Semi-parametric tail inference through Probability-Weighted Moments

Frederico Caeiro*
New University of Lisbon and CMA
fac@fct.unl.pt

and

M. Ivette Gomes†
University of Lisbon, DEIO, CEAUL and FCUL
ivette.gomes@fc.ul.pt

Abstract: In this paper, for heavy-tailed models, and working with the sample of the k largest observations, we present Probability Weighted Moments (PWM) estimators for the first-order tail parameters. Under regular variation conditions on the right-tail of the underlying distribution function F we prove the consistency and asymptotic normality of these estimators. Their performance, for finite sample sizes, is illustrated through a small-scale Monte Carlo simulation.

Keywords and phrases: Heavy tails, extreme value index, first order scale parameter, semi-parametric estimation, statistics of extremes

AMS 2000 subject classifications: Primary 62G32, 62E20; secondary 65C05

*Faculdade de Ciências e Tecnologia, 2829−516 Caparica, Portugal (Corresponding author.)
†Faculdade de Ciências, Campo Grande, 1749−016 Lisboa, Portugal

1 Introduction and preliminaries

Let $X_1, X_2, \ldots, X_n$ be a set of n independent and identically distributed (i.i.d.) random variables (r.v.'s), from a population with distribution function (d.f.) F. Let us arrange them in ascending order, to get the order statistics (o.s.) $X_{1:n} \le X_{2:n} \le \ldots \le X_{n:n}$. Suppose that we are interested in making inference about extreme values of X. Extreme Value Theory (EVT) provides a great variety of results that enable us to to deal with alternative approaches in the statistical analysis of extremes. We shall present a brief review of the most important approaches: the block maxima method, the peaks-over-threshold (POT) method and the largest observations method.

The first method is based on the unified version of the possible non-degenerate limit distributions of the normalized maximum, the Extreme Value (EV) distribution,

$$G_\gamma(x) = \begin{cases} \exp(-(1+\gamma x)^{-1/\gamma}), & 1+\gamma x > 0 \quad \text{if} \quad \gamma \neq 0 \\ \exp(-\exp(-x)), & x \in \mathbb{R} \quad \text{if} \quad \gamma = 0. \end{cases} \tag{1.1}$$

When such a non-degenerate limit exists, we say that F belongs to the max-domain of attraction of G_γ and denote this by $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$. The shape parameter γ is related with the heaviness of the right tail $\overline{F} := 1 - F$ and it is often called the *extreme value index* (EVI). This distribution unifies the possible non-degenerate limit distributions in Gnedenko (1943): Weibull ($\gamma < 0$), Gumbel ($\gamma = 0$) and Fréchet ($\gamma > 0$). The block maxima method is of a parametric nature: we work with the sample of maxima and estimate the parameters $(\lambda, \delta, \gamma)$ of the EV distribution, $G_\gamma((x-\lambda)/\delta)$, $\lambda \in \mathbb{R}$, $\delta > 0$, with $G_\gamma(x)$ given in (1.1). This method is known to be inefficient, due to the fact that the loss of information in each block can be catastrophic (the sample of the r maximum values in each of the r blocks does not necessarily contain the largest r observations in the sample). If we have access to all observations, and not only to block maxima, other methods may be more efficient.

In the second approach, the POT method, inference is performed through the use of the sample of exceedances from i.i.d. sequences over a high threshold u. The limit distribution of these exceedances is the Generalized Pareto (GP) distribution (Balkema and de Haan, 1974;

Pickands, 1975) defined by,

$$GP_\gamma(x) = \begin{cases} 1-(1+\gamma x)^{-1/\gamma}, & 1+\gamma x>0, \quad x>0 \qquad \text{if} \quad \gamma \neq 0 \\ 1-\exp(-x), & x>0 \qquad \text{if} \quad \gamma = 0. \end{cases} \tag{1.2}$$

This distribution unifies all possible non-degenerate limit distributions: Beta ($\gamma < 0$), Exponential ($\gamma = 0$) and Pareto ($\gamma > 0$). Note that the high threshold can also be a random value, leading to the peaks over random threshold (PORT) methodology, a terminology introduced in Araújo Santos *et al.* (2006).

The third approach is the one we shall consider in this paper. It uses the largest k observations to make inference about the right tail $\overline{F}$. For heavy-tailed models, i.e., models with a positive EVI, we assume that F has a Pareto-type tail, i.e., as $t \to \infty$ and with the notation $a(t) \sim b(t)$ if and only if $a(t)/b(t) \to 1$,

$$\overline{F}(x) = 1 - F(x) \ \sim \ (x/C)^{-1/\gamma} \quad \Longleftrightarrow \quad U(t) \sim C\, t^\gamma, \quad C,\ \gamma > 0, \tag{1.3}$$

where C is a scale parameter and $U(t) := F^{\leftarrow}(1-1/t)$, $t > 1$, with $F^{\leftarrow}(x) := \inf\{y : F(y) \geq x\}$ denoting the generalized inverse function of F. Note that models with a right Pareto-type tail have a regularly varying right tail with a negative EVI equal to $-1/\gamma$ and belong to the max-domain of attraction of G_γ, in (1.1), with $\gamma > 0$. More generally, we have, for all $x > 0$,

$$F \in \mathcal{D}_{\mathcal{M}}(G_{\gamma>0}) \quad \Leftrightarrow \quad \lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-1/\gamma} \quad \Leftrightarrow \quad \lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma.$$

To guarantee the consistency of many semi-parametric estimators we usually need to assume that k is intermediate, i.e., that k is a sequence of integers in $[1, n[$, such that

$$k = k_n \to \infty, \qquad k/n \to 0, \qquad n \to \infty. \tag{1.4}$$

To obtain information on the non-degenerate distributional behaviour of semi-parametric estimators of parameters of extreme events, we shall also assume a second order condition,

$$\lim_{t\to\infty} \frac{U(tx)/U(t) - x^\gamma}{A(t)} = x^\gamma\ \frac{x^\rho - 1}{\rho} \Longleftrightarrow \quad \lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}, \tag{1.5}$$

for all $x > 0$, where $\rho \leq 0$ is a second order parameter controlling the speed of convergence of $U(tx)/U(t)$ to x^γ, as $t \to \infty$. The validity of condition (1.5), with $\rho < 0$, implies the validity of condition (1.3).

1.1 Estimators under study

Under the largest observations framework, and whenever dealing with heavy-tailed models, the classical semi-parametric EVI and scale estimators are the Hill estimator (Hill, 1975) and Weissman's estimator (Weissman, 1978), with functional expressions

$$\hat{\gamma}^H_{n,k} := \frac{1}{k}\sum_{i=1}^{k}\left(\ln X_{n-i+1:n} - \ln X_{n-k:n}\right), \qquad k = 1, 2, \ldots, n-1, \tag{1.6}$$

and

$$\hat{C}^{W|H}_{n,k} = X_{n-k:n}\left(\frac{k}{n}\right)^{\hat{\gamma}^H_{n,k}}, \qquad k = 1, 2, \ldots, n-1, \tag{1.7}$$

respectively, which are pseudo-maximum likelihood estimators, consistent for k intermediate in the whole $\mathcal{D}_{\mathcal{M}}(G_\gamma)$, $\gamma > 0$. Under the second order framework in (1.5) and for intermediate k, we can guarantee, for an adequate k, the asymptotic normality of the estimators $\hat{\gamma}^H_{n,k}$ as well as $\hat{C}^{W|H}_{n,k}$ in (1.6) and (1.7), respectively.

Most of the times, this type of estimators exhibits a large variance for small k, a strong bias for moderate k and sample paths with very short stability regions around the target value. This has led researchers to search for alternative estimators with smaller mean square error.

