# Tail index estimation for heavy tails: accommodation of bias in the excesses over a high threshold*

M. Ivette Gomes

University of Lisbon, Faculty of Science, DEIO and CEAUL, Portugal

and

Lígia Henriques Rodrigues

Tomar Polytechnic, College of Technology, CEAUL, Portugal

February 12, 2008

**Abstract.** In statistics of extremes, inference is often based on the excesses over a high random threshold. Those excesses are approximately distributed as the set of order statistics associated to a sample from a generalized Pareto model. We then get the so-called "maximum likelihood" (ML) estimators of the tail index $\gamma$. In this paper, we are interested in the derivation of the asymptotic distributional properties of a similar "maximum likelihood" estimator of a positive tail index $\gamma$, based also on the excesses over a high random threshold, but with a trial of accommodation of bias in the Pareto model underlying those excesses. We next proceed to an asymptotic comparison of the two estimators at their optimal levels. An illustration of the finite sample behaviour of the estimators is provided through a small-scale Monte Carlo simulation study.

**AMS 2000 subject classification**. Primary 62G32; Secondary 65C05.

**Keywords and phrases**. *Statistics of extremes*, *semi-parametric estimation*, *generalized Pareto models*, *maximum likelihood estimation.*

---

*Research partially supported by FCT / POCTI and POCI / FEDER.

# 1 Introduction

Heavy-tailed models appear often in practice in fields like telecommunication traffic and finance. A model $F$ is said to be heavy-tailed whenever the *tail function*, $\overline{F} := 1 - F$, is a regularly varying function with a negative index of regular variation $-1/\gamma$, i.e., for every $x > 0$, $\lim_{t\to\infty} \overline{F}(tx)/\overline{F}(t) = x^{-1/\gamma}$. Then we are in the domain of attraction for maxima of an *extreme value* (EV) distribution function (d.f.),

$$EV_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}), \qquad x > -1/\gamma, \quad \gamma > 0,$$

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

We are going to base inference on the $k$ top-order statistics (o.s.), and as usual in semi-parametric estimation of parameters of extreme events, we shall assume that $k$ is an intermediate sequence of integers, i.e., $k = k_n$ is such that

$$k = k_n \to \infty, \quad k_n = o(n), \text{ as } n \to \infty. \tag{1.1}$$

In this paper, inference on $\gamma$ is going to be based on the excesses over the random threshold $X_{n-k:n}$, represented by

$$W_{ik} := X_{n-i+1:n} - X_{n-k:n}, \quad 1 \le i \le k < n, \tag{1.2}$$

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 the $k$ top o.s. in a sample of size $k$ from a *generalized Pareto* (GP) model, with d.f.

$$GP(x;\gamma,\alpha) = 1 - (1+\alpha\, x)^{-1/\gamma}, \quad x > 0 \quad (\alpha,\ \gamma > 0), \tag{1.3}$$

a re-parametrization due to Davison (Davison, 1984). The *maximum likelihood* (ML) estimator of $\gamma$ has, with such a re-parametrization, an explicit expression as a function of the ML-estimator $\widehat{\alpha} = \widehat{\alpha}_{ML}$ of $\alpha$ and the sample of the excesses. We have

$$\widehat{\gamma}_n^{ML}(k) = \widehat{\gamma}_{n,\widehat{\alpha}}^{ML}(k) := \frac{1}{k}\sum_{i=1}^{k} \ln\left(1 + \widehat{\alpha}\, W_{ik}\right), \tag{1.4}$$

the so-called PORT-ML tail index estimator, with PORT standing for *peaks over random threshold*, a terminology introduced in Araújo Santos *et al.* (2006). The parameter $\alpha$ is such that $\alpha W_{ik} \approx Y^{\gamma}_{k-i+1:k} - 1$, with $Y$ a unit Pareto random variable (r.v.), with d.f. $F_Y(y) = 1-1/y,\ y \ge 1$. As we shall see later on, in Section 2, an obvious choice for an estimator of $\alpha$ is $1/X_{n-k:n}$. If we consider $\hat\alpha = 1/X_{n-k:n}$ in (1.4), $1 + \hat\alpha\ W_{ik} = X_{n-i+1:n}/X_{n-k:n}$, and the estimator in (1.4) gives rise to the Hill estimator (Hill, 1975)

$$\widehat{\gamma}^{H}_n(k) := \frac{1}{k}\sum_{i=1}^{k}\{\ln X_{n-i+1:n} - \ln X_{n-k:n}\} =: \frac{1}{k}\sum_{i=1}^{k} V_{ik}, \tag{1.5}$$

the average of the log-excesses $V_{ik},\ 1 \le i \le k$. Gomes *et al.* (2008) suggested the use of an adequate weighting of the log-excesses $V_{ik}$ instead of the Hill estimator in (1.5). Analogously, 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 $i$,

$$\alpha W_{ik} - \big(Y^{\gamma/p_{ik}}_{k-i+1:k} - 1\big) = o_p\big(\alpha W_{ik} - \big(Y^{\gamma}_{k-i+1:k} - 1\big)\big). \tag{1.6}$$

The validity of (1.6) leads us to expect to possibly be able to get a "better" estimator of $\gamma$ if we apply the approximation $\alpha W_{ik} \approx Y^{\gamma/p_{ik}}_{k-i+1:k} - 1$, instead of the approximation $\alpha W_{ik} \approx Y^{\gamma}_{k-i+1:k} - 1,\ 1 \le i \le k$, used to support the PORT-ML estimator in (1.4). Then, 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 $\ln(1 + \hat\alpha W_{ik}),\ 1 \le i \le k$, i.e., by

$$\widehat{\gamma}^{MP}_n(k) \equiv \widehat{\gamma}^{MP}_{n,\widehat\alpha,\widehat\beta,\widehat\rho}(k) := \frac{1}{k}\sum_{i=1}^{k} p_{ik}(\hat\beta, \hat\rho)\ln(1 + \widehat\alpha W_{ik}), \tag{1.7}$$

here called the PORT-MP tail index estimator, with $MP$ standing for *modified Pareto*, and with $\widehat\alpha = \widehat\alpha_{MP}$, the PORT-MP estimator of $\alpha$. The estimators $(\hat\beta, \hat\rho)$ need to be adequate consistent estimators of the second order parameters $(\beta, \rho)$.

In Section 2, we shall present a few introductory technical details in the field of *statistics of extremes* and introduce the new class of tail index estimators, the PORT-MP estimators, providing further motivation for their consideration, under the assumption that all the model parameters, but the tail index $\gamma$, are known. 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 derived in Section 3. In Section 4, we shall exhibit the performance of the new PORT-MP estimator, comparatively to the classical PORT-ML estimator, through the use of simulation techniques. Finally, in Section 5, we shall provide an appendix, with the asymptotic behaviour of the r.v.'s under play in the main theorems, the proofs of the main results in the paper and a modified version of Grimshaw's algorithm (Grimshaw, 1993).

## 2 Further introductory results

### 2.1 First and second order framework

In a context of heavy tails, and with the notation $U(t) = F^{\leftarrow}(1 - 1/t),\ t \geq 1,\ F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$ the generalized inverse function of the underlying model $F$, the first order parameter (or tail index) $\gamma$ $(> 0)$ appears, for every $x > 0$, as the limiting value, as $t \to \infty$, of the quotient $(\ln U(tx) - \ln U(t))/\ln x$ (de Haan, 1970). Indeed, with the usual notation $RV_\alpha$ for the class of *regularly varying* functions with index of regularly variation $\alpha$, we can further say

$$F \in \mathcal{D}(EV_{\gamma>0}) \quad \text{iff} \quad U \in RV_\gamma \quad \text{iff} \quad 1 - F \in RV_{-1/\gamma} \text{ (Gnedenko, 1943).} \tag{2.1}$$

The second order parameter $\rho$ $(\leq 0)$ rules the rate of convergence in the first order condition (2.1) and is the non-positive value which appears in the limiting relation

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho} \quad \text{iff} \quad \lim_{t\to\infty} \frac{U(tx)/U(t) - x^\gamma}{A(t)} = x^\gamma\ \frac{x^\rho - 1}{\rho}, \tag{2.2}$$

which we assume to hold for every $x > 0$, and where $|A|$ is of regular variation with index $\rho$ (Geluk and de Haan, 1987).

Throughout the paper, and unless otherwise stated, we shall assume the validity of a condition slightly less general than condition (2.2). More specifically, we shall assume that we are working in Hall-Welsh class of models (Hall and Welsh, 1985), with a tail function

$$\overline{F}(x) = 1 - F(x) = \Big(\frac{x}{C}\Big)^{-1/\gamma}\Big(1 + \frac{\beta}{\rho}\Big(\frac{x}{C}\Big)^{\rho/\gamma} + o(x^{\rho/\gamma})\Big), \quad \text{as } x \to \infty, \tag{2.3}$$

with $C > 0$, $\beta \neq 0$ and $\rho < 0$. Equivalently, we can say that (2.2) holds with

$$A(t) = \gamma\, \beta\, t^{\rho}, \;\; \rho < 0, \tag{2.4}$$

being $(\beta, \rho)$ a vector of second order parameters. We are thus excluding models like for instance the ones with a Paretian logarithmic tail, of the type $\overline{F}(x) \sim \alpha x^{-1/\gamma}(\ln x)^{\delta}$, as $x \to \infty$, with $\gamma,\ \alpha > 0$ and $\delta \neq 0$, for which (2.2) holds, with $\rho = 0$. For the standard Pareto, with d.f. $F_Z(z) = 1 - z^{-1/\gamma}$, $z \geq 1$, we get $U(tx)/U(t) = x^{\gamma}$ for all $x$, and we might have considered by convention that (2.2) held with $A(t) \equiv 0$, i.e., $\rho = -\infty$ or $\beta = 0$ in (2.4). However, due to the bad finite sample behaviour of the $\rho$-estimates for this model, where $\rho$ can indeed take any value, provided that $\beta = 0$, we have also excluded it from our study. Fortunately, in this ideal case, the estimation of the tail index is almost trivial: the Hill estimator in (1.5) is unbiased for all $k$. Note however that most heavy-tailed models useful in applications, like the Fréchet, the generalized Pareto, the Burr and the Student's $t$ satisfy (2.3).

## 2.2 Excesses over a high threshold and the GP model

From the definition of the function $U$ and on the basis of the universal uniform transformation, we get the representation $X_{i:n} = U(Y_{i:n})$, again with $Y$ a unit Pareto r.v. On the other hand, as for $j > i$, $Y_{j:n}/Y_{i:n} \stackrel{d}{=} Y_{j-i:n-i}$, $\ln Y_{i:n} \stackrel{d}{=} E_{i:n}$, where $E$ denotes a standard exponential r.v., and $Y_{n-k:n} \stackrel{p}{\sim} n/k$ whenever (1.1) holds, we can indeed write, whenever we are under the first order framework in (2.1),

$$W_{ik} \stackrel{d}{=} X_{n-k:n} \left\{ \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})} - 1 \right\} \stackrel{d}{=} U(n/k) \left\{ Y^{\gamma}_{k-i+1:k} - 1 \right\} (1 + o_p(1)). \tag{2.5}$$

Then, with $\alpha = 1/U(n/k)$, we have for intermediate $k$,

$$W_{ik} = X_{n-i+1:n} - X_{n-k:n} \approx \frac{Y^{\gamma}_{k-i+1:k} - 1}{\alpha}, \quad 1 \leq i \leq k, \tag{2.6}$$

i.e., as mentioned before, the $k$ excesses $W_{ik}$, $1 \leq i \leq k$, in (1.2), are approximately the $k$ o.s. from the *GP* model in (1.3), and the maximum likelihood methodology leads us to the tail index estimator in (1.4), with $\widehat{\alpha} = \widehat{\alpha}_{ML}$. Smith (1987) has got the asymptotic behaviour of the estimator in (1.4) for a fixed threshold, replacing the r.v. $X_{n-k:n}$ by a deterministic value $u$. A more general result, related with a random threshold $X_{n-k:n}$, is available in Drees *et al.* (2004).

