Asymptotic and Finite Sample Comparison of Two "Maximum Likelihood" Tail Index Estimators

M. Ivette Gomes

Universidade de Lisboa, Portugal

(joint work with Lígia Henriques Rodrigues)

Heavy-tailed models are quite useful in several areas of application, like computer science, telecommunication networks, insurance and finance, among others. 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.

In *Statistics of Extremes*, a model F is said to be *heavy-tailed* whenever, for some $\gamma > 0$,

$$\boxed{\overline{F} := 1 - F \in RV_{-1/\gamma} \iff U \in RV_\gamma, \quad \text{with } U(t) = F^{\leftarrow}(1 - 1/t).}$$

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 such that $\lim_{t\to\infty} g(tx)/g(t) = x^\alpha$, for all $x > 0$.

Then, we are in the domain of attraction for maxima of an *Extreme Value* distribution function (d.f.),

$$\boxed{EV_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}),\ x \geq -1/\gamma,}$$

and we write $F \in \mathcal{D}_{\mathcal{M}}(EV_{\gamma>0})$. The parameter γ is the *tail index*, the primary parameter of extreme events.

For the semi-parametric estimation of the tail index γ we need to work with an *intermediate* number k of top order statistics (o.s.'s), i.e., we should consider a sequence of integers $k = k_n$, $k \in [1, n)$, such that

$$\boxed{k = k_n \to \infty, \quad \text{and} \quad k_n = o(n), \quad \text{as} \quad n \to \infty.}$$

- In Statistics of Extremes, inference is often based on the excesses over a high random threshold $X_{n-k:n}$, represented by

$$\boxed{W_{ik} := X_{n-i+1:n} - X_{n-k:n}, \quad 1 \le i \le k < n,}$$

where $X_{i:n}$ denotes, as usual, the i-th ascending o.s., $1 \le i \le n$, associated to a random sample $(X_1, X_2, \cdots, X_n)$.

- These excesses are approximately distributed as the whole set of order statistics associated with a sample from a Generalized Pareto (GP) model, with d.f.

$$\boxed{GP(x;\gamma,\alpha) = 1 - (1 + \alpha\ x)^{-1/\gamma}, \quad x > 0 \quad (\alpha,\ \gamma > 0),}$$

a re-parametrization due to Davison [Davison, 1984]. Indeed,

$$\boxed{\alpha W_{ik} \ \approx \ Y^{\gamma}_{k-i+1:k} - 1,}$$

with Y a unit Pareto r.v., with d.f. $F_Y(y) = 1 - 1/y,\ y \ge 1$.

- We then get the so-called “maximum likelihood” (ML) extreme value index estimators [Smith, 1987; Drees, Ferreira and de Haan, 2004]. The ML estimator of γ has, with Davison’s re-parametrization, an explicit expression as a function of the ML-estimator $\widehat{\alpha} = \widehat{\alpha}_{ML}$ of α and the sample of the excesses. We have

$$\boxed{\widehat{\gamma}_n^{ML}(k) = \widehat{\gamma}_{n,\widehat{\alpha}}^{ML}(k) := \frac{1}{k}\sum_{i=1}^{k} \ln\left(1 + \widehat{\alpha}\ W_{ik}\right),}$$

  the PORT-ML tail index estimator, with PORT standing for *peaks over random threshold*, a terminology introduced in Araújo Santos, Fraga Alves and Gomes (2006).

- Dealing with heavy tails only, we are interested in the derivation of the asymptotic distributional properties of a similar maximum likelihood estimator, based also on the excesses over a high random threshold, but with a trial of accommodation of bias on the GP model underlying those excesses.

An obvious choice for an estimator of α is $1/X_{n-k:n}$. If we consider $\widehat{\alpha} = 1/X_{n-k:n}$,

$$\boxed{1 + \widehat{\alpha}\ W_{ik} = X_{n-i+1:n}/X_{n-k:n},}$$

and

$$\begin{aligned} \widehat{\gamma}^{ML}_{n,\widehat{\alpha}}(k) &:= \frac{1}{k}\sum_{i=1}^{k} \ln\left(1 + \widehat{\alpha}\ W_{ik}\right) \\ &= \frac{1}{k}\sum_{i=1}^{k} \left\{\ln X_{n-i+1:n} - \ln X_{n-k:n}\right\} =: \frac{1}{k}\sum_{i=1}^{k} V_{ik} \end{aligned}$$

is the average of the log-excesses V_{ik}, $1 \le i \le k$, i.e., it is the classical Hill estimator [Hill, 1975], denoted by $\widehat{\gamma}^{H}_{n}(k)$.

Gomes, de Haan and Henriques Rodrigues (2008) suggested the use of an adequate weighting of the log-excesses V_{ik} instead of the Hill estimator. Analagously, we shall show here that there exist weights $p_{ik} = p_{ik}(\beta, \rho)$, converging towards 1, as $k \to \infty$, dependent on a vector of second order unknown parameters $(\beta, \rho) \in \mathbb{R} \setminus \{0\} \times \mathbb{R}^-$, and such that, uniformly in i,

$$\boxed{\alpha W_{ik} - \left(Y_{k-i+1:k}^{\gamma/p_{ik}} - 1\right) = o_p\left(\alpha W_{ik} - \left(Y_{k-i+1:k}^{\gamma} - 1\right)\right).}$$

The validity of this result leads us to expect to possibly be able to get a "better" estimator of γ if we apply the approximation

$$\boxed{\alpha W_{ik} \approx Y_{k-i+1:k}^{\gamma/p_{ik}} - 1 \text{ instead of } \alpha W_{ik} \approx Y_{k-i+1:k}^{\gamma} - 1, \ 1 \le i \le k,}$$

used to support the PORT-ML.

The maximization of the log-likelihood associated to the k excesses, W_{ik}, $1 \le i \le k$, leads us to suggest the replacement of the PORT-ML estimator by a weighted combination of the statistics $\ln(1+\widehat{\alpha} W_{ik})$, $1 \le i \le k$, i.e., by

$$\boxed{\widehat{\gamma}_n^{MP}(k) \equiv \widehat{\gamma}^{MP}_{n,\widehat{\alpha},\widehat{\beta},\widehat{\rho}}(k) := \frac{1}{k}\sum_{i=1}^{k} p_{ik}(\widehat{\beta},\widehat{\rho}) \ln(1+\widehat{\alpha}_{MP} W_{ik}),}$$

here called the PORT-MP tail index estimator, with MP standing for *modified Pareto*. The estimators $(\widehat{\beta},\widehat{\rho})$ need to be adequate consistent estimators of the second order parameters (β,ρ), essentially such that $\widehat{\rho}-\rho = o_p(1/\ln n)$, as $n \to \infty$.

- We shall present a few introductory details in the field of *statistics of extremes* and introduce the new class of PORT-MP tail index estimators.

- We further provide a motivation for their consideration, assuming that all the model parameters, but the tail index γ, are known.

- One of the interesting problems to be dealt with is related with the estimation of the second order parameters β and ρ. We shall only briefly review the estimation of the second order parameters.

- The asymptotic behaviour of the PORT-MP estimator, together with the asymptotic comparison of the PORT-ML and the PORT-MP estimators at optimal levels, will be considered.

- We shall show the performance of the new PORT-MP estimator, comparatively to the classical PORT-ML estimator, through the use of simulation techniques.

- Finally, if time allows, we shall provide an overall comparison at optimal levels of a few comparable tail index estimators.

First and second order framework. In a context of heavy tails, and with the notation

$$\boxed{U(t) = F^{\leftarrow}(1 - 1/t),\ t \geq 1, \qquad F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}}$$

the generalized inverse function of the underlying model F, the first order parameter (or *tail index*) γ (> 0) appears, for every $x > 0$, as the limiting value, as $t \to \infty$, of the quotient

$$\boxed{\gamma = \lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t)}{\ln x} \quad \text{[de Haan, 1970].}}$$

Indeed, with the usual notation RV_α for the class of regularly varying functions with index of regularly variation α, we can further say

$$\boxed{F \in \mathcal{D}(EV_{\gamma>0}) \iff U \in RV_\gamma \iff 1 - F \in RV_{-1/\gamma} \quad \text{[Gnedenko, 1943].}}$$

In order to obtain information on the asymptotic behaviour of semi-parametric tail index estimators, we need further assuming a second order condition, ruling the rate of convergence in the first order condition. The *second order parameter*, ρ (≤ 0), rules such a rate of convergence, and is the parameter appearing in

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

which we often assume to hold for every $x > 0$, and where $|A|$ must then be in RV_{ρ} [Geluk and de Haan, 1987]. This condition has been widely accepted as an appropriate condition to specify the tail of a Pareto-type distribution in a semi-parametric way, and it holds for most common Pareto-type models. For technical simplicity, we shall further assume that $\rho < 0$.

