Asymptotic comparison of the mixed moment and classical extreme value index estimators[†]

M. Ivette Gomes[‡]
DEIO (FCUL) and CEAUL, University of Lisbon
and
Cláudia Neves
UIMA, Department of Mathematics, University of Aveiro

A new promising extreme value index estimator, the *mixed-moment* estimator, has been recently introduced in the literature. This estimator uses not only the first moment of the top excesses of the log-observations in the sample, the basis of the classical *Hill* and *moment* estimators, but also the first moment of another type of statistics, dependent on quotients of top order statistics. In this paper we shall compare, asymptotically at optimal levels, the mixed moment estimator with the classical Hill, the moment and the usually denoted "maximum likelihood" extreme value index estimator, associated to an approximation for the excesses over a high observation provided by the generalised Pareto distribution. Again, the MM estimator cannot always dominate the alternatives, but its asymptotic performance is quite interesting.

Keywords and phrases: Statistics of extremes; semi-parametric estimation; extreme value index; asymptotic theory.

1 Introduction and preliminaries

In *Statistics of Extremes*, it is well-known that the possible non-degenerate limiting distributions of the sequence of maximum values, linearly normalized, associated to independent, identically distributed (i.i.d.) random variables (r.v.'s), $X_1, X_2, \cdots, X_n \cdots$, are the so-called *extreme value* (EV) distribution functions (d.f.'s), with the functional form,

$$EV_\gamma(x) := \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right), \ 1+\gamma x > 0 & \text{if} \quad \gamma \neq 0 \\ \exp(-\exp(-x)), \ x \in \mathbb{R} & \text{if} \quad \gamma = 0, \end{cases} \tag{1}$$

[†]Research partially supported by FCT / POCTI and POCI / FEDER.

[‡]Corresponding author: Faculdade de Ciências de Lisboa (DEIO), Bloco C6, Piso 4, 1749-016 Lisboa, Portugal

dependent on a "shape parameter" $\gamma \in \mathbb{R}$, the so-called *extreme value index*. We then say that the d.f. F, underlying the r.v.'s $X_1, X_2, \cdots$, belongs to the *domain of attraction* for maximum values of EV_γ and we use the notation $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$.

The estimation of the unknown parameter γ has been worked out by several authors. Among the most popular γ-estimators, we refer the Hill estimator, valid in the region $\gamma > 0$, the often called POT-ML (with POT-ML standing for *peaks over threshold-maximum likelihood*) or simply ML estimator, an approximate maximum likelihood estimator, valid in the region $\gamma > -1/2$, and the moment estimator, valid for all $\gamma \in \mathbb{R}$. The *Hill* estimator (H), has the functional form

$$\widehat{\gamma}_{n,k}^H := \frac{1}{k}\sum_{i=1}^{k} \ln X_{n-i+1:n} - \ln X_{n-k:n} \equiv M_{n,k}^{(1)}, \tag{2}$$

where $X_{i:n},\ 1 \le i \le n$, denotes the sample of ascending order statistics (o.s.) associated to the sample $X_i,\ 1 \le i \le n$, of i.i.d. r.v.'s with d.f. $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$. As mentioned in de Haan and Ferreira (2006), the class of d.f.'s $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$, for some $\gamma \in \mathbb{R}$, cannot be parameterized with a finite number of parameters, and consequently, there does not exist a ML estimator for γ, in such a wide class of models. There exists however an estimator, usually denoted *"maximum likelihood"* estimator, associated to an approximation provided by the *generalised Pareto* (GP) model,

$$GP_\gamma(x) = \begin{cases} 1 - (1+\gamma x)^{-1/\gamma}\,,\ x \ge 0,\ 1+\gamma x > 0 & \text{if} \quad \gamma \neq 0 \\ 1 - \exp(-x)\,,\ x \ge 0 & \text{if} \quad \gamma = 0, \end{cases} \tag{3}$$

for the excesses over a high observation. The ML estimator is based on the excesses $V_{ik} := X_{n-i+1:n} - X_{n-k:n},\ 1 \le i \le k$, which are approximately the k top o.s. associated to a sample of size k from a d.f. $GP_\gamma(\alpha x/\gamma),\ \alpha > 0$, with $GP_\gamma(x)$ given in (3). The solution of the ML equations associated to the above mentioned set-up (Davison, 1984) gives rise to,

$$\widehat{\gamma}_{n,k}^{ML} := \frac{1}{k}\sum_{i=1}^{k} \ln(1 + \hat{\alpha}\ V_{ik}), \tag{4}$$

where $\hat{\alpha}$ is the implicit ML estimator of the unknown scale parameter α. The *moment* estimator (M), has the functional expression

$$\widehat{\gamma}_{n,k}^M := M_{n,k}^{(1)} + \tfrac{1}{2}\Big\{1 - \Big(M_{n,k}^{(2)}/[M_{n,k}^{(1)}]^2 - 1\Big)^{-1}\Big\}, \quad M_{n,k}^{(j)} := \frac{1}{k}\sum_{i=1}^{k}\Big(\ln\frac{X_{n-i+1:n}}{X_{n-k:n}}\Big)^j,\ j \ge 1. \tag{5}$$

Recently, Fraga Alves, Gomes, de Haan and Neves (2007b) introduced a new estimator, based not only on the moment statistic $M_{n,k}^{(1)}$, both in (2) and (5), but also on

$$L_{n,k}^{(1)} := \frac{1}{k}\sum_{i=1}^{k}\big(1 - X_{n-k:n}/X_{n-i+1:n}\big). \tag{6}$$

They got

$$\widehat{\varphi}_{n,k} := \frac{M_{n,k}^{(1)} - L_{n,k}^{(1)}}{\left(L_{n,k}^{(1)}\right)^2} \underset{n\to\infty}{\overset{p}{\longrightarrow}} \varphi(\gamma) := \begin{cases} 1+\gamma & \text{if} \quad \gamma > 0 \\ \frac{1-\gamma}{1-2\gamma} & \text{if} \quad \gamma \le 0, \end{cases} \tag{7}$$

a monotone function of γ, and next they easily obtained, through inversion, an estimator for any *extreme value index* $\gamma \in \mathbb{R}$, called the *mixed moment* (MM) estimator, defined by

$$\widehat{\gamma}_{n,k}^{MM} := \frac{\widehat{\varphi}_{n,k} - 1}{1 + 2\min\left(\widehat{\varphi}_{n,k} - 1, 0\right)}. \tag{8}$$