## 2.3 Accommodating bias in the Paretian excesses

Under the second order framework in (2.2), we can say that there exist $\gamma$ and $\alpha$ such that, for $1 \le i \le k$,

$$\alpha\, W_{ik} \stackrel{d}{=} Y_{k-i+1:k}^{\gamma} - 1 + A(n/k) Y_{k-i+1:k}^{\gamma} \frac{Y_{k-i+1:k}^{\rho} - 1}{\rho}(1 + o_p(1)). \tag{2.7}$$

The use of Taylor's formula for $e^x$, as $x \to 0$, and $\ln x$, as $x \to 1$, enables us to rewrite the equation (2.7) as

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

Notice that, from (2.7),

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

Since we can approximately write

$$\frac{Y_{k-i+1:k}^{\rho} - 1}{\rho \ln Y_{k-i+1:k}} \approx -\frac{(i/k)^{-\rho} - 1}{\rho \ln(i/k)} =: \psi_{ik} \equiv \psi(i/k) \equiv \psi_{ik}(\rho)\ [\psi_{kk} \equiv 1], \tag{2.8}$$

with $\psi_{ik}$ 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 $\gamma$, as it is usually done, but dependent on $i$ (and $k$) and given by

$$\gamma_{ik} := \gamma\, e^{\beta\, \left(\frac{n}{k}\right)^{\rho} \psi_{ik}}, \quad 1 \le i \le k, \tag{2.9}$$

for models in (2.3).

We are thus going to base inference on the fact that there exists a parameter $\alpha$ such that $W_{ik} = X_{n-i+1:n} - X_{n-k:n}$ comes from a $GP$ model, with d.f. $GP(x; \gamma_{ik}, \alpha)$ in (1.3), $\gamma_{ik}$ defined in (2.9), for every $1 \le i \le k$. The likelihood function of $\underline{W} = \big(W_{ik},\ 1 \le i \le k\big)$ is then proportional to

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

and consequently we have

$$\ln L(\gamma,\alpha,\beta,\rho;\ \underline{W}) = k\ln\alpha - k\ln\gamma - \beta(n/k)^{\rho}\sum_{i=1}^{k}\psi_{ik} - \sum_{i=1}^{k}\ln(1+\alpha\ W_{ik})$$
$$-\frac{1}{\gamma}\sum_{i=1}^{k} e^{-\beta(n/k)^{\rho}\psi_{ik}}\ln(1+\alpha\ W_{ik}). \quad (2.10)$$

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

$$\widehat{\gamma}_n^{MP}(k) \equiv \widehat{\gamma}_{n,\widehat{\alpha},\widehat{\beta},\widehat{\rho}}^{MP}(k) := \frac{1}{k}\sum_{i=1}^{k} e^{-\widehat{\beta}\ (n/k)^{\widehat{\rho}}\ \widehat{\psi}_{ik}}\ \ln(1+\widehat{\alpha}W_{ik}), \quad \widehat{\psi}_{ik} = -\frac{(i/k)^{-\widehat{\rho}}-1}{\widehat{\rho}\ln(i/k)}, \quad (2.11)$$

and with $\widehat{\alpha} = \widehat{\alpha}_{MP}$, the PORT-MP estimator of $\alpha$. Consequently, the $p_{ik}(\beta,\rho)$ in (1.7) are given by $\exp\{-\beta(n/k)^{\rho}\psi_{ik}\}$, $1 \le i \le k$.

If we look at (2.5) and (2.6), we see that, as mentioned before in Section 1, $\hat{\alpha}$ can be chosen equal to $1/X_{n-k:n}$. If we replace in (2.11) $\hat{\alpha}$ by $1/X_{n-k:n}$, we get the *weighted log-excesses* or *weighted-Hill* (WH) estimator,

$$\widehat{\gamma}_n^{WH}(k) \equiv \widehat{\gamma}_{n,\widehat{\beta},\widehat{\rho}}^{WH}(k) := \frac{1}{k}\sum_{i=1}^{k} e^{-\widehat{\beta}\ (n/k)^{\widehat{\rho}}\ \widehat{\psi}_{ik}}\ \ln\Big(\frac{X_{n-i+1:n}}{X_{n-k:n}}\Big), \quad (2.12)$$

introduced and studied in Gomes *et al.* (2008). This is a second-order reduced-bias estimator with an asymptotic variance equal to $\gamma^2$, the asymptotic variance of the Hill estimator in (1.5), for adequate levels $k$ and an adequate estimation of the second order parameters. We shall consider only an external estimation of the second order parameters $\rho$ and $\beta$, involving a number $k_1$ of top o.s. larger than the number $k$ of top o.s. used for the tail index estimation. Such a decision is related to the discussion on the advantages of an external estimation of the second order parameters (or even their mispecification) versus an internal estimation at the same level $k$, initiated in Gomes and Martins (2002). In the following, we shall make explicit the estimators of the second order parameters to be used in this paper.

## 2.4 Estimation of the second order parameters

### 2.4.1 The estimation of $\rho$

We shall first address the estimation of $\rho$. We have nowadays two general classes of $\rho$-estimators, which work well in practice, introduced in Gomes *et al.* (2002) and Fraga Alves *et al.* (2003). The estimator of $\rho$ to be considered in this study is a particular member of the class of estimators proposed by Fraga Alves *et al.* (2003), where an heuristic choice of the threshold seems to provide interesting results. Under adequate general conditions, such a class provides semi-parametric asymptotically normal estimators of $\rho$, which show highly stable sample paths as functions of $k$, the number of top o.s. used, for a wide range of large $k$-values. Such a class of estimators is parameterized in a *tuning* parameter $\tau$, not necessarily non-negative but real, as detected in Caeiro and Gomes (2006), and depends on the statistics

$$T_n^{(\tau)}(k) := \begin{cases} \dfrac{\ln\left(M_n^{(1)}(k)\right)-\frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right)}{\frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right)-\frac{1}{3}\ln\left(M_n^{(3)}(k)/6\right)}, & \text{if} \quad \tau=0 \\ \dfrac{\left(M_n^{(1)}(k)\right)^{\tau}-\left(M_n^{(2)}(k)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k)/2\right)^{\tau/2}-\left(M_n^{(3)}(k)/6\right)^{\tau/3}}, & \text{if} \quad \tau\neq 0 \end{cases},$$

where, with $V_{ik}$ given in (1.5),

$$M_n^{(j)}(k) = \frac{1}{k}\sum_{i=1}^{k} V_{ik}^j, \ j \geq 1 \qquad [M_n^{(1)} = \widehat{\gamma}_n^H].$$

The statistic $T_n^{(\tau)}(k)$ converges towards $3(1-\rho)/(3-\rho)$, independently of $\tau$, whenever the second order condition (2.2) holds and $k$ is such that $k=o(n)$ and $\sqrt{k}\,A(n/k) \to \infty$, as $n \to \infty$. We can thus get a class of consistent estimators of $\rho$, given by

$$\widehat{\rho}(k;\tau) := -\left|3(T_n^{(\tau)}(k)-1)/(T_n^{(\tau)}(k)-3)\right|. \tag{2.13}$$

### 2.4.2 Estimation of $\beta$ based on the scaled log-spacings

We shall consider the estimator of $\beta$ obtained in Gomes and Martins (2002) and based on the scaled log-spacings $U_i = i\,\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}$, $1 \leq i \leq k$. Such an estimator is given by

$$\widehat{\beta}(k;\widehat{\rho}) := \frac{\left(\frac{k}{n}\right)^{\widehat{\rho}}\left[\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^{k}U_i\right)-\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}U_i\right)\right]}{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}U_i\right)-\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\widehat{\rho}}U_i\right)}. \tag{2.14}$$

**Remark 2.1.** Caeiro *et al.* (2005), Gomes and Pestana (2007a,b), Gomes *et al.* (2007b) *and* Gomes *et al.* (2008) *presented different heuristic algorithms for the estimation of the second order parameters. Asymptotic considerations as well as simulated results led these authors to consider a high level* $k_1$*, given by* $k_1 := [n^{1-\epsilon}]$ *where* $\epsilon \sim 0^+$ *and* $[x]$ *denotes, as usual, the integer part of* $x$*. This level has not been chosen in any optimal way, but works well in practice for the* $\rho$*-estimation, provided we pay first attention to the adequate choice of* $\tau$ *in* (2.13)*. Moreover, with a very slight restriction in the class of models where the asymptotic results hold, such a* $k_1$ *provides the crucial property for the* $\rho$*-estimator needed in the above mentioned papers, the condition* $\hat{\rho} - \rho = o_p(1/\ln n)$*. Despite of the fact that the tuning parameter* $\tau = 1$ *can be used for all models, we improve the final results if we consider a tuning parameter* $\tau = 0$ *for models with* $|\rho| \le 1$*. Indeed, the theoretical and simulated results in* Fraga Alves *et al.* (2003) *and in* Gomes and Martins (2002) *lead us to advise in practice the consideration of the tuning parameter* $\tau = 0$ *for the region* $\rho \in [-1, 0)$ *and* $\tau = 1$ *for the region* $\rho \in (-\infty, -1)$*.*

We have here implemented simulation experiments based on the estimation of $\beta$ at the same level $k_1 = n^{0.995}$ we have used for the estimation of $\rho$. We use the notations $\widehat{\rho}_\tau = \widehat{\rho}(k_1; \tau)$ and $\widehat{\beta}_\tau = \widehat{\beta}(k_1; \widehat{\rho}_\tau)$ for any real $\tau$, and with $\widehat{\rho}(k; \tau)$ and $\widehat{\beta}(k; \widehat{\rho})$ given in (2.13) and (2.14), respectively.

## 2.5 Motivation for the new estimators — only $\gamma$ is unknown

Let us assume that everything is known, apart from $\gamma$. Then,

**Theorem 2.1.** *Under the second order framework in* (2.3)*, with* $A(t)$ *given in* (2.4) *and for levels* $k$ *such that* (1.1) *holds, we get for* $\widehat{\gamma}^{MP}_{n,\alpha,\beta,\rho}(k)$*, with* $\widehat{\gamma}^{MP}_{n,\widehat{\alpha},\widehat{\beta},\widehat{\rho}}(k)$ *provided in* (2.11)*, an asymptotic distributional representation of the type*

$$\widehat{\gamma}^{MP}_{n,\alpha,\beta,\rho}(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, N_k + o_p(A(n/k)), \tag{2.15}$$

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

The main problems to be dealt with are related with the study of how the estimation of $(\alpha,\beta,\rho)$ affects the asymptotic distributional behaviour of $\widehat{\gamma}^{MP}_{n,\alpha,\beta,\rho}(k)$, given in (2.15).