Since heavy-tailed models only have mean value if $\gamma < 1$, methods based on sample moments are usually rarely considered when we work with such a type of distributions. But in many practical fields like in finance or insurance, for example, we usually have an EVI smaller than one, and even $\gamma < 1/2$. In this article we propose the use of a Probability Weighted Moments (PWM) method based on the largest observations for the estimation of tail parameters.

The PWM method is a generalization of the Method of Moments. It also consists in equating sample moments with their corresponding theoretical moments, and then solving those equations in order to obtain estimates of the different parameters under play. The PWM of a r.v. X are defined by

$$M_{p,r,s} := E(X^p (F(X))^r (1 - F(X))^s),$$

where p, r and s are any real numbers (Greenwood *et al.*, 1979). When $r = s = 0$, $M_{p,0,0}$ are the usual noncentral moments of order p. Hosking *et al.* (1985) advise the use of $M_{1,r,s}$, because then the relations between parameters and moments have usually a much simpler form. Also, when r and s are integers $F^r(1-F)^s$ can be written as a linear combination of powers of F or $1-F$. So it is usual to work with one of the two special cases:

$$a_r := M_{1,0,r} = E(X(1-F(X))^r) \text{ or } \qquad b_r := M_{1,r,0} = E(X(F(X))^r). \quad (1.8)$$

Given a sample of size n, the unbiased estimators of a_r and b_r in (1.8) are, respectively,

$$\hat{a}_r = \frac{1}{n}\sum_{i=1}^{n-r} \frac{(n-1-r)!(n-i)!}{(n-1)!(n-i-r)!} X_{i:n} = \frac{1}{n}\sum_{i=1}^{n} \frac{(n-i)(n-i-1)\dots(n-i-r+1)}{(n-1)(n-2)\dots(n-r)} X_{i:n}, \quad (1.9)$$

and

$$\hat{b}_r = \frac{1}{n}\sum_{i=r+1}^{n} \frac{(n-1-r)!(i-1)!}{(n-1)!(i-r-1)!} X_{i:n} \quad = \frac{1}{n}\sum_{i=1}^{n} \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} X_{i:n}.$$

For the Pareto d.f., $F(x) = 1-(x/C)^{-1/\gamma}, \quad x > C, \quad \gamma > 0$, the PWM are $M_{p,r,s} = C^p \mathcal{B}(s+1-\gamma\ p, r+1), \quad s-\gamma\ p > -1, \quad r > -1$, where $\mathcal{B}$ represents the complete beta function. In particular, $a_r = M_{1,0,r} = C/(r+1-\gamma), \quad \gamma < r+1$. Consequently, if we consider the theoretical moments, $a_0 = C/(1-\gamma)$, the mean of X, and $a_1 = C/(2-\gamma)$, with $\gamma < 1$, the Pareto PWM (PPWM) estimators of γ and C are, respectively,

$$\hat{\gamma}^{PPWM} = 1 - \left(\frac{\hat{a}_1}{\hat{a}_0 - \hat{a}_1}\right) \quad \text{and} \quad \hat{C}^{PPWM} = \hat{a}_0 \left(\frac{\hat{a}_1}{\hat{a}_0 - \hat{a}_1}\right), \quad \gamma < 1, \quad (1.10)$$

where $\hat{a}_0$ and $\hat{a}_1$ are given in (1.9).

We shall consider in this paper, the PPWM estimators for the parameters of a Pareto tail in (1.3), based on the top k largest o.s. If we work with the sample of the k largest observations, $X_{n-k+1:n} \le X_{n-k+2:n} \le \dots \le X_{n:n}$, the estimators $\hat{a}_0$ and $\hat{a}_1$ in (1.10) should be replaced by,

$$\hat{a}_0(k) := \frac{1}{k}\sum_{i=1}^{k} X_{n-i+1:n}, \qquad \text{and} \qquad \hat{a}_1(k) := \frac{1}{k}\sum_{i=1}^{k} \frac{i-1}{k-1} X_{n-i+1:n}, \quad (1.11)$$

respectively. The EVI and scale PPWM estimators, based on the largest values are:

$$\hat{\gamma}_{n,k}^{PPWM} = 1 - \frac{\hat{a}_1(k)}{\hat{a}_0(k) - \hat{a}_1(k)}, \qquad \hat{C}_{n,k}^{PPWM} = \frac{\hat{a}_0(k)\ \hat{a}_1(k)}{\hat{a}_0(k) - \hat{a}_1(k)} \left(\frac{k}{n}\right)^{\hat{\gamma}_{n,k}^{PPWM}}, \quad (1.12)$$

with $k = 1, 2, \cdots, n-1$ and $\gamma < 1$.

If we consider the GP distribution with scale parameter δ, i.e, $GP_\gamma(x/\delta) = 1-(1+\gamma x/\delta)^{-1/\gamma}$, $1+\gamma x/\delta > 0,\ x > 0,\ \delta > 0$ and $\gamma < 1$, with $GP_\gamma(x)$ defined in (1.2), the Generalized Pareto PWM (GPPWM) estimators of γ and δ are

$$\hat{\gamma}^{GPPWM} = 1 - \frac{2\hat{a}_1}{\hat{a}_0 - 2\hat{a}_1}, \qquad \text{and} \qquad \hat{\delta}^{GPPWM} = \frac{2\hat{a}_0\hat{a}_1}{\hat{a}_0 - 2\hat{a}_1}, \quad \gamma < 1, \tag{1.13}$$

respectively (Hosking and Wallis, 1987). Notice that the GP scale parameter δ is usually different from C. Since $1 - GP_\gamma(x/\delta) = (\gamma x/\delta)^{-1/\gamma}$, we have $\delta = C\gamma$. An obvious GPPWM estimator for the scale parameter C, is

$$\hat{C}^{GPPWM} = \frac{2\hat{a}_0\hat{a}_1}{\hat{a}_0 - 4\hat{a}_1}, \quad \gamma < 1.$$

De Haan and Ferreira (2006) considered, also for $\gamma < 1$, the previous GPPWM estimators of γ and δ based on the sample of exceedances over the high random level $X_{n-k:n}$. Since we are more interested in the estimation of the scale parameter C, we shall also consider the GPPWM estimators of γ and C based on the sample of exceedances over the high random level $X_{n-k:n}$, i.e.,

$$\hat{\gamma}_{n,k}^{GPPWM} = 1 - \frac{2\hat{a}_1^\star(k)}{\hat{a}_0^\star(k) - 2\hat{a}_1^\star(k)}, \text{and} \quad \hat{C}_{n,k}^{GPPWM} = \frac{2\hat{a}_0^\star(k)\,\hat{a}_1^\star(k)}{\hat{a}_0^\star(k) - 4\hat{a}_1^\star(k)} \left(\frac{k}{n}\right)^{\hat{\gamma}_{n,k}^{GPPWM}}, \tag{1.14}$$

with $k = 1, 2, \ldots, n-1$, $\gamma < 1$, and

$$\hat{a}_s^\star(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i-1}{k-1}\right)^s (X_{n-i+1:n} - X_{n-k:n}), \quad s = 0, 1.$$

We shall also study the GPPWM estimators in (1.14) and compare them with the PPWM estimators in (1.12), based on the largest observations.

Remark 1.1. *Note that all EVI estimators here referred are scale invariant. The two GPPWM estimators, in* (1.13) *and* (1.14), *are also shift invariant.*

1.2 Scope of the article

In Section 2, after the study of the asymptotic behaviour of two auxiliary statistics, we state a few results already proved in the literature and a generalization of Proposition 5 in Caeiro and Gomes (2009). Next, in Section 3, we derive the asymptotic properties of the PPWM and GPPWM-estimators of the first-order parameters and proceed to an asymptotic comparison at their optimal levels of the tail-index estimators under consideration. Finally, in Section 4, we perform a small-scale Monte-Carlo simulation, in order to compare the behaviour of the estimators under study for finite samples.

