



HAL
open science

Reduced-bias estimation of the extreme conditional tail expectation for Box-Cox transforms of heavy-tailed distributions

Michaël Allouche, Jonathan El Methni, Stéphane Girard

► To cite this version:

Michaël Allouche, Jonathan El Methni, Stéphane Girard. Reduced-bias estimation of the extreme conditional tail expectation for Box-Cox transforms of heavy-tailed distributions. 2023. hal-04243953v1

HAL Id: hal-04243953

<https://inria.hal.science/hal-04243953v1>

Preprint submitted on 16 Oct 2023 (v1), last revised 11 Apr 2024 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Reduced-bias estimation of the extreme conditional tail expectation for Box-Cox transforms of heavy-tailed distributions

Michaël Allouche⁽¹⁾, Jonathan El Methni⁽²⁾ and Stéphane Girard^(2,*)

⁽¹⁾ Kaiko - Quantitative Data. 2 rue de Choiseul, 75002 Paris, France.

⁽²⁾ Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.

^(*) Corresponding author: `Stephane.Girard@inria.fr`

Abstract

Conditional tail expectation (CTE) is a coherent risk measure defined as the mean of the loss distribution above a high quantile. The existence of the CTE as well as the asymptotic properties of associated estimators however require integrability conditions that may be violated when dealing with heavy-tailed distributions. We introduce Box-Cox transforms of the CTE that have two benefits. First, they alleviate these theoretical issues. Second, they enable to recover a number of risk measures such as conditional tail expectation, expected shortfall, conditional value-at-risk or conditional tail variance. The construction of dedicated estimators is based on the investigation of the asymptotic relationship between Box-Cox transforms of the CTE and quantiles at extreme probability levels, as well as on an extrapolation formula established in the heavy-tailed context. We quantify and estimate the bias induced by the use of these approximations and then introduce reduced-bias estimators whose asymptotic properties are rigorously shown. Their finite-sample properties are assessed on a simulation study and illustrated on real data, highlighting the practical interest of both the bias reduction and the Box-Cox transform.

Keywords: Conditional tail expectation, Heavy-tailed distributions, Asymptotic normality, Risk measures, Extreme-value statistics.

MSC classification codes: 62G32, 62G30, 62E20.

1 Introduction

Value-at-Risk (VaR) is one of the most popular risk measures introduced in the 1990's, see [33] for a review. In statistical terms, the VaR at level $\alpha \in (0, 1)$ is the upper α -quantile of the loss distribution associated with a random variable X :

$$\text{VaR}(\alpha) = q(\alpha) = F^{\leftarrow}(1 - \alpha),$$

with $F^{\leftarrow}(\cdot) = \inf\{x \in \mathbb{R}; F(x) \geq \cdot\}$ the generalized inverse of the cumulative distribution function of X . The VaR however suffers from several weaknesses, see [21, Section 1.3.3] for an overview.

It may be insensitive to the magnitude of extreme events since it only depends on the frequency of tail losses and not on their values. Similarly, random losses with light-tailed distributions or heavy-tailed distributions may have the same Value-at-Risk [42]. Finally, Value-at-Risk is not a coherent risk measure [2] since it is not subadditive in general.

As a consequence, it has been proposed [2, 40] to switch from the usual VaR to the Expected Shortfall (ES). Recall that the ES, also known as Conditional Value at Risk or Average Value at Risk, is defined as the average of the quantile function above a given confidence level $\alpha \in (0, 1)$:

$$\text{ES}(\alpha) = \frac{1}{\alpha} \int_0^\alpha q(u) du. \quad (1)$$

Assuming that X is continuous with density f with respect to the Lebesgue measure, $\text{ES}(\alpha)$ coincides with the conditional expectation of X given that it exceeds the VaR:

$$\text{ES}(\alpha) = \text{CTE}(\alpha) = \mathbb{E}(X|X > q(\alpha)) = \frac{1}{\alpha} \int_{q(\alpha)}^\infty tf(t) dt, \quad (2)$$

and it is referred to as Conditional Tail Expectation (CTE). Unlike the VaR, it can be shown that the CTE satisfies all of the requirements to be a coherent risk measure, namely translation invariance, monotonicity, positive homogeneity and subadditivity [4]. Explicit derivations are available in some parametric families, namely elliptical [35], or phase-type distributions [12], opening the way to parametric estimation. Non-parametric estimators have been introduced in [10, 19] basing on the empirical counterpart of (2):

$$\widehat{\text{CTE}}(\alpha) = \frac{1}{[n\alpha]} \sum_{j=1}^{[n\alpha]} X_{n-j+1,n}, \quad (3)$$

where $X_{1,n} \leq \dots \leq X_{n,n}$ denote the order statistics associated with the initial n -sample and $[\cdot]$ is the integer part.

Let us stress that, from (2), the existence of the CTE requires the random variable X to be integrable while the asymptotic normality of the empirical estimator (3) assumes the existence of the second moment, see [10, Theorem 3.1]. In heavy-tailed models that are the focus of this work, these moment restrictions can be reformulated in terms of conditions on the extreme-value index γ of X . Recall that the random variable X is said to be heavy-tailed with index $\gamma > 0$ if its distribution behaves approximately like a power law distribution with exponent $-1/\gamma$ (see Section 2 for a rigorous definition). For instance, in such a model, the integrability assumption does not hold as soon as the extreme-value index γ driving the right tail heaviness of X is larger than 1 and the finite second moment condition is violated when $\gamma > 1/2$. This can be the case in insurance, see for instance [26] where the extreme-value index is estimated on commercial fire losses at $\hat{\gamma} \simeq 0.7$ and [6, Example 1.2] where $\hat{\gamma} \simeq 0.8$ on Norwegian fire data in 1976 (studied in Section 5 for the year 1990).

As an alternative, we propose to consider the CTE for Box-Cox transforms [9] of X :

$$\mathcal{M}_a(\alpha) = \mathbb{E}(K_a(X)|X > q(\alpha)), \text{ with } K_a(x) = \int_1^x u^{a-1} du, \quad (4)$$

and $a \in \mathbb{R}$. Clearly, considering \mathcal{M}_a permits to recover various risk measures since $\text{CTE}(\alpha) = \mathcal{M}_1(\alpha) + 1$, the conditional tail variance [43] is given by $\text{CTV}(\alpha) = 2(\mathcal{M}_2(\alpha) - \mathcal{M}_1(\alpha)) - \mathcal{M}_1^2(\alpha)$ and

more generally, the conditional tail moment introduced in [24] defined when $a \neq 0$ by $\text{CTM}_a(\alpha) = \mathbb{E}(X^a | X > q(\alpha))$ can be rewritten as $\text{CTM}_a(\alpha) = 1 + a\mathcal{M}_a(\alpha)$. Moreover, we shall show that working with \mathcal{M}_a with $a < 1$ allows us to weaken the integrability assumptions on X linked to the CTE, see Lemma 5(i) in Appendix A. We refer to Table 8 and Appendix B for calculations of \mathcal{M}_a associated with some heavy-tailed distributions, together with the underlying existence conditions. Explicit derivations of the CTM for elliptical and skew-elliptical distributions are provided in [34] and [23] respectively.

Here, we propose non-parametric estimators of $\mathcal{M}_a(\alpha_n)$ in the challenging situation where $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$ and when X is heavy-tailed. Two cases are investigated as $n \rightarrow \infty$: The so-called intermediate case $n\alpha_n \rightarrow \infty$ where $\mathcal{M}_a(\alpha_n)$ can be estimated using order statistics similarly to (3) and the extreme case where $n\alpha_n \rightarrow 0$ where the estimation requires dedicated extrapolation techniques. In each two cases, two estimators are introduced exploiting the heavy tail properties of X (Section 2). The advantages and drawbacks of the proposed estimators are highlighted thanks to asymptotic normality results providing the asymptotic biases and variances (Section 3). Reduced-bias estimators are then introduced featuring no increase of variance. The performance of the proposed estimators are compared on a large scale simulation study in Section 4 and an illustration on a real data set is provided in Section 5. Proofs are postponed to Appendix A.

2 Statistical framework

Let X_1, X_2, \dots, X_n be independent and identically distributed random variables with cumulative distribution function F and let $X_{1,n} \leq \dots \leq X_{n,n}$ denote the associated order statistics. Let (k_n) be an intermediate sequence of integers *i.e.* such that $k_n \in \{1, \dots, n-1\}$, $k_n \rightarrow \infty$ and $k_n/n \rightarrow 0$ as $n \rightarrow \infty$. A direct estimator of the \mathcal{M}_a at the intermediate level k_n/n is obtained by considering the empirical counterpart of (4):

$$\widetilde{\mathcal{M}}_{a,n}^{(D)}(k_n/n) = \frac{1}{k_n} \sum_{i=1}^{k_n} K_a(X_{n-i+1,n}).$$

A similar estimator (3) has been introduced by [10] in the particular case $a = 1$ to estimate the CTE for fixed probability levels, see also [19] for bias corrections. A kernel version is proposed in [41], while the asymptotic properties are established in [13] in a dependent framework but still in the fixed probability level setting.

Let (p_n) be an extreme probability sequence, *i.e.* such that $p_n \rightarrow 0$ and $np_n \rightarrow 0$ as $n \rightarrow \infty$. In such a case, $\mathbb{P}(X_{n,n} < q(p_n)) \rightarrow 1$ as $n \rightarrow \infty$ meaning that the quantile in (4) is out of the sample with probability tending to one. The expectation cannot be estimated by its empirical counterpart anymore and an extrapolation technique is thus necessary. To this end, we suppose that F belongs to the maximum domain attraction of Fréchet with extreme-value index $\gamma > 0$. From [30, Corollary 1.2.10], this is equivalent to assuming that the tail quantile function $x > 1 \mapsto U(x) = q(1/x)$ is regularly-varying with index γ , this property being denoted by $U \in \mathcal{RV}_\gamma$. See [8] for a detailed account on regular variation. It can then be shown that $\mathcal{M}_a(\alpha)$ exists for α small enough provided that $a\gamma < 1$, which will be our minimal working assumption in the sequel, and, moreover,

$$\lim_{\alpha \rightarrow 0} \frac{\mathcal{M}_a(\alpha) - K_a(q(\alpha))}{1 + aK_a(q(\alpha))} = \frac{\gamma}{1 - a\gamma}. \quad (5)$$

See Lemma 5(i) for details and [25, Proposition 1(ii)] for a similar result dedicated to the CTM in the conditional case. The equality in (5) holds for all $\alpha \in (0, 1)$ in the case of strict Pareto distributions. Clearly, the existence condition $a\gamma < 1$ for \mathcal{M}_a is weaker than for the CTE when $a < 1$, and in particular, the existence of $\mathcal{M}_0(\alpha)$ does not impose any condition on γ . It follows from (5) that $x \mapsto K_a^{-1}((1 - a\gamma)\mathcal{M}_a(1/x) - \gamma) \simeq U(x)$ is regularly-varying with index γ and therefore, when n is large,

$$\frac{K_a^{-1}((1 - a\gamma)\mathcal{M}_a(p_n) - \gamma)}{K_a^{-1}((1 - a\gamma)\mathcal{M}_a(k_n/n) - \gamma)} \simeq d_n^\gamma,$$

where $d_n = k_n/(np_n)$ is the extrapolation factor. Remarking that $K_a(\lambda K_a^{-1}(x)) = K_a(\lambda) + x\lambda^a$, for all $x, \lambda > 0$, one has

$$\mathcal{M}_a(p_n) \simeq \mathcal{M}_a(k_n/n) d_n^{a\gamma} + K_a(d_n^\gamma). \quad (6)$$

this formula being exact for strict Pareto distributions. The same extrapolation principle is used in Weissman estimator [44] dedicated to extreme quantiles and in [16, 17] to estimate extreme extremiles and expectiles. As a consequence, one can introduce the following (direct) estimator for $\mathcal{M}_a(p_n)$ at the extreme level p_n :

$$\widehat{\mathcal{M}}_{a,n}^{(D)}(p_n; \hat{\gamma}_n) = \widetilde{\mathcal{M}}_{a,n}^{(D)}(k_n/n) d_n^{a\hat{\gamma}_n} + K_a(d_n^{\hat{\gamma}_n}) = \frac{1}{k_n} \sum_{i=1}^{k_n} K_a(X_{n-i+1,n} d_n^{\hat{\gamma}_n}),$$

where $\hat{\gamma}_n$ is an estimator of γ , for instance the Hill estimator [32]

$$\widehat{\gamma}_n^{(H)}(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \log(X_{n-i+1,n}) - \log(X_{n-k_n,n}), \quad (7)$$

or a biased corrected version such as [14]. A similar estimator is proposed in [24] in the conditional case while this extrapolation principle is applied to multivariate versions of the CTE in [11, 20].

It is also possible to derive from (5) that

$$\mathcal{M}_a(\alpha) \simeq \frac{1}{1 - a\gamma} (K_a(q(\alpha)) + \gamma), \quad (8)$$

and to build an indirect estimator of \mathcal{M}_a at the intermediate level k_n/n by estimating the intermediate quantile $q(k_n/n)$ by the associated order statistic:

$$\widetilde{\mathcal{M}}_{a,n}^{(I)}(k_n/n; \hat{\gamma}_n) = \frac{1}{1 - a\hat{\gamma}_n} (K_a(X_{n-k_n,n}) + \hat{\gamma}_n).$$

Note that $\widetilde{\mathcal{M}}_{0,n}^{(I)}(k_n/n; \hat{\gamma}_n) = \widetilde{\mathcal{M}}_{0,n}^{(D)}(k_n/n)$ if $\hat{\gamma}_n$ is the Hill estimator. This estimator can also be used to build an indirect estimator of \mathcal{M}_a at the extreme level p_n using the same extrapolation principle as in (6):

$$\widehat{\mathcal{M}}_{a,n}^{(I)}(p_n; \hat{\gamma}_n) = \widetilde{\mathcal{M}}_{a,n}^{(I)}(k_n/n; \hat{\gamma}_n) d_n^{a\hat{\gamma}_n} + K_a(d_n^{\hat{\gamma}_n}) = \frac{K_a(X_{n-k_n,n}) d_n^{a\hat{\gamma}_n} + K_a(d_n^{\hat{\gamma}_n}) + \hat{\gamma}_n}{1 - a\hat{\gamma}_n}.$$

Interestingly, remarking that $K_a(\lambda x) = K_a(x)\lambda^a + K_a(\lambda)$ for all $x, \lambda > 0$, one can also rewrite the latter estimator as

$$\widehat{\mathcal{M}}_{a,n}^{(I)}(p_n; \hat{\gamma}_n) = \frac{1}{1 - a\hat{\gamma}_n} (K_a(X_{n-k_n,n} d_n^{\hat{\gamma}_n}) + \hat{\gamma}_n) = \frac{1}{1 - a\hat{\gamma}_n} (K_a(\hat{q}_n(p_n)) + \hat{\gamma}_n),$$

where

$$\hat{q}_n(p_n) = X_{n-k_n, n} d_n^{\hat{\gamma}_n} \quad (9)$$

is the Weissman estimator [44] of the extreme quantile $q(p_n)$, so that $\widetilde{\mathcal{M}}_{a,n}^{(I)}(k_n/n; \hat{\gamma}_n)$ and $\widehat{\mathcal{M}}_{a,n}^{(I)}(p_n; \hat{\gamma}_n)$ share the same structure: Both indirect estimators can be interpreted as a consequence of (8).