# 3 Asymptotic behaviour of the PORT-MP tail index estimator

Let us assume that we have access to the sample of the excesses, $\underline{W}=(W_{ik}, 1\le i\le k)$, and we are interested in the "maximum likelihood" estimator $\hat{\gamma}^{MP}_{n,\hat{\alpha},\hat{\beta},\hat{\rho}}(k)$, in (2.11), an explicit function of $\widehat{\alpha}=\widehat{\alpha}_{_{MP}}$, the PORT-MP estimator of $\alpha$ in a modified generalized Pareto model, and external estimators of the second order parameters $(\beta,\rho)$.

## 3.1 The r.v.'s under play and asymptotic normality of the PORT-MP tail index estimator

Let us introduce the following notations:

$$B := \tfrac{1}{k}\textstyle\sum_{i=1}^k\ \ln(1+\alpha\ W_{ik}),\quad B_{(j)} := \frac{1}{k}\sum_{i=1}^k\ \psi_{ik}^{j-1}e^{-\beta(n/k)^\rho\psi_{ik}}\ \ln(1+\alpha\ W_{ik}), \tag{3.1}$$

$$C := \tfrac{1}{k}\textstyle\sum_{i=1}^k\ \frac{\alpha\ W_{ik}}{1+\alpha W_{ik}},\qquad C_{(j)} := \frac{1}{k}\sum_{i=1}^k\ \psi_{ik}^{j-1}e^{-\beta(n/k)^\rho\psi_{ik}}\ \frac{\alpha\ W_{ik}}{1+\alpha W_{ik}}, \tag{3.2}$$

$$D := \tfrac{1}{k}\textstyle\sum_{i=1}^k\ \frac{\alpha\ W_{ik}}{(1+\alpha\ W_{ik})^2},\qquad D_{(j)} := \frac{1}{k}\sum_{i=1}^k\ \psi_{ik}^{j-1}e^{-\beta(n/k)^\rho\psi_{ik}}\ \frac{\alpha\ W_{ik}}{(1+\alpha\ W_{ik})^2}, \tag{3.3}$$

together with the obvious notations for $\widehat{B}$, $\widehat{C}$, $\widehat{D}$, $\widehat{B}_{(j)}$, $\widehat{C}_{(j)}$ and $\widehat{D}_{(j)}$, $j\ge 1$. The log-likelihood in (2.10) can thus be written as

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

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

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

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

Consequently,

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

As the PORT-MP estimator of $\gamma$, in (2.11), is $\widehat{\gamma}_n^{MP} = \widehat{B}_{(1)}$, the PORT-MP estimator of $\alpha$ is solution of the equation

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

Since $\partial\left(C + C_{(1)}/B_{(1)} - 1\right)/\partial\alpha = (D + D_{(1)}/B_{(1)} - (C_{(1)}/B_{(1)})^2)/\alpha$, $\widehat{\alpha}_{MP} \equiv \widehat{\alpha}_{MP}(k)$ is such that

$$\widehat{C} + \frac{\widehat{C}_{(1)}}{\widehat{B}_{(1)}} - 1 \equiv 0 = C + \frac{C_{(1)}}{B_{(1)}} - 1 + \frac{\widehat{\alpha}_{MP} - \alpha}{\alpha}\left(D + \frac{D_{(1)}}{B_{(1)}} - \left(\frac{C_{(1)}}{B_{(1)}}\right)^2\right)(1 + o_p(1)),$$

i.e.,

$$\frac{\widehat{\alpha}_{MP} - \alpha}{\alpha} = \frac{1 - C - C_{(1)}/B_{(1)}}{D + D_{(1)}/B_{(1)} - (C_{(1)}/B_{(1)})^2}\,(1 + o_p(1)). \tag{3.4}$$

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

$$\widehat{\gamma}^{MP}_{n,\hat{\alpha},\hat{\beta},\hat{\rho}}(k) = \widehat{B}_{(1)} = B_{(1)} + C_{(1)}\,\frac{\widehat{\alpha}_{MP} - \alpha}{\alpha}\,(1 + o_p(1)), \tag{3.5}$$

As can be seen from (3.4) and (3.5), the distributional behaviour of the PORT-MP estimators of $\alpha$ and $\gamma$ is thus strongly related with the distributional behaviour of the r.v.s in (3.1), (3.2) and (3.3), to be studied in Section 5.1.

As seen before in Section 2, if only $\gamma$ is unknown, Theorem 2.1 holds for $\widehat{\gamma}^{MP}_{n,\alpha,\beta,\rho}$. The same result holds for $\widehat{\gamma}^{MP}_{n,\alpha,\hat{\beta},\hat{\rho}}$ if we assume $\alpha$ known and we estimate $\beta$ and $\rho$ 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 $\widehat{\gamma}^{MP}_{n,\alpha,\hat{\beta},\hat{\rho}}(k)$, i.e., values of $k$ such that $k = o(n)$ and $\sqrt{k}\;A(n/k) \to \lambda$, finite, as $n \to \infty$. If we estimate $\alpha$ and $\gamma$ jointly through the maximum likelihood procedure, we can state the following:

**Theorem 3.1.** *Under the second order framework in* (2.3)*, with $A(t)$ given in* (2.4) *and for levels $k$ such that* (1.1) *holds, we have the asymptotic distributional representation,*

$$\widehat{\gamma}^{MP}_{n,\hat{\alpha},\hat{\beta},\hat{\rho}}(k) \overset{d}{=} \gamma + \frac{(1+\gamma)}{\sqrt{k}} S_k + b_{MP} A(n/k)(1 + o_p(1)), \tag{3.6}$$

*for the estimator $\widehat{\gamma}^{MP}_{n,\hat{\alpha},\hat{\beta},\hat{\rho}}(k)$ in* (2.11)*, with $S_k$ asymptotically standard normal and the notation*

$$b_{MP} := -\frac{(1+\gamma)(1+2\gamma)\left(\frac{1}{\rho}\ln\frac{(1+\gamma)(1-\rho)}{1+\gamma-\rho} + \frac{\gamma}{1+\gamma-\rho}\right)}{\gamma^3}.$$

*For the estimator* $\widehat{\gamma}_n^{ML}(k)$ *in* (1.4)*, we have, for any model satisfying* (2.2)*, the asymptotic distributional representation*

$$\widehat{\gamma}_n^{ML}(k) \stackrel{d}{=} \gamma + \frac{(1+\gamma)}{\sqrt{k}} M_k + \frac{(1+\gamma)(\gamma+\rho)\ A(n/k)}{\gamma(1-\rho)(1+\gamma-\rho)}\ (1+o_p(1)), \tag{3.7}$$

*with* $M_k$ *asymptotically standard normal.*

**Remark 3.1.** *As can be seen from* Theorem 3.1*, equation* (3.6)*, we no longer have a second-order reduced-bias tail index estimator, i.e., the estimation of* $\alpha$ *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 equation* (3.7)*, i.e., a context in which there is a 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 in* (1.4)*, being a second-order reduced-bias estimator of* $\gamma$*, is expected to outperform the* PORT-MP *estimator in* (2.11)*.*

## 3.2 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 *et al.* (2005), Gomes *et al.* (2007c) and Gomes and Neves (2007). Suppose that $\widehat{\gamma}_n^{\bullet}(k)$ is a general semi-parametric estimator of the tail index estimator, with the distributional representation,

$$\widehat{\gamma}_n^{\bullet}(k) = \gamma + \frac{\sigma_\bullet}{\sqrt{k}} Z_n^{\bullet} + b_\bullet A(n/k) + o_p(A(n/k)), \tag{3.8}$$

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

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

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

The *asymptotic mean-squared error* (AMSE) is given by

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

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 so-called optimal level for the estimation of $\gamma$ through $\widehat{\gamma}_n^\bullet(k)$, i.e., the level associated to a minimum asymptotic mean-squared error, 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 *et al.*, 1987) enabled Dekkers and de Haan (1993) to prove that, whenever $b_\bullet \neq 0$, there exists a function $\varphi(n) = \varphi(n;\rho,\gamma)$, dependent only on the underlying model, and not on the estimator, such that

$$\lim_{n\to\infty} \varphi(n) AMSE[\widehat{\gamma}_{n0}^\bullet] = \frac{2\rho-1}{2\rho} \left(\sigma_\bullet^2\right)^{-\frac{2\rho}{1-2\rho}} \left(b_\bullet^2\right)^{\frac{1}{1-2\rho}} =: LMSE[\widehat{\gamma}_{n0}^\bullet], \tag{3.9}$$

where $LMSE$ stands for *limiting mean-squared error.* Indeed, for each $\rho \leq 0$, there exists a positive decreasing function $s(.) \in RV_{2\rho-1}$, such that, as $t\to\infty$, $A^2(t) \sim \int_t^{+\infty} s(u)du$ (Proposition 1.7.3 of Geluk and de Haan, 1987). With $r := n/k$, $b\neq 0$ and by Lemma 2.9 of Dekkers and de Haan (1993), we have,

$$\inf_{r>0}\left\{r\frac{\sigma_\bullet^2}{n} + b_\bullet^2 A^2(r)\right\} = \inf_{r>0}\left\{r\frac{\sigma_\bullet^2}{n} + b_\bullet^2\int_r^{+\infty} s(u)du\right\}(1+o(1)) = \int_0^{\frac{\sigma_\bullet^2}{b_\bullet^2 n}} s^{\leftarrow}(u)du(1+o(1)),$$

and, with $r_0^\bullet = n/k_0^\bullet$, since $s^{\leftarrow}(1/t) \in RV_{-\frac{1}{2\rho-1}}$,

$$r_0^\bullet = \arg\inf_{r>0}\left\{r\frac{\sigma_\bullet^2}{n} + b_\bullet^2 A^2(r)\right\} = s^{\leftarrow}\left(\frac{\sigma_\bullet^2}{b_\bullet^2 n}\right)(1+o(1)) = \left(\frac{b_\bullet^2}{\sigma_\bullet^2}\right)^{\frac{1}{2\rho-1}} s^{\leftarrow}\left(\frac{1}{n}\right)(1+o(1)).$$

Hence,

$$k_0^\bullet := \arg\inf_k AMSE[\widehat{\gamma}_n^\bullet(k)] = \left(\frac{\sigma_\bullet^2}{b_\bullet^2}\right)^{\frac{1}{1-2\rho}} \frac{n}{s^{\leftarrow}(1/n)}(1+o(1)) \in RV_{\frac{2\rho}{2\rho-1}},$$

and with $\varphi(n) = n/s^{\leftarrow}(1/n)$ we get $\lim_{n\to\infty}\varphi(n)\, Var_\infty[\widehat{\gamma}_{n0}^\bullet] = \left(\sigma_\bullet^2\right)^{\frac{-2\rho}{1-2\rho}}\left(b_\bullet^2\right)^{\frac{1}{1-2\rho}}$, as well as, $\lim_{n\to\infty}|\sqrt{\varphi(n)}\; Bias_\infty[\widehat{\gamma}_{n0}^\bullet]| = (-2\rho)^{-1/2}\left(\sigma_\bullet^2\right)^{\frac{-\rho}{1-2\rho}}\left(b_\bullet\right)^{\frac{1}{1-2\rho}}$, and (3.9) follows.

