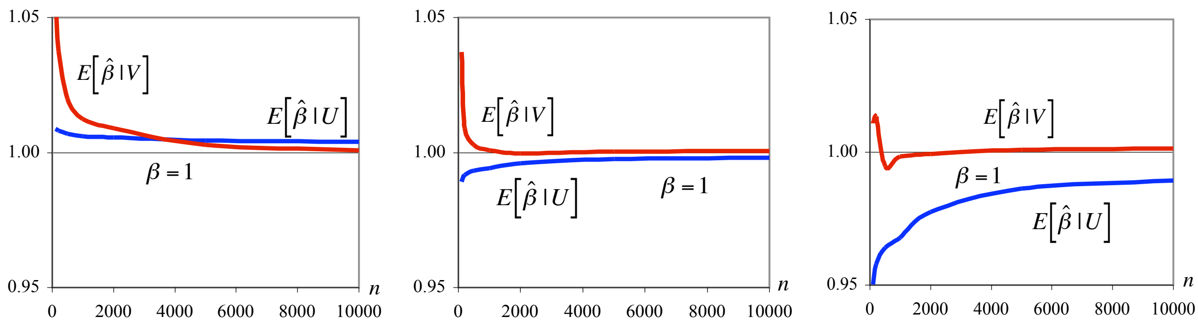


Figure 2: Mean values of $\hat\beta|U$ and $\hat\beta|V$, for Burr samples with $(\gamma, \rho) = (1, -0.5)$ (*left*), $(\gamma, \rho) = (1, -1)$ (*center*) and $(\gamma, \rho) = (1, -2)$ (*right*).

It is clear from Figure 2 that regarding mean values (and consequently bias) the new β-estimator based on the log-excesses, i.e., the estimator $\hat\beta_{k;\hat\rho|V}$ in (2.5), provides better results than the old estimator $\hat{\beta}_{k;\hat{\rho}|U}$ in Gomes and Martins (2002), unless ρ is quite close to 0 and n small (see Figure 2, left). Regarding mean squared errors, and despite of the slightly higher asymptotic standard deviation of the estimator in (2.5), comparatively with the estimator $\hat{\beta}_{k;\hat{\rho}|U}$ (see Figure 1), the new estimator is still able to overpass the initial one for large n when ρ is close to 0 and for small n when ρ is smaller than -1, as can be inferred from Figure 3, related with the behavior of the β-estimators under comparison, for the same Burr parents as before.

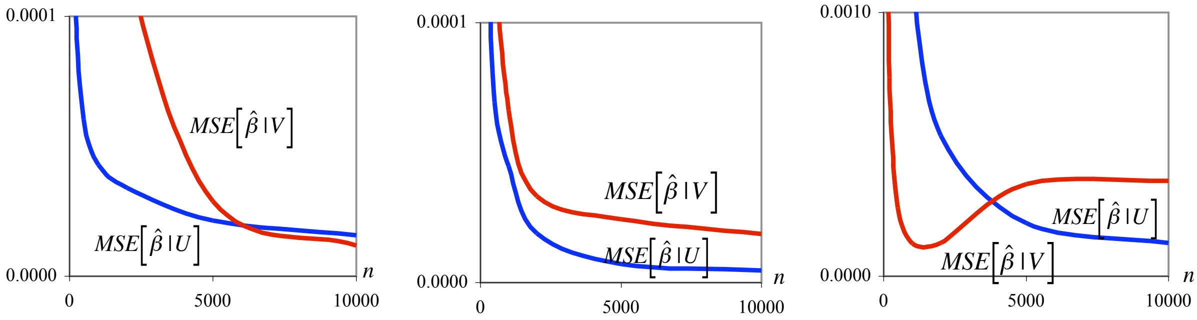


Figure 3: Mean squared errors of $\hat{\beta}|U$ and $\hat{\beta}|V$, for Burr samples with $\gamma = 1$ and $\rho = -0.5$ (*left*), $\rho = -1$ (*center*) and $\rho = -2$ (*right*).

As mentioned before, the most interesting feature of the new β-estimator in (2.5) is its small bias at optimal levels or at the level k_1 in (4.2), even for values of ρ close to 0, provided that the sample size n is not too small. Similarly to what happens with the estimation of ρ, the level k_1 in (4.2), although not necessarily optimal, seems to be also appropriate for the estimation of β through the new estimator $\hat{\beta}_{k;\hat{\rho}|V}$ in (2.5).