2 Preliminary Asymptotic Properties

2.1 Auxiliary statistics

To study the asymptotic properties of the EVI and scale PPWM and GPPWM estimators introduced in this paper, we first need to study the behaviour of the statistics,

$$\hat{q}_s(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i-1}{k-1}\right)^s \frac{X_{n-i+1:n}}{X_{n-k:n}}, \qquad s \geq 0, \tag{2.1}$$

and

$$\hat{q}^\star_s(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i-1}{k-1}\right)^s \left(\frac{X_{n-i+1:n}}{X_{n-k:n}} - 1\right), \qquad s \geq 0. \tag{2.2}$$

Proposition 2.1. *Under the second order framework in* (1.5)*, and for intermediate* k*, i.e., whenever* (1.4) *holds, we can guarantee the asymptotic normality of* $q_s(k)$*. Indeed, we can write, for* $s > \gamma - 1/2$*,*

$$\hat{q}_s(k) \stackrel{d}{=} \frac{1}{1+s-\gamma} + \frac{\sigma_s}{\sqrt{k}} W_k^{(s)} + \frac{A(n/k)(1+o_p(1))}{(1+s-\gamma)(1+s-\gamma-\rho)}, \tag{2.3}$$

and

$$\hat{q}^\star_s(k) \stackrel{d}{=} \frac{\gamma}{(1+s-\gamma)(1+s)} + \frac{\sigma_s}{\sqrt{k}} W_k^{(s)} + \frac{A(n/k)(1+o_p(1))}{(1+s-\gamma)(1+s-\gamma-\rho)}, \tag{2.4}$$

where $W_k^{(s)}$ *is an asymptotically standard normal r.v., and*

$$\sigma_s^2 := \frac{\gamma^2}{(1+s-\gamma)^2(1+2s-2\gamma)}. \tag{2.5}$$

Proof. Since $U(X_{i:n}) \stackrel{d}{=} Y_{i:n}$, where Y is a standard Pareto r.v. with d.f. $F_Y(y) = 1 - 1/y$, $y > 1$, $Y_{n-i+1:n}/Y_{n-k:n} \stackrel{d}{=} Y_{k-i+1:k}$, and under the second order framework in (1.5),

$$\frac{X_{n-i+1:n}}{X_{n-k:n}} \stackrel{d}{=} \frac{U(\frac{Y_{n-i+1:n}}{Y_{n-k:n}} Y_{n-k:n})}{U(Y_{n-k:n})} \stackrel{d}{=} (Y_{k-i+1:k})^\gamma \left(1 + \frac{Y_{k-i+1:k}^\rho - 1}{\rho} A(Y_{n-k:n})(1 + o_p(1))\right).$$

Consequently, since $\frac{i-1}{k-1} \sim \frac{i}{k}$,

$$\begin{aligned}\hat{q}_s(k) &\stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^s (Y_{k-i+1:k})^\gamma + \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^s (Y_{k-i+1:k})^\gamma \frac{(Y_{k-i+1:k})^\rho - 1}{\rho} A(n/k)(1 + o_p(1)) \\ &\stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}\left(1 - \frac{i}{k}\right)^s Y_{i:k}^\gamma + \frac{1}{k}\sum_{i=1}^{k}\left(1 - \frac{i}{k}\right)^s Y_{i:k}^\gamma \frac{Y_{i:k}^\rho - 1}{\rho} A(n/k)(1 + o_p(1)).\end{aligned}$$

Using the asymptotic results for linear functions of o.s. (David and Nagaraja, 2003), with the notation $L_k^{(s)} := \sum_{i=1}^{k} (1 - i/k)^s\, Y_{i:k}^\gamma / k$, we have

$$W_k^{(s)} := \sqrt{k}\, \frac{L_k^{(s)} - \mu_s}{\sigma_s} \underset{n\to\infty}{\overset{d}{\longrightarrow}} N(0,1), \qquad s > \gamma - \frac{1}{2}, \tag{2.6}$$

with

$$\mu_s := \int_0^1 (1-u)^{s-\gamma} du = \frac{1}{1+s-\gamma}, \quad s > \gamma - 1, \tag{2.7}$$

and

$$\sigma_s^2 = 2 \iint\limits_{0<u<v<1} u(1-u)^{s-\gamma-1}(1-v)^{s-\gamma} du dv, \quad s > \gamma - 1/2,$$

the value given in (2.5). The covariance, $Cov(\sigma_s W_k^{(s)}, \sigma_r W_k^{(r)})$, can also be easily computed and is equal to

$$k\; Cov(L_k^{(s)}, L_k^{(r)}) = \frac{g_{r,s} + g_{s,r}}{2} = \frac{\gamma^2}{(1+s+r-2\gamma)(1+s-\gamma)(1+r-\gamma)}, \tag{2.8}$$

where $g_{r,s} := 2\gamma^2 \iint\limits_{0<u<v<1} u(1-u)^{r-\gamma-1}(1-v)^{s-\gamma} du dv$ (Hosking *et al.*, 1985). Consequently, since $L_k^{(s)} \overset{\mathbb{P}}{\longrightarrow} \mu_s$, and

$$\frac{1}{k}\sum_{i=1}^{k}\left(1 - \frac{i}{k}\right)^s Y_{i:k}^\gamma\, \frac{Y_{i:k}^\rho - 1}{\rho} \overset{\mathbb{P}}{\longrightarrow} \frac{1}{(1+s-\gamma)(1+s-\gamma-\rho)},$$

equation (2.3) follows, with σ_s^2 given in (2.5).

Next, since

$$q_s^\star(k) = q_s(k) - \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i-1}{k-1}\right)^s \quad \text{and} \quad \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i-1}{k-1}\right)^s \sim \int_0^1 x^s dx = \frac{1}{1+s},$$

equation (2.4) follows straightforwardly. □

We still refer the following:

Proposition 2.2 (Caeiro and Gomes, 2009, Corollary 1)**.** *Under the conditions of* Proposition 2.1, but with $\rho < 0$,

$$\frac{X_{n-k:n}}{(n/k)^\gamma} \stackrel{d}{=} C\left(1 + \frac{\gamma}{\sqrt{k}}B_k + \frac{A(n/k)}{\rho} + o_p(A(n/k))\right),$$

with B_k an asymptotically standard normal r.v. and $Cov(B_r, B_s) = \sqrt{r\,s}(1 - s/n)/(s-1)$, $r < s$.

2.2 Asymptotic behaviour of the classical estimators

Proposition 2.3 (de Haan and Peng, 1998)**.** *Under the second-order framework in* (1.5)*, and for intermediate k, the asymptotic distributional representation of $\hat\gamma_{n,k}^H$, in* (1.6)*, is given by,*

$$\hat\gamma_{n,k}^H \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}Z_k + \frac{A(n/k)}{1-\rho}(1 + o_p(1)),$$

with $Z_k = \sqrt{k}\left(\sum_{i=1}^k E_i/k - 1\right)$, and $\{E_i\}$ i.i.d. standard exponential r.v.'s. Consequently, if we choose k such that $\sqrt{k}\,A(n/k) \to \lambda$, finite and not necessarily null, then,

$$\sqrt{k}(\hat\gamma_{n,k}^H - \gamma) \stackrel{d}{\longrightarrow} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right), \qquad \text{as } n \to \infty.$$

More generally than Proposition 5 in Caeiro and Gomes (2009), but with a similar proof, we now state the following proposition.

