

Semi-Parametric Probability-Weighted Moments Estimation Revisited*

Frederico Caeiro

Universidade Nova de Lisboa, and CMA

M. Ivette Gomes

Universidade de Lisboa, DEIO, CEAUL and FCUL

and

Björn Vandewalle

Universidade Nova de Lisboa, ISEGI, and CEAUL

October 30, 2010

Abstract

In this paper, for heavy-tailed models and through the use of probability weighted moments based on the largest observations, we deal with the semi-parametric estimation of the *Value-at-Risk* at a level p , the size of the loss occurred with a small probability p , as well as the dual problem of estimation of the *probability of exceedance* of a high level x . These estimation procedures depend crucially on the estimation of the *extreme value index*, the primary parameter in *Statistics of Extremes*, also done on the basis of the same weighted moments. Under regular variation conditions on the right-tail of the underlying distribution function F , we prove the consistency and asymptotic normality of the estimators under consideration in this paper, linking their asymptotic behaviour to the asymptotic behaviour of the extreme value index estimator they are based on. The performance of these estimators, for finite samples, is illustrated through Monte-Carlo simulations. An adaptive choice of thresholds is put forward. Applications to a real data set in the field of insurance as well as to simulated data are also provided.

AMS 2000 subject classification. Primary 62G32, 62E20; Secondary 65C05.

Keywords and phrases. Heavy tails, value-at-risk or high quantiles, probability of exceedance of a high level, semi-parametric estimation.

*Research partially supported by FCT / POCI 2010 and PTDC/ FEDER.

1 Introduction, preliminaries and scope of the article

Let X_1, X_2, \dots, 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.'s) $X_{1:n} \leq \dots \leq X_{n:n}$. 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 ,

$$\text{VaR}_p \equiv \chi_{1-p} := F^{\leftarrow}(1 - p) = \inf\{x : F(x) \geq 1 - p\}, \quad (1.1)$$

with the notation F^{\leftarrow} standing thus for the generalized inverse function of F . Moreover, we are also interested in the estimation of the *probability of exceedance* of a high level $x = x_n$,

$$p = p_x := 1 - F(x) =: \bar{F}(x). \quad (1.2)$$

Extreme Value Theory (EVT) provides a great variety of results that enable us to deal with alternative approaches in the statistical analysis of extreme events. Those approaches are essentially based on the well-established limiting results described in the following.

1.1 Main limiting results in EVT

The main limiting result in EVT can be attributed to Gnedenko (1943), who fully characterized the possible non-degenerate limiting distribution of the linearly normalized maximum, $(X_{n:n} - b_n)/a_n$, $a_n > 0$, $b_n \in \mathbb{R}$. Such a limit is of the type of the general *extreme value distribution* (EVD),

$$EV_\gamma(x) := \begin{cases} \exp(-(1 + \gamma x)^{-1/\gamma}), & 1 + \gamma x > 0 & \text{if } \gamma \neq 0 \\ \exp(-\exp(-x)), & x \in \mathbb{R} & \text{if } \gamma = 0. \end{cases} \quad (1.3)$$

When such a non-degenerate limit exists, we say that F belongs to the max-domain of attraction of EV_γ and denote this by $F \in \mathcal{D}_M(EV_\gamma)$. The shape parameter γ is related with the heaviness of the right-tail $\bar{F} = 1 - F$ and it is often called the *extreme value index* (EVI).

Another seminal result in the field of EVT is due to Balkema and de Haan (1974) and Pickands (1975). If we properly scale the excesses over a high threshold u , the limit distribution of those scaled excesses is the *Generalized Pareto distribution* (GPD), strongly related with the d.f., $EV_\gamma(x)$, in (1.3), and defined by,

$$GP_\gamma(x) := 1 + \ln EV_\gamma(x) = \begin{cases} 1 - (1 + \gamma x)^{-1/\gamma}, & 1 + \gamma x > 0, & x > 0 & \text{if } \gamma \neq 0 \\ 1 - \exp(-x), & x > 0 & & \text{if } \gamma = 0 \end{cases} \quad (1.4)$$

(see, for instance, Embrechts *et al.*, 1997, Section 3.4, and Reiss and Thomas, 2007, Section 1.4, for more details).

1.2 Most relevant approaches in the field of Statistics of Extremes

We shall briefly review the three most important approaches in the area of *Statistics of Extremes*: the *block maxima* (BM) method, the *peaks-over-threshold* (POT) or even the *peaks-over-random-threshold* (PORT) methods and the *largest observations* (LOB) method. For a more detailed review, with extensive associated references, see Gomes *et al.* (2008).

- The first method, the BM method, is of a parametric nature: we work with a sample of maxima of adequate blocks of observations, and estimate the parameters $(\lambda, \delta, \gamma)$ of the EVD, $EV_\gamma((x - \lambda)/\delta)$, $\lambda \in \mathbb{R}$, $\delta > 0$, $\gamma \in \mathbb{R}$, with $EV_\gamma(x)$ given in (1.3). This method is known to be possibly inefficient, due to the fact that the loss of information in each block can be catastrophic.
- In the second approach, the POT method, inference is performed through the use of the sample of excesses over a high deterministic threshold u . The limiting d.f. of these excesses is, up to a scale factor, the distribution $GP_\gamma(x)$, in (1.4), and the method can be of a parametric or a semi-parametric nature. Note that the high threshold can also be a random value, leading to the PORT methodology, a terminology recently introduced in Araújo Santos *et al.* (2006).
- The third approach, the LOB method, is the one we shall consider in this paper. It uses the *largest k observations* to make inference about the right tail $\bar{F} = 1 - F$, assuming only that $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$.

1.3 Estimators under study

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

$$\hat{\gamma}_{k,n}^H := \frac{1}{k} \sum_{i=1}^k (\ln X_{n-i+1:n} - \ln X_{n-k:n}) \quad (1.5)$$

and

$$\hat{Q}_{k,n}^H(p) := X_{n-k:n} c_k^{\hat{\gamma}_{k,n}^H}, \quad c_k \equiv c_k(p) := \frac{k}{np}, \quad k = 1, 2, \dots, n-1, \quad (1.6)$$

respectively, which are pseudo-maximum likelihood estimators, consistent in the whole $\mathcal{D}_{\mathcal{M}}^+ := \mathcal{D}_{\mathcal{M}}(EV_\gamma)_{\gamma>0}$, provided that k is intermediate, i.e. if

$$k = k_n \rightarrow \infty \quad \text{and} \quad k/n \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (1.7)$$

In a way dual to (1.6), and given a high level $x = x_n$, the probability $p = p_n$ of exceedance of such a level can be estimated by

$$\hat{p}_{k,n}^H(x) := \left(\frac{k}{n}\right) \tilde{C}_k^{-1/\hat{\gamma}_{k,n}^H}, \quad \tilde{C}_k \equiv \tilde{C}_k(x) := \frac{x}{X_{n-k:n}}, \quad k = 1, 2, \dots, n-1. \quad (1.8)$$

Under further adequate restrictions on k , we can guarantee the asymptotic normality of the estimators $\hat{\gamma}_{k,n}^H$, $\hat{Q}_{k,n}^H(p)$ and $\hat{p}_{k,n}^H(x)$, in (1.5), (1.6) and (1.8), respectively. But most of the times, these estimators exhibit a large variance for small k , a strong bias for moderate k , sample paths with very short stability regions around the target value and a very peaked mean square error (MSE) structure, as a function of k . This has led researchers to the search of alternative estimators, with a smaller MSE.

Since heavy-tailed models only have mean value if $\gamma < 1$, methods based on sample moments have been 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 a positive EVI smaller than one, and even smaller than $1/2$. In this article, and for the estimation of the above mentioned parameters of extreme events, we now revisit the use of a *probability weighted moments* (PWM) method based on the largest observations, developed in Caeiro and Gomes (2010) for the EVI.

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 the particular case,

$$a_r := M_{1,0,r} = E(X(1-F(X))^r),$$

and the associated estimator,

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

For $\gamma < 1$ and for d.f.'s like the EVD, $EV_\gamma((x-\lambda)/\delta)$, with $EV_\gamma(x)$ given in (1.3), the *Pareto* d.f.,

$$P_\gamma(x; \delta) = 1 - (x/\delta)^{-1/\gamma}, \quad x > \delta, \quad (1.10)$$

and the GPD, $GP_\gamma(x/\delta)$, with $GP_\gamma(x)$ defined in (1.4), the PWM have simple expressions, which allow a straightforward estimation of the EVI, γ . For the EVD, see Hosking *et al.* (1985) and the improved versions in Diebolt *et al.* (2007, 2008). As an example, the Pareto PWM (PPWM) and the generalized Pareto PWM (GPPWM) estimators of γ are valid for $\gamma < 1$, and given by

$$\hat{\gamma}^{PPWM} = 1 - \left(\frac{\hat{a}_1}{\hat{a}_0 - \hat{a}_1} \right) \quad \text{and} \quad \hat{\gamma}^{GPPWM} = 1 - \frac{2\hat{a}_1}{\hat{a}_0 - 2\hat{a}_1}, \quad (1.11)$$

respectively, where \hat{a}_0 and \hat{a}_1 are given in (1.9). The estimator $\hat{\gamma}^{GPPWM}$, in (1.11), was introduced and studied in Hosking and Wallis (1987).

We shall consider in this paper, the PPWM estimators of $VarR_p$ and p_x , defined in (1.1) and (1.2), respectively, associated with the PPWM EVI-estimators studied in Caeiro and Gomes (2010). Those estimators are semi-parametric in nature and, for comparison with the equivalent estimators based on the Hill EVI-estimator, in (1.5), are based on the top $k + 1$ largest o.s.'s, $X_{n-k:n} \leq X_{n-k+1:n} \leq \dots \leq X_{n:n}$. Under such a framework, the estimators \hat{a}_0 and \hat{a}_1 , in (1.9), should be replaced by,

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

respectively. The PPWM EVI, VaR and p -estimators, based on the largest values are

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

$$\hat{Q}_{k,n}^{PPWM}(p) := \frac{\hat{a}_0(k) \hat{a}_1(k)}{\hat{a}_0(k) - \hat{a}_1(k)} \left(\frac{k}{np} \right)^{\hat{\gamma}_{k,n}^{PPWM}} \quad (1.13)$$

and

$$\hat{p}_{k,n}^{PPWM}(x) := \left(\frac{k}{n} \right) \left(\frac{x(\hat{a}_0(k) - \hat{a}_1(k))}{\hat{a}_0(k)\hat{a}_1(k)} \right)^{-1/\hat{\gamma}_{k,n}^{PPWM}}, \quad (1.14)$$

respectively, with $k = 1, 2, \dots, n - 1$, and are consistent whenever $\gamma < 1$.

De Haan and Ferreira (2006) considered, also for $\gamma < 1$, the semi-parametric GPPWM EVI-estimator, based on the sample of excesses over the high random level $X_{n-k:n}$, i.e.,

$$\hat{\gamma}_{k,n}^{GPPWM} := 1 - \frac{2\hat{a}_1^*(k)}{\hat{a}_0^*(k) - 2\hat{a}_1^*(k)}, \quad (1.15)$$

with $k = 1, 2, \dots, n - 1$, and $\hat{a}_s^*(k) := \sum_{i=1}^k (i/k)^s (X_{n-i+1:n} - X_{n-k:n})/k$, $s = 0, 1$. For a comparison between the PPWM EVI-estimators in (1.12) and the GPPWM EVI-estimators in (1.15), see Caeiro and Gomes (2010).

1.4 Scope of the article

In Section 2, after reviewing a few results already available in the literature, we state a lemma and a theorem related with the asymptotic properties of the PPWM-estimators, defined in (1.13) and (1.14), of the above mentioned parameters of extreme events, the Value-at-Risk at the level p and the probability p_x of exceedance of a high level x , defined in (1.1) and (1.2), respectively. The performance of these estimators, for finite samples, is illustrated, in Section 3, through a Monte-Carlo simulation study. In

Section 4, we put forward an adaptive choice of thresholds, again on the basis of bootstrap computer-intensive methods. Applications to a real data set in the field of insurance as well as to a simulated data set are provided in Section 5. Finally, in Section 6, we prove the main results in this paper, stated in Section 2.