The estimator in (8) seems to be an interesting alternative to the most common *extreme value index* estimators. As pointed out in Fraga Alves *et al.* (2007b), some of the most attractive features of this estimator are the following ones:

- It is valid for all $\gamma \in \mathbb{R}$ and, contrarily to the ML estimator (valid only for $\gamma > -1/2$), has a simple functional form, of the type of the M-estimator, an estimator that is also valid in the whole parameter space, $\gamma \in \mathbb{R}$.
- It exhibits properties similar to the ones of the ML estimator for a large class of models with $\gamma \ge 0$.
- For $\gamma > 0$, the asymptotic variance of the MM estimator is equal to the asymptotic variance of the ML estimator. For $\gamma \le 0$, the asymptotic variance of the MM estimator equals the one of the M estimator.
- Fraga Alves *et al.* (2007b) also propose versions of the MM estimator that are invariant for changes in location, and with similar asymptotic properties. Such estimators are based on the *peaks over random threshold* ($PORT$) methodology, as named in Araújo Santos, Fraga Alves and Gomes (2006). The new class of estimators has the same functional form as the MM estimator in (8), but $X_{i:n}$ is replaced everywhere by $X_{i:n} - X_{[np]+1:n}$, $0 < p < 1$, $1 \le i \le n$. The asymptotic variance of these location invariant versions coincides with the one of the MM estimator and the dominant component of asymptotic bias is also the same for $\gamma > 0$, being always smaller for $\gamma \le 0$, provided that we restrict ourselves to adequate values of k and p.

In Section 2 of this paper, we shall approach a second order condition, true for a general $\gamma \in \mathbb{R}$ and the one commonly used for heavy tails, i.e., for $\gamma > 0$, providing the link between the two second order conditions, in the lines of the work developed in Fraga Alves *et al.* (2007a). In Section 3, we shall briefly refer some of the asymptotic results of the different estimators under study, and finally, in Section 4, we shall proceed to the asymptotic comparison of these estimators at their optimal levels.

2 First and second order conditions

In order to obtain weak consistency of an estimator of $\gamma \in \mathbb{R}$, it is usual to assume the so-called *generalized regular variation* condition (de Haan, 1984), a necessary and sufficient condition to have $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$, i.e., with $U(t) := (1/(1-F))^{\leftarrow}(t) = \inf\left\{x : F(x) \geq 1 - \frac{1}{t}\right\}$,

$$F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma) \quad \Longleftrightarrow \quad \lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \begin{cases} \frac{x^\gamma - 1}{\gamma} & \text{if} \quad \gamma \neq 0 \\ \ln x & \text{if} \quad \gamma = 0, \end{cases} \tag{9}$$

for all $x > 0$ and for a positive measurable function a.

Beyond the first order condition in (9), and to obtain a Central Limit Theorem for the estimators of γ, under a semi-parametric framework, it is usual to assume, whenever we work with the general case $\gamma \in \mathbb{R}$, a *second order generalised condition*, i.e., it is common to assume the existence of a function A, converging towards 0 as $t \to \infty$, and such that

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = H_{\gamma,\rho}(x) := \frac{1}{\rho}\left(\frac{x^{\gamma+\rho} - 1}{\gamma + \rho} - \frac{x^\gamma - 1}{\gamma}\right) \tag{10}$$

for all $x > 0$, where $\rho \leq 0$ is a *second order* parameter, which controls the rate of convergence of maximum values, linearly normalized, towards the limiting law in (1). Note that we then have necessarily $|A| \in RV_\rho$. Since for a great variety of models we have $\rho < 0$, we add the following:

Proposition 1. *Let us assume that there exist $a(\cdot)$ and $A(\cdot)$ such that* (10) *holds, with $\rho < 0$. Then, there exist $a_0(\cdot)$ and $A_0(\cdot)$ such that*

$$\frac{\frac{U(tx)-U(t)}{a_0(t)} - \frac{x^\gamma - 1}{\gamma}}{A_0(t)} \underset{t\to\infty}{\longrightarrow} \frac{x^{\gamma+\rho} - 1}{\gamma + \rho}, \quad \textit{with } A_0(t) = \frac{A(t)}{\rho},\ a_0(t) = a(t)\left(1 - A_0(t)\right). \tag{11}$$

Proof. Indeed, from (10), we get straightforwardly

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

Consequently,

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

We have thus the validity of (11). □

From Theorem A in Draisma, de Haan, Peng and Pereira (1999), jointly with some additions in Ferreira, de Haan and Peng (2003) and de Haan and Ferreira (2006), we state the following:

Theorem 1. *If $x^F = U(\infty) > 0$ and if there exist $a(\cdot)$ and $A(\cdot)$ such that* (10) *holds, with $\rho \le 0$, $\gamma \ne \rho$, then, with $\gamma_+ := \max(0,\ \gamma)$,*

$$\overline{A}(t) := \left(\frac{a(t)}{U(t)} - \gamma_+\right) \tag{12}$$

converging towards 0 as $t \to \infty$, and with

$$l := \lim_{t\to\infty} \left(U(t) - \frac{a(t)}{\gamma}\right) \quad \textit{for } \gamma + \rho < 0, \tag{13}$$

$$\frac{\overline{A}(t)}{A(t)} \underset{t\to\infty}{\longrightarrow} c = \begin{cases} 0 & \textit{if} \quad \gamma < \rho \le 0 \\ \frac{\gamma}{\gamma+\rho} & \textit{if} \quad 0 \le -\rho < \gamma \ \textit{or} \ (0 < \gamma < -\rho \ \textit{and} \ l = 0) \\ \pm\infty & \textit{if} \quad \gamma + \rho = 0 \ \textit{or} \ (0 < \gamma < -\rho \ \textit{and} \ l \ne 0) \ \textit{or} \ \rho < \gamma \le 0. \end{cases} \tag{14}$$

The possible values of c in the (γ, ρ)-plane are pictured in Figure 1.