The asymptotic normality of all four estimators $\widetilde{\mathcal{M}}_{a,n}^{(D)}(k_n/n)$, $\widetilde{\mathcal{M}}_{a,n}^{(I)}(k_n/n; \hat{\gamma}_n)$, $\widehat{\mathcal{M}}_{a,n}^{(D)}(p_n; \hat{\gamma}_n)$ and $\widehat{\mathcal{M}}_{a,n}^{(I)}(p_n; \hat{\gamma}_n)$ is established in the next section. It will appear that three of them suffer from an asymptotic bias. In each case, a reduced-bias version of the estimator is then introduced.

3 Asymptotic results

To establish the asymptotic distribution of an extreme-value estimator, a second-order condition is usually introduced, it is recalled in Paragraph 3.1. This condition is used to establish the asymptotic normality of direct and indirect estimators of \mathcal{M}_a at the intermediate level k_n/n (Paragraph 3.2) and at the extreme level p_n (Paragraph 3.3).

3.1 Second-order condition

The cornerstone in extreme-value analysis for bias assessment is the following second-order condition that refines the heavy-tail assumption $U \in \mathcal{RV}_\gamma$, $\gamma > 0$. There exist $\gamma > 0$, $\rho \leq 0$ and a positive or negative function A with $A(t) \rightarrow 0$ as $t \rightarrow \infty$ such that

$$\frac{1}{A(t)} \left(x^{-\gamma} \frac{U(tx)}{U(t)} - 1 \right) \rightarrow K_\rho(x), \quad (10)$$

for all $x > 0$ as $t \rightarrow \infty$. Since $A(t) \rightarrow 0$ as $t \rightarrow \infty$, the limit (10) can be equivalently rewritten as

$$\frac{1}{A(t)} (\log U(tx) - \log U(t) - \gamma \log x) \rightarrow K_\rho(x), \quad (11)$$

for all $x > 0$ as $t \rightarrow \infty$. Moreover, it can be shown that $|A|$ is regularly-varying with index ρ . Let us also recall that [22] has shown that the second-order condition (11) implies

$$\left| \frac{\log U(tx) - \log U(t) - \gamma \log x}{A_0(t)} - K_\rho(x) \right| \leq \varepsilon x^{\rho+\varepsilon}, \quad (12)$$

where $A_0(t) \sim A(t)$ as $t \rightarrow \infty$, $\varepsilon > 0$, $t \geq t_0$, $x \geq 1$, see also (3.2.7) in [30].

3.2 Estimators of \mathcal{M}_a at an intermediate level

Let us first focus on the direct estimator of $\mathcal{M}_a(k_n/n)$ for several orders $a \in \{a_1, \dots, a_J\}$. Theorem 1 establishes the joint asymptotic normality of the remormalized random vector $(\widetilde{\mathcal{M}}_{a_j, n}^{(D)}(k_n/n))_{j=1, \dots, J}$.

Theorem 1 (Direct estimator, intermediate level) *Let $(a_1, \dots, a_J) \in \mathbb{R}^J$. Assume the second-order condition (10) holds with $2a_j\gamma < 1$ for all $j \in \{1, \dots, J\}$ and $\rho < 0$. Let (k_n) be an intermediate sequence such that*

$$(C_1) \quad \sqrt{k_n} A(n/k_n) \rightarrow \lambda \in \mathbb{R} \text{ as } n \rightarrow \infty.$$

Then, as $n \rightarrow \infty$, one has the following asymptotic representation:

$$\sqrt{k_n} \left(\frac{\widetilde{\mathcal{M}}_{a_j, n}^{(D)}(k_n/n) - \mathcal{M}_{a_j}(k_n/n)}{1 + a_j \mathcal{M}_{a_j}(k_n/n)} \right)_{j=1, \dots, J} = \xi^{(1)} + o_P(1),$$

where $\xi^{(1)}$ is a centered Gaussian random vector with covariance matrix defined for all $(j, \ell) \in \{1, \dots, J\}^2$ by $C_{j, \ell}^{(D)}(a_1, \dots, a_J, \gamma) = \gamma^2(2 - (a_j + a_\ell)\gamma)/(1 - (a_j + a_\ell)\gamma)$.

It appears that $\widetilde{\mathcal{M}}_{a_j, n}^{(D)}(k_n/n)$ is asymptotically Gaussian and unbiased under the condition $2a_j\gamma < 1$ which is necessary for the existence of the asymptotic variance

$$v^{(D)}(a_j, \gamma) = C_{j, j}^{(D)}(a_1, \dots, a_J, \gamma) = 2\gamma^2 \left(\frac{1 - a_j\gamma}{1 - 2a_j\gamma} \right), \quad j \in \{1, \dots, J\}.$$

Let us note that condition **(C₁)** is used to balance the squared bias of order $A^2(n/k_n)$ and the variance of order $1/k_n$. A similar result is established in [24, Theorem 1] to deal with an estimator of the CTM in a conditional setting, under the stronger assumption $\sqrt{k_n}A(n/k_n) \rightarrow 0$ as $n \rightarrow \infty$. Interestingly, the indirect estimator has different properties in terms of bias and variance:

Theorem 2 (Indirect estimator, intermediate level) *Let $(a_1, \dots, a_J) \in \mathbb{R}^J$. Assume the second-order condition (10) holds with $a_j\gamma < 1$ for all $j \in \{1, \dots, J\}$ and $\rho < 0$. Let (k_n) be an intermediate sequence. Suppose **(C₁)** holds and*

(C₂) $\hat{\gamma}_n$ is an estimator of γ such that $\sqrt{k_n}(\hat{\gamma}_n - \gamma) \xrightarrow{d} \mathcal{N}(\lambda\mu, \sigma^2)$ for some $\mu \in \mathbb{R}$ and $\sigma > 0$.

Then, as $n \rightarrow \infty$, one has the following asymptotic representation:

$$\begin{aligned} \sqrt{k_n} \left(\frac{\widetilde{\mathcal{M}}_{a_j, n}^{(I)}(k_n/n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(k_n/n)}{1 + a_j \mathcal{M}_{a_j}(k_n/n)} \right)_{j=1, \dots, J} &= \gamma u_J \xi^{(2)} + \left(\frac{\sigma}{1 - a_j\gamma} \right)_{j=1, \dots, J} \xi^{(3)} \\ &+ \lambda \left(\frac{\mu}{1 - a_j\gamma} - \frac{1}{1 - a_j\gamma - \rho} \right)_{j=1, \dots, J} + o_P(1), \end{aligned}$$

with $u_J = (1, \dots, 1) \in \mathbb{R}^J$ and where $\xi^{(2)}$ and $\xi^{(3)}$ are two standard Gaussian random variables.

One can first note that, for the indirect estimator, the condition $a_j\gamma < 1$ is weaker than the one associated with the direct estimator. Unlike the direct estimator, $\widetilde{\mathcal{M}}_{a_j, n}^{(I)}(k_n/n; \hat{\gamma}_n)$ suffers from an asymptotic bias that has two sources: the bias induced by the estimation of γ and the bias coming from the approximation (8). Let us consider the example where $\hat{\gamma}_n = \widehat{\gamma}_n^{(H)}$ the Hill estimator (7). In this case, $\mu = 1/(1 - \rho)$ and the asymptotic bias is given by $-a_j\lambda\rho\gamma/((1 - \rho)(1 - a_j\gamma)(1 - a_j\gamma - \rho))$. Besides, $\xi^{(2)}$ and $\xi^{(3)}$ are independent, $\sigma^2 = \gamma^2$, and the variance of the asymptotic marginal Gaussian distribution is given by

$$v^{(I)}(a_j, \gamma) = \gamma^2 \left(1 + \frac{1}{(1 - a_j\gamma)^2} \right) \leq v^{(D)}(a_j, \gamma), \quad j \in \{1, \dots, J\},$$

see Figure 1 for an illustration. To summarize, $\widetilde{\mathcal{M}}_{a_j, n}^{(I)}(k_n/n; \hat{\gamma}_n)$ has a smaller asymptotic variance than $\widetilde{\mathcal{M}}_{a_j, n}^{(D)}(k_n/n)$ but has a non zero asymptotic bias.

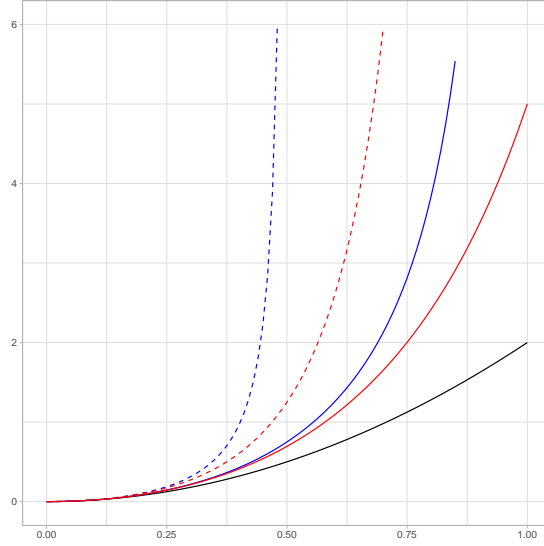


Figure 1: Asymptotic variances $v^{(I)}(a, \gamma)$ (in red) and $v^{(D)}(a, \gamma)$ (in blue) as functions of $\gamma \in (0, 1)$ for $a = 1/2$ (solid lines) and $a = 1$ (dashed lines). The black solid line corresponds to the case $a = 0$ where $v^{(I)}(0, \gamma) = v^{(D)}(0, \gamma) = 2\gamma^2$.

This asymptotic bias may however be cancelled by first, choosing a reduced-bias estimator $\widehat{\gamma}_n^{(\text{RB})}$ of γ so that $\mu = 0$ and second, estimating and removing the bias component given by $-\lambda/(1 - a_j\gamma - \rho)$. To this end, suppose that the auxiliary function in (10) is given by

$$A(t) = \beta\gamma t^\rho(1 + o(1)), \text{ as } t \rightarrow \infty, \quad (13)$$

with $\beta \neq 0$. This situation arises for instance in the Hall-Welsh class of heavy-tailed distributions [31] defined by

$$U(t) = Ct^\gamma(1 + \beta\gamma t^\rho/\rho + o(t^\rho)), \text{ with } C > 0. \quad (14)$$

It is then possible to build from Theorem 2 a bias corrected version of the indirect estimator of \mathcal{M}_a at the intermediate level:

$$\widetilde{\mathcal{M}}_{a,n}^{(\text{I,RB})} \left(k_n/n; \widehat{\gamma}_n^{(\text{RB})} \right) = \widetilde{\mathcal{M}}_{a,n}^{(\text{I})} \left(k_n/n; \widehat{\gamma}_n^{(\text{RB})} \right) \left(1 + \frac{a\hat{A}_n(n/k_n)}{1 - a\widehat{\gamma}_n^{(\text{RB})} - \hat{\rho}_n} \right) + \frac{\hat{A}_n(n/k_n)}{1 - a\widehat{\gamma}_n^{(\text{RB})} - \hat{\rho}_n}$$

with $\hat{A}_n(n/k) = \hat{\beta}_n \widehat{\gamma}_n^{(\text{RB})} (n/k)^{\hat{\rho}_n}$ where $\hat{\rho}_n$ and $\hat{\beta}_n$ are estimators of the second-order parameters ρ and β respectively. Examples of estimators $\widehat{\gamma}_n^{(\text{RB})}$, $\hat{\rho}_n$ and $\hat{\beta}_n$ are provided in Section 4. The next theorem establishes that $\widetilde{\mathcal{M}}_{a,n}^{(\text{I,RB})} \left(k_n/n; \widehat{\gamma}_n^{(\text{RB})} \right)$ is indeed a reduced-bias estimator of $\mathcal{M}_a(k_n/n)$.

Theorem 3 (Indirect estimator, reduced-bias, intermediate level) *Let $(a_1, \dots, a_J) \in \mathbb{R}^J$. Assume the second-order condition (10) holds with $a_j\gamma < 1$ for all $j \in \{1, \dots, J\}$, $\rho < 0$ and with auxiliary function A given by (13). Let (k_n) be an intermediate sequence such that (C1) holds. Suppose moreover, as $n \rightarrow \infty$,*

(C₂') $\widehat{\gamma}_n^{(\text{RB})}$ is a reduced-bias estimator of γ such that $\sqrt{k_n}(\widehat{\gamma}_n^{(\text{RB})} - \gamma) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ for some $\sigma > 0$,

(C₃) $\widehat{\rho}_n$ is an estimator of $\rho < 0$ such that $(\log(n/k_n))(\widehat{\rho}_n - \rho) = o_P(1)$ and $\widehat{\beta}_n$ a consistent estimator of β .

Then, as $n \rightarrow \infty$, one has the following asymptotic representation:

$$\sqrt{k_n} \left(\frac{\widetilde{\mathcal{M}}_{a_j, n}^{(\text{I, RB})} \left(k_n/n; \widehat{\gamma}_n^{(\text{RB})} \right) - \mathcal{M}_{a_j}(k_n/n)}{1 + a_j \mathcal{M}_{a_j}(k_n/n)} \right)_{j=1, \dots, J} = \gamma u_J \xi^{(2)} + \left(\frac{\sigma}{1 - a_j \gamma} \right)_{j=1, \dots, J} \xi^{(3)} + o_P(1),$$

where $\xi^{(2)}$ and $\xi^{(3)}$ are two standard Gaussian random variables.

It is worth noticing that the bias reduction does not yield any increase of the asymptotic variance since the random component $\gamma u_J \xi^{(2)} + \left(\frac{\sigma}{1 - a_j \gamma} \right)_j \xi^{(3)}$ remains the same in Theorem 3 as in Theorem 2.

3.3 Estimators of \mathcal{M}_a at an extreme level

This paragraph is dedicated to the estimation of $\mathcal{M}_a(p_n)$ when $np_n \rightarrow 0$. A key quantity is the extrapolation factor $d_n = k_n/(np_n)$ defined as the ratio between the intermediate level k_n/n and the extreme one p_n . Condition (C₄) hereafter requires that $d_n \rightarrow \infty$ as $n \rightarrow \infty$ not too fast: The extrapolation cannot be conducted too far from the sample, see [30, Remark 4.3.3]. Let us begin with the asymptotic distribution of the direct estimator.