Unless otherwise stated, we shall assume that we are working in Hall-Welsh class of models [Hall and Welsh, 1985], with a tail function

$$\boxed{\overline{F}(x) := 1 - F(x) = \left(\frac{x}{C}\right)^{-1/\gamma}\left(1 + \frac{\beta}{\rho}\left(\frac{x}{C}\right)^{\rho/\gamma} + o(x^{\rho/\gamma})\right), \text{ as } x \to \infty,}$$

with $C > 0$, $\beta \neq 0$ and $\rho < 0$. Equivalently, we can say that, with (β, ρ) a vector of second order parameters, the general second order condition holds with

$$\boxed{A(t) = \gamma\ \beta\ t^{\rho},\ \ \rho < 0.}$$

Models like the log-gamma and the log-Pareto $(\rho = 0)$ are thus excluded from our study. The standard Pareto is also excluded. But most heavy-tailed models used in applications, like the Fréchet, the generalized Pareto, the Burr and the Student's t d.f.'s belong to Hall-Welsh class of distributions.

Excesses over a high threshold and the GP model. On the basis of the definition of U and the universal uniform transformation, we get $X_{i:n} \overset{d}{=} U(Y_{i:n})$ again with Y a unit Pareto r.v. As for $j > i$,

$$Y_{j:n}/Y_{i:n} \overset{d}{=} Y_{j-i:n-i}, \qquad \ln Y_{i:n} \overset{d}{=} E_{i:n},$$

where E denotes a standard exponential r.v., and $Y_{n-k:n} \overset{p}{\sim} n/k$, we may indeed write, whenever we are under the first order framework,

$$\boxed{W_{ik} \overset{d}{=} X_{n-k:n} \left\{ \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})} - 1 \right\} \overset{p}{\sim} U(n/k) \left\{ Y^{\gamma}_{k-i+1:k} - 1 \right\}.}$$

Then, with $\alpha = 1/U(n/k)$ (and here is the justification for a possible choice $\hat{\alpha} = 1/X_{n-k:n}$), we have for intermediate k,

$$\boxed{W_{ik} = X_{n-i+1:n} - X_{n-k:n} \approx \left(Y^{\gamma}_{k-i+1:k} - 1 \right)/\alpha, \quad 1 \le i \le k,}$$

i.e., as mentioned before, the k excesses W_{ik}, $1 \le i \le k$, are approximately the k o.s. from the above mentioned GP model.

Accommodating bias in the Paretian excesses. Under the general second order framework, we may say that, for $1 \le i \le k$,

$$\boxed{\alpha\ W_{ik} \overset{d}{=} Y_{k-i+1:k}^{\gamma} - 1 + A(n/k) Y_{k-i+1:k}^{\gamma} \frac{Y_{k-i+1:k}^{\rho} - 1}{\rho}(1 + o_p(1)).}$$

The use of Taylor's formula for e^x, as $x \to 0$, and $\ln x$, as $x \to 1$, enables us to rewrite this equation as

$$\begin{aligned}
1 + \alpha\ W_{ik} &\overset{d}{=} Y_{k-i+1:k}^{\gamma}\left(1 + A(n/k)\frac{Y_{k-i+1:k}^{\rho} - 1}{\rho}(1 + o_p(1))\right) \\
&= e^{\gamma \ln Y_{k-i+1:k} + A(n/k)\frac{Y_{k-i+1:k}^{\rho}-1}{\rho}(1+o_p(1))} \\
&= e^{\gamma \ln Y_{k-i+1:k}\left(1+\frac{A(n/k)}{\gamma}\frac{Y_{k-i+1:k}^{\rho}-1}{\rho \ln Y_{k-i+1:k}}(1+o_p(1))\right)} \\
&= Y_{k-i+1:k}^{\gamma\left(1+\frac{A(n/k)}{\gamma}\frac{Y_{k-i+1:k}^{\rho}-1}{\rho \ln Y_{k-i+1:k}}(1+o_p(1))\right)} \approx Y_{k-i+1:k}^{\gamma\ e^{\frac{A(n/k)}{\gamma}\frac{Y_{k-i+1:k}^{\rho}-1}{\rho \ln Y_{k-i+1:k}}}}
\end{aligned}$$

Consequently,

$$\boxed{\alpha W_{ik} - \left(Y_{k-i+1:k}^{\gamma e^{\frac{A(n/k)}{\gamma} \frac{Y_{k-i+1:k}^{\rho}-1}{\rho \ln Y_{k-i+1:k}}}} - 1 \right) = o_p \left(\alpha W_{ik} - \left(Y_{k-i+1:k}^{\gamma} - 1 \right) \right).}$$

Since we can approximately write

$$\boxed{\frac{Y_{k-i+1:k}^{\rho} - 1}{\rho \ln Y_{k-i+1:k}} \approx -\frac{(i/k)^{-\rho} - 1}{\rho \ln(i/k)} =: \psi_{ik} \equiv \psi(i/k) \equiv \psi_{ik}(\rho) \ \ [\psi_{kk} \equiv 1],}$$

with ψ_{ik} a limited function, we expect to get a less biased estimator if we assume that the random excess W_{ik} comes from a GP model with a shape parameter not equal to γ, as it is usually done, but dependent on i (and k) and given by

$$\boxed{\gamma_{ik} := \gamma \ e^{\beta \ (\frac{n}{k})^{\rho} \psi_{ik}}, \qquad 1 \le i \le k,}$$

for models in Hall-Welsh class.

We are thus going to base inference on the fact that there exists a parameter α such that $W_{ik} = X_{n-i+1:n} - X_{n-k:n}$ comes from a GP model, with d.f. $GP(x;\gamma_{ik},\alpha)$, for every $1 \le i \le k$. The likelihood function of $\underline{W} = \left(W_{ik},\ 1 \le i \le k\right)$ is then proportional to

$$L(\gamma, \beta, \rho;\ \underline{W}) = \frac{\alpha^k}{\gamma^k} \prod_{i=1}^{k} e^{-\beta(n/k)^\rho \psi_{ik}} \left(1 + \alpha\ W_{ik}\right)^{-\frac{1}{\gamma} e^{-\beta(n/k)^\rho \psi_{ik}} - 1},$$

and consequently we have

$$\ln L(\gamma, \beta, \rho;\ \underline{W}) = k \ln \alpha - k \ln \gamma - \beta(n/k)^\rho \sum_{i=1}^{k} \psi_{ik} - \sum_{i=1}^{k} \ln(1 + \alpha\ W_{ik})$$
$$- \frac{1}{\gamma} \sum_{i=1}^{k} e^{-\beta(n/k)^\rho \psi_{ik}} \ln(1 + \alpha\ W_{ik}).$$

The maximization of $\ln L(\gamma, \beta, \rho;\ \underline{W})$ leads us to an explicit expression for the tail index estimator, given by

$$\boxed{\widehat{\gamma}_n^{MP}(k) \equiv \widehat{\gamma}^{MP}_{n,\widehat{\alpha},\widehat{\beta},\widehat{\rho}}(k) := \frac{1}{k}\sum_{i=1}^{k} e^{-\widehat{\beta}\ (n/k)^{\widehat{\rho}}\ \widehat{\psi}_{ik}}\ \ln(1+\widehat{\alpha}W_{ik}),}$$

with $\widehat{\psi}_{ik} = -\frac{(i/k)^{-\widehat{\rho}}-1}{\widehat{\rho}\ln(i/k)}$. Consequently, the weights $p_{ik}(\beta,\rho)$ suggested at the beginning are given by $\exp\{-\beta(n/k)^{\rho}\psi_{ik}\}$, $1 \le i \le k$. If we now replace here $\widehat{\alpha}$ by $1/X_{n-k:n}$, we get the *weighted log-excesses* or *weighted-Hill* (WH) estimator,

$$\boxed{\widehat{\gamma}_n^{WH}(k) \equiv \widehat{\gamma}^{WH}_{n,\widehat{\beta},\widehat{\rho}}(k) := \frac{1}{k}\sum_{i=1}^{k} e^{-\widehat{\beta}\ (n/k)^{\widehat{\rho}}\ \widehat{\psi}_{ik}}\ \ln\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right),}$$

introduced and studied in Gomes *et al.* (2008). This is a minimum-variance second-order reduced-bias estimator, for adequate levels k and an adequate external estimation of the second order parameters. We shall next make explicit the estimators of the second order parameters to be used in this paper.

