PORT-PPWM extreme value index estimation

Ivette Gomes, *Universidade de Lisboa*, ivette.gomes@fc.ul.pt
Frederico Caeiro, *Universidade Nova de Lisboa*, fac@fct.unl.pt
Lígia Henriques-Rodrigues, *Instituto Politécnico de Tomar*, Ligia.Rodrigues@aim.estt.ipt.pt

Abstract. Making use of the peaks over random threshold (PORT) methodology and the Pareto probability weighted moments (PPWM) of the largest observations, and moreover dealing with the extreme value index (EVI), the primary parameter in statistics of extremes, new classes of location-invariant EVI-estimators are built. These estimators, the so-called PORT-PPWM EVI-estimators, are compared with the generalised Pareto probability weighted moments (GPPWM) and a recent class of minimum-variance reduced-bias (MVRB) EVI-estimators, for finite samples, through a Monte-Carlo simulation study.

Keywords. Heavy tails, Statistics of extremes, Extreme value index, Location/scale invariant estimation.

1 Introduction and preliminaries

The primary parameter in *statistics of univariate extremes* is the *extreme value index* (EVI), the 'shape' parameter $\gamma \in \mathbb{R}$ in the general *extreme value* distribution function (d.f.), with the functional form

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

Let $\underline{X}_n := (X_1, \ldots, X_n)$ denote a sample of size n from either independent, identically distributed or even stationary weakly dependent random variables from an underlying model F, and let us use the notation $(X_{1:n} \leq \cdots \leq X_{n:n})$ for the associated sample of ascending order statistics (o.s.'s). The EV_γ d.f. appears as the limiting d.f., whenever such a non-degenerate limit exists, of the maximum $X_{n:n}$, suitably linearly normalised. When such a non-degenerate limit exists, we say that F is in the *domain of attraction* for maximum values of the general EV_γ d.f., in (1), and use the notation $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$.

We shall deal with heavy-tails, quite common in the most diverse fields, like finance, insurance and telecommunications, i.e. with a positive EVI. Then, as first proved in [11], the right-tail function is of regular variation with an index of regular variation equal to $-1/\gamma$, i.e.

$$F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)_{\gamma>0} \quad \Longleftrightarrow \quad \overline{F} := 1 - F \in RV_{-1/\gamma}, \tag{2}$$

where the notation RV_a stands for the class of *regularly varying* functions at infinity with an *index of regular variation* equal to $a \in \mathbb{R}$, i.e. positive measurable functions g such that $\lim_{t\to\infty} g(tx)/g(t) = x^a$, for all $x > 0$.

The first class of semi-parametric estimators of a positive EVI was considered in [23]. These estimators are based on the log-excesses over an o.s., $X_{n-k:n}$, and have the functional form

$$\widehat{\gamma}^H_{k,n} \equiv \widehat{\gamma}^H_{k,n}(\underline{X}_n) := \frac{1}{k}\sum_{i=1}^{k} \left\{\ln X_{n-i+1:n} - \ln X_{n-k:n}\right\}, \qquad k = 1, \ldots, n-1. \tag{3}$$

Apart from the Hill estimator, in (3), we shall also consider the *Pareto probability weighted moments* (PPWM) EVI-estimators, recently introduced in [4]. They are valid for heavy right-tails with $\gamma < 1$, a quite relevant region in the field of extremes, compare favourably with the Hill estimator, in (3), for a wide variety of underlying models F, and are given by

$$\widehat{\gamma}^{PPWM}_{k,n} \equiv \widehat{\gamma}^{PPWM}_{k,n}(\underline{X}_n) := 1 - \frac{\hat{a}_1(k;\underline{X}_n)}{\hat{a}_0(k;\underline{X}_n) - \hat{a}_1(k;\underline{X}_n)}, \tag{4}$$

with

$$\hat{a}_r(k;\underline{X}_n) \equiv \hat{a}_r(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^r X_{n-i+1:n}, \qquad r = 0, 1.$$

Both classes of estimators, in (3) and (4), are scale-invariant, but not location-invariant, as often desired, and this contrarily to the PORT-Hill estimators, introduced in [1] and further studied in [17], with PORT standing for *peaks over random thresholds.* The class of PORT-Hill estimators is based on a *sample of excesses* over a random threshold $X_{n_q:n}$, $n_q := [nq] + 1$, where $[x]$ denotes, as usual, the integer part of x, i.e. it is based on

$$\underline{X}^{(q)}_n := \big(X_{n:n} - X_{n_q:n}, \ldots, X_{n_q+1:n} - X_{n_q:n}\big). \tag{5}$$

We can have $0 < q < 1$, for d.f.'s with finite or infinite left endpoint $x_F := \inf\{x : F(x) > 0\}$ (*the random threshold is then any empirical quantile*), and $0 \le q < 1$, for d.f.'s with a finite left endpoint x_F (*the random threshold can also be the minimum*). The PORT-Hill EVI-estimators are thus given by

$$\widehat{\gamma}^{H|q}_{k,n} := \widehat{\gamma}^H_{k,n}(\underline{X}^{(q)}_n) = \frac{1}{k}\sum_{i=1}^{k} \ln \frac{X_{n-i+1:n} - X_{n_q:n}}{X_{n-k:n} - X_{n_q:n}}, \qquad 0 \le q < 1, \qquad k < n - n_q, \tag{6}$$

i.e. they have the same functional form of the Hill estimator in (3), but with the original sample $\underline{X}_n$ replaced by the sample of excesses $\underline{X}^{(q)}_n$ in (5). Other results on PORT EVI-estimation can be found in [9], [19] and [20].