2 Asymptotic Behaviour of the Estimators

2.1 Most common first and second-order frameworks for heavy tails

For heavy-tailed models, i.e., models with a positive EVI, we assume that F has a Pareto-type right-tail, i.e., with the notation $a(t) \sim b(t)$ if and only if $a(t)/b(t) \rightarrow 1$, as $t \rightarrow \infty$,

$$\bar{F}(x) = 1 - F(x) \sim (x/C)^{-1/\gamma} \iff U(t) \sim C t^\gamma, \quad C, \gamma > 0, \quad (2.1)$$

where C is a scale parameter and $U(t) := F^\leftarrow(1 - 1/t)$, $t > 1$. Note that models with the Pareto-type right-tail, in (2.1), have a regularly varying (RV) right-tail with a negative index of regular variation equal to $-1/\gamma$ (with the notation $\bar{F} \in RV_{-1/\gamma}$), and belong to the max-domain of attraction $\mathcal{D}_{\mathcal{M}}^+$. Indeed, more specifically and for all $x > 0$, we have,

$$F \in \mathcal{D}_{\mathcal{M}}^+ \iff \lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-1/\gamma} \quad (\bar{F} \in RV_{-1/\gamma}) \iff \lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma \quad (U \in RV_\gamma), \quad (2.2)$$

i.e. $\bar{F}(x) = x^{-1/\gamma} L_{\bar{F}}(x)$ and $U(t) = x^\gamma L_U(t)$, where $L_{\bar{F}}$ and L_U are both slowly varying functions (i.e. they both belong to RV_0), not necessarily converging to constants, as happens if condition (2.1) holds.

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 (1.7) holds. To obtain information on the non-degenerate distributional behaviour of semi-parametric estimators of parameters of extreme events, we often assume a second-order condition, like

$$\lim_{t \rightarrow \infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho} \iff \lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}, \quad (2.3)$$

valid 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 $t \rightarrow \infty$. We then say that U is of second-order regular variation, with parameters γ and ρ , and denote such a fact by $U \in 2RV_\gamma^\rho$. If the limits in (2.3) exist, they are necessarily of the above mentioned types and $|A| \in RV_\rho$ (Geluk and de Haan, 1987; see also de Haan and Ferreira, 2006). Moreover, we have

$$U \in 2RV_\gamma^\rho \iff \bar{F} \in 2RV_{-1/\gamma}^{\tilde{\rho}}, \quad \text{with } \tilde{\rho} = \rho/\gamma \quad (2.4)$$

and a rate function \tilde{A} related with A , in (2.3), through the relation $\tilde{A}(t) = A(1/\bar{F}(t))/\gamma^2$. The validity of condition (2.3), with $\rho < 0$, is equivalent to condition (2.1).

2.2 Auxiliary results on intermediate order statistics

In the sequel, let us denote (Y_1, \dots, Y_n) a random sample of size n from a strict Pareto model, with d.f. $P_\gamma(x; 1)$, $P_\gamma(x; \delta)$ given in (1.10). Let $(Y_{1:n}, \dots, Y_{n:n})$ denote the sample of associated ascending o.s.'s. We first state without proof the following results on the asymptotic behaviour of intermediate o.s.'s (see de Haan and Ferreira, 2006, for details).

Lemma 2.1. *If (1.7) holds, then for the intermediate Pareto o.s. $Y_{n-k:n}$, $k Y_{n-k:n}/n \xrightarrow{p} 1$ and consequently, $Y_{n-k:n} \xrightarrow{p} \infty$, as $n \rightarrow \infty$. Moreover,*

$$B_{k,n} := \sqrt{k}(k Y_{n-k:n}/n - 1) \xrightarrow{d} \mathcal{B}, \quad \text{a standard normal r.v.}, \quad (2.5)$$

and $\text{Cov}(B_{r,n}, B_{s,n}) = \sqrt{rs} (1 - s/n)/(s - 1)$, $r < s$.

Lemma 2.2. *Under the first-order framework in (2.2), let us assume that k is intermediate, i.e., (1.7) holds. Then, for the intermediate o.s. $X_{n-k:n}$ ($\stackrel{d}{=} U(Y_{n-k:n})$), $X_{n-k:n}/U(n/k) \xrightarrow{p} 1$, and consequently, $X_{n-k:n} \xrightarrow{p} \infty$, as $n \rightarrow \infty$. If we further assume the validity of the second-order framework in (2.3), then, as $n \rightarrow \infty$, the asymptotic distributional representation*

$$X_{n-k:n}/U(n/k) \stackrel{d}{=} 1 + \gamma B_{k,n}/\sqrt{k} + o_p(A(n/k))$$

holds, where $B_{k,n}$ is the asymptotically standard normal sequence of r.v.'s in (2.5).

2.3 Asymptotic behaviour of the EVI-estimators under play

Regarding the Hill estimator, $\hat{\gamma}_{k,n}^H$, in (1.5):

Proposition 2.1 (de Haan and Peng, 1998, Theorem 1). *Under the second-order framework in (2.3), and for intermediate k , i.e. if (1.7) holds, the asymptotic distributional representation*

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

holds, with Z_k^H asymptotically standard normal. Consequently, if we choose k such that $\sqrt{k} A(n/k) \rightarrow \lambda_A$, finite and not necessarily null,

$$\sqrt{k}(\hat{\gamma}_{k,n}^H - \gamma) \xrightarrow[n \rightarrow \infty]{d} N\left(\frac{\lambda_A}{1 - \rho}, \gamma^2\right).$$

We next refer the following results related with the asymptotic behaviour of the PPWM EVI-estimator, $\hat{\gamma}_{k,n}^{PPWM}$, in (1.12):

Proposition 2.2 (Caeiro and Gomes, 2010, Proposition 2.4). *Under the conditions of Proposition 2.1, and with Z_k^{PPWM} asymptotically standard normal, the asymptotic distributional representation*

$$\hat{\gamma}_{k,n}^{PPWM} \stackrel{d}{=} \gamma + \frac{\sigma_{PPWM} Z_k^{PPWM}}{\sqrt{k}} + b_{PPWM} A(n/k)(1 + o_p(1))$$

holds for any $\gamma < 1/2$, as $n \rightarrow \infty$, where

$$\sigma_{PPWM}^2 := \frac{\gamma^2(1-\gamma)(2-\gamma)^2}{(1-2\gamma)(3-2\gamma)}, \quad \text{and} \quad b_{PPWM} := \frac{(1-\gamma)(2-\gamma)}{(1-\gamma-\rho)(2-\gamma-\rho)}. \quad (2.6)$$

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

$$\sqrt{k} \left(\hat{\gamma}_{k,n}^{PPWM} - \gamma \right) \xrightarrow[n \rightarrow \infty]{d} N(\lambda_A b_{PPWM}, \sigma_{PPWM}^2).$$

Remark 2.1. *It is obvious that $\sigma_H^2 := \gamma^2 < \sigma_{PPWM}^2$, for every $0 < \gamma < 1/2$. On the other hand, $b_{PPWM} < b_H = 1/(1-\rho)$, unless $\rho = 0$. An asymptotic comparison of H and $PPWM$ EVI-estimators at optimal levels can be seen in Caeiro and Gomes (2010).*

2.4 Main asymptotic results

We first state the following lemma:

Lemma 2.3. *For intermediate k , i.e., whenever (1.7) holds, and under the second-order framework in (2.3), but with $\rho < 0$, (2.1) holds and, if $p = p_n$ is a sequence of probabilities such that*

$$c_k \equiv c_k(p) = k/(np) \rightarrow \xi \in (0, \infty], \quad \text{as } n \rightarrow \infty, \quad (2.7)$$

then, with

$$C_k \equiv C_k(p) := (pY_{n-k:n})^{-1} = \frac{nc_k}{kY_{n-k:n}}, \quad (2.8)$$

and $VaR_p = U(1/p)$,

$$\frac{VaR_p / C_k^\gamma - 1}{A(n/k)} \xrightarrow{p} \frac{\xi^\rho - 1}{\rho} =: h_\rho(\xi), \quad (2.9)$$

as $n \rightarrow \infty$. If we furthermore assume that $x = x_n$ is a sequence of high levels such that

$$\tilde{C}_k \equiv \tilde{C}_k(x) = x/X_{n-k:n} \xrightarrow{p} \tilde{\xi} \in (0, \infty], \quad \text{as } n \rightarrow \infty, \quad (2.10)$$

and the rate function \tilde{A} satisfies $\gamma^2 \tilde{A}(t) = A(1/\bar{F}(t))$, with A given in (2.3), then, as $n \rightarrow \infty$,

$$\frac{\bar{F}(x) / \tilde{C}_k^{-1/\gamma} - 1}{\tilde{A}(X_{n-k:n})} \xrightarrow{p} \frac{\tilde{\xi}^{\tilde{\rho}} - 1}{\tilde{\rho}} = h_{\tilde{\rho}}(\tilde{\xi}) \quad (2.11)$$

with $h_\rho(\xi)$ defined in (2.9).

Corollary 2.1. *Under the conditions of Lemma 2.3, if $p = p_n$ is a sequence of probabilities and $x = x_n$ a sequence of high thresholds such that (2.7) and (2.10) hold, then, as $n \rightarrow \infty$,*

$$\frac{\tilde{C}_k(\text{VaR}_p)}{c_k^\gamma(p)} = 1 + o_p(1) \quad \text{as well as} \quad \frac{\tilde{C}_k(x)}{c_k^\gamma(\overline{F}(x))} = 1 + o_p(1).$$

More generally than Theorem 4.1 in Beirlant *et al.* (2008), and Proposition 5 in Caeiro and Gomes (2009), we now state the following theorem.

Theorem 2.1. *Under the conditions of Lemma 2.3, let $\hat{\gamma}_{k,n}^\bullet$ be any semi-parametric estimator of γ , such that, with Z_k^\bullet asymptotically standard normal, $\sigma_\bullet > 0$ and $b_\bullet \in \mathbb{R}$,*

$$\hat{\gamma}_{k,n}^\bullet \stackrel{d}{=} \gamma + \frac{\sigma_\bullet Z_k^\bullet}{\sqrt{k}} + b_\bullet A(n/k)(1 + o_p(1)), \quad \text{as } n \rightarrow \infty. \quad (2.12)$$

Consequently, if $\sqrt{k}A(n/k) \rightarrow \lambda_A$, finite

$$\sqrt{k}(\hat{\gamma}_{k,n}^\bullet - \gamma) \xrightarrow[n \rightarrow \infty]{d} N(\lambda_A b_\bullet, \sigma_\bullet^2) =: \mathcal{G}_\bullet. \quad (2.13)$$

Let us further assume that $p = p_n$ is a sequence of probabilities such that, as $n \rightarrow \infty$,

- $c_k \equiv c_k(p) = k/(np) \rightarrow \xi$, $\xi \in (0, \infty]$,
- $\ln c_k = o(\sqrt{k})$,
- $r_k A(n/k) \rightarrow \eta_A \in \mathbb{R}$,

then, with $\text{VaR}_p = F^-(1-p) = U(1/p)$, $\hat{Q}_{k,n}^H(p)$, $\hat{Q}_{k,n}^{PPWM}(p)$, generally denoted $\hat{Q}_{k,n}^\bullet(p)$, \mathcal{B} and \mathcal{G}_\bullet given in (1.1), (1.6), (1.13), (2.5) and (2.13), respectively, $\rho < 0$ and $h_\rho(\xi)$ given in (2.9),

$$r_k (\ln \hat{Q}_{k,n}^\bullet(p) - \ln \text{VaR}_p) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_p = \begin{cases} (\ln \xi) \mathcal{G}_\bullet + \gamma \mathcal{B} - \eta_A h_\rho(\xi) & \text{if } \xi \in (0, \infty) \\ \mathcal{G}_\bullet + \eta_A / \rho & \text{if } \xi = \infty, \end{cases}$$

with

$$r_k = \begin{cases} \sqrt{k} & \text{if } \xi \in (0, \infty) \\ \sqrt{k} / \ln c_k & \text{if } \xi = \infty. \end{cases}$$