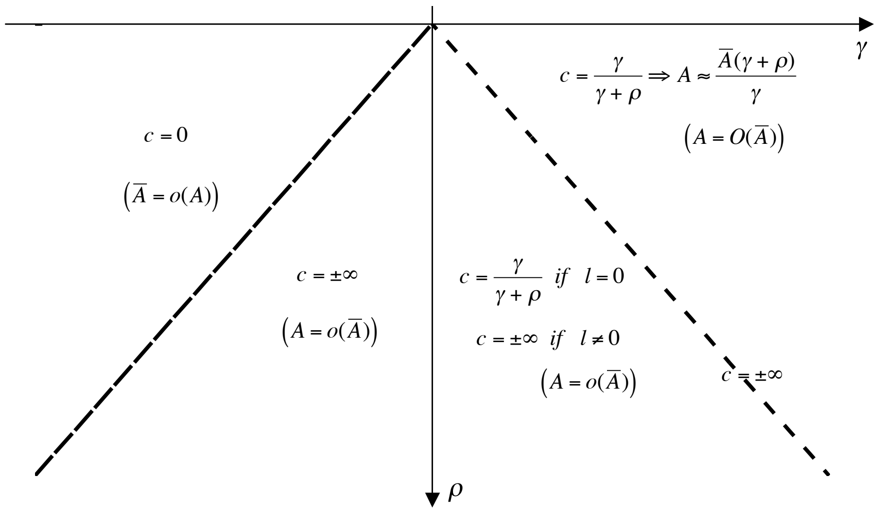

Figure 1: Possible values of c in the (γ, ρ)-plane.

For heavy tails, i.e., whenever $\gamma > 0$, the second order condition appears usually written at expenses of the behaviour of $\ln U(t)$. The link between the second order condition in (10) and the most usual second order condition for $\gamma > 0$ has been obtained in de Haan and Ferreira (2006) and Fraga Alves *et al.* (2007a):

Proposition 2 (de Haan and Ferreira, 2006; Fraga Alves *et al.*, 2007a)**.** *Let us assume that condition* (10) *holds and let c be the limit in* (14)*. If $\gamma > 0$, and with*

$$\widetilde{\rho} = \begin{cases} \rho & \textit{if} \quad c = \frac{\gamma}{\gamma+\rho} \\ -\gamma & \textit{if} \quad c = \pm\infty, \end{cases} \tag{15}$$

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{\widetilde{A}(t)} = \widetilde{K}_{\gamma,\rho}(x) := \begin{cases} \frac{x^\rho - 1}{\rho} & \textit{if} \quad c = \frac{\gamma}{\gamma+\rho} \\ \frac{x^{-\gamma}-1}{-\gamma} & \textit{if} \quad c = \pm\infty \end{cases}$$

$$=: \widetilde{K}_{\tilde\rho}(x) := \frac{x^{\tilde\rho} - 1}{\tilde\rho}, \tag{16}$$

for all $x > 0$*, where, with* $\overline{A}(t)$ *given in* (12),

$$\widetilde{A}(t) := \begin{cases} c\ A(t) = \overline{A}(t)(1 + o(1)) & \textit{if} \quad c = \frac{\gamma}{\gamma+\rho} \\ \overline{A}(t) & \textit{if} \quad c = \pm\infty. \end{cases} \tag{17}$$

We have then necessarily $\big|\widetilde{A}(t)\big| \in RV_{\widetilde\rho}$*, with* $\tilde\rho$ *given in* (15).

Remark 1. *Note first that the region* $\{(\gamma,\rho) : 0 < \gamma < -\rho < 0, l \neq 0\}$ *in the* (γ,ρ)*-plane, as well as the line* $\rho = -\gamma$*, are transformed in the line* $\widetilde\rho = -\gamma$*, in the* $(\gamma,\widetilde\rho)$*-plane. All heavy-tailed models with* $l \neq 0$ *lie on the line* $\widetilde\rho = -\gamma$*, where* $c = \pm\infty$*. Outside that line,* $c = \gamma/(\gamma+\widetilde\rho) \equiv \gamma/(\gamma+\rho)$*.*

Remark 2. *The second order condition in* (16) *is the usual second order condition for heavy tails, i.e., for* $\gamma > 0$*. Note that, from the expression of* (17)*,* $A(t) = o\big(\widetilde{A}(t)\big)$ *when* $c = \pm\infty$*. Consequently, and when* $n \to \infty$*, if* $\sqrt{k}\ \widetilde{A}(n/k) \to \lambda \neq 0$*, finite, the usual condition that enables us to get, for heavy tails, a Central Limit Theorem for classical extreme value index estimators,* $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \lambda(\gamma+\rho)/\gamma \neq 0$ *if* $c = \gamma/(\gamma+\rho)$*, but* $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} 0$ *if* $c = \pm\infty$*.*

3 Asymptotic behaviour of the estimators

Let us denote $\widehat\gamma^{\bullet}_{n,k}$ any of the estimators under study in this paper, i.e., the Hill estimator, H (Hill, 1975; de Haan and Peng, 1998), the moment estimator, M (Dekkers, Einmahl and de Haan, 1989), the "maximum likelihood" estimator, ML (Smith, 1987; Drees, de Haan and Ferreira, 2004) and the mixed moment estimator, MM (Fraga Alves *et al.*, 2007b). Under the validity of the second order condition (10), trivial adaptations of the results in de Haan and Ferreira (2006) and in Fraga Alves *et al.* (2007b) enable us to state the following:

Theorem 2. *Assume that condition* (10) *holds, with the restriction* $\gamma \neq \rho$*, and that* $x^F > 0$*. Let* $k = k_n$ *be an intermediate sequence, let* $\overline{A}$ *be the function in* (12) *and let* c *be the limit in* (14)*. If we additionally assume that we are working with values of* k *such that*

$$\lambda := \lim_{n\to\infty} \begin{cases} \sqrt{k}\ A(n/k) & \textit{if} & \gamma < \rho \leq 0\ \big(\textit{i.e.,}\ \overline{A} = o(A)\big) \\ \sqrt{k}\ \overline{A}(n/k) & \textit{otherwise} & \big(\textit{i.e.,}\ c = \frac{\gamma}{\gamma+\rho}\ \textit{or}\ c = \pm\infty\big) \end{cases} \tag{18}$$

is finite, we may guarantee that $\sqrt{k}\left(\widehat{\gamma}^{\bullet}_{n,k}-\gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} N\left(\lambda b_{\bullet}, \sigma^2_{\bullet}\right)$, *where*

$$b_{H} = \frac{1}{1-\tilde{\rho}} = \begin{cases} \frac{1}{1-\rho} & \text{if} \quad c=\frac{\gamma}{\gamma+\rho} \\ \frac{1}{1+\gamma} & \text{if} \quad c=\pm\infty, \end{cases} \quad b_{M} := \begin{cases} \frac{(1-\gamma)(1-2\gamma)}{(1-\gamma-\rho)(1-2\gamma-\rho)} & \text{if} \quad \gamma<\rho\le 0 \\ -\frac{\gamma(1+\gamma)}{(1-\gamma)(1-3\gamma)} & \text{if} \quad \rho<\gamma\le 0 \\ \frac{\gamma-\gamma\rho+\rho}{\gamma(1-\rho)^2} & \text{if} \quad c=\frac{\gamma}{\gamma+\rho} \\ \frac{\gamma}{(1+\gamma)^2} & \text{if} \quad c=\pm\infty \text{ and } \gamma>0, \end{cases}$$

$$b_{ML} :=:= \begin{cases} \frac{1+\gamma}{(1-\rho)(1-\rho+\gamma)} & \text{if} \quad -\frac{1}{2}<\gamma<\rho\le 0 \\ 0 & \text{if} \quad \rho<\gamma\le 0 \text{ and } \gamma>-\frac{1}{2} \\ \frac{(1+\gamma)(\gamma+\rho)}{\gamma(1-\rho)(1+\gamma-\rho)} & \text{if} \quad c=\frac{\gamma}{\gamma+\rho} \\ 0 & \text{if} \quad c=\pm\infty \text{ and } \gamma>0, \end{cases},$$

$$b_{MM} := \begin{cases} b_{M} & \text{if} \quad \gamma<\rho\le 0 \\ \frac{-4\gamma(1-\gamma)}{(1-3\gamma)} & \text{if} \quad \rho<\gamma\le 0 \\ b_{ML} & \text{if} \quad c=\frac{\gamma}{\gamma+\rho} \\ 0 & \text{if} \quad c=\pm\infty \text{ and } \gamma>0, \end{cases}$$

$$\sigma^2_{H} = \gamma^2, \quad \gamma>0, \qquad \sigma^2_{M} := \begin{cases} 1+\gamma^2 & \text{if} \quad \gamma\ge 0 \\ \frac{(1-\gamma)^2(1-2\gamma)(1-\gamma+6\gamma^2)}{(1-3\gamma)(1-4\gamma)} & \text{if} \quad \gamma<0, \end{cases}$$

and

$$\sigma^2_{ML} := (1+\gamma)^2, \gamma>-\tfrac{1}{2}, \quad \sigma^2_{MM} := \begin{cases} \sigma^2_{ML} & \text{if} \quad \gamma\ge 0 \\ \sigma^2_{M} & \text{if} \quad \gamma<0 \end{cases}.$$

In Figure 2 we summarise, in the plane (γ,ρ), the asymptotic bias of the different estimators.

Remark 3. *Note that for heavy tails, and regarding bias, the* MM *estimator beats the classical Hill estimator,* $\widehat{\gamma}^H_n(k) \equiv M^{(1)}_{n,k}$, *for all plane* $(\gamma,\ \rho)$. *Indeed, for values of* k *such that* $\sqrt{k}\,\overline{A}(n/k) \to \lambda \neq 0$, *the* k*-values usually considered under a heavy-tailed framework* (*see Proposition* 2), *the asymptotic bias of* $\sqrt{k}\big(\widehat{\gamma}^H_{n,k}-\gamma\big)$ *is* $\lambda/(1-\tilde{\rho})>0$, *whereas we are here able to get a null bias in the region* $\gamma+\rho\le 0,\ l\neq 0$. *Moreover,* $(1+\gamma)(\gamma+\rho)/(\gamma(1+\gamma-\rho))<1$ *for all* $\gamma>0,\ \rho<0$, *i.e.,* $b_{ML}<b_{H}$ *in all semi-plane* $\gamma>0$. *A similar comment applies to the moment estimator.*

3.1 A few comments on the null asymptotic bias

3.1.1 The ML estimator

In Theorem 2, the region of the (γ,ρ)-plane where $b_{ML}=0$ is given by:

$$\mathcal{R}_{ML} := \left\{\rho<\gamma\le 0, \text{ and } \gamma>-\tfrac{1}{2}\right\} \cup \{0<\gamma<-\rho \text{ and } l\neq 0\} \cup \{0<\gamma=-\rho\}.$$