In this paper, we consider the application of the PORT methodology to the PPWM EVI-estimators, deriving the so-called PORT-PPWM estimators. Such EVI-estimators have the same functional form of the PPWM estimators in (4), but with the original sample $\underline{X}_n$ replaced everywhere by the sample of excesses $\underline{X}_n^{(q)}$, in (5). Let us use the notation χ_q for the q-quantile of the d.f. F. Then (see [2], among others),

$$X_{n_q:n} \underset{n\to\infty}{\overset{p}{\longrightarrow}} \chi_q = F^{\leftarrow}(q) := \inf\{x : F(x) \geq q\}, \quad \text{for} \quad 0 \leq q < 1 \quad \big(F^{\leftarrow}(0) = x_F\big). \tag{7}$$

Consequently, such estimators, also valid only for $0 < \gamma < 1$, just as happens with the PPWM estimators, and provided that χ_q, in (7), is finite, are given by the functional equation,

$$\widehat{\gamma}_{k,n}^{PPWM|q} := \widehat{\gamma}_{k,n}^{PPWM}(\underline{X}_n^{(q)}), \tag{8}$$

with $\widehat{\gamma}_{k,n}^{PPWM}(\underline{X}_n)$ and $\underline{X}_n^{(q)}$ given in (4) and (5), respectively. Just as the PORT-Hill EVI-estimators in (6), these estimators are now invariant for both changes of location and scale, and depend on the *tuning parameter* q, which only influences the asymptotic bias of $\widehat{\gamma}_{k,n}^{PPWM}$, in (4), making this new class highly flexible, and able to compare favourably with the *generalised Pareto probability weighted moment* estimators (GPPWM), for a large variety of underlying models F in the domain of attraction for maxima of the EV_γ d.f., in (1), with $\gamma > 0$.

The GPPWM EVI-estimators have been studied in [22], are scale and location invariant, are valid also only for $0 < \gamma < 1$, and are given by

$$\widehat{\gamma}_{k,n}^{GPPWM} := 1 - \frac{2\hat{a}_1^\star(k)}{\hat{a}_0^\star(k) - 2\hat{a}_1^\star(k)}, \tag{9}$$

with $k = 1, \ldots, n-1$, and

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

See also [4], for an asymptotic comparison at optimal levels of the PPWM and GPPWM EVI-estimators, in (4) and (9), respectively.

We shall further consider one of the best EVI-estimators in the literature, the corrected-Hill estimator in [6], given by

$$\widehat{\gamma}_{k,n}^{CH}(\hat{\beta}, \hat{\rho}) := \widehat{\gamma}_{k,n}^{H}\big(1 - \hat{\beta}(n/k)^{\hat{\rho}}/(1-\hat{\rho})\big), \tag{10}$$

with $(\hat{\beta}, \hat{\rho})$ an adequate estimator of a vector of 'scale' and 'shape' second-order parameters (β, ρ), in a parameterisation $A(t) = \gamma\beta t^\rho$ for a function $A(\cdot)$, that measures the rate of convergence of maximum values to its non-degenerate limit, to be specified in Section 2. The estimators in (10) are indeed minimum-variance reduced-bias (MVRB) EVI-estimators. We have again used the class of β-estimators in [14] and the simplest class of ρ-estimators in [8], made explicit in **Algorithm 3.1**, provided in Section 3. In Section 2 of this paper, we make a brief reference to the asymptotic properties of the EVI-estimators under consideration, and in Section 3, apart from writing the aforementioned algorithm, we perform a small-scale Monte-Carlo simulation, in order to compare the behaviour of the estimators under study for finite samples. Finally, in Section 4, we draw some overall conclusions.

2 Asymptotic behaviour of EVI-estimators under a semi-parametric framework

With $F^{\leftarrow}(\cdot)$ standing for the generalised inverse function of $F(\cdot)$, defined in (7), let us further use the notation

$$U(t) := F^{\leftarrow}(1-1/t),\ t \geq 1, \tag{11}$$

for the reciprocal quantile function. Condition (2) is equivalent to saying that $U \in RV_\gamma$ ([21]), i.e. we often assume the validity of the so-called first-order condition

$$F \in \mathcal{D_M}(EV_\gamma)_{\gamma>0} \quad \Longleftrightarrow \quad \overline{F} \in RV_{-1/\gamma} \quad \Longleftrightarrow \quad U \in RV_\gamma. \tag{12}$$

Consistency of any of the aforementioned EVI-estimators is achieved in the $\mathcal{D_M}(EV_\gamma)_{\gamma \in S_\bullet}$, with $S_H = S_{CH} = (0,\infty)$ and $S_{PPWM} = S_{GPPWM} = (0,1)$, provided that $X_{n-k:n}$ is an *intermediate* o.s., i.e. we need to have

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

Whenever working with heavy right-tails, and in order to derive the asymptotic normality of any semi-parametric EVI-estimator, it is often assumed the validity of a second-order condition either on $\overline{F}$, in (2), or on U, in (11), like

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \begin{cases} \frac{x^\rho - 1}{\rho} & \text{if} \quad \rho < 0 \\ \ln x & \text{if} \quad \rho = 0, \end{cases} \tag{13}$$

where $\rho \leq 0$ is a second-order parameter, which measures the rate of convergence in the first-order condition, in (12). If the limit in (13) exists, it is necessarily of the above mentioned type and $|A| \in RV_\rho$ (see [10]).