If $\xi = \infty$ and $\sqrt{k}A(n/k) \rightarrow \lambda_A$, finite, then,

$$\frac{\sqrt{k}}{\ln c_k} \left(\frac{\hat{Q}_{k,n}^\bullet(p)}{\text{VaR}_p} - 1 \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{G}_\bullet, \quad (2.14)$$

given in (2.13). We can indeed write the asymptotic distributional representation,

$$\hat{Q}_{k,n}^\bullet(p) \stackrel{d}{=} \text{VaR}_p \left\{ 1 + \ln c_k (\hat{\gamma}_{k,n}^\bullet - \gamma) (1 + o_p(1)) + \frac{\gamma \mathcal{B}}{\sqrt{k}} + \frac{A(n/k)(1 + o_p(1))}{\rho} \right\}. \quad (2.15)$$

If on the other hand, $x = x_n$ is a sequence of levels such that, as $n \rightarrow \infty$,

- $\tilde{C}_k \equiv \tilde{C}_k(x) := x/X_{n-k:n} \xrightarrow{p} \tilde{\xi} \in (0, \infty]$,
- $\ln \tilde{C}_k = o_p(\sqrt{k})$,
- $\tilde{R}_k A(n/k) \xrightarrow{p} \tilde{\eta}_A \in \mathbb{R}$,

then, similarly as before, with p_x , $\hat{p}_{k,n}^H(x)$ and $\hat{p}_{k,n}^{PPWM}(x)$, generally denoted $\hat{p}_{k,n}^\bullet(x)$, defined in (1.2), (1.8) and (1.14), respectively,

$$\tilde{R}_k(\ln \hat{p}_{k,n}^\bullet(x) - \ln p) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_x = \begin{cases} (\ln \tilde{\xi})\mathcal{G}_\bullet/\gamma^2 + \mathcal{B} - \tilde{\eta}_A h_{\tilde{\rho}}(\tilde{\xi})/\gamma^2 & \text{if } \tilde{\xi} \in (0, \infty) \\ (\mathcal{G}_\bullet + \tilde{\eta}_A/\tilde{\rho})/\gamma^2 & \text{if } \tilde{\xi} = \infty, \end{cases}$$

with

$$\tilde{R}_k = \begin{cases} \sqrt{k} & \text{if } \tilde{\xi} \in (0, \infty) \\ \sqrt{k}/\ln \tilde{C}_k & \text{if } \tilde{\xi} = \infty. \end{cases}$$

Just as before, if $\tilde{\xi} = \infty$ and $\sqrt{k}A(n/k) \rightarrow \lambda_A$, finite,

$$\frac{\sqrt{k}}{\ln \tilde{C}_k} \left(\frac{\hat{p}_{k,n}^\bullet(x)}{p} - 1 \right) \xrightarrow[n \rightarrow \infty]{d} \frac{\mathcal{G}_\bullet}{\gamma^2}. \quad (2.16)$$

Indeed, as $n \rightarrow \infty$, we have the validity of the asymptotic distributional representation,

$$\hat{p}_{k,n}^\bullet(x) \stackrel{d}{=} p \left\{ 1 + \frac{\ln \tilde{C}_k}{\gamma^2} (\hat{\gamma}_{k,n}^\bullet - \gamma)(1 + o_p(1)) + \frac{\mathcal{B}}{\sqrt{k}} + \frac{\tilde{A}(n/k)(1 + o_p(1))}{\tilde{\rho}} \right\}. \quad (2.17)$$

Remark 2.2. The asymptotic dominant behaviour of $\hat{Q}_{k,n}^\bullet(p)$ and $\hat{p}_{k,n}^\bullet(x)$ is thus fully determined by the asymptotic behaviour of $\hat{\gamma}_{k,n}^\bullet$.

Remark 2.3. In a certain dualistic way, the above conditions on the sequence of probabilities $p = p_x$ and the Value-at-Risk, VaR_p , regarded as a level $\text{VaR}_p = x = x_n$ can be interchanged. From the second half of Corollary 2.1, for instance, when $x = x_n$ is such that $\tilde{C}_k = \tilde{C}_k(x)$ satisfies the conditions of Theorem 2.1, it is readily understood that the sequence $c_k \equiv c_k(\bar{F}(x)) = k/(n \bar{F}(x))$ satisfies the condition $\tilde{C}_k^{1/\gamma}(x)/c_k(\bar{F}(x)) \xrightarrow{p} 1$, as $n \rightarrow \infty$, and consequently, also the conditions on c_k assumed in Theorem 2.1. Alternatively, when $p = p_n$ is such that $c_k = c_k(p)$ satisfies the conditions in Theorem 2.1, from the first half of Corollary 2.1 we see that $\tilde{C}_k(U(1/p))/c_k^\gamma(p) \xrightarrow{p} 1$. Consequently, the conditions on $\tilde{C}_k(U(1/p))$ assumed in Theorem 2.1 also hold.

Remark 2.4. Because of this dualistic interpretation, i.e. more specifically from the fact that $\tilde{C}_k(x)/c_k^\gamma(\bar{F}(x)) \xrightarrow{p} 1$, as $n \rightarrow \infty$, it is also easily understood how, for the exceedance probability estimator, $\hat{p}_{k,n}^\bullet(x)$, and with

$$r_k(x) = \begin{cases} \sqrt{k} & \text{if } \tilde{\xi} \in (0, \infty) \\ \sqrt{k}/\ln c_k(\bar{F}(x)) & \text{if } \tilde{\xi} = \infty, \end{cases}$$

we can state in an equivalent way that, as $n \rightarrow \infty$, $r_k(x)(\ln \hat{p}_{k,n}^\bullet - \ln p) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_x/\gamma$.

3 Simulation results

In this section, we have implemented a multi-sample Monte Carlo simulation experiment of size 5000×10 , to obtain the distributional behaviour of the new VaR-estimators $\hat{Q}_{k,n}^{PPWM}(p)$, in (1.13), comparatively with the behaviour of the classical estimators $\hat{Q}_{k,n}^H(p)$, in (1.6), for $p = 1/n$, $n = 50, 100, 200, 500, 1000, 2000, 5000, 10000$ and 20000 , from the following underlying parents: the Fréchet(γ) parent, with d.f. $F(x) = \exp(-x^{-1/\gamma})$, $x > 0$, $\gamma > 0$ ($\rho = -1$) and the Burr(γ, ρ) parent, with d.f. $F(x) = 1 - (1 + x^{-\rho/\gamma})^{-1/\rho}$, $x > 0$. For the same models, we have also dealt with the dual problem of estimation of p (known to be equal to $1/n$), the probability of exceedance of a high level $x = VaR_p$, through the new estimator $\hat{p}_{k,n}^{PPWM}(x)$, in (1.14), comparatively with the behaviour of the classical estimators $\hat{p}_{k,n}^H(x)$, in (1.8).

3.1 Simulated mean values and root mean squared errors of the VaR-estimators

To illustrate the finite sample behaviour of the VaR-estimators, we present, in Figures 1 and 2, the simulated mean values (E) and root mean square errors (RMSE) patterns of the normalized values of $\hat{Q}_{k,n}^H(p)$ and $\hat{Q}_{k,n}^{PPWM}(p)$, denoted \bar{Q}^H and \bar{Q}^{PPWM} , respectively, as functions of k , the number of top o.s. used, for $p = 1/n$ and $n = 500$, for Fréchet(0.25) and Burr(0.25, -0.75) parents, respectively.

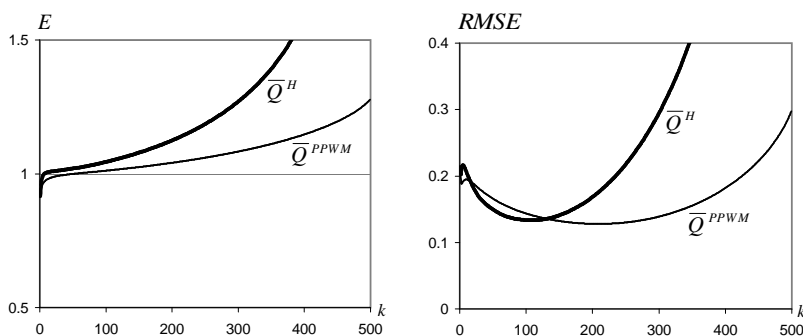


Figure 1: Fréchet(γ), with $\gamma = 0.25$, for $n = 500$ and $p = 1/n$.

3.1.1 Finite sample behaviour of the VaR-estimators at simulated optimal levels

In Table 1, we present the simulated mean values of the above mentioned normalized VaR_p -estimators, H and PPWM, at their simulated optimal levels k_0^H and k_0^{PPWM} , respectively, for $p = 1/n$. In Table 2 and Table 3 we present the simulated relative efficiencies (*REFF*) of $\hat{Q}_{k,n}^{PPWM}(p)$, $p = 1/n$, comparatively with the Weissman-Hill estimator, whenever computed at their simulated optimal levels, i.e., the

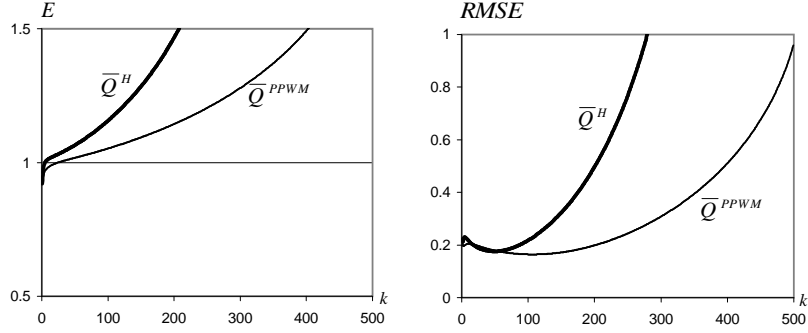


Figure 2: Burr(γ, ρ), with $(\gamma, \rho) = (0.25, -0.75)$ for $n = 500$ and $p = 1/n$.

simulated values of

$$REFF_{PPWM|H} = RMSE\left(\hat{Q}_{n,k_0^H}^H(p)\right) / RMSE\left(\hat{Q}_{n,k_0^{PPWM}}^{PPWM}\right),$$

and the RMSE's, respectively, of the above mentioned normalized VaR_p -estimators, again at their simulated optimal levels, for $p = 1/n$. In all tables the “best” values are written in **bold**.

Table 1: Simulated mean values of the normalized VaR_p -estimators under study, at their simulated optimal levels, for $p = 1/n$

n	50	100	200	500	1000	2000	5000	10000	20000
Burr(0.25, -0.2)									
H	0.921	0.931	0.938	0.945	0.948	0.950	0.980	1.083	1.087
$PPWM$	0.824	0.985	1.036	1.032	1.040	1.050	1.062	1.068	1.072
Burr(0.5, -0.5)									
H	0.932	0.928	0.932	1.017	1.175	1.164	1.148	1.137	1.125
$PPWM$	0.770	0.857	1.034	1.079	1.049	1.005	1.031	1.019	1.104
Burr(0.25, -0.75)									
H	0.959	0.962	1.058	1.066	1.062	1.057	1.050	1.046	1.042
$PPWM$	1.013	1.041	1.053	1.057	1.057	1.055	1.049	1.045	1.040
Fréchet(0.25)									
H	1.001	1.044	1.050	1.050	1.045	1.042	1.036	1.030	1.027
$PPWM$	1.031	1.038	1.042	1.044	1.043	1.041	1.035	1.031	1.027
Burr(0.25, -1.5)									
H	1.035	1.051	1.052	1.045	1.040	1.034	1.028	1.023	1.020
$PPWM$	1.033	1.039	1.044	1.044	1.040	1.036	1.029	1.024	1.021
Burr(0.75, -1.5)									
H	0.990	1.247	1.220	1.168	1.144	1.117	1.090	1.075	1.063
$PPWM$	0.724	0.769	0.823	0.900	0.962	1.024	1.109	1.069	0.896

Table 2: Values of the indicator $REFF_{PPWM|H}$ for Var_p estimation, $p = 1/n$.