It is then sensible to consider the following:

**Definition 3.1.** *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 type* (3.8) *hold with constants* $(\sigma_1, b_1)$

*and* $(\sigma_2, b_2)$, $b_1$, $b_2 \neq 0$, *respectively, both computed at their optimal levels, the asymptotic root efficiency* $(AREFF)$ *of* $\widehat{\gamma}_n^{(1)}$ *relatively to* $\widehat{\gamma}_n^{(2)}$ *is*

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

*with LMSE given in* (3.9).

**Remark 3.2.** *Note that this measure was devised so that the higher AREFF measure, the better the first estimator is.*

**Remark 3.3.** *The optimal levels for the estimation of* $\gamma$ *through* $\widehat{\gamma}_n^{MP}$ *and* $\widehat{\gamma}_n^{ML}$ *are denoted by* $k_0^{ML}$ *and* $k_0^{MP}$ *and are given by*

$$k_0^{ML} = k_0^{ML}(\rho, \beta) := \Big[\frac{(1-\rho)(1+\gamma-\rho)n^{-\rho}}{\sqrt{-2\rho}(\gamma+\rho)|\beta|}\Big]^{\frac{2}{1-2\rho}} \tag{3.10}$$

*and*

$$k_0^{MP} = k_0^{MP}(\rho, \beta) := \left[\frac{\gamma^2 n^{-\rho}}{\sqrt{-2\rho}(1+2\gamma)|\beta|\left(-\frac{1}{\rho}\ln\left(\frac{(1+\gamma)(1-\rho)}{1+\gamma-\rho}\right) - \frac{\gamma}{1+\gamma-\rho}\right)}\right]^{\frac{2}{1-2\rho}}, \tag{3.11}$$

*respectively.*

**Proposition 3.1.** *The* $AREFF$*-indicator of* $\widehat{\gamma}_n^{MP}$ *relatively to* $\widehat{\gamma}_n^{ML}$ *is*

$$AREFF_{MP|ML} = \left|\frac{\gamma^2(\gamma+\rho)}{(1+2\gamma)(1-\rho)(1+\gamma-\rho)\left(-\frac{1}{\rho}\ln\left(\frac{(1+\gamma)(1-\rho)}{1+\gamma-\rho}\right) - \frac{\gamma}{1+\gamma-\rho}\right)}\right|^{\frac{1}{1-2\rho}}.$$

This $AREFF_{MP|ML}$ measure is presented in Figure 1.

As can be seen, the gain in efficiency for the PORT-MP estimator happens for two regions of values of $(\gamma, \rho)$. 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 a topic of research out of the scope of this paper.

| ρ \ γ | 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 | 0.00 | 1.56 | 2.65 | 3.58 | 4.41 | 5.18 | 5.89 | 6.56 | 7.21 | 7.83 | 8.43 | 9.02 | 9.59 | 10.15 | 10.69 | 11.23 | 11.76 | 12.28 | 12.80 | 13.31 |
| -0.20 | 0.94 | 0.00 | 0.85 | 1.36 | 1.76 | 2.11 | 2.43 | 2.72 | 2.99 | 3.24 | 3.48 | 3.71 | 3.93 | 4.15 | 4.36 | 4.56 | 4.76 | 4.95 | 5.14 | 5.32 |
| -0.30 | 1.13 | 0.70 | 0.00 | 0.65 | 0.98 | 1.24 | 1.46 | 1.65 | 1.82 | 1.98 | 2.13 | 2.27 | 2.40 | 2.53 | 2.65 | 2.77 | 2.88 | 2.99 | 3.10 | 3.20 |
| -0.40 | 1.19 | 0.91 | 0.60 | 0.00 | 0.57 | 0.82 | 1.01 | 1.17 | 1.31 | 1.43 | 1.54 | 1.65 | 1.74 | 1.83 | 1.92 | 2.00 | 2.08 | 2.16 | 2.23 | 2.30 |
| -0.50 | 1.20 | 1.00 | 0.80 | 0.55 | 0.00 | 0.53 | 0.73 | 0.89 | 1.01 | 1.12 | 1.22 | 1.30 | 1.38 | 1.46 | 1.53 | 1.59 | 1.65 | 1.71 | 1.77 | 1.82 |
| -0.60 | 1.21 | 1.05 | 0.90 | 0.73 | 0.52 | 0.00 | 0.51 | 0.69 | 0.81 | 0.92 | 1.01 | 1.09 | 1.16 | 1.22 | 1.28 | 1.34 | 1.39 | 1.44 | 1.49 | 1.53 |
| -0.70 | 1.20 | 1.08 | 0.96 | 0.83 | 0.69 | 0.51 | 0.00 | 0.50 | 0.65 | 0.77 | 0.86 | 0.94 | 1.00 | 1.06 | 1.12 | 1.17 | 1.21 | 1.26 | 1.30 | 1.34 |
| -0.80 | 1.19 | 1.09 | 0.99 | 0.89 | 0.79 | 0.66 | 0.50 | 0.00 | 0.49 | 0.64 | 0.74 | 0.82 | 0.89 | 0.94 | 1.00 | 1.04 | 1.09 | 1.13 | 1.17 | 1.20 |
| -0.90 | 1.18 | 1.10 | 1.02 | 0.93 | 0.85 | 0.75 | 0.65 | 0.50 | 0.00 | 0.49 | 0.62 | 0.72 | 0.79 | 0.85 | 0.90 | 0.95 | 0.99 | 1.03 | 1.07 | 1.10 |
| -1.00 | 1.18 | 1.10 | 1.03 | 0.96 | 0.89 | 0.82 | 0.73 | 0.63 | 0.50 | 0.00 | 0.49 | 0.62 | 0.70 | 0.77 | 0.82 | 0.87 | 0.91 | 0.95 | 0.99 | 1.02 |
| -1.10 | 1.17 | 1.10 | 1.04 | 0.98 | 0.92 | 0.86 | 0.79 | 0.72 | 0.63 | 0.50 | 0.00 | 0.50 | 0.61 | 0.69 | 0.75 | 0.80 | 0.85 | 0.89 | 0.92 | 0.95 |
| -1.20 | 1.16 | 1.10 | 1.05 | 0.99 | 0.94 | 0.89 | 0.84 | 0.78 | 0.71 | 0.62 | 0.51 | 0.00 | 0.50 | 0.61 | 0.68 | 0.74 | 0.79 | 0.83 | 0.87 | 0.90 |
| -1.30 | 1.15 | 1.10 | 1.05 | 1.00 | 0.96 | 0.91 | 0.87 | 0.82 | 0.76 | 0.70 | 0.62 | 0.51 | 0.00 | 0.51 | 0.61 | 0.68 | 0.73 | 0.78 | 0.82 | 0.85 |
| -1.40 | 1.14 | 1.10 | 1.05 | 1.01 | 0.97 | 0.93 | 0.89 | 0.85 | 0.80 | 0.75 | 0.69 | 0.62 | 0.52 | 0.00 | 0.51 | 0.61 | 0.68 | 0.73 | 0.77 | 0.81 |
| -1.50 | 1.14 | 1.09 | 1.05 | 1.02 | 0.98 | 0.95 | 0.91 | 0.87 | 0.84 | 0.79 | 0.75 | 0.69 | 0.62 | 0.52 | 0.00 | 0.52 | 0.61 | 0.68 | 0.73 | 0.76 |
| -1.60 | 1.13 | 1.09 | 1.06 | 1.02 | 0.99 | 0.96 | 0.93 | 0.89 | 0.86 | 0.82 | 0.79 | 0.74 | 0.69 | 0.62 | 0.53 | 0.00 | 0.52 | 0.62 | 0.68 | 0.72 |
| -1.70 | 1.13 | 1.09 | 1.06 | 1.03 | 1.00 | 0.97 | 0.94 | 0.91 | 0.88 | 0.85 | 0.82 | 0.78 | 0.74 | 0.69 | 0.63 | 0.53 | 0.00 | 0.53 | 0.62 | 0.68 |
| -1.80 | 1.12 | 1.09 | 1.06 | 1.03 | 1.00 | 0.97 | 0.95 | 0.92 | 0.90 | 0.87 | 0.84 | 0.81 | 0.77 | 0.74 | 0.69 | 0.63 | 0.54 | 0.00 | 0.54 | 0.62 |
| -1.90 | 1.12 | 1.08 | 1.06 | 1.03 | 1.00 | 0.98 | 0.96 | 0.93 | 0.91 | 0.89 | 0.86 | 0.83 | 0.80 | 0.77 | 0.73 | 0.69 | 0.63 | 0.54 | 0.00 | 0.54 |
| -2.00 | 1.11 | 1.08 | 1.06 | 1.03 | 1.01 | 0.99 | 0.96 | 0.94 | 0.92 | 0.90 | 0.88 | 0.85 | 0.83 | 0.80 | 0.77 | 0.73 | 0.69 | 0.63 | 0.55 | 0.00 |

$1.00 < AREFF_{MP\,|\,ML} \le 1.20$ | $1.20 < AREFF_{MP\,|\,ML} \le 1.40$ | $1.40 < AREFF_{MP\,|\,ML} \le 2.00$ | $AREFF_{MP\,|\,ML} > 2.00$

Figure 1: Asymptotic relative efficiency of $\widehat{\gamma}_n^{MP}$ relatively to $\widehat{\gamma}_n^{ML}$ in the $(\gamma, \rho)$-plane.

## 4 Simulated behaviour of the estimators

In order to obtain the PORT-MP estimates introduced in this paper we present, in the appendix, the developed (and implemented) 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 $10 \times 100$ (10 replicates with 100 runs each), 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. The multi-sample simulation is a common practice in Monte Carlo procedures whenever we do not have an easy way to estimate measures of dispersion of a statistic, like for instance the $MSE$ (see Fishman (1972), for details). The idea is reasonably simple: in a multi-sample simulation of size $r \times m$, instead of generating a sample of a very large size (let us say $N = r \times m$) of observed values of a statistic, we collect $m$ observations of the statistic (here $m = 100$) on each of $r$ independent replications of the experiment (here $r = 10$). The value of $m$ also needs to be large enough to reduce bias, and provide asymptotic normality. We next take as an overall estimate of the population parameter of interest the average of the $r$ corresponding estimates computed on the

$r$ independent replications. Then, under very broad conditions, that overall estimator (which is a sample mean) converges to normality as $r$ increases. Moreover we can estimate the standard error of this overall estimate, even if $r$ is small. For small values of $r$, and whenever we can guarantee the asymptotic normality of the estimator of the characteristic under study, we can use the $t$-distribution with $r-1$ degrees of freedom to approximate its true distribution, and to derive a confidence interval for the parameter of interest. For further details on multi-sample simulation see also Gomes and Oliveira (2001).