Theorem 4 (Direct estimator, extreme level) Let $(a_1, \dots, a_J) \in \mathbb{R}^J$. Assume the second-order condition (10) holds with $2a_j \gamma < 1$ for all $j \in \{1, \dots, J\}$ and $\rho < 0$. Let (k_n) be an intermediate sequence and introduce (p_n) a probability sequence. Suppose, (C₁), (C₂) hold and

(C₄) $d_n \rightarrow \infty$ and $\log(d_n)/\sqrt{k_n} \rightarrow 0$ as $n \rightarrow \infty$.

Then, as $n \rightarrow \infty$, one has the following asymptotic representation:

$$\frac{\sqrt{k_n}}{\log(d_n)} \left(\frac{\widehat{\mathcal{M}}_{a_j, n}^{(\text{D})} (p_n; \widehat{\gamma}_n) - \mathcal{M}_{a_j}(p_n)}{1 + a_j \mathcal{M}_{a_j}(p_n)} \right)_{j=1, \dots, J} = \sigma u_J \xi^{(4)} + \lambda \left(\mu + \frac{1 - a_j \gamma}{\rho(1 - a_j \gamma - \rho)} \frac{1}{\log(d_n)} \right)_{j=1, \dots, J} + o_P(1),$$

where $\xi^{(4)}$ is a standard Gaussian random variable.

In contrast to the direct estimator at the intermediate level, $\widehat{\mathcal{M}}_{a_j, n}^{(\text{D})} (p_n; \widehat{\gamma}_n)$ has an asymptotic bias involving two components: A first one inherited from the estimator of the extreme-value index and a second one due to the extrapolation method. The same observations have been made in the case of the Weissman estimator dedicated to extreme quantiles in [3]. Let us also note that a similar result has been established in [24, Theorem 2] for the CTM risk measure, in the conditional case, and when $\lambda = 0$, *i.e.* without studying the bias terms. As previously, the asymptotic bias can thus be removed by using a reduced-bias estimator of γ (with $\mu = 0$) and by estimating and removing

the bias term due to the second-order terms. More specifically, the considered reduced-bias version of the direct estimator of \mathcal{M}_a at an extreme level p_n is given by

$$\widehat{\mathcal{M}}_{a,n}^{(D, RB)}(p_n; \widehat{\gamma}_n^{(RB)}) = \widehat{\mathcal{M}}_{a,n}^{(D)}(p_n; \widehat{\gamma}_n^{(RB)}) \left(1 - \frac{a(1 - a\widehat{\gamma}_n^{(RB)}) \widehat{A}_n(n/k_n)}{\widehat{\rho}_n(1 - a\widehat{\gamma}_n^{(RB)} - \widehat{\rho}_n)} \right) - \frac{(1 - a\widehat{\gamma}_n^{(RB)}) \widehat{A}_n(n/k_n)}{\widehat{\rho}_n(1 - a\widehat{\gamma}_n^{(RB)} - \widehat{\rho}_n)}.$$

The next asymptotic normality result illustrates that the asymptotic bias terms have indeed been removed without increase of the asymptotic variance.

Theorem 5 (Direct estimator, reduced-bias, extreme level) *Let $(a_1, \dots, a_J) \in \mathbb{R}^J$. Assume the second-order condition (10) holds with $2a_j\gamma < 1$ for all $j \in \{1, \dots, J\}$, $\rho < 0$ and with auxiliary function A given by (13). Let (k_n) be an intermediate sequence and introduce (p_n) a probability sequence. Suppose (\mathbf{C}_1) , (\mathbf{C}'_2) , (\mathbf{C}_3) and (\mathbf{C}_4) hold. Then, as $n \rightarrow \infty$, one has the following asymptotic representation:*

$$\frac{\sqrt{k_n}}{\log(d_n)} \left(\frac{\widehat{\mathcal{M}}_{a_j,n}^{(D, RB)}(p_n; \widehat{\gamma}_n^{(RB)}) - \mathcal{M}_{a_j}(p_n)}{1 + a_j\mathcal{M}_{a_j}(p_n)} \right)_{j=1, \dots, J} = \sigma u_J \xi^{(4)} + o_P(1),$$

where $\xi^{(4)}$ is a standard Gaussian random variable.

Let us now turn to the indirect estimator that features different bias properties.

Theorem 6 (Indirect estimator, extreme level) *Let $(a_1, \dots, a_J) \in \mathbb{R}^J$. Assume the second-order condition (10) holds with $a_j\gamma < 1$ for all $j \in \{1, \dots, J\}$ and $\rho < 0$. Let (k_n) be an intermediate sequence and introduce (p_n) a probability sequence. Suppose (\mathbf{C}_1) , (\mathbf{C}_2) and (\mathbf{C}_4) hold. Then, as $n \rightarrow \infty$, one has the following asymptotic representation:*

$$\frac{\sqrt{k_n}}{\log(d_n)} \left(\frac{\widehat{\mathcal{M}}_{a_j,n}^{(I)}(p_n; \widehat{\gamma}_n) - \mathcal{M}_{a_j}(p_n)}{1 + a_j\mathcal{M}_{a_j}(p_n)} \right)_{j=1, \dots, J} = \sigma u_J \xi^{(5)} + \lambda \mu \left(1 + \frac{1}{1 - a_j\gamma \log(d_n)} \right)_{j=1, \dots, J} + o_P(1),$$

where $\xi^{(5)}$ is a standard Gaussian random variable.

Interestingly, the asymptotic bias of $\widehat{\mathcal{M}}_{a_j,n}^{(I)}(p_n; \widehat{\gamma}_n)$ is only a consequence of the asymptotic bias of $\widehat{\gamma}_n$. It can be simply removed by considering a reduced-bias estimator of γ as illustrated in the following corollary.

Corollary 1 (Indirect estimator, reduced-bias, extreme level) *Let $(a_1, \dots, a_J) \in \mathbb{R}^J$. Assume the second-order condition (10) holds with $a_j\gamma < 1$ for all $j \in \{1, \dots, J\}$ and $\rho < 0$. Let (k_n) be an intermediate sequence and introduce (p_n) a probability sequence. Suppose (\mathbf{C}_1) , (\mathbf{C}'_2) and (\mathbf{C}_4) hold. Then, as $n \rightarrow \infty$, one has the following asymptotic representation:*

$$\frac{\sqrt{k_n}}{\log(d_n)} \left(\frac{\widehat{\mathcal{M}}_{a_j,n}^{(I)}(p_n; \widehat{\gamma}_n^{(RB)}) - \mathcal{M}_{a_j}(p_n)}{1 + a_j\mathcal{M}_{a_j}(p_n)} \right)_{j=1, \dots, J} = \sigma u_J \xi^{(5)} + o_P(1),$$

where $\xi^{(5)}$ is a standard Gaussian random variable.

The effectiveness of the proposed bias corrections is illustrated on simulated data in the next section.

4 Validation on simulations

The finite sample behavior of five estimators of $\mathcal{M}_a(p_n)$ at the extreme level p_n is first illustrated on simulated data. We consider the following estimators: the direct $\widehat{\mathcal{M}}_{a,n}^{(D)}(p_n; \hat{\gamma}_n)$, the direct with a reduced-bias estimator of the extreme-value index plugged in, namely $\widehat{\mathcal{M}}_{a,n}^{(D)}(p_n; \widehat{\gamma}_n^{(\text{RB})})$, the reduced-bias direct $\widehat{\mathcal{M}}_{a,n}^{(D,\text{RB})}(p_n; \widehat{\gamma}_n^{(\text{RB})})$, the indirect $\widehat{\mathcal{M}}_{a,n}^{(I)}(p_n; \hat{\gamma}_n)$ and the reduced-bias indirect $\widehat{\mathcal{M}}_{a,n}^{(I)}(p_n; \widehat{\gamma}_n^{(\text{RB})})$. In the sequel, they are respectively denoted by $\widehat{\mathcal{M}}_a^{(D)}$, $\widehat{\mathcal{M}}_a^{(D,\text{GRB})}$, $\widehat{\mathcal{M}}_a^{(D,\text{RB})}$, $\widehat{\mathcal{M}}_a^{(I)}$ and $\widehat{\mathcal{M}}_a^{(I,\text{RB})}$. The considered reduced-bias estimator of γ is the Corrected-Hill estimator [14] defined as

$$\widehat{\gamma}_n^{(\text{RB})}(k_n) = \widehat{\gamma}_n^{(\text{H})}(k_n) \left(1 - \frac{\hat{\beta}_n}{1 - \hat{\rho}_n} \left(\frac{n}{k_n} \right)^{\hat{\rho}_n} \right), \quad (15)$$

with $\hat{\rho}_n$ and $\hat{\beta}_n$ estimators of the second-order parameter ρ and β respectively, both dedicated to the Hall-Welsh class of heavy-tailed distributions, see (14). The estimation of ρ has been extensively discussed in the extreme-value literature [15, 28, 29, 31, 38]. See also [18] for a review in the particular case of heavy-tailed distributions. Here, we adopt the estimator $\hat{\rho}_n(\tilde{k})$ of the second-order parameter ρ introduced in [27, Equation (2.18)] and implemented in the package `evt0` of the R software [37] with $\tilde{k} = \lfloor n^{0.999} \rfloor + 1$ as tuning parameter. This package also provides an associated estimator $\hat{\beta}_n$.

In the following, we describe the implementation framework and the experimental design including the simulated heavy-tailed distributions as well as the performance criterion used to compare all considered estimators.

4.1 Implementation

All numerical experiments have been conducted on the Cholesky computing cluster from Ecole Polytechnique http://meso-ipp.gitlab.labos.polytechnique.fr/user_doc. It is composed by 4 nodes, where each one includes 2 CPU Intel Xeon Gold 6230 @ 2.1GHz, 20 cores and 4 Nvidia Tesla v100 graphics card. The code was implemented in Python 3.10.10 and is available for download at <https://github.com/michael-allouche/bctm-extreme.git>

4.2 Experimental design

The comparative study is achieved on seven heavy-tailed distributions: Pareto, Student's t, Fréchet, Inverse gamma, Burr, generalized Pareto distribution (GPD), and Fisher-Snedecor (with $\nu_1 = 1$ degrees of freedom). All aforementioned distributions belong to the Hall-Welsh class (14) and satisfy the second-order condition (10) with (13), see Table 8 for their definitions and associated values of γ and ρ . For all considered distributions, five values of the extreme-value index are considered: $\gamma \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. The Burr distribution has a free second-order parameter, four large values are investigated: $\rho \in \{-1, -1/2, -1/4\}$, while the ρ parameter associated with the six remaining distributions directly depends on γ . We are interested in estimating $\mathcal{M}_a(p_n)$ for $a \in \{1/2, 1\}$ at the extreme level $p_n = 1/(2n)$. All previous considered distributions benefit from a closed-form expression of $\mathcal{M}_a(p_n)$, see Table 8.

For each of these 100 considered configurations, $R = 500$ replicated data sets of size $n = 500$ are simulated and $\mathcal{M}_a(1/(2n))$ is estimated using the five above described estimators. Their performance is assessed using the Relative median-squared error (RMedSE) defined as:

$$\text{RMedSE} \left(\widehat{\mathcal{M}}_a \left(\frac{1}{2n} \right) \right) = \text{median}_{r \in \{1, \dots, R\}} \left(\frac{\widehat{\mathcal{M}}_a^{(r)} \left(\frac{1}{2n} \right)}{\mathcal{M}_a \left(\frac{1}{2n} \right)} - 1 \right)^2, \quad (16)$$

where $\widehat{\mathcal{M}}_a^{(r)} \left(\frac{1}{2n} \right)$ denotes an estimator of $\mathcal{M}_a \left(\frac{1}{2n} \right)$ computed with the anchor index $k_n^*(r)$ selected using the data-driven approach introduced in [3, Algorithm 1] (with initial points $a^{(0)} = \lceil 3n/100 \rceil$ and $c^{(0)} = \lceil 3n/4 \rceil$) on the r -th replication, $r \in \{1, \dots, R\}$.

4.3 Results

The results for $a = 1/2$ (resp. $a = 1$) are provided in Tables 2, 4 and 6 (resp. in Tables 3, 5 and 7).

First observe that, in the Pareto case, both $\widehat{\mathcal{M}}^{(D)}$ the direct and $\widehat{\mathcal{M}}^{(I)}$ the indirect estimators benefit from the perfect parametric form ($\rho = -\infty$), with γ the only parameter to estimate, explaining their close and good results. In contrast, for all the other considered distributions with finite ρ , they nearly never provide the best RMedSE (only twice out of 90 configurations). These results emphasize the interest for reduced-bias estimators.

Second, when $a = 1$, for all distributions with large second-order parameter ($\rho \geq -1/4$), almost all estimators fail to estimate $\mathcal{M}_1(p_n) = \text{CTE}(p_n) - 1$: In 12 situations out of 50, the RMedSE is larger than 1 for all five considered estimators. This phenomena is mitigated in the case $a = 1/2$ (the RMedSE is larger than 1 for all five considered estimators in only 4 situations out of 50) where the asymptotic variance of the estimators is smaller, highlighting the benefit of the Box-Cox transform.

Third, when $\rho \neq -\infty$ and at least an estimator provides a RMedSE smaller than 1, the direct reduced-bias $\widehat{\mathcal{M}}^{(D, \text{RB})}$ and the indirect reduced-bias $\widehat{\mathcal{M}}^{(I, \text{RB})}$ estimators provide the best results 32 and 28 times respectively out of the remaining 74 cases. One can also remark that $\widehat{\mathcal{M}}^{(D, \text{GRB})}$ yields the best results 12 times, essentially on the Burr distribution with $\rho = -3/4$. In this case, $\widehat{\mathcal{M}}^{(D, \text{RB})}$ provides similar results so that we do not recommend the use of $\widehat{\mathcal{M}}^{(D, \text{GRB})}$ in practice.

To conclude, in view of the previous results, we advise that the indirect reduced-biased estimator $\widehat{\mathcal{M}}^{(I, \text{RB})}$ should be used in the challenging large bias situation where $\rho \geq -0.6$ with $\gamma \leq 0.6$ and that the direct reduced-biased estimator $\widehat{\mathcal{M}}^{(D, \text{RB})}$ should be used in the remaining cases.

As an illustration, the behaviour of both the bias and the RMedSE associated with the five estimators is illustrated in Figure 2 for $a = 1/2$ and Figure 3 for $a = 1$. It appears that, in all cases, the indirect bias-reduced estimator $\widehat{\mathcal{M}}^{(I, \text{RB})}$ enjoys the best performance in terms of bias and a nice RMedSE stability over a wide range of k_n values.

5 Illustration on real data

We consider fire insurance claims of a Norwegian insurance company for the year 1990 analyzed in [7]. It consists of $n = 628$ claims in 1000 kr (Norwegian crown) larger than the threshold 500,000 kr.