Proposition 2.4. *Under the conditions of* Proposition 2.3*, let $\hat\gamma_{n,k}^\bullet$ be any semi-parametric estimator of γ, such that, with $Z_k^\bullet$ asymptotically standard normal, $\sigma_\bullet > 0$ and $b_\bullet \in \mathbb{R}$,*

$$\hat\gamma_{n,k}^\bullet \stackrel{d}{=} \gamma + \frac{\sigma_\bullet Z_k^\bullet}{\sqrt{k}} + b_\bullet A(n/k)(1 + o_p(1)), \quad \text{as} \quad n \to \infty. \tag{2.9}$$

Then, with $\hat{C}^{W|\bullet}_{n,k} := X_{n-k:n}(k/n)^{\hat{\gamma}^{\bullet}_{n,k}}$, *and for* $\rho < 0$,

$$\hat{C}^{W|\bullet}_{n,k} \stackrel{d}{=} C\left\{1 + \ln\left(\tfrac{k}{n}\right)(\hat{\gamma}^{\bullet}_{n,k} - \gamma)(1 + o_p(1)) + \tfrac{\gamma}{\sqrt{k}}B_k + \tfrac{A(n/k)(1+o_p(1))}{\rho}\right\}.$$

If we further assume that $\sqrt{k}A(n/k) \to \lambda$, *as* $n \to \infty$,

$$\frac{\sqrt{k}}{\ln(k/n)}\left(\frac{\hat{C}^{W|\bullet}_{n,k}}{C} - 1\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} N\left(\lambda\ b_\bullet, \sigma^2_\bullet\right).$$

3 Asymptotic behaviour of the PPWM and GPPWM estimators

We can now study the asymptotic behaviour of the semi-parametric PPWM and GPPM estimators in (1.12) and (1.14), respectively.

3.1 Limiting behaviour at k

We next state the main result in this paper:

Theorem 3.1. *Under the second order framework in* (1.5), *with* $0 < \gamma < 1/2$, *and for intermediate* k *such that* $\sqrt{k}\,A(n/k) \to \lambda$, *finite, we can guarantee the asymptotic normality of* $\hat{\gamma}^{PPWM}_{n,k}$ *and* $\hat{\gamma}^{GPPWM}_{n,k}$, *in* (1.12) *and* (1.14), *respectively. Indeed, with* $\bullet$ *denoting either PPWM or GPPWM, the distributional representation in* (2.9) *holds, where*

$$\sigma^2_{PPWM} := \frac{\gamma^2(1-\gamma)(2-\gamma)^2}{(1-2\gamma)(3-2\gamma)}, \quad b_{PPWM} := \frac{(1-\gamma)(2-\gamma)}{(1-\gamma-\rho)(2-\gamma-\rho)},$$

and

$$\sigma^2_{GPPWM} := \frac{(1-\gamma+2\gamma^2)(1-\gamma)(2-\gamma)^2}{(1-2\gamma)(3-2\gamma)}, \quad b_{GPPWM} := \frac{(\gamma+\rho)\ b_{PPWM}}{\gamma}.$$

Moreover, with σ^2_s, $W^{(s)}_k$, *and* μ_s *given in* (2.5), (2.6), *and* (2.7), *respectively, and with* $\overline{\sigma}_s := \sigma_s/\mu_s$

$$Z^{PPWM}_k = \frac{(1-\gamma)(2-\gamma)(\overline{\sigma}_0 W^{(0)}_k - \overline{\sigma}_1 W^{(1)}_k)}{\sigma_{PPWM}}$$

and

$$Z_k^{GPPWM} = \frac{(1-\gamma)(2-\gamma)(\overline{\sigma}_0 W_k^{(0)} - 2\overline{\sigma}_1 W_k^{(1)})}{\gamma\ \sigma_{PPWM}}$$

are asymptotically standard Norma r.v.'s. Consequently,

$$\sqrt{k}\Big(\hat{\gamma}^{\bullet}_{n,k} - \gamma\Big) \underset{n\to\infty}{\overset{d}{\longrightarrow}} N\left(\lambda\ b_{\bullet}, \sigma^2_{\bullet}\right) \tag{3.1}$$

and, for $\rho < 0$,

$$\frac{\sqrt{k}}{\ln(k/n)}\left(\frac{\hat{C}^{\bullet}_{n,k}}{C} - 1\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} N\left(\lambda\ b_{\bullet}, \sigma_{\bullet}{}^2\right). \tag{3.2}$$

Proof. We can write,

$$\hat{\gamma}_{n,k}^{PPWM} = 1 - \left(\frac{\hat{q}_0(k)}{\hat{q}_1(k)} - 1\right)^{-1} \quad \text{and} \quad \hat{\gamma}_{n,k}^{GPPWM} = 1 - \left(\frac{\hat{q}^{\star}_0(k)}{2\hat{q}^{\star}_1(k)} - 1\right)^{-1},$$

with $\hat{q}_s(k)$ and $\hat{q}^{\star}_s(k)$, $s = 0, 1$, defined in (2.1) and (2.2), respectively. Using (2.3) with $s = 1$ and Taylor's expansion $(1+x)^{-1} = 1 - x + o(x)$, as $x \to 0$, we get

$$\hat{q}_1(k)^{-1} \overset{d}{=} (2-\gamma)\left\{1 - \frac{\overline{\sigma}_1}{\sqrt{k}} W_k^{(1)} - \frac{A(n/k)}{2-\gamma-\rho}(1 + o_p(1))\right\}.$$

Consequently, using again the previous Taylor expansion,

$$\left(\frac{\hat{q}_0(k)}{\hat{q}_1(k)} - 1\right)^{-1} \overset{d}{=} (1-\gamma)\Big\{1 - \frac{(2-\gamma)}{\sqrt{k}}\Big(\overline{\sigma}_0 W_k^{(0)} - \overline{\sigma}_1 W_k^{(1)}\Big) - \frac{(2-\gamma)A(n/k)}{(1-\gamma-\rho)(2-\gamma-\rho)}(1+o_p(1))\Big\}, \tag{3.3}$$

and (2.9) as well as (3.1) follow easily for $\bullet$ = PPWM.

Analogously, we have

$$(2\hat{q}^{\star}_1(k))^{-1} \overset{d}{=} \frac{(2-\gamma)}{\gamma}\left\{1 - \frac{2\ \overline{\sigma}_1}{\sqrt{k}\ \gamma} W_k^{(1)} - \frac{2A(n/k)}{\gamma(2-\gamma-\rho)}(1 + o_p(1))\right\}.$$

Consequently,

$$\left(\frac{\hat{q}^{\star}_0(k)}{2\hat{q}^{\star}_1(k)} - 1\right)^{-1} \overset{d}{=} (1-\gamma)\Big\{1 - \frac{(2-\gamma)}{\sqrt{k}\ \gamma}\Big(\overline{\sigma}_0 W_k^{(0)} - 2\overline{\sigma}_1 W_k^{(1)}\Big) - \frac{(2-\gamma)(\gamma+\rho)A(n/k)}{\gamma(1-\gamma-\rho)(2-\gamma-\rho)}(1 + o_p(1))\Big\},$$

and (2.9) as well as (3.1) follow easily, now for $\bullet =$ GPPWM.

The asymptotic normality of Z_k^{PPWM} and Z_k^{GPPWM} follows from (2.6) and (2.8).

For the scale parameter estimator, we can easily derive that, both for the PPWM and the GPPWM scale-estimators

$$\hat{C}_{n,k}^{\bullet} = \hat{C}_{n,k}^{W|\bullet} \left(1 + o_p\left(\frac{\ln(k/n)}{\sqrt{k}}\right)\right),$$

with $\hat{C}_{n,k}^{W|H}$ defined in (1.7).