5 A case-study and a few final comments

We have analyzed a data set on $n = 80$ males diagnosed with cancer of the tongue, with observations of Z=time to death or on-study time, in weeks (Klein and Moeschberger, 2005, Section 1.11; Sickle-Santanello *et al.*, 1988). Despite of the fact that the data of interest, X = time to death after detection of illness, is random censored, we can obviously estimate $\gamma = \gamma_Z$ conscious that $\gamma_X \neq \gamma_Z$. For the extreme value index estimation associated with data under random censoring, see Beirlant *et al.* (2007) and Einmahl *et al.* (2007).

For this data-set, there was an indication of a right tail clearly heavier than the normal, and also slightly heavier than the Gumbel, not too easy to find in other sets of survival data. This can be seen in Figure 4, where we plot the Gumbel and normal QQ-plots associated with the data. The same comment applies if we distinguish the two kinds of tumors available in the data set considered (aneuploid and diploid tumors).

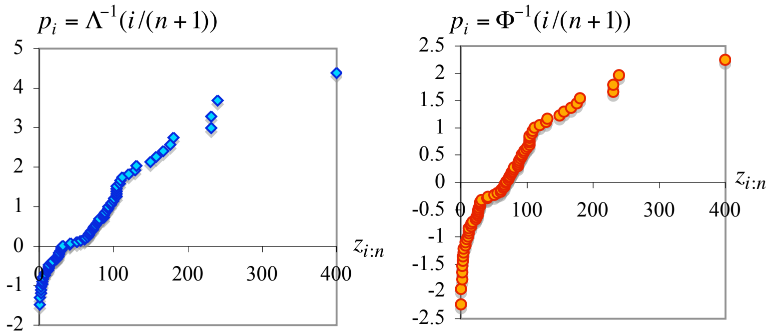


Figure 4: Gumbel probability paper (*left*) and normal probability paper (*right*) for the data under analysis.

The application of the algorithm in Section 4.1 led us to the estimate $\hat{\rho} = \hat{\rho}_0 = -0.654$, and to the estimates $\hat{\beta}|V = 1.151$ and $\hat{\beta}|U = 1.108$. The estimates of the second order parameters ρ and β, as a function of k, the number of top o.s.'s used in the estimation, are provided in Figure 5. It is clear the stability of $\hat{\rho}_{k,0}$ $(\tau = 0)$ for large values of k. The same comment doest not apply here to the β-estimates.

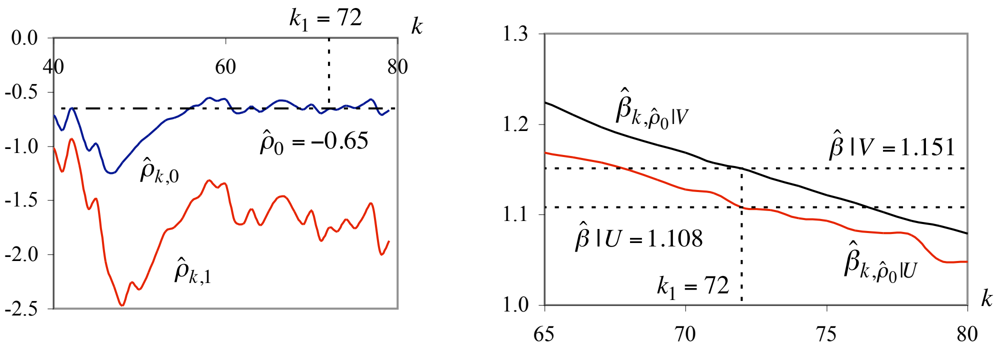


Figure 5: Estimates of second order parameters (β, ρ) for the data under analysis.

In Figure 6 we provide estimates for $\gamma = \gamma_Z$, again as a function of k. As mentioned before, apart from the Hill (H) estimator in (1.4), we have considered the moment (M), the mixed moment (MM), the generalized Hill (GH) and the bias-corrected Hill ($\overline{H}$) estimators, in (1.7), (1.8), (1.9) and (1.13), respectively. For the bias-corrected Hill estimators, we use the obvious notation $\overline{H}_0|V$ and $\overline{H}_0|U$. As expected, due to the closeness between $\hat{\beta}|V$ and $\hat{\beta}|U$, both reduced-bias statistics are quite close to each other, and highly stable in a wide region of k-values. Any stability criterion for moderate values of k, such as the ones used in Gomes *et al.* (2004) and Gomes and Pestana (2007a), among others, leads us to the choice of the tail index estimator $\overline{H}_0$, among the tail index estimators considered. Whenever we consider the tail index estimates with one decimal figure only, we get, both for $\overline{H}_0|V$ and $\overline{H}_0|U$, the estimate $\hat{\gamma} = 0.3$, the value pictured in Figure 6.