An algorithm for the estimation of second order parameters β and ρ. We propose the following **Algorithm**:

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

$$\boxed{\hat{\rho}_\tau(k) := \min\left\{0,\ \left(3(T_n^{(\tau)}(k) - 1)\right) / \left(T_n^{(\tau)}(k) - 3\right)\right\},}$$

where, with $\boxed{M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left\{\ln \frac{X_{n-i+1:n}}{\ln X_{n-k:n}}\right\}^j,\ j = 1, 2, 3,}$ and the notation $a^{b\tau} = b \ln a$ whenever $\tau = 0$,

$$\boxed{T_n^{(\tau)}(k) := \frac{\left(M_n^{(1)}(k)\right)^\tau - \left(M_n^{(2)}(k)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k)/2\right)^{\tau/2} - \left(M_n^{(3)}(k)/6\right)^{\tau/3}}, \quad \tau \in \Re.}$$

2. Consider $\{\widehat{\rho}_\tau(k)\}_{k\in\mathcal{K}}$, for large k, say values $k \in \mathcal{K} = \left(\left[n^{0.995}\right],\ \left[n^{0.999}\right]\right)$, and compute their median, denoted ρ_τ. Next choose the *tuning parameter* $\tau^* := \arg\min_\tau \sum_{k\in\mathcal{K}} (\widehat{\rho}_\tau(k) - \rho_\tau)^2$.
3. Work then with

$$\boxed{(\widehat{\rho}_{\tau^*}, \widehat{\beta}_{\tau^*}) := (\widehat{\rho}_{\tau^*}(k_1), \widehat{\beta}_{\widehat{\rho}_{\tau^*}}(k_1)), \qquad k_1 = \left[n^{0.995}\right],}$$

and

$$\boxed{\widehat{\beta}_{\widehat{\rho}}(k) := \left(\frac{k}{n}\right)^{\widehat{\rho}} \ \frac{d_k(\widehat{\rho})\ D_k(0) - D_k(\widehat{\rho})}{d_k(\widehat{\rho})\ D_k(\widehat{\rho}) - D_k(2\widehat{\rho})}, \quad d_k(\alpha) := \frac{1}{k}\sum_{i=1}^{k}(i/k)^{-\alpha},}$$

$$\boxed{D_k(\alpha) := \frac{1}{k}\sum_{i=1}^{k}(i/k)^{-\alpha}U_i, \quad U_i := i\left\{ln\frac{X_{n-i+1_n}}{X_{n-i:n}}\right\}.}$$

For asymptotic and finite sample details on these estimators of ρ, see Fraga Alves, Gomes and de Haan (2003). The above-mentioned estimator of β has been introduced in Gomes and Martins (2002), where conditions that enable its asymptotic normality have been set, whenever ρ is estimated at a level k_1 of a larger order than the level k used for the estimation of β. Details on the asymptotic distribution of $\widehat{\beta}_{\widehat{\rho}(k)}(k)$ may be found in Gomes *et al.* (2008).

Steps 1. and 2. of the algorithm lead in almost all situations to the *tuning parameter* $\tau^* = 0$ whenever $|\rho| \leq 1$ and $\tau^* = 1$, otherwise. Such an educated guess usually provides better results than a possibly "noisy" estimation of τ, and it is highly recommended in practice. For details on this and similar algorithms for the ρ-estimation, see Gomes and Pestana (2007).

Motivation for the new estimators — only γ is unknown. Let us assume that everything is known, apart from γ. Then,

Theorem 1. *For models in Hall-Welsh class, and for intermediate levels k, we get for $\widehat{\gamma}^{MP}_{n,\alpha,\beta,\rho}(k)$, with $\widehat{\gamma}^{MP}_{n,\widehat{\alpha},\widehat{\beta},\widehat{\rho}}(k)$ provided before, an asymptotic distributional representation of the type*

$$\widehat{\gamma}^{MP}_{n,\alpha,\beta,\rho}(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, N_k + o_p(A(n/k)),$$

where N_k is asymptotically a standard normal r.v. Consequently $\sqrt{k}\left(\widehat{\gamma}^{MP}_{n,\alpha,\beta,\rho}(k) - \gamma\right)$ is asymptotically normal not only when $\sqrt{k}\ A(n/k) \longrightarrow 0$, but also when $\sqrt{k}\ A(n/k) \longrightarrow \lambda \neq 0$, finite, as $n \to \infty$.

The main problems to be dealt with are related with the study of how the estimation of (α, β, ρ) affects the asymptotic distributional behaviour of $\widehat{\gamma}^{MP}_{n,\alpha,\beta,\rho}(k)$.

Asymptotic behaviour of the PORT-MP tail index estimator.
Let us assume that we have access to the sample of the excesses, $\underline{W} = (W_{ik}, 1 \leq i \leq k)$, and we are interested in the PORT-MP estimator $\widehat{\gamma}^{MP}_{n,\widehat{\alpha},\widehat{\beta},\widehat{\rho}}(k)$, an explicit function of $\widehat{\alpha} = \widehat{\alpha}_{MP}$, the PORT-MP estimator of α in a modified generalized Pareto model, and external estimators of the second order parameters (β, ρ). The asymptotic behaviour of both the PORT-MP estimator depends essentially on the behaviour of the r.v.'s:

$$B := \tfrac{1}{k}\textstyle\sum_{i=1}^{k} \ \ln(1+\alpha\ W_{ik}), \quad B_{(j)} := \frac{1}{k}\sum_{i=1}^{k} \ \psi_{ik}^{j-1} e^{-\beta(n/k)^{\rho}\psi_{ik}} \ \ln(1+\alpha\ W_{ik}),$$

$$C := \tfrac{1}{k}\textstyle\sum_{i=1}^{k} \ \tfrac{\alpha\ W_{ik}}{1+\alpha W_{ik}}, \qquad C_{(j)} := \frac{1}{k}\sum_{i=1}^{k} \ \psi_{ik}^{j-1} e^{-\beta(n/k)^{\rho}\psi_{ik}} \ \frac{\alpha\ W_{ik}}{1+\alpha W_{ik}},$$

$$D := \tfrac{1}{k}\textstyle\sum_{i=1}^{k} \ \tfrac{\alpha\ W_{ik}}{(1+\alpha\ W_{ik})^2}, \qquad D_{(j)} := \frac{1}{k}\sum_{i=1}^{k} \ \psi_{ik}^{j-1} e^{-\beta(n/k)^{\rho}\psi_{ik}} \ \frac{\alpha\ W_{ik}}{(1+\alpha\ W_{ik})^2}.$$

The log-likelihood associated to the MP-scheme can thus be written as

$$\boxed{\ln L(\gamma, \beta, \rho;\ \underline{W}) = k \ln \alpha - k \ln \gamma - \beta (n/k)^{\rho} \sum_{i=1}^{k} \psi_{ik} - kB - \frac{kB_{(0)}}{\gamma}.}$$

Notice next that, also for $j \geq 1$,

$$\frac{\partial B}{\partial \alpha} = \frac{C}{\alpha}, \quad \frac{\partial C}{\partial \alpha} = \frac{D}{\alpha}, \quad \frac{\partial B_{(j)}}{\partial \alpha} = \frac{C_{(j)}}{\alpha}, \quad \frac{\partial C_{(j)}}{\partial \alpha} = \frac{D_{(j)}}{\alpha}.$$

$$\frac{\partial B}{\partial \beta} = \frac{\partial C}{\partial \beta} = 0, \quad \frac{\partial B_{(j)}}{\partial \beta} = -\frac{A(n/k)\ B_{(j+1)}}{\beta\gamma},$$

$$\frac{\partial C_{(j)}}{\partial \beta} = -\frac{A(n/k)\ C_{(j+1)}}{\beta\gamma}.$$

Consequently,

$$\frac{\partial \ln L(\gamma, \beta, \rho;\ \underline{W})}{\partial \alpha} = \frac{k}{\alpha}\left(1 - C - \frac{C_{(1)}}{\gamma}\right).$$

As the MP estimator of γ is, under the above mentioned assumptions, $\widehat{\gamma}^{MP} = \hat{B}_{(1)}$, the MP estimator of α is solution of the equation

$$\boxed{\widehat{C} + \widehat{C}_{(1)}/\widehat{B}_{(1)} - 1 \equiv 0.}$$

If we decide for an external consistent estimation of β, as well as of ρ, with the additional condition $\widehat{\rho} - \rho = o_p(1/\ln n)$, we may write:

$$\boxed{\widehat{\gamma}^{MP}_{n,\widehat{\alpha},\widehat{\beta},\widehat{\rho}}(k) = \widehat{B}_{(1)} = B_{(1)} + C_{(1)}\ \frac{\widehat{\alpha}_{MP} - \alpha}{\alpha}\ (1 + o_p(1)),}$$

and, since

$$\boxed{\frac{\partial\left(C + C_{(1)}/B_{(1)} - 1\right)}{\partial\alpha} = \frac{D + D_{(1)}/B_{(1)} - (C_{(1)}/B_{(1)})^2}{\alpha},}$$

$\widehat{\alpha}_{MP} \equiv \widehat{\alpha}_{MP}(k)$ is such that

$$\widehat{C} + \widehat{C}_{(1)}/\widehat{B}_{(1)} - 1 \equiv 0$$

$$= C + \frac{C_{(1)}}{B_{(1)}} - 1 + \frac{\widehat{\alpha}_{MP} - \alpha}{\alpha} \left(D + \frac{D_{(1)}}{B_{(1)}} - \left(\frac{C_{(1)}}{B_{(1)}} \right)^2 \right) (1 + o_p(1)),$$

i.e.,

$$\boxed{\frac{\widehat{\alpha}_{MP} - \alpha}{\alpha} = \frac{1 - C - C_{(1)}/B_{(1)}}{D + D_{(1)}/B_{(1)} - (C_{(1)}/B_{(1)})^2} (1 + o_p(1)),}$$

and, as seen before,

$$\boxed{\widehat{\gamma}^{MP}(k) = \widehat{B}_{(1)} = B_{(1)} + C_{(1)} \frac{\widehat{\alpha}_{MP} - \alpha}{\alpha} (1 + o_p(1)).}$$

The asymptotic distributional behavior of the MP-estimators comes then easily from the results above.