The Hill plot $k_n \mapsto \widehat{\gamma}_n^{(H)}(k_n)$ presented in Figure 4a features a nice stability on the range $k_n \in [150, 300]$ indicating a heavy-tail behaviour of the data. More specifically, [3, Algorithm 1] presented in Section 4.2 yields $\widehat{\gamma}_n^{(H)}(k_n^*) = 0.62$ with $k_n^* = 279$. As a comparison, [7] used the procedure described in [5] to select the threshold $k_n^\dagger = 290$ and obtained $\widehat{\gamma}_n^{(H)}(k_n^\dagger) = 0.62$. Additionally, the log quantile-quantile plot of the pairs $(\log((n+1)/i), \log X_{n-i+1,n})$ for $i \in \{1, \dots, k_n^*\}$ drawn on Figure 4b is approximately linear, which constitutes an empirical evidence that the heavy-tail assumption makes sense and that $k_n^* = 279$ is a reasonable choice to estimate the extreme-value index. Besides, the estimated second-order parameter provided by the package `evt0` of the R software [37] is $\hat{\rho}_n = -1.85$, corresponding to a relatively low bias situation.

Let us denote by $\mathcal{X}^{\text{test}} = \{X_i^{\text{test}} = X_{n-i+1,n}, i = 1, \dots, \lceil n(1-\xi) \rceil\}$ the testing set composed by the $\lceil n(1-\xi) \rceil$ largest order statistics with $\xi = 0.9$ and $\mathcal{X}^{\text{train}}$ the remaining $\lceil n\xi \rceil$ smallest data. Clearly, $X_{\lceil n(1-\xi) \rceil, \lceil n(1-\xi) \rceil}^{\text{test}} = X_{n-\lceil n(1-\xi) \rceil+1,n}$ can be interpreted as the empirical estimator computed on $\mathcal{X}^{\text{test}}$ of the extreme quantile $q(p_n = 1/\lceil n\xi \rceil)$ while

$$\widehat{\mathcal{M}}_a^{(\text{EMP})}(1/\lceil n\xi \rceil) = \frac{1}{\lceil n(1-\xi) \rceil} \sum_{i=1}^{\lceil n(1-\xi) \rceil} K_a(X_i^{\text{test}}) = \frac{1}{\lceil n(1-\xi) \rceil} \sum_{i=1}^{\lceil n(1-\xi) \rceil} K_a(X_{n-i+1,n}) \quad (17)$$

is the direct estimator of $\mathcal{M}_a(p_n = 1/\lceil n\xi \rceil)$ computed on $\mathcal{X}^{\text{test}}$.

Our first objective is to compute $\mathcal{M}_a(p_n)$ with $a \in \{1/2, 1\}$ at the extreme risk level $p_n = 1/\lceil n\xi \rceil$ on $\mathcal{X}^{\text{train}}$ using the five above competitors and to compare the results with the empirical estimation computed on $\mathcal{X}^{\text{test}}$ by (17). As it can be seen on Figures 5a and 5b, for respectively $a = 1/2$ and $a = 1$, both estimators $\widehat{\mathcal{M}}^{(I)}$ and $\widehat{\mathcal{M}}^{(D)}$ tend to widely overestimate the empirical estimation $\widehat{\mathcal{M}}^{(\text{EMP})}$, while the three reduced-bias ones slightly underestimate it and benefit from a nice stable behaviour in a large neighborhood around their respective estimated values. In this application, the direct reduced-bias estimator seems to provide the closest estimations to the empirical ones (17) with respectively $\widehat{\mathcal{M}}_{1/2}^{(D, \text{RB})} = 168.0$ versus $\widehat{\mathcal{M}}_{1/2}^{(\text{EMP})} = 167.2$ and $\widehat{\mathcal{M}}_1^{(D, \text{RB})} = 7265.9$ versus $\widehat{\mathcal{M}}_1^{(\text{EMP})} = 8731.0$. The good performance of the direct reduced-bias estimator here is probably due to the low estimated bias ($\hat{\rho}_n = -1.85$) and the large estimated extreme-value index ($\hat{\gamma}_n > 0.6$), which is a behavior already observed on simulated data (Section 4).

Our second goal is to deduce from the previous results some estimates of the CTE at the extreme level $p_n = 1/\lceil n\xi \rceil$. Recalling from Section 1 that, since $\text{CTE}(p_n) = 1 + \mathcal{M}_1(p_n)$, one can straightforwardly set

$$\widehat{\text{CTE}}(p_n; 1) = 1 + \widehat{\mathcal{M}}_1(p_n), \quad (18)$$

to get five estimators of the CTE with $\widehat{\mathcal{M}}_1 \in \{\widehat{\mathcal{M}}_1^{(D)}, \widehat{\mathcal{M}}_1^{(D, \text{GRB})}, \widehat{\mathcal{M}}_1^{(D, \text{RB})}, \widehat{\mathcal{M}}_1^{(I)}, \widehat{\mathcal{M}}_1^{(I, \text{RB})}\}$. Moreover, from (5), one has

$$1 + \frac{1}{2}\mathcal{M}_{1/2}(p_n) \sim \frac{q^{1/2}(p_n)}{1 - \gamma/2}, \quad \text{while } \text{CTE}(p_n) \sim \frac{q(p_n)}{1 - \gamma},$$

as $p_n \rightarrow 0$, so that combining the two approximations yields

$$\text{CTE}(p_n) \sim \frac{1}{1 - \gamma} \left(\left(1 - \frac{\gamma}{2}\right) \left(1 + \frac{1}{2}\mathcal{M}_{1/2}(p_n)\right) \right)^2.$$

Taking advantage of the estimated extreme-value index $\widehat{\gamma}_n^{(H)}(k_n^*) = 0.44$ on $\mathcal{X}^{\text{train}}$, one obtains five additional estimators of the CTE:

$$\widehat{\text{CTE}}(p_n; 1/2) = \frac{1}{1 - \widehat{\gamma}_n^{(H)}(k_n^*)} \left(\left(1 - \frac{\widehat{\gamma}_n^{(H)}(k_n^*)}{2} \right) \left(1 + \frac{1}{2} \widehat{\mathcal{M}}_{1/2}(p_n) \right) \right)^2. \quad (19)$$

All ten resulting estimates are compared in Table 1 to the empirical estimation computed on the testing set by

$$\widehat{\text{CTE}}^{(\text{EMP})}(p_n = 1/\lceil n\xi \rceil) = \frac{1}{\lceil n(1-\xi) \rceil} \sum_{i=1}^{\lceil n(1-\xi) \rceil} X_i^{\text{test}} = \frac{1}{\lceil n(1-\xi) \rceil} \sum_{i=1}^{\lceil n(1-\xi) \rceil} X_{n-i+1,n}. \quad (20)$$

It appears that $\widehat{\text{CTE}}^{(\text{D,RB})}(p_n; 1/2)$ is the closest to the empirical estimator. This highlights the benefit of using the Box-Cox transform with $a < 1$ to obtain a reduced-variance estimate and then back transform the result to the targetted CTE.

	(EMP)	(D)	(D, GRB)	(D, RB)	(I)	(I, RB)
$\widehat{\text{CTE}}(p_n; 1)$		14724	6052	7267	15624	4963
$\widehat{\text{CTE}}(p_n; 1/2)$		16606	6508	7852	17056	5307
$\widehat{\text{CTE}}(p_n)$	8732					

Table 1: Estimation of $\text{CTE}(p_n)$ obtained from (18), (19) and (20) at the risk level $p_n = 1/\lceil n\xi \rceil$ with $\xi = 0.9$.

6 Conclusion

We have introduced reduced-bias estimators of the CTE at intermediate and extreme levels for Box-Cox transformed heavy-tailed random variables. The asymptotic normality of all estimators is established highlighting the interests of this approach. First, the use of Box-Cox transform permits to define and consistently estimate the CTE even for distributions with large extreme-value indices. It also allows to derive estimators for other risk measures such as the CTM [24]. Second, the bias reduction (without increase of the variance) leads to estimators with stable sample path (with respect to the sample fraction) easy to use in practice. This phenomenon is illustrated both on simulated and real data thanks to a Python code freely available for download.

Acknowledgements

This work is further supported by the French National Research Agency (ANR) in the framework of the Investissements d’Avenir Program (ANR-15-IDEX-02). S. Girard 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.

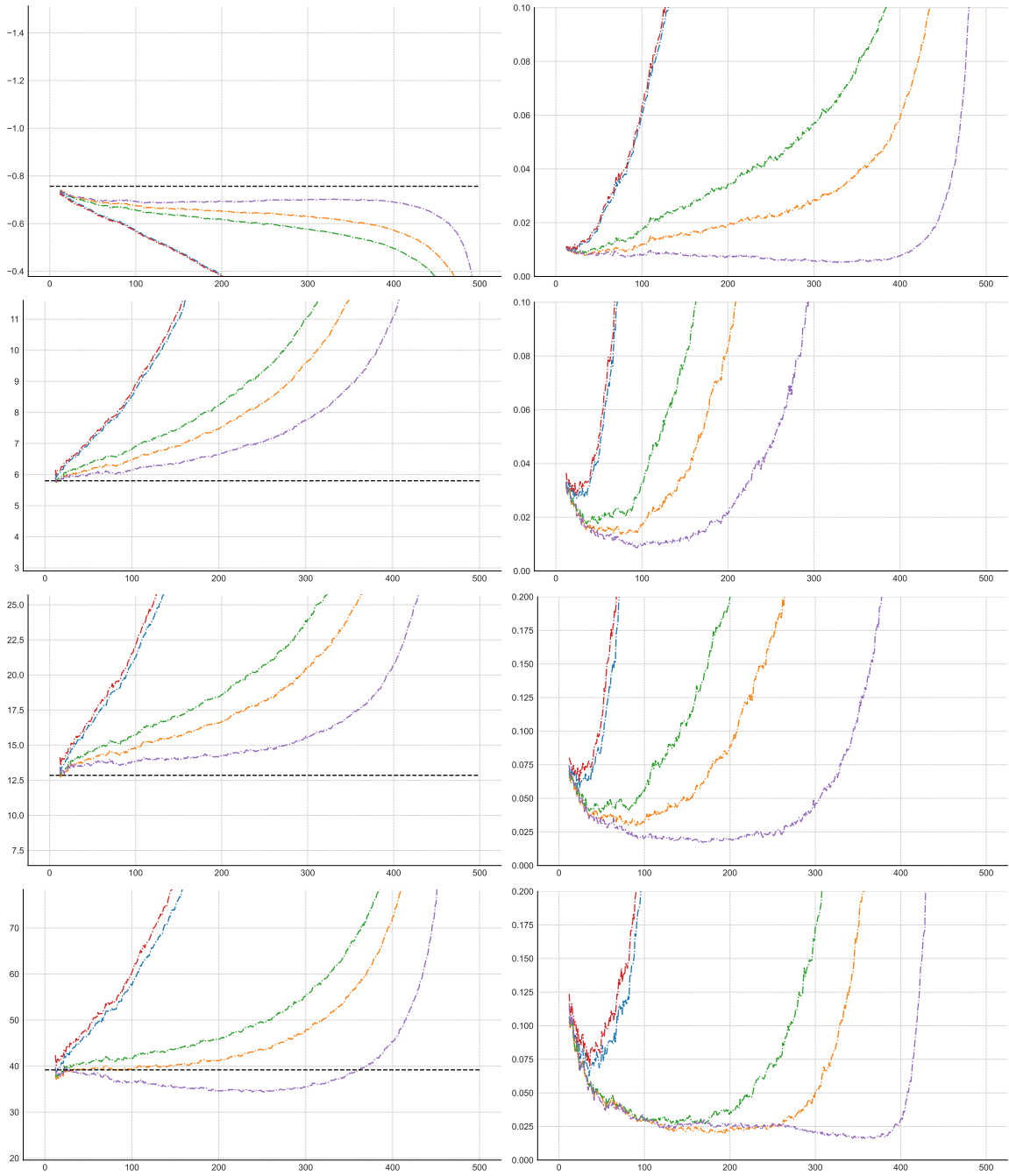


Figure 2: Illustration on simulated data sets of size $n = 500$ from Inverse gamma ($\gamma = 0.1$, $\rho = -0.1$), Student's t ($\gamma = 0.3$, $\rho = -0.6$), Burr ($\gamma = 0.5$, $\rho = -0.5$), GPD ($\gamma = 0.7$, $\rho = -0.7$) distributions from top to bottom. Median of the estimators (left panel) of $\mathcal{M}_{1/2}(p_n = 1/(2n))$ and RMedSE (right panel) as functions of $k_n \in \{12, \dots, n-1\}$ (black dashed line) computed on $R = 500$ replications, associated with $\widehat{\mathcal{M}}_{1/2}^{(D)}$ (blue), $\widehat{\mathcal{M}}_{1/2}^{(D,GRB)}$ (orange), $\widehat{\mathcal{M}}_{1/2}^{(D,RB)}$ (green), $\widehat{\mathcal{M}}_{1/2}^{(I)}$ (red) and $\widehat{\mathcal{M}}_{1/2}^{(I,RB)}$ (purple) estimators.

