A refined Weissman estimator for extreme quantiles

Michaël Allouche, Jonathan El Methni, Stéphane Girard

A refined Weissman estimator for extreme quantiles

Michaël Allouche[(1)], Jonathan El Methni[(2,⋆)] and Stéphane Girard[(3)]

[(1)] Centre de Mathématiques Appliquées (CMAP), CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau Cedex, France

[(2)] Université Paris Cité, CNRS, MAP5, UMR 8145, 75006 Paris, France

[(⋆)] Corresponding author: jonathan.el-methni@parisdescartes.fr

[(3)] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.

Abstract

Weissman extrapolation methodology for estimating extreme quantiles from heavy-tailed distributions is based on two estimators: an order statistic to estimate an intermediate quantile and an estimator of the tail-index. The common practice is to select the same intermediate sequence for both estimators. In this work, we show how an adapted choice of two different intermediate sequences leads to a reduction of the asymptotic bias associated with the resulting refined Weissman estimator. The asymptotic normality of the latter estimator is established and a data-driven method is introduced for the practical selection of the intermediate sequences. Our approach is compared to Weissman estimator and to six bias reduced estimators of extreme quantiles on a large scale simulation study. It appears that the refined Weissman estimator outperforms its competitors in a wide variety of situations, especially in the challenging high bias cases. Finally, an illustration on an actuarial real data set is provided.

Key words: extreme quantile, bias reduction, heavy-tailed distribution, extreme-value statistics, asymptotic normality.

Mathematics Subject Classification: 60G70, 62G32, 62G20.

Acknowledgments: he authors would like to thank the two referees and the Associate Editor for their valuable suggestions, which have significantly improved the paper. The authors would also like to thank Frederico Caeiro for providing the R code associated with Algorithm 4.1 and Algorithm 4.2 from [27]. This work is supported by the French National Research Agency (ANR) in the framework of the Investissements d'Avenir Program (ANR-15-IDEX-02). S. Girard also acknowledges the support of the Chair Stress Test, Risk Management and Financial Steering, led by the French Ecole Polytechnique and its Foundation and sponsored by BNP Paribas.

Code and data availability: Code is available at https://github.com/michael-allouche/refined-weissman.git. The Secura Belgian reinsurance data set is available in the package CASdatasets of the R software [17].

Declaration - Conflict of interest: The authors declare that they have no conflict of interest.

1 Introduction

Assessing the extreme behaviour of a random phenomenon is a major issue in quantitative finance, assurance and environmental science. For instance, extreme weather events may have strong negative and simultaneous impacts, including loss of life, damages to buildings, decrease of agricultural production, as well as longer term economic consequences [36, 43]. Assuming the phenomenon of interest is modelled by a quantitative random variable X, the associated risk is usually represented by a quantile $q(\alpha)$ such that $\mathbb{P}(X > q(\alpha)) = \alpha$. In finance or insurance, $q(\alpha)$ is referred to as the Value at Risk (VaR), while in environmental sciences, $q(\alpha)$ is referred to as the return level. Focusing on extreme risks, when α is small, the quantile of interest may be larger than the maximal observation. Indeed, denoting by n the size of an independent and identically distributed (i.i.d.) sample and by $X_{n,n}$ the maximal observation, it is easily seen that $\mathbb{P}(X_{n,n} \leq q(\alpha)) = \exp(-n\alpha(1 + o(1))) \to 1$ provided that $n\alpha \to 0$. The empirical cumulative distribution function of X being (relatively) non consistent in such a situation, extrapolation methods are then necessary to estimate the so-called *extreme* quantile $q(\alpha)$.

The celebrated Weissman estimator [46] assumes that the distribution of X is heavy-tailed, *i.e.* the associated survival function $\bar{F}(x) = \mathbb{P}(X > x)$ decays like a power function $x^{-1/\gamma}$ as $x \to \infty$, with $\gamma > 0$, see (1) in the next section for a formal definition. As a consequence, $q(\alpha)$ can be estimated by combining two ingredients: an order statistic and an estimator of the tail-index γ. This extrapolation principle has been adapted to a great variety of situations: light-tailed distributions [18], conditional distributions (to account for covariates) [9, 24] and other risk measures including expectiles [11, 13], M-quantiles [12], Wang risk measures [19], extremiles [10], marginal expected shortfall [7], to cite a few.

Since the reliability of the extrapolations provided by Weissman method heavily depends on the quality of the estimation of the tail-index, a lot of efforts have been made to improve the original Hill estimator [35]. A number of bias reduction techniques for estimating γ have been introduced [6, 25, 26] and their consequences on Weissman estimator have been investigated in [27, 28, 30]. These estimators are described in further details in Section 3. Let us note that all of them are dedicated to the particular Hall-Welsh class of heavy-tailed distributions, see (8) below and [33, 34].

In this work, a different direction is explored to reduce the bias of Weissman estimator. We show that the biases associated with Weissman extrapolation method and the tail-index estimator may asymptotically cancel out in the extreme quantile estimator thanks to an appropriate tuning of the number of upper order statistics involved in the tail-index estimator. The construction of the resulting estimator is presented in Section 2 both from a theoretical and a practical point of view. In particular, an asymptotic normality result is provided, emphasizing that the proposed extreme quantile estimator is asymptotically unbiased in contrast to the original Weissman estimator. Its performances are illustrated on simulated data in Section 3 and compared to state-of-the-art competitors. An illustration on an actuarial real data set is provided in Section 4. Finally, a small conclusion is proposed in Section 5 and the proofs are postponed to the Appendix.

2 A refined Weissman estimator

2.1 Statistical framework

Let $X_1, \ldots, X_n$ be an i.i.d. sample from a cumulative distribution function F and let $X_{1,n} \leq \ldots \leq X_{n,n}$ denote the order statistics associated with this sample. We denote by U the associate tail quantile function defined as $U(t) = F^{\leftarrow}(1 - 1/t)$ for all $t > 1$, where $F^{\leftarrow}(\cdot) = \inf\{x \in \mathbb{R},\ F(x) > \cdot\}$ denotes the generalized inverse of F. In the following it is assumed that F belongs to the maximum domain attraction of Fréchet, which is equivalent to assuming that U is regularly varying with index $\gamma > 0$:

$$\lim_{t\to\infty} \frac{U(ty)}{U(t)} = y^{\gamma}, \tag{1}$$

for all $y > 0$. Recall that, equivalently, U can be rewritten as $U(t) = t^{\gamma} L(t)$ where L is a slowly-varying function, *i.e.* a regularly-varying function with index zero. See [4] for more details on regular variation theory. In such a situation, the distribution associated with U is said to be heavy-tailed and γ is called the tail-index. The goal is to estimate the extreme quantile $q(\alpha_n) = U(1/\alpha_n)$ where $\alpha_n \to 0$ as $n \to \infty$ basing on an intermediate quantile $U(n/k_n)$ where (k_n) is an intermediate sequence *i.e.* such that $k_n \in \{1, \ldots, n-1\}$, $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. In view of the regular variation property (1), one has

$$\frac{U(1/\alpha_n)}{U(n/k_n)} \simeq \left(\frac{k_n}{n\alpha_n}\right)^{\gamma} =: d_n^{\gamma},$$

as $n \to \infty$, where $d_n = k_n/(n\alpha_n)$ is the extrapolation factor. Estimating $U(n/k_n)$ by its empirical counterpart $X_{n-k_n,n}$ and γ by a convenient estimator $\hat{\gamma}_n(k'_n)$ depending on another intermediate sequence (k'_n) yields

$$\hat{q}_n(\alpha_n, k_n, k'_n) = X_{n-k_n,n}\ d_n^{\hat{\gamma}_n(k'_n)}. \tag{2}$$

One can for instance use Hill estimator [35] defined as

$$\mathrm{H}(k'_n) = \frac{1}{k'_n} \sum_{i=1}^{k'_n} \log(X_{n-i+1,n}) - \log(X_{n-k'_n,n}), \tag{3}$$

and choose $k'_n = k_n$ to get the original Weissman estimator [46]:

$$\hat{q}_n(\alpha_n, k_n, k_n) = X_{n-k_n,n}\ d_n^{\mathrm{H}(k_n)}. \tag{4}$$

The asymptotic properties of $\hat{q}_n(\alpha_n, k_n, k_n)$ are established for instance in [31, Theorem 4.3.8]. In the next paragraph, we show that choosing $k'_n \neq k_n$ can yield better results from an asymptotic point of view.

2.2 Asymptotic analysis

The cornerstone in extreme-value analysis for bias assessment is the following second-order condition that refines the initial heavy-tail assumption (1). The tail quantile function U is assumed to be second-order regularly varying with index $\gamma > 0$, second-order parameter $\rho < 0$ and an auxiliary function A having constant sign and converging to 0 at infinity, *i.e.*

$$\lim_{t\to\infty} \frac{1}{A(t)} \left(\frac{U(ty)}{U(t)} - y^\gamma \right) = y^\gamma \frac{y^\rho - 1}{\rho}, \tag{5}$$

for all $y > 0$. The auxiliary function A drives the bias of most extreme-value estimators. In particular, the asymptotic sign of A determines the sign of the asymptotic bias. Besides, necessarily, $|A|$ is regularly varying with index ρ, see for instance [31, Theorem 2.3.9], so that $|A(t)| = t^\rho \ell(t)$ with ℓ a slowly-varying function. The larger ρ is, the larger the (absolute) asymptotic bias. Numerous distributions satisfying assumption (5) can be found in [2], see also Table 1 for examples.

Distribution (parameters)	Density function	γ	ρ
Generalised Pareto ($\xi > 0$)	$(1+\xi t)^{-1-1/\xi}$, $t > 0$	ξ	$-\xi$
Burr ($\zeta, \theta > 0$)	$\zeta\theta t^{\zeta-1}\left(1+t^\zeta\right)^{-\theta-1}$, $t > 0$	$1/(\zeta\theta)$	$-1/\theta$
Fréchet ($\zeta > 0$)	$\zeta t^{-\zeta-1} \exp\left(-t^{-\zeta}\right)$, $t > 0$	$1/\zeta$	-1
Fisher ($\nu_1, \nu_2 > 0$)	$\dfrac{(\nu_1/\nu_2)^{\nu_1/2}}{B(\nu_1/2, \nu_2/2)} t^{\nu_1/2-1} \left(1 + \dfrac{\nu_1}{\nu_2} t\right)^{-(\nu_1+\nu_2)/2}$, $t > 0$	$2/\nu_2$	$-2/\nu_2$
Inverse Gamma ($\zeta > 0$)	$\dfrac{1}{\Gamma(\zeta)} t^{-\zeta-1} \exp(-1/t)$, $t > 0$	$1/\zeta$	$-1/\zeta$
Student ($\nu > 0$)	$\dfrac{1}{\sqrt{\nu} B(\nu/2, 1/2)} \left(1 + \dfrac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	$1/\nu$	$-2/\nu$

Table 1: Examples of heavy-tailed distributions satisfying the second-order condition (5) with the associated values of γ and ρ. Here, $\Gamma(\cdot)$ and $B(\cdot,\cdot)$ denote respectively the Gamma and Beta functions.

Our first result is a refinement of [31, Theorem 4.3.8]. It provides an asymptotic normality result for the extreme quantile estimator (2) based on two intermediate sequences (k_n) and (k'_n).

