A class of semi-parametric probability weighted moment estimators

Frederico Caeiro and M. Ivette Gomes

Abstract In this paper we deal with the semi-parametric estimation of the right tail $1-F$. Through the use of probability weighted moments based on the largest observations, we study a class of estimators for the extreme value index γ, the scale parameter C and the *Value-at-Risk* at a level p, the size of the loss occurred with a small probability p.

1 Introduction

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. We assume that F is a heavy tailed model with a Pareto-type tail, i.e.,

$$\overline{F}(x) := 1 - F(x) \sim (C/x)^{1/\gamma}, \qquad x \to \infty, \tag{1}$$

where C and γ are unknown scale and shape parameters, respectively. Then $\overline{F}$ is a regularly varying function with a negative index of regular variation equal to $-1/\gamma$, and we are in the max-domain of attraction of the Extreme Value distribution

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

Suppose that we are interested in the estimation of a *high quantile* of probability $1-p$, or equivalently, in the estimation of the *Value-at-Risk* (VaR) at a level p, the size of the loss occurred with a small probability p,

Frederico Caeiro
Faculdade de Ciências e Tecnologia da UNL, 2829-516 Caparica and CMA, e-mail: fac@fct.unl.pt

M. Ivette Gomes
FCUL, Campo Grande, 1749-016 Lisboa, and CEAUL e-mail: ivette.gomes@fc.ul.pt

$$VaR_p \equiv \chi_{1-p} := F^{\leftarrow}(1-p) = \inf\{x : F(x) \geq 1-p\}, \tag{2}$$

with the notation $F^{\leftarrow}$ standing thus for the generalized inverse function of F.

In Section 2 we present some already studied semi-parametric estimators and introduce a new class, to be studied in this paper. In Section 3, after a few technical details, we study the asymptotic behaviour of the estimators under consideration. Finally, Section 4 is dedicated to a small-scale simulation study.

2 Estimators under study

Under the largest observations framework, and whenever dealing with Pareto-type tailed models, the classical semi-parametric estimators of γ and C are the Hill estimator ([4]) and Weissman estimator ([7]), with functional expressions

$$\hat{\gamma}^H_{k,n} := \frac{1}{k}\sum_{i=1}^{k}\left(\ln X_{n-i+1:n} - \ln X_{n-k:n}\right), \qquad k = 1,2,\ldots,n-1, \tag{3}$$

and

$$\hat{C}^{W,H}_{k,n} := X_{n-k:n}\left(\frac{k}{n}\right)^{\hat{\gamma}^H_{k,n}}, \quad k = 1,2,\ldots,n-1, \tag{4}$$

where $X_{i:n}$ denotes the *i-th* ascending order statistic. These estimators are pseudo-maximum likelihood estimators and have usually a high asymptotic bias which makes the choice of k very difficult. This problem led researchers to deal with bias reduction and study new estimators with smaller mean squared error (MSE).
Since heavy-tailed models only have mean value if $\gamma < 1$, methods based on sample moments are rarely considered when we work with such distributions. But in many practical fields like in finance or insurance, for example, we usually have a positive EVI smaller than one, and even smaller than $1/2$. In this article, we again consider the *probability weighted moments* (PWM) method, a generalization of the Method of Moments (Greenwood *et al.* [3]). This method is know for being more efficient than the maximum likelihood method for small to moderate sample sizes (Landwher *et al.* [6], Hosking and Wallis [5]). 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. When $r = s = 0$, $M_{p,0,0}$ are the usual noncentral moments. It is usual to work with one of the two particular and simple cases:

$$a_r := M_{1,0,r} = E(X(1-F(X))^r) \quad \text{or} \quad b_r := M_{1,r,0} = E(X(F(X))^r). \tag{5}$$

Given a sample of size n, the unbiased estimators of a_r and b_r are, respectively,

$$\hat{a}_r = \frac{1}{n}\sum_{i=1}^{n-r}\frac{\binom{n-i}{r}}{\binom{n-1}{r}}X_{i:n} = \frac{1}{n}\sum_{i=1}^{n}\frac{(n-i)(n-i-1)\ldots(n-i-r+1)}{(n-1)(n-2)\ldots(n-r)}X_{i:n}, \tag{6}$$

and

$$\hat{b}_r = \frac{1}{n}\sum_{i=r+1}^{n} \frac{\binom{i-1}{r}}{\binom{n-1}{r}} X_{i:n} = \frac{1}{n}\sum_{i=1}^{n} \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} X_{i:n}. \tag{7}$$