Then, using the results in Proposition 2.4, (3.2) follows. □

Remark 3.1. *In* Figure 1*, we provide a picture of the asymptotic variances of the EVI estimators given in* Propositions 2.3 and 3.1*.*

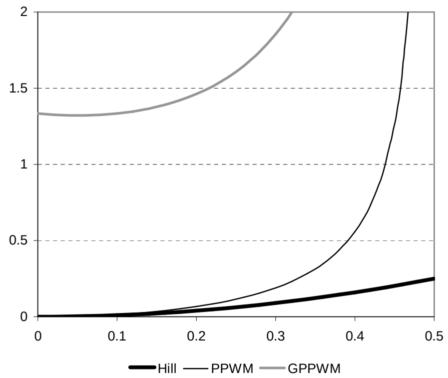

Figure 1: Asymptotic variances of $\sqrt{k}(\hat{\gamma}_{n,k}^{H} - \gamma)$, $\sqrt{k}(\hat{\gamma}_{n,k}^{PPWM} - \gamma)$ and $\sqrt{k}(\hat{\gamma}_{n,k}^{GPPWM} - \gamma)$.

It is obvious that $\sigma_H^2 := \gamma^2 < \sigma_{PPWM}^2 < \sigma_{GPPWM}^2$*, for every* $0 < \gamma < 1/2$*.*

Remark 3.2. *For levels such that* $\sqrt{k}A(n/k) \to \lambda$ *and for* $\gamma + \rho = 0$*, both* $\hat{\gamma}_{n,k}^{GPPWM}$ *and* $\hat{C}_{n,k}^{GPPWM}$ *are asymptotically unbiased.*

Remark 3.3. *The asymptotic dominant behaviour of* $\hat{C}_{n,k}^{\bullet}$ *is determined by the asymptotic behaviour of* $\hat{\gamma}_{n,k}^{\bullet}$*.*

3.2 Asymptotic comparison of the EVI estimators at optimal levels

We now proceed to an asymptotic comparison of the EVI estimators at their optimal levels in the lines of de Haan and Peng (1998), Gomes and Martins (2001), Gomes *et al.* (2005, 2007) and Gomes and Neves (2008). Similar results holds for the scale estimators, at their optimal levels, since they have the same asymptotic behaviour as the EVI estimators, although with a slower convergence rate. For details see Proposition 2.4 and Theorem 3.1. Suppose that $\widehat{\gamma}^{\bullet}_{n,k}$ is a general semi-parametric estimator of the tail index γ, such that the distributional representation (2.9) holds. Then we have,

$$\sqrt{k}(\widehat{\gamma}^{\bullet}_{n,k} - \gamma) \ \stackrel{d}{\longrightarrow}\ N(\lambda b_{\bullet}, \sigma^2_{\bullet}), \text{ as } n \to \infty,$$

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

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

where $Bias_{\infty}(\widehat{\gamma}^{\bullet}_{n,k}) := b_{\bullet} A(n/k)$ and $Var_{\infty}(\widehat{\gamma}^{\bullet}_{n,k}) := \sigma^2_{\bullet}/k$.

Let $k^{\bullet}_0 \equiv k^{\bullet}_0(n)$:=$\arg\min_k AMSE(\widehat{\gamma}^{\bullet}_{n,k})$ be the so-called optimal level for the estimation of γ through $\widehat{\gamma}^{\bullet}_{n,k}$, i.e., the level associated to a minimum asymptotic mean square error, and let us denote $\widehat{\gamma}^{\bullet}_{n0} := \widehat{\gamma}^{\bullet}_{n,k^{\bullet}_0}$, 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

$$\lim_{n\to\infty} \varphi(n) AMSE(\widehat{\gamma}^{\bullet}_{n0}) = \frac{2\rho - 1}{\rho} \left(\sigma^2_{\bullet}\right)^{-\frac{2\rho}{1-2\rho}} \left(b^2_{\bullet}\right)^{\frac{1}{1-2\rho}} =: LMSE(\widehat{\gamma}^{\bullet}_{n0}). \tag{3.4}$$

It is then sensible to consider the following:

Definition 3.1. *Given two biased estimators* $\widehat{\gamma}^{(1)}_{n,k}$ *and* $\widehat{\gamma}^{(2)}_{n,k}$, *for which distributional representations of the type* (2.9) *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,* $k^{(1)}_0$ *and* $k^{(2)}_0$, *the Asymptotic Root Efficiency* ($AREFF$) *indicator is defined as*

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

with LMSE given in (3.4), $\widehat{\gamma}^{(1)}_{n0} = \widehat{\gamma}^{\bullet}_{n,k^{(1)}_0}$ *and* $\widehat{\gamma}^{(2)}_{n0} = \widehat{\gamma}^{\bullet}_{n,k^{(2)}_0}$.

Remark 3.4. *Note that this measure was devised so that the higher the AREFF indicator is, the better the first estimator is.*

The indicator $AREFF_{PPWM|GPPWM}$ is presented in Figure 1, and in Figure 2 we provide an indication of the best estimator at optimal levels, with H, P and G standing for the Hill, the PPWM and the GPPWM estimators, respectively. For $\rho = 0$, the estimators H, P and G are asymptotically equivalent at optimal levels.

ρ \ γ	0.05	0.07	0.09	0.11	0.13	0.15	0.17	0.19	0.21	0.23	0.25	0.27	0.29	0.31	0.33	0.35	0.37	0.39	0.41	0.43	0.45
0.00	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
-0.05	0.0	0.4	0.6	0.7	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	1.0	1.0	1.0	1.0	1.0	1.0
-0.07	0.6	0.0	0.4	0.5	0.6	0.7	0.8	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
-0.09	1.3	0.5	0.0	0.3	0.5	0.6	0.7	0.7	0.8	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
-0.11	2.0	1.0	0.4	0.0	0.3	0.5	0.6	0.7	0.7	0.8	0.8	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
-0.13	2.7	1.5	0.9	0.4	0.0	0.3	0.5	0.6	0.6	0.7	0.7	0.8	0.8	0.8	0.8	0.8	0.9	0.9	0.9	0.9	0.9
-0.15	3.4	2.0	1.3	0.8	0.4	0.0	0.3	0.4	0.5	0.6	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.8	0.9	0.9	0.9
-0.17	4.1	2.5	1.7	1.1	0.7	0.4	0.0	0.3	0.4	0.5	0.6	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.8	0.8	0.9
-0.19	4.8	3.0	2.1	1.4	1.0	0.6	0.3	0.0	0.3	0.4	0.5	0.6	0.6	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.8
-0.21	5.5	3.5	2.5	1.8	1.3	0.9	0.6	0.3	0.0	0.3	0.4	0.5	0.6	0.6	0.7	0.7	0.7	0.8	0.8	0.8	0.8
-0.23	6.1	4.0	2.9	2.1	1.6	1.2	0.8	0.6	0.3	0.0	0.3	0.4	0.5	0.6	0.6	0.7	0.7	0.7	0.7	0.8	0.8
-0.25	6.8	4.5	3.2	2.4	1.8	1.4	1.1	0.8	0.5	0.3	0.0	0.3	0.4	0.5	0.6	0.6	0.6	0.7	0.7	0.7	0.8
-0.27	7.4	5.0	3.6	2.7	2.1	1.7	1.3	1.0	0.7	0.5	0.3	0.0	0.3	0.4	0.5	0.5	0.6	0.6	0.7	0.7	0.7
-0.29	8.0	5.4	4.0	3.0	2.4	1.9	1.5	1.2	0.9	0.7	0.5	0.3	0.0	0.3	0.4	0.5	0.5	0.6	0.6	0.7	0.7
-0.31	8.6	5.9	4.3	3.3	2.6	2.1	1.7	1.4	1.1	0.9	0.7	0.5	0.3	0.0	0.3	0.4	0.5	0.5	0.6	0.6	0.7
-0.33	9.2	6.3	4.6	3.6	2.9	2.3	1.9	1.6	1.3	1.1	0.9	0.7	0.5	0.3	0.0	0.3	0.4	0.5	0.5	0.6	0.6
-0.35	9.8	6.7	5.0	3.9	3.1	2.5	2.1	1.7	1.5	1.2	1.0	0.8	0.6	0.5	0.3	0.0	0.3	0.4	0.5	0.5	0.6
-0.37	10.3	7.1	5.3	4.1	3.3	2.7	2.3	1.9	1.6	1.4	1.1	1.0	0.8	0.6	0.5	0.3	0.0	0.3	0.4	0.5	0.5
-0.39	10.8	7.4	5.6	4.4	3.5	2.9	2.4	2.1	1.8	1.5	1.3	1.1	0.9	0.8	0.6	0.5	0.3	0.0	0.3	0.4	0.5
-0.41	11.3	7.8	5.8	4.6	3.7	3.1	2.6	2.2	1.9	1.6	1.4	1.2	1.0	0.9	0.7	0.6	0.5	0.3	0.0	0.3	0.4
-0.43	11.8	8.1	6.1	4.8	3.9	3.3	2.8	2.4	2.0	1.8	1.5	1.3	1.2	1.0	0.9	0.7	0.6	0.4	0.3	0.0	0.3
-0.45	12.2	8.5	6.4	5.0	4.1	3.4	2.9	2.5	2.2	1.9	1.7	1.5	1.3	1.1	1.0	0.8	0.7	0.6	0.4	0.3	0.0