Model(γ, ρ)	$\gamma + \rho$	50	100	200	500	1000	2000	5000	10000	20000
Burr(0.25, -0.2)	+0.05	1.381	1.999	1.299	1.207	1.181	1.181	1.219	1.180	1.158
Burr(0.5, -0.5)	-0.00	1.659	2.181	2.535	1.516	1.310	1.144	0.977	0.873	0.800
Burr(0.25, -0.75)	-0.50	1.107	1.121	1.102	1.067	1.046	1.032	1.027	1.024	1.021
Fréchet(0.25, -1)	-0.75	1.118	1.091	1.064	1.041	1.019	1.005	0.998	0.996	0.995
Burr(0.25, -1.5)	-1.25	1.103	1.071	1.029	1.001	0.992	0.980	0.976	0.974	0.971
Burr(0.75, -1.5)	-0.75	2.209	2.176	1.945	1.586	1.289	0.995	0.679	0.491	0.393

Table 3: Simulated RMSE's of the normalized Var_p -estimators under study, at their simulated optimal levels, for $p = 1/n$

	50	100	200	500	1000	2000	5000	10000	20000
	Burr(0.25, -0.2)								
H	0.372	0.340	0.316	0.292	0.279	0.268	0.254	0.235	0.217
$PPWM$	0.270	0.170	0.243	0.242	0.236	0.227	0.211	0.199	0.188
	Burr(0.5, -0.5)								
H	0.537	0.516	0.505	0.485	0.444	0.391	0.329	0.290	0.255
$PPWM$	0.324	0.237	0.199	0.320	0.339	0.342	0.337	0.333	0.321
	Burr(0.25, -0.75)								
H	0.220	0.217	0.204	0.175	0.155	0.136	0.113	0.098	0.085
$PPWM$	0.199	0.194	0.185	0.164	0.149	0.132	0.110	0.095	0.083
	Fréchet(0.25)								
H	0.204	0.182	0.160	0.133	0.115	0.098	0.079	0.067	0.057
$PPWM$	0.183	0.167	0.151	0.128	0.113	0.098	0.080	0.067	0.057
	Burr(0.25, -1.5)								
H	0.200	0.178	0.156	0.128	0.109	0.091	0.071	0.058	0.048
$PPWM$	0.181	0.167	0.152	0.128	0.110	0.093	0.073	0.060	0.050
	Burr(0.75, -1.5)								
H	0.810	0.729	0.603	0.466	0.383	0.309	0.234	0.189	0.154
$PPWM$	0.367	0.335	0.310	0.294	0.297	0.311	0.345	0.385	0.391

Some remarks:

1. Regarding mean values at optimal levels, the PPWM Var_p estimator performs better than the classical Weissman-Hill estimator, in a large variety of situations.
2. Regarding RMSE, or equivalently REFF-indicators, $\hat{Q}_{k,n}^{PPWM}$ performs better than the Weissman-Hill quantile estimator at optimal levels, for all n , provided that $\gamma + \rho > 0$.
3. The new quantile estimator $\hat{Q}_{k,n}^{PPWM}$ can be used as an alternative to the Weissman-Hill estimator, specially for small to moderate sample sizes, but also for large samples, in most situations. All the simulated results suggest that its RMSE is never much bigger than the one of Weissman-Hill's

estimator RMSE and it has often a smaller bias than the Weissman-Hill estimator. At their optimal levels and for small sample sizes, $\hat{Q}_{k,n}^{PPWM}$ is usually much more efficient than Weissman-Hill's estimator.

4. The asymptotic properties of the new PPWM quantile estimators do not hold for the Burr model with $(\gamma, \rho) = (0.75, -1.5)$. But even in this example, the PPWM estimators are more efficient than the Weissman-Hill estimators at their simulated optimal levels, for small up to moderate sample sizes.

3.2 Simulated mean values and root mean squared errors of the p_x -estimators

Figures 3 and 4 are equivalent to Figures 1 and 2, respectively, but for the estimation of $p = p_x = 1 - F(x) = 1/n$, i.e. $x = (-\ln(1 - 1/n))^{-\gamma}$ for the Fréchet(γ) parent, and $x = (n^{-\rho} - 1)^{-\gamma/\rho}$ for the Burr(γ, ρ) parent. In Figure 3 and Figure 4, we have considered a sample size $n = 500$, from Fréchet(0.25) and Burr(0.5, -0.5) parents, respectively.

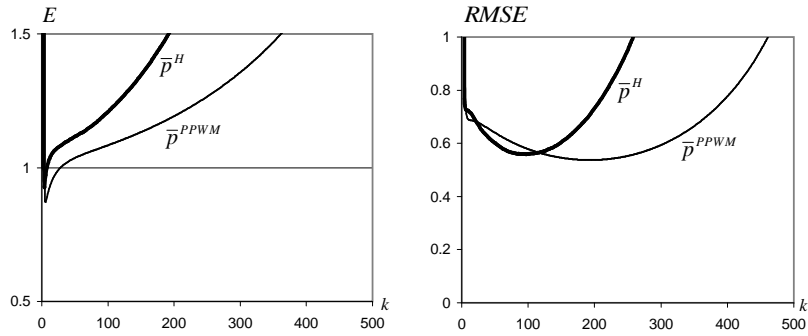


Figure 3: Fréchet(γ), with $\gamma = 0.25$, for $n = 500$ and $p = 1/n$.

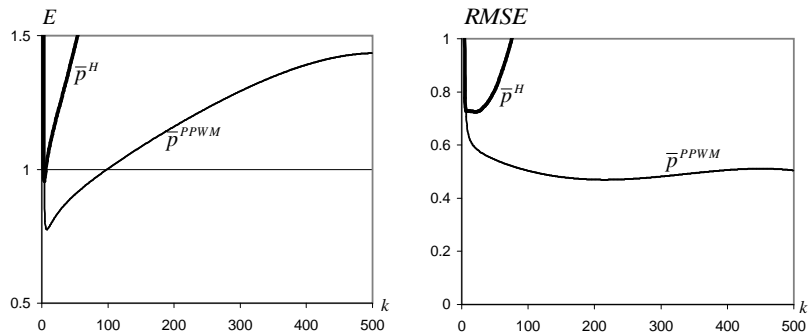


Figure 4: Burr(γ, ρ), with $(\gamma, \rho) = (0.5, -0.5)$, for $n = 500$ and $p = 1/n$.

3.2.1 Finite sample behaviour of the p_x -estimators at simulated optimal levels

Tables 4, 5 and 6 are equivalent to Tables 1, 2 and 3, respectively. For the estimation of p_x , we have now considered $x = F^{\leftarrow}(1 - 1/n)$, and have decided to work with the normalized estimators $\bar{p}^H = \hat{p}_{n,k_0}^H(x)/p_x$ and $\bar{p}^{PPWM} = \hat{p}_{n,k_0}^{PPWM}(x)/p_x$.

Table 4: Simulated mean values of the normalized p_x -estimators under study, at their simulated optimal levels, for $x = F^{\leftarrow}(1 - 1/n)$

	50	100	200	500	1000	2000	5000	10000	20000
	Burr(0.25, -0.2)								
H	1.039	1.040	1.027	1.154	1.230	1.256	1.252	1.252	1.237
$PPWM$	0.671	0.823	1.039	1.182	1.139	1.127	1.138	1.152	1.176
	Burr(0.5, -0.5)								
H	1.039	1.040	1.027	1.154	1.230	1.256	1.252	1.252	1.237
$PPWM$	0.671	0.823	1.039	1.182	1.139	1.127	1.138	1.152	1.176
	Burr(0.25, -0.75)								
H	1.037	1.053	1.176	1.230	1.237	1.236	1.211	1.197	1.180
$PPWM$	1.154	1.146	1.170	1.209	1.229	1.228	1.211	1.194	1.175
	Fréchet(0.25)								
H	1.069	1.171	1.192	1.196	1.185	1.177	1.150	1.130	1.117
$PPWM$	1.125	1.136	1.160	1.184	1.185	1.176	1.154	1.132	1.114
	Burr(0.25, -1.5)								
H	1.132	1.176	1.197	1.185	1.169	1.145	1.116	1.098	1.083
$PPWM$	1.145	1.153	1.176	1.187	1.176	1.156	1.126	1.104	1.087
	Burr(0.75, -1.5)								
H	1.132	1.176	1.197	1.185	1.169	1.145	1.116	1.098	1.083
$PPWM$	0.648	0.704	0.772	0.867	0.944	1.022	1.129	1.187	1.091

Table 5: Values of the indicator $REFF_{PPWM|H}$, for p_x estimation, $x = F^{\leftarrow}(1 - 1/n)$.

Model(γ, ρ)	$\gamma + \rho$	50	100	200	500	1000	2000	5000	10000	20000
Burr(0.25, -0.2)	+0.05	1.809	3.054	1.584	1.325	1.252	1.206	1.181	1.162	1.148
Burr(0.5, -0.5)	-0.00	1.638	2.391	2.935	1.538	1.323	1.183	1.038	0.938	0.861
Burr(0.25, -0.75)	-0.50	1.187	1.128	1.097	1.065	1.042	1.024	1.017	1.015	1.016
Fréchet(0.25, -1)	-0.75	1.122	1.098	1.070	1.038	1.011	0.998	0.990	0.989	0.991
Burr(0.25, -1.5)	-1.25	1.105	1.076	1.034	0.994	0.983	0.968	0.966	0.966	0.966
Burr(0.75, -1.5)	-0.75	1.505	1.624	1.657	1.539	1.341	1.085	0.760	0.550	0.420

Despite of the fact that there is not a clear agreement between the behaviour of the PPWM p_x and VaR_p -estimators, the final conclusions do not differ too much from the one we have previously drawn for the PPWM VaR_p -estimator.

1. The pattern for the PPWM p_x estimator is not so clear-cut as the one of the PPWM VaR_p

Table 6: Simulated RMSE's of the normalized p_x -estimators under study, at their simulated optimal levels, for $x = F^{\leftarrow}(1 - 1/n)$

	50	100	200	500	1000	2000	5000	10000	20000
	Burr(0.25, -0.2)								
H	0.673	0.700	0.718	0.722	0.707	0.668	0.603	0.549?	0.497?
$PPWM$	0.411	0.293	0.245	0.470	0.534	0.565	0.581	0.586	0.579
	Burr(0.5, -0.5)								
H	0.673	0.700	0.718	0.722	0.707	0.668	0.603	0.549	0.497
$PPWM$	0.411	0.293	0.245	0.470	0.534	0.565	0.581	0.586	0.579
	Burr(0.25, -0.75)								
H	0.693	0.715	0.722	0.683	0.634	0.570	0.482	0.420	0.366
$PPWM$	0.584	0.634	0.658	0.642	0.609	0.557	0.474	0.414	0.360
	Fréchet(0.25)								
H	0.710	0.688	0.641	0.558	0.488	0.423	0.341	0.287	0.241
$PPWM$	0.632	0.626	0.599	0.537	0.483	0.424	0.345	0.290	0.243
	Burr(0.25, -1.5)								
H	0.703	0.684	0.635	0.542	0.467	0.390	0.303	0.248	0.203
$PPWM$	0.636	0.635	0.614	0.546	0.475	0.403	0.314	0.257	0.210
	Burr(0.75, -1.5)								
H	0.703	0.684	0.635	0.542	0.467	0.390	0.303	0.248	0.203
$PPWM$	0.467	0.421	0.383	0.352	0.348	0.360	0.400	0.451	0.485

estimator. The estimator \hat{p}_x^H outperforms \hat{p}_x^{PPWM} for a large variety of values of n (both small and large), whenever $|\rho| \geq 1$. For $|\rho| < 1$, and large n , \hat{p}_x^{PPWM} performs generally better than \hat{p}_x^H .