Caeiro and Gomes ([1]) studied the PWM estimators for the parameters of a Pareto tail, based on the top k largest observations, $X_{n:n} \geq X_{n-1:n} \geq \dots \geq X_{n-k+1:n}$. The PWM estimators, valid for $\gamma < 1$, and based on the largest values, are

$$\hat{\gamma}_{k,n}^{PWM} = 1 - \frac{\hat{a}_1(k)}{\hat{a}_0(k) - \hat{a}_1(k)}, \qquad \hat{C}_{k,n}^{PWM} = \frac{\hat{a}_0(k)\ \hat{a}_1(k)}{\hat{a}_0(k) - \hat{a}_1(k)} \left(\frac{k}{n}\right)^{\hat{\gamma}_{k,n}^{PWM}}, \tag{8}$$

with $k = 2, \cdots, n$ and $\hat{a}_s(k) := \frac{1}{k}\sum_{i=1}^{k} \left(\frac{i-1}{k-1}\right)^s X_{n-i+1:n}, \quad s = 0, 1.$

To overcome the restriction $\gamma < 1$ of the previous estimators, we will study the PWM estimators based on the moments

$$a_{r,s} := E(X^r(1-F(X))^s) \quad \text{or} \quad b_{r,s} := E(X^r(F(X))^s). \tag{9}$$

The constant r will be a parameter that allow us to extend the domain of validity of the previous PWM estimators. The PWM estimators, valid for $\gamma < \frac{1}{r}$, and based on the largest values, are

$$\hat{\gamma}_{k,n}^{PWM(r)} = \frac{1}{r}\left(1 - \frac{\hat{a}_{r,1}(k)}{\hat{a}_{r,0}(k) - \hat{a}_{r,1}(k)}\right), \quad k = 2, \cdots, n, \tag{10}$$

and

$$\hat{C}_{k,n}^{PWM(r)} = \left(\frac{\hat{a}_{r,0}(k)\ \hat{a}_{r,1}(k)}{\hat{a}_{r,0}(k) - \hat{a}_{r,1}(k)}\right)^{\frac{1}{r}} \left(\frac{k}{n}\right)^{\hat{\gamma}_{k,n}^{PWM(r)}}, \quad k = 2, \cdots, n, \tag{11}$$

with $\hat{a}_{r,s}(k) := \frac{1}{k}\sum_{i=1}^{k} \left(\frac{i-1}{k-1}\right)^s (X_{n-i+1:n})^r,\ s = 0, 1$. The parameter r can also be used as a tuning parameter, controlled at our ease to reduce the bias or the MSE.

Since $\chi_{1-p} := F^{\leftarrow}(1-p) \sim Cp^{-\gamma}$, as $p \to 0$, the classical and PWM high quantile estimators, based on the largest values, are

$$\hat{Q}_{k,n}^{W,H}(p) := X_{n-k:n} \left(\frac{k}{np}\right)^{\hat{\gamma}_{k,n}^{H}}, \tag{12}$$

and

$$\hat{Q}_{k,n}^{PWM(r)}(p) = \left(\frac{\hat{a}_{r,0}(k)\ \hat{a}_{r,1}(k)}{\hat{a}_{r,0}(k) - \hat{a}_{r,1}(k)}\right)^{\frac{1}{s}} \left(\frac{k}{np}\right)^{\hat{\gamma}_{k,n}^{PWM(r)}}. \tag{13}$$

3 Asymptotic properties

To guarantee the consistency of many semi-parametric estimators, we usually need to assume that k is intermediate, i.e.,

$$k = k_n \to \infty \quad \text{and} \quad k/n \to 0, \quad \text{as } n \to \infty. \tag{14}$$

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

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho} \quad \Leftrightarrow \quad \lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma\, \frac{x^\rho - 1}{\rho}, \tag{15}$$

valid for all $x > 0$, where $\rho \le 0$ is a second-order parameter controlling the speed of convergence of $U(tx)/U(t)$ to x^γ.