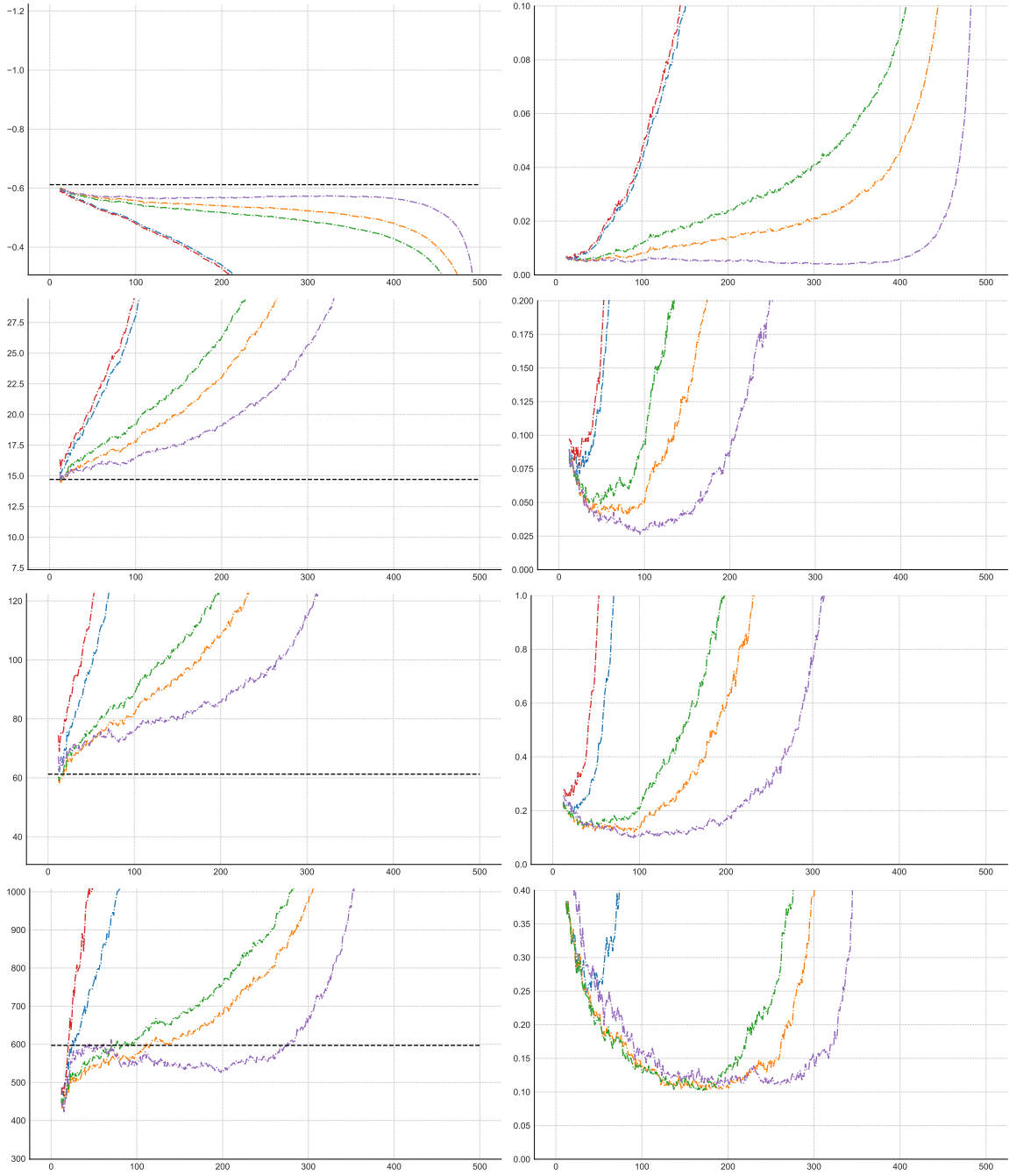


Figure 3: Illustration on simulated data sets of size $n = 500$ from Inverse gamma ($\gamma = 0.1$, $\rho = -0.1$), Student's t ($\gamma = 0.3$, $\rho = -0.6$), Burr ($\gamma = 0.5$, $\rho = -0.5$), GPD ($\gamma = 0.7$, $\rho = -0.7$) distributions from top to bottom. Median of the estimators (left panel) of $\mathcal{M}_1(p_n = 1/(2n))$ and RMedSE (right panel) as functions of $k_n \in \{12, \dots, n-1\}$ (black dashed line) computed on $R = 500$ replications, associated with $\widehat{\mathcal{M}}_1^{(D)}$ (blue), $\widehat{\mathcal{M}}_1^{(D,GRB)}$ (orange), $\widehat{\mathcal{M}}_1^{(D,RB)}$ (green), $\widehat{\mathcal{M}}_1^{(I)}$ (red) and $\widehat{\mathcal{M}}_1^{(I,RB)}$ (purple) estimators.

	$\widehat{\mathcal{M}}_{1/2}^{(D)}$	$\widehat{\mathcal{M}}_{1/2}^{(D,GRB)}$	$\widehat{\mathcal{M}}_{1/2}^{(D,RB)}$	$\widehat{\mathcal{M}}_{1/2}^{(I)}$	$\widehat{\mathcal{M}}_{1/2}^{(I,RB)}$
Pareto ($\rho = -\infty$)					
$\gamma = 0.1$	0.0020	0.0068	0.0043	0.0021	0.0080
$\gamma = 0.3$	0.0042	0.0138	0.0089	0.0042	0.0159
$\gamma = 0.5$	0.0080	0.0253	0.0171	0.0077	0.0302
$\gamma = 0.7$	0.0141	0.0435	0.0298	0.0133	0.0506
$\gamma = 0.9$	0.0224	0.0677	0.0486	0.0221	0.0822
Student's t ($\rho = -2\gamma$)					
$\gamma = 0.1$	-	-	-	-	0.6981
$\gamma = 0.3$	-	0.1520	0.2967	-	0.0409
$\gamma = 0.5$	-	0.0103	0.0168	-	0.0310
$\gamma = 0.7$	0.6753	0.0746	0.0373	0.8036	0.1619
$\gamma = 0.9$	0.3449	0.1792	0.1285	0.4516	0.2995

Table 2: RMedSE associated with five estimators of $\mathcal{M}_{1/2}(p_n = 1/(2n))$ on two heavy-tailed distributions. The best result is emphasized in bold. RMedSEs larger than 1 are not reported.

	$\widehat{\mathcal{M}}_1^{(D)}$	$\widehat{\mathcal{M}}_1^{(D,GRB)}$	$\widehat{\mathcal{M}}_1^{(D,RB)}$	$\widehat{\mathcal{M}}_1^{(I)}$	$\widehat{\mathcal{M}}_1^{(I,RB)}$
Pareto ($\rho = -\infty$)					
$\gamma = 0.1$	0.0030	0.0099	0.0062	0.0030	0.0112
$\gamma = 0.3$	0.0107	0.0331	0.0232	0.0102	0.0398
$\gamma = 0.5$	0.0287	0.0842	0.0616	0.0277	0.1026
$\gamma = 0.7$	0.0731	0.1794	0.1444	0.0705	0.2372
$\gamma = 0.9$	0.2649	0.3652	0.3600	0.2655	0.6777
Student's t ($\rho = -2\gamma$)					
$\gamma = 0.1$	-	-	-	-	-
$\gamma = 0.3$	-	0.5923	-	-	0.1660
$\gamma = 0.5$	-	0.0385	0.0525	-	0.0784
$\gamma = 0.7$	-	0.2270	0.1662	-	0.4350
$\gamma = 0.9$	-	0.6481	0.6091	-	0.8152

Table 3: RMedSE associated with five estimators of $\mathcal{M}_1(p_n = 1/(2n))$ on two heavy-tailed distributions. The best result is emphasized in bold. RMedSEs larger than 1 are not reported.

	$\widehat{\mathcal{M}}_{1/2}^{(D)}$	$\widehat{\mathcal{M}}_{1/2}^{(D,GRB)}$	$\widehat{\mathcal{M}}_{1/2}^{(D,RB)}$	$\widehat{\mathcal{M}}_{1/2}^{(I)}$	$\widehat{\mathcal{M}}_{1/2}^{(I,RB)}$
Fréchet ($\rho = -1$)					
$\gamma = 0.1$	0.0217	0.0057	0.0046	0.0221	0.0092
$\gamma = 0.3$	0.0460	0.0123	0.0093	0.0494	0.0185
$\gamma = 0.5$	0.0919	0.0226	0.0178	0.1032	0.0327
$\gamma = 0.7$	0.1649	0.0386	0.0304	0.1988	0.0551
$\gamma = 0.9$	0.2804	0.0605	0.0525	0.3866	0.0852
Inverse gamma ($\rho = -\gamma$)					
$\gamma = 0.1$	0.3720	0.0227	0.0437	0.3809	0.0061
$\gamma = 0.3$	-	0.0279	0.0658	-	0.0140
$\gamma = 0.5$	0.6263	0.0173	0.0214	0.7044	0.0222
$\gamma = 0.7$	0.4971	0.0295	0.0264	0.6003	0.0447
$\gamma = 0.9$	0.4016	0.0604	0.0512	0.5472	0.0873

Table 4: RMedSE associated with five estimators of $\mathcal{M}_{1/2}(p_n = 1/(2n))$ on two heavy-tailed distributions included in the IGG family (see Appendix B). The best result is emphasized in bold. RMedSEs larger than 1 are not reported.

	$\widehat{\mathcal{M}}_1^{(D)}$	$\widehat{\mathcal{M}}_1^{(D,GRB)}$	$\widehat{\mathcal{M}}_1^{(D,RB)}$	$\widehat{\mathcal{M}}_1^{(I)}$	$\widehat{\mathcal{M}}_1^{(I,RB)}$
Fréchet ($\rho = -1$)					
$\gamma = 0.1$	0.0316	0.0086	0.0065	0.0337	0.0130
$\gamma = 0.3$	0.1232	0.0298	0.0229	0.1454	0.0438
$\gamma = 0.5$	0.3515	0.0738	0.0658	0.5216	0.1061
$\gamma = 0.7$	0.6863	0.1798	0.1589	-	0.2329
$\gamma = 0.9$	0.2447	0.4084	0.3955	-	0.5489
Inverse gamma ($\rho = -\gamma$)					
$\gamma = 0.1$	0.3037	0.0165	0.0311	0.3329	0.0042
$\gamma = 0.3$	-	0.0766	0.1645	-	0.0310
$\gamma = 0.5$	-	0.0696	0.0725	-	0.0677
$\gamma = 0.7$	-	0.1395	0.1252	-	0.1883
$\gamma = 0.9$	0.3632	0.3943	0.3843	-	0.5484

Table 5: RMedSE associated with five estimators of $\mathcal{M}_1(p_n = 1/(2n))$ on two heavy-tailed distributions included in the IGG family (see Appendix B). The best result is emphasized in bold. RMedSEs larger than 1 are not reported.

	$\widehat{\mathcal{M}}_{1/2}^{(D)}$	$\widehat{\mathcal{M}}_{1/2}^{(D,GRB)}$	$\widehat{\mathcal{M}}_{1/2}^{(D,RB)}$	$\widehat{\mathcal{M}}_{1/2}^{(I)}$	$\widehat{\mathcal{M}}_{1/2}^{(I,RB)}$
Burr ($\rho = -1$)					
$\gamma = 0.1$	0.0913	0.0086	0.0036	0.0922	0.0245
$\gamma = 0.3$	0.2118	0.0157	0.0075	0.2224	0.0446
$\gamma = 0.5$	0.4543	0.0255	0.0145	0.5241	0.0735
$\gamma = 0.7$	0.9492	0.0399	0.0244	-	0.1130
$\gamma = 0.9$	-	0.0589	0.0396	-	0.1613
Burr ($\rho = -3/4$)					
$\gamma = 0.1$	0.2219	0.0029	0.0060	0.2268	0.0081
$\gamma = 0.3$	0.5865	0.0060	0.0124	0.6356	0.0148
$\gamma = 0.5$	-	0.0115	0.0222	-	0.0258
$\gamma = 0.7$	-	0.0230	0.0427	-	0.0383
$\gamma = 0.9$	-	0.0398	0.0621	-	0.0551
Burr ($\rho = -1/2$)					
$\gamma = 0.1$	0.7199	0.0265	0.0667	0.7380	0.0048
$\gamma = 0.3$	-	0.0672	0.1513	-	0.0110
$\gamma = 0.5$	-	0.1563	0.3203	-	0.0227
$\gamma = 0.7$	-	0.3472	0.6760	-	0.0525
$\gamma = 0.9$	-	0.7042	-	-	0.1437
Burr ($\rho = -1/4$)					
$\gamma = 0.1$	-	0.5296	0.8414	-	0.2492
$\gamma = 0.3$	-	-	-	-	0.7763
$\gamma = 0.5$	-	-	-	-	-
$\gamma = 0.7$	-	-	-	-	-
$\gamma = 0.9$	-	-	-	-	-
GPD ($\rho = -\gamma$)					
$\gamma = 0.1$	-	-	-	-	0.8246
$\gamma = 0.3$	-	0.5532	0.9113	-	0.2266
$\gamma = 0.5$	-	0.1430	0.2930	-	0.0207
$\gamma = 0.7$	-	0.0305	0.0625	-	0.0264
$\gamma = 0.9$	-	0.0420	0.0303	-	0.1242
Fisher-Snedecor ($\rho = -\gamma$)					
$\gamma = 0.1$	-	0.1803	0.1646	-	-
$\gamma = 0.3$	-	-	-	-	-
$\gamma = 0.5$	-	-	-	-	0.4105
$\gamma = 0.7$	-	0.2909	0.6065	-	0.0477
$\gamma = 0.9$	-	0.0433	0.1248	-	0.0321

Table 6: RMedSE associated with five estimators of $\mathcal{M}_{1/2}(p_n = 1/(2n))$ on three heavy-tailed distributions included in the BFS family (see Appendix B). The best result is emphasized in bold. RMedSEs larger than 1 are not reported.

	$\widehat{\mathcal{M}}_1^{(D)}$	$\widehat{\mathcal{M}}_1^{(D,GRB)}$	$\widehat{\mathcal{M}}_1^{(D,RB)}$	$\widehat{\mathcal{M}}_1^{(I)}$	$\widehat{\mathcal{M}}_1^{(I,RB)}$
Burr ($\rho = -1$)					
$\gamma = 0.1$	0.1393	0.0119	0.0053	0.1445	0.0333
$\gamma = 0.3$	0.6579	0.0319	0.0187	0.7986	0.0918
$\gamma = 0.5$	-	0.0715	0.0495	-	0.1887
$\gamma = 0.7$	-	0.1398	0.1178	-	0.3304
$\gamma = 0.9$	-	0.4262	0.3890	-	0.5720
Burr ($\rho = -3/4$)					
$\gamma = 0.1$	0.3611	0.0045	0.0086	0.3792	0.0111
$\gamma = 0.3$	-	0.0160	0.0313	-	0.0324
$\gamma = 0.5$	-	0.0518	0.0742	-	0.0662
$\gamma = 0.7$	-	0.1109	0.1220	-	0.1257
$\gamma = 0.9$	-	0.2287	0.2145	-	-
Burr ($\rho = -1/2$)					
$\gamma = 0.1$	-	0.0427	0.0996	-	0.0073
$\gamma = 0.3$	-	0.2329	0.4668	-	0.0345
$\gamma = 0.5$	-	0.9919	-	-	0.2427
$\gamma = 0.7$	-	-	-	-	-
$\gamma = 0.9$	-	-	0.3001	-	-
Burr ($\rho = -1/4$)					
$\gamma = 0.1$	-	0.9341	-	-	0.4432
$\gamma = 0.3$	-	-	-	-	-
$\gamma = 0.5$	-	-	-	-	-
$\gamma = 0.7$	-	-	-	-	-
$\gamma = 0.9$	-	-	-	-	-
GPD ($\rho = -\gamma$)					
$\gamma = 0.1$	-	-	-	-	-
$\gamma = 0.3$	-	-	-	-	-
$\gamma = 0.5$	-	0.9759	-	-	0.2388
$\gamma = 0.7$	-	0.1279	0.1957	-	0.1239
$\gamma = 0.9$	-	0.3384	0.3055	-	0.4862
Fisher-Snedecor ($\rho = -\gamma$)					
$\gamma = 0.1$	-	-	-	-	-
$\gamma = 0.3$	-	-	-	-	-
$\gamma = 0.5$	-	-	-	-	-
$\gamma = 0.7$	-	-	-	-	-
$\gamma = 0.9$	-	0.1803	0.1646	-	-

Table 7: RMedSE associated with five estimators of $\mathcal{M}_1(p_n = 1/(2n))$ on three heavy-tailed distributions included in the BFS family (see Appendix B). The best result is emphasized in bold. RMedSEs larger than 1 are not reported.