$AREFF_{PPWM|GPPWM} < 1$ $1 \leq AREFF_{PPWM|GPWWM} < 2$ $AREFF_{PPWM|GPPWM} \geq 2$

Figure 2: Asymptotic relative efficiency of $\widehat{\gamma}_{n0}^{PPWM}$ relatively to $\widehat{\gamma}_{n0}^{GPPWM}$.

As can be seen, the gain in efficiency for the PPWM-estimator (comparatively with the GPPWM-estimator) happens for a large region of values of (γ, ρ) such that $\gamma + \rho < 0$. In the region $\gamma + \rho = 0$, the GPPWM estimator is a second-order reduced-bias tail index estimator and consequently outperforms the PPWM estimator. These results claim now, not for a semi-parametric test of the hypothesis $H_0 : \eta = \gamma + \rho = 0$ as in Gomes and Henriques-Rodrigues (2009), but for a semi-parametric test of the hypothesis $H_0 : \eta = \gamma + \rho < 0$, a topic out of the scope of this paper.

ρ \ γ	0.05	0.07	0.09	0.11	0.13	0.15	0.17	0.19	0.21	0.23	0.25	0.27	0.29	0.31	0.33	0.35	0.37	0.39	0.41	0.43	0.45
0.00	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
-0.05	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
-0.07	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
-0.09	P	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
-0.11	P	P	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
-0.13	P	P	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
-0.15	P	P	P	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
-0.17	P	P	P	P	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
-0.19	P	P	P	P	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
-0.21	P	P	P	P	P	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
-0.23	P	P	P	P	P	P	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
-0.25	P	P	P	P	P	P	P	G	G	G	G	G	G	G	G	G	G	G	G	G	G
-0.27	P	P	P	P	P	P	P	G	G	G	G	G	G	G	G	G	G	G	G	G	G
-0.29	P	P	P	P	P	P	P	P	G	G	G	G	G	G	G	G	G	G	G	G	G
-0.31	P	P	P	P	P	P	P	P	P	G	G	G	G	G	G	G	G	G	G	G	G
-0.33	P	P	P	P	P	P	P	P	P	P	G	G	G	G	G	G	G	G	G	G	G
-0.35	P	P	P	P	P	P	P	P	P	P	P	G	G	G	G	G	G	G	G	G	G
-0.37	P	P	P	P	P	P	P	P	P	P	P	G	G	G	G	G	G	G	G	G	G
-0.39	P	P	P	P	P	P	P	P	P	P	P	P	G	G	G	G	G	G	G	G	G
-0.41	P	P	P	P	P	P	P	P	P	P	P	P	P	G	G	G	G	G	G	G	G
-0.43	P	P	P	P	P	P	P	P	P	P	P	P	P	P	G	G	G	G	G	G	G
-0.45	P	P	P	P	P	P	P	P	P	P	P	P	P	P	G	G	G	G	G	G	G

Figure 3: Best estimator at optimal levels.

4 Simulated behaviour of the EVI estimators

In this section, we have implemented a multi-sample Monte Carlo simulation experiment of size 5000×10, to obtain the distributional behaviour of the EVI estimators $\hat{\gamma}^{H}_{n,k}$, $\hat{\gamma}^{PPWM}_{n,k}$ and $\hat{\gamma}^{GPPWM}_{n,k}$ in (1.6), (1.12) and (1.14), respectively, for the following underlying parents: Student's t with $\nu = 4$ degrees of freedom ($\gamma = 0.25$, $\rho = -0.5$, $C = \sqrt[4]{3}$), the Fréchet parent, with d.f. $F(x) = \exp(-x^{-1/\gamma})$, $x > 0$, $\gamma > 0$ with $\gamma = 0.25$ ($\rho = -1$, $C = 1$) and the Burr parent, with d.f. $F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}$, $x > 0$, with $(\gamma, \rho) \in \{(0.25, -0.2), (0.25, -0.75), (0.25, -1.5), (0.5, -0.5), (0.75, -1.5)\}$ ($C = 1$). Notice that the Burr model with $\rho = -\gamma$ is the GP_γ distribution, in (1.2).

To ilustraste the finite sample behaviour of the EVI estimators, we present, in Figures 4, 5, 6, 7, 8, 9 and 10, the simulated mean values (E) and root mean square errors (RMSE) patterns of $\hat{\gamma}^{H}_{n,k}$, $\hat{\gamma}^{PPWM}_{n,k}$ and $\hat{\gamma}^{GPPWM}_{n,k}$, as functions of k, the number of top o.s. used, for a sample size $n = 1000$.

In Table 1 we present the simulated mean values of the above mentioned EVI estimators, at their simulated optimal levels k_0. Again, with $\bullet$ denoting either PPWM or GPPWM, Table

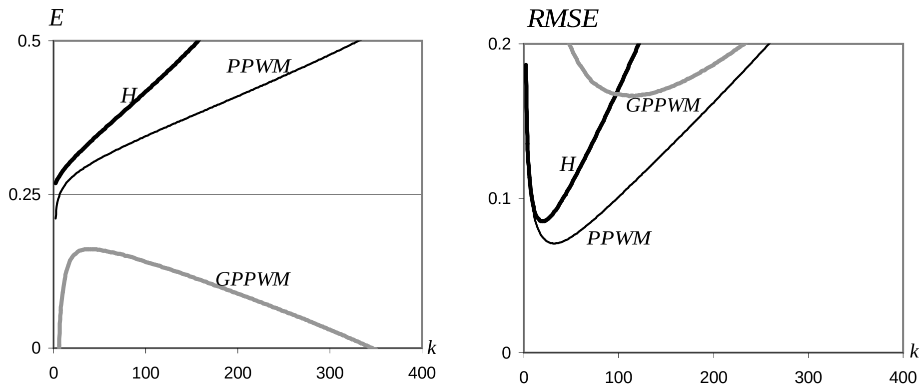

Figure 4: Student's t parent with $\nu = 4$ degrees of freedom, $(\gamma, \rho) = (0.25, -0.5)$.