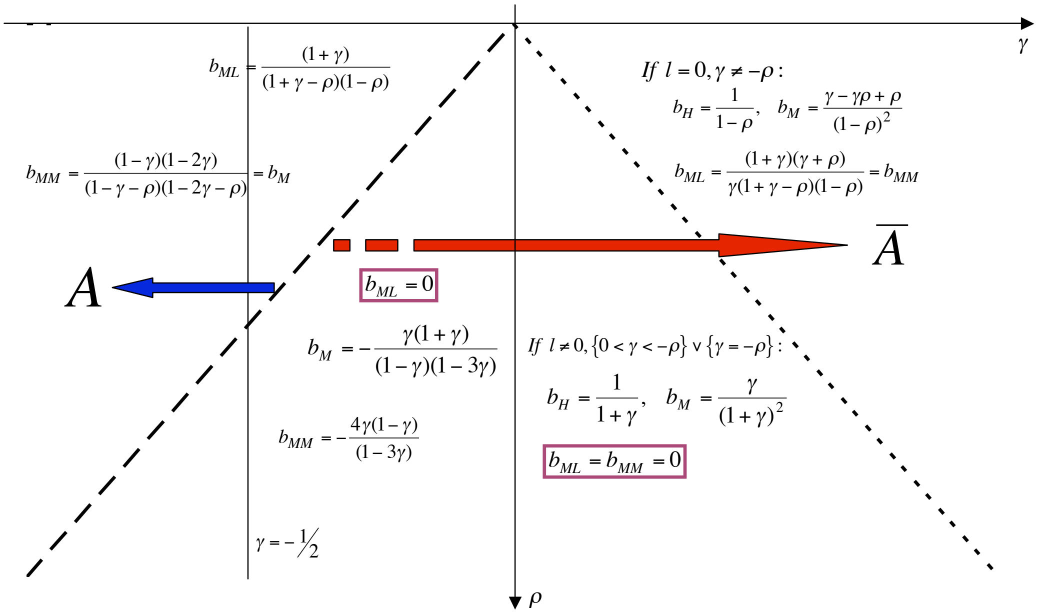

Figure 2: Asymptotic bias in the plane (γ, ρ)

It is obviously a region of the semi-plane $\gamma > -1/2$ where the ML estimator is valid and $c = \pm\infty$, i.e., $A = o(\overline{A})$. Drees *et al.* (2004) have proved that, if we consider more top o.s., or more precisely, if we consider higher values of k such that $\sqrt{k}\ A(n/k) \to \lambda < \infty$, we get a bias $b_{ML} = \frac{1+\gamma}{(1-\rho)(1-\rho+\gamma)}$, the same value we get, in the region where $c = \gamma/(\gamma+\rho)$, if we consider this same type of k-values, instead of the k-values in Theorem 2 (see Remark 2).

3.1.2 The MM estimator

We shall now try to specify, for the MM estimator, a non-null bias in the region

$$\mathcal{R}_{MM} := \{\rho < \gamma = 0\} \cup \{0 < \gamma < -\rho \text{ and } l \neq 0\} \ \cup \{0 < \gamma = -\rho\},$$

where $b_{MM} = 0$. Analogously to what has been done in Fraga Alves *et al.* (2007b), it is then necessary to split the region $\mathcal{R}_{MM}$ in 3 regions, $\mathcal{R}_{MM} = \mathcal{R}_0 \cup \mathcal{R}_1 \cup \mathcal{R}_2$, with $\mathcal{R}_0 := \{(\gamma, \rho) : \gamma = -\rho/2\}$,

$$\begin{aligned} \mathcal{R}_1 &:= \{(\gamma, \rho) : \rho < \gamma = 0 \text{ or } (0 < \gamma < -\rho/2,\ l \neq 0)\}, \\ \mathcal{R}_2 &:= \{(\gamma, \rho) : (-\rho/2 < \gamma < -\rho,\ l \neq 0) \text{ or } \gamma = -\rho\}. \end{aligned}$$

Note that in $\mathcal{R}_1$, $A = o(\overline{A}^2)$ and in $\mathcal{R}_2$, $\overline{A}^2 = o(A)$.

Theorem 3 (Fraga Alves *et al.*, 2007b)**.** *Under the condition in Theorem 2, if we assume that, in the region* $\mathcal{R}_1 \cup \mathcal{R}_2$*,*

$$\lambda := \lim_{n\to\infty} \begin{cases} \sqrt{k}\ \overline{A}^2(n/k) & \text{if} \quad (\gamma,\rho) \in \mathcal{R}_1 \\ \sqrt{k}\ A(n/k) & \text{if} \quad (\gamma,\rho) \in \mathcal{R}_2 \end{cases}$$

is finite, we can guarantee that $\sqrt{k}\left(\widehat{\gamma}^{MM}_{n,k} - \gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} N\left(\lambda\ b_{MM}, \sigma^2_{MM}\right)$*, where*

$$b_{MM} := \begin{cases} \frac{2(1+\gamma)}{(1+2\gamma)^2(1+3\gamma)} & \text{if} \quad (\gamma,\rho) \in \mathcal{R}_1 \\ \frac{1+\gamma}{(1-\rho)(1+\gamma-\rho)} \equiv b_{ML} & \text{if} \quad (\gamma,\rho) \in \mathcal{R}_2 \end{cases}$$

with σ^2_{MM} *the variance provided in Teorema 2.*

We may still state:

Theorem 4 (Fraga Alves, et al., 2007b)**.** *Let* $\widehat{\gamma}^{ML}_{n,k}$ *be a solution of the maximum likelihood equations associated to the above mentioned Generalized Pareto set-up. Let us also assume that* $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$ *with* $\gamma \geq 0$*, that (10) holds and that* $k = k_n$ *is an intermediate sequence. If* $\sqrt{k}\ \overline{A}(n/k) = O(1)$*, as* $n \to \infty$*, then* $\sqrt{k}\left(\widehat{\gamma}^{MM}_{n,k} - \widehat{\gamma}^{ML}_{n,k}\right) \underset{n\to\infty}{\overset{P}{\longrightarrow}} 0$*. This same result still holds if* $\sqrt{k}\ A(n/k) = O(1)$*, as* $n \to \infty$*, provided that* $l = 0$ *or* $\{\gamma > -\rho/2,\ l \neq 0\}$*.*