	Distribution (parameters)	Density function f Risk measure $\mathcal{M}_a(\alpha)$, $a > 0$	γ	ρ
	Pareto ($\theta > a$)	$f(t) = \theta t^{-\theta-1}$, $t \geq 1$ $\mathcal{M}_a(\alpha) = \frac{1}{a} \left(\frac{\theta \alpha^{-a/\theta}}{\theta - a} - 1 \right) = K_a \left(\frac{\theta \alpha^{-1/\theta}}{\theta - a} \right)$	$1/\theta$	$-\infty$
	Student's t ($\nu > a$)	$f(t) = \frac{1}{\sqrt{\nu} B(\nu/2, 1/2)} \left(1 + \frac{t^2}{\nu} \right)^{-(\nu+1)/2}$, $t \in \mathbb{R}$ $\mathcal{M}_a(\alpha) = \frac{1}{a \alpha 2 B(\nu/2, 1/2)} \left(\nu^{\frac{\alpha}{2}} B_{\ell} \left(\frac{1}{1+q^2(\alpha)/\nu}; \frac{\nu-a}{2}, \frac{a+1}{2} \right) - B_{\ell} \left(\frac{1}{1+q^2(\alpha)/\nu}; \frac{\nu}{2}, \frac{1}{2} \right) \right)$	$1/\nu$	$-2/\nu$
IGG	Fréchet ($\theta > a$)	$f(t) = \theta t^{-\theta-1} \exp(-t^{-\theta})$, $t > 0$ $\mathcal{M}_a(\alpha) = \frac{1}{a} \left(\frac{1}{\alpha} \Gamma_{\ell}(1 - a/\theta, -\log(1 - \alpha)) - 1 \right)$	$1/\theta$	-1
	Inverse gamma ($\zeta > a$)	$f(t) = \frac{1}{\Gamma(\zeta)} t^{-\zeta-1} \exp(-1/t)$, $t > 0$ $\mathcal{M}_a(\alpha) = \frac{1}{a} \left(\frac{1}{\alpha} \frac{\Gamma_{\ell}(\zeta - a, 1/q(\alpha))}{\Gamma(\zeta)} - 1 \right)$	$1/\zeta$	$-1/\zeta$
BFS	Burr* ($\zeta > 0, \zeta\theta > a$)	$f(t) = \zeta \theta t^{\zeta-1} (1 + t^{\zeta})^{-\theta-1}$, $t > 0$ $\mathcal{M}_a(\alpha) = \frac{1}{a} \left(\frac{\theta \zeta q(\alpha)^a}{(\theta \zeta - a)} {}_2F_1 \left(-\frac{a}{\zeta}, 1; \theta + 1 - a/\zeta; \frac{1}{1 - \alpha^{-1/\theta}} \right) - 1 \right)$	$1/(\zeta\theta)$	$-1/\theta$
	GPD ($0 < a\xi < 1$)	$f(t) = (1 + \xi t)^{-1-1/\xi}$, $t > 0$ $\mathcal{M}_a(\alpha) = \frac{1}{a} \left(\frac{(\alpha^{-\xi} - 1)^{a-1/\xi}}{\alpha \xi^a (1 - a\xi)} {}_2F_1 \left(\frac{1}{\xi}, \frac{1}{\xi} - a, \frac{1}{\xi} - a + 1; \frac{1}{1 - \alpha^{-\xi}} \right) - 1 \right)$	ξ	$-\xi$
	Fisher-Snedecor ($\nu_1 > 0, \nu_2 > 2a$)	$f(t) = \frac{(\nu_1/\nu_2)^{\nu_1/2}}{B(\nu_1/2, \nu_2/2)} t^{\nu_1/2-1} \left(1 + \frac{\nu_1}{\nu_2} t \right)^{-(\nu_1+\nu_2)/2}$, $t > 0$ $\mathcal{M}_a(\alpha) = \frac{1}{a} \left(\frac{(\nu_1/\nu_2)^{-\nu_2/2} q(\alpha)^{a-\nu_2/2}}{\alpha B(\nu_1/2, \nu_2/2) (\nu_2/2 - a)} {}_2F_1 \left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_2}{2} - a; \frac{\nu_2}{2} - a + 1; -\frac{\nu_2}{\nu_1 q(\alpha)} \right) - 1 \right)$	$2/\nu_2$	$-2/\nu_2$

Table 8: Examples of heavy-tailed distributions (divided in four classes) with the associated values of $\mathcal{M}_a(\alpha)$, γ and ρ . Here, $\Gamma(\cdot)$, $\Gamma_{\ell}(\cdot, \cdot)$, $B(\cdot, \cdot)$, $B_{\ell}(\cdot; \cdot, \cdot)$ and ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$ respectively denote the gamma, lower incomplete gamma, beta, lower incomplete beta and hypergeometric functions. See Appendix B for details.

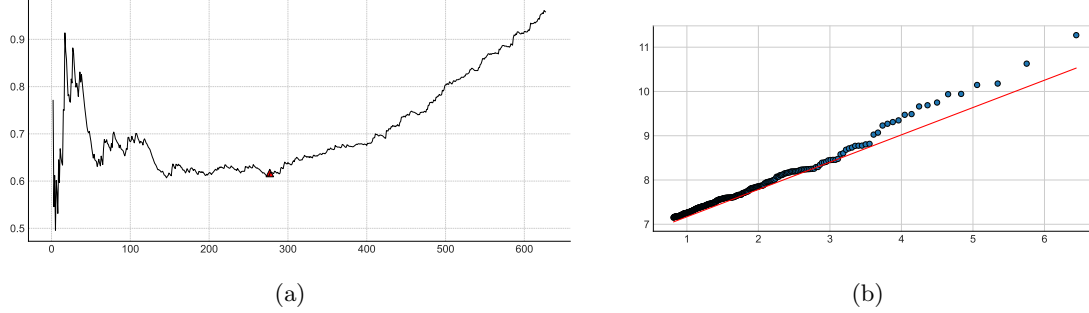


Figure 4: Illustration on real data. (a): Hill estimator $\widehat{\gamma}_n^{(H)}(k_n)$ as a function of k_n and pair $(k_n^*, \widehat{\gamma}_n^{(H)}(k_n^*))$ selected by [3, Algorithm 1] emphasized by a red triangle. (b): Log quantile-quantile plot (horizontally: $\log((n+1)/i)$, vertically: $\log(X_{n-i+1,n})$ for $i \in \{1, \dots, k_n^*\}$). The approximating line with slope $\widehat{\gamma}_n^{(H)}(k_n^*)$ is superimposed in red.

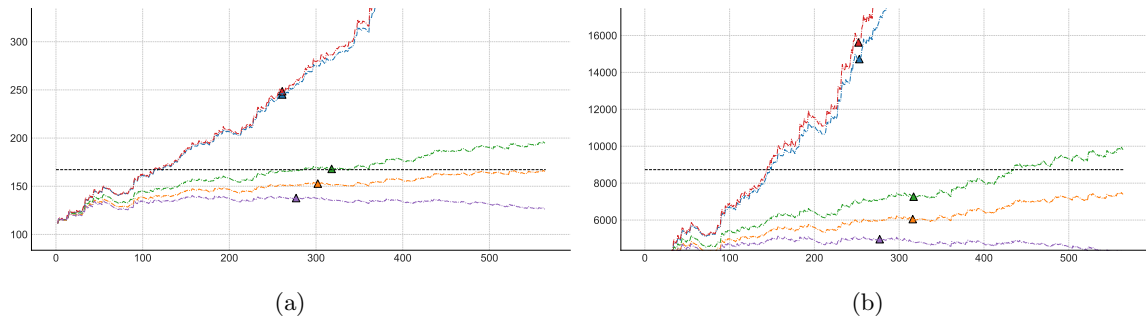


Figure 5: Illustration on real data. (a): estimation of $\mathcal{M}_{1/2}(p_n = 1/\lceil n\xi \rceil)$ and (b) estimation of $\mathcal{M}_1(p_n = 1/\lceil n\xi \rceil)$ associated with $\widehat{\mathcal{M}}^{(\text{EMP})}$ (black), $\widehat{\mathcal{M}}^{(\text{D})}$ (blue), $\widehat{\mathcal{M}}^{(\text{D,GRB})}$ (orange), $\widehat{\mathcal{M}}^{(\text{D,RB})}$ (green), $\widehat{\mathcal{M}}^{(\text{I})}$ (red) and $\widehat{\mathcal{M}}^{(\text{I,RB})}$ (purple) estimators as functions of $k_n \in \{2, \dots, \lceil n\xi \rceil\}$ with $\xi = 0.9$. Estimations at the selected level k_n^* are emphasized by a triangle.

References

- [1] Abramowitz, M. and Stegun, I. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. National Bureau of Standards Applied Mathematics Series, No. 55. U.S. Government Printing Office, Washington, D.C.
- [2] Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26(7):1505–1518.
- [3] Allouche, M., El Methni, J., and Girard, S. (2023). A refined Weissman estimator for extreme quantiles, *Extremes*, 26, 545–572.
- [4] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.
- [5] Beirlant, J., Dierckx, G., Goegebeur, Y., and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes*, 2, 177–200.
- [6] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.
- [7] Beirlant J., Matthys, G., and Dierckx, G. (2001). Heavy-tailed distributions and rating. *ASTIN Bulletin: The Journal of the IAA*, 31(1), 37–58.
- [8] Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1989). *Regular variation*, Cambridge university press.
- [9] Box, G.E. and Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B*, 26, 211–243.
- [10] Brazauskas, V., Jones, B. L., Puri, M. L., and Zitikis, R. (2008). Estimating conditional tail expectation with actuarial applications in view. *Journal of Statistical Planning and Inference*, 138(11), 3590–3604.
- [11] Cai, J. J., Einmahl, J. H. J., 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*, 77(2), 417–442.
- [12] Cai, J., and Li, H. (2005). Conditional tail expectations for multivariate phase-type distributions. *Journal of Applied Probability*, 42(3), 810–825.
- [13] Chen, S. X. (2008). Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, 6(1), 87–107.
- [14] Caeiro, F., Gomes, M. I., and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat – Statistical Journal*, 3(2), 113–136.
- [15] Ciuperca, G., and Mercadier, C. (2010). Semi-parametric estimation for heavy tailed distributions. *Extremes*, 13, 55–87.

- [16] 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.
- [17] 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(2), 263–292.
- [18] Deme, E. H., Gardes, L., and Girard, S. (2013). On the estimation of the second order parameter for heavy-tailed distributions. *Revstat – Statistical Journal*, 11, 277–299.
- [19] Deme, E. H., Girard, S., and Guillou, A. (2015). Reduced-bias estimator of the Conditional Tail Expectation of heavy-tailed distributions. In M. Hallin et al, editors, *Mathematical Statistics and Limit Theorems: Festschrift in Honour of Paul Deheuvels*, 105–123, Springer.
- [20] Di Bernardino, E., and Prieur, C. (2018). Estimation of the multivariate conditional tail expectation for extreme risk levels: Illustration on environmental data sets. *Environmetrics*, 29(7), e2510.
- [21] Dowd, K. (2007). *Measuring Market Risk*, John Wiley & Sons.
- [22] Drees, H. (1998). On smooth statistical tail functionals. *Scandinavian Journal of Statistics*, 25, 187–210.
- [23] Eini, E. J., and Khaloozadeh, H. (2021). Tail conditional moment for generalized skew-elliptical distributions. *Journal of Applied Statistics*, 48(13–15), 2285–2305.
- [24] El Methni, J., Gardes, L., and Girard, S. (2014). Nonparametric estimation of extreme risks from conditional heavy-tailed distributions, *Scandinavian Journal of Statistics*, 41, 988–1012.
- [25] El Methni, J., Gardes, L., and Girard, S. (2018). Kernel estimation of extreme regression risk measures, *Electronic Journal of Statistics*, 12, 359–398.
- [26] 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.
- [27] 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.
- [28] 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.
- [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] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer-Verlag, New-York.
- [31] Hall, P., and Welsh, A. W. (1985). Adaptive estimates of parameters of regular variation. *Annals of Statistics*, 13, 331–341.

- [32] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, 3, 1163–1174.
- [33] Jorion, P. (2007) Value at risk: the new benchmark for managing financial risk, McGraw-Hill, New-York.
- [34] Landsman, Z. M., Makov, U., and Shushi, T. (2016). Tail conditional moments for elliptical and log-elliptical distributions. *Insurance: Mathematics and Economics*, 71, 179–188.
- [35] Landsman, Z. M., and Valdez, E. A. (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7(4), 55–71.
- [36] Mao, A., Stupfler, G., and Yang, F. (2023). Asymptotic properties of generalized shortfall risk measures for heavy-tailed risks. *Insurance: Mathematics and Economics*, 111, 173–192,
- [37] 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*.
- [38] Peng, L. (1998). Asymptotic unbiased estimators for the extreme value index. *Statistics and Probability Letters*, 38, 107–115.
- [39] Ramos, P.L., Mota, A.L., Ferreira, P.H., Ramos, E., Tomazella, V.L.D. and Louzada, F. (2021). Bayesian analysis of the inverse generalized gamma distribution using objective priors. *Journal of Statistical Computation and Simulation*, 91(4):786–816.
- [40] Rockafellar, R.T., and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26(7):1443–1471.
- [41] Scaillet, O. (2004). Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance*, 14, 115–129.
- [42] Tasche, D. (2002). Expected shortfall and beyond. *Journal of Banking and Finance*, 26, 1519–1533.
- [43] Valdez, E.A. (2005). Tail conditional variance for elliptically contoured distributions. *Belgian Actuarial Bulletin*, 5, 26–36.
- [44] 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.

A Appendix: Proofs

Paragraph [A.1](#) collects some preliminary lemmas that will reveal useful in the proofs of the main results provided in Paragraph [A.2](#).

A.1 Preliminary results

Let us begin with a multivariate central limit theorem dedicated to sums involving ratios of Pareto order statistics.