- If only γ is unknown, Theorem 1 holds for $\widehat{\gamma}^{MP}_{n,\alpha,\beta,\rho}$, i.e., $\widehat{\gamma}^{MP}_{n,\alpha,\beta,\rho}$ is a MVRB estimator.

- The same result holds for $\widehat{\gamma}^{MP}_{n,\alpha,\widehat{\beta},\widehat{\rho}}$ if we assume α known and we estimate β and ρ externally, in an adequate way, i.e., so that $\widehat{\rho}-\rho = o_p(1/\ln n)$ and $\widehat{\beta}-\beta = o_p(1)$ for all k on which we usually base $\widehat{\gamma}^{MP}_{n,\alpha,\widehat{\beta},\widehat{\rho}}(k)$, i.e., such that $k = o(n)$ and $\sqrt{k}\ A(n/k) = O(1)$, as $n \to \infty$.

- If we estimate α and γ jointly through the maximum likelihood procedure, we may state the following:

Theorem 2. *For intermediate* k *and in Hall-Welsh class of models, then with* S_k *asymptotically standard normal r.v. and the notation*

$$\boxed{b_{MP} := -\frac{(1+\gamma)(1+2\gamma)}{\gamma^3}\Big(\frac{1}{\rho}\ln\frac{(1+\gamma)(1-\rho)}{1+\gamma-\rho} + \frac{\gamma}{1+\gamma-\rho}\Big),}$$

we have the asymptotic distributional representation,

$$\boxed{\widehat{\gamma}^{MP}_{n,\widehat{\alpha},\widehat{\beta},\widehat{\rho}}(k) \overset{d}{=} \gamma + \frac{(1+\gamma)}{\sqrt{k}} S_k + b_{MP} A(n/k)(1+o_p(1)).}$$

For the estimator $\widehat{\gamma}^{ML}_n(k)$*, we have the asymptotic distributional representation*

$$\boxed{\widehat{\gamma}^{ML}_n(k) \overset{d}{=} \gamma + \frac{(1+\gamma)}{\sqrt{k}} M_k + \frac{(1+\gamma)(\gamma+\rho)\ A(n/k)}{\gamma(1-\rho)(1-\rho+\gamma)}\ (1+o_p(1)),}$$

with M_k *also asymptotically standard normal.*

Remark 1. *As can be seen from Theorem 2, the PORT-MP tail index estimator is no longer a MVRB estimator or even a second-order reduced-bias tail index estimator, i.e., the estimation of α through maximum-likelihood gives rise to a dominant component of bias of the order of $A(n/k)$.*

Relatively to Smith's result, rephrased in this context in Theorem 2 (i.e. with the replacement of a fixed threshold u by a random threshold $X_{n-k:n}$), we have the same asymptotic variance, $(1+\gamma)^2$, but a change in bias, although both bias are of the same order if $\gamma+\rho \neq 0$. If $\gamma+\rho=0$ the PORT-ML estimator, being a second-order reduced-bias estimator of γ, is expected to outperform the PORT-MP estimator.

Asymptotic comparison at optimal levels. We now proceed to an asymptotic comparison of the estimators at their optimal levels in the lines of de Haan and Peng (1998), Gomes and Martins (2001), Gomes *el al.* (2005), Gomes *el al.* (2007) and Gomes and Neves (2007). Suppose that $\widehat{\gamma}_n^\bullet(k)$ is a general semi-parametric estimator of the tail index estimator, with distributional representation,

$$\boxed{\widehat{\gamma}_n^\bullet(k) = \gamma + \frac{\sigma_\bullet}{\sqrt{k}} Z_n^\bullet + b_\bullet A(n/k) + o_p(A(n/k)),}$$

which hold for any intermediate k, and where $Z_n^\bullet$ is an asymptotically standard normal r.v.. Then we have,

$$\boxed{\sqrt{k}[\widehat{\gamma}_n^\bullet(k) - \gamma] \xrightarrow{d} \mathbb{N}(\lambda b_\bullet, \sigma_\bullet^2), \text{ as } n \to \infty,}$$

provided k is such that $\sqrt{k} A(n/k) \to \lambda$, finite, as $n \to \infty$.

The Asymptotic Mean Square Error (AMSE) is given by

$$\boxed{AMSE[\widehat{\gamma}_n^{\bullet}(k)] := \frac{\sigma_{\bullet}^2}{k} + b_{\bullet}^2 A^2(n/k),}$$

where $Bias_\infty[\widehat{\gamma}_n^{\bullet}(k)] := b_{\bullet}A(n/k)$ and $Var_\infty[\widehat{\gamma}_n^{\bullet}(k)] := \sigma_{\bullet}^2/k$.

Let $k_0^{\bullet}$:=arg inf$_k$ $AMSE[\widehat{\gamma}_n^{\bullet}(k)]$ be the optimal level for the estimation of γ through $\widehat{\gamma}_n^{\bullet}(k)$, i.e., the level associated to a minimum AMSE, and let us denote $\widehat{\gamma}_{n0}^{\bullet} := \widehat{\gamma}_n^{\bullet}(k_0^{\bullet}(n))$, the estimator computed at its optimal level. The use of regular variation theory [Bingham, Goldie and Teugels, 1987] enabled Dekkers and de Haan (1993) to prove that, whenever $b_{\bullet} \neq 0$, $\exists\ \varphi(n) = \varphi(n;\rho,\gamma)$, dependent only on the underlying model, and not on the estimator, such that

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

It is then sensible to consider the following:

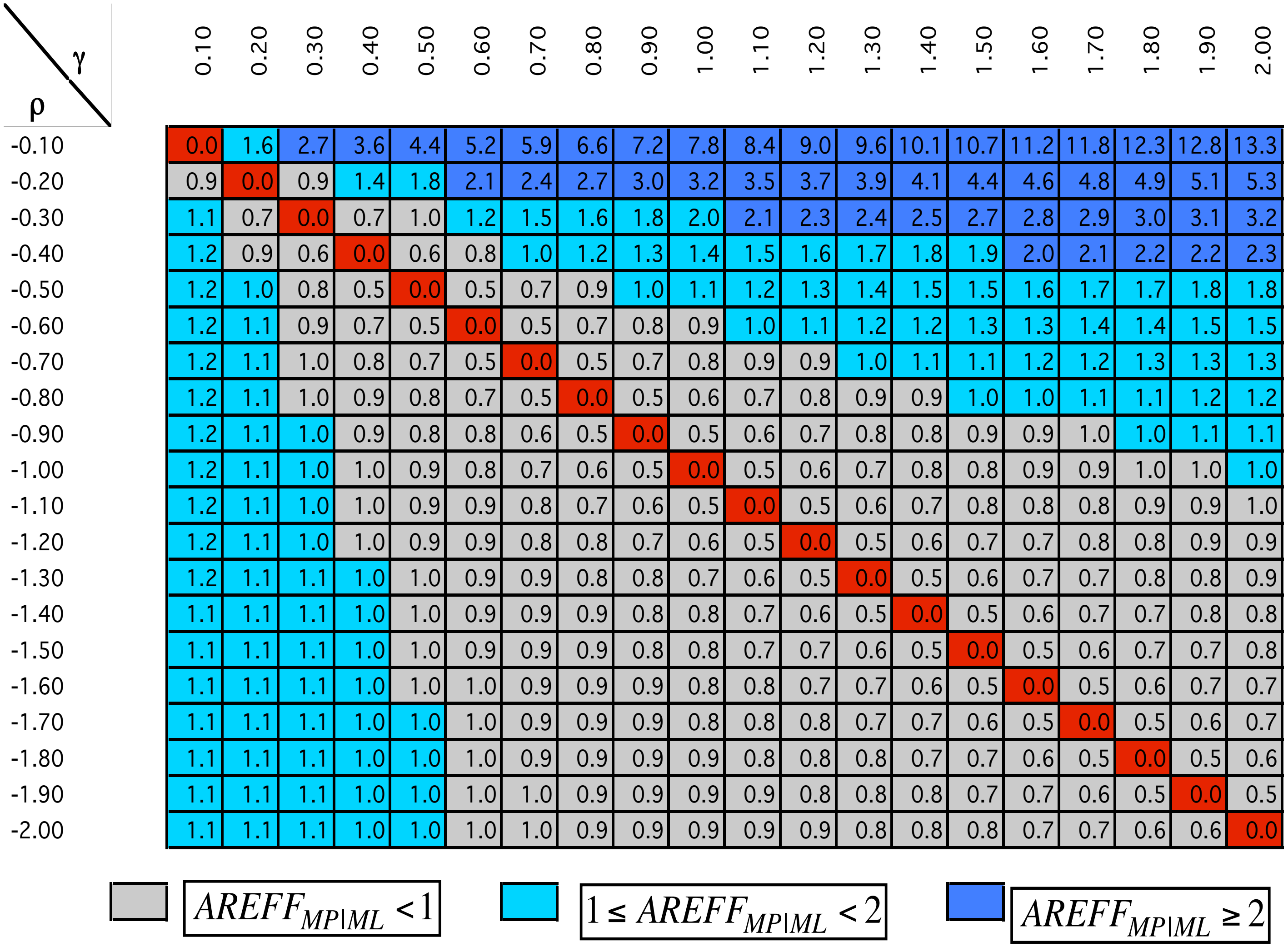
Definition 1. *Given* $\widehat{\gamma}_{n0}^{(1)} = \widehat{\gamma}_n^{\bullet}(k_0^{(1)}(n))$ *and* $\widehat{\gamma}_{n0}^{(2)} = \widehat{\gamma}_n^{\bullet}(k_0^{(2)}(n))$, *two biased estimators* $\widehat{\gamma}_n^{(1)}$ *and* $\widehat{\gamma}_n^{(2)}$ *for which distributional representations of the above-mentioned type hold 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* $\widehat{\gamma}_n^{(1)}$ *relatively to* $\widehat{\gamma}_n^{(2)}$ *is*

$$\boxed{AREFF_{1|2} \equiv AREFF_{\gamma_n^{(1)}|\gamma_n^{(2)}} := \sqrt{LMSE\left[\widehat{\gamma}_{n0}^{(2)}\right]/LMSE\left[\widehat{\gamma}_{n0}^{(1)}\right]},}$$

with LMSE given before.

Remark 2. *Note that this measure was devised so that the higher AREFF measure, the better the first estimator is.*

The $AREFF$ of $\widehat{\gamma}_n^{MP}$ relatively to $\widehat{\gamma}_n^{ML}$ is presented in Figure 1.



γ ρ	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
-0.10	0.0	1.6	2.7	3.6	4.4	5.2	5.9	6.6	7.2	7.8	8.4	9.0	9.6	10.1	10.7	11.2	11.8	12.3	12.8	13.3
-0.20	0.9	0.0	0.9	1.4	1.8	2.1	2.4	2.7	3.0	3.2	3.5	3.7	3.9	4.1	4.4	4.6	4.8	4.9	5.1	5.3
-0.30	1.1	0.7	0.0	0.7	1.0	1.2	1.5	1.6	1.8	2.0	2.1	2.3	2.4	2.5	2.7	2.8	2.9	3.0	3.1	3.2
-0.40	1.2	0.9	0.6	0.0	0.6	0.8	1.0	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.2	2.3
-0.50	1.2	1.0	0.8	0.5	0.0	0.5	0.7	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.5	1.6	1.7	1.7	1.8	1.8
-0.60	1.2	1.1	0.9	0.7	0.5	0.0	0.5	0.7	0.8	0.9	1.0	1.1	1.2	1.2	1.3	1.3	1.4	1.4	1.5	1.5
-0.70	1.2	1.1	1.0	0.8	0.7	0.5	0.0	0.5	0.7	0.8	0.9	0.9	1.0	1.1	1.1	1.2	1.2	1.3	1.3	1.3
-0.80	1.2	1.1	1.0	0.9	0.8	0.7	0.5	0.0	0.5	0.6	0.7	0.8	0.9	0.9	1.0	1.0	1.1	1.1	1.2	1.2
-0.90	1.2	1.1	1.0	0.9	0.8	0.8	0.6	0.5	0.0	0.5	0.6	0.7	0.8	0.8	0.9	0.9	1.0	1.0	1.1	1.1
-1.00	1.2	1.1	1.0	1.0	0.9	0.8	0.7	0.6	0.5	0.0	0.5	0.6	0.7	0.8	0.8	0.9	0.9	1.0	1.0	1.0
-1.10	1.2	1.1	1.0	1.0	0.9	0.9	0.8	0.7	0.6	0.5	0.0	0.5	0.6	0.7	0.8	0.8	0.8	0.9	0.9	1.0
-1.20	1.2	1.1	1.0	1.0	0.9	0.9	0.8	0.8	0.7	0.6	0.5	0.0	0.5	0.6	0.7	0.7	0.8	0.8	0.9	0.9
-1.30	1.2	1.1	1.1	1.0	1.0	0.9	0.9	0.8	0.8	0.7	0.6	0.5	0.0	0.5	0.6	0.7	0.7	0.8	0.8	0.9
-1.40	1.1	1.1	1.1	1.0	1.0	0.9	0.9	0.9	0.8	0.8	0.7	0.6	0.5	0.0	0.5	0.6	0.7	0.7	0.8	0.8
-1.50	1.1	1.1	1.1	1.0	1.0	0.9	0.9	0.9	0.8	0.8	0.7	0.7	0.6	0.5	0.0	0.5	0.6	0.7	0.7	0.8
-1.60	1.1	1.1	1.1	1.0	1.0	1.0	0.9	0.9	0.9	0.8	0.8	0.7	0.7	0.6	0.5	0.0	0.5	0.6	0.7	0.7
-1.70	1.1	1.1	1.1	1.0	1.0	1.0	0.9	0.9	0.9	0.8	0.8	0.8	0.7	0.7	0.6	0.5	0.0	0.5	0.6	0.7
-1.80	1.1	1.1	1.1	1.0	1.0	1.0	0.9	0.9	0.9	0.9	0.8	0.8	0.8	0.7	0.7	0.6	0.5	0.0	0.5	0.6
-1.90	1.1	1.1	1.1	1.0	1.0	1.0	1.0	0.9	0.9	0.9	0.9	0.8	0.8	0.8	0.7	0.7	0.6	0.5	0.0	0.5
-2.00	1.1	1.1	1.1	1.0	1.0	1.0	1.0	0.9	0.9	0.9	0.9	0.9	0.8	0.8	0.8	0.7	0.7	0.6	0.6	0.0

▨ $AREFF_{MP|ML} < 1$ ▤ $1 \leq AREFF_{MP|ML} < 2$ ■ $AREFF_{MP|ML} \geq 2$

As may be seen, the gain in efficiency for the PORT-MP estimator happens for two regions of values of (γ, ρ). In the first region we have $\gamma \leq -a\rho$, with $a < 1/2$ and in the second one we have $\gamma \geq -b\rho$ with $b \geq 2$.