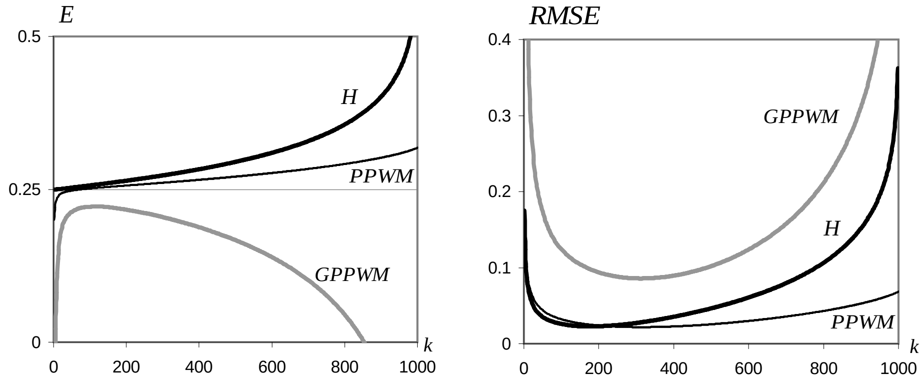

Figure 5: Fréchet(0.25) parent with $(\gamma, \rho) = (0.25, -1)$.

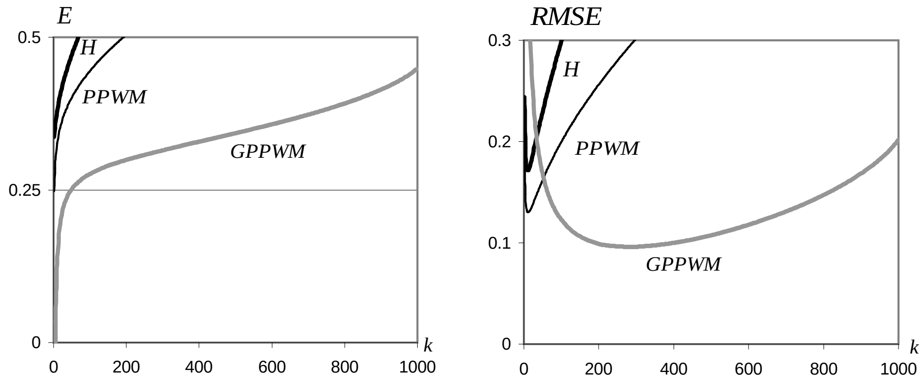

Figure 6: Burr parent with $(\gamma, \rho) = (0.25, -0.2)$.

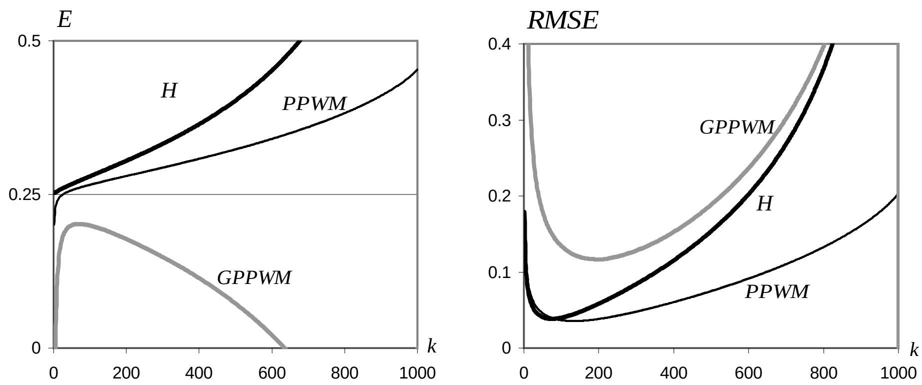

Figure 7: Burr parent with $(\gamma, \rho) = (0.25, -0.75)$.

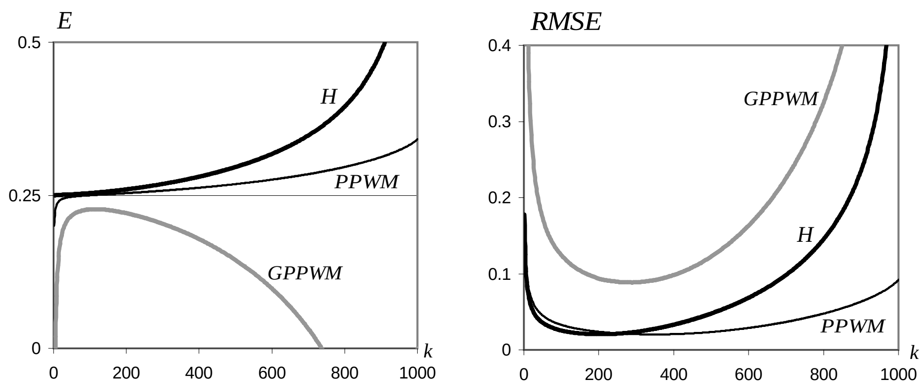

Figure 8: Burr parent with $(\gamma, \rho) = (0.25, -1.5)$.

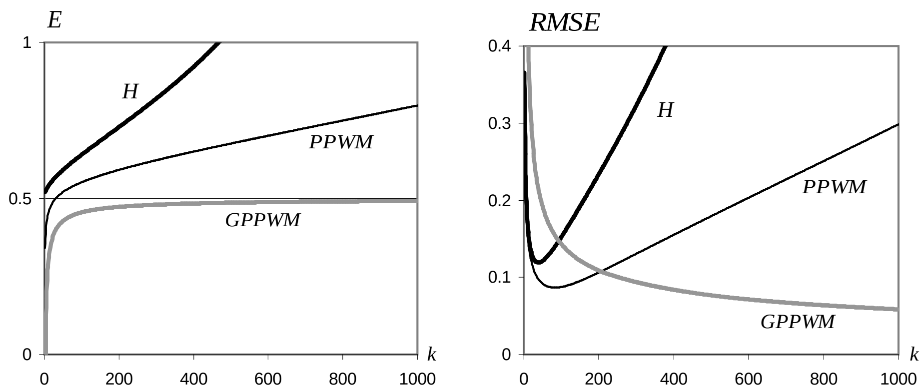

Figure 9: Burr parent with $(\gamma, \rho) = (0.5, -0.5)$.

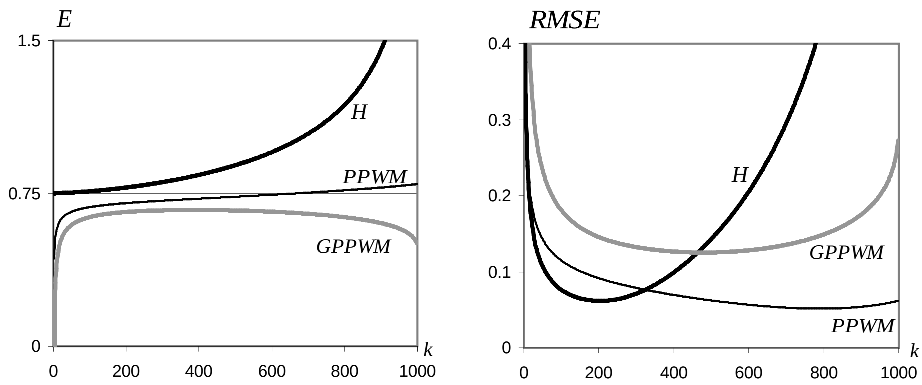

Figure 10: Burr parent with $(\gamma, \rho) = (0.75, -1.5)$.

2 has the simulated relative efficiencies of $\hat{\gamma}^{\bullet}_{n,k}$ comparatively with the Hill estimator, whenever computed at their simulated optimal levels, i.e., the simulated values of

$$REFF_{\bullet|H} = \frac{RMSE\left[\hat{\gamma}^{H}_{n,k_0^H}\right]}{RMSE\left[\hat{\gamma}^{\bullet}_{n,k_0^{\bullet}}\right]}.$$

Some remarks:

1. For Burr models with $\gamma + \rho \geq 0$, or equivalently, GP_γ models, $\hat{\gamma}^{GPPWM}_{n,k}$ is the best EVI estimator at its optimal level, unless n is small. In all other cases it shows a quite bad performance, essentialy due its high variance.