Lemma 1 *Let us define for all $(x, y) \in \mathbb{R}^2$,*

$$S_n(x, y) = \frac{1}{k_n} \sum_{i=0}^{k_n-1} K_x \left(\left(\frac{Y_{n-i,n}}{Y_{n-k_n,n}} \right)^y \right),$$

where $Y_{1,n} \leq \dots \leq Y_{n,n}$ denote the order statistic associated with n independent standard Pareto random variables and (k_n) is an intermediate sequence. Let $(x_1, \dots, x_J) \in \mathbb{R}^J$ and $y > 0$ such that $x_j y < 1/2$ for all $j \in \{1, \dots, J\}$. Then,

$$\sqrt{k_n} \left(S_n(x_j, y) - \frac{y}{1 - x_j y} \right)_{j=1, \dots, J} \xrightarrow{d} \mathcal{N}_J(0, C(x_1, \dots, x_J, y)),$$

with $C(x_1, \dots, x_J, y)$ the covariance matrix defined for all $(j, \ell) \in \{1, \dots, J\}^2$ by

$$C_{j,\ell}(x_1, \dots, x_J, y) = \frac{y^2}{(1 - x_j y)(1 - x_\ell y)(1 - (x_j + x_\ell)y)}.$$

Proof. Let $E_{1,n} \leq \dots \leq E_{n,n}$ denote the order statistic associated with n independent standard exponential random variables E_1, \dots, E_n . Clearly, Renyi's representation yields for all $j \in \{1, \dots, J\}$,

$$\begin{aligned} S_n(x_j, y) &= \frac{1}{k_n} \sum_{i=0}^{k_n-1} K_{x_j}(\exp(y(\log(Y_{n-i,n}) - \log(Y_{n-k_n,n})))) \\ &\stackrel{d}{=} \frac{1}{k_n} \sum_{i=0}^{k_n-1} K_{x_j}(\exp(y(E_{n-i,n} - E_{n-k_n,n}))) \\ &\stackrel{d}{=} \frac{1}{k_n} \sum_{i=0}^{k_n-1} K_{x_j}(\exp(yE_{k_n-i,k_n})) = \frac{1}{k_n} \sum_{i=0}^{k_n-1} K_{x_j}(\exp(yE_i)). \end{aligned}$$

The multivariate Central Limit Theorem entails that $\{\sqrt{k_n}[S_n(x_j, y) - \mathbb{E}(K_{x_j}(\exp(yE_1)))]\}_{j=1, \dots, J}$ is asymptotically centered, Gaussian, with covariance matrix defined for all $(j, \ell) \in \{1, \dots, J\}^2$ by $C_{j,\ell}(x) = \text{cov}(K_{x_j}(\exp(yE_1)), K_{x_\ell}(\exp(yE_1)))$. \blacksquare

The second lemma is of analytical nature, it provides the rate of convergence to 1 for ratios of slowly-varying quantities.

Lemma 2 *Assume the second-order condition [\(10\)](#) holds with $\rho < 0$. Let (p_n) be a probability sequence and (k_n) be an intermediate sequence such that $d_n := k_n/(np_n) \rightarrow \infty$ as $n \rightarrow \infty$. Then, for all $a \in \mathbb{R}$,*

$$K_a \left(\frac{\ell(n/k_n)}{\ell(1/p_n)} \right) = \frac{1}{\rho} A(n/k_n)(1 + o(1)),$$

as $n \rightarrow \infty$.

Proof. The consequence of the second-order condition (12) can be written as

$$\left| \frac{\log \ell(tx) - \log \ell(t)}{A_0(t)} - K_\rho(x) \right| \leq \varepsilon x^{\rho+\varepsilon}.$$

Letting $t = n/k_n \geq 1$ and $x = d_n \rightarrow \infty$, it follows

$$\log \ell(n/k_n) - \log \ell(1/p_n) = \frac{1}{\rho} A_0(n/k_n)(1 + o(1)) = \frac{1}{\rho} A(n/k_n)(1 + o(1)),$$

and thus the result is proved for $a = 0$. As a straightforward consequence, one has, for $a \neq 0$,

$$a(\log \ell(n/k_n) - \log \ell(1/p_n)) = \frac{a}{\rho} A(n/k_n)(1 + o(1)),$$

and consequently,

$$\left(\frac{\ell(n/k_n)}{\ell(1/p_n)} \right)^a = \exp \left(\frac{a}{\rho} A(n/k_n)(1 + o(1)) \right) = 1 + \frac{a}{\rho} A(n/k_n)(1 + o(1)).$$

The result is proved. ■

Lemma 3 is a direct consequence of [30, Theorem 2.4.1] linking an intermediate quantile with its empirical counterpart.

Lemma 3 *Assume the second-order condition (10) holds. Let (k_n) be an intermediate sequence such that (\mathbf{C}_1) holds. Then, for every $a \in \mathbb{R}$,*

$$K_a \left(\frac{X_{n-k_n, n}}{U(n/k_n)} \right) = \frac{\gamma}{\sqrt{k_n}} \xi_n^{(1)} (1 + o_P(1)),$$

where $\xi_n^{(1)} \xrightarrow{d} \mathcal{N}(0, 1)$ as $n \rightarrow \infty$ does not depend on a .

Proof. From [30, Theorem 2.4.1], one has

$$\frac{X_{n-k_n, n}}{U(n/k_n)} = 1 + \frac{\gamma}{\sqrt{k_n}} \xi_n^{(1)},$$

where $\xi_n^{(1)} \xrightarrow{d} \mathcal{N}(0, 1)$. Taking account of $K_a(1) = 0$ and $K'_a(1) = 1$, a first-order Taylor expansion yields

$$K_a \left(\frac{X_{n-k_n, n}}{U(n/k_n)} \right) = \frac{\gamma}{\sqrt{k_n}} \xi_n^{(1)} (1 + o_P(1)),$$

and the result is proved. ■

The next lemma collects some useful asymptotic representations for non-random functions of an asymptotically Gaussian estimator $\hat{\gamma}_n$ of the extreme-value index γ .

Lemma 4 *Assume the second-order condition (10) holds with $a\gamma < 1$. Let (k_n) be an intermediate sequence such that (\mathbf{C}_1) and (\mathbf{C}_2) hold.*

(i) *The following expansion holds:*

$$\frac{\hat{\gamma}_n}{1 - a\hat{\gamma}_n} = \frac{\gamma}{1 - a\gamma} + \frac{\mu}{(1 - a\gamma)^2} A(n/k_n)(1 + o_P(1)) + \frac{\sigma}{(1 - a\gamma)^2} \frac{\xi_n^{(2)}}{\sqrt{k_n}} (1 + o_P(1)),$$

where $\xi_n^{(2)} \xrightarrow{d} \mathcal{N}(0, 1)$ as $n \rightarrow \infty$ does not depend on a .

(ii) If, moreover, (\mathbf{C}_4) holds, then

$$K_a(d_n^{\hat{\gamma}_n - \gamma}) = \frac{\log(d_n)}{\sqrt{k_n}} \left(\mu\sqrt{k_n}A(n/k_n) + \sigma\xi_n^{(2)} \right) + O_P\left(\frac{(\log d_n)^2}{k_n}\right).$$

(iii) If, moreover, $\rho < 0$ and \hat{q}_n is the Weissman estimator (9), then

$$K_a\left(\frac{\hat{q}_n(p_n)}{q(p_n)}\right) = \frac{\log(d_n)}{\sqrt{k_n}} (\mu\sqrt{k_n}A(n/k_n) + \sigma\xi_n^{(2)}) + O_P\left(\frac{(\log d_n)^2}{k_n}\right),$$

Proof. (i) Under (\mathbf{C}_2) , we get

$$\begin{aligned} \frac{1 - a\gamma}{1 - a\hat{\gamma}_n} &= \frac{1}{1 - \frac{a\mu}{1-a\gamma}A(n/k_n) - \frac{a\sigma}{1-a\gamma}\frac{\xi_n^{(2)}}{\sqrt{k_n}}} \\ &= 1 + \frac{a\mu}{1-a\gamma}A(n/k_n)(1 + o_P(1)) + \frac{a\sigma}{1-a\gamma}\frac{\xi_n^{(2)}}{\sqrt{k_n}}(1 + o_P(1)), \end{aligned}$$

and therefore

$$\frac{1}{1 - a\hat{\gamma}_n} = \frac{1}{1 - a\gamma} + \frac{a\mu}{(1 - a\gamma)^2}A(n/k_n)(1 + o_P(1)) + \frac{a\sigma}{(1 - a\gamma)^2}\frac{\xi_n^{(2)}}{\sqrt{k_n}}(1 + o_P(1)).$$

Recalling that

$$\hat{\gamma}_n = \gamma + \mu A(n/k_n) + \frac{\xi_n^{(2)}}{\sqrt{k_n}},$$

and multiplying the two above expansions, the result is proved.

(ii) Observe that

$$K_a(d_n^{\hat{\gamma}_n - \gamma}) = K_a\left(\exp\left(\frac{\log(d_n)}{\sqrt{k_n}}\sqrt{k_n}(\hat{\gamma}_n - \gamma)\right)\right) = K_a\left(\exp\left(\frac{\log(d_n)}{\sqrt{k_n}}(\mu\sqrt{k_n}A(n/k_n) + \sigma\xi_n^{(2)})\right)\right),$$

where $\xi_n^{(2)} \xrightarrow{d} \mathcal{N}(0, 1)$. Recalling that $\log(d_n)/\sqrt{k_n} \rightarrow 0$ and $\sqrt{k_n}A(n/k_n) \rightarrow \lambda$ as $n \rightarrow \infty$ under (\mathbf{C}_1) and (\mathbf{C}_4) , the following first-order expansion holds

$$K_a(d_n^{\hat{\gamma}_n - \gamma}) = \frac{\log(d_n)}{\sqrt{k_n}} (\mu\sqrt{k_n}A(n/k_n) + \sigma\xi_n^{(2)}) + O_P\left(\frac{(\log d_n)^2}{k_n}\right),$$

since $K_a(\exp(u)) = u + O(u^2)$ as $u \rightarrow 0$. The result is thus proved.

(iii) [3, Theorem 1] together with a first-order Taylor expansion entail that

$$\begin{aligned} K_a\left(\frac{\hat{q}_n(p_n)}{q(p_n)}\right) &= K_a\left(1 + \frac{\log(d_n)}{\sqrt{k_n}}(\mu\sqrt{k_n}A(n/k_n) + \sigma\xi_n^{(2)})\right) \\ &= \frac{\log(d_n)}{\sqrt{k_n}} (\mu\sqrt{k_n}A(n/k_n) + \sigma\xi_n^{(2)}) + O_P\left(\frac{(\log d_n)^2}{k_n}\right), \end{aligned}$$

since $K_a(1 + u) = u + O(u^2)$ as $u \rightarrow 0$, and the result is proved. \blacksquare

Lemma 5 introduces first- and second-order approximations of $\mathcal{M}_a(\alpha)$ basing on the quantile evaluated at the same probability level α that can be either intermediate or extreme. See also [36] for first- and second-order asymptotic expansions of generalized shortfall risk measures.

Lemma 5 Suppose $U \in \mathcal{RV}_\gamma$ with $\gamma > 0$ such that $a\gamma < 1$.

(i) Then, $\mathcal{M}_a(\alpha)$ exists for α small enough and, as $\alpha \rightarrow 0$,

$$\frac{\mathcal{M}_a(\alpha) - K_a(U(1/\alpha))}{1 + aK_a(U(1/\alpha))} \rightarrow \frac{\gamma}{1 - a\gamma}.$$

(ii) If, moreover, the second-order condition (10) holds, then, as $\alpha \rightarrow 0$,

$$\frac{K_a(U(1/\alpha)) - \mathcal{M}_a(\alpha)}{1 + a\mathcal{M}_a(\alpha)} = -\gamma - \frac{1 - a\gamma}{1 - a\gamma - \rho} A(1/\alpha)(1 + o(1)).$$

Proof. (i) Let us first remark, that since $U \in \mathcal{RV}_\gamma$ with $\gamma > 0$, one has that $q(\alpha) = U(1/\alpha) \rightarrow \infty$ for $\alpha \rightarrow 0$, so that $K_a(X)$ in (4) is well-defined conditionally on $X > q(\alpha)$ for α small enough. Moreover, denoting by f the density function of X , one has

$$\mathcal{M}_a(\alpha) = \mathbb{E}(K_a(X)|X > q(\alpha)) = \int_{q(\alpha)}^{+\infty} \frac{K_a(u)f(u)}{\mathbb{P}(X > q(\alpha))} du = \frac{1}{\alpha} \int_{q(\alpha)}^{+\infty} K_a(u)f(u) du, \quad (21)$$

and the change of variable $x \mapsto u = U(x/\alpha)$ leads to

$$\mathcal{M}_a(\alpha) = \int_1^{+\infty} x^{-2} K_a(U(x/\alpha)) dx.$$

Note that $x \mapsto x^{-2} K_a(U(x/\alpha))$ is regularly-varying with index $a\gamma - 2 < -1$ and is thus integrable on $[1, \infty)$. This proves the existence of $\mathcal{M}_a(\alpha)$ when $a\gamma < 1$ and α is small enough. It follows that

$$J_a(\alpha) := \frac{\mathcal{M}_a(\alpha) - K_a(U(1/\alpha))}{1 + aK_a(U(1/\alpha))} = \int_1^{+\infty} K_a\left(\frac{U(x/\alpha)}{U(1/\alpha)}\right) \frac{dx}{x^2},$$

and, since U is regularly-varying with index γ , one has $K_a(U(x/\alpha)/U(1/\alpha)) \rightarrow K_a(x^\gamma)$ as $\alpha \rightarrow 0$. Potter bounds [8, Theorem 1.5.6] and Lebesgue's dominated convergence theorem yield

$$J_a(\alpha) \rightarrow \int_1^{+\infty} K_a(x^\gamma) \frac{dx}{x^2} = \frac{\gamma}{1 - a\gamma},$$

as $\alpha \rightarrow 0$, and the first part of the result is proved.