Theorem 1. *Assume the second-order condition (5) holds. Let (k_n) and (k'_n) be two intermediate sequences such that $k'_n \le k_n$ and introduce (α_n) a sequence in $(0,1)$ such that $\alpha_n \to 0$ as $n \to \infty$. Suppose, as $n \to \infty$,*

(i) $\sqrt{k'_n} A(n/k'_n) \to \lambda \in \mathbb{R}$,

(ii) *$\hat\gamma_n(\cdot)$ is an estimator of γ such that $\sqrt{k'_n}(\hat\gamma_n(k'_n) - \gamma) \xrightarrow{d} \mathcal{N}(\lambda\mu, \sigma^2)$ where $\mu, \sigma > 0$,*

(iii) *$d_n \to \infty$, $(\log d_n)/\sqrt{k'_n} \to 0$ and $(k'_n/k_n)^\rho / \log d_n \to c \ge 0$.*

Then, as $n \to \infty$,

$$\frac{\sqrt{k'_n}}{\log d_n} \left(\frac{\hat q_n(\alpha_n, k_n, k'_n)}{q(\alpha_n)} - 1 \right) \xrightarrow{d} \mathcal{N}(\lambda(\mu + c/\rho), \sigma^2). \tag{6}$$

Assumptions (i) and (ii) are inherited from [31, Theorem 4.3.8]: they ensure the asymptotic normality of $\hat{\gamma}_n(k'_n)$ with balanced bias and variance. The role of (iii) is to control the extrapolation factor d_n appearing in (2). Compared to [31, Theorem 4.3.8], it involves the extra condition $(k'_n/k_n)^\rho/\log d_n \to c \geq 0$ which is used to balance the extrapolation bias with the bias of $\hat{\gamma}_n(k'_n)$. In view of (6), the refined Weissman estimator inherits its asymptotic distribution from $\hat{\gamma}_n(k'_n)$ with an additional bias component $\lambda c/\rho$ compared to the original Weissman estimator. Let us note that, in the particular case where $k'_n = k_n$, the above extra condition is satisfied with $c = 0$, and we recover the classical Weissman asymptotic normality result.

Corollary 1. *Assume the assumptions of Theorem 1 hold. If, moreover,* $(\log d_n)/\sqrt{k'_n} = o(k_n^{-1/4})$ *and the auxiliary function in (5) is given by*

$$A(t) = \beta\gamma t^\rho(1 + o(1)), \text{ as } t \to \infty, \tag{7}$$

with $\beta \neq 0$*, i.e.* $\ell(t) \to C$ *as* $t \to \infty$ *for some* $C > 0$*, then the asymptotic squared bias of* $\hat{q}_n(\alpha_n, k_n, k'_n)$ *is given by*

$$AB^2(k_n, k'_n, \alpha_n) = (\log d_n)^2 A^2(n/k'_n)\left(\mu + (k'_n/k_n)^\rho\left(\frac{1 - d_n^\rho}{\rho \log d_n}\right)\right)^2.$$

This situation (7) arises for instance in the Hall-Welsh class of heavy-tailed distributions [33, 34] defined by

$$U(t) = Ct^\gamma(1 + \beta\gamma t^\rho/\rho + o(t^\rho)), \text{ with } C > 0. \tag{8}$$

The crucial point is that the asymptotic bias can be cancelled (disregarding smaller order terms) by letting

$$k_n^\star := k'_n(\rho, \alpha_n, k_n) = k_n\left(-\rho\mu\frac{\log d_n}{1 - d_n^\rho}\right)^{1/\rho}. \tag{9}$$

The next lemma describes the behavior of $k_n^\star$ as a function of k_n.

Lemma 1.

(i) *For all* $d_n \geq 1$ *and* $\rho < 0$*,* $k_n^\star$ *is an increasing function of* k_n *and* $k_n^\star \leq \mu^{1/\rho}\, k_n$*.*

(ii) *For all* $\rho < 0$*,* $k_n^\star \sim \tau k_n (\log d_n)^{1/\rho}$ *as* $n \to \infty$*, where* $\tau := (-\rho\mu)^{1/\rho}$*.*

(iii) *If, moreover,* $\log(n\alpha_n)/\log(k_n) \to c' \leq 0$ *as* $n \to \infty$*, then* $k_n^\star \sim \tau' k_n(\log k_n)^{1/\rho}$*, where* $\tau' = \tau(1 - c')^{1/\rho}$*.*

From (ii), it appears that $k_n^\star/k_n \to 0$ as $n \to \infty$, meaning that the number of upper order statistics used in the tail-index estimator should be asymptotically small compared to k_n. See Paragraph 2.3 for a detailed discussion in the case of Hill estimator.

The next result provides the asymptotic distribution of the extreme quantile estimator (2) computed with $k'_n = k_n^\star$.

Corollary 2. *Assume the second-order condition (5) holds with auxiliary function* A *given by (7). Let* (k_n) *be an intermediate sequence and* (α_n) *a sequence in* $(0, 1)$ *such that* $\alpha_n \to 0$ *as* $n \to \infty$*. Let* $d_n = k_n/(n\alpha_n) \to \infty$ *such that*

$$\sqrt{k_n}(\log d_n)^{1/(2\rho)-1}A(n/k_n) \to \lambda' \in \mathbb{R} \text{ and } (\log d_n)^{1-1/(2\rho)}/\sqrt{k_n} \to 0, \tag{10}$$

as $n \to \infty$. Define $k_n^\star$ as in (9) for some $\mu > 0$ and let $\hat\gamma_n(\cdot)$ be an estimator of γ such that $\sqrt{k_n^\star}(\hat\gamma_n(k_n^\star) - \gamma) \xrightarrow{d} \mathcal{N}(\lambda'\mu, \sigma^2)$ where $\sigma > 0$. Then, as $n \to \infty$,

$$\frac{\sqrt{k_n^\star}}{\log d_n}\left(\frac{\hat q_n(\alpha_n, k_n, k_n^\star)}{q(\alpha_n)} - 1\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2). \tag{11}$$

As expected, the resulting asymptotic Gaussian distribution in (11) is centered, in contrast to (6). A possible choice of sequences is $k_n = n^{-2\rho/(1-2\rho)}$ and $\alpha_n = n^{-a}$ for all $a > 1/(1-2\rho)$, leading to $c' = (1-a)(1-2\rho)/(-2\rho)$ in Lemma 1(iii). These sequences yield the usual rate of convergence of order $n^{-\rho/(1-2\rho)}$ in (11), up to a logarithmic factor.

In practice, the refined Weissman estimator is computed using $\hat k_n^\star$, an estimation of the intermediate sequence given in (9):

$$\hat k_n^\star := \left\lfloor k_n \left(-\hat\rho_n \mu(\hat\rho_n) \frac{\log d_n}{1 - d_n^{\hat\rho_n}}\right)^{1/\hat\rho_n} \right\rfloor, \tag{12}$$

with $\hat\rho_n$ an estimator of the second order parameter ρ and where $\lfloor\cdot\rfloor$ denotes the integer part. The estimation of ρ has been extensively discussed in the extreme-value literature, we refer to [15] for a review in the particular case of heavy-tailed distributions and to Paragraph 2.4 for implementation details. In (12), we assumed that the asymptotic bias μ of $\hat\gamma_n$ only depends on ρ. It can then be shown that the estimated intermediate sequence $\hat k_n^\star$ is asymptotically equivalent to the theoretical one $k_n^\star$.

Lemma 2. *Let $\log(n\alpha_n)/\log(k_n) \to c' \leq 0$ as $n \to \infty$ and consider $\hat\rho_n$ an estimator of $\rho < 0$ such that $(\log k_n)(\hat\rho_n - \rho) = O_P(1)$. Assume that $\mu(\cdot)$ in (12) is a positive continuous function. Then, the estimated intermediate sequence verifies $\hat k_n^\star = k_n^\star(1 + o_P(1))$.*

The required consistency condition on $\hat\rho_n$ is rather weak. It is fulfilled for instance by estimators from [23, 34, 47] as a consequence of [15, Lemma 1 and Theorem 2] and by estimators introduced in [8, 21, 22, 29, 40] as a consequence of [15, Lemma 2 and Theorem 2]. The refined Weissman estimator $\hat q_n(\alpha_n, k_n, \hat k_n^\star)$ using the estimated intermediate sequence thus inherits its asymptotic distribution from its theoretical counterpart $\hat q_n(\alpha_n, k_n, k_n^\star)$.

Corollary 3. *Assume the second-order condition (5) holds with auxiliary function A given by (7). Let (k_n) be an intermediate sequence and (α_n) in $(0, 1)$ such that $\alpha_n \to 0$, $\log(n\alpha_n)/\log(k_n) \to c' \leq 0$ and $\sqrt{k_n}(\log k_n)^{1/(2\rho)-1}A(n/k_n) \to \lambda' \in \mathbb{R}$ as $n \to \infty$. Define $k_n^\star$, $\hat k_n^\star$ as in (9), (12) respectively (with $\mu(\cdot)$ a positive continuous function) and assume that*

- *$\hat\gamma_n(\cdot)$ is an estimator of γ such that $\sqrt{k_n^\star}(\hat\gamma_n(k_n^\star) - \gamma) \xrightarrow{d} \mathcal{N}(\lambda'\mu(\rho), \sigma^2)$ where $\sigma > 0$,*
- *$\hat\rho_n$ is an estimator of $\rho < 0$ such that $(\log k_n)(\hat\rho_n - \rho) = O_P(1)$.*

Then, as $n \to \infty$,

$$\frac{\sqrt{k_n^\star}}{\log d_n}\left(\frac{\hat q_n(\alpha_n, k_n, \hat k_n^\star)}{q(\alpha_n)} - 1\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2). \tag{13}$$

Let us highlight that convergence (13) also holds with the random rate of convergence $\sqrt{\hat k_n^\star}/\log d_n$ (see (29) in the Appendix) so that asymptotic confidence intervals on $q(\alpha_n)$ can easily be derived. Some examples of tail-index estimators satisfying the conditions of the above theoretical results are provided in the next paragraph, as well as their associated asymptotic mean $\mu(\cdot)$ and variance σ^2.

2.3 Examples

Hill estimator. Let us first focus on the case where $\hat{\gamma}_n(\cdot)$ is the Hill estimator (3). The assumptions of the above results are satisfied with $\mu(\rho) = 1/(1-\rho)$ and $\sigma = \gamma$, see for instance [31, Theorem 3.2.5], leading to

$$k_n^{\mathrm{H},\star} = k_n^{\mathrm{H},\star}(\rho, \alpha_n, k_n) = k_n \left(\frac{-\rho}{1-\rho} \frac{\log d_n}{1 - d_n^{\rho}} \right)^{1/\rho}. \tag{14}$$

Remark that, when $\rho \to -\infty$, then $k_n^{\mathrm{H},\star}(-\infty, \alpha_n, k_n) = k_n$ and we find back the original Weissman estimator (4). At the opposite, when $\rho \to 0^-$, one has $k_n^{\mathrm{H},\star}(0^-, \alpha_n, k_n) = ek_n/\sqrt{d_n} \sim ek_n^{\frac{1+c'}{2}+o(1)}$ as $n \to \infty$ under the condition $\log(n\alpha_n)/\log(k_n) \to c' \leq 0$. It thus appear that, in the low bias situation, $k_n^{\mathrm{H},\star}$ and k_n are of the same order. Conversely, in the high bias situation, $k_n^{\mathrm{H},\star}$ is significantly smaller than k_n, see Figure 1 for an illustration in the case $\alpha = 1/n$ *i.e.* $d_n = k_n$ and $n = 500$. It appears that, the larger ρ is, the smaller $k_n^{\mathrm{H},\star}$ is, in order to dampen the extrapolation bias.

Other examples. Similarly, Zipf estimator, based on a least-squares regression on the quantile-quantile plot and proposed simultaneously by [37, 44], fulfills the assumptions of Theorem 1 with $\mu(\rho) = 1/(1-\rho)^2$ and $\sigma^2 = 2\gamma^2$, see [1]. Finally, the moment [14] and Pickands estimators [41] satisfy the assumptions of the above theorem even though they address the estimation of γ without positivity assumption. We refer to [32, Theorem 1] for the associated values of (μ, σ^2) in the particular case where $\gamma > 0$.

In the sequel, we focus on the Hill estimator, which is the tail-index estimator with smallest variance among the above mentioned ones.