Remark 4. *The statements in Theorem 2 and Theorem 4 enable us to guarantee that whenever* $\gamma > 0$ *and* $l = 0$ *the* ML *and the* MM *estimators are equivalent. If* $l \neq 0$ *and* $\gamma > -\rho/2$ *we still have asymptotic equivalence of the two estimators, being* MM *preferable to* ML*, due to its simplicity. The unique case of heavy tails where* ML *beats* MM*, is the case* $0 < \gamma < -\rho/2$ *and* $l \neq 0$ $(\tilde{\rho} = -\gamma)$*. Then, we have to consider levels* k *such that* $\sqrt{k}\ \overline{A}^2(n/k) \to \lambda$*, finite, in order to have the asymptotic normality of* $\sqrt{k}(\widehat{\gamma}^{MM}_{n,k} - \gamma)$*, with a positive bias. As in such a region,* $\overline{A}^2 = o(A)$*, we get* $b_{ML} = 0$*.*

3.2 Asymptotic comparison of the estimators at their optimal levels

We shall next proceed to the comparison of the estimators under study at their optimal levels, in a way similar to the one used in de Haan and Peng (1998), Gomes and Martins (2001), Gomes, Miranda and Pereira (2005) and Gomes, Miranda and Viseu (2006). Let us assume that $\widehat{\gamma}^{\bullet}_{n,k}$ denotes any arbitrary semi-parametric estimator of the extreme value index γ, for which we have

$$\widehat{\gamma}^{\bullet}_{n,k} = \gamma + \frac{\sigma_\bullet}{\sqrt{k}} Z^{\bullet}_n + b_\bullet\ A(n/k) + o_p(A(n/k)) \tag{19}$$

for any intermediate sequence of integers $k = k_n$, and where $Z^{\bullet}_n$ is an asymptotically standard normal r.v. Then, we have $\sqrt{k}\left[\widehat{\gamma}^{\bullet}_{n,k} - \gamma\right] \overset{d}{\to} N(\lambda b_\bullet, \sigma^2_\bullet)$, as $n \to \infty$, provided that k is such that

$\sqrt{k}\ A(n/k) \to \lambda$, finite, as $n \to \infty$. We then write $Bias_\infty\big[\widehat{\gamma}^\bullet_{n,k}\big] := b_\bullet\ A(n/k)$ e $Var_\infty\big[\widehat{\gamma}^\bullet_{n,k}\big] := \sigma^2_\bullet/k$. The so-called Asymptotic Mean Square Error ($AMSE$) is then given by

$$AMSE\left[\widehat{\gamma}^\bullet_{n,k}\right] := \frac{\sigma^2_\bullet}{k} + b^2_\bullet\ A^2(n/k).$$

Regular variation theory (Bingham, Goldie and Teugels, 1998), enables us to show that, whenever $b_\bullet \neq 0$, there exists a function $\varphi(n) = \varphi(n, \gamma, \rho)$, such that

$$\lim_{n\to\infty} \varphi(n)\ AMSE\left[\widehat{\gamma}^\bullet_{n0}\right] = \left(\sigma^2_\bullet\right)^{-\frac{2\rho}{1-2\rho}} \left(b^2_\bullet\right)^{\frac{1}{1-2\rho}} =: LMSE\left[\widehat{\gamma}^\bullet_{n0}\right],$$

where $\widehat{\gamma}^\bullet_{n0} := \widehat{\gamma}^\bullet_{n,k_0^\bullet(n)}$ and $k_0^\bullet(n) := \arg\inf_k AMSE\left[\widehat{\gamma}^\bullet_{n,k}\right]$.

It is then usual to consider the following:

Definition 1. *Given two biased estimators $\widehat{\gamma}^{(1)}_{n,k}$ and $\widehat{\gamma}^{(2)}_{n,k}$, for which a distributional representation of the type of the one in* (19) *holds, with constants (σ_1, b_1) and (σ_2, b_2), $b_1, b_2 \neq 0$, respectively, both computed at their optimal levels, the Asymptotic Root Efficiency ($AREFF$) of $\widehat{\gamma}^{(1)}_{n0}$ relatively to $\widehat{\gamma}^{(2)}_{n0}$ is*

$$AREFF_{1|2} \equiv AREFF_{\widehat{\gamma}^{(1)}_{n0}|\widehat{\gamma}^{(2)}_{n0}} := \sqrt{LMSE\big[\widehat{\gamma}^{(2)}_{n0}\big]/LMSE\big[\widehat{\gamma}^{(1)}_{n0}\big]} = \left(\left(\frac{\sigma_2}{\sigma_1}\right)^{-2\rho}\left|\frac{b_2}{b_1}\right|\right)^{\frac{1}{1-2\rho}}. \qquad (20)$$

Remark 5. *Note that this $AREFF$ indicator has been conceived so that the highest the $AREFF$ indicator is, the better is the first estimator.*

Since the M and MM estimators are, among the ones considered, the only ones valid for all $\gamma \in \mathbb{R}$, we first present in Figure 3, the measure $AREFF_{MM|M}$, for the region $l \neq 0$ ($\gamma > 0$). Figure 4 is equivalent to Figure 3, but for $l = 0$ ($\gamma > 0$). Differences in the two figures obviously occur only in the region $\gamma + \rho < 0,\ \gamma > 0$. The general behaviour is shown in Figures 5 and 6, where we register, for heavy-tailed models ($\gamma > 0$) with $l \neq 0$ and $l = 0$, respectively, the estimator with the highest efficiency among the ones considered, i.e., the one with maximum $AREFF$, or equivalently, minimum $LMSE$. If there appear ties among the best estimators, and the MM-estimator is included in those ties, we shall place it in the figures. As mentioned before, none of the estimators can always dominate the alternatives, but the asymptotic performance of the MM estimator is quite interesting. On the basis of initial estimates of (γ, ρ, l), the following figures can help us to choose the most adequate estimator.