- Regarding RMSE, or equivalently REFF-indicators, \hat{p}_x^{PPWM} performs better than \hat{p}_x^H at optimal levels, for small values of n .
- Just as before, the new quantile estimator \hat{p}_x^{PPWM} can be used as an alternative to \hat{p}_x^H , specially for small to moderate sample sizes, but also for large samples, in most situations. All the simulated results suggest that the RMSE of \hat{p}_x^{PPWM} is never much bigger than the one of \hat{p}_x^H . At their optimal levels and for small sample sizes, \hat{p}_x^{PPWM} is usually much more efficient than \hat{p}_x^H .
- Again as before, for the PPWM Var_p estimation, the asymptotic properties of \hat{p}_x^{PPWM} do not hold for the Burr model with $(\gamma, \rho) = (0.75, -1.5)$. But even in this example, the \hat{p}_x^{PPWM} estimators are more efficient than the \hat{p}_x^{PPWM} estimators at their simulated optimal levels, for small to moderate sample sizes.
- Note also that for a fixed ρ and the Burr(γ, ρ) model, $\hat{p}_{k,n}^H(x)$ does not depend on γ .

4 The bootstrap methodology in the estimation of optimal thresholds

We now put forward an adaptive choice of thresholds, again on the basis of bootstrap computer-intensive methods. Indeed, due to the specificity of these PPWM estimators, and contrarily to what happens with the Hill estimators, in (1.5), the most common estimators of a positive EVI, a direct estimation of the optimal sample fraction (OSF), done on the basis of estimates of scale and shape second-order parameters, is problematic. The use of bootstrap computer intensive methods helps us to provide an adaptive choice of the optimal number of o.s.'s to be used in the estimation, and will be the topic of discussion in this section.

Again with $\hat{\gamma}_{k,n}^\bullet$ denoting either the Hill or the PPWM EVI-estimators, in (1.5) and (1.12), respectively, let us use the notation

$$k_0 \equiv k_0^\bullet(n) := \arg \min_k MSE(\hat{\gamma}_{k,n}^\bullet).$$

Under a semi-parametric framework, and under the validity of the second-order condition, in (2.3), with $\rho < 0$, let us parameterize the function $A(\cdot)$ as $A(t) = \gamma\beta t^\rho$, where $\beta \neq 0$ and $\rho < 0$ are generalized scale and shape second-order parameters. Then, with \mathbb{E} denoting the mean value operator, a possible substitute for the MSE of any classical EVI-estimator $\hat{\gamma}_{k,n}^\bullet$ is, cf. equation (2.12),

$$AMSE(\hat{\gamma}_{k,n}^\bullet) := \mathbb{E}(\sigma_\bullet Z_k^\bullet / \sqrt{k} + b_\bullet A(n/k))^2 = \sigma_\bullet^2/k + b_\bullet^2 \gamma^2 \beta^2 (n/k)^{2\rho},$$

depending on n and k , and with $AMSE$ standing for *asymptotic mean squared error*. We get (Dekkers and de Haan, 1993),

$$k_{0|\bullet}(n) := \arg \min_k AMSE(\hat{\gamma}_{k,n}^\bullet) = ((-2\rho) b_\bullet^2 \gamma^2 \beta^2 n^{2\rho} / \sigma_\bullet^2)^{-1/(1-4\rho)} = k_0^\bullet(n)(1 + o(1)).$$

For the Hill estimator, we have, in (2.12), $\sigma_H = \gamma$ and $b_H = 1/(1-\rho)$. Consequently, with $(\hat{\beta}, \hat{\rho})$ any consistent estimator of the vector (β, ρ) of second-order parameters, we have an asymptotic justification for the estimator

$$\hat{k}_0^H := [((1-\hat{\rho})^2 n^{-2\hat{\rho}} / (-2\hat{\rho}\hat{\beta}^2))^{1/(1-2\hat{\rho})}], \quad (4.1)$$

where, as usual, $[x]$ denotes the integer part of x . Moreover, provided that $\sqrt{k} (n/k)^\rho \rightarrow \lambda$, finite, and with $b_{k,n,\rho} = 1 + \beta(n/k)^\rho / (1-\rho)$, $\sqrt{k} \{\hat{\gamma}_{k,n}^H / \gamma - b_{k,n,\rho}\}$ is approximately $\mathcal{N}(0, 1)$. We may then get approximate 100(1- α)% confidence intervals (CI's) for γ ,

$$\left(\hat{\gamma}_{k,n}^H / (b_{k,n,\rho} + \xi_{1-\alpha/2} / \sqrt{k}), \hat{\gamma}_{k,n}^H / (b_{k,n,\rho} - \xi_{1-\alpha/2} / \sqrt{k}) \right), \quad (4.2)$$

where ξ_p is the p -quantile of a $\mathcal{N}(0, 1)$ d.f. If $\lambda = 0$, we may replace in (4.2) the bias summand $\beta(n/k)^\rho / (1-\rho)$ by 0, i.e., $b_{k,n,\rho} = 1$.

The same does not happen with the PPWM EVI-estimators, with an asymptotic variance (σ_{PPWM}^2) and a dominant component of bias (b_{PPWM}) dependent on γ (see equation (2.6)). In this situation, it is sensible to use the bootstrap methodology for the adaptive PPWM EVI-estimation. Just as in Gomes and Oliveira (2001), for the estimation of γ through the Hill estimator, let us more generally consider the auxiliary statistics,

$$T_{k,n}^\bullet := \hat{\gamma}_{[k/2],n}^\bullet - \hat{\gamma}_{k,n}^\bullet, \quad k = 2, \dots, n-1, \quad (4.3)$$

which converge in probability to the known value zero, for any intermediate k , enabling thus easily the simulation of their MSE through the non-central moment of order two. On the basis of results similar to the ones in Gomes *et al.* (2000) and Gomes and Oliveira (2001), we can get, for the auxiliary statistic $T_{k,n}$, in (4.3), the asymptotic distributional representation,

$$T_{k,n} \equiv T_{k,n}^\bullet \stackrel{d}{=} \sigma_\bullet Q_k^\bullet / \sqrt{k} + b_\bullet (2^\rho - 1) A(n/k) + o_p(A(n/k)),$$

with Q_k asymptotically standard normal, and $(b_\bullet, \sigma_\bullet)$ given in (2.12). The *AMSE* of $T_{k,n}$ is thus minimal at a level $k_{0|T}(n)$ such that $\sqrt{k} A(n/k) \rightarrow \lambda'_A \neq 0$, i.e. a level such that

$$k_{0|\bullet}(n) = k_{0|T}(n) (1 - 2^\rho)^{\frac{1}{1-2\rho}} (1 + o(1)).$$

Then (see the above mentioned papers for further details), given the sample $\underline{X}_n = (X_1, \dots, X_n)$ from an unknown model F , for any $n_1 = o(n)$, and a bootstrap sample $\underline{X}_{n_1}^* = (X_1^*, \dots, X_{n_1}^*)$, denote T_{k_1, n_1}^* , $1 < k_1 < n_1$, the corresponding bootstrap auxiliary statistic. With the obvious notation $k_{0|T}^*(n_1) = \arg \min_{k_1} AMSE(T_{k_1, n_1}^*)$,

$$k_{0|T}^*(n_1) / k_{0|T}(n) = (n_1/n)^{-\frac{2\rho}{1-2\rho}} (1 + o(1)),$$

Consequently, for another sample size $n_2 = n_1^2/n$, we have $(k_{0|T}^*(n_1))^2 / k_{0|T}^*(n_2) = k_{0|T}(n) (1 + o(1))$, as $n \rightarrow \infty$. On the basis of the estimation of $k_{0|T}^*$, we are now able to estimate $k_{0|T}$, and next $k_0^\bullet(n)$. With $\hat{k}_{0|T}^*$ denoting the sample counterpart of $k_{0|T}^*$, we can build the k_0 -estimate,

$$\hat{k}_0^{\bullet*} \equiv \hat{k}_0^{\bullet*}(n; n_1) := \min \left(n-1, \left[\frac{(1-2^\rho)^{\frac{1}{1-2\rho}} (\hat{k}_{0|T}^*(n_1))^2}{\hat{k}_{0|T}^*([n_1^2/n] + 1)} \right] + 1 \right), \quad (4.4)$$

and the γ -estimate

$$\hat{\gamma}^{\bullet*} \equiv \hat{\gamma}^{\bullet*}(n; n_1) := \hat{\gamma}_{\hat{k}_0^{\bullet*}(n; n_1), n}^\bullet. \quad (4.5)$$

Note also that, with $c_k(p)$ defined in (1.6), on the basis of (2.14), for the estimation of Var_p , and on the basis of (2.16) with \tilde{c}_k the observed value of \tilde{C}_k , in (1.8), for the estimation of $p_x = \bar{F}(x)$, we get

$$k_{0|p}^\bullet := \arg \min_k AMSE(\hat{Q}_{k,n}^\bullet(p)) = \arg \min_k (\ln c_k)^2 AMSE(\hat{\gamma}_{k,n}^\bullet),$$

and

$$k_{0|x}^{\bullet} := \arg \min_k AMSE(\hat{p}_{k,n}^{\bullet}(x)) = \arg \min_k (\ln \tilde{c}_k)^2 AMSE(\hat{\gamma}_{k,n}^{\bullet}).$$

A few practical questions, some of them with answers out of the scope of this paper, may be raised under the set-up developed: How does the asymptotic method work for moderate sample sizes? What is the type of the sample path of the new estimator for different values of n_1 ? Is the method strongly dependent on the choice of n_1 ? Although aware of the theoretical need to have $n_1 = o(n)$, what happens if we choose $n_1 = n$? Answers to these questions are expected not to be a long way from the ones given in previous papers (see Hall, 1990; Draisma *et al.*, 1999; Danielsson *et al.* 2001; Gomes and Oliveira, 2001), and some of them will be given in Section 5 of this article, on the basis of the analysis of real and simulated data.

4.1 An algorithm for adaptive estimation of parameters

The estimates $\hat{\rho}$, of the second-order parameter ρ , are the ones already used in previous papers. See for instance the algorithm provided in Gomes and Pestana (2007). Now, with \bullet denoting either H or PWM , the algorithm is the following:

1. Given an observed sample (x_1, x_2, \dots, x_n) , compute the estimates $\hat{\gamma}_{k,n}^{\bullet}$, $\hat{Q}_{k,n}^{\bullet}(p)$ and $\hat{p}_{k,n}^{\bullet}(x)$, $k = 1, 2, \dots, n - 1$.
2. Compute, for the tuning parameters $\tau = 0$ and $\tau = 1$, the observed values of the most simple ρ -estimators introduced and studied in Fraga Alves *et al.* (2003), and also used in the algorithm of Gomes and Pestana (2007). Such estimators depend on the statistics

$$V_{k,n}^{(\tau)} := \begin{cases} \frac{(M_{k,n}^{(1)}) - (M_{k,n}^{(2)}/2)^{1/2}}{(M_{k,n}^{(2)}/2)^{1/2} - (M_{k,n}^{(3)}/6)^{1/3}} & \text{if } \tau = 1 \\ \frac{\ln(M_{k,n}^{(1)}) - \frac{1}{2} \ln(M_{k,n}^{(2)}/2)}{\frac{1}{2} \ln(M_{k,n}^{(2)}/2) - \frac{1}{3} \ln(M_{k,n}^{(3)}/6)} & \text{if } \tau = 0, \end{cases}$$

where

$$M_{k,n}^{(j)} := \frac{1}{k} \sum_{i=1}^k (\ln X_{n-i+1:n} - \ln X_{n-k:n})^j, \quad j \geq 1,$$

$M_{k,n}^{(1)} \equiv H(k)$, in (1.5). They have the functional form

$$\hat{\rho}_{\tau}(k) := \min \left(0, \frac{3(V_{k,n}^{(\tau)} - 1)}{V_{k,n}^{(\tau)} - 3} \right).$$