If we assume the validity of the second-order framework in (13), the EVI-estimators in (3), for any $\gamma > 0$, and the estimators in (4) and (9), for $\gamma < 1/2$, are asymptotically normal, provided that $\sqrt{k}A(n/k) \to \lambda_A$, finite, as $n \to \infty$, with $A(\cdot)$ given in (13). Indeed, if we denote $\widehat{\gamma}^\bullet_{k,n}$, either the Hill, the PPWM or the GPPWM EVI-estimators, we have, with $Z^\bullet_k$ asymptotically standard normal and for adequate $(b_\bullet,\ \sigma_\bullet) \in (\mathbb{R},\ \mathbb{R}^+)$, the validity of the asymptotic distributional representation

$$\widehat{\gamma}^\bullet_{k,n} \stackrel{d}{=} \gamma + \sigma_\bullet Z^\bullet_k/\sqrt{k} + b_\bullet\ A(n/k)(1 + o_p(1)), \quad \text{as} \quad n \to \infty.$$

Consequently, if we choose k such that $\sqrt{k}\,A(n/k) \to \lambda_A$, finite and not necessarily null, then,

$$\sqrt{k}(\widehat{\gamma}^\bullet_{k,n} - \gamma) \underset{n\to\infty}{\overset{d}{\longrightarrow}} N\left(\lambda_A b_\bullet, \sigma^2_\bullet\right).$$

For the same type of k-values, if we consider the MVRB EVI-estimator $\widehat{\gamma}^{CH}_{k,n}$, in (10), and $A(t) = \gamma\beta t^\rho$, $\rho < 0$,

$$\sqrt{k}(\widehat{\gamma}^{CH}_{k,n} - \gamma) \underset{n\to\infty}{\overset{d}{\longrightarrow}} N\left(0, \sigma^2_H = \gamma^2\right),$$

i.e. $\widehat{\gamma}^{CH}_{k,n}$ outperforms $\widehat{\gamma}^{H}_{k,n}$ for all k.

However, if there is a possible shift s in the model, i.e. if the d.f. $F(x) = F(x; s)$ depends on (x, s) through the difference $x - s$, the function U, the parameter ρ and the A-function, in (13), depend on such a shift s, i.e. $U = U_s = U_0 + s$, $A = A_s$, $\rho = \rho_s$,

$$A_s(t) := \begin{cases} -\gamma s/U_0(t), & \text{if} \quad \gamma + \rho_0 < 0 \ \wedge \ s \neq 0 \\ A_0(t) - \gamma s/U_0(t), & \text{if} \quad \gamma + \rho_0 = 0 \\ A_0(t), & \quad \text{otherwise}, \end{cases} \tag{14}$$

and

$$\rho_s := \begin{cases} -\gamma & \text{if} \quad \gamma + \rho_0 < 0 \wedge s \neq 0 \\ \rho_0 & \text{if} \quad \gamma + \rho_0 \geq 0 \vee s = 0. \end{cases}$$

Just as happens for the PORT-Hill EVI-estimators, in (6), if we consider the PORT-PPWM EVI-estimator $\hat{\gamma}_{k,n}^{PPPM(q)}$, in (8), we expect a change in the dominant component of the bias term comparatively with the one of the PPWM EVI-estimator. Such a dominant component of bias is expected to be no longer related with the $A(\cdot)$-function, in (13), but with the behaviour of $A_s(\cdot)$, $s = -\chi_q$, with χ_q and $A_s(\cdot)$ given in (7) and (14), respectively. Indeed, for the PORT-Hill EVI-estimators, we have $\hat{\gamma}_{k,n}^{H|q} \stackrel{d}{=} \gamma + \gamma Z_k^H/\sqrt{k} + \big(b_{_H} A_0(n/k) + \gamma\chi_q/U_0(n/k)\big)(1 + o_p(1))$ (see [1]). A full theoretical study of the PORT-PPWM estimators, with detailed information on the dominant component of bias, is however out of the scope of this paper.

3 Behaviour of the EVI-estimators: a Monte-Carlo simulation.

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 $\hat{\gamma}_{k,n}^{H}$, $\hat{\gamma}_{k,n}^{PPWM}$, $\hat{\gamma}_{k,n}^{PPWM|q}$, $\hat{\gamma}_{k,n}^{GPPWM}$ and $\hat{\gamma}_{k,n}^{CH}$ in (3), (4), (8), (9) and (10), respectively, for the following underlying parents, all with $|\rho| \leq 1$, the region where the Hill EVI-estimator has often a problematic behaviour:

(i) Student's t_ν with $\nu = 4$ degrees of freedom ($\gamma = 0.25$, $\rho = -0.5$);

(ii) Fréchet parents, with d.f. $F(x) = \exp(-x^{-1/\gamma})$, $x > 0$, $\gamma > 0$, also with $\gamma = 0.25$ ($\rho = -1$);

(iii) Extreme value d.f.s, in (1), with $\gamma = 0.5$ ($\rho = -\gamma = -0.5$), a case in which we cannot guarantee the asymptotic normality of the PPWM EVI-estimators.