In Figures 2 and 3 we show, on the basis of the first replicate (100 runs), 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 $\widehat{\gamma}_n^{MP}(k)$, in (1.4) and (2.11), respectively.

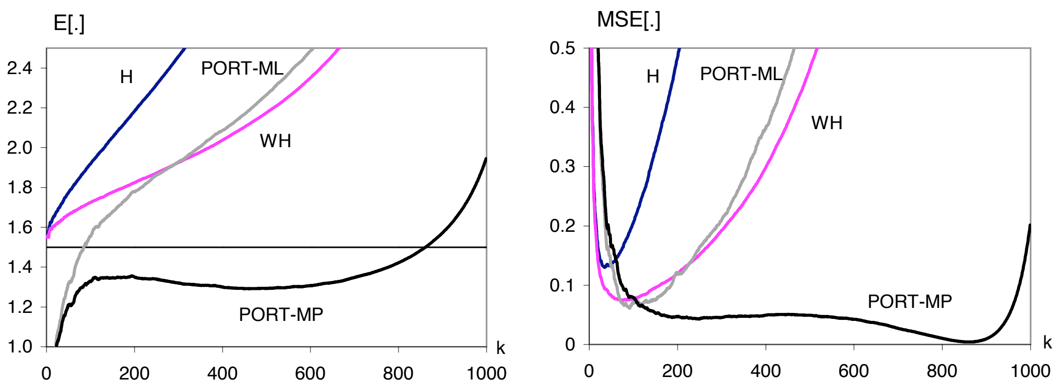


Figure 2: Mean values and mean-squared errors of the estimators under study for a sample of size $n = 1000$, from a *Burr* parent with $\rho = -0.5$ and $\gamma = 1.5$.

The simulations show that the PORT-MP tail index estimator 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 $(\gamma,\rho)$-region the bias is always smaller than the corresponding one of the PORT-ML estimator, for all $k$. The mean-squared error of the PORT-MP estimator is, in general, smaller than the mean-squared error of the PORT-ML estimator for a large region of values of $k$, as well as at

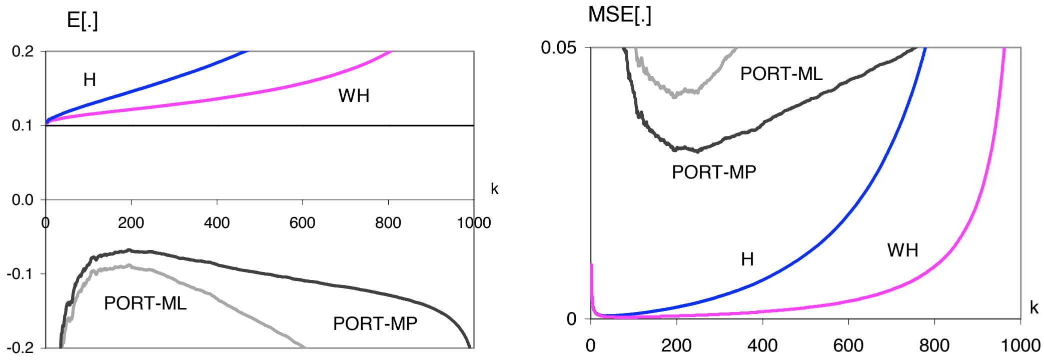


Figure 3: Mean values and mean-squared errors of the estimators under study for a sample of size $n = 1000$, from a *Burr* parent with $\rho = -0.5$ and $\gamma = 0.1$.

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 (2.12).

In Figure 4 we present the mean value and mean-squared of the estimators for a *Burr* model with $(\gamma, \rho) = (0.5, -0.5)$. When $\gamma + \rho = 0$ the PORT-ML estimator is second-order asymptotically unbiased for the estimation of $\gamma$, 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 $k$.

In Table 1 we present, for $n = 100, 200, 500$ and 1000, the optimal sample fractions, bias and mean-squared errors at optimal $k$-levels, i.e., levels where mean-squared errors are minimal, as functions of $k$, as well as the corresponding 95% confidence intervals, computed on the basis of the 10 replicates with 100 runs each.

We have next considered, for models with $\gamma + \rho \neq 0$, the simulated relative efficiencies of

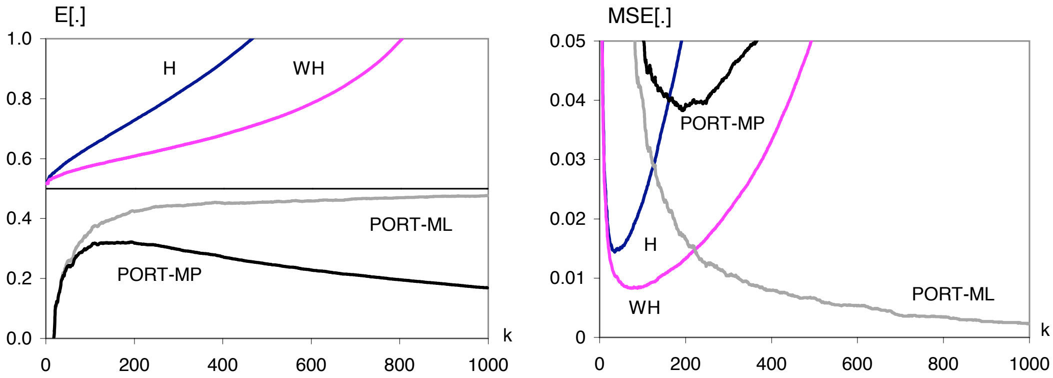


Figure 4: Mean values and mean-squared errors of the estimators under study for a sample of size $n = 1000$, from a *Burr* parent with $\rho = -0.5$ and $\gamma = 0.5$.

$\widehat{\gamma}_n^{MP}|\widehat{\gamma}_n^{ML}$ at optimal levels, denoted by $\overline{R}_0$, and the estimated relative efficiencies of $\widehat{\gamma}_n^{MP}|\widehat{\gamma}_n^{ML}$ at estimated optimal levels, denoted by $\widehat{R}_0$. The indicators of relative efficiency ($REFF$) are given by

$$\overline{R}_0 := \sqrt{\frac{MSE[\widehat{\gamma}_n^{ML}(k_{0s}^{ML})]}{MSE[\widehat{\gamma}_n^{MP}(k_{0s}^{MP})]}} =: \sqrt{\frac{MSE_0^{ML}}{MSE_0^{MP}}} \quad \text{and} \quad \widehat{R}_0 := \sqrt{\frac{MSE[\widehat{\gamma}_n^{ML}(\widehat{k}_0^{ML})]}{MSE[\widehat{\gamma}_n^{MP}(\widehat{k}_0^{MP})]}},$$

where $\widehat{\gamma}_n^{ML}(k)$ and $\widehat{\gamma}_n^{MP}(k)$ are the estimators given in (1.4) and (2.11), respectively, $k_{0s}^{ML}$ and $k_{0s}^{MP}$ are the simulated optimal levels of the estimators, $\widehat{k}_0^{ML} := k_0^{ML}(\widehat{\rho}, \widehat{\beta})$ and $\widehat{k}_0^{MP} := k_0^{MP}(\widehat{\rho}, \widehat{\beta})$ are their estimated optimal levels, with $k_0^{ML}(\rho, \beta)$ and $k_0^{MP}(\rho, \beta)$ provided in (3.10) and (3.11), respectively. We have considered the estimators of the second order parameters $\rho$ and $\beta$, $\widehat{\rho} := \widehat{\rho}(k_1, \tau)$ and $\widehat{\beta} := \widehat{\beta}(k_1; \widehat{\rho})$, computed at the level $k_1 = n^{0.995}$, with $\widehat{\rho}(k; \tau)$ and $\hat{\beta}(k; \hat{\rho})$ given in (2.13) and (2.14), respectively. As mentioned before, as $|\rho| < 1$ in all simulated models, we have chosen the tuning parameter $\tau = 0$. When $\gamma + \rho = 0$, the PORT-ML estimator is a second order reduced bias estimator and we therefore need to consider heuristic adaptive choices of the threshold, as in lines of Gomes *et al.* (2007a). The adaptive choices considered in this work are: $\hat{k}_{01}^{ML} := \hat{k}_0^{MP}$ and $\hat{k}_{02}^{ML} := n - 1$, with indicators,

$$\widehat{R}_1 := \sqrt{\frac{MSE[\widehat{\gamma}_n^{ML}(\widehat{k}_0^{MP})]}{MSE[\widehat{\gamma}_n^{MP}(\widehat{k}_0^{MP})]}} \quad \text{and} \quad \widehat{R}_2 := \sqrt{\frac{MSE[\widehat{\gamma}_n^{ML}(n-1)]}{MSE[\widehat{\gamma}_n^{MP}(\widehat{k}_0^{MP})]}}.$$

| $n$ | 100 | 200 | 500 | 1000 |
|---|---|---|---|---|
| $(\gamma, \rho) = (0.1, -0.5)$ | | | | |
| $k_{0s}^{ML}/n$ | 0.4080 ± 0.0221 | 0.3185 ± 0.0247 | 0.2342 ± 0.0275 | 0.1980 ± 0.0129 |
| $k_{0s}^{MP}/n$ | 0.5010 ± 0.0483 | 0.3600 ± 0.0223 | 0.2550 ± 0.0359 | 0.2354 ± 0.0130 |
| $E_0^{ML}$ | -0.2096 ± 0.0153 | -0.1500 ± 0.0094 | -0.0985 ± 0.0165 | -0.0888 ± 0.0011 |
| $E_0^{MP}$ | -0.1061 ± 0.0125 | -0.0857 ± 0.0046 | -0.0669 ± 0.0112 | -0.0690 ± 0.0009 |
| $MSE_0^{ML}$ | 0.1279 ± 0.0119 | 0.0809 ± 0.0043 | 0.0483 ± 0.0052 | 0.0406 ± 0.0001 |
| $MSE_0^{MP}$ | 0.0519 ± 0.0046 | 0.0415 ± 0.0018 | 0.0320 ± 0.0027 | 0.0307 ± 0.0003 |
| $(\gamma, \rho) = (0.5, -0.5)$ | | | | |
| $k_{0s}^{ML}/n$ | 0.9720 ± 0.0184 | 0.9865 ± 0.0084 | 0.9866 ± 0.0172 | 0.9955 ± 0.0024 |
| $k_{0s}^{MP}/n$ | 0.4650 ± 0.0411 | 0.3350 ± 0.0220 | 0.2306 ± 0.0255 | 0.1745 ± 0.0173 |
| $E_0^{ML}$ | 0.4588 ± 0.0135 | 0.4704 ± 0.0065 | 0.4746 ± 0.0068 | 0.4763 ± 0.0042 |
| $E_0^{MP}$ | 0.3025 ± 0.0189 | 0.3153 ± 0.0082 | 0.3261 ± 0.0112 | 0.3211 ± 0.0025 |
| $MSE_0^{ML}$ | 0.0241 ± 0.0038 | 0.0112 ± 0.0010 | 0.0043 ± 0.0005 | 0.0025 ± 0.0002 |
| $MSE_0^{MP}$ | 0.0568 ± 0.0072 | 0.0479 ± 0.0025 | 0.0397 ± 0.0033 | 0.0385 ± 0.0007 |
| $(\gamma, \rho) = (1.5, -0.5)$ | | | | |
| $k_{0s}^{ML}/n$ | 0.2340 ± 0.0379 | 0.1785 ± 0.0243 | 0.1492 ± 0.0129 | 0.1052 ± 0.0123 |
| $k_{0s}^{MP}/n$ | 0.7670 ± 0.0202 | 0.8190 ± 0.0095 | 0.8488 ± 0.0046 | 0.8592 ± 0.0035 |
| $E_0^{ML}$ | 1.7315 ± 0.0619 | 1.6566 ± 0.0486 | 1.6375 ± 0.0327 | 1.5375 ± 0.0665 |
| $E_0^{MP}$ | 1.4976 ± 0.0152 | 1.4960 ± 0.0063 | 1.4969 ± 0.0028 | 1.4972 ± 0.0028 |
| $MSE_0^{ML}$ | 0.3448 ± 0.0330 | 0.1933 ± 0.0124 | 0.0968 ± 0.0074 | 0.0546 ± 0.0046 |
| $MSE_0^{MP}$ | 0.0481 ± 0.0045 | 0.0246 ± 0.0026 | 0.0092 ± 0.0009 | 0.0051 ± 0.0006 |

Table 1: Optimal sample fractions, mean values and mean-squared errors of $\widehat{\gamma}_n^{ML}(k)$ and $\widehat{\gamma}_n^{MP}(k)$ for *Burr* parents with $\gamma = \{0.1, 0.5, 1.5\}$ and $\rho = -0.5$.