3. Consider $\{\hat{\rho}_\tau(k)\}_{k \in \mathcal{K}}$, with $\mathcal{K} = ([n^{0.995}], [n^{0.999}])$, compute their median, denoted η_τ , and compute $I_\tau := \sum_{k \in \mathcal{K}} (\hat{\rho}_\tau(k) - \eta_\tau)^2$, $\tau = 0, 1$. Next choose the *tuning parameter* $\tau^* = 0$ if $I_0 \leq I_1$; otherwise, choose $\tau^* = 1$.
4. Work with $\hat{\rho} \equiv \hat{\rho}_{\tau^*} = \hat{\rho}_{\tau^*}(k_1)$, $k_1 = [n^{0.999}]$.
5. Next, consider a sub-sample size $n_1 = o(n)$, and $n_2 = [n_1^2/n] + 1$.
6. For l from 1 until $B = 250$, generate independently B bootstrap samples $(x_1^*, \dots, x_{n_2}^*)$ and $(x_1^*, \dots, x_{n_2}^*, x_{n_2+1}^*, \dots, x_{n_1}^*)$, of sizes n_2 and n_1 , respectively, from the empirical d.f. $F_n^*(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}}$ associated with the observed sample (x_1, \dots, x_n) .
7. Denoting $T_{k,n}^* \equiv T_{k,n}^{\bullet*}$ the bootstrap counterpart of $T_{k,n}^\bullet$, defined in (4.3), obtain $t_{k,n_1,l}^*$, $1 < k < n_1$, $t_{k,n_2,l}^*$, $1 < k < n_2$, $1 \leq l \leq B$, the observed values of the statistic T_{k,n_i}^* , $i = 1, 2$. For $k = 2, \dots, n_i - 1$, compute

$$MSE_1^*(n_i, k) = \frac{1}{B} \sum_{l=1}^B (t_{k,n_i,l}^*)^2, \quad i = 1, 2,$$

as well as

$$MSE_p^*(n_i, k) = (\ln c_k)^2 MSE_1^*(n_i, k), \quad MSE_x^*(n_i, k) = (\ln \tilde{c}_k^2) MSE_1^*(n_i, k), \quad i = 1, 2.$$

8. Obtain, for $i = 1, 2$,

$$\hat{k}_{0|T}^*(n_i) := \arg \min_{1 < k < n_i} MSE_1^*(n_i, k), \quad (4.6)$$

$$\hat{k}_{0|p}^*(n_i) := \arg \min_{1 < k < n_i} MSE_p^*(n_i, k), \quad (4.7)$$

$$\hat{k}_{0|x}^*(n_i) := \arg \min_{1 < k < n_i} MSE_x^*(n_i, k). \quad (4.8)$$

9. On the basis of (4.6), compute the threshold estimate $\hat{k}_0^{\bullet*} \equiv \hat{k}_0^{\bullet*}(n; n_1)$, in (4.4), for the adaptive EVI-estimation. Compute also the threshold estimates, $\hat{k}_{0|p}^{\bullet*} \equiv \hat{k}_{0|p}^{\bullet*}(n; n_1)$ and $\hat{k}_{0|x}^{\bullet*} \equiv \hat{k}_{0|x}^{\bullet*}(n; n_1)$, on the basis of an equation similar to (4.4), but with $\hat{k}_{0|T}^*$ replaced by $\hat{k}_{0|p}^*$ and $\hat{k}_{0|x}^*$, in (4.7) and (4.8), according as we are interested in the adaptive VaR_p -estimation or the p_x -estimation, respectively.
10. Obtain the adaptive EVI-estimate, $\hat{\gamma}_*^\bullet \equiv \hat{\gamma}_*^\bullet(n, n_1) := \hat{\gamma}_{\hat{k}_0^{\bullet*}, n}^\bullet$, already provided in (4.5).
11. Finally, compute the adaptive VaR_p - and p_x -estimates,

$$\hat{Q}_p^{\bullet*} \equiv \hat{Q}_{p|\hat{\gamma}}^{\bullet*} := \hat{Q}_{\hat{k}_{0|p}^{\bullet*}, n}^\bullet, \quad \hat{p}_x^{\bullet*} \equiv \hat{p}_{x|\hat{\gamma}}^{\bullet*} := \hat{p}_{\hat{k}_{0|x}^{\bullet*}, n}^\bullet,$$

with $\hat{k}_{0|p}^{\bullet*}$ and $\hat{k}_{0|x}^{\bullet*}$ obtained in Step 9.

Remarks:

- (i) If there are negative elements in the sample, n should be replaced by $n^+ = \sum_{i=1}^n I_{[X_i > 0]}$ (the number of positive values in the sample). Analogously for n_1 and for n_2 .
- (ii) The Monte-Carlo procedure in Steps **6.** to **11.** of the Algorithm can be replicated r times if we want to associate standard errors to the OSF's estimates and to the γ , VaR_p and p_x bootstrap estimates. The value of B can also be adequately chosen.

5 Applications to simulated and real data

5.1 A simulated data set

In order to have an indication on the way the algorithm in Section 4.1 performs, and to further motivate its use, prior to its validation through a surely highly time-consuming simulation, a topic out of the scope of this paper, we first apply it to an arbitrarily simulated sample, with size $n = 500$, from a Fréchet model, with $\gamma = 0.25$. The algorithm here presented led to the ρ -estimate $\hat{\rho} \equiv \hat{\rho}_0 = -1.173$, obtained at the level $k_1 = [n_0^{0.999}] = 496$. The associated β -estimate is $\hat{\beta} \equiv \hat{\beta}_0 = 0.901$. Then, we got $\hat{k}_0^H = 102$, with \hat{k}_0^H provided in (4.1), and an associated EVI estimate equal to 0.285, clearly over-estimating the true value $\gamma = 0.25$. The associated approximate 95% confidence interval, in (4.2), is (0.2266, 0.3277), with a size 0.1011.

The application of the algorithm presented in Section 4.1 of this paper, with a sub-sample size $n_1 = [n^{0.955}] = 378$, and $B = 250$ bootstrap generations, led us to $\hat{k}_0^{PPWM*} = 52$ and to the adaptive PPWM EVI-estimate $PPWM^* \equiv \hat{\gamma}_*^{PPWM} = 0.254$. This same algorithm applied to the Hill estimators leads us to $\hat{k}_0^{H*} = 167$ and to the adaptive Hill EVI-estimate $H^* \equiv \hat{\gamma}_*^H = 0.284$. These values are pictured in Figure 5, where we also present the estimates under study as a function of k . The most adequate estimate seems neatly to be the one associated with the PPWM methodology, particularly due to the smoothness of the stability region of the estimates as a function of k .

For the estimation of $VaR_{1/(2n)} = VaR_{0.001}$, equal to 5.623 for this Fréchet sample, and through the PPWM estimators, we were led to a choice of k equal to 52 and to the $VaR_{0.001}$ estimate, $\hat{Q}_{0.001} = 5.986$, slightly overestimating the true value of $VaR_{0.001}$. For the estimation of $p = 1/(2n) = 0.001$, the PPWM methodology led us to a k -value equal to 51 and $\hat{p} = 0.0013$, quite close to p . Regarding the classical estimators of VaR_p and p , we were led to a k -value equal to 167, in both cases, and to the estimates 5.638 and 0.002, respectively.

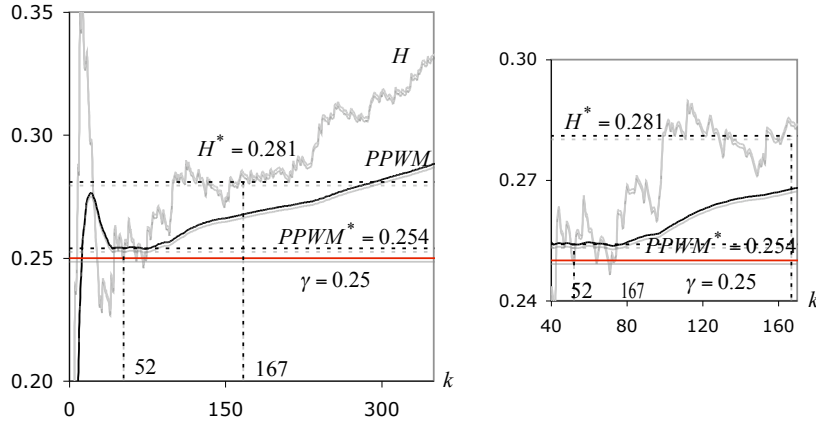


Figure 5: Non-adaptive (as a function of k) and bootstrap adaptive Hill (H) and PPWM EVI-estimates, for a Fréchet sample of size $n = 500$ and $\gamma = 0.25$ (left), and a zoomed figure for $40 \leq k \leq 170$ (right).

5.1.1 Resistance of the methodology to changes in the sub-sample size n_1

In Figure 6, we picture at the left, as a function of the sub-sample size n_1 , ranging from $n_1 = \lceil n^{0.95} \rceil = 366$ until $n_1 = \lceil n^{0.9999} \rceil = 499$, the bootstrap EVI-estimates associated with the Hill and the PPWM estimators, in (1.5) and (1.12), respectively. The pictures in the center and the right are similar to the picture at the left, but related with the corresponding adaptive bootstrap estimation of $Var_{0.001} = 5.623$ and $p = 0.001$, respectively.

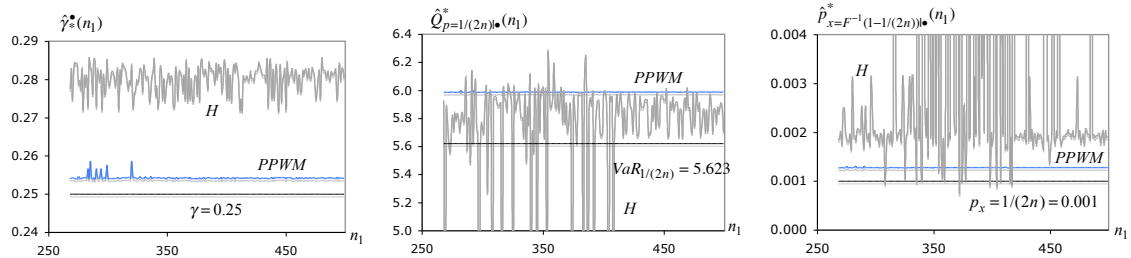


Figure 6: Bootstrap Hill (H) and PPWM EVI-estimates of γ (left), $Var_{0.001} = 5.623$ (center) and $p = 1/(2n) = 0.001$ (right), as functions of n_1 .

The low sensitivity of the PPWM estimates to changes of the subsample size n_1 seems also to be a point in favour of these new estimators. Whereas the bootstrap PPWM EVI, VaR_p and p -estimates lie in the intervals $(0.254, 0.259)$, $(5.977, 6.001)$ and $(0.00127, 0.00131)$, respectively, the bootstrap Hill-type EVI, VaR_p and p -estimates lie in the intervals $(0.272, 0.286)$, $(3.728, 6.254)$ and $(0.0009, 0.0361)$, respectively.

5.2 A case study in the field of insurance

We shall next consider an illustration of the performance of the adaptive PPWM EVI-estimates under study, comparatively with the same methodology applied to the Hill EVI-estimates, again through the analysis of automobile claim amounts exceeding 1,200,000 Euro over the period 1988-2001, gathered from several European insurance companies co-operating with the same re-insurer (Secura Belgian Re). This data set was already studied in Beirlant *et al.* (2004), Vandewalle and Beirlant (2006) and Beirlant *et al.* (2008) as an example to excess-of-loss reinsurance rating and heavy-tailed distributions in car insurance. See also Gomes *et al.* (2009). A preliminary graphical analysis of the data, x_i , $1 \leq i \leq n$, $n = 371$, leads us to an immediate conclusion that data have been censored to the left and that the right-tail of the underlying model is quite heavy. The sample paths of the ρ -estimates associated with $\tau = 0$ and $\tau = 1$ lead us to choose, on the basis of any stability criterion for large k , the estimate associated with $\tau = 0$. The algorithm here presented led us to the ρ -estimate $\hat{\rho}_0 = -0.74$, obtained at the level $k_1 = [n^{0.999}] = 368$. The associated β -estimate, based on the β -estimator in Gomes and Martins (2002), was $\hat{\beta}_0 = 0.80$. Then, we got the estimate $\hat{k}_0^H = 55$, with \hat{k}_0^H provided in (4.1), and an associated adaptive EVI estimate equal to 0.291. The associated approximate 95% confidence interval, in (4.2), is (0.2115, 0.3432), with a size 0.1317.