Details on multi-sample simulation are available in [15], among others. The multi-sample simulation is a common practice in Monte-Carlo procedures, when we do not have a readily easy way to estimate measures of dispersion of a statistic, like the variance, the MSE or $\arg\min_k MSE(k)$. In a multi-sample simulation of size $m \times r$ instead of generating a unique sample of large size $N = m \times r$, we consider r independent replications of the experiment, all with a size m. We then take as overall estimate of the population parameter under study the average of the r corresponding estimates computed on the basis of the independent replicates. Under very broad conditions, the overall estimator (which is a sample mean) converges to normality as r increases. We can thus estimate the standard error of this overall estimate. For small r, and whenever we can guarantee the asymptotic normality of the estimator of the parameter under consideration, we can then use the t-distribution with $r - 1$ degrees of freedom to approximate its true distribution, and to derive a confidence interval (CI) to the simulated paramater of interest.

We have run the following algorithm $r = 10$ independent times:

Algorithm 3.1.
Repeat the following procedure 5000 *times:*

Step 1 *Randomly generate a sample* $\underline{x}_n := (x_1, \ldots, x_n)$ *from a model* F*, consider the ascending o.s.'s,* $x_{1:n} \le \cdots \le x_{n:n}$*, and for* $q = 0, 0.1, 0.25$*, obtain the associated sample of excesses* $\underline{x}_n^{(q)}$*, with* $\underline{X}_n^{(q)}$ *given in* (5)*.*

Step 2 *With* m_0 *generally denoting the number of positive elements in any of the generated samples, compute* $k_1 = [m_0^{0.995}]$*, and obtain* $\hat\rho \equiv \hat\rho_0(k_1) := -\left|3(T_{m_0}^{(0)}(k_1) - 1)/\left(T_{m_0}^{(0)}(k_1) - 3\right)\right|$*, where*

$$T_n^{(0)}(k) := \frac{\ln\left(M_n^{(1)}(k)\right) - \frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right)}{\frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right) - \frac{1}{3}\ln\left(M_n^{(3)}(k)/6\right)},$$

with

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\{\ln x_{n-i+1:n} - \ln x_{n-k:n}\}^j, \quad j = 1, 2, 3.$$

Step 3 *Next compute*

$$\hat\beta \equiv \hat\beta_{\hat\rho}(k_1) := \left(\frac{k_1}{n}\right)^{\hat\rho}\frac{d_{\hat\rho}(k_1)\ D_0(k_1) - D_{\hat\rho}(k_1)}{d_{\hat\rho}(k_1)\ D_{\hat\rho}(k_1) - D_{2\hat\rho}(k_1)},$$

dependent on the estimator $\hat\rho = \hat\rho_0(k_1)$*, obtained in* ***Step 2****, and where, for any* $\alpha \le 0$*,*

$$d_\alpha(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\alpha} \quad \text{and} \quad D_\alpha(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\alpha} u_i,$$

with u_i*,* $1 \le i \le k$*, the observed values of the scaled log-spacings associated with any of the generated random samples, given by* $u_i := i\left(\ln x_{n-i+1:n} - \ln x_{n-i:n}\right)$*,* $1 \le i \le k$*,* $1 \le k < m_0$*.*

Step 4 *Obtain the simulated samples of* 5000 *EVI-estimates associated with* $\hat\gamma_{k,n}^{H}$*,* $\hat\gamma_{k,n}^{PPWM}$*,* $\hat\gamma_{k,n}^{PPWM|q}$*,* $\hat\gamma_{k,n}^{GPPWM}$ *and* $\hat\gamma_{k,n}^{CH}$ *in* (3)*,* (4)*,* (8)*,* (9) *and* (10)*, respectively. Compute the mean and root mean square error of those* 5000 *estimates, generally denoted* $E(T_{k,n})$ *and* $RMSE(T_{k,n})$*, respectively, with* $T_{k,n} \equiv \hat\gamma_{k,n}^{T}$ *denoting thus any of the aforementioned EVI-estimators. Further compute the optimal values* $k_{0|T} := \arg\min_k RMSE(T_{k,n})$*, the associated estimates at optimal levels,* $T_0 := T_{k_{0|T},n}$*, the root mean square errors* $(RMSE)$ *at optimal levels,* $RMSE_{0|T} := RMSE(T_{k_{0|T},n})$ *and the REFF-indicator*

$$REFF_{T_0|H_0} := RMSE_{0|H}/RMSE_{0|T}. \tag{15}$$

Remark 1.
Further details on the estimation of (β, ρ) *can be found in [16], among others. Interesting alternative classes of* ρ*-estimators have recently been introduced in* [12]*, [13], [7] and [5]. Alternative estimators of* β *can be found in [3] and [18].*

To illustrate the finite sample behaviour of the EVI-estimators, as a function of k, we present, in Figures 1, 2 and 3, the simulated mean values (E) and $RMSE$ patterns of $\widehat{\gamma}_{k,n}^{H}$, $\widehat{\gamma}_{k,n}^{PPWM}$, $\widehat{\gamma}_{k,n}^{PPWM|q}$, $q = 0, 0.1, 0.25$, $\widehat{\gamma}_{k,n}^{GPPWM}$ and $\widehat{\gamma}_{k,n}^{CH}$, as functions of k, the number of top o.s.'s used in the estimation, for a sample size $n = 1000$ and on the basis of the 5000 runs. For simplicity, we shall denote these EVI-estimators by H, $PPWM$, $PPWM|q$, $GPPWM$ and CH.