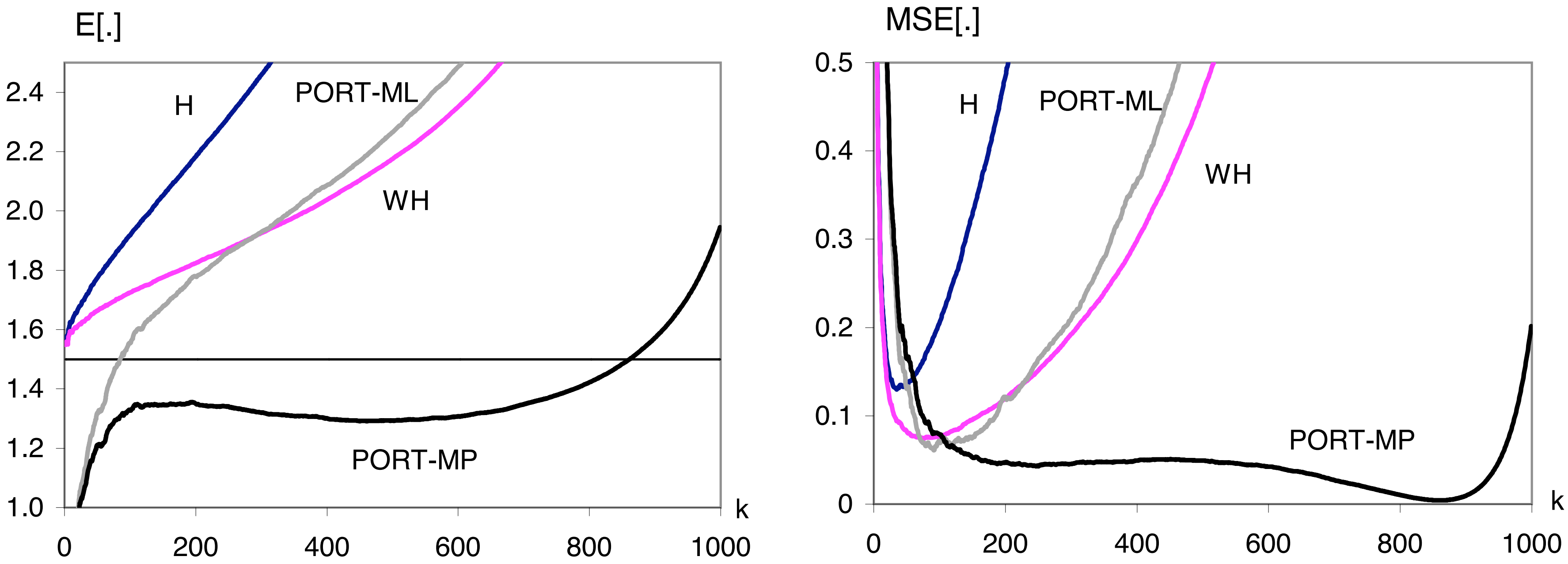
In the region $\gamma + \rho = 0$, the PORT-ML estimator is a second-order reduced-bias tail index estimator and consequently is expected to outperform the PORT-MP estimator at optimal levels.

These results claim for a semi-parametric test of the hypothesis $H_0 : \eta = \gamma + \rho = 0$. The non-rejection of such an hypothesis would lead us to the consideration of the PORT-ML estimator, things working in favor of the PORT-MP estimator, in case of rejection of H_0. This is however out of the scope of this seminar.

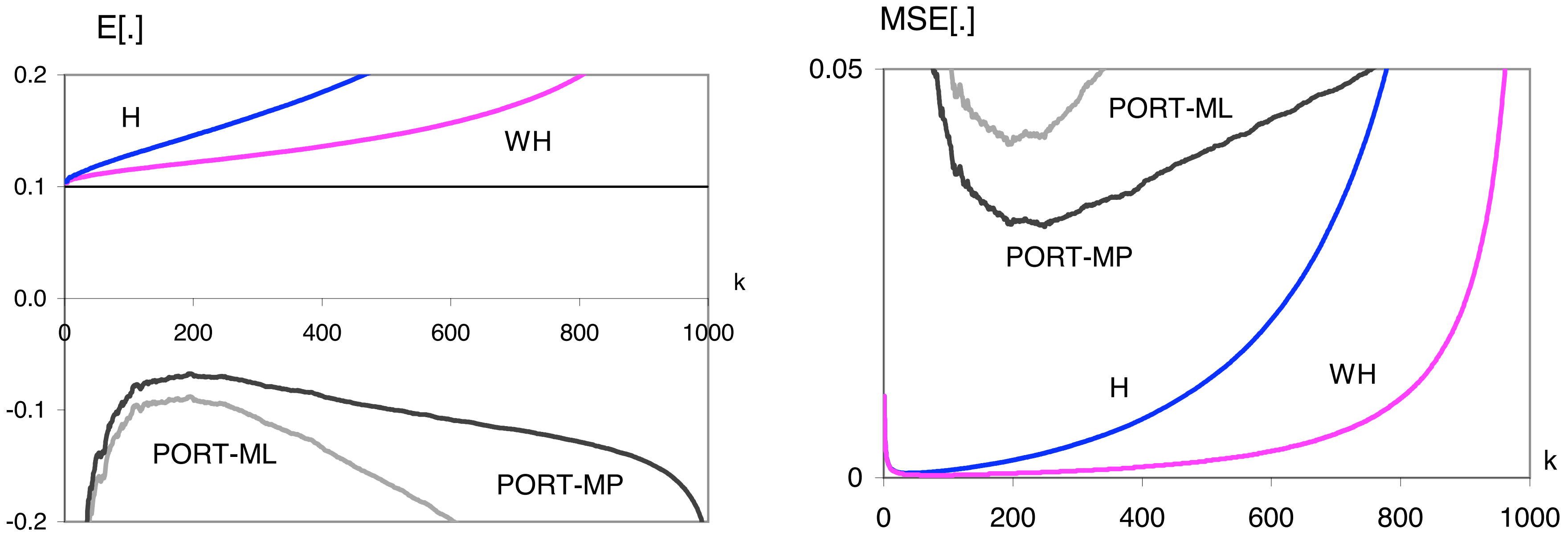
Simulated behavior of the estimators. In order to obtain the PORT-MP estimates, we have implemented a modified version of Grimshaw's method [Grimshaw, 1993].

Due to the high computation time of the general comparison algorithm, we have based our simulations on a multi-sample simulation of size 10×100, for samples with size n up to $n = 1000$, and we have chosen the value 100 for the maximum number of iterations in the modified Newton-Raphson algorithm.

In Figures 2 and 3 we show, on the basis of the first replicate, the simulated patterns of mean values, E[.], and mean squared errors, MSE[.], of the estimators under study for an underlying *Burr* parent, $F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}, x \geq 0$ with $(\gamma, \rho) = (1.5, -0.5)$ and $(0.1, -0.5)$, respectively. In all figures, PORT-ML and PORT-MP denote the estimators $\widehat{\gamma}_n^{ML}(k)$ and $\widehat{\gamma}_n^{MP}(k)$, respectively.



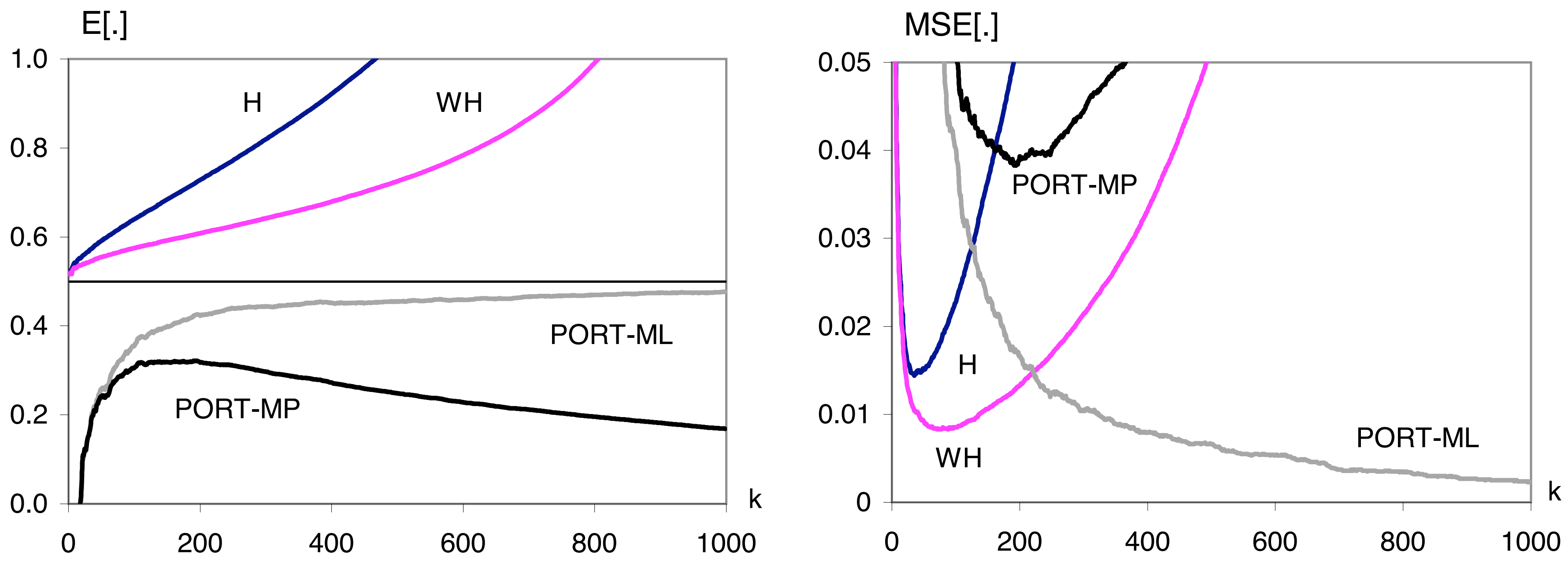
Mean values and mean squared errors of the estimators under study for a sample of size $n = 1000$, from a *Burr* parent with $\rho = -0.5$ and $\gamma = 1.5$.