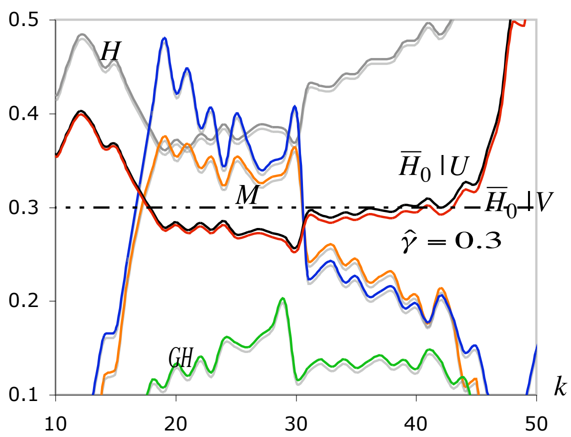


Figure 6: Estimation of parameters associated with data related with cancer of the tongue.

Due to fact that we already have estimates of the second order parameters, we shall here use a different heuristic adaptive choice of k for $\overline{H}$, already suggested in Gomes and Pestana (2007b) and sketched in the sequel. On the basis of the second-order parameters' estimator, $(\hat{\rho}_0, \hat{\beta}_0) = (-0.654, 1.151)$ we have got for the optimal sample fraction k_0^H in (1.11) the estimate $\hat{k}_0^H = 14$ and an associated tail index estimate, $H_{\hat{k}_0^H,n} = 0.45$, much larger than the value 0.3, hinted from the behaviour of the $MVRB$ estimates. Indeed, as we can see from Figure 6, the moment, the mixed moment and the Hill tend to overestimate the estimated value $\hat{\gamma} = 0.3$ whereas the generalized Hill estimator tends to underestimate it, in a region of k-values between 20 and 30. Taking into account the asymptotic behaviour of the Hill estimator in (1.4), i.e.

considering that

$$\sqrt{k}\left(\frac{H_{k,n}}{\gamma}-1-\frac{\beta\left(\frac{n}{k}\right)^{\rho}}{1-\rho}\right) \approx \text{Normal}(0,1),$$

provided that $\sqrt{k}(n/k)^{\rho} \to \lambda$, finite (de Haan and Peng, 1998), we can get approximate 95% confidence intervals for γ, given by

$$\left(\frac{H_{k,n}}{1+\frac{\beta(n/k)^{\rho}}{1-\rho}+\frac{1.96}{\sqrt{k}}},\ \frac{H_{k,n}}{1+\frac{\beta(n/k)^{\rho}}{1-\rho}-\frac{1.96}{\sqrt{k}}}\right) =: (LCL_H(k), UCL_H(k)), \tag{5.1}$$

where LCL and UCL stand for *lower control limit* and *upper control limit*, respectively. If $\lambda = 0$, we can replace the bias summand $\beta(n/k)^{\rho}/(1-\rho)$ in (5.1) by 0. The replacement of β and ρ by their estimates leads us to the approximate 95% confidence interval $(0.257, 0.643)$, with a size equal to 0.386. For this same type of levels, and for the *MVRB* estimator in (1.13), we get a null asymptotic bias, and consequently, we get 95% confidence intervals for γ of the type:

$$(LCL_{\overline{H}}(k), UCL_{\overline{H}}(k)) = \left(\frac{\overline{H}_{k,n,\hat{\beta},\hat{\rho}}}{1+\frac{1.96}{\sqrt{k}}},\ \frac{\overline{H}_{k,n,\hat{\beta},\hat{\rho}}}{1-\frac{1.96}{\sqrt{k}}}\right). \tag{5.2}$$

Regarding reduced-bias tail index estimation, we do not have up to now simple techniques to estimate the optimal threshold, but we know that such a k-value should be larger than k_0^H in (1.11). If we plot the 95% approximate confidence region in (5.1), as a function of k, the Hill estimate is sooner or later going to cross this region. We have used such a k-value for the tail index estimation through the *MVRB* estimator $\overline{H}$ in (1.13). Such a crossing level is given by

$$k_{01} \equiv k_{01}(n;\beta,\rho) = \left[(1.96(1-\rho)n^{-\rho}/|\beta|)^{2/(1-2\rho)}\right]. \tag{5.3}$$