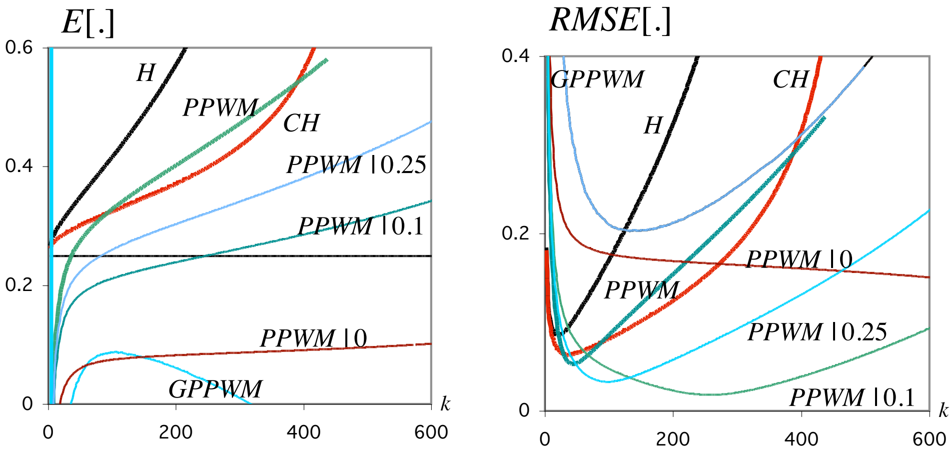

Figure 1: Mean values (E) and root mean square errors ($RMSE$) of the EVI-estimators under study for a Student-t_ν underlying parent with $\nu = 4$ ($\gamma = 1/\nu = 0.25, \rho = -2/\nu = -0.5$).

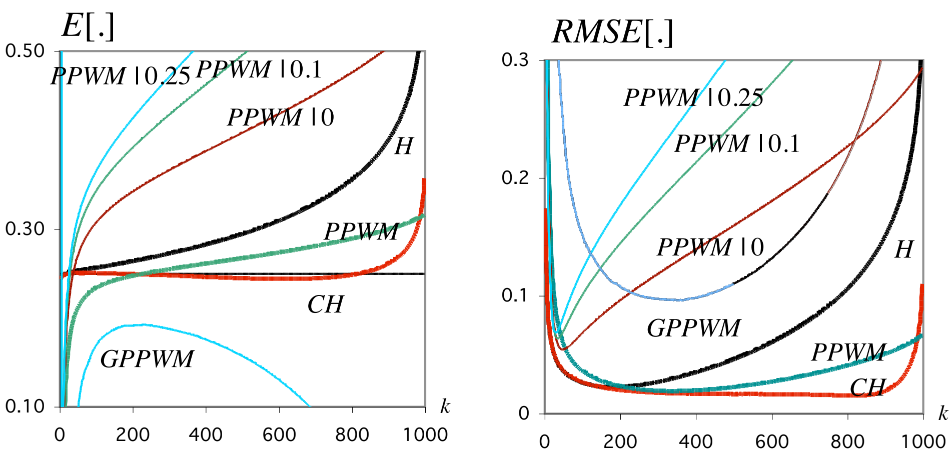

Figure 2: Mean values (E) and root mean square errors ($RMSE$) of the EVI-estimators under study for a Fréchet underlying parent with $\gamma = 0.25$ ($\rho = -1$).

After $r = 10$ repetitions of **Algorithm 3.1**, we have thus computed the averages of all the indicators under consideration, $k_{0|T}$, T_0, $RMSE_{0|T}$, $REFF_{T_0|H_0}$, and the associated 95% CIs. In Table 1, we present the simulated mean values of the above mentioned EVI-estimators, at their simulated optimal levels, for the parents in (i), (ii) and (iii), respectively. For each model,

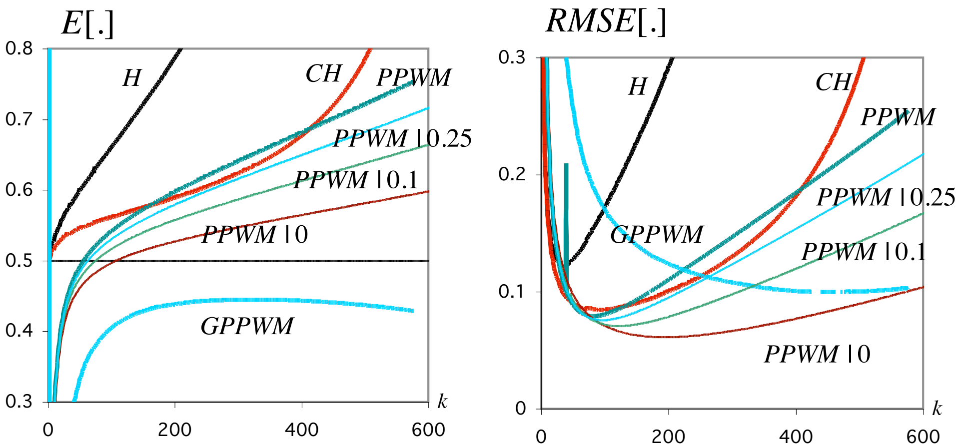

Figure 3: Mean values (E) and root mean square errors ($RMSE$) of the EVI-estimators under study for a EV underlying parent with $\gamma = 0.5$ ($\rho = -\gamma = -0.5$).

the less biased estimator is written in **bold**. Confidence intervals based on the 10 replications are not shown in the tables, but are available from the authors.