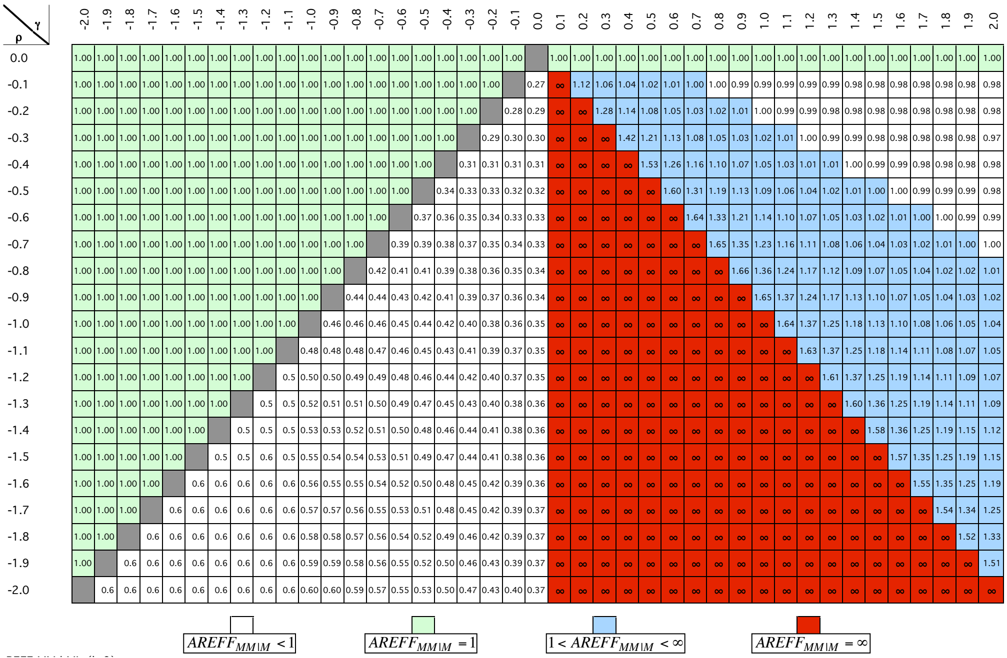

Figure 3: $AREFF_{MM|M}$, for models where $l \neq 0$ when $\gamma > 0$.

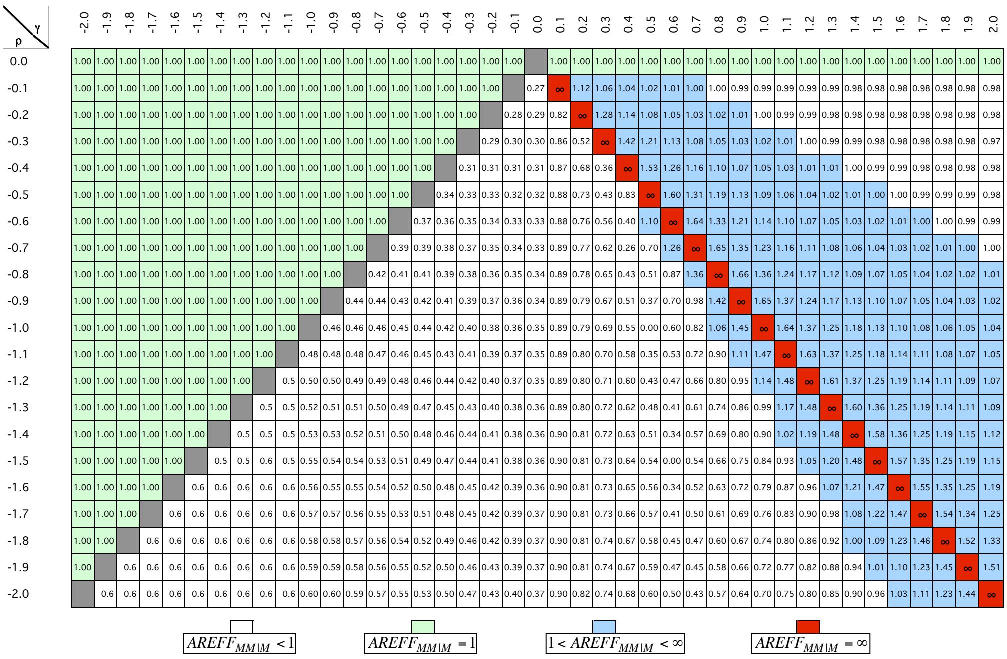

Figure 4: $AREFF_{MM|M}$, for models where $l = 0$ when $\gamma > 0$.

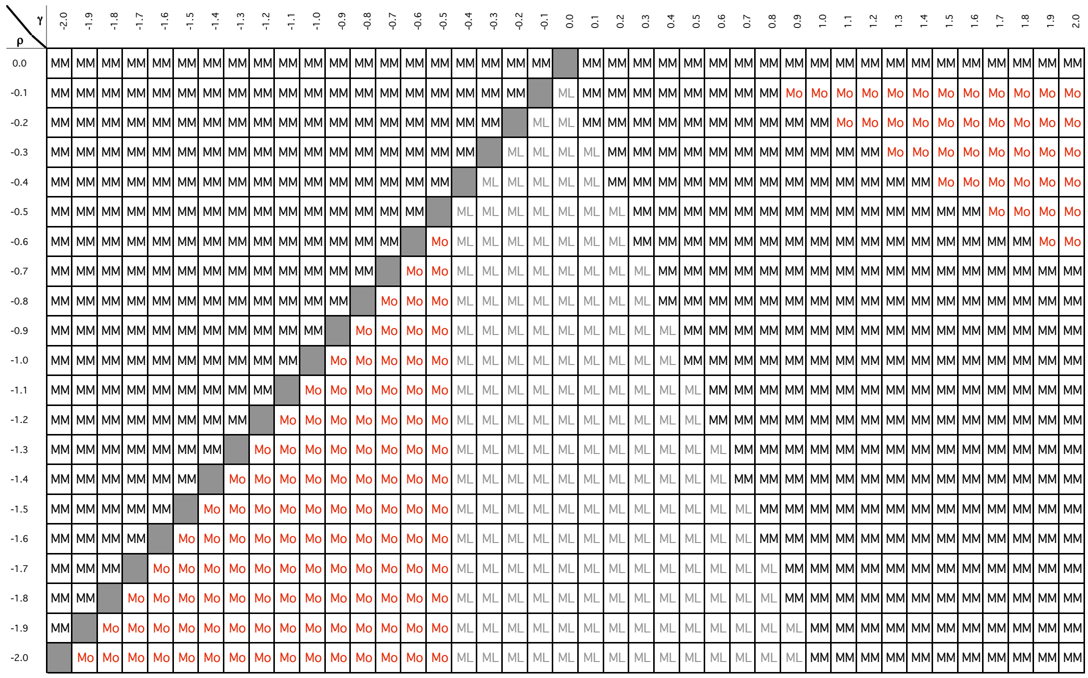

Figure 5: Estimator with minimum $LMSE$, in the class of models where $l \neq 0$ when $\gamma > 0$.