The results, obtained on the basis of a simulation experiment with 10 replicates of 100 runs each, are presented in Table 2. The underlined values correspond to REFF-indicators smaller than one.

| $n$ | 100 | 200 | 500 | 1000 |
|---|---|---|---|---|
| $(\gamma, \rho) = (0.1, -0.5)$ | | | | |
| $\overline{R}_0$ | 1.5699 ± 0.0218 | 1.3959 ± 0.0131 | 1.2261 ± 0.0145 | 1.1502 ± 0.0040 |
| $\widehat{R}_0$ | 1.5521 ± 0.0277 | 1.3894 ± 0.0358 | 1.2231 ± 0.0215 | 1.1231 ± 0.0124 |
| $(\gamma, \rho) = (0.5, -0.5)$ | | | | |
| $\overline{R}_0$ | <u>0.6508</u> ± 0.0388 | <u>0.4829</u> ± 0.0279 | <u>0.3315</u> ± 0.0258 | <u>0.2525</u> ± 0.0102 |
| $\widehat{R}_1$ | <u>0.7472</u> ± 0.0385 | <u>0.6468</u> ± 0.0258 | <u>0.5520</u> ± 0.0202 | <u>0.5189</u> ± 0.0101 |
| $\widehat{R}_2$ | <u>0.6125</u> ± 0.5598 | <u>0.4429</u> ± 0.4760 | <u>0.3086</u> ± 0.3974 | <u>0.2409</u> ± 0.3511 |
| $(\gamma, \rho) = (1.5, -0.5)$ | | | | |
| $\overline{R}_0$ | 2.6843 ± 0.1344 | 2.8264 ± 0.1956 | 3.2506 ± 0.1316 | 3.2911 ± 0.2829 |
| $\widehat{R}_0$ | 3.9339 ± 0.8311 | 3.1137 ± 0.1895 | 2.1662 ± 0.1482 | 1.8628 ± 0.1097 |

Table 2: Relative efficiency indicators for *Burr* parents with $\gamma = \{0.1, 0.5, 1.5\}$ and $\rho = -0.5$.

# 5 Appendix

## 5.1 The asymptotic behaviour of the r.v.'s under play

With $\{E_i\}_{i\geq 1}$ and $\{Y_i\}_{i\geq 1}$ denoting sequences of independent, identically distributed (i.i.d.), standard exponential and unit Pareto r.v.'s, respectively, and with the usual notation for the o.s., let us further introduce the following notations:

$$P_k^{(j)} := \frac{1}{k}\sum_{i=1}^{k} \psi_{ik}^{j}\ E_{k-i+1:k},\ j \geq 0, \tag{5.1}$$

$$Q_k^{(j)} := \frac{1}{k}\sum_{i=1}^{k} \psi_{ik}^{j-1}\ \frac{Y_{k-i+1:k}^{\rho}-1}{\rho},\ j \geq 1, \tag{5.2}$$

$$R_k^{(j)} := \frac{1}{k}\sum_{i=1}^{k} \psi_{ik}^{j}\ \left(1-e^{-\gamma E_{k-i+1:k}}\right),\ j \geq 0, \tag{5.3}$$

$$Z_k^{(j)} := \frac{1}{k}\sum_{i=1}^{k} \psi_{ik}^{j}\ Y_{k-i+1:k}^{-\gamma}\ \frac{Y_{k-i+1:k}^{\rho}-1}{\rho},\ j \geq 0, \tag{5.4}$$

and

$$
\begin{aligned}
b_j &:= \frac{(-1)^{j-1}}{\rho^j}\int_0^1 \frac{(x^{-\rho}-1)^j}{\ln^{j-1} x}dx,\ j\geq 1,\ b_0=1, && (5.5)\\
c_{j,1} &:= \frac{1}{|\rho|^j}\int_0^1 \frac{(x^{-\rho}-1)^j(1-x^{\gamma})}{\ln^j x}dx,\ j\geq 0, && (5.6)\\
c_{j,2} &:= -\frac{1}{|\rho|^{j+1}}\int_0^1 \frac{x^{\gamma}(x^{-\rho}-1)^{j+1}}{\ln^j x}\ dx,\ j\geq 0, && (5.7)
\end{aligned}
$$

where $\psi_{ik} := -((i/k)^{-\rho}-1)/(\rho\ln(i/k))$ has been defined in (2.8) and $P_k^{(j)}$, $Q_k^{(j)}$, $R_k^{(j)}$ and $Z_k^{(j)}$ are asymptotically normal r.v's.

We first state the following result:

**Lemma 5.1.** *The first order structure of the r.v's $P_k^{(j)}$, $Q_k^{(j)}$, $R_k^{(j)}$ and $Z_k^{(j)}$, provided in* (5.1), (5.2), (5.3) *and* (5.4), *respectively, is given by*

$$\mathbb{E}(P_k^{(0)}) = b_0, \quad \mathbb{E}(Q_k^{(1)}) = b_1, \quad \mathbb{E}(R_k^{(0)}) = c_{0,1}, \quad \mathbb{E}(Z_k^{(0)}) = c_{0,2},$$

*and, for $j\geq 1$,*

$$\lim_{k\to\infty}\mathbb{E}(P_k^{(j)}) = \lim_{k\to\infty}\mathbb{E}(Q_k^{(j)}) = b_j, \quad \lim_{k\to\infty}\mathbb{E}(R_k^{(j)}) = c_{j,1}, \quad \lim_{k\to\infty}\mathbb{E}(Z_k^{(j)}) = c_{j,2},$$

*with $b_j$, $c_{j,1}$ and $c_{j,2}$ defined in* (5.5) (5.6) *and* (5.7), *respectively. The second order structure for $j=0$ is given by*

$$k\mathbb{V}ar(P_k^{(0)}) = 1; \quad k\mathbb{V}ar(R_k^{(0)}) = \frac{\gamma^2}{(1+\gamma)^2(1+2\gamma)}; \quad k\mathbb{C}ov(P_k^{(0)}, R_k^{(0)}) = \frac{\gamma}{(1+\gamma)^2}. \tag{5.8}$$

*Proof.* For $j=0$, $P_k^{(0)} = \frac{1}{k}\sum_{i=1}^k E_{k-i+1:k} = \frac{1}{k}\sum_{i=1}^k E_i$. Consequently, $\mathbb{E}[P_k^{(0)}] = 1 = b_0$, in (5.5). On the basis of Rényi's representation (Rényi, 1953) of standard exponential o.s. as a linear combination of independent standard exponential r.v.'s, given by

$$E_{i:n} \stackrel{d}{=} \sum_{j=1}^{i}\frac{E_j}{n-j+1}, \quad 1\leq i\leq n,$$

we can write for $j\geq 1$,

$$\begin{aligned}
\mathbb{E}[P_k^{(j)}] &= \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j}\ \mathbb{E}[E_{k-i+1:k}] = \frac{(-1)^j}{\rho^j}\Big(\frac{1}{k}\sum_{i=1}^{k}\Big(\frac{(i/k)^{-\rho}-1}{\ln(i/k)}\Big)^j\Big(\frac{1}{k}\sum_{j=i}^{k}\frac{1}{j/k}\Big)\Big)\\
&\underset{k\to\infty}{\longrightarrow} \frac{(-1)^j}{\rho^j}\int_{[0,1]}\left(\frac{x^{-\rho}-1}{\ln x}\right)^j\ dx\int_x^1\frac{1}{y}\ dy\\
&= \frac{(-1)^{j-1}}{\rho^j}\int_0^1\frac{(x^{-\rho}-1)^j}{\ln^{j-1}x}\ dx = b_j,\quad \text{as given in (5.5).}
\end{aligned}$$

Regarding $Q_k^{(j)}$, $R_k^{(j)}$ and $Z_k^{(j)}$ in (5.2), (5.3) and (5.4), respectively, we get

$$Q_k^{(1)} = \frac{1}{k}\sum_{i=1}^{k}\frac{Y_i^\rho-1}{\rho},\quad R_k^{(0)} = \frac{1}{k}\sum_{i=1}^{k}\Big(1-e^{-\gamma E_i}\Big),\quad Z_k^{(0)} = \frac{1}{k}\sum_{i=1}^{k}Y_i^{-\gamma}\frac{Y_i^\rho-1}{\rho}.$$

Consequently,

$$\mathbb{E}[Q_k^{(1)}] = \mathbb{E}\Big(\frac{Y^\rho-1}{\rho}\Big) = \frac{1}{1-\rho} = b_1,\quad \mathbb{E}[R_k^{(0)}] = 1-\mathbb{E}[e^{-\gamma E}] = 1-\mathbb{E}(Y^{-\gamma}) = \frac{\gamma}{1+\gamma} = c_{0,1}$$

and

$$\mathbb{E}[Z_k^{(0)}] = \mathbb{E}(Y^{-\gamma}(Y^\rho-1)/\rho) = \frac{1}{(1+\gamma)(1+\gamma-\rho)} = c_{0,2}.$$

Asymptotically, we have for $j>1$,

$$\begin{aligned}
\mathbb{E}[Q_k^{(j)}] &\sim \frac{1}{k}\sum_{i=1}^{k}\ \psi_{ik}^{j-1}\frac{(i/k)^{-\rho}-1}{\rho}\\
&= \frac{(-1)^{j-1}}{\rho^j}\Big(\frac{1}{k}\sum_{i=1}^{k}\ \frac{((i/k)^{-\rho}-1)^j}{\ln^{j-1}(i/k)}\Big) \underset{k\to\infty}{\longrightarrow} b_j,\quad \text{as given in (5.5).}
\end{aligned}$$