Hill's estimator is well studied in the literature. Under the above second order condition in (15) and for intermediate k, we get (de Haan and Peng [2]):

$$\hat{\gamma}^H_{k,n} \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} Z^H_k + \frac{A(n/k)}{1-\rho}(1 + o_p(1)), \tag{16}$$

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.

More generally than Theorem 3.1 in Caeiro and Gomes [1], but with a similar proof, we now state the following theorem.

Theorem 1. *Under the second-order framework, in* (15)*, and for intermediate k, the asymptotic distributional representation*

$$\hat{\gamma}^{PWM(r)}_{k,n} \stackrel{d}{=} \gamma + \frac{\sigma_{PWM(r)}}{\sqrt{k}} Z^{PWM(r)}_k + b_{PWM(r)} A(n/k)(1 + o_p(1)) \tag{17}$$

holds, for $\gamma < 1/2r$ *,*$r > 0$*, where* $Z^\bullet_k$ *is a standard normal r.v., and*

$$\sigma^2_{PWM(r)} = \frac{\gamma^2(1-r\gamma)(2-r\gamma)^2}{(1-2r\gamma)(3-2r\gamma)}, \qquad b_{PWM(r)} = \frac{(1-r\gamma)(2-r\gamma)}{(1-r\gamma-\rho)(2-r\gamma-\rho)}. \tag{18}$$

Corollary 1. *If we further assume that* $\sqrt{k}A(n/k) \to \lambda$*, finite and not necessarily null,*

$$\sqrt{k}(\hat{\gamma}^\bullet_{k,n} - \gamma) \xrightarrow{d} N\left(\lambda\, b_\bullet, \sigma^2_\bullet\right), \qquad as \quad n \to \infty, \tag{19}$$

with $\bullet$ *denoting H or* $PWM(r)$*,* $\sigma^2_H = \gamma^2$ *and* $b_H = \frac{1}{1-\rho}$*.*

Remark 1. Notice that $\sigma^2_H < \sigma^2_{PPWM(r)}$, for every $\gamma > 0$ and $r > 0$. On the other hand, $b_{PPWM(r)} < b_H$, unless $\rho = 0$.

In Figure 3 we provide a picture of $\sigma_{PWM(r)}$ (left) and $b_{PWM(r)}$ (right) as function of r. These functions have opposite behaviour: The variance increases with r, but the bias decreases with r. The choice of the "optimal" r that minimizes the MSE is not obvious and is a subject outside the scope of this paper.

We now state the following two theorems, related with the asymptotic behaviour, at optimal levels, of the estimators of the scale parameter C and high quantiles.

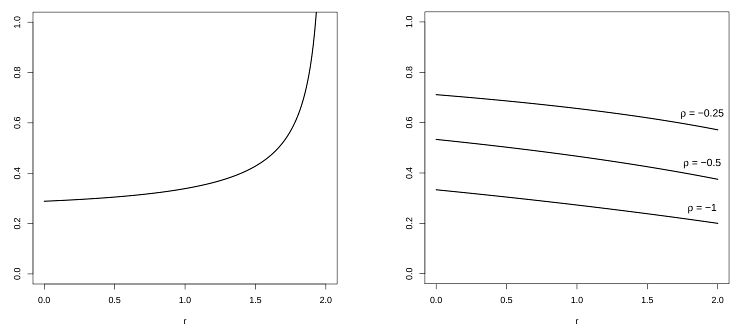

Fig. 1 Left: $\sigma_{_{PWM(r)}}$, as function of r for $\gamma = 0.25$, Right: $b_{_{PWM(r)}}$, as function of r for $\gamma = 0.25$ and three different values of ρ.