2. The EVI estimator $\hat{\gamma}^{PPWM}_{n,k}$ can be used as an alternative to the Hill estimator, specially for small to moderate sample sizes, but also for large samples. All the figures suggest that its RMSE is never much bigger than Hill's estimator RMSE and it has smaller bias than the Hill estimator. At their optimal levels and for small sample sizes, $\hat{\gamma}^{PPWM}_{n,k}$ is usually much more efficient than Hill's estimator.

3. The asymptotic properties of new extreme value index estimators do not hold for the Burr model with $(\gamma, \rho) = (0.75, -1.5)$. But even in this example, the PPWM EVI estimator is more efficient than the Hill estimator at their simulated optimal levels, for small to moderate sample sizes.

Table 1: Simulated mean values of EVI estimators under study, at their simulated optimal levels

n	50	100	200	500	1000	2000	5000	10000	20000
					Student-t_4				
H	0.388	0.361	0.338	0.316	0.305	0.296	0.286	0.279	0.275
$PPWM$	**0.325**	**0.308**	**0.304**	**0.297**	**0.292**	**0.287**	**0.280**	**0.276**	**0.272**
$GPPWM$	-0.077	0.007	0.059	0.110	0.136	0.155	0.179	0.192	0.200
					Fréchet(0.25)				
H	0.283	0.275	0.270	0.265	0.262	0.260	0.257	0.256	0.255
$PPWM$	**0.274**	**0.270**	**0.267**	**0.264**	**0.261**	**0.259**	**0.257**	**0.256**	**0.255**
$GPPWM$	0.089	0.132	0.160	0.187	0.203	0.213	0.223	0.229	0.234
					Burr(0.25, -0.2)				
H	0.511	0.467	0.434	0.401	0.381	0.364	0.344	0.334	0.325
$PPWM$	0.356	0.350	0.348	0.337	0.333	0.327	0.320	0.316	0.310
$GPPWM$	**0.353**	**0.342**	**0.333**	**0.320**	**0.313**	**0.304**	**0.296**	**0.291**	**0.287**
					Burr(0.25, -0.75)				
H	0.310	0.298	0.287	0.279	0.274	0.268	0.264	0.261	0.259
$PPWM$	**0.292**	**0.286**	**0.281**	**0.274**	**0.271**	**0.267**	**0.263**	**0.261**	**0.259**
$GPPWM$	0.032	0.087	0.125	0.159	0.178	0.193	0.208	0.216	0.222
					Burr(0.25, -1.5)				
H	0.280	0.272	0.268	0.263	0.260	0.258	0.255	0.254	0.253
$PPWM$	**0.273**	**0.269**	**0.266**	**0.262**	**0.260**	**0.258**	**0.255**	**0.254**	**0.253**
$GPPWM$	0.085	0.131	0.164	0.193	0.207	0.219	0.228	0.233	0.237
					Burr(0.5, -0.5)				
H	0.686	0.650	0.622	0.594	0.579	0.564	0.550	0.542	0.535
$PPWM$	**0.558**	0.553	0.550	0.546	0.543	0.540	0.535	0.532	0.528
$GPPWM$	0.428	**0.456**	**0.474**	**0.486**	**0.492**	**0.495**	**0.498**	**0.499**	**0.499**
					Burr(0.75, -1.5)				
H	0.839	0.816	0.805	0.789	0.779	0.773	0.766	0.762	0.760
$PPWM$	**0.749**	**0.765**	**0.769**	**0.766**	**0.764**	**0.762**	**0.761**	**0.759**	**0.758**
$GPPWM$	1.000	1.000	0.605	0.645	0.667	0.684	0.700	0.709	0.717

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.

Table 2: Relative efficiencies

	50	100	200	500	1000	2000	5000	10000	20000
					Student-t_4				
$PPWM\|H$	**1.631**	**1.478**	**1.363**	**1.260**	**1.201**	**1.161**	**1.121**	**1.100**	**1.083**
$GPPWM\|H$	0.506	0.508	0.507	0.511	0.512	0.518	0.520	0.521	0.521
					Fréchet(0.25)				
$PPWM\|H$	**1.188**	**1.144**	**1.108**	**1.067**	**1.041**	**1.023**	**1.013**	**1.009**	**1.006**
$GPPWM\|H$	0.237	0.244	0.253	0.261	0.265	0.270	0.275	0.277	0.279
					Burr(0.25, −0.2)				
$PPWM\|H$	**2.247**	1.992	1.791	1.594	1.469	1.370	1.272	1.209	1.159
$GPPWM\|H$	2.222	**2.069**	**1.951**	**1.836**	**1.761**	**1.710**	**1.675**	**1.640**	**1.621**
					Burr(0.25, −0.75)				
$PPWM\|H$	**1.282**	**1.206**	**1.148**	**1.104**	**1.080**	**1.060**	**1.048**	**1.040**	**1.035**
$GPPWM\|H$	0.299	0.309	0.316	0.323	0.330	0.333	0.334	0.338	0.339
					Burr(0.25, −1.5)				
$PPWM\|H$	**1.160**	**1.112**	**1.064**	**1.027**	**1.011**	0.997	0.988	0.984	0.980
$GPPWM\|H$	0.211	0.219	0.224	0.230	0.232	0.234	0.237	0.238	0.239
					Burr(0.5, −0.5)				
$PPWM\|H$	**2.154**	**1.898**	1.699	1.493	1.370	1.272	1.173	1.110	1.057
$GPPWM\|H$	1.589	1.688	**1.786**	**1.914**	**2.033**	**2.174**	**2.426**	**2.644**	**2.917**
					Burr(0.75, −1.5)				
$PPWM\|H$	**2.466**	**2.163**	**1.789**	**1.418**	**1.195**	**1.011**	0.819	0.694	0.593
$GPPWM\|H$	0.803	0.601	0.561	0.525	0.492	0.452	0.402	0.364	0.328

[2] Balkema, A.A., and Haan, L. de (1974). Residual life time at great age. *Annals of Probability* **2**, 792-804.

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

[4] Caeiro, F., Gomes, M. I. and Pestana, D. (2005). Direct Reduction of Bias of the Classical Hill Estimator. *Revstat* **3**(2), 113-136.

[5] Caeiro, F. and Gomes, M.I. (2009). Semi-parametric second-order reduced-bias high quantile estimation. *Test*, **18**(2), 392-413.

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

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

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

[9] Gomes, M.I., and Henriques Rodrigues, L. (2008). Tail index estimation for heavy tails: accommodation of bias in the excesses over a high threshold. *Extremes*, **11**(3), 303-328.

[10] Gomes, M.I., and Martins, M.J. (2001). Generalizations of the Hill estimator - asymptotic versus finite sample behaviour. *J. Statist. Planning and Inference* 93, 161-180.

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

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

[13] Gomes, M.I. and Neves, C. (2008). Asymptotic comparison of the mixed moment and classical extreme value index estimators, *Statistics and Probability Letters*, **78**, 6, 643-653.

[14] Greenwood, J. A,, Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form. *Water Resources Research*, 15, 1049-1054.

[15] de Haan, L. and Peng, L. (1998). Comparison of Tail Index Estimators. *Statistica Neerlandica*, 52, 60-70.

[16] de Haan, L. and Ferreira, A., (2006). Extreme Value Theory: an Introduction. Springer Science+Business Media, LLC, New York.

[17] Hall, P. (1982). On some Simple Estimates of an Exponent of Regular Variation. *J. R. Statist. Soc.* **44**(1), 37-42.

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

[19] Hill, B.M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *Ann. Statist.* **3**(5), 1163-1174.

[20] Hosking, J, Wallis, J, and Wood E. (1985). Estimation of the Generalized Extreme Value Distribution by the Method of Probability-Weighted Moments. Technometrics, 27(3), 251-261.

[21] Hosking, J. and Wallis, J. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, 29(3), 339-349.

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

[23] Weissman, I. (1978). Estimation of parameters of large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73**, 812-815.