For $j\geq 1$,

$$\begin{aligned}
\mathbb{E}[R_k^{(j)}] &\sim \frac{(-1)^j}{\rho^j}\Big(\frac{1}{k}\sum_{i=1}^{k}\ \Big(\frac{(i/k)^{-\rho}-1}{\ln(i/k)}\Big)^j\ (1-(i/k)^\gamma)\Big)\\
&\underset{k\to\infty}{\longrightarrow} \frac{1}{|\rho|^j}\int_0^1\frac{(x^{-\rho}-1)^j(1-x^\gamma)}{\ln^j x} = c_{j,1},\ \text{as given in (5.6),}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[Z_k^{(j)}] &\sim \frac{(-1)^j}{\rho^{j+1}}\Big(\frac{1}{k}\sum_{i=1}^{k}\ \frac{((i/k)^{-\rho}-1)^{j+1}}{\ln^j(i/k)}(i/k)^\gamma\Big)\\
&\underset{k\to\infty}{\longrightarrow} -\frac{1}{|\rho|^{j+1}}\int_0^1\frac{x^\gamma(x^{-\rho}-1)^{j+1}}{\ln^j x} = c_{j,2},\ \text{as given in (5.7).}
\end{aligned}$$

Also, we trivially have $k\mathbb{V}ar\left(P_k^{(0)}\right) = \mathbb{V}ar[E] = 1$. The other results in (5.8) come also straightforwardly from the fact that $\mathbb{E}(e^{-\gamma E}) = 1/(1+\gamma)$ and $\mathbb{E}(E\ e^{-\gamma E}) = 1/(1+\gamma)^2$. Consequently $\mathbb{C}ov(E, 1-e^{-\gamma E}) = \gamma/(1+\gamma)^2$. $\square$

**Remark 5.1.** *For $j = 0,\ 1$, the values of $b_j$, $c_{j,1}$ and $c_{j,2}$, defined in* (5.5) (5.6) *and* (5.7)*, respectively, are given by*

$$\begin{aligned}
b_0 &= 1; \qquad b_1 = \frac{1}{1-\rho}; \\
c_{0,1} &= \frac{\gamma}{1+\gamma}; \qquad c_{1,1} = -\frac{1}{\rho}\ln\frac{(1+\gamma)(1-\rho)}{1+\gamma-\rho}; \\
c_{0,2} &= \frac{1}{(1+\gamma)(1+\gamma-\rho)}; \qquad c_{1,2} = -\frac{1}{\rho^2}\ln\frac{(1+\gamma)(1+\gamma-2\rho)}{(1+\gamma-\rho)^2}.
\end{aligned}$$

Knowing that $\psi_{ik}$ is a limited function and $k$ an intermediate sequence, the r.v.'s $B_{(j)}$, $C_{(j)}$ and $D_{(j)}$ defined in (3.1), (3.2) and (3.3), respectively, can be written as:

$$\begin{aligned}
B_{(j)} &= \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j-1}\ e^{-\beta(n/k)^\rho\psi_{ik}}\ln(1+\alpha\ W_{ik}) \\
&= \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j-1}\ln(1+\alpha\ W_{ik}) - \frac{A(n/k)}{\gamma}\frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j}\ln(1+\alpha\ W_{ik})(1+o_p(1)) \\
&=: B_j - \frac{A(n/k)}{\gamma}\ B_{j+1}(1+o_p(1)).
\end{aligned} \tag{5.9}$$

$$\begin{aligned}
C_{(j)} &= \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j-1}\ e^{-\beta(n/k)^\rho\psi_{ik}}\frac{\alpha\ W_{ik}}{1+\alpha\ W_{ik}} \\
&= \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j-1}\frac{\alpha\ W_{ik}}{1+\alpha\ W_{ik}} - \frac{A(n/k)}{\gamma}\frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j}\frac{\alpha\ W_{ik}}{1+\alpha\ W_{ik}}(1+o_p(1)) \\
&=: C_j - \frac{A(n/k)}{\gamma}\ C_{j+1}(1+o_p(1)).
\end{aligned} \tag{5.10}$$

$$\begin{aligned}
D_{(j)} &= \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j-1}\ e^{-\beta(n/k)^\rho\psi_{ik}}\frac{\alpha\ W_{ik}}{(1+\alpha\ W_{ik})^2} \\
&= \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j-1}\frac{\alpha\ W_{ik}}{(1+\alpha\ W_{ik})^2} - \frac{A(n/k)}{\gamma}\frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j}\frac{\alpha\ W_{ik}}{(1+\alpha\ W_{ik})^2} \\
&=: D_j - \frac{A(n/k)}{\gamma}\ D_{j+1}(1+o_p(1)).
\end{aligned} \tag{5.11}$$

**Remark 5.2.** *From equations* (5.9), (5.10) *and* (5.11) *we observe that:* $B_1 \equiv B$, $C_1 \equiv C$ *and* $D_1 \equiv D$, *with* $B$, $C$ *and* $D$ *defined at the beginning of Section 3.1, in* (3.1), (3.2) *and* (3.3), *respectively.*

For the r.v.'s under play we have the validity of the following distributional representations:

**Lemma 5.2.** *If the second order condition* (2.3) *holds, if* $k = k_n$ *is a sequence of intermediate positive integers, i.e.,* (1.1) *holds, and with* $A(t)$ *given in* (2.4), *we have the validity of the distributional representations,*

$$B_j \stackrel{d}{=} \gamma b_{j-1} + \gamma\left(P_k^{(j-1)} - b_{j-1}\right) + b_j\ A(n/k)(1+o_p(1)), \tag{5.12}$$

$$B_{(j)} \stackrel{d}{=} \gamma b_{j-1} + \gamma\left(P_k^{(j-1)} - b_{j-1}\right) + o_p(A(n/k)),$$

$$C_j \stackrel{d}{=} c_{j-1,1} + \left(R_k^{(j-1)} - c_{j-1,1}\right) + c_{j-1,2}\ A(n/k)(1+o_p(1)), \tag{5.13}$$

$$C_{(j)} \stackrel{d}{=} c_{j-1,1} + \left(R_k^{(j-1)} - c_{j-1,1}\right) + \left(c_{j-1,2} - \frac{c_{j,1}}{\gamma}\right)\ A(n/k)(1+o_p(1)),$$

*with* $b_j$, $c_{j,1}$ *and* $c_{j,2}$ *defined in* (5.5), (5.6) *and* (5.7), *respectively. Moreover,*

$$D \equiv D_1 \stackrel{d}{=} \frac{\gamma}{(1+\gamma)(1+2\gamma)} + o_p(1). \tag{5.14}$$

*Consequently, with* $c_{1,1}$, *given explicitly in Remark 5.1, we have*

$$1 - C - \frac{C_{(1)}}{B_{(1)}} \stackrel{d}{=} \frac{\gamma}{(1+\gamma)\sqrt{1+2\gamma}}\ \frac{S_k}{\sqrt{k}} + \frac{c_{1,1} - \frac{\gamma}{1+\gamma-\rho}}{\gamma^2}\ A(n/k)(1+o_p(1)), \tag{5.15}$$

*with* $S_k$ *asymptotically standard normal. We also have*

$$D + \frac{D_{(1)}}{B_{(1)}} - \left(\frac{C_{(1)}}{B_{(1)}}\right)^2 \stackrel{d}{=} \frac{\gamma^2}{(1+\gamma)^2(1+2\gamma)} + o_p(1). \tag{5.16}$$

*Proof.* Let us think first on $B_j$ in (5.9). Since

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

we can write

$$\begin{aligned} B_j &= \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j-1}\ \ln(1+\alpha\ W_{ik}) \\ &\stackrel{d}{=} \frac{\gamma}{k}\sum_{i=1}^{k}\psi_{ik}^{j-1}\ E_{k-i+1:k} + \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j-1}\ \frac{Y_{k-i+1:k}^{\rho} - 1}{\rho}\ A(n/k)\ (1+o_p(1)). \end{aligned}$$

Then, from the results related to the convergence in probability of $P_k^{(j)}$ and $Q_k^{(j)}$ in Lemma 5.1, (5.12) follows. Analogously, we have

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

Consequently, now for the r.v. $C_j$ in (5.10),

$$\begin{aligned}
C_j = \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j-1}\ \frac{\alpha\ W_{ik}}{1+\alpha\ W_{ik}} &\stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}\ \psi_{ik}^{j-1}(1-e^{-\gamma\ E_{k-i+1:k}}) \\
&\quad + \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j-1}\ Y_{k-i+1:k}^{-\gamma}\ \frac{Y_{k-i+1:k}^{\rho}-1}{\rho}\ A(n/k)(1+o_p(1)),
\end{aligned}$$

and (5.13) follows as well. The results for $B_{(j)}$ and $C_{(j)}$ follow straightforwardly from the previous results for $B_j$ and $C_j$, together with (5.9) and (5.10). Finally,

$$\begin{aligned}
\frac{\alpha\ W_{ik}}{(1+\alpha\ W_{ik})^2} &\stackrel{d}{=} \frac{1 - Y_{k-i+1:k}^{-\gamma} + A(n/k)\ \frac{Y_{k-i+1:k}^{\rho}-1}{\rho}(1+o_p(1))}{Y_{k-i+1:k}^{\gamma}\left(1 + A(n/k)\ \frac{Y_{k-i+1:k}^{\rho}-1}{\rho}(1+o_p(1))\right)^2} \\
&= \left(Y_{k-i+1:k}^{-\gamma}(1-Y_{k-i+1:k}^{-\gamma}) + A(n/k)Y_{k-i+1:k}^{-\gamma}\frac{Y_{k-i+1:k}^{\rho}-1}{\rho}(1+o_p(1))\right) \\
&\qquad \times\left(1 - 2A(n/k)\ \frac{Y_{k-i+1:k}^{\rho}-1}{\rho}(1+o_p(1))\right) \\
&= Y_{k-i+1:k}^{-\gamma}(1-Y_{k-i+1:k}^{-\gamma}) + A(n/k)\left(2Y_{k-i+1:k}^{-2\gamma} - Y_{k-i+1:k}^{-\gamma}\right)\frac{Y_{k-i+1:k}^{\rho}-1}{\rho}(1+o_p(1)).
\end{aligned}$$

Consequently, with $D_j$ given in (5.11),

$$\begin{aligned}
D \equiv D_1 &= \frac{1}{k}\sum_{i=1}^{k}\frac{\alpha\ W_{ik}}{(1+\alpha\ W_{ik})^2} \\
&\stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}e^{-\gamma\ E_i}\left(1-e^{-\gamma\ E_i}\right) + \frac{1}{k}\sum_{i=1}^{k}\left(2Y_i^{-2\gamma} - Y_i^{-\gamma}\right)\frac{Y_i^{\rho}-1}{\rho}A(n/k)(1+o_p(1))
\end{aligned}$$

and (5.14) follows, as well as the remaining of the theorem. Indeed,

$$1-C-\frac{C_{(1)}}{B_{(1)}} \stackrel{d}{=} \Big(\frac{P_k^{(0)}-b_0}{1+\gamma}-\frac{(1+\gamma)(R_k^{(0)}-c_{0,1})}{\gamma}\Big)+\frac{c_{1,1}-\gamma(1+\gamma)c_{0,2}}{\gamma^2}\,A(n/k)(1+o_p(1)), \quad (5.17)$$

where $c_{0,1}$, $c_{0,2}$ and $c_{1,1}$ are made explicit in Remark 5.1, and from Lemma 5.1,

$$\mathbb{V}ar\Big(\frac{P_k^{(0)}-b_0}{1+\gamma}-\frac{(1+\gamma)(R_k^{(0)}-c_{0,1})}{\gamma}\Big)=\frac{\gamma^2}{k(1+\gamma)^2(1+2\gamma)}.$$