Levels of this type are still levels such that $\sqrt{k}(n/k)^{\rho} \to \lambda$, finite, are not yet optimal for the tail index estimation through second-order reduced-bias tail index estimators, but induce already interesting estimates, as shown through simulation methods in Gomes *et al.* (2008b). More than this, since with *MVRB* estimators, like the ones in (1.12) and (1.13), we are always safe, we again suggest, after the implementation of the Algorithm in Section 4.1, which enable us to get $(\hat{\beta}, \hat{\rho})$, the use of the following steps:

Algorithm (continued; tail index estimation)

4. Plot the classical Hill estimates $H_{k,n}$ given in (1.4) and adaptively consider $H(\hat{k}_0^H)$, $\hat{k}_0^H = k_0^H(n;\hat{\beta},\hat{\rho})$, $k_0^H(n;\beta,\rho)$ given in (1.11), together with the 95% confidence interval $\big(LCL_H(\hat{k}_0^H), UCL_H(\hat{k}_0^H)\big)$ given in (5.1) for a general k.

5. Plot also the *MVRB* tail index estimates $\overline{H}(k)$ given in (1.13), associated with the estimates $(\hat{\beta},\hat{\rho})$ obtained in steps 2 and 3., and adaptively consider $\overline{H}(\hat{k}_{01})$, $\hat{k}_{01} = k_{01}(n;\hat{\beta},\hat{\rho})$, $k_{01}(n;\beta,\rho)$ given in (5.3), together with the 95% confidence interval $\big(LCL_{\overline{H}}(\hat{k}_{01}), UCL_{\overline{H}}(\hat{k}_{01})\big)$ given in (5.2) for a general k.

6. Choose the tail index estimate providing the smallest 95% confidence interval.

From a theoretical point of view, the chosen estimate in Step 6. of the **Algorithm** should be $\overline{H}(\hat{k}_{01})$, and this indeed happens for the data set under analysis. We got $\hat{k}_{01} = 31$, $\overline{H} = 0.29$ and an approximate confidence interval $(0.215, 0.448)$, with a size equal to $0.233 < 0.386$. The confidence regions for γ, as functions of k, are pictured in Figure 7.

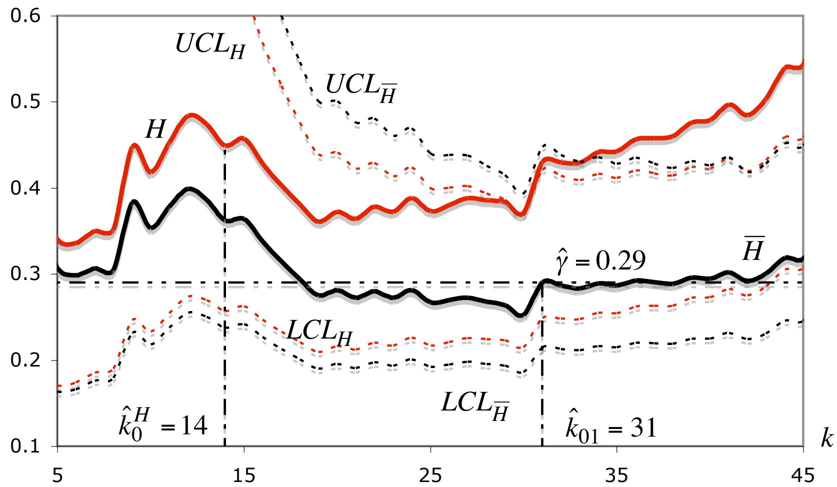


Figure 7: Estimates and confidence intervals for $\gamma = \gamma_Z$ based on the data related with cancer of the tongue.

5.1 Final comments

A thorough investigation of tailweight and of the tail index, aside from being an interesting problem of mathematical statistics, has important applications in Insurance, Finance, Telecommunication Networks and Biostatistics, namely in the area of survival analysis and pharmacokinetics, where extreme secondary effects of treatments can be much more than a simple nuisance to the pharmacy industry. The analysis of the tailweight of the parent distribution and unbiased/efficient estimation of the tail index is therefore an important practical problem.