Mean values and mean squared errors of the estimators under study for a sample of size $n = 1000$, from a *Burr* parent with $\rho = -0.5$ and $\gamma = 0.1$.

The simulations show that the tail index estimator PORT-MP has, in general, very stable sample paths and works quite well for values of $\gamma \geq 1.0$ and values of $|\rho| < 1.0$. For this (γ, ρ)-region the bias is always smaller than the corresponding one of the PORT-ML estimator, for all k. The mean square error of the PORT-MP estimator is, in general, smaller than the mean square error of the PORT-ML estimator for a large region of values of k, as well as at optimal levels. When $\gamma < 0.5$ and $|\rho| < 1$, the PORT-MP estimator does not work as expected, but it has a smaller bias and a smaller mean squared error than the PORT-ML estimator, for all k. However, both the PORT-ML and the PORT-MP are a long way from the Hill, and the best performance is achieved by the WH-estimator.

In Figure 4 we present the mean value and mean squared of the estimators for a *Burr* model with $(\gamma, \rho) = (0.5, -0.5)$.



Mean values and mean squared errors of the estimators under study for a sample of size $n = 1000$, from a *Burr* parent with $\rho = -0.5$ and $\gamma = 0.5$.

When $\gamma + \rho = 0$ the PORT-ML estimator is second-order asymptotically unbiased for the estimation of γ, and we were indeed expecting such an out-performance of the PORT-ML comparatively to the PORT-MP estimator. Indeed, for this model, the PORT-ML estimator has a squared bias and a mean squared error smaller than those of the PORT-MP, for all values of k. Also, the PORT-ML, looking almost like a true "unbiased" estimator for large k and for this particular model, outperforms the WH-estimator for large values of k.

Overall comparison at optimal levels of a few comparable tail index estimators. Apart from the Hill we shall also consider the moment estimator [Dekkers, Einmhal and de Haan, 1989],

$$\boxed{\hat{\gamma}_n^M(k) := M_n^{(1)}(k) + \tfrac{1}{2}\left\{1 - \left(\frac{M_n^{(2)}(k)}{[M_n^{(1)}(k)]^2} - 1\right)^{-1}\right\},}$$

and the mixed moment estimator [Fraga Alves, Gomes, de Haan and Neves, 2006], asymptotically equivalent to the ML-estimator if $\gamma + \rho \neq 0$, and with the simple functional form:

$$\boxed{\hat{\gamma}_n^{MM}(k) := \frac{\hat{\varphi}_n(k) - 1}{1 + 2\min\left(\hat{\varphi}_n(k) - 1, 0\right)}, \quad \hat{\varphi}_n(k) := \frac{M_n^{(1)}(k) - L_n^{(1)}(k)}{\left(L_n^{(1)}(k)\right)^2},}$$

where

$$\boxed{L_n^{(1)}(k) := 1 - \frac{1}{k}\sum_{i=1}^{k} \frac{X_{n-k:n}}{X_{n-i+1:n}}.}$$

The Moment can outperform the Hill . . .

γ ρ	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
0.00	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
-0.10	H	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.20	H	H	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.30	H	H	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.40	H	H	H	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.50	H	H	H	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.60	H	H	H	H	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.70	H	H	H	H	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.80	H	H	H	H	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.90	H	H	H	H	H	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-1.00	H	H	H	H	H	Mo	Mo	Mo	Mo	H	H	H	H	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-1.10	H	H	H	H	H	Mo	Mo	Mo	Mo	H	H	H	H	H	H	Mo	Mo	Mo	Mo	Mo	Mo
-1.20	H	H	H	H	H	Mo	Mo	Mo	H	H	H	H	H	H	H	H	Mo	Mo	Mo	Mo	Mo
-1.30	H	H	H	H	H	H	Mo	Mo	H	H	H	H	H	H	H	H	H	H	Mo	Mo	Mo
-1.40	H	H	H	H	H	H	Mo	Mo	H	H	H	H	H	H	H	H	H	H	H	Mo	Mo
-1.50	H	H	H	H	H	H	Mo	Mo	H	H	H	H	H	H	H	H	H	H	H	H	Mo
-1.60	H	H	H	H	H	H	Mo	Mo	H	H	H	H	H	H	H	H	H	H	H	H	H
-1.70	H	H	H	H	H	H	Mo	Mo	H	H	H	H	H	H	H	H	H	H	H	H	H
-1.80	H	H	H	H	H	H	Mo	Mo	H	H	H	H	H	H	H	H	H	H	H	H	H
-1.90	H	H	H	H	H	H	H	Mo	H	H	H	H	H	H	H	H	H	H	H	H	H
-2.00	H	H	H	H	H	H	H	Mo	H	H	H	H	H	H	H	H	H	H	H	H	H

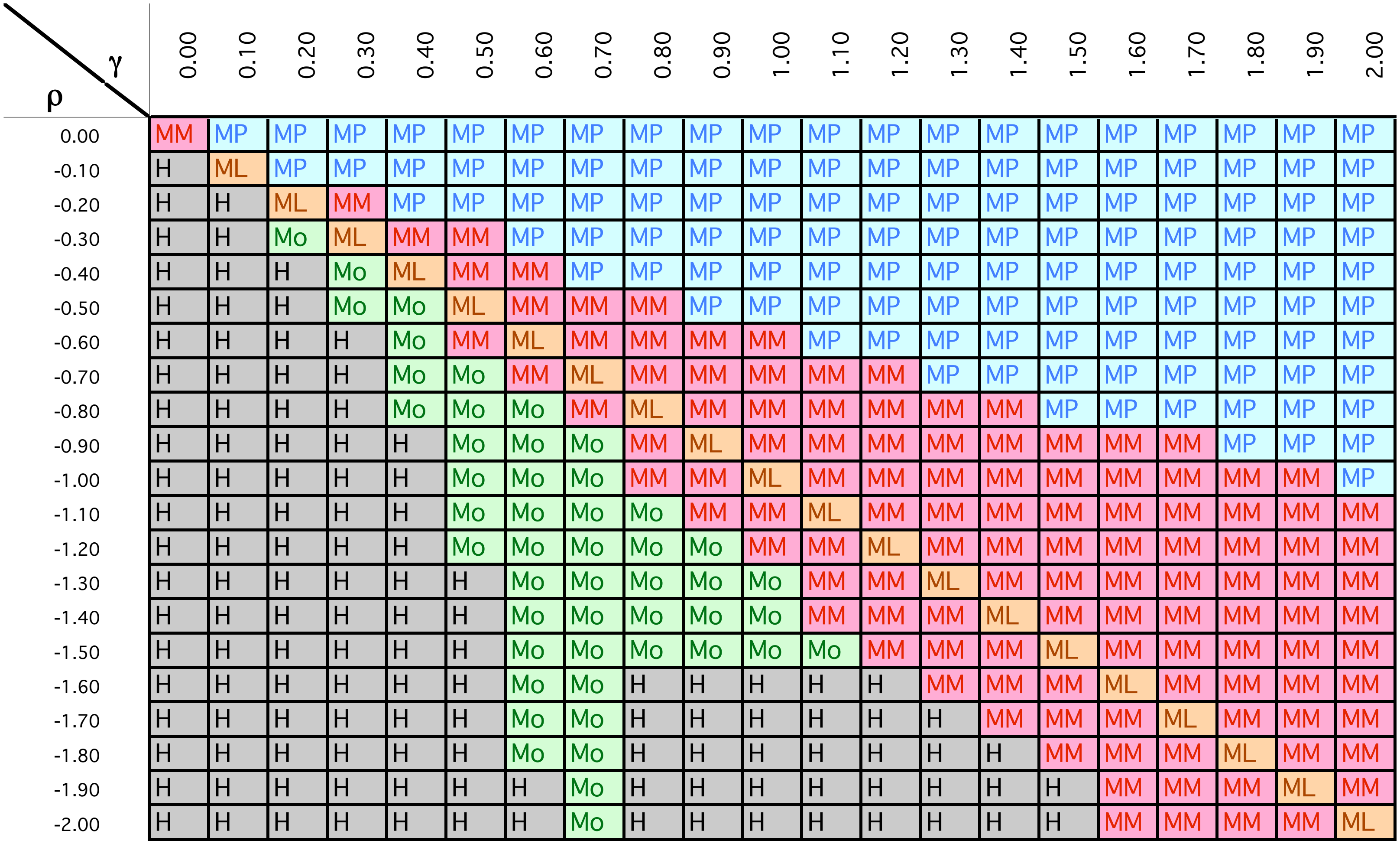
The MM($\equiv$ML, unless $\gamma+\rho \neq 0$, $(\gamma,\rho) \neq (0,0)$) can outperform the Moment and the Hill . . .