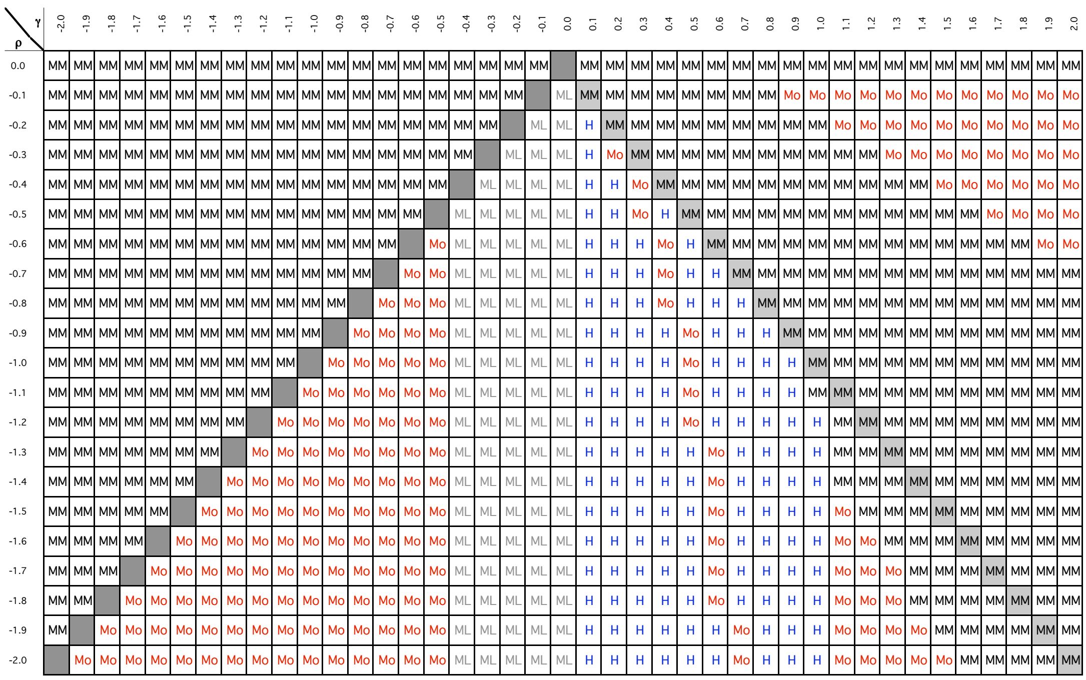

Figure 6: Estimator with minimum $LMSE$, in the class of models where $l = 0$ when $\gamma > 0$.

A few extra comments.

- For $\gamma \leq 0$, the MM and the M estimators are asymptotically equivalent if $\gamma < \rho \leq 0$, being the ML estimator the "best" one if $-1/2 < \gamma \leq 0$ and $\rho < \gamma \leq 0$.
- For $\gamma > 0$, the comparative behaviour of the estimators is less clear-cut. For models with $l = 0$, either the Hill, the moment or the mixed moment estimators can play a relevant role. If $l \neq 0$, the situation is more clear: the ML estimator is the "best" if $0 < \gamma < -\rho/2$ and the M estimator is the "best" for large values of γ and small ρ. Otherwise, the MM estimator should be eligible as the "best" one, among the estimators considered.

References

[1] Araújo Santos, P., Fraga Alves, M.I. and Gomes, M.I. (2006), Peaks over random threshold methodology for tail index and quantile estimation, *RevStat* **4** (3), 227-247.

[2] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987), *Regular Variation.* Cambridge Univ. Press.

[3] Davison, A. C. (1984), Modelling excesses over high thresholds, In Tiago de Oliveira, J. (ed.), *Statistical Extremes and Applications*, Dordrecht: D. Reidel, 461-482.

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

[5] Draisma, G., de Haan, L., Peng, L. and Themido Pereira, T. (1999), A bootstrap-based method to achieve optimality in estimating the extreme value index, *Extremes*, **2** (4): 367-404.

[6] Drees, H., Ferreira, A. and de Haan, L. (2004), On maximum likelihood estimation of the extreme value index, *Ann. Appl. Probab.* **14**: 1179-1201.

[7] Ferreira, A., de Haan, L. and Peng, L. (2003), On optimizing the estimation of high quantiles of a probability distribution, *Statistics* **37** (5): 401-434.

[8] Fraga Alves, M.I., Gomes M.I., de Haan, L. and Neves, C. (2007a), *A Note on Second Order Conditions in Extreme Value Theory: Linking General and Heavy Tails Conditions*, Notas e Comunicações CEAUL. Submitted.

[9] Fraga Alves, M.I., Gomes M.I., de Haan, L. and Neves, C. (2007b), *Mixed Moment Estimator and Location Invariant Alternatives*, Notas e Comunicações CEAUL. Submitted.

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

[11] Gomes, M.I., Miranda, C. and Pereira, H. (2005), Revisiting the role of the Jackknife methodology in the estimation of a positive extreme value index, *Comm. in Statistics – Theory and Methods* **34**, 1-20.

[12] Gomes, M.I., Miranda, C. and Viseu, C. (2006), Reduced bias extreme value index estimation and the Jackknife methodology, *Statistica Neerlandica* **60**, 4, 1-28.

[13] Haan, L. de (1984), Slow variation and characterization of domains of attraction, In Tiago de Oliveira, J. (ed.), *Statistical Extremes and Applications*, D. Reidel, Dordrecht, 31-48.

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

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

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

[17] Smith, R.L. (1987), Estimating tails of probability distributions, *Ann. Statist.* **15**: 1174-1207.