But the theoretical importance of this issue goes far beyond its applications in extreme values analysis, and can be regarded as a tool when the central o.s.'s are at the core of the investigation. In fact, in many populations it is very difficult to obtain exact sampling distributions. Computer intensive methods are a useful tool to deal with concrete cases, but in many situations fail to provide a general theoretical framework, which is the ultimate aim in Science. Hence, in our view, a general methodology is the use of asymptotic results, that provide a sound framework to use approximations. The central limit theorem and the extremal limit theorem are two ways of dealing with the information in the sample, respectively when central o.s.'s and extremal o.s.'s prevail. Domains of attraction, both in the central limit theory and in the theory of extreme o.s.'s, are characterized in terms of the regular variation of the tails of the parent distribution. However, while in the extreme value theory for each $\gamma \in \mathbb{R}$ there is an asymptotic law, in the additive scheme different stable asymptotic laws do exist for extremely heavy tail sums of index $\gamma > \frac{1}{2}$, while finite support laws or laws with infinite support but moderately heavy tail sums of index $\gamma \in [0, \frac{1}{2}]$ are in the huge domain of attraction of the Gaussian law. This protagonism of the gaussian law as asymptotic approximation of central o.s.'s carries however an extra burden in evaluating the rate of convergence, and therefore the sample size needed to confidently vouch for the gaussian approximation. Our results on the estimation of the tail index have an evident bearing in dilucidating this problem. Indeed, the estimation of the tail index can even help on the assessment of rates of convergence in the central limit theorem.

References

[1] Beirlant, J., Dierckx, G., and Guillou, A. (2005). Estimation of the extreme-value index and generalized quantile plots. *Bernoulli* **11**:6, 949-970.

[2] Beirlant, J., Figueiredo, F., Gomes, M.I. and Vandewalle, B. (2008). Improved reduced-bias tail index and quantile estimators. *J. Statist. Plann. and Inference* **138**, 1851-1870.

[3] Beirlant, J., Guillou, A., Dierckx, G. and Fils-Viletard, A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes* **10**(3), 151-174.

[4] Beirlant, J., Vynckier, P. and Teugels, J. (1996). Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc* **91** (1996), 1659-1667.

[5] Caeiro, F. and Gomes, M.I. (2006). A new class of estimators of a "scale" second order parameter. *Extremes* **9**, 193-211, 2007.

[6] Caeiro, F., Gomes, M.I. and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat* **3**(2), 111-136.

[7] Dekkers, A., Einmahl, J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17** (1989), 1833-1855

[8] Einmahl, J.H.J, Fils-Villetard, A. and Guillou, A. (2007). Statistics of extremes under random censoring. *Bernoulli* **14**(1) (2007), 207-227.

[9] Fraga Alves, M.I., Gomes, M.I. and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* **60**(2), 194-213.

[10] Fraga Alves, M.I., Gomes, M.I., de Haan, L. and Neves, C. (2007). The Mixed Moment Estimator & Location Invariant Alternatives. Accepted at *Extremes* (Notas e Comunicações CEAUL 14/07).

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

[12] Gomes, M.I., de Haan, L. and Henriques Rodrigues, L. (2008a). Tail Index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *J. Royal Statistical Society* **B70**, Issue 1: 31-52.

[13] Gomes, M.I., Henriques Rodrigues, L. , Vandewalle, B. and Viseu, C. (2008b). A heuristic adaptive choice of the threshold for bias-corrected Hill estimators. *J. Statist. Comput. and Simulation* **78**:2, 133-150.

[14] Gomes, M.I. and Martins, M.J. (2002). "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes* **5**, 1-311.

[15] Gomes, M.I., Martins, M.J. and Neves, M.M. (2007). Improving second-order reduced-bias tail index estimation. *Revstat* **5**(2), 177-207.

[16] Gomes, M.I. and Pestana, D. (2007a). A simple second order reduced-bias tail index estimator. *J. Statist. Comp. & Simul.* **77**, No. 6, 487-504.

[17] Gomes, M.I. and Pestana, D. (2007b). A sturdy reduced bias extreme quantile (VaR) estimator. *J. Amer. Statist. Assoc.* **102**, No. 477, 280-292.

[18] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction.* Springer Verlag.

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

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

[21] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, No. 5: 1163-1174.

[22] Klein, J.P. and Moeschberger, M.L. (2005). *Datasets for Survival Analysis – Techniques for Censored and Truncated Data*, Second Edition, Springer.

[23] Pareto, V. (1965). *Oeuvres Complètes.* Geneva, Droz.

[24] Sickle-Santanello B.J., Farrar W.B., DeCenzo J.F., Keyhani-Rofagha S., Klein J., Pearl D., Laufman H. and O'Toole R.V. (1988). Technical and statistical improvements for flow cytometric DNA analysis of paraffin-embedded tissue. *Cytometry* **9**:6 (1988), 594-9

[25] Zipf, G. K. (1941). *National Unity and Disunity.* Blomington, IN: Principia Press.