2.4 Implementation

In practice, the refined Weissman estimator is computed using the Hill estimator $\mathrm{H}(\hat{k}_n^{\mathrm{H},\star})$ where

$$\hat{k}_n^{\mathrm{H},\star} := k_n^{\mathrm{H},\star}(\hat{\rho}_n, \alpha_n, k_n) = \left\lfloor k_n \left(\frac{-\hat{\rho}_n}{1-\hat{\rho}_n} \frac{\log d_n}{1 - d_n^{\hat{\rho}_n}} \right)^{1/\hat{\rho}_n} \right\rfloor \tag{15}$$

is an estimation of the intermediate sequence given in (14). Here, we adopt the estimator $\hat{\rho}_n(\tilde{k})$ of the second-order parameter introduced in [21, Equation (2.18)] and implemented in the package evt0 of the R software [38] which features a satisfying behaviour in practice using the default level parameter $\tilde{k} = \lfloor n^{0.999} \rfloor + 1$. The resulting refined Weissman estimator is denoted by

$$\mathrm{RW}(k_n) = \hat{q}_n(\alpha_n, k_n, \hat{k}_n^{\mathrm{H},\star}) = \hat{q}_n(\alpha_n, k_n, k_n^{\mathrm{H},\star}(\hat{\rho}_n, \alpha_n, k_n)).$$

This data-driven choice of $\hat{k}_n^{\mathrm{H},\star}$ as a function of k_n is illustrated on Figure 2 on data sets of size $n = 500$ together with its consequences on the estimation of the tail-index (left panel) and extreme quantiles (right panel). Two Burr distributions are considered with $\gamma = 1/4$ and $\rho \in \{-2, -3/4\}$, see Table 1. The Relative mean-squared error (RMSE) is computed on $N = 1000$ replications, see (20) below. It appears that, in the low bias situation ($\rho = -2$, top panel), $\hat{k}_n^{\mathrm{H},\star}$ is automatically limited to the range $\{1, \dots, 151\}$ as $k_n \in \{1, \dots, n-1\}$. In the high bias situation ($\rho = -3/4$, bottom panel), $\hat{k}_n^{\text{H},\star}$ is further limited to the range $\{1,\dots,135\}$. In both cases, the bias associated with Hill estimator $\text{H}(\hat{k}_n^{\text{H},\star})$ remains acceptable in the considered ranges. As a consequence, the relative mean-squared error associated with $\text{RW}(k_n) = \hat{q}_n(\alpha_n = 1/n, k_n, \hat{k}_n^{\text{H},\star})$ is small for a wide range of values of k_n, in contrast to the original Weissman estimator. These preliminary numerical experiments are conducted in a more systematic way in the following section.

3 Validation on simulated data

The proposed refined Weissman estimator is compared on simulated data to the original Weissman estimator and to six other bias-reduced estimators of the extreme quantile.

3.1 Experimental design

The comparison is achieved on seven heavy-tailed distributions. The first six distributions: Burr, Fréchet, Fisher, generalized Pareto distribution (GPD), Inverse Gamma, and Student belong to the Hall-Welsh class (8), they satisfy the second-order condition (5) with (7), see Table 1 for their definitions and associated values of γ and ρ. The last distribution, denoted by $\text{NHW}(\gamma,\rho)$, is defined for all $\gamma \geq \exp(-2)/2$ and $\rho < 0$ by its tail quantile function $U(t) = t^\gamma \exp(A(t)/\rho)$ where $A(t) = \rho t^\rho \log(t)/2$, $t \geq 1$, is the auxiliary function associated with the second-order condition (5). It thus appears from (8) that the NHW distribution does not belong to the Hall-Welsh class.

For Fisher, Inverse Gamma and Student distributions, four tail-index values $\gamma \in \{1/8, 1/4, 1/2, 1\}$ are investigated. For Fréchet and generalized Pareto distributions, we focus on $\gamma \in \{1/8, 1/4, 1/2\}$ since, when $\gamma = 1$, they coincide respectively with the Inverse Gamma and Burr distributions, see Table 1. The choice of the second-order parameter ρ depends on the considered distribution:

- Fréchet distribution: the second order parameter is fixed to $\rho = -1$.
- Fisher, GPD and Inverse Gamma: the second order parameter is fixed to $\rho = -\gamma$. In the case of the Fisher distribution, we set $\nu_1 = 3$.
- Student distribution: the second order parameter is fixed to $\rho = -2\gamma$.

In the case of Burr and NHW distributions, we restrict ourselves to three tail-index values $\gamma \in \{1/8, 1/4, 1/2\}$ since preliminary experiments showed that all estimators failed to estimate extreme quantiles in an accurate way when $\gamma = 1$ (RMSE≥ 1, see (20)).

- Burr and NHW distributions: the second order parameter can be chosen independently from the tail-index, five values are tested $\rho \in \{-1/8, -1/4, -1/2, -1, -2\}$.

In each case, we simulate $N = 1000$ replications of a data set of $n = 500$ i.i.d. realisations from the 48 considered parametric models. Finally, two cases are investigated for the order of the extreme quantile: $\alpha_n \in \{1/n, 1/(2n)\}$. Summarizing, this experimental design includes $48 \times 2 = 96$ configurations.

3.2 Competitors

Since the original Weissman estimator inherits its asymptotic distribution from $\mathrm{H}(k_n)$, the main idea of most reduced bias estimators of the extreme quantile is to replace the Hill estimator $\mathrm{H}(k_n)$ in (4) by a bias-reduced version. We shall first consider the Corrected-Hill (CH) [6]:

$$\mathrm{CH}(k_n) = \mathrm{H}(k_n)\left(1 - \frac{\hat{\beta}_n}{1-\hat{\rho}_n}\left(\frac{n}{k_n}\right)^{\hat{\rho}_n}\right), \tag{16}$$

where $\hat{\rho}_n$ and $\hat{\beta}_n$ are estimators of the second-order parameters ρ and β, see (7). The associated bias reduced Weissman estimator is studied in [30]. Second, let us introduce

$$\mathrm{H}_p(k_n) = \begin{cases} \dfrac{1}{p}\left(1 - \left(\dfrac{1}{k_n}\displaystyle\sum_{i=1}^{k_n}\left(\dfrac{X_{n-i+1,n}}{X_{n-k_n,n}}\right)^p\right)^{-1}\right) & \text{if } p < 1/\gamma \text{ and } p \neq 0, \\ \mathrm{H}(k_n) & \text{if } p = 0, \end{cases} \tag{17}$$

the mean-of-order-p estimator of γ proposed almost simultaneously in [3, 5, 39], where p is some tuning parameter. A bias-reduced version of the previous estimator (17), referred to as reduced-bias mean-of-order-p and denoted by CH_p, is considered in [26]:

$$\mathrm{CH}_p(k_n) = \mathrm{H}_p(k_n)\left(1 - \frac{\hat{\beta}_n(1-p\mathrm{H}_p(k_n))}{1-\hat{\rho}_n - p\mathrm{H}_p(k_n)}\left(\frac{n}{k_n}\right)^{\hat{\rho}_n}\right), \tag{18}$$

following the same principle as in (16), so that $\mathrm{CH}_0(k_n) = \mathrm{CH}(k_n)$. An alternative bias-reduced version of (17) is proposed in [25] by replacing in the bias correction term of (18) an optimal value of p in terms of asymptotic efficiency. This gives rise to the Partially reduced-bias mean-of-order-p (PRB_p) estimator defined by

$$\mathrm{PRB}_p(k_n) = \mathrm{H}_p(k_n)\left(1 - \frac{\hat{\beta}_n(1-\varphi_{\hat{\rho}_n})}{1-\hat{\rho}_n - \varphi_{\hat{\rho}_n}}\left(\frac{n}{k_n}\right)^{\hat{\rho}_n}\right), \tag{19}$$

with $\varphi_\rho = 1 - \rho/2 - \sqrt{(1-\rho/2)^2 - 1/2}$. Both CH_p and PRB_p estimators are plugged in Weissman estimator to obtain extreme quantile estimators, see [27]. In the following, for the sake of simplicity, the extreme quantile estimators derived from (16), (18) and (19) are denoted by their associated tail-index estimators, namely CH, CH_p and PRB_p. Finally, a Corrected Weissman (CW) estimator is introduced in [30] implementing two bias corrections: a first one in the tail-index estimator and a second one in the extrapolation factor:

$$\mathrm{CW}(k_n) = X_{n-k_n,n}\left(\frac{k_n}{n\alpha_n}\exp\left(\hat{\beta}_n\left(\frac{n}{k_n}\right)^{\hat{\rho}_n}\frac{(k_n/(n\alpha_n))^{\hat{\rho}_n}-1}{\hat{\rho}_n}\right)\right)^{\mathrm{CH}(k_n)}.$$

3.3 Selection of hyperparameters

All considered extreme quantile estimators (Weissman, RW, CH, CH_p, PRB_p and CW) depend on the intermediate sequence k_n. The selection of k_n is a crucial point which has been widely discussed in the extreme-value literature. A standard practice is to pick out a value of k_n in the first stable part of the plot $k_n \mapsto \hat{\gamma}(k_n)$ where $\hat{\gamma}(\cdot)$ is the tail-index estimator of interest, see [31, Chapter 3]. Some attempts at formalizing this procedure can be found in [16, 42] and, more recently, in [19, 20].

We propose a new algorithm for the selection of k_n based on a bisection method inspired from random forests. The objective is to find the region with the smallest variance in a given series $\{Z_1, \dots, Z_n\}$. However, this region may be hard to find, depending on the variations on both large and small scales. Our proposed method starts by randomly sampling two points $a^{(0)} < c^{(0)}$ from the series $\{Z_1, \dots, Z_n\}$. Next, split the series $\{Z_i \in [a^{(0)}, c^{(0)}], i = 1, \dots, n\}$ in two, compute the variance in each sub-region and repeat the action in the one with smallest variance until getting a final single point (Algorithm 2). Because we may start by sampling two initial points in a non-optimal region, the above procedure is embedded in a bootstrap technique to dampen this vexing effect (Algorithm 1). Finally, we take the median accross the T replications to select the final $k_n^\dagger$. In the simulations we used $T = 10000$, $a^{(0)} = 15$, $c^{(0)} = 3n/4$.

Besides, CH_p and PRB_p involve an extra parameter p which also has to be selected. Two solutions are possible. First, following [27, p. 1739], one can choose the optimal value of p in terms of efficiency:

$$p^\star = \varphi_{\hat\rho_n}/\mathrm{CH}(\hat k_0) \quad \text{where} \quad \hat k_0 = \min\left(n-1, \left\lfloor \left((1-\hat\rho_n)^2 n^{-2\hat\rho_n}/(-2\hat\rho_n\hat\beta_n^2)\right)^{1/(1-2\hat\rho_n)} \right\rfloor + 1\right),$$

which gives rise to two estimators $\mathrm{CH}_{p^\star}$ and $\mathrm{PRB}_{p^\star}$. Second, one may select p using the sample path stability criterion of [27, Algorithm 4.2]. The resulting estimators are still denoted by CH_p and PRB_p. To summarize, eight extreme quantile estimators are compared in the following: Weissman, RW, CH, CH_p, PRB_p, $\mathrm{CH}_{p^\star}$, $\mathrm{PRB}_{p^\star}$ and CW.

3.4 Results

The performance of the extreme quantile estimators is assessed using

$$\mathrm{RMSE}\left(\hat q_n(\alpha_n)\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat q_n^{(i)}(\alpha_n)}{q_n(\alpha_n)} - 1\right)^2, \tag{20}$$

where $\hat q_n^{(i)}(\alpha_n)$ denotes the estimator computed on the ith replication, $i \in \{1, \dots, N\}$. The results are provided in Table 3 ($\alpha_n = 1/n$), Table 6 ($\alpha_n = 1/(2n)$), for the Burr distribution, in Table 4 ($\alpha_n = 1/n$), Table 7 ($\alpha_n = 1/(2n)$), for the NHW distribution and in Table 5 ($\alpha_n = 1/n$), Table 8 ($\alpha_n = 1/(2n)$) for Fréchet, Fisher, GPD, Inverse Gamma, and Student distributions.