We can thus write (5.15). The result in (5.16) follows straightforwardly from (5.12), (5.13) and (5.14). □

## 5.2 Proofs of the main results

*Proof.* [Theorem 2.1] From (2.7), we can write

$$\ln(1+\alpha W_{ik}) \;\stackrel{d}{=}\; \gamma \ln Y_{k-i+1:k}\Big(1+\frac{A(n/k)}{\gamma}\,\frac{Y^{\rho}_{k-i+1:k}-1}{\rho\ln Y_{k-i+1:k}}(1+o_p(1))\Big),$$

and $\widehat{\gamma}^{MP}_{n,\alpha,\beta,\rho}(k)$ can be written as

$$\frac{\gamma}{k}\sum_{i=1}^{k} E_{k-i+1:k}\Big(1+\frac{A(n/k)}{\gamma}\,\frac{Y^{\rho}_{k-i+1:k}-1}{\rho\;E_{k-i+1:k}}(1+o_p(1))\Big)\Big(1-\frac{A(n/k)}{\gamma}\psi_{ik}(1+o(1))\Big).$$

Consequently, with $P_k^{(j)}$ and $Q_k^{(j)}$ given in (5.1) and (5.2), respectively, let us denote

$$T_k := \frac{1}{k}\sum_{i=1}^{k}\frac{Y^{\rho}_{k-i+1:k}-1}{\rho}-\frac{1}{k}\sum_{i=1}^{k}\psi_{ik}\;E_{k-i+1:k}=Q_k^{(1)}-P_k^{(1)}. \quad (5.18)$$

We have

$$\widehat{\gamma}^{MP}_{n,\alpha,\beta,\rho}(k)=\frac{\gamma}{k}\sum_{i=1}^{k}E_i+A(n/k)\;T_k\;(1+o_p(1)).$$

Let us study the asymptotic behaviour of the r.v. $T_k$ in (5.18): the weak law of large numbers enables to say that both $Q_k^{(1)}$ and $P_k^{(1)}$ converge in probability towards their mean values. We have

$$\mathbb{E}[Q_k^{(1)}]=\mathbb{E}[(Y^{\rho}-1)/\rho]=1/(1-\rho).$$

On the other hand, as seen before in Lemma 5.1,

$$\mathbb{E}[P_k^{(1)}]\underset{k\to\infty}{\longrightarrow}\frac{1}{1-\rho}=\mathbb{E}[Q_k^{(1)}].$$

Consequently $T_k$ converges in probability towards 0, as $k \to \infty$, and, with $N_k = \sqrt{k}\big(\frac{1}{k}\sum_{i=1}^{k} E_i - 1\big)$, (2.15) follows, i.e., the usual dominant component of bias, which is for the classical tail index estimators of the order of $A(n/k)$ is now of smaller order. □

*Proof.* [Theorem 3.1] From (5.17), together with (3.4) and (5.16), we get

$$\frac{\widehat{\alpha}_{MP}(k)-\alpha}{\alpha} \stackrel{d}{=} \frac{(1+\gamma)^2(1+2\gamma)}{\gamma^2\sqrt{k}}\Big(\frac{P_k^{(0)}-b_0}{1+\gamma} - \frac{(1+\gamma)(R_k^{(0)}-c_{0,1})}{\gamma}\Big) + \frac{(1+\gamma)^2(1+2\gamma)(c_{1,1}-\gamma(1+\gamma)c_{0,2})}{\gamma^4}\,A(n/k)(1+o_p(1)).$$

From the covariance structure between $P_k^{(0)}$ and $R_k^{(0)}$, given in (5.8), we can write

$$\frac{\widehat{\alpha}_{MP}(k)-\alpha}{\alpha} \stackrel{d}{=} \frac{(1+\gamma)\sqrt{1+2\gamma}}{\gamma\sqrt{k}}\,S_k + \frac{(1+\gamma)^2(1+2\gamma)(c_{1,1}-\gamma(1+\gamma)c_{0,2})}{\gamma^4}\,A(n/k)(1+o_p(1)),$$

with $S_k$ asymptotically standard normal. Finally, since we have, from (3.5),

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

we get

$$\widehat{\gamma}^{MP}_{n,\hat{\alpha},\hat{\beta},\hat{\rho}}(k) \stackrel{d}{=} \gamma + \gamma(P_k^{(0)}-b_0) + \frac{\gamma}{1+\gamma}\,\frac{\widehat{\alpha}_{MP}(k)-\alpha}{\alpha}(1+o_p(1)),$$

and (3.6) follows. The asymptotic distributional representation in (3.7) follows the same lines as before, with the replacement of $B_{(1)}$, $C_{(1)}$ and $D_{(1)}$, with $B_{(j)}$, $C_{(j)}$ and $D_{(j)}$ defined in (3.1), (3.2) and (3.3), respectively, by $B_1$, $C_1$ and $D_1$, with $B_j$, $C_j$ and $D_j$ defined in (5.9), (5.10) and (5.11), respectively. Indeed, we can write $\widehat{\gamma}_n^{ML}(k) = \widehat{B}_1$ and similarly to (3.4) we have

$$\frac{\widehat{\alpha}_{ML}-\alpha}{\alpha} = \frac{1-C-C_1/B_1}{D+D_1/B_1-(C_1/B_1)^2}\,(1+o_p(1)).$$

□

## 5.3 The Modified Version of Grimshaw's Algorithm on the Computation of ML Estimates for the GP Distribution

The log-likelihood of the excesses is, up to an additive constant, equal to

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

where $\underline{W} = (W_{ik},\ 1 \le i \le k)$ is the ordered sample of the excesses from the generalized Pareto model in (1.3), with $\gamma = \gamma_{ik}$ given in (2.9), and $p_{ik} = e^{-\beta(n/k)^\rho \psi_{ik}}$. Therefore, the $h$ function used in the estimation procedure described by Grimshaw is, now, given by

$$h_1(\theta) := \frac{1}{k}\sum_{i=1}^{k} p_{ik}(1-\theta\ W_{ik})^{-1} - \frac{1}{k}\sum_{i=1}^{k} p_{ik} + \Big(\frac{1}{k}\sum_{i=1}^{k} p_{ik}\ln(1-\theta\ W_{ik})\Big)\Big(\frac{1}{k}\sum_{i=1}^{k}(1-\theta\ W_{ik})^{-1}\Big), \tag{5.19}$$

where $\theta = -\alpha$. This function satisfies the necessary properties for the determination of the zeros, described in the following:

**(i)** $\lim_{\theta\to 1/W_{kk}^-} h_1(\theta) = -\infty$;

**(ii)** $h_1(\theta) < 0,\ \forall\, \theta\ <\ \theta_L := \frac{2(W_{1k}-\overline{W}_k)}{(W_{1k})^2}$;

**(iii)** $h_1'(\theta) := \frac{1}{\theta}\Big(\frac{1}{k}\sum_{i=1}^{k} p_{ik}(1-\theta\ W_{ik})^{-2} - \frac{1}{k}\sum_{i=1}^{k} p_{ik}(1-\theta\ W_{ik})^{-1} +$
$\frac{1}{k}\Big(\sum_{i=1}^{k} p_{ik} - \sum_{i=1}^{k} p_{ik}(1-\theta\ W_{ik})^{-1}\Big) \times \Big(\frac{1}{k}\sum_{i=1}^{k}(1-\theta\ W_{ik})^{-1}\Big)$
$-\Big[\frac{1}{k}\sum_{i=1}^{k} p_{ik}\ln(1-\theta\ W_{ik})\Big] \times \Big[\frac{1}{k}\sum_{i=1}^{k}(1-\theta\ W_{ik})^{-1} - \frac{1}{k}\sum_{i=1}^{k}(1-\theta\ W_{ik})^{-2}\Big]\Big)$;

**(iv)** $\lim_{\theta\to 0} h_1'(\theta) = 0$;

**(v)** $\lim_{\theta\to 0} h_1''(\theta) := \frac{1}{k}\sum_{i=1}^{k} p_{ik}W_{ik}^2 - 2\overline{W}_k\Big(\frac{1}{k}\sum_{i=1}^{k} p_{ik}W_{ik}\Big)$.

## 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] Caeiro, F. and Gomes, M.I. (2006). A new class of estimators of the “scale” second order parameter. *Extremes* **9**, 193-211, 2007.

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

[5] 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.

[6] Dekkers, A. and de Haan, L. (1993). Optimal sample fraction in extreme value estimation. *J. Multivariate Analysis* **47**(2), 173-195.

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

[8] Fishman G.S. (1972). *Concepts and Methods in Discrete Event Digital Simulation.* Wiley: New York.

[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, 193-213.

[10] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423-453.

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

[12] Gomes, M.I., de Haan, L. and Henriques Rodrigues, L. (2008). Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *J. Royal Statist. Soc.* **B70**, Issue 1, 31-52, DOI: 10.1007/s 10687-007-0048-9, 2007.

[13] Gomes, M.I., de Haan, L. and Peng, L. (2002). Semi-parametric estimation of the second order parameter — asymptotic and finite sample behaviour. *Extremes* **5**:4, 387-414.

[14] Gomes, M.I., Henriques Rodrigues, L., Vandewalle, B. and Viseu, C. (2007a). A Heuristic Adaptive Choise of the Threshold for Bias-Corrected Hill Estimators. *J. Statist. Comp. and Simul.*, DOI: 10.1080/10629360600954000, 2007.

[15] 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.

[16] 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, 5-31.

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

[18] 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.

[19] Gomes, M.I., Miranda, C. and Viseu, C. (2007c). Reduced bias extreme value index estimation and the Jackknife methodology, *Statistica Neerlandica* **61**:2, 243-270.

[20] 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.

[21] Gomes, M.I. and Oliveira, O. (2001). The bootstrap methodology in statistical extremes — the choice of the optimal sample fraction. *Extremes* **4**:4, 331-358.

[22] Gomes, M.I. and Pestana, D. (2007a). A simple second order reduced-bias' tail index estimator. *J. Statist. Comp. and Simul.*, **77**:6, 487-504.

[23] Gomes, M.I. and Pestana, D. (2007b). A sturdy second order reduced bias' Value at Risk estimator. *J. Amer. Statist. Assoc.*, Vol. **102**, No. 477, 280-292.

[24] Grimshaw, S.D. (1993). Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution. *Technometrics* **35**:2, 185-191.

[25] de Haan, L. (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*. Mathematical Centre Tract 32, Amsterdam.

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

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

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

[29] Rényi, A. (1953). On the theory of order statistics. *Acta Mathematica Scient. Hungar.*, Tomus **IV**, 191-227.

[30] Smith, R.L. (1987). Estimating tails of probability distributions. *The Annals of Statistics* **15**:3, 1174-1207.