In Table 2, we present the REFF-indicators of the different EVI-estimators under consideration and the $RMSE$ of H_0, for the same models as above, respectively, so that we can recover the $RMSE_{0|T}$ of any T_0. Again, confidence intervals are available from the authors and, for each underlying parent, the highest REFF-indicator is written in **bold**.

For Student parents, the value $q = 0$ was not included in the tables, due to the inconsistency of such PORT-PPWM EVI-estimators (see [17], for details on the subject).

For a better visualisation of the tables, we present, in Figures 4, 5 and 6, the mean value ($E_{0|\bullet}$), the $RMSE_{0|\bullet}$ at simulated optimal levels and the $REFF_{\bullet|H_0}$ indicators in (15), for the different EVI-estimators under study, and again for the parents in (i), (ii) and (iii), respectively.

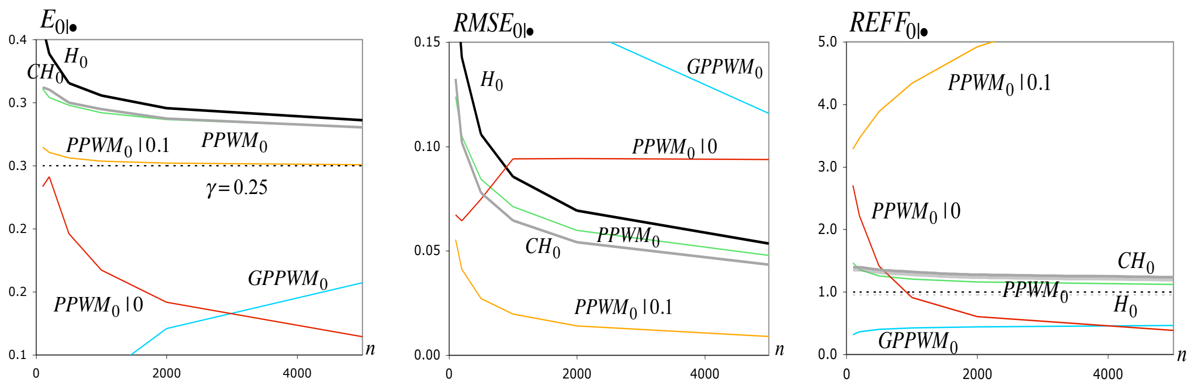

Figure 4: Mean values (*left*), root mean square errors (*center*) at optimal levels and REFF-indicators, in (15) (*right*), for a Student t_4 underlying parent $(\gamma, \rho) = (0.25, -0.5)$ samples.

Table 1: Simulated mean values of the semi-parametric EVI-estimators under consideration, at their simulated optimal levels for underlying Student t_ν, Fréchet and EV_γ parents.

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$
Student t_4 parent ($\gamma = 0.25$)						
H_0	0.3603	0.3386	0.3154	0.3056	0.2956	0.2860
$PPWM_0$	0.3108	0.3042	0.2978	0.2919	0.2867	0.2805
$GPPWM_0$	0.2403	-0.0912	0.0285	0.0826	0.1209	0.1572
$PPWM_0\|0.1$	**0.2648**	**0.2607**	**0.2561**	**0.2536**	**0.2520**	**0.2509**
$PPWM_0\|0.25$	0.2877	0.2819	0.2744	0.2694	0.2651	0.2604
CH_0	0.3121	0.3104	0.3000	0.2948	0.2874	0.2804
Fréchet parent ($\gamma = 0.25$)						
H_0	0.2771	0.2712	0.2658	0.2622	0.2595	0.2574
$PPWM_0$	0.2725	0.2688	0.2643	0.2616	0.2594	0.2570
$GPPWM_0$	0.0223	0.1013	0.1592	0.1860	0.2027	0.2183
$PPWM_0\|0$	0.3216	0.3152	0.3078	0.3017	0.2963	0.2894
$PPWM_0\|0.1$	0.3277	0.3229	0.3160	0.3106	0.3055	0.2988
$PPWM_0\|0.25$	0.3309	0.3273	0.3199	0.3151	0.3095	0.3025
CH_0	**0.2460**	**0.2466**	**0.2488**	**0.2497**	**0.2501**	**0.2499**
EV_γ parent ($\gamma = 0.5$)						
H_0	0.6536	0.6250	0.5961	0.5800	0.5647	0.5509
$PPWM_0$	0.5528	0.5488	0.5445	0.5419	0.5390	0.5345
$GPPWM_0$	0.2811	0.3711	0.4178	0.4411	0.4556	0.4687
$PPWM_0\|0$	**0.5437**	**0.5404**	**0.5359**	**0.5327**	**0.5294**	**0.5250**
$PPWM_0\|0.1$	0.5476	0.5451	0.5408	0.5380	0.5349	0.5307
$PPWM_0\|0.25$	0.5509	0.5469	0.5431	0.5403	0.5374	0.5331
CH_0	0.6375	0.6204	0.5945	0.5797	0.5645	0.5509