Let us first remark that, in the case $\gamma = 1$, for all distributions with $\rho > -1/2$, all estimators fail in estimating the extreme quantiles $q(1/n)$ and $q(1/(2n))$: this corresponds to the most difficult situations where both γ and ρ are large. In particular, in the case of a Burr distribution with $\gamma = 1/2$ and $\rho = -1/8$, none of the eight considered estimators gave acceptable results for $\alpha_n = 1/n$ and $\alpha_n = 1/(2n)$, in the sense that RMSE ≥ 1.

Second, in the 94 remaining situations, the original Weissman estimator yields the best result 8 times, all corresponding to the NHW distribution with $\rho \leq -1$, indicating that bias reduction is useful in general. At the opposite, the proposed RW estimator is the most accurate one since it provides the best result in 30 out of 94 times. Note that, in the particular case of the NHW distribution, it yields the best result in 16 out of 30 times. The RW estimator is especially efficient in the challenging cases where ρ is large. As an example, it yields the best result 15 out of 19 times when $\rho = -1/8$. Let us also highlight that, when γ is fixed, it associated RMSE is a decreasing function of $|\rho|$, see Tables 3, 4, 6 and 7.

Algorithm 1: Selection of k_n

Input: *series:* $\mathcal{Z} = \{Z_1, \ldots, Z_n\}$,
number of trees: $T \in \mathbb{N} \setminus \{0\}$,
initial left point: $a^{(0)} \in \{1, \ldots, n-3\}$,
initial right point: $c^{(0)} \in \{2, \ldots, n-1\}$

Output: *selected point:* $k_n^\dagger$

for $t = 1 : T$ **do**
  $a^{(t)} \sim \texttt{randint}(a^{(0)}, c^{(0)} - 1)$
  $c^{(t)} \sim \texttt{randint}(a^{(t)} + 1, c^{(0)})$
  $k_n^{(t)} \leftarrow \mathrm{Tree}(\mathcal{Z}, a^{(t)}, c^{(t)})$

$k_n^\dagger \leftarrow \texttt{median}(k_n^{(1)}, \ldots, k_n^{(T)})$

Algorithm 2: Tree

Input: *series:* $\mathcal{Z} = \{Z_1, \ldots, Z_n\}$,
initial left point: $a \in \{1, \ldots, n-2\}$,
initial right point: $c \in \{2, \ldots, n-1\}$

Output: *selected point:* b

$b \leftarrow \left[\frac{a+c}{2}\right]$

while $(b - a) > 1$ **do**
  $\bar{Z}_{\mathrm{L}} \leftarrow \frac{1}{b-a+1} \sum_{k=a}^{b} Z_k, \quad \mathrm{v}_{\mathrm{L}} \leftarrow \frac{1}{b-a+1} \sum_{k=a}^{b} \left(Z_k - \bar{Z}_{\mathrm{L}}\right)^2$
  $\bar{Z}_{\mathrm{R}} \leftarrow \frac{1}{c-b+1} \sum_{k=b}^{c} Z_k, \quad \mathrm{v}_{\mathrm{R}} \leftarrow \frac{1}{c-b+1} \sum_{k=b}^{c} \left(Z_k - \bar{Z}_{\mathrm{R}}\right)^2$
  if $\mathrm{v}_{\mathrm{L}} < \mathrm{v}_{\mathrm{R}}$ **then**
    $c \leftarrow b$
  else
    $a \leftarrow b$
  $b \leftarrow \left[\frac{a+c}{2}\right]$

The second most accurate estimator is CW which provides the best result in 24 out of 94 times. It is particularly efficient in the Fréchet case where it obtains the best result in the 6 considered cases. $\mathrm{CH}_{p^\star}$ and $\mathrm{PRB}_{p^\star}$ obtain worst and similar results with respectively 16 and 14 best results out of 94 situations. Finally, CH, CH_p and PRB_p yield very poor estimations, with respectively 2, 0 and 0 best results.

To conclude, it appears on these experiments on simulated data that, in average, the RW estimator performs the best within the eight considered estimators. Unsurprisingly, its main competitor is CW, which, similarly to RW, considers the two sources of bias (associated with the tail-index estimator and the extrapolation). Note that $\mathrm{CH}_{p^\star}$ and $\mathrm{PRB}_{p^\star}$ can also reveal useful out of the NHW family of distributions. One may also conclude that the refined Weissman estimator is more robust to the Hall-Welsh assumption than the other bias reduced estimators.

As an illustration, the behaviour of the bias and RMSEs associated with the five above mentioned estimators computed on a Burr distribution with $\gamma = 1/8$, $\rho \in \{-1/8, -1/4, -1/2, -1, -2\}$ and $\alpha_n = 1/(2n)$ is depicted on Figure 3. It is clearly seen that the refined Weissman estimator enjoys the best performance in terms of bias, with stable sample paths for all considered values of ρ. In the difficult cases where ρ is large ($\rho = -1/8$ and $\rho = -1/4$), the RMSE associated with the refined Weissman estimator also benefits from a nice stable behaviour for a wide range of k_n values. The behaviour of all estimators on real data is illustrated in the next section.

4 Illustration on an actuarial data set

We consider here the Secura Belgian reinsurance data set [17] on automobile claims from 1998 until 2001, introduced in [2] as an example of heavy-tailed data set and further analyzed in [19] from an extreme risk measures perspective. This data set consists of $n = 371$ claims which were at least as large as 1.2 million Euros and were corrected for inflation. See the top left panel of Figure 4 for a histogram representation of the data distribution. Our goal is to estimate the extreme quantile $q(1/n)$ (with $1/n \simeq 0.0027$) and to compare it to the maximum of the sample $x_{n,n} = 7.898$ million Euros.

The first step is to estimate the second order parameter. We get $\hat{\rho}_n = -0.756$ (see Paragraph 2.4 for implementation details) which corresponds to a relatively high bias situation. We refer to Figure 2 for an illustration in a similar simulated situation with $\rho = -3/4$. Second, the estimated intermediate sequence is then computed from (15): $\hat{k}_n^{\mathrm{H},\star} \in \{1, \dots, 106\}$ as $k_n \in \{1, \dots, n-1 = 370\}$. The associated Hill plots $\mathrm{H}(k_n)$ and $\mathrm{H}(\hat{k}_n^{\mathrm{H},\star})$ are displayed on the top right panel of Figure 4. Algorithm 1 selects $k_n^\dagger = 210$ leading to $\hat{k}_n^{\mathrm{H},\star} = 69$ and $\mathrm{H}(\hat{k}_n^{\mathrm{H},\star}) = 0.2801$ as estimated tail-index. As a visual check, a quantile-quantile plot of the log-excesses $\log(X_{n-i+1,n}) - \log(X_{n-\hat{k}_n^{\mathrm{H},\star},n})$ against the quantiles of the unit exponential distribution $\log(\hat{k}_n^{\mathrm{H},\star}/i)$ for $i = 1, \dots, \hat{k}_n^{\mathrm{H},\star}$ is drawn on the bottom left panel of Figure 4. The relationship appearing in this plot is approximately linear, which constitutes an empirical evidence that the heavy-tail assumption makes sense and that $\hat{k}_n^{\mathrm{H},\star} = 69$ is a reasonable choice to estimate the tail-index.

The eight estimates of the tail-index (see Section 3) are reported in Table 2. For the last two estimators, the automatic selection procedure provided the value $p^\star = 0.765$. The original Hill estimator yields the largest value $\hat{\gamma}_n = 0.3345$ while the remaining six estimators provide smaller estimated tail-indices $\hat{\gamma}_n \simeq 0.25$. The corresponding estimated extreme quantiles $\hat{q}_n(1/n)$ are also reported. Note that in [19], the authors obtained $\hat{q}_n(0.005) = 7.163$ and $\hat{q}_n(0.001) = 10.899$. It appears from Table 2 that the estimations provided by CH, CH_p, PRB_p, $\text{CH}_{p^\star}$ and $\text{PRB}_{p^\star}$ are not coherent with the previous results since, in these cases $\hat{q}_n(1/n) \leq \hat{q}_n(0.005)$ while $1/n < 0.005$. Underestimation can then be suspected for these estimators.

Similarly, the Weissman estimator seems to overestimate the extreme quantile $q(1/n)$ since, in this case $\hat{q}_n(1/n) \geq \hat{q}_n(0.001)$ while $1/n > 0.001$. Summarizing, the only two plausible estimations are provided by the proposed refined Weissman and CW estimators who give the closest estimation of the maximum value of the sample: $\text{RW}(1/n) = 8.298$ and $\text{CW}(1/n) = 8.203$ while $x_{n,n} = 7.898$. Both sample paths associated with $k_n \mapsto \text{RW}(1/n)$ and $k_n \mapsto \text{CW}(1/n)$ enjoy a stable behaviour in a large neighbourhood of $k_n^\dagger$, see the bottom right panel of Figure 4. Besides, the maximum value does belong to the asymptotic 95% confidence interval $[5.366, 11.231]$ associated with $\text{RW}(1/n)$, which is thus a reasonable estimation of $q(1/n)$.

As a conclusion, according to $\text{RW}(1/n)$ estimate, one can expect a claim larger than 8.298 million Euros to occur in average once every four years.

	Weissman	RW	CW	CH	CH_p	PRB_p	$\text{CH}_{p^\star}$	$\text{PRB}_{p^\star}$
$\hat{\gamma}_n$	0.3445	0.2801	0.2469	0.2494	0.2494	0.2468	0.2501	0.2498
$\hat{q}_n(1/n)$	11.759	8.298	8.203	7.136	7.136	7.104	7.091	7.108
$k_n^\dagger$	177	210	203	199	199	196	202	190

Table 2: Comparison of eight estimators on the Secura Belgian reinsurance actuarial data set. Tail-index γ and extreme quantile $q(1/n)$ (in million Euros) estimations. The selected intermediate sequence $k_n^\dagger$ is also provided for each estimator.

5 Conclusion

We believe that the refined Weissman estimator is an efficient tool for estimating extreme quantiles in a variety of heavy-tailed situations. In contrast to usual bias reduced estimators, our proposition is not based on a preliminary reduction of the bias associated with some tail-index estimator. It relies on an original idea consisting in selecting carefully two intermediate sequences to make the asymptotic bias vanish. A data-driven method is proposed for the practical selection of these intermediate sequences. Our further work will consist in extending this bias reduction principle in the more general context of an arbitrary maximum domain of attraction.