Theorem 2. *Under the conditions of* Theorem 1, but with $\rho < 0$, if we further assume $\ln(k/n) = o(\sqrt{k})$ and $\sqrt{k}A(n/k) \longrightarrow \lambda$, then

$$\frac{\sqrt{k}}{\ln(k/n)}\left(\frac{\hat{C}^{W,H}_{k,n}}{C} - 1\right) \stackrel{d}{=} \sqrt{k}\left(\hat{\gamma}^{H}_{k,n} - \gamma\right)(1 + o_p(1)), \tag{20}$$

and

$$\frac{\sqrt{k}}{\ln(k/n)}\left(\frac{\hat{C}^{PWM(r)}_{k,n}}{C} - 1\right) \stackrel{d}{=} \sqrt{k}\left(\hat{\gamma}^{PWM(r)}_{k,n} - \gamma\right)(1 + o_p(1)), \quad r > 0. \tag{21}$$

Theorem 3. *Under the conditions of* Theorem 1*, but with* $\rho < 0$*, if* $p = p_n$ *is a sequence of probabilities such that* $c_n = k/(np) \underset{n\to\infty}{\longrightarrow} \infty$*,* $\ln c_n = o(\sqrt{k})$ *and* $\sqrt{k}A(n/k) \longrightarrow \lambda$*, then,*

$$\frac{\sqrt{k}}{\ln c_n}\left(\frac{\hat{Q}^{W,H}_{k,n}(p)}{\chi_{1-p}} - 1\right) \stackrel{d}{=} \sqrt{k}\left(\hat{\gamma}^{H}_{k,n} - \gamma\right)(1 + o_p(1)), \tag{22}$$

and

$$\frac{\sqrt{k}}{\ln c_n}\left(\frac{\hat{Q}^{PWM(r)}_{k,n}(p)}{\chi_{1-p}} - 1\right) \stackrel{d}{=} \sqrt{k}\left(\hat{\gamma}^{PWM(r)}_{k,n} - \gamma\right)(1 + o_p(1)), \quad r > 0. \tag{23}$$

Remark 2. The previous theorems allow us to conclude that the asymptotic dominant behaviour of the scale parameter and high quantiles estimators is thus fully determined by the asymptotic behaviour of $\hat{\gamma}^{\bullet}_{k,n}$.

4 Finite sample behaviour: small-scale simulation study

We have implemented a Monte Carlo simulation of 5000 runs for the shape parameter estimators $\gamma^{H}_{k^*,n}$, $\hat{\gamma}^{PWM(r)}_{k^*,n}$ and high quantile normalized estimators $\bar{Q}^{W,H}_{k^*,n}(p) := \hat{Q}^{W,H}_{k^*,n}(p)/\chi_{1-p}$ and $\bar{Q}^{PWM(r)}_{k^*,n}(p) := \hat{Q}^{PWM(r)}_{k^*,n}(p)/\chi_{1-p}$ with $p = 1/n$ and $r \in \{0.25, 1, 1.25\}$. To work with the exact same number of top o.s., we have considered $k^* = k$ for the classical estimators and $k^* = k+1$ for the PWM estimators. We have only considered Burr underlying parents with d.f. $F(x) = 1-(1+x^{-\rho/\gamma})^{1/\rho}$, $x > 0$ with $(\gamma, \rho) \in \{(0.25, -1.0), (1.25, -1.0)\}$. For every estimator we have simulated, the **mean value**, the **root mean squared error** (RMSE) the **optimal level**, $k_0^* = \arg\min_k RMSE$, and the **optimal sample fraction**, k_0^*/n. For simplicity, we shall denote the shape parameter estimators by H and PWM(r) and the normalized high quantile estimators by W-H and PWM(r).

To illustrate the finite sample behaviour of the estimators, we present, in Figures 2, 3, 4, and 5, the simulated mean values (E) and root mean square errors (RMSE) patterns as functions of k^* for a sample size $n = 500$. In Table 4 we present the simulated sample fraction and mean values/RMSE of the above mentioned estimators, at their simulated optimal levels. For Burr models with $(\gamma, \rho) = (1.25, -1)$, we present the simulated results of the PWM(r) estimators with r=0.25 and 1. But for this model, we need to have $r < 0.8$ to assure the consistency of the PWM(r) estimators.