The application of the algorithm presented in Section 4.1 of this paper, with a sub-sample size $n_1 = [n^{0.955}] = 284$, and $B = 250$ bootstrap generations, led us to $\hat{k}_0^{PPWM*} = 58$ and to the adaptive PPWM EVI-estimate $PPWM^* \equiv \hat{\gamma}_*^{PPWM} = 0.272$. This same algorithm applied to the Hill estimates leads us to $\hat{k}_0^{H*} = 52$ and to the adaptive Hill EVI-estimate $H^* \equiv \hat{\gamma}_*^H = 0.299$. These values are pictured in Figure 7, where we also present the estimates under study as a function of k .

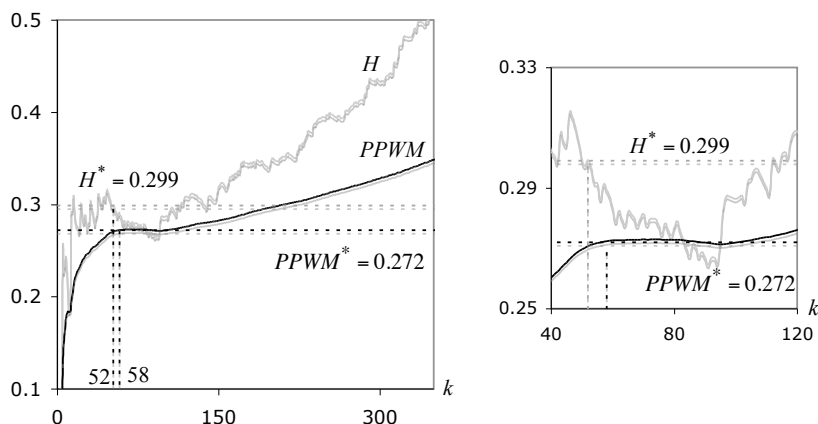


Figure 7: Hill (H) and PPWM EVI-estimates for the SECURA data, as a function of k (left), and a zoomed figure for $40 \leq k \leq 120$ (right).

5.2.1 Resistance of the methodology to changes in the sub-sample size n_1

In Figure 8, we picture at the left, as a function of the sub-sample size n_1 , ranging from $n_1 = [n^{0.95}] = 275$ until $n_1 = [n^{0.9999}] = 370$, the estimates of the OSF for the adaptive bootstrap estimation of γ through the Hill and the PPWM estimators, in (1.5) and (1.12), respectively. The associated bootstrap EVI-estimates are pictured at the right.

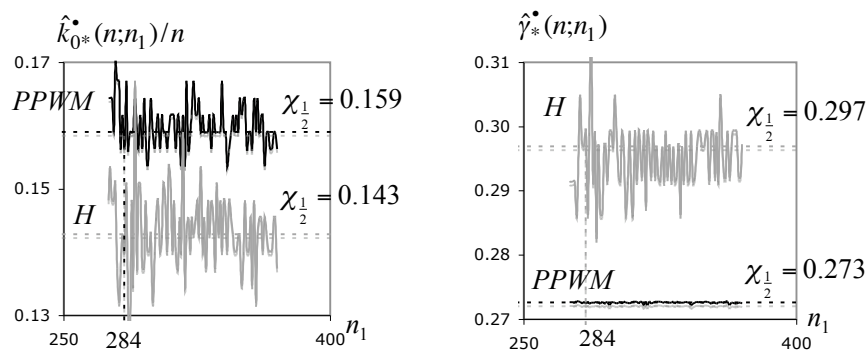


Figure 8: Estimates of the OSF's $\hat{k}_{0^*}^{PPWM}/n$ (left) and the bootstrap adaptive extreme value index estimates $\hat{\gamma}_*^{PPWM}$ (right), as functions of the sub-sample size n_1 , for the SECURA data.

The bootstrap PPWM EVI-estimates are indeed quite stable as a function of the sub-sample size n_1 (see Figure 8, at the right), varying from a minimum value equal to 0.272 until 0.273, with a median equal to 0.273, not a long way from the value obtained for the bootstrap EVI estimate associated to the arbitrarily chosen sub-sample size $n_1 = [n^{0.955}] = 284$, equal to 0.272. We can indeed guarantee the two decimal figures, i.e. the estimate 0.27. The bootstrap estimates of the OSF for the estimation of γ through the PPWM estimator, in (1.12), vary from 15.4% until 17%, with a median equal to 15.9%. For the bootstrap Hill EVI-estimates the volatility is higher. We get estimates from a minimum value equal to 0.283 until a maximum value equal to 0.315. The median of these values is 0.297, also not a long way from the value we obtain for the bootstrap γ -estimate associated to the same arbitrarily chosen sub-sample size $n_1 = 284$, equal to 0.299. The volatility of the OSF's for the Hill estimation is similar to the one we get for the PPWM EVI estimation, ranging from 12.4% until 16.7%, with a median equal to 14.3%. Similar results have been obtained for the PPWM VaR_p and p_x -estimators in this paper. As already detected in previous papers, and in the most diversified comparisons, the Hill estimates are clearly over-estimating the true value of the EVI. As mentioned before, the most adequate estimate seems neatly to be the one associated with the PPWM methodology, particularly due to the smoothness of the stability region of the estimates as a function of k .

6 Proofs

Lemma 2.3. If we combine the results of Lemma 2.1 with the fact that $X_{n-k:n} \stackrel{d}{=} U(Y_{n-k:n})$ and the results in Drees (1998), since

$$\frac{VaR_p}{X_{n-k:n}} = \frac{U(1/p)}{X_{n-k:n}} = \frac{U(C_k Y_{n-k:n})}{U(Y_{n-k:n})},$$

with C_k defined in (2.8), we get

$$\frac{\frac{U(1/p)}{X_{n-k:n}}/C_k^\gamma - 1}{A(Y_{n-k:n})} = \frac{C_k^\rho - 1}{\rho} + o_p(1),$$

as $n \rightarrow \infty$. The first part of the lemma, i.e. equation (2.9), follows from the fact that if (2.7) holds, then, with C_k given in (2.8),

$$C_k = \left(\frac{k Y_{n-k:n}}{n} \right)^{-1} \frac{k}{np} \xrightarrow[n \rightarrow \infty]{p} \xi,$$

as $n \rightarrow \infty$.

Conversely, and due to the equivalence, already mentioned in (2.4), between the condition $U \in 2RV_\gamma^\rho$, with a rate function A and the condition $\bar{F} \in 2RV_{-1/\gamma}^{\tilde{\rho}}$, with $\tilde{\rho} = \rho/\gamma$ and a rate function \tilde{A} that satisfies $\gamma^2 \tilde{A}(t) = A(1/\bar{F}(t))$, a combination of the first part of Lemma 2.2 and the results in Drees (1998), together with the fact that $1/Y_{n-k:n} \stackrel{d}{=} \bar{F}(X_{n-k:n})$, leads to

$$\frac{\frac{\bar{F}(x)}{1/Y_{n-k:n}}/\tilde{C}_k^{-1/\gamma} - 1}{\tilde{A}(X_{n-k:n})} = \frac{\tilde{C}_k^{\tilde{\rho}} - 1}{\tilde{\rho}} + o_p(1),$$

as $n \rightarrow \infty$, with \tilde{C}_k defined in (2.10). Knowing that $|\tilde{A}| \in RV_{\tilde{\rho}}$, $\tilde{\rho} < 0$, the same argument as before enables us to derive (2.11). \square

Theorem 2.1. Regarding the first part of the theorem, and from the definition of $\hat{Q}_{k,n}^H(p)$ and $\hat{Q}_{k,n}^{PPWM}(p)$, given in equation (1.6) and (1.13), respectively, denoted $\hat{Q}_{k,n}^\bullet(p)$, it is clear that

$$\ln \hat{Q}_{k,n}^\bullet(p) - \ln VaR_p = \sqrt{k}(\hat{\gamma}_{k,n}^\bullet - \gamma) \frac{\ln c_k}{\sqrt{k}} - \ln \left(\frac{c_k^{-\gamma} VaR_p}{X_{n-k:n}} \right), \quad (6.1)$$

where $\sqrt{k}(\hat{\gamma}_{k,n}^\bullet - \gamma)$, in the first term on the right-hand side, has been assumed to converge in distribution to a normal r.v. \mathcal{G}_\bullet , in (2.13), with known mean value and variance, whenever $\sqrt{k}A(n/k) \rightarrow \lambda_A$, finite, as $n \rightarrow \infty$.

If $c_k = k/(np) \rightarrow \xi \in (0, \infty]$, as $n \rightarrow \infty$,

$$C_k = \left(\frac{k Y_{n-k:n}}{n} \right)^{-1} \frac{k}{np} \xrightarrow[n \rightarrow \infty]{p} \xi.$$

Using Lemma 2.1 and Cramér's delta method, it follows that

$$\left(\frac{c_k}{C_k}\right)^{-\gamma} = \left(\frac{k Y_{n-k:n}}{n}\right)^{-\gamma} = 1 - \frac{\gamma \mathcal{B}}{\sqrt{k}} + o_p(1/\sqrt{k}),$$

as $n \rightarrow \infty$, where \mathcal{B} is the standard normal r.v. in (2.5). This, together with an application of Lemma 2.3, gives

$$\frac{c_k^{-\gamma} VaR_p}{X_{n-k:n}} = \left(\frac{c_k}{C_k}\right)^{-\gamma} \frac{C_k^{-\gamma} U(1/p)}{X_{n-k:n}} = 1 - \frac{\gamma \mathcal{B}}{\sqrt{k}} + A(n/k)h_\rho(\xi) + o_p(1/\sqrt{k}) + o_p(A(n/k)),$$

which converges to one in probability, as $n \rightarrow \infty$. Making use of the fact that $\ln(1 + \kappa_n) = \kappa_n(1 + o_p(1))$, when $\kappa_n \xrightarrow{p} 0$, as $n \rightarrow \infty$, it can furthermore easily be understood that the second term in the right hand-side of (6.1) can be written as

$$\ln\left(\frac{c_k^{-\gamma} VaR_p}{X_{n-k:n}}\right) = -\frac{\gamma \mathcal{B}}{\sqrt{k}} + A(n/k)h_\rho(\xi) + o_p(1/\sqrt{k}) + o_p(A(n/k)).$$

This result, together with the condition $\ln c_k = o(\sqrt{k})$, immediately leads to the consistency of the quantile estimators $\hat{Q}_{k,n}^\bullet(p)$. In what concerns the asymptotic normality, on the other hand, plugging the last result in equation (6.1), it follows that

$$\begin{aligned} \sqrt{k}(\ln \hat{Q}_{k,n}^\bullet(p) - \ln VaR_p) &= \sqrt{k}(\hat{\gamma}_{k,n}^\bullet - \gamma) \ln c_k + \gamma \mathcal{B} + o_p(1) \\ &\quad - \sqrt{k}A(n/k)h_\rho(\xi) + o_p(\sqrt{k}A(n/k)), \end{aligned} \quad (6.2)$$

which, under the conditions $\xi \in (0, \infty)$ and $\sqrt{k}A(n/k) \rightarrow \eta_A \in \mathbb{R}$, converges in distribution to $(\ln \xi)\mathcal{G}_\bullet + \gamma \mathcal{B} - \eta_A h_\rho(\xi)$, as $n \rightarrow \infty$. In a similar way, it can also be seen that if $\xi = \infty$, the rate of convergence is no longer $1/\sqrt{k}$, but $\ln c_k/\sqrt{k}$, and on the basis of (6.2), we can write

$$\begin{aligned} \frac{\sqrt{k}}{\ln c_k}(\ln \hat{Q}_{k,n}^\bullet(p) - \ln VaR_p) &= \sqrt{k}(\hat{\gamma}_{k,n}^\bullet - \gamma) + \frac{\gamma \mathcal{B}}{\ln c_k} + o_p(1/\ln c_k) \\ &\quad - \frac{\sqrt{k}}{\ln c_k}A(n/k)h_\rho(\xi) + o_p\left(\frac{\sqrt{k}}{\ln c_k}A(n/k)\right). \end{aligned}$$

If $\sqrt{k}A(n/k)/\ln c_k \rightarrow \eta_A \in \mathbb{R}$, since $h_\rho(\xi) = (\xi^\rho - 1)/\rho \rightarrow -1/\rho$, as $\xi \rightarrow \infty$, we get the distributional convergence to $\mathcal{G} + \eta_A/\rho$, as $n \rightarrow \infty$. In this situation, if we assume that $\sqrt{k}A(n/k) \rightarrow \lambda_A$, finite, we get $\eta_A = \lim_{n \rightarrow \infty} \sqrt{k}A(n/k)/\ln c_k = 0$, and a straightforward application of the above-mentioned results leads to (2.14). Also, a straightforward application of Cramer's delta-method leads to (2.15).