Appendix: proofs

Proof of Theorem 1. It follows the same lines as the one of [31, Theorem 4.3.8]. Let us consider the expansion

$$\frac{\sqrt{k_n'}}{\log d_n}\left(\frac{\hat{q}_n(\alpha_n,k_n,k_n')}{q(\alpha_n)}-1\right)=\frac{\sqrt{k_n'}}{\log d_n}\left(\frac{X_{n-k_n,n}\,d_n^{\hat{\gamma}_n(k_n')}}{U(1/\alpha_n)}-1\right)=\frac{T_{1,n}+T_{2,n}+T_{3,n}}{T_{0,n}},$$

where we have introduced:

$$\begin{aligned}
T_{0,n} &= d_n^{-\gamma}\,\frac{U(1/\alpha_n)}{U(n/k_n)},\\
T_{1,n} &= \frac{\sqrt{k_n'}}{\log d_n}\left(\frac{X_{n-k_n,n}}{U(n/k_n)}-1\right)d_n^{\hat{\gamma}_n(k_n')-\gamma},\\
T_{2,n} &= \frac{\sqrt{k_n'}}{\log d_n}\left(d_n^{\hat{\gamma}_n(k_n')-\gamma}-1\right),\\
T_{3,n} &= \frac{\sqrt{k_n'}}{\log d_n}(1-T_{0,n}).
\end{aligned}$$

Let us first focus on $T_{0,n}$. From [31, Theorem 2.3.9], it follows from the second-order condition (5) that, for any ε, $\delta>0$, there exists $t_0>1$ such that for all $t\geq t_0$ and $x\geq 1$,

$$\left|\frac{1}{A_0(t)}\left(\frac{U(tx)}{U(t)}-x^{\gamma}\right)-x^{\gamma}\frac{x^{\rho}-1}{\rho}\right|\leq\varepsilon x^{\gamma+\rho+\delta},$$

where A_0 is asymptotically equivalent to A. Letting $x=d_n$ and $t=n/k_n$ then yields

$$\left|\frac{T_{0,n}-1}{A_0(n/k_n)}-\frac{d_n^{\rho}-1}{\rho}\right|\leq\varepsilon d_n^{\rho+\delta},$$

or equivalently,

$$T_{0,n}=1+A_0(n/k_n)\left(\frac{d_n^{\rho}-1}{\rho}+\varepsilon R_n\right),$$

where $|R_n|\leq d_n^{\rho+\delta}$. Now, writing $|A_0|(t)=t^{\rho}\ell(t)$, where ℓ is a slowly-varying function, it follows,

$$A_0(n/k_n)=A_0(n/k_n')(k_n'/k_n)^{\rho}\frac{\ell(n/k_n)}{\ell(n/k_n')},$$

as $n\to\infty$. As a consequence, we obtain

$$T_{0,n}=1+(k_n'/k_n)^{\rho}A_0(n/k_n')\left(\frac{d_n^{\rho}-1}{\rho}+\varepsilon R_n\right)\frac{\ell(n/k_n)}{\ell(n/k_n')},$$

and letting $\varepsilon\to 0$ yields

$$T_{0,n}=1+(k_n'/k_n)^{\rho}A_0(n/k_n')\left(\frac{d_n^{\rho}-1}{\rho}\right)\frac{\ell(n/k_n)}{\ell(n/k_n')}. \tag{21}$$

Second, under the assumption $\sqrt{k_n'}(\hat{\gamma}_n(k_n')-\gamma)\xrightarrow{d}\mathcal{N}(\lambda\mu,\sigma^2)$ as $\sqrt{k_n'}A(n/k_n')\to\lambda$, we get

$$d_n^{\hat{\gamma}_n(k_n')-\gamma}=\exp\left(\frac{(\log d_n)}{\sqrt{k_n'}}\sqrt{k_n'}(\hat{\gamma}_n(k_n')-\gamma)\right)=\exp\left(\frac{(\log d_n)}{\sqrt{k_n'}}(\lambda\mu+\sigma\xi_n)\right),$$

where $\xi_n \xrightarrow{d} \mathcal{N}(0,1)$. Recalling that $(\log d_n)/\sqrt{k_n'} \to 0$ as $n \to \infty$, the following first order expansion holds

$$d_n^{\hat{\gamma}_n(k_n')-\gamma} = 1 + \frac{(\log d_n)}{\sqrt{k_n'}}(\lambda\mu + \sigma\xi_n) + O_P\left(\frac{(\log d_n)^2}{k_n'}\right). \tag{22}$$

In particular, $d_n^{\hat{\gamma}_n(k_n')-\gamma} \xrightarrow{\mathbb{P}} 1$ and therefore,

$$T_{1,n} = \frac{\sqrt{k_n'/k_n}}{\log d_n}\sqrt{k_n}\left(\frac{X_{n-k_n,n}}{U(n/k_n)} - 1\right)(1 + o_P(1)) = \frac{\sqrt{k_n'/k_n}}{\log d_n}\gamma\xi_n'(1 + o_P(1)), \tag{23}$$

where $\xi_n' \xrightarrow{d} \mathcal{N}(0,1)$, from [31, Theorem 2.2.1]. Third, it immediately follows from (22) that

$$T_{2,n} = \lambda\mu + \sigma\xi_n + O_P\left(\frac{\log d_n}{\sqrt{k_n'}}\right). \tag{24}$$

Finally, in view of (21) and recalling that $\sqrt{k_n'}A_0(n/k_n') \to \lambda$, one has

$$T_{3,n} = \lambda(k_n'/k_n)^\rho\left(\frac{1 - d_n^\rho}{\rho\log d_n}\right)\frac{\ell(n/k_n)}{\ell(n/k_n')}(1 + o(1)). \tag{25}$$

Collecting (21), (23), (24) and (25) yields

$$\begin{aligned}\frac{\sqrt{k_n'}}{\log d_n}\left(\frac{\hat{q}_n(\alpha_n, k_n, k_n')}{q(\alpha_n)} - 1\right) &= \lambda\mu + \lambda(k_n'/k_n)^\rho\left(\frac{1 - d_n^\rho}{\rho\log d_n}\right)\frac{\ell(n/k_n)}{\ell(n/k_n')}(1 + o(1)) \\ &+ \sigma\xi_n + \frac{\sqrt{k_n'/k_n}}{\log d_n}\gamma\xi_n'(1 + o_P(1)) + O_P\left(\frac{\log d_n}{\sqrt{k_n'}}\right),\end{aligned} \tag{26}$$

since $T_{0,n} = 1 + o(T_{3,n})$. Besides, assumptions $(k_n'/k_n)^\rho/(\log d_n) \to c \ge 0$ and $d_n \to \infty$ as $n \to \infty$ imply

$$\frac{\sqrt{k_n'}}{\log d_n}\left(\frac{\hat{q}_n(\alpha_n, k_n, k_n')}{q(\alpha_n)} - 1\right) \xrightarrow{d} \mathcal{N}(\lambda(\mu + c/\rho), \sigma^2),$$

and the result is proved. ■

Proof of Corollary 1. Under assumption (7), equation (26) in the proof of Theorem 1 can be simplified as

$$\begin{aligned}\frac{\sqrt{k_n'}}{\log d_n}\left(\frac{\hat{q}_n(\alpha_n, k_n, k_n')}{q(\alpha_n)} - 1\right) &= \lambda\mu + \lambda(k_n'/k_n)^\rho\left(\frac{1 - d_n^\rho}{\rho\log d_n}\right)(1 + o(1)) \\ &+ \sigma\xi_n + \frac{\sqrt{k_n'/k_n}}{\log d_n}\gamma\xi_n'(1 + o_P(1)) + O_P\left(\frac{\log d_n}{\sqrt{k_n'}}\right),\end{aligned}$$

since ℓ is asymptotically constant. Let us moreover note that

$$\frac{\log d_n}{\sqrt{k_n'}} = o\left(\frac{\sqrt{k_n'/k_n}}{\log d_n}\right) \iff \frac{\sqrt{k_n}(\log d_n)^2}{k_n'} = o(1) \iff \frac{\log d_n}{\sqrt{k_n'}} = o(k_n^{-1/4})$$

and therefore

$$\frac{\sqrt{k_n'}}{\log d_n}\left(\frac{\hat{q}_n(\alpha_n, k_n, k_n')}{q(\alpha_n)} - 1\right) = \lambda\mu + \lambda(k_n'/k_n)^\rho\left(\frac{1 - d_n^\rho}{\rho\log d_n}\right)(1 + o(1)) + \sigma\xi_n + \frac{\sqrt{k_n'/k_n}}{\log d_n}\gamma\xi_n'(1 + o_P(1)),$$

which proves the result. ■

Proof of Lemma 1. **(i)** Letting $f(x) = -\rho(\log x)/(1-x^\rho)$ for all $x \geq 1$ and $\rho < 0$, from (9) one has $k_n^\star = \mu^{1/\rho} k_n (f(d_n))^{1/\rho}$ with $d_n = k_n/(n\alpha_n) \geq 1$. First, routine calculations give:

$$\frac{\partial k_n^\star}{\partial k_n} = \mu^{1/\rho}(f(d_n))^{1/\rho}\left(1 + \frac{d_n}{\rho}\frac{f'(d_n)}{f(d_n)}\right) = \mu^{1/\rho}(f(d_n))^{1/\rho}\left(\frac{1}{\rho\log d_n} + \frac{1}{1-d_n^\rho}\right) \geq 0,$$

for all $d_n \geq 1$. As a conclusion, $\partial k_n^\star/\partial k_n \geq 0$ which proves that $k_n^\star$ is an increasing function of k_n. Second, it is easily shown that f is increasing and $f(1) = 1$, leading to $f(d_n) \geq 1$ and thus $k_n^\star \leq \mu^{1/\rho} k_n$.
(ii) is a consequence of $f(x) \sim -\rho \log x$ as $x \to \infty$.
(iii) Remark that assumption $\log(n\alpha_n)/\log(k_n) \to c' \leq 0$ implies that $\log d_n \sim (1-c')\log k_n$ as $n \to \infty$. The conclusion follows. ■

Proof of Corollary 2. It is sufficient to prove that assumptions (i) and (iii) of Theorem 1 hold true. First, Lemma 1(ii) entails that $k_n^\star/k_n \sim \tau(\log d_n)^{1/\rho}$ as $n \to \infty$. Besides, from (7), we have $A(n/k_n^\star)/A(n/k_n) \sim (k_n^\star/k_n)^{-\rho}$, so that

$$\sqrt{k_n^\star}A(n/k_n^\star) \sim (k_n^\star/k_n)^{1/2-\rho}\sqrt{k_n}A(n/k_n) \sim \tau^{1/2-\rho}(\log d_n)^{1/(2\rho)-1}\sqrt{k_n}A(n/k_n) \to \lambda'\tau^{1/2-\rho}$$

as $n \to \infty$ in view of the first part of (10). Assumption (i) of Theorem 1 thus holds true with $\lambda = \lambda'\tau^{1/2-\rho}$. Second,

$$\frac{\log d_n}{\sqrt{k_n^\star}} = \frac{\log d_n}{\sqrt{k_n}}(k_n^\star/k_n)^{-1/2} \sim \tau^{-1/2}\frac{(\log d_n)^{1-1/(2\rho)}}{\sqrt{k_n}} \to 0 \tag{27}$$

as $n \to \infty$ in view of the second part of (10). Third,

$$\frac{(k_n^\star/k_n)^\rho}{\log d_n} \to \tau^\rho \tag{28}$$

as $n \to \infty$. Collecting (27) and (28) proves that assumption (iii) of Theorem 1 thus holds true with $c = \tau^\rho$. ■

Proof of Lemma 2. Recall that, from the proof of Lemma 1(iii), $\log d_n \sim (1-c')\log k_n$ as $n \to \infty$. Let us then observe that

$$\frac{\log d_n}{1-d_n^{\hat\rho_n}} = \frac{\log d_n}{1-\exp(\rho\log d_n + O_P(1))} = \frac{\log d_n}{1-O_P(d_n^\rho)} = (\log d_n)(1+O_P(d_n^\rho))$$

and consequently

$$\begin{aligned}
\left(\frac{\log d_n}{1-d_n^{\hat\rho_n}}\right)^{1/\hat\rho_n} &= \exp\left(\frac{\log\log d_n + O_P(d_n^\rho)}{\rho + O_P(1/\log d_n)}\right)\\
&= \exp\left(\frac{\log\log d_n}{\rho} + O_P\left(\frac{\log\log d_n}{\log d_n}\right) + O_P(d_n^\rho)\right)\\
&= (\log d_n)^{1/\rho}(1+o_P(1)).
\end{aligned}$$

Besides, since $\hat{\rho}_n$ is a consistent estimator of ρ and $\mu(\cdot)$ is continuous, it follows that $(-\hat{\rho}_n\mu(\hat{\rho}_n))^{1/\hat{\rho}_n} \xrightarrow{\mathbb{P}} (-\rho\mu(\rho))^{1/\rho}$ and therefore

$$\begin{aligned} k_n \left(-\hat{\rho}_n\mu(\hat{\rho}_n)\frac{\log d_n}{1-d_n^{\hat{\rho}_n}}\right)^{1/\hat{\rho}_n} &= k_n \left(-\rho\mu(\rho)(\log d_n)\right)^{1/\rho}(1+o_P(1)) \\ &= k_n \left(-\rho\mu(\rho)(1-c')(\log k_n)\right)^{1/\rho}(1+o_P(1)) \\ &= k_n^\star(1+o_P(1)), \end{aligned}$$

in view of Lemma 1(iii). Remarking that the right hand side term tends to infinity in probability, one immediately has

$$\hat{k}_n^\star = \left\lfloor k_n \left(-\hat{\rho}_n\mu(\hat{\rho}_n)\frac{\log d_n}{1-d_n^{\hat{\rho}_n}}\right)^{1/\hat{\rho}_n} \right\rfloor = k_n^\star(1+o_P(1)),$$

and the result is proved. ■

Proof of Corollary 3. The first step is to prove that

$$\frac{\sqrt{\hat{k}_n^\star}}{\log d_n}\left(\frac{\hat{q}_n(\alpha_n,k_n,\hat{k}_n^\star)}{q(\alpha_n)}-1\right) \xrightarrow{d} \mathcal{N}(0,\sigma^2). \tag{29}$$