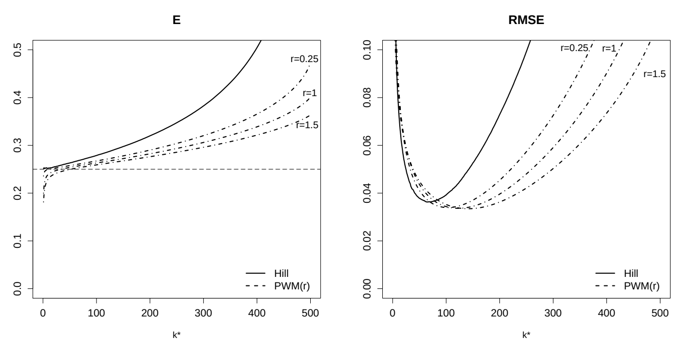

Fig. 2 Simulated mean values (left) and root mean squared errors (right), as functions of k^*, of the estimators $\hat{\gamma}^{H}_{k^*,n}$ and $\hat{\gamma}^{PWM(r)}_{k^*,n}$, $n = 500$, from a Burr$(0.25, -1)$ parent.

Acknowledgements Research partially supported by FCT / OE, Financiamento Base 2009 ISFL-1-297, POCI 2010 and PTDC / FEDER.

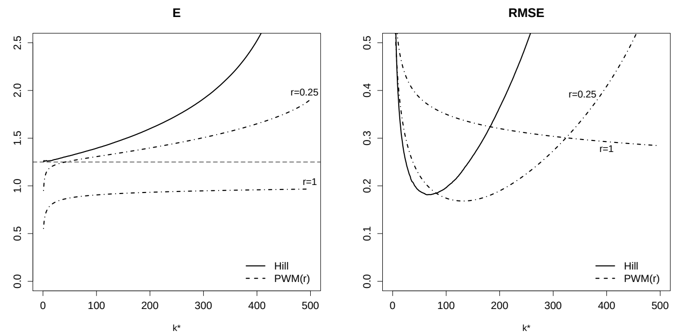

Fig. 3 Simulated mean values (left) and root mean squared errors (right), as functions of k^*, of the estimators $\hat{\gamma}^{H}_{k^*,n}$ and $\hat{\gamma}^{PWM(r)}_{k^*,n}$, $n = 500$, from a $\text{Burr}(1.25, -1)$ parent.

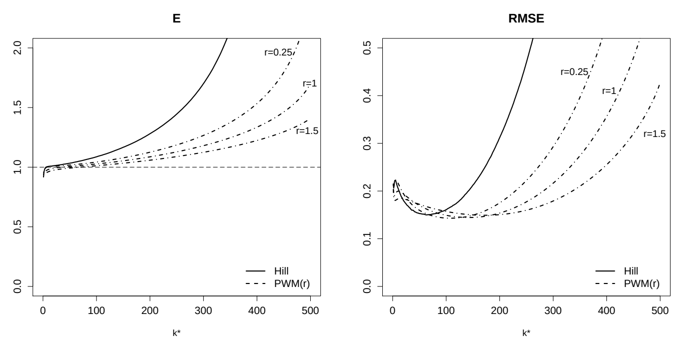

Fig. 4 Simulated mean values (left) and root mean squared errors (right), as functions of k^*, of the normalized estimators $\bar{Q}^{W,H}_{k^*,n}(p)$ and $\bar{Q}^{PWM(r)}_{k^*,n}(p)$, $n = 500$, from a Burr$(0.25, -1)$ parent.

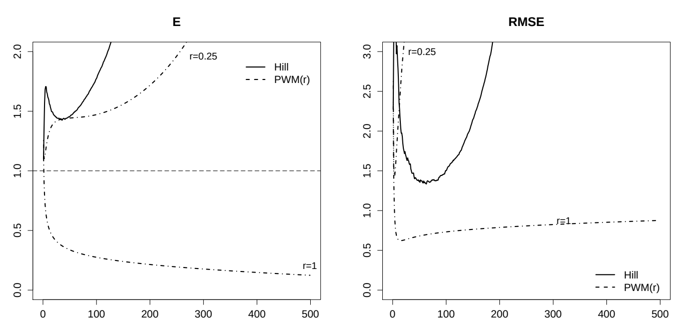

Fig. 5 Simulated mean values (left) and root mean squared errors (right), as functions of k^*, of the normalized estimators $\bar{Q}^{W,H}_{k^*,n}(p)$ and $\bar{Q}^{PWM(r)}_{k^*,n}(p)$, $n = 500$, from a Burr$(1.25, -1)$ parent.