Concerning the second part of Theorem 2.1, from the definition of $\hat{p}_{k,n}^H(x)$ and $\hat{p}_{k,n}^{PPWM}(x)$, defined in (1.8) and (1.14), respectively, and generally denoted $\hat{p}_{k,n}^\bullet(x)$, we see that

$$\ln \hat{p}_{k,n}^\bullet(x) - \ln p = \sqrt{k}(\hat{\gamma}_{k,n}^\bullet - \gamma) \frac{\ln \tilde{C}_k}{\gamma^2 \sqrt{k}} - \sqrt{k}(\hat{\gamma}_{k,n}^\bullet - \gamma) \frac{\ln \tilde{C}_k}{\gamma^2 \sqrt{k}} \frac{\hat{\gamma}_{k,n}^\bullet - \gamma}{\hat{\gamma}_{k,n}^\bullet} - \ln\left(\frac{\tilde{C}_k^{1/\gamma} np}{k}\right), \quad (6.3)$$

where $\sqrt{k}(\hat{\gamma}_{k,n}^\bullet - \gamma)$ has been assumed to converge in distribution to the normal r.v. \mathcal{G}_\bullet , whenever $\sqrt{k}A(n/k) \rightarrow \lambda_A$, finite, as $n \rightarrow \infty$. As a result, and due to the assumption that $\ln \tilde{C}_k = o_p(\sqrt{k})$, as $n \rightarrow \infty$, the first and second terms on the right-hand side of (6.3) converge in probability to zero, as $n \rightarrow \infty$. Let us now think on the last term in (6.3). Using Lemma 2.1 in combination with Cramér's delta method, and Lemma 2.3 under the assumption that $x/X_{n-k:n} \xrightarrow{p} \tilde{\xi} \in (0, \infty]$, it is clear that, with $p = \bar{F}(x)$,

$$\tilde{C}_k^{1/\gamma} \left(\frac{np}{k} \right) = \left(\frac{k Y_{n-k:n}}{n} \right)^{-1} \tilde{C}_k^{1/\gamma} \frac{\bar{F}(x)}{1/Y_{n-k:n}} = 1 - \frac{\mathcal{B}}{\sqrt{k}} + \tilde{A}(n/k)h_{\tilde{\rho}}(\tilde{\xi}) + o_p(1/\sqrt{k}) + o_p(\tilde{A}(n/k)),$$

which converges to one in probability, as $n \rightarrow \infty$. Note that \mathcal{B} is the same standard normal distributed r.v. as before. Again,

$$\ln \left(\tilde{C}_k^{1/\gamma} \frac{np}{k} \right) = -\frac{\mathcal{B}}{\sqrt{k}} + \tilde{A}(n/k)h_{\tilde{\rho}}(\tilde{\xi}) + o_p(1/\sqrt{k}) + o_p(\tilde{A}(n/k)),$$

and we then have the consistency of the probability estimators $\hat{p}_{k,n}^\bullet(x)$. In what concerns asymptotic normality, the plugging of this last result in equation (6.3) gives

$$\begin{aligned} \sqrt{k}(\ln \hat{p}_{k,n}^\bullet(x) - \ln p) &= \sqrt{k}(\hat{\gamma}_{k,n}^\bullet - \gamma) \frac{\ln \tilde{C}_k}{\gamma^2} - \sqrt{k}(\hat{\gamma}_{k,n}^\bullet - \gamma) \frac{\ln \tilde{C}_k}{\gamma^2} \frac{\hat{\gamma}_{k,n}^\bullet - \gamma}{\hat{\gamma}_{k,n}^\bullet} \\ &\quad + \mathcal{B} - \sqrt{k}\tilde{A}(n/k)h_{\tilde{\rho}}(\tilde{\xi}) + o_p(1) + o_p(\sqrt{k}\tilde{A}(n/k)). \end{aligned}$$

Under the assumptions that $\tilde{\xi} \in (0, \infty)$ and $\sqrt{k}A(n/k) \rightarrow \tilde{\eta}_A \in \mathbb{R}$, $\sqrt{k}\tilde{A}(n/k) \rightarrow \tilde{\eta}_A/\gamma^2$, and we get convergence in distribution to $(\ln \tilde{\xi})\mathcal{G}_\bullet/\gamma^2 + \mathcal{B} - \tilde{\eta}_A h_{\tilde{\rho}}(\tilde{\xi})/\gamma^2$, as $n \rightarrow \infty$. As before, if $\tilde{\xi} = \infty$,

$$\begin{aligned} \frac{\sqrt{k}}{\ln \tilde{C}_k} (\ln \hat{p}_{k,n}^\bullet(x) - \ln p) &= \sqrt{k}(\hat{\gamma}_{k,n}^\bullet - \gamma) \frac{1}{\gamma^2} - \sqrt{k}(\hat{\gamma}_{k,n}^\bullet - \gamma) \frac{1}{\gamma^2} \frac{\hat{\gamma}_{k,n}^\bullet - \gamma}{\hat{\gamma}_{k,n}^\bullet} \\ &\quad + \frac{\mathcal{B}}{\ln \tilde{C}_k} - \frac{\sqrt{k}\tilde{A}(n/k)h_{\tilde{\rho}}(\tilde{\xi})}{\ln \tilde{C}_k} + o_p(1/\ln \tilde{C}_k) + o_p\left(\frac{\sqrt{k}\tilde{A}(n/k)}{\ln \tilde{C}_k}\right), \end{aligned}$$

which, given that $\tilde{\xi} = \infty$ and $\sqrt{k}A(n/k)/\ln \tilde{C}_k \xrightarrow{p} \tilde{\eta}_A \in \mathbb{R}$, converges in distribution to $\mathcal{G}_\bullet/\gamma^2 + \tilde{\eta}_A \tilde{\rho}/\gamma^2$, as $n \rightarrow \infty$. Just as before, if $\tilde{\xi} = \infty$ and $\sqrt{k}A(n/k) \rightarrow \lambda_A$, finite, $\tilde{\eta}_A = 0$ and (2.16) follows. More specifically, we have then the validity of (2.17). \square

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] Balkema, A.A. and de Haan, L. (1974). Residual Life Time at Great Age. *Annals of Probability*, 2, 792–804.

- [3] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes. Theory and Applications*. Wiley.
- [4] Beirlant, J., Figueiredo, F., Gomes, M.I. and Vandewalle, B. (2008). Improved Reduced-Bias Tail Index and Quantile Estimators. *J. Statist. Plann. and Inference*, 138(6), 1851–1870.
- [5] Caeiro, F. and Gomes, M.I. (2009). Semi-Parametric Second-Order Reduced-Bias High Quantile Estimation. *Test*, 18(2), 392–413.
- [6] Caeiro, F. and Gomes, M.I. (2010). Semi-Parametric Tail Inference through Probability-Weighted Moments. *J. Statist. Plann. Inference*, doi:10.1016/j.jspi.2010.08.015.
- [7] Danielsson, J., Haan, L. de, Peng, L. and de Vries, C.G. (2001). Using a Bootstrap Method to Choose the Sample Fraction in the Tail Index Estimation. *J. Multivariate Analysis*, 76, 226–248.
- [8] Dekkers, A., and de Haan, L. (1993). Optimal Sample Fraction in Extreme Value Estimation. *J. Multivariate Analysis* 47(2), 173–195.
- [9] Diebolt, J., Guillou, A. and Rached, I. (2007). Approximation of the Distribution of Excesses through a Generalized Probability-Weighted Moments Method. *J. Statist Plann. Inference*, 137, 841–857.
- [10] Diebolt, J., Guillou, A., Naveau, P. and Ribereau, P. (2008). Improving Probability-Weighted Moment Methods for the Generalized Extreme Value Distribution. *Revstat*, 6(1), 33–50.
- [11] Draisma, G., de Haan, L., Peng, L. and Pereira, M.T. (1999). A Bootstrap-Based Method to Achieve Optimality in Estimating the Extreme Value Index. *Extremes*, 2 (4), 367–404.
- [12] Drees, H. (1998). On Smooth Statistical Tail Functions. *Scand. J. Statist.*, 25, 187–210.
- [13] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin, Heidelberg.
- [14] 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.
- [15] 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.
- [16] Gnedenko, B.V. (1943). Sur la Distribution Limite du Terme Maximum d’une Série Aléatoire. *Ann. Math.*, 44, 423–453.
- [17] 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.
- [18] 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.
- [19] Gomes, M.I. and Pestana, D. (2007). A Sturdy Reduced-Bias Extreme Quantile (VaR) Estimator. *J. American Statistical Association*, 102(477), 280–292.

- [20] Gomes, M.I., Martins, M.J. and Neves, M.M. (2000). Alternatives to a Semi-Parametric Estimator of Parameters of Rare Events — the Jackknife Methodology. *Extremes*, 3(3), 207–229.
- [21] Gomes, M.I., Canto e Castro, L., Fraga Alves, M.I. and Pestana, D.D. (2008). Statistics of Extremes for IID Data and Breakthroughs in the Estimation of the Extreme Value Index: Laurens de Haan Leading Contributions. *Extremes*, 11(1), 3–34.
- [22] Gomes, M.I., Mendonça, S. and Pestana D. (2009). Adaptive Reduced-Bias Tail Index and Value-at-Risk Estimation. In Sakalauskas L., Skiadas C. and Zavadskas E.K. (eds.). *Applied Stochastic Models and Data Analysis — ASMDA 2009*, IMI & VGTU Editions, 41–44.
- [23] 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 Expressible in Inverse Form. *Water Resources Research*, 15, 1049–1054.
- [24] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: an Introduction*. Springer Science+Business Media, LLC, New York.
- [25] de Haan, L. and Peng, L. (1998). Comparison of Tail Index Estimators. *Statistica Neerlandica*, 52, 60–70.
- [26] Hall, P. (1990). Using the Bootstrap to Estimate Mean-Square Error and Selecting Smoothing Parameter in Non-Parametric Problems. *J. Multivariate Anal.*, 32, 177-203.
- [27] Hill, B.M. (1975). A Simple General Approach to Inference about the Tail of a Distribution. *Ann. Statist.*, 3(5), 1163–1174.
- [28] 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.
- [29] Hosking, J. and Wallis, J. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, 29(3), 339–349.
- [30] Pickands III, J. (1975). Statistical Inference Using Extreme Order Statistics. *Ann. Statist.*, 3, 119–131.
- [31] Reiss, R.-D. and Thomas, M. (2007). *Statistical Analysis of Extreme Values, with Application to Insurance, Finance, Hydrology and Other Fields*, 3rd edition. Birkhäuser Verlag.
- [32] Vandewalle B., Beirlant J. (2006). On Univariate Extreme Value Statistics and the Estimation of Reinsurance Premiums. *Insurance Mathematics and Economics*, 38:3, 444–459.
- [33] Weissman, I. (1978). Estimation of Parameters of Large Quantiles Based on the k Largest Observations. *J. Amer. Statist. Assoc.*, 73, 812–815.