To this end, recall that $\log(n\alpha_n)/\log(k_n) \to c' \le 0$ implies that $\log d_n \sim (1-c')\log k_n$ as $n\to\infty$ and therefore condition (10) of Corollary 2 is fulfilled under the assumptions of Corollary 3. Besides, recalling that $k_n/n \to 0$ as $n\to\infty$, Lemma 2 entails that $\hat{k}_n^\star \xrightarrow{\mathbb{P}} \infty$ and $\hat{k}_n^\star/n \xrightarrow{\mathbb{P}} 0$. Therefore, for n large enough, $\hat{k}_n^\star < n$ almost surely. Besides, for all $m_n \in \{1,\ldots,n\}$,

$$\frac{\sqrt{\hat{k}_n^\star}}{\log d_n}\left(\frac{\hat{q}_n(\alpha_n,k_n,\hat{k}_n^\star)}{q(\alpha_n)}-1\right)|\{\hat{k}_n^\star = m_n\} \stackrel{d}{=} \frac{\sqrt{m_n}}{\log d_n}\left(\frac{\hat{q}_n(\alpha_n,k_n,m_n)}{q(\alpha_n)}-1\right).$$

By [45, Lemma 8], since $\hat{k}_n^\star \in \{1,\ldots,n-1\}$ and $\hat{k}_n^\star \xrightarrow{\mathbb{P}} \infty$, it is enough to show that the desired convergence (29) holds with $\hat{q}_n(\alpha_n,k_n,\hat{k}_n^\star)$ replaced by its de-conditioned version $\hat{q}_n(\alpha_n,k_n,m_n)$. This is a direct consequence of Corollary 2. The second and final step consists in replacing $\hat{k}_n^\star$ by its non random version $k_n^\star$ in the rate of convergence of (29). This can be achieved using Lemma 2 and Slutsky's lemma. ■

References

[1] Beirlant, J., Dierckx, G., and Guillou, A. (2005). Estimation of the extreme-value index and generalized quantile plots. *Bernoulli*, 11(6), 949–970.

[2] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley.

[3] Beran, J., Schell, D., and Stehlík, M. (2014). The harmonic moment tail index estimator: asymptotic distribution and robustness. *Annals of the Institute of Statistical Mathematics*, 66(1), 193–220.

[4] Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1989). *Regular variation*, Cambridge university press.

[5] Brilhante, M. F., Gomes, M. I., and Pestana, D. (2013). A simple generalisation of the Hill estimator. *Computational Statistics and Data Analysis*, 57(1), 518–535.

[6] Caeiro, F., Gomes, M. I., and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat*, 3(2), 113–136.

[7] Cai, J. J., Einmahl, J. H., de Haan, L., and Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *Journal of the Royal Statistical Society: Series B*, 417–442.

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

[9] Daouia, A., Gardes, L., Girard, S., and Lekina, A. (2011). Kernel estimators of extreme level curves. *Test*, 20(2), 311–333.

[10] Daouia, A., Gijbels, I., and Stupfler, G. (2019). Extremiles: A new perspective on asymmetric least squares. *Journal of the American Statistical Association*, 114(527), 1366–1381.

[11] Daouia, A., Girard, S., and Stupfler, G. (2018). Estimation of Tail Risk based on Extreme Expectiles. *Journal of the Royal Statistical Society: Series B*, 80, 262–292.

[12] Daouia, A., Girard, S., and Stupfler, G. (2019). Extreme M-quantiles as risk measures: From L^1 to L^p optimization. *Bernoulli*, 25, 264–309.

[13] Daouia, A., Girard, S., and Stupfler, G. (2020). Tail expectile process and risk assessment. *Bernoulli*, 26(1), 531–556.

[14] Dekkers, A. L. M., Einmahl, J. H. J., and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 17, 1833–1855.

[15] Deme, E., Gardes, L., and Girard, S. (2013). On the estimation of the second order parameter for heavy-tailed distributions. *Revstat*, 11, 277–299.

[16] Drees, H., de Haan, L., and Resnick, S. (2000). How to make a Hill plot. *Annals of Statistics*, 28, 254–274.

[17] Dutang, C. and Charpentier, A. (2020). CASdatasets: Insurance datasets. *R package version 1.0-11.*

[18] El Methni, J., Gardes, L., Girard, S., and Guillou, A. (2012). Estimation of extreme quantiles from heavy and light tailed distributions. *Journal of Statistical Planning and Inference*, 142(10), 2735–2747.

[19] El Methni, J., and Stupfler, G. (2017). Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions. *Statistica Sinica*, 27, 907–930.

[20] El Methni, J., and Stupfler, G. (2018). Improved estimators of extreme Wang distortion risk measures for very heavy-tailed distributions. *Econometrics and Statistics*, 6, 129–148.

[21] Fraga Alves, M. I., Gomes, M. I., and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, 60(2), 193–214.

[22] Fraga Alves, M. I., de Haan, L., and Lin, T. (2003). Estimation of the parameter controlling the speed of convergence in extreme value theory. *Mathematical Methods of Statistics*, 12(2), 155–176.

[23] Goegebeur, Y., Beirlant, J., and de Wet, T. (2010). Kernel estimators for the second order parameter in extreme value statistics. *Journal of Statistical Planning and Inference*, 140, 2632–2652.

[24] Goegebeur, Y., Guillou, A., and Schorgen, A. (2014). Nonparametric regression estimation of conditional tails: the random covariate case. *Statistics*, 48(4), 732–755.

[25] Gomes, M. I., Brilhante, M. F., Caeiro, F., and Pestana, D. (2015). A new partially reduced-bias mean-of-order p class of extreme value index estimators. *Computational Statistics and Data Analysis*, 82, 223–237.

[26] Gomes, M. I., Brilhante, M. F., and Pestana, D. (2016). New reduced-bias estimators of a positive extreme value index. *Communications in Statistics-Simulation and Computation*, 45(3), 833–862.

[27] Gomes, M. I., Caeiro, F., Figueiredo, F., Henriques-Rodrigues, L., and Pestana, D. (2020). Reduced-bias and partially reduced-bias mean-of-order-p value-at-risk estimation: a Monte-Carlo comparison and an application. *Journal of Statistical Computation and Simulation*, 90(10), 1735–1752.

[28] Gomes, M. I., Caeiro, F., Figueiredo, F., Henriques-Rodrigues, L., and Pestana, D. (2020). Corrected-Hill versus partially reduced-bias value-at-risk estimation. *Communications in Statistics-Simulation and Computation*, 49(4), 867–885.

[29] Gomes, M. I., de Haan, L., and Peng, L. (2002). Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes*, 5, 387–414.

[30] Gomes, M. I., and Pestana, D. (2007). A sturdy reduced-bias extreme quantile (VaR) estimator. *Journal of the American Statistical Association*, 102(477), 280–292.

[31] de Haan, L., and Ferreira, A. (2007). *Extreme value theory: an introduction.* Springer Science and Business Media.

[32] de Haan, L. and Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, 52(1), 60–70.

[33] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B*, 44(1), 37–42.

[34] Hall, P., and Welsh, A. W. (1985). Adaptive estimates of parameters of regular variation. *Annals of Statistics*, 13, 331–341.

[35] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3, 1163–1174.

[36] Kazama, S., Sato, A., and Kawagoe, S. (2009). Evaluating the cost of flood damage based on changes in extreme rainfall in Japan. *Sustainability Science*, 4(61), `https://doi.org/10.1007/s11625-008-0064-y`

[37] Kratz, M., and Resnick, S. I. (1996). The QQ-estimator and heavy tails. *Stochastic Models*, 12(4), 699–724.

[38] Manjunath, B. G., and Caeiro, F. (2013). `evt0`: Mean of order p, peaks over random threshold Hill and high quantile estimates. *R package version 1.1-3.*

[39] Paulauskas, V., and Vaičiulis, M. (2013). On an improvement of Hill and some other estimators. *Lithuanian Mathematical Journal*, 53(3), 336–355.

[40] Peng, L. (1998). Asymptotic unbiased estimators for the extreme value index. *Statistics and Probability Letters*, 38, 107–115.

[41] Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1), 119–131.

[42] Resnick, S., and Stărică, C. (1997). Smoothing the Hill estimator. *Advances in Applied Probability*, 29, 271–293.

[43] Rootzén, H., and Tajvidi, N. (1997). Extreme value statistics and wind storm losses: a case study. *Scandinavian Actuarial Journal*, 1, 70–94.

[44] Schultze, J., and Steinebach, J. (1996). On least squares estimates of an exponential tail coefficient. *Statistics and Risk Modeling*, 14(4), 353–372.

[45] Stupfler, G. (2019). On a relationship between randomly and non-randomly thresholded empirical average excesses for heavy tails. *Extremes*, 22, 749–769.

[46] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364), 812–815.

[47] de Wet, T., Goegebeur, Y., and Munch, M. R. (2012). Asymptotically unbiased estimation of the second order tail parameter. *Statistics and Probability Letters*, 82, 565–573.

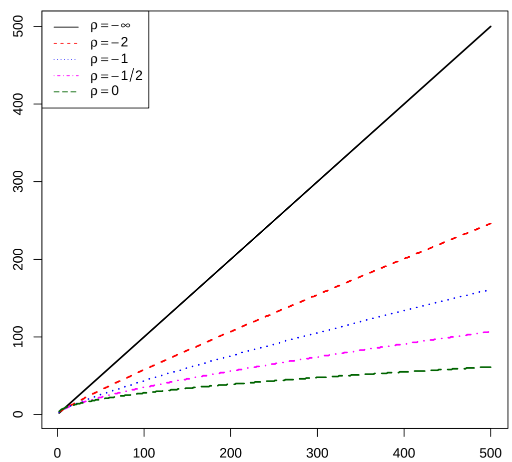

Figure 1: $k_n^{\mathrm{H},\star}$ as a function of k_n for $\rho \in \{-\infty, -2, -1, -1/2, 0\}$ when $\alpha_n = 1/n$ *i.e.* $d_n = k_n$ and $n = 500$.