ρ \ γ	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
0.00	MM	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
-0.10	H	ML	MM	MM	MM	MM	MM	MM	MM	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.20	H	H	ML	MM	MM	MM	MM	MM	MM	MM	MM	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.30	H	H	Mo	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.40	H	H	H	Mo	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	Mo	Mo	Mo	Mo	Mo	Mo
-0.50	H	H	H	Mo	Mo	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	Mo	Mo	Mo	Mo
-0.60	H	H	H	H	Mo	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	Mo	Mo
-0.70	H	H	H	H	Mo	Mo	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
-0.80	H	H	H	H	Mo	Mo	Mo	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
-0.90	H	H	H	H	H	Mo	Mo	Mo	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
-1.00	H	H	H	H	H	Mo	Mo	Mo	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
-1.10	H	H	H	H	H	Mo	Mo	Mo	Mo	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM
-1.20	H	H	H	H	H	Mo	Mo	Mo	Mo	Mo	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM
-1.30	H	H	H	H	H	H	Mo	Mo	Mo	Mo	Mo	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM
-1.40	H	H	H	H	H	H	Mo	Mo	Mo	Mo	Mo	MM	MM	MM	ML	MM	MM	MM	MM	MM	MM
-1.50	H	H	H	H	H	H	Mo	Mo	Mo	Mo	Mo	Mo	MM	MM	MM	ML	MM	MM	MM	MM	MM
-1.60	H	H	H	H	H	H	Mo	Mo	H	H	H	H	H	MM	MM	MM	ML	MM	MM	MM	MM
-1.70	H	H	H	H	H	H	Mo	Mo	H	H	H	H	H	H	MM	MM	MM	ML	MM	MM	MM
-1.80	H	H	H	H	H	H	Mo	Mo	H	H	H	H	H	H	H	MM	MM	MM	ML	MM	MM
-1.90	H	H	H	H	H	H	H	Mo	H	H	H	H	H	H	H	H	MM	MM	MM	ML	MM
-2.00	H	H	H	H	H	H	H	Mo	H	H	H	H	H	H	H	H	MM	MM	MM	MM	ML

The MP can outperform the MM . . .

γ ρ	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
0.00	MM	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.10	MM	ML	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.20	MM	MM	ML	MM	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.30	MM	MP	MM	ML	MM	MM	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.40	MM	MP	MM	MM	ML	MM	MM	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.50	MM	MP	MP	MM	MM	ML	MM	MM	MM	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.60	MM	MP	MP	MM	MM	MM	ML	MM	MM	MM	MM	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.70	MM	MP	MP	MM	MM	MM	MM	ML	MM	MM	MM	MM	MM	MP	MP	MP	MP	MP	MP	MP	MP
-0.80	MM	MP	MP	MM	MM	MM	MM	MM	ML	MM	MM	MM	MM	MM	MM	MP	MP	MP	MP	MP	MP
-0.90	MM	MP	MP	MP	MM	MM	MM	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MP	MP	MP
-1.00	MM	MP	MP	MP	MM	MM	MM	MM	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MP
-1.10	MM	MP	MP	MP	MM	MM	MM	MM	MM	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM
-1.20	MM	MP	MP	MP	MM	MM	MM	MM	MM	MM	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM
-1.30	MM	MP	MP	MP	MP	MM	MM	MM	MM	MM	MM	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM
-1.40	MM	MP	MP	MP	MP	MM	MM	MM	MM	MM	MM	MM	MM	MM	ML	MM	MM	MM	MM	MM	MM
-1.50	MM	MP	MP	MP	MP	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	ML	MM	MM	MM	MM	MM
-1.60	MM	MP	MP	MP	MP	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	ML	MM	MM	MM	MM
-1.70	MM	MP	MP	MP	MP	MP	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	ML	MM	MM	MM
-1.80	MM	MP	MP	MP	MP	MP	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	ML	MM	MM
-1.90	MM	MP	MP	MP	MP	MP	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	ML	MM
-2.00	MM	MP	MP	MP	MP	MP	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	ML

Overall . . .



γ ρ	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
0.00	MM	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.10	H	ML	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.20	H	H	ML	MM	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.30	H	H	Mo	ML	MM	MM	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.40	H	H	H	Mo	ML	MM	MM	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.50	H	H	H	Mo	Mo	ML	MM	MM	MM	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.60	H	H	H	H	Mo	MM	ML	MM	MM	MM	MM	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
-0.70	H	H	H	H	Mo	Mo	MM	ML	MM	MM	MM	MM	MM	MP	MP	MP	MP	MP	MP	MP	MP
-0.80	H	H	H	H	Mo	Mo	Mo	MM	ML	MM	MM	MM	MM	MM	MM	MP	MP	MP	MP	MP	MP
-0.90	H	H	H	H	H	Mo	Mo	Mo	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MP	MP	MP
-1.00	H	H	H	H	H	Mo	Mo	Mo	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MP
-1.10	H	H	H	H	H	Mo	Mo	Mo	Mo	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM
-1.20	H	H	H	H	H	Mo	Mo	Mo	Mo	Mo	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM
-1.30	H	H	H	H	H	H	Mo	Mo	Mo	Mo	Mo	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM
-1.40	H	H	H	H	H	H	Mo	Mo	Mo	Mo	Mo	MM	MM	MM	ML	MM	MM	MM	MM	MM	MM
-1.50	H	H	H	H	H	H	Mo	Mo	Mo	Mo	Mo	Mo	MM	MM	MM	ML	MM	MM	MM	MM	MM
-1.60	H	H	H	H	H	H	Mo	Mo	H	H	H	H	H	MM	MM	MM	ML	MM	MM	MM	MM
-1.70	H	H	H	H	H	H	Mo	Mo	H	H	H	H	H	H	MM	MM	MM	ML	MM	MM	MM
-1.80	H	H	H	H	H	H	Mo	Mo	H	H	H	H	H	H	H	MM	MM	MM	ML	MM	MM
-1.90	H	H	H	H	H	H	H	Mo	H	H	H	H	H	H	H	H	MM	MM	MM	ML	MM
-2.00	H	H	H	H	H	H	H	Mo	H	H	H	H	H	H	H	H	MM	MM	MM	MM	ML

AND THAT’s ALL . . .

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. (1984). Modeling excesses over high threshold with an application. In J. Tiago de Oliveira ed., *Statistical Extremes and Applications*, D. Reidel, 461-482.
4. Dekkers, A. and de Haan, L. (1993). Optimal sample fraction in extreme value estimation. *J. Multivariate Analysis* **47**(2), 173-195.
5. Dekkers, A., Einmahl and de Haan, L. (1989). A momment estimator for the index of an extreme-value distribution. *Annals of Statistics* **17**, 1833-1855.
6. Drees, H., Ferreira, A., and Haan, L. de (2004). On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.* **14**, 1179-1201.
7. 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), 193-214.
8. Fraga Alves, M.I., Gomes, M.I., de Haan, L. and Neves (2007). *Mixed Moment Estimator and its Invariant Alternatives*. Notas e Comunicações CEAUL 14/07. Submitted.

9. Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423-453.
10. Geluk, J. and L. de Haan (1987). *Regular Variation, Extensions and Tauberian Theorems*. CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam, Netherlands.
11. Gomes, M. I., de Haan, L. and Rodrigues, L. (2008). Tail index estimation through accommodation of bias in the weighted log-excesses. *J. Royal Statistical Society* **B70**, Issue 1, 31-52.
12. 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.
13. Gomes, M. I. and M. J. Martins (2002). "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes* **5**:1, 5-31.
14. 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.
15. Gomes, M.I., Miranda, C., and Viseu, C. (2007). Reduced bias extreme value index estimation and the Jackknife methodology, *Statistica Neerlandica* **61**:2, 243-270.

16. Gomes, M.I., and Neves, C. (2007). Asymptotic comparison of the mixed moment and classical extreme value index estimators. *Statistics and Probability Letters*, in press.
17. Gomes, M. I. and Pestana, D. (2007). A sturdy reduced-bias extreme quantile (VaR) estimator. *J. American Statistical Association*, Vol. **102**, No. 477, 280-292.
18. Haan, L. de (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*. Mathematical Centre Tract 32, Amsterdam.
19. Haan, L. de, and Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica* **52**, 60-70.
20. Hall, P. and A.H. Welsh (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13**, 331-341.
21. Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist*. **3**, 1163-1174.
22. Smith, R.L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15**:3, 1174-1207.