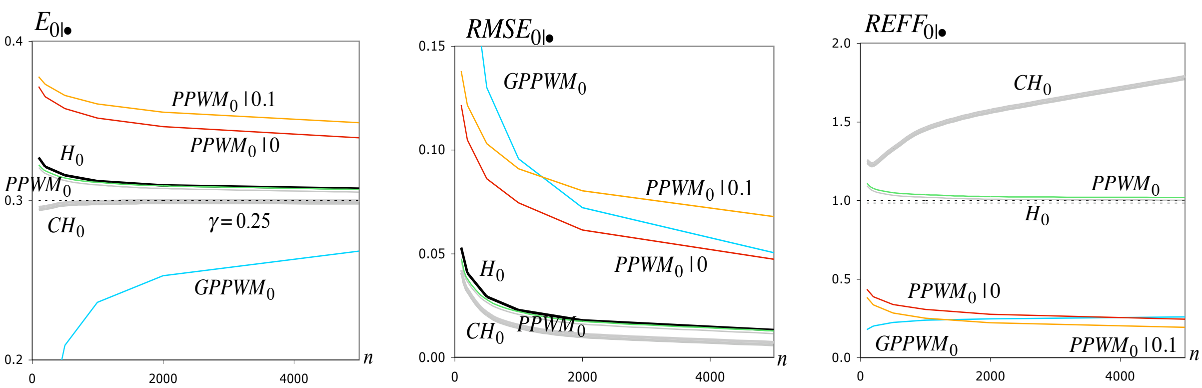

Figure 5: Mean values (*left*), root mean square errors (*center*) at optimal levels and REFF-indicators, in (15) (*right*), for a Fréchet($\gamma = 0.25$) underlying parent.

Table 2: Simulated values of the $\mathrm{REFF}_{\bullet|H_0}$ (for all rows but the last one) and $RMSE_{0|H}$ for underlying Student t_ν, Fréchet and EV_γ parents.

	$n=100$	$n=200$	$n=500$	$n=1000$	$n=2000$	$n=5000$
Student t_4 parent ($\gamma=0.25$)						
$PPWM_0$	1.4697	1.3624	1.2571	1.2014	1.1591	1.1201
$GPPWM_0$	0.3169	0.3627	0.4023	0.4243	0.4413	0.4626
$PPWM_0\|0.1$	**3.2870**	**3.4616**	**3.8865**	**4.3398**	**4.9224**	**5.9161**
$PPWM_0\|0.25$	2.1438	2.0869	2.0758	2.1212	2.1950	2.3505
CH_0	1.3953	1.3995	1.3600	1.3249	1.2811	1.2360
$RMSE_{0\|H}$	0.1819	0.1428	0.1058	0.0856	0.0694	0.0536
Fréchet parent ($\gamma=0.25$)						
$PPWM_0$	1.1105	1.0762	1.0488	1.0368	1.0244	1.0188
$GPPWM_0$	0.1785	0.2012	0.2242	0.2382	0.2466	0.2583
$PPWM_0\|0$	0.4346	0.3880	0.3383	0.3059	0.2753	0.2436
$PPWM_0\|0.1$	0.3828	0.3347	0.2831	0.2508	0.2217	0.1920
$PPWM_0\|0.25$	0.3530	0.3080	0.2599	0.2299	0.2030	0.1755
CH_0	**1.2562**	**1.2378**	**1.3368**	**1.4628**	**1.5752**	**1.7902**
$RMSE_{0\|H}$	0.0529	0.0407	0.0292	0.0228	0.0178	0.0130
EV_γ parent ($\gamma=0.5$)						
$PPWM_0$	1.9925	1.7633	1.5295	1.3971	1.2919	1.1863
$GPPWM_0$	0.8351	1.0292	1.1512	1.2182	1.2781	1.3467
$PPWM_0\|0$	**2.4246**	**2.2017**	**1.9811**	**1.8656**	**1.7796**	**1.6983**
$PPWM_0\|0.1$	2.2157	1.9731	1.7267	1.5883	1.4800	1.3709
$PPWM_0\|0.25$	2.0797	1.8447	1.6049	1.4699	1.3635	1.2560
CH_0	1.4952	1.5034	1.4770	1.4503	1.4174	1.3577
$RMSE_{0\|H}$	0.2566	0.2029	0.1502	0.1218	0.0995	0.0768

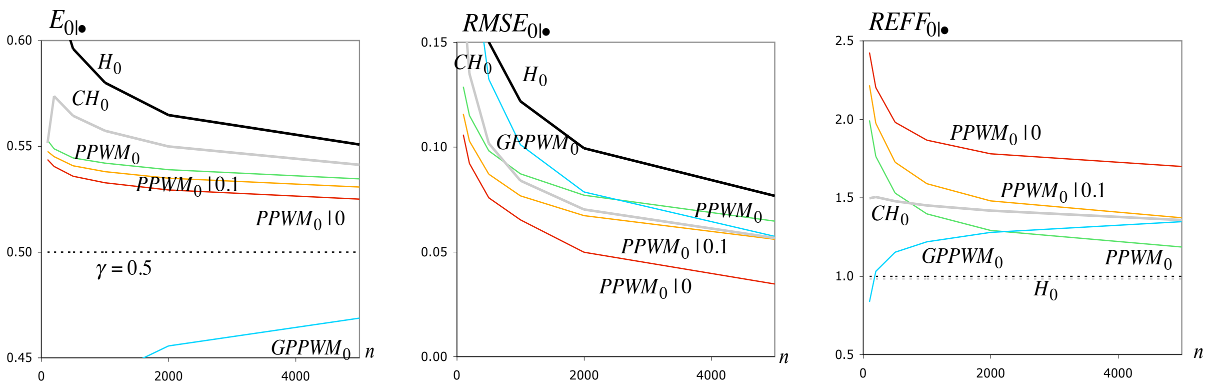

Figure 6: Mean values (*left*), root mean square errors (*center*) at optimal levels and REFF-indicators, in (15) (*right*), for a EV_γ underlying parent ($\rho=-\gamma=-0.5$).