Table 1 Burr parent: Simulated optimal sample fraction and mean values/RMSE, at their simulated optimal levels for the high quantile normalized estimators $\bar{Q}^{W-H}_{k^*,n}(p)$ and $\bar{Q}^{PWM(r)}_{k^*,n}(p)$ with $p = 1/n$.

	$(\gamma,\rho) = (0.25,-1.0)$				$(\gamma,\rho) = (1.25,-1.0)$			
	n	k_0^*/n	E	$RMSE$	n	k_0^*/n	E	$RMSE$
W-H	50	0.0400	0.9569	0.2086	50	0.0400	1.3431	3.1957
	100	0.2100	1.0471	0.1984	100	0.2000	1.7620	2.4345
	200	0.1750	1.0505	0.1800	200	0.1750	1.6919	1.9448
	500	0.1280	1.0460	0.1499	500	0.1260	1.5083	1.3346
	1000	0.1100	1.0479	0.1313	1000	0.1100	1.4479	1.0342
	2000	0.0905	1.0443	0.1128	2000	0.0755	1.3237	0.8255
	5000	0.0710	1.0407	0.0926	5000	0.0698	1.2974	0.6254
PWM(1)	50	0.5600	1.0295	0.1889	50	0.4200	0.3722	0.7233
	100	0.4600	1.0413	0.1813	100	0.2200	0.4005	0.6844
	200	0.3750	1.0471	0.1674	200	0.0900	0.4421	0.6446
	500	0.2840	1.0487	0.1445	500	0.0300	0.4738	0.6222
	1000	0.2310	1.0483	0.1280	1000	0.0170	0.4648	0.6248
	2000	0.1845	1.0449	0.1111	2000	0.0100	0.4471	0.6336
	5000	0.1374	1.0399	0.0912	5000	0.0050	0.4274	0.6483
PWM(0.25)	50	0.0400	0.9674	0.2098	50	0.1000	1.1564	1.8126
	100	0.3500	1.0491	0.1895	100	0.0300	1.0993	1.6142
	200	0.2950	1.0527	0.1703	200	0.0150	1.1045	1.5402
	500	0.2160	1.0480	0.1430	500	0.0060	1.1096	1.4340
	1000	0.1790	1.0478	0.1264	1000	0.0040	1.1548	1.4836
	2000	0.1420	1.0428	0.1085	2000	0.2600	1.4928	1.0606
	5000	0.1120	1.0400	0.0890	5000	0.1588	1.3177	0.6786
PWM(1.25)	50	0.8600	1.0402	0.1621				
	100	0.6000	1.0346	0.1681				
	200	0.4650	1.0398	0.1629				
	500	0.3440	1.0439	0.1492				
	1000	0.2860	1.0473	0.1356				
	2000	0.2335	1.0465	0.1206				
	5000	0.1722	1.0415	0.0984				

References

1. Caeiro, F., Gomes, M.I.: Semi-Parametric Tail Inference through Probability-Weighted Moments. J. Statist. Plann. Inference **141**, 937–950 (2011)
2. de Haan, L., Peng, L.: Comparison of Tail Index Estimators. *Statistica Neerlandica* **52**, 60–70 (1988)
3. Greenwood, J. A,, Landwehr, J. M., Matalas, N. C., Wallis, J.R.: Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form. *Water Resources Research* **15**, 1049–1054 (1979)
4. Hill, B.M.: A simple general approach to inference about the tail of a distribution. Ann. Statist. **3**, 1163–1174 (1975)
5. Hosking, J., Wallis, J.: Parameter and quantile estimation for the Generalized Pareto distribution. *Technometrics* **29**(3), 339–349 (1987)
6. Landwher, J., Matalas, N., Wallis, J.: Probability Weighted Moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resour. Res.* **15**, 1055–1064 (1979)
7. Weissman, I.: Estimation of Parameters of Large Quantiles Based on the k Largest Observations. *J. Amer. Statist. Assoc.* **73**, 812–815 (1978)