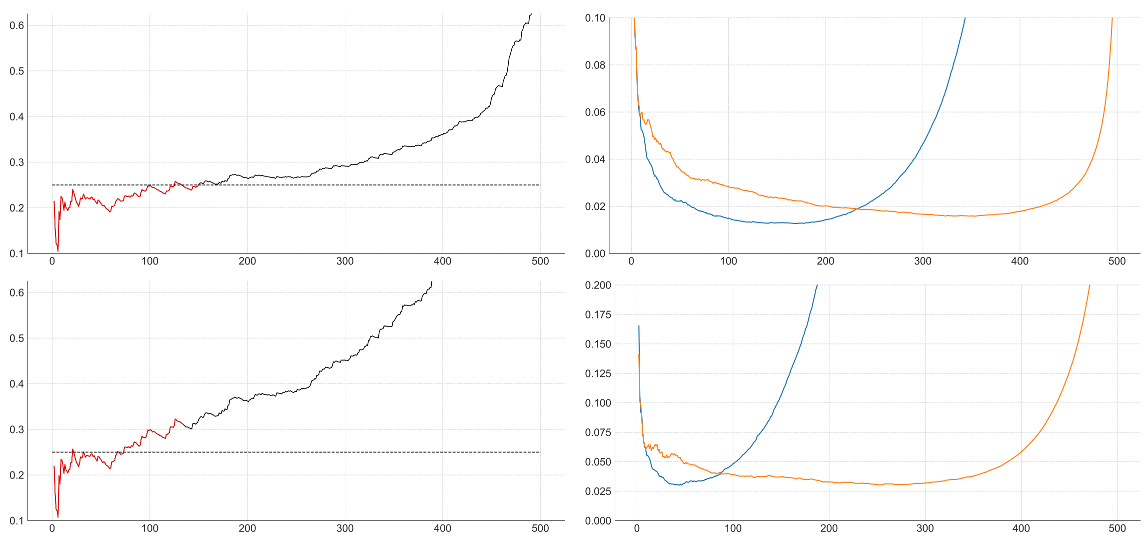

Figure 2: Illustration on simulated data sets of size $n = 500$ from a Burr distribution with $\gamma = 1/4$. Top: $\rho = -2$ and bottom: $\rho = -3/4$. Left panel: Hill estimators $\mathrm{H}(k_n)$ (black) and $\mathrm{H}(\hat{k}_n^{\mathrm{H},\star})$ (red) as functions of k_n. The true value of $\gamma = 1/4$ is depicted by a black horizontal dashed line. Right panel: RMSEs as functions of k_n computed on $N = 1000$ replications associated with Weissman estimator $\hat{q}_n(\alpha_n = 1/n, k_n, k_n)$ (blue) and the refined version $\hat{q}_n(\alpha_n = 1/n, k_n, \hat{k}_n^{\mathrm{H},\star})$ (orange).

Burr	Weissman	RW	CW	CH	CH_p	PRB_p	$CH_{p^\star}$	$PRB_{p^\star}$
$\gamma = 1/8$								
$\rho = -1/8$	-	0.0653	-	0.2376	0.1291	0.2340	0.1433	**0.0429**
$\rho = -1/4$	-	0.0229	0.1752	0.0224	0.0205	0.0409	0.0154	**0.0069**
$\rho = -1/2$	0.2262	0.0105	0.0152	0.0041	0.0060	0.0111	**0.0039**	0.0073
$\rho = -1$	0.0327	0.0068	**0.0047**	0.0138	0.0131	0.0142	0.0121	0.0146
$\rho = -2$	0.0047	**0.0046**	0.0053	0.0096	0.0096	0.0095	0.0088	0.0092
$\gamma = 1/4$								
$\rho = -1/8$	-	**0.2709**	-	-	-	-	0.9261	0.3134
$\rho = -1/4$	-	0.0898	-	0.1116	0.0765	0.1488	0.0771	**0.0336**
$\rho = -1/2$	-	0.0404	0.0701	0.0163	0.0304	0.0433	**0.0156**	0.0272
$\rho = -1$	0.1524	0.0261	**0.0197**	0.0504	0.0529	0.0615	0.0446	0.0528
$\rho = -2$	0.0205	**0.0180**	0.0197	0.0342	0.0361	0.0409	0.0313	0.0327
$\gamma = 1/2$								
$\rho = -1/8$	-	-	-	-	-	-	-	-
$\rho = -1/4$	-	0.4209	-	0.8842	0.7139	0.7783	0.5756	**0.3070**
$\rho = -1/2$	-	0.1637	0.3828	0.0693	0.0889	0.1269	**0.0652**	0.1006
$\rho = -1$	0.8447	0.1035	**0.0936**	0.1785	0.1830	0.1809	0.1589	0.1821
$\rho = -2$	0.0966	0.0725	**0.0715**	0.1106	0.1129	0.1158	0.1022	0.1059

Table 3: RMSEs associated with eight estimators of the extreme quantile $q(\alpha_n = 1/n)$ on a Burr distribution. The best result is emphasized in bold. RMSEs larger than 1 are not reported.

NHW	Weissman	RW	CW	CH	CH_p	PRB_p	$CH_{p^\star}$	$PRB_{p^\star}$
$\gamma = 1/8$								
$\rho = -1/8$	-	**0.0571**	0.1524	0.2897	0.2677	0.3271	0.2376	0.3440
$\rho = -1/4$	-	**0.0169**	0.0260	0.0662	0.0637	0.0835	0.0398	0.1416
$\rho = -1/2$	0.3843	0.0199	0.0155	**0.0040**	0.0068	0.0141	0.0032	0.0187
$\rho = -1$	**0.0020**	0.0151	0.0145	0.0262	0.0236	0.0248	0.0237	0.0257
$\rho = -2$	0.0212	**0.0080**	0.0292	0.0325	0.0245	0.0264	0.0285	0.0284
$\gamma = 1/4$								
$\rho = -1/8$	-	**0.1268**	0.2394	0.3740	0.3534	0.4050	0.3354	0.4123
$\rho = -1/4$	-	**0.0469**	0.0738	0.1880	0.1808	0.2307	0.1533	0.2497
$\rho = -1/2$	0.4341	0.0326	**0.0151**	0.0539	0.0598	0.0785	0.0496	0.0753
$\rho = -1$	**0.0071**	0.0276	0.0286	0.0523	0.0528	0.0588	0.0487	0.0522
$\rho = -2$	0.0240	**0.0185**	0.0344	0.0389	0.0345	0.0360	0.0355	0.0355
$\gamma = 1/2$								
$\rho = -1/8$	-	**0.3991**	0.4637	0.5514	0.5255	0.5508	0.5167	0.5530
$\rho = -1/4$	-	**0.1885**	0.2533	0.4005	0.3923	0.4282	0.3743	0.4359
$\rho = -1/2$	0.5580	0.0986	**0.0925**	0.1936	0.2045	0.2295	0.1865	0.2144
$\rho = -1$	**0.0332**	0.0644	0.0646	0.1060	0.1211	0.1196	0.1026	0.1064
$\rho = -2$	**0.0382**	0.0564	0.0642	0.0796	0.0809	0.0817	0.0734	0.0738

Table 4: RMSEs associated with eight estimators of the extreme quantile $q(\alpha_n = 1/n)$ on a NHW distribution. The best result is emphasized in bold. RMSEs larger than 1 are not reported.

	Weissman	RW	CW	CH	CH_p	PRB_p	$CH_{p^\star}$	$PRB_{p^\star}$
Fréchet ($\rho = -1$)								
$\gamma = 1/8$	0.0098	0.0051	**0.0036**	0.0065	0.0076	0.0075	0.0063	0.0067
$\gamma = 1/4$	0.0438	0.0206	**0.0143**	0.0242	0.0324	0.0343	0.0235	0.0247
$\gamma = 1/2$	0.2199	0.0869	**0.0585**	0.0864	0.1056	0.1061	0.0843	0.0877
Fisher ($\rho = -\gamma$)								
$\gamma = 1/8$	-	**0.050**	0.6780	0.0923	0.0622	0.1225	0.0625	0.0208
$\gamma = 1/4$	-	0.0750	0.4680	0.0413	0.0512	0.0948	0.0332	**0.0252**
$\gamma = 1/2$	-	0.1580	0.1946	0.0931	0.1011	0.1229	**0.0833**	0.1173
$\gamma = 1$	-	**0.5047**	0.5424	0.6111	0.5770	0.5364	0.5708	0.6064
GPD ($\rho = -\gamma$)								
$\gamma = 1/8$	-	0.0653	-	0.2376	0.1128	0.2098	0.1433	**0.0429**
$\gamma = 1/4$	-	0.0898	-	0.1116	0.0890	0.1853	0.0771	**0.0336**
$\gamma = 1/2$	-	0.1637	0.3828	0.0693	0.0898	0.1267	**0.0652**	0.1006
Inverse Gamma ($\rho = -\gamma$)								
$\gamma = 1/8$	0.3024	0.0176	0.0224	0.0069	0.0130	0.0201	**0.0066**	0.0088
$\gamma = 1/4$	0.5980	0.0477	0.0415	**0.0278**	0.0341	0.0459	0.2650	0.0318
$\gamma = 1/2$	-	0.1423	**0.0881**	0.0998	0.1127	0.1203	0.0957	0.1053
$\gamma = 1$	-	0.4552	**0.2800**	0.3188	0.8205	0.6219	0.3163	0.3235
Student ($\rho = -2\gamma$)								
$\gamma = 1/8$	-	0.0379	0.7327	0.1494	0.0937	0.1420	0.0815	**0.0227**
$\gamma = 1/4$	-	0.0556	0.3587	0.0332	0.0323	0.0738	0.0214	**0.0147**
$\gamma = 1/2$	-	0.1155	**0.0533**	0.1165	0.1230	0.1492	0.1031	0.1471
$\gamma = 1$	-	**0.3597**	0.3859	0.4987	0.4804	0.4687	0.4662	0.4875

Table 5: RMSEs associated with eight estimators of the extreme quantile $q(\alpha_n = 1/n)$ on five heavy-tailed distributions. The best result is emphasized in bold. RMSEs larger than 1 are not reported.

Burr	Weissman	RW	CW	CH	CH_p	PRB_p	$CH_{p^\star}$	$PRB_{p^\star}$
$\gamma = 1/8$								
$\rho = -1/8$	-	**0.0898**	-	0.6860	0.4337	0.6315	0.4416	0.1721
$\rho = -1/4$	-	0.0314	0.3085	0.0642	0.0474	0.0906	0.0465	**0.0180**
$\rho = -1/2$	0.3547	0.0148	0.0248	0.0042	0.0054	0.0116	**0.0041**	0.0058
$\rho = -1$	0.0471	0.0095	**0.0063**	0.0155	0.0149	0.0150	0.0135	0.0164
$\rho = -2$	0.0063	**0.0063**	0.0070	0.0118	0.0120	0.0114	0.0108	0.0113
$\gamma = 1/4$								
$\rho = -1/8$	-	**0.4408**	-	-	-	-	-	-
$\rho = -1/4$	-	0.1313	-	0.3484	0.2495	0.4215	0.2467	**0.0982**
$\rho = -1/2$	-	0.0573	0.1179	0.0185	0.0295	0.0505	**0.0176**	0.0229
$\rho = -1$	0.2264	0.0368	**0.0267**	0.0573	0.0588	0.0707	0.0506	0.0602
$\rho = -2$	0.0277	**0.0247**	0.0259	0.0414	0.0434	0.0479	0.0380	0.0396
$\gamma = 1/2$								
$\rho = -1/8$	-	-	-	-	-	-	-	-
$\rho = -1/4$	-	**0.7907**	-	-	-	-	-	-
$\rho = -1/2$	-	0.2372	0.7028	0.0998	0.1017	0.1486	**0.0946**	0.1041
$\rho = -1$	-	0.1473	**0.1360**	0.2136	0.2110	0.2005	0.1897	0.2167
$\rho = -2$	0.1350	0.0998	**0.0931**	0.1319	0.1383	0.1416	0.1224	0.1267

Table 6: RMSEs associated with eight estimators of the extreme quantile $q(\alpha_n = 1/(2n))$ on a Burr distribution. The best result is emphasized in bold. RMSEs larger than 1 are not reported.