4 Some overall comments

We think sensible to provide the following comments, which in a certain sense justify the parents chosen in Section 3.

- The PORT-PPWM EVI-estimators can be unable to improve the performance of $\widehat{\gamma}^{PPWM}$, in (4), as had already happened with the PORT-Hill estimators when compared with the Hill estimator, H(see Figure 2, associated with Fréchet models). Indeed, this hapens for all models with a left endpoint greater than or equal to zero.
- However, the PORT-PPWM estimators can even outperform the MVRB-estimator, CH, (see both Figure 1 and Figure 3, associated with a Student t_4 and an $EV_{0.5}$ underlying parent, respectively) and have always outperformed the GPPWM estimator.
- For models with a left endpoint equal to infinity, like the Student model, the value $q = 0$ should be discarded due to inconsistency (see the patterns of $PPWM|0$ in Figure 1, and Gomes *et al.*, 2008, for further details on the subject).
- We can often find a value of q that provides the best estimator of γ, regarding for instance minimum $RMSE$, through the use of the new class of estimators $\widehat{\gamma}_{k,n}^{PPWM|q}$, in (8) (the value $q = 0.1$, in Figure 1, and the value $q = 0$, in Figure 3).
- An adaptive choice of k and q is thus an important topic, out of the scope of this paper. But either a heuristic technique, similar to the ones used in [19] and [20], or a generalisation of the double bootstrap methodology used in [15], among others, can surely provide such a data-driven choice of the tuning parameters under play.

Acknowledgement

Research partially supported by FCT – Fundação para a Ciência e a Tecnologia, projects PEst-OE/MAT/UI0006/2011 (CEAUL), PEst-OE/MAT/UI0297/2011 (CMA/UNL), PTDC/FEDER, EXTREMA project and grant SFRH/BPD/77319/2011. We also thank a reviewer, for helpful suggestions.

Bibliography

[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] Arnold, B.C., Balakrishnan, N. and Nagaraja H.N. (1992; 2008) *A first course in order statistics.* John Wiley and Sons; SIAM.

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

[4] Caeiro, F. and Gomes M.I. (2011) *Semi-parametric tail inference through probability-weighted moments.* J. Statist. Plann. and Inference, **141**:2, 937–950.

[5] Caeiro, F. and Gomes, M.I. (2012) *A semi-parametric estimator of a shape second order parameter.* Notas e Comunicações CEAUL 07/2012.

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

[7] Ciuperca, G. and Mercadier, C. (2010) *Semi-parametric estimation for heavy tailed distributions*. Extremes, **13**:1, 55–87.

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

[9] Fraga Alves, M.I., Gomes, M.I., de Haan, L. and Neves, C. (2009) *Mixed moment estimators and location invariant alternatives*. Extremes, **12**, 149–185.

[10] 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.

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

[12] Goegebeur, Y., Beirlant, J. and de Wet, T. (2008) *Linking Pareto-tail kernel goodness-of-fit statistics with tail index at optimal threshold and second order estimation*. Revstat, **6**:1, 51–69.

[13] Goegebeur, Y., Beirlant, J. and de Wet, T. (2010) *Kernel estimators for the second order parameter in extreme value statistics*. J. Statist. Plann. Inference, **140**:9, 2632–2654.

[14] Gomes, M.I. and Martins, M.J. (2002) *"Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter*. Extremes, **5**:1, 5–31.

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

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

[17] Gomes, M.I., Fraga Alves, M.I. and Araújo Santos, P. (2008) *PORT Hill and moment estimators for heavy-tailed models*. Comm. Statist.—Simul. and Comput., **37**, 1281–1306.

[18] Gomes, M.I., Henriques-Rodrigues, L., Pereira, H. and Pestana, D. (2010) *Tail index and second order parameters' semi-parametric estimation based on the log-excesses*. J. Statist. Comput. and Simul., **80**:6, 653–666.

[19] Gomes, M.I., Henriques-Rodrigues, L. and Miranda, C. (2011) *Reduced-bias location-invariant extreme value index estimation: a simulation study*. Comm. Statist.—Simul. and Comput., **40**:3, 424–447.

[20] Gomes, M.I., Henriques-Rodrigues, L., Fraga Alves, M.I. and Manjunath, B.G. (2012) *Adaptive PORT-MVRB estimation: an empirical comparison of two heuristic algorithms*. J. Statist. Comput. and Simul., DOI:10.1080/00949655.2011.652113

[21] Haan, L. de (1984) *Slow variation and characterization of domains of attraction*. In Tiago de Oliveira, ed., *Statistical extremes and applications*. D. Reidel, Dordrecht, 31–48.

[22] Haan, L. de and Ferreira, A. (2006) *Extreme value theory: an introduction*. Springer Science+Business Media, LLC, New York, USA.

[23] Hill, B.M. (1975) *A simple general approach to inference about the tail of a distribution*. Ann. Statist., **3**:5, 1163–1174.