NHW	Weissman	RW	CW	CH	CH_p	PRB_p	$\mathrm{CH}_{p^\star}$	$\mathrm{PRB}_{p^\star}$
$\gamma = 1/8$								
$\rho = -1/8$	-	**0.0772**	0.1661	0.2854	0.2596	0.3155	0.2274	0.3433
$\rho = -1/4$	-	**0.0232**	0.0482	0.0527	0.0544	0.0750	0.0260	0.1283
$\rho = -1/2$	0.5514	0.0320	0.0207	0.0041	0.0064	0.0140	**0.0030**	0.0203
$\rho = -1$	**0.0036**	0.0203	0.0241	0.0382	0.0348	0.0360	0.0350	0.0371
$\rho = -2$	0.0293	**0.0107**	0.0398	0.0435	0.0333	0.0360	0.0383	0.0383
$\gamma = 1/4$								
$\rho = -1/8$	-	**0.1817**	0.2945	0.3952	0.3686	0.4196	0.3517	0.4325
$\rho = -1/4$	-	**0.0694**	0.0901	0.1837	0.1700	0.2137	0.1453	0.2492
$\rho = -1/2$	0.6246	0.0492	**0.0197**	0.0615	0.0669	0.0884	0.0560	0.0857
$\rho = -1$	**0.0093**	0.0386	0.0428	0.0694	0.0717	0.0821	0.0649	0.0692
$\rho = -2$	0.0325	**0.0246**	0.0461	0.0511	0.0466	0.0489	0.0467	0.0468
$\gamma = 1/2$								
$\rho = -1/8$	-	**0.5501**	0.7052	0.7224	0.6151	0.6756	0.6621	0.6973
$\rho = -1/4$	-	**0.3065**	0.3254	0.4455	0.4393	0.4634	0.4181	0.4808
$\rho = -1/2$	0.8179	0.1486	**0.1230**	0.2276	0.2591	0.2552	0.2192	0.2489
$\rho = -1$	**0.0404**	0.0858	0.0867	0.1308	0.1458	0.1420	0.1266	0.1312
$\rho = -2$	**0.0490**	0.0767	0.0827	0.0991	0.1077	0.0999	0.0916	0.0921

Table 7: RMSEs associated with eight estimators of the extreme quantile $q(\alpha_n = 1/(2n))$ on a NHW distribution. The best result is emphasized in bold. RMSEs larger than 1 are not reported.

	Weissman	RW	CW	CH	CH_p	PRB_p	$\mathrm{CH}_{p^\star}$	$\mathrm{PRB}_{p^\star}$
Fréchet ($\rho = -1$)								
$\gamma = 1/8$	0.0137	0.0071	**0.0047**	0.0078	0.0091	0.0090	0.0076	0.0080
$\gamma = 1/4$	0.0623	0.0285	**0.0188**	0.0291	0.0352	0.0408	0.0283	0.0298
$\gamma = 1/2$	0.3252	0.1229	**0.0789**	0.1051	0.1551	0.1325	0.1030	0.1069
Fisher ($\rho = -\gamma$)								
$\gamma = 1/8$	-	**0.0715**	-	0.2806	0.1536	0.2339	0.1992	0.0780
$\gamma = 1/4$	-	0.1117	0.8871	0.1282	0.1093	0.1995	0.1030	**0.0458**
$\gamma = 1/2$	-	0.2239	0.3508	0.1120	0.1244	0.1510	**0.1002**	0.1267
$\gamma = 1$	-	0.8794	0.8992	0.8880	0.8556	0.8602	**0.8234**	0.8627
GPD ($\rho = -\gamma$)								
$\gamma = 1/8$	-	**0.0898**	-	0.6860	0.4115	0.5707	0.4416	0.1721
$\gamma = 1/4$	-	0.1313	-	0.3484	0.2413	0.3816	0.2467	**0.0982**
$\gamma = 1/2$	-	0.2372	0.7028	0.0998	0.1137	0.1864	**0.0946**	0.1041
Inverse Gamma ($\rho = -\gamma$)								
$\gamma = 1/8$	0.5005	0.0272	0.0396	0.0091	0.0127	0.0223	**0.0086**	0.0090
$\gamma = 1/4$	-	0.0731	0.0697	0.0343	0.0333	0.0461	**0.0326**	0.0362
$\gamma = 1/2$	-	0.2002	0.1348	0.1253	0.1476	0.1571	**0.1206**	0.1304
$\gamma = 1$	-	0.7207	**0.4044**	0.4274	-	-	0.4261	0.4331
Student ($\rho = -2\gamma$)								
$\gamma = 1/8$	-	**0.0464**	-	0.3724	0.2465	0.3389	0.2225	0.0847
$\gamma = 1/4$	-	0.0736	0.6190	0.0961	0.0573	0.1129	0.0627	**0.0231**
$\gamma = 1/2$	-	0.1582	**0.0756**	0.1177	0.1211	0.1338	0.1054	0.1496
$\gamma = 1$	-	0.5833	**0.5171**	0.6009	0.7234	0.5630	0.5625	0.5860

Table 8: RMSEs associated with eight estimators of the extreme quantile $q(\alpha_n = 1/(2n))$ on five heavy-tailed distributions. The best result is emphasized in bold. RMSEs larger than 1 are not reported.

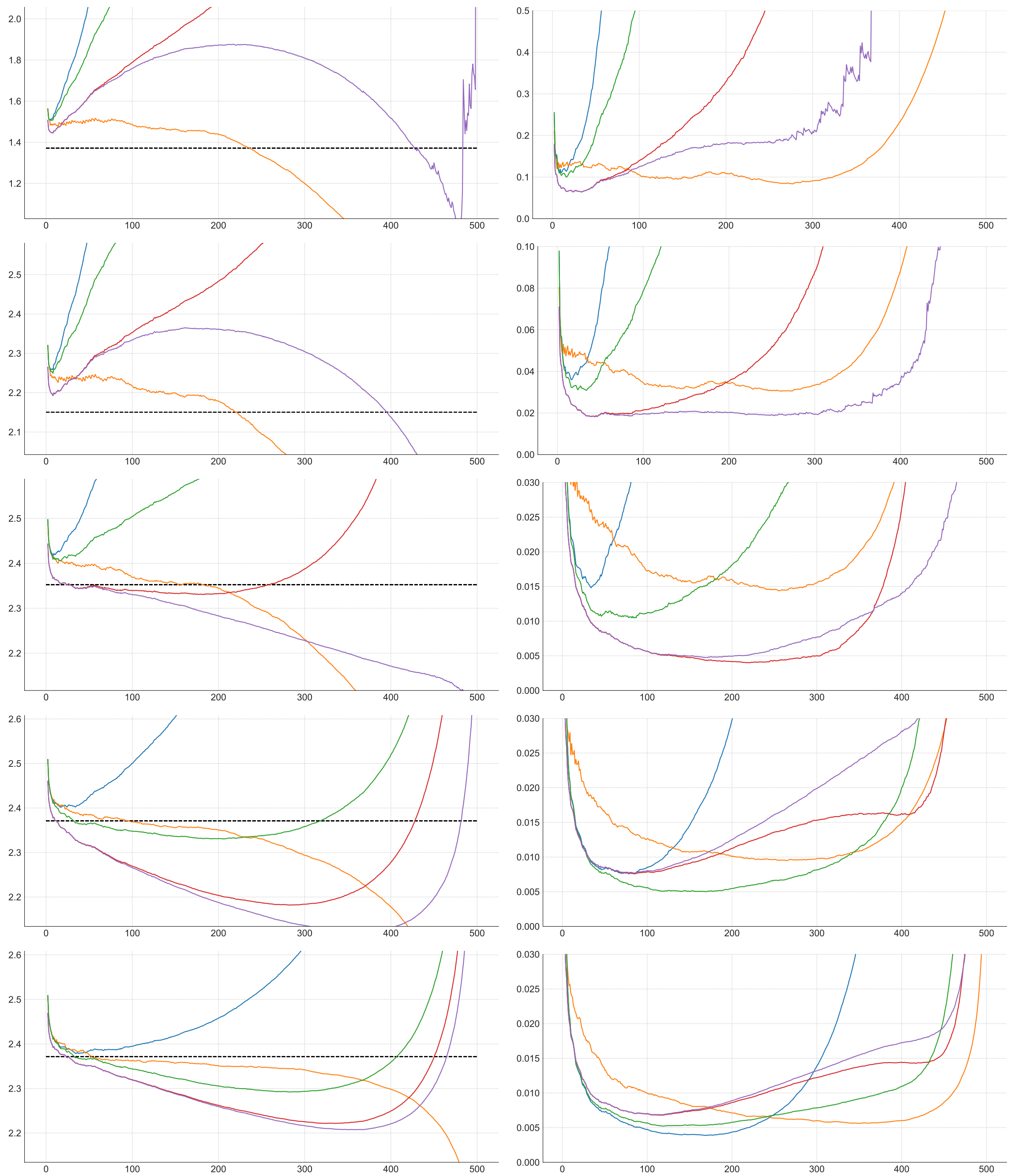

Figure 3: Illustration on simulated data sets of size $n = 500$ from a Burr distribution with $\gamma = 1/8$ and $\rho \in \{-1/8, -1/4, -1/2, -1, -2\}$ (from top to bottom) computed on $N = 1000$ replications. Bias (left panel) and RMSEs (right panel) as functions of $k_n \in \{1, \dots, n-1\}$, associated with estimators Weissman (blue), RW (yellow), CW (green), $\mathrm{CH}_{p^\star}$ (red), $\mathrm{PRB}_{p^\star}$ (purple) of the true extreme quantile $q(\alpha_n = 1/(2n))$ (black dashed line).

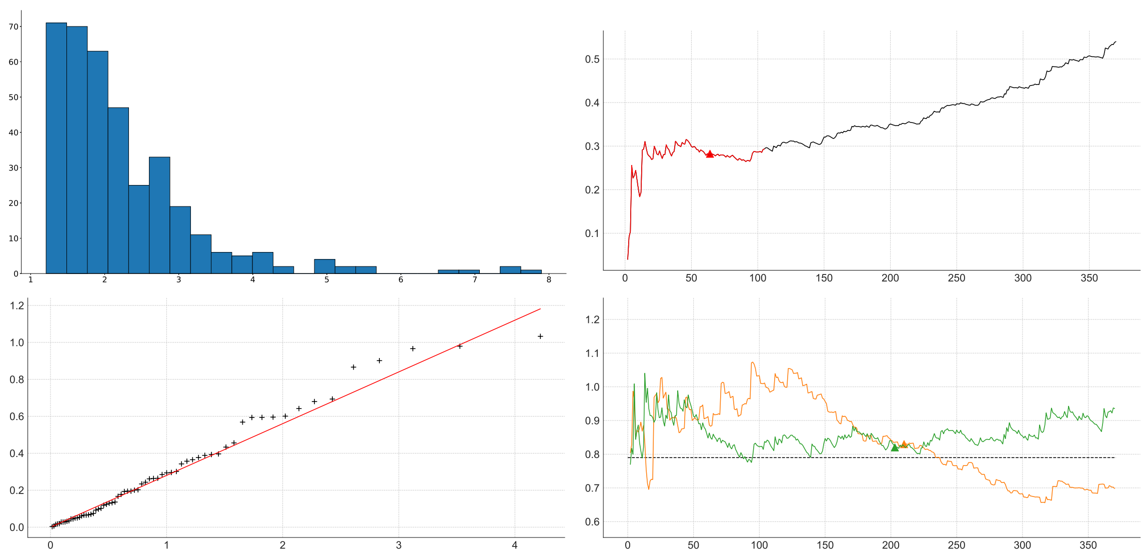

Figure 4: Illustration on the Secura Belgian reinsurance actuarial data set. Top left: Histogram of the data set. The horizontal axis is scaled by 10^{-6}. Top right: Hill estimators $\mathrm{H}(k_n)$ (black) and $\mathrm{H}(\hat{k}_n^{\mathrm{H},\star})$ (red) as functions of k_n. The pair $(\hat{k}_n^{\mathrm{H},\star}, \mathrm{H}(\hat{k}_n^{\mathrm{H},\star}))$ associated with the value of $k_n^\dagger$ selected by Algorithm 1 is emphasized by a triangle. Bottom left: quantile-quantile plot (horizontally: $\log(\hat{k}_n^{\mathrm{H},\star}/i)$, vertically: $\log(X_{n-i+1,n}) - \log(X_{n-\hat{k}_n^{\mathrm{H},\star},n})$ for $i = 1, \dots, \hat{k}_n^{\mathrm{H},\star}$). The regression line with the estimated value of γ as slope is superimposed in red. Bottom right: Estimations of the extreme quantile $q(\alpha_n = 1/n)$ by RW (orange) and CW (green) as functions of k_n with their associated $k_n^\dagger$ emphasized by a triangle. The sample maximum $x_{n,n}$ is depicted by a horizontal black dashed line. The vertical axis is scaled by 10^{-7}.