(ii) In view of (12), we have for all $t \geq t_0$ and $x \geq 1$,

$$\log\left(\frac{U(tx)}{U(t)}\right) = \gamma \log(x) + A_0(t)(K_\rho(x) + R(t, x)), \quad (22)$$

with $|R(t, x)| \leq \varepsilon x^{\rho+\varepsilon}$ for all $\varepsilon > 0$, which yields

$$\begin{aligned} K_a\left(\frac{U(tx)}{U(t)}\right) &= K_a(x^\gamma) + K_a(x^\gamma (\exp(A_0(t)(K_\rho(x) + R(t, x)))) - K_a(x^\gamma) \\ &= K_a(x^\gamma) + x^{a\gamma} K_a(\exp(A_0(t)(K_\rho(x) + R(t, x))))). \end{aligned} \quad (23)$$

Taking account of $a\gamma < 1$ and integrating, it follows that

$$\begin{aligned} J_a(1/t) &= \int_1^{+\infty} K_a \left(\frac{U(tx)}{U(t)} \right) \frac{dx}{x^2} \\ &= \int_1^{+\infty} K_a(x^\gamma) \frac{dx}{x^2} + \int_1^{+\infty} x^{a\gamma-2} K_a(\exp(A_0(t)(K_\rho(x) + R(t, x)))) dx \\ &= \frac{\gamma}{1-a\gamma} \end{aligned} \quad (24)$$

$$+ A_0(t) \int_1^{+\infty} x^{a\gamma-2} K_\rho(x) dx \quad (25)$$

$$+ A_0(t) \int_1^{+\infty} x^{a\gamma-2} R(t, x) dx \quad (26)$$

$$+ \int_1^{+\infty} x^{a\gamma-2} (K_a(\exp(u(t, x))) - u(t, x)) dx \quad (27)$$

with $u(t, x) := A_0(t)(K_\rho(x) + R(t, x))$. Straightforward calculations yield:

$$(25) = \frac{A_0(t)}{\rho} \int_1^{+\infty} x^{a\gamma-2+\rho} - x^{a\gamma-2} dx = \frac{A_0(t)}{\rho} \left(\frac{1}{1-a\gamma-\rho} - \frac{1}{1-a\gamma} \right) = \frac{A_0(t)}{(1-a\gamma-\rho)(1-a\gamma)}$$

and, as $t \rightarrow \infty$,

$$|(26)| \leq |A_0(t)|\varepsilon \int_1^{+\infty} x^{a\gamma-2+\rho+\varepsilon} dx = o(A_0(t)),$$

since ε can be chosen arbitrarily small. Moreover, for all $x \geq 1$,

$$|au(t, x)| \leq |aA_0(t)| \left(\varepsilon - \frac{1}{\rho} \right) \leq \log(2),$$

for t large enough, so that $|K_a(\exp(u(t, x))) - u(t, x)| \leq |a|u(t, x)^2$ and therefore

$$\begin{aligned} |(27)| &\leq |a|A_0^2(t) \int_1^{+\infty} x^{a\gamma-2} (K_\rho(x) + R(t, x))^2 dx \\ &\leq 2|a|A_0^2(t) \int_1^{+\infty} x^{a\gamma-2} (K_\rho^2(x) + R^2(t, x)) dx \\ &\leq 2|a|A_0^2(t) \left(\frac{1}{\rho^2} + \varepsilon^2 \right) \int_1^{+\infty} x^{a\gamma-2} dx = o(A_0(t)). \end{aligned}$$

as $t \rightarrow \infty$. Collecting the above results on (24)–(27), it follows that, as $\alpha = 1/t \rightarrow 0$,

$$\begin{aligned} J_a(\alpha) &= \frac{\gamma}{1-a\gamma} + \frac{A_0(1/\alpha)}{(1-a\gamma-\rho)(1-a\gamma)}(1+o(1)), \\ &= \frac{\gamma}{1-a\gamma} + \frac{A(1/\alpha)}{(1-a\gamma-\rho)(1-a\gamma)}(1+o(1)), \end{aligned} \quad (28)$$

since $A_0(1/\alpha) \sim A(1/\alpha)$ as $\alpha \rightarrow 0$. Finally, remarking that

$$\frac{K_a(U(1/\alpha)) - \mathcal{M}_a(\alpha)}{1 + a\mathcal{M}_a(\alpha)} = -\frac{J_a(\alpha)}{1 + aJ_a(\alpha)},$$

and replacing $J_a(\alpha)$ by its expansion (28) conclude the proof. \blacksquare

The next lemma is the random counterpart of Lemma 5, it establishes an asymptotic link between the direct estimator of \mathcal{M}_a at an intermediate level and the associated intermediate order statistic.

Lemma 6 *Assume the second-order condition (10) holds with $2a_j\gamma < 1$ for all $j \in \{1, \dots, J\}$ and $\rho < 0$. Let (k_n) be an intermediate sequence. Then, for all $j \in \{1, \dots, J\}$,*

$$\frac{\widetilde{\mathcal{M}}_{a_j, n}^{(D)}(k_n/n) - K_{a_j}(X_{n-k_n, n})}{1 + a_j K_{a_j}(X_{n-k_n, n})} = \frac{1}{\sqrt{k_n}} \xi_{j, n}^{(3)} + \frac{\gamma}{1 - a_j \gamma} \left(1 + \frac{A(n/k_n)}{\gamma(1 - a_j \gamma - \rho)} (1 + o_P(1)) \right),$$

with $\xi_n^{(3)} = (\xi_{j, n}^{(3)})_{j=1, \dots, J} \xrightarrow{d} \mathcal{N}_J(0, C(a_1, \dots, a_J, \gamma))$ independent from $\xi_n^{(1)}$ and where C is the covariance matrix defined in Lemma 1.

Proof. Let $j \in \{1, \dots, J\}$. First, by definition of $\widetilde{\mathcal{M}}_{a_j, n}^{(D)}(k_n/n)$, one has:

$$\frac{\widetilde{\mathcal{M}}_{a_j, n}^{(D)}(k_n/n) - K_{a_j}(X_{n-k_n, n})}{1 + a_j K_{a_j}(X_{n-k_n, n})} = \frac{1}{k_n} \sum_{i=0}^{k_n-1} K_{a_j} \left(\frac{X_{n-i, n}}{X_{n-k_n, n}} \right) \stackrel{d}{=} \frac{1}{k_n} \sum_{i=0}^{k_n-1} K_{a_j} \left(\frac{U(Y_{n-i, n})}{U(Y_{n-k_n, n})} \right),$$

where $Y_{1, n} \leq \dots \leq Y_{n, n}$ denote the order statistic associated with n independent standard Pareto random variables. Applying the consequence (23) of the second-order condition with $t = Y_{n-k_n, n} \xrightarrow{P} \infty$ and $x = Y_{n-i, n}/Y_{n-k_n, n} \geq 1$, we obtain

$$\frac{\widetilde{\mathcal{M}}_{a_j, n}^{(D)}(k_n/n) - K_{a_j}(X_{n-k_n, n})}{1 + a_j K_{a_j}(X_{n-k_n, n})} = S_n(a_j, \gamma) + \frac{1}{k_n} \sum_{i=0}^{k_n-1} \left(\frac{Y_{n-i, n}}{Y_{n-k_n, n}} \right)^{a_j \gamma} K_{a_j}(\exp(\Delta_{i, n})),$$

with

$$S_n(a_j, \gamma) = \frac{1}{k_n} \sum_{i=0}^{k_n-1} K_{a_j} \left(\left(\frac{Y_{n-i, n}}{Y_{n-k_n, n}} \right)^\gamma \right), \text{ see Lemma 1, and}$$

$$\Delta_{i, n} := A_0(Y_{n-k_n, n}) \left(K_\rho \left(\frac{Y_{n-i, n}}{Y_{n-k_n, n}} \right) + R \left(Y_{n-k_n, n}, \frac{Y_{n-i, n}}{Y_{n-k_n, n}} \right) \right).$$

The two terms are studied separately. First one, condition $a_j \gamma < 1/2$ and Lemma 1 yield

$$S_n(a_j, \gamma) \stackrel{d}{=} \frac{\gamma}{1 - a_j \gamma} + \frac{1}{\sqrt{k_n}} \xi_{j, n}^{(3)}, \quad (29)$$

where $\xi_n^{(3)} = (\xi_{j, n}^{(3)})_{j=1, \dots, J} \xrightarrow{d} \mathcal{N}_J(0, C(a_1, \dots, a_J, \gamma))$. Second, it follows from the consequence of the second-order condition that

$$\max_{i=1, \dots, n} |\Delta_{i, n}| \leq |A_0(Y_{n-k_n, n})| \left(\varepsilon - \frac{1}{\rho} \right),$$

with $A_0(Y_{n-k_n, n}) \stackrel{P}{\sim} A_0(n/k_n) \rightarrow 0$ as $n \rightarrow \infty$. As a consequence, $\Delta_{i, n} \xrightarrow{P} 0$ and $K_{a_j}(\exp(\Delta_{i, n})) \stackrel{P}{\sim} \Delta_{i, n}$ leading to

$$\begin{aligned} \frac{1}{k_n} \sum_{i=0}^{k_n-1} \left(\frac{Y_{n-i, n}}{Y_{n-k_n, n}} \right)^{a_j \gamma} K_{a_j}(\exp(\Delta_{i, n})) &\stackrel{P}{\sim} \frac{1}{k_n} \sum_{i=0}^{k_n-1} \left(\frac{Y_{n-i, n}}{Y_{n-k_n, n}} \right)^{a_j \gamma} \Delta_{i, n} \\ &= \frac{A_0(n/k_n)}{k_n} \sum_{i=0}^{k_n-1} \left(\frac{Y_{n-i, n}}{Y_{n-k_n, n}} \right)^{a_j \gamma} K_\rho \left(\frac{Y_{n-i, n}}{Y_{n-k_n, n}} \right) \end{aligned} \quad (30)$$

$$+ \frac{A_0(n/k_n)}{k_n} \sum_{i=0}^{k_n-1} \left(\frac{Y_{n-i, n}}{Y_{n-k_n, n}} \right)^{a_j \gamma} R \left(Y_{n-k_n, n}, \frac{Y_{n-i, n}}{Y_{n-k_n, n}} \right). \quad (31)$$

Let us consider

$$T_n(x) := \frac{1}{k_n} \sum_{i=0}^{k_n-1} \left(\frac{Y_{n-i,n}}{Y_{n-k_n,n}} \right)^x = xS_n(x, 1) + 1,$$

where $S_n(\cdot, \cdot)$ is defined in Lemma 1. It follows from the latter lemma that

$$\begin{aligned} (30) &= \frac{A_0(n/k_n)}{\rho} (T_n(a_j\gamma + \rho) - T_n(a_j\gamma)) \\ &= \frac{A_0(n/k_n)}{\rho} ((a_j\gamma + \rho)S_n(a_j\gamma + \rho, 1) - a_j\gamma S_n(a_j\gamma, 1)) \\ &= \frac{A_0(n/k_n)}{(1 - a_j\gamma - \rho)(1 - a_j\gamma)} (1 + o_P(1)), \end{aligned} \quad (32)$$

while (12) entails for all $\varepsilon > 0$,

$$|31| \leq \varepsilon |A_0(n/k_n)| \frac{1}{k_n} \sum_{i=0}^{k_n-1} \left(\frac{Y_{n-i,n}}{Y_{n-k_n,n}} \right)^{a_j\gamma + \rho + \varepsilon} = \varepsilon |A_0(n/k_n)| T_n(a_j\gamma + \rho + \varepsilon) = o_P(A_0(n/k_n)). \quad (33)$$

Collecting (29), (32), and (33), it follows

$$\frac{\widetilde{\mathcal{M}}_{a_j,n}^{(D)}(k_n/n) - K_{a_j}(X_{n-k_n,n})}{1 + a_j K_{a_j}(X_{n-k_n,n})} = \frac{1}{\sqrt{k_n}} \xi_{j,n}^{(3)} + \frac{\gamma}{1 - a_j\gamma} \left(1 + \frac{A_0(n/k_n)}{\gamma(1 - a_j\gamma - \rho)} (1 + o_P(1)) \right),$$

and recalling that $A_0(n/k_n) \sim A(n/k_n)$ as $n \rightarrow \infty$ concludes the proof. \blacksquare

Finally, Lemma 7 is used to build bias corrected versions of \mathcal{M}_a estimators.

Lemma 7 *Let $a \in \mathbb{R}$, $\rho < 0$ and (k_n) be an intermediate sequence such that $\sqrt{k_n}(n/k_n)^\rho = O(1)$ as $n \rightarrow \infty$. Let $(\hat{\gamma}_n, \hat{\rho}_n, \hat{\beta}_n)$ be an estimator of (γ, ρ, β) such that $\hat{\gamma}_n \xrightarrow{P} \gamma$ and **(C₃)** holds. Let $\varphi_a : \mathbb{R}_+ \times \mathbb{R}_- \times \mathbb{R} \rightarrow \mathbb{R}$ be a continuous function. Then, the following expansion holds:*

$$\varphi_a(\hat{\gamma}_n, \hat{\rho}_n, \hat{\beta}_n)(n/k_n)^{\hat{\rho}_n} = \varphi_a(\gamma, \rho, \beta)(n/k_n)^\rho + o_P\left(\frac{1}{\sqrt{k_n}}\right),$$

as $n \rightarrow \infty$.

Proof. One has

$$\begin{aligned} &\sqrt{k_n} \left(\varphi_a(\hat{\gamma}_n, \hat{\rho}_n, \hat{\beta}_n)(n/k_n)^{\hat{\rho}_n} - \varphi_a(\gamma, \rho, \beta)(n/k_n)^\rho \right) \\ &= \varphi_a(\gamma, \rho, \beta) \sqrt{k_n} (n/k_n)^\rho \left(\frac{\varphi_a(\hat{\gamma}_n, \hat{\rho}_n, \hat{\beta}_n)(n/k_n)^{\hat{\rho}_n}}{\varphi_a(\gamma, \rho, \beta)(n/k_n)^\rho} - 1 \right), \end{aligned}$$

and, since $\varphi_a(\cdot)$ is continuous,

$$\begin{aligned} \frac{\varphi_a(\hat{\gamma}_n, \hat{\rho}_n, \hat{\beta}_n)(n/k_n)^{\hat{\rho}_n}}{\varphi_a(\gamma, \rho, \beta)(n/k_n)^\rho} - 1 &= \frac{(n/k_n)^{\hat{\rho}_n}}{(n/k_n)^\rho} (1 + o_P(1)) - 1 \\ &= \exp((\hat{\rho}_n - \rho) \log(n/k_n)) (1 + o_P(1)) - 1 \\ &= o_P(1), \end{aligned}$$

by assumption. As a conclusion, taking account of $\sqrt{k_n}(n/k_n)^\rho = O(1)$ as $n \rightarrow \infty$, it follows that

$$\sqrt{k_n} \left(\varphi_a(\hat{\gamma}_n, \hat{\rho}_n, \hat{\beta}_n)(n/k_n)^{\hat{\rho}_n} - \varphi_a(\gamma, \rho, \beta)(n/k_n)^\rho \right) = o_P(1),$$

and the result is proved. \blacksquare

A.2 Proofs of main results

The proofs of Theorem 1–6 all share the same structure, they rely on successive applications of the above lemmas.

Proof of Theorem 1. The following expansion holds for all $j \in \{1, \dots, J\}$:

$$\frac{\widetilde{\mathcal{M}}_{a_j, n}^{(D)}(k_n/n) - \mathcal{M}_{a_j}(k_n/n)}{1 + a_j \mathcal{M}_{a_j}(k_n/n)} = L_{j, n}^{[3]} + \left(L_{j, n}^{[5(ii)]} + L_{j, n}^{[6]} + a_j L_{j, n}^{[5(ii)]} L_{j, n}^{[6]} \right) \left(1 + a_j L_{j, n}^{[3]} \right), \quad (34)$$

and each of the three terms in the previous expansion is studied in a dedicated lemma:

$$\begin{aligned} L_{j, n}^{[3]} &= K_{a_j} \left(\frac{X_{n-k_n, n}}{U(n/k_n)} \right) = \frac{\gamma}{\sqrt{k_n}} \xi_n^{(1)} (1 + o_P(1)) && \text{from Lemma 3,} \\ L_{j, n}^{[5(ii)]} &= \frac{K_{a_j}(U(n/k_n)) - \mathcal{M}_{a_j}(k_n/n)}{1 + a_j \mathcal{M}_{a_j}(k_n/n)} = -\gamma - \frac{1 - a_j \gamma}{1 - a_j \gamma - \rho} A(k_n/n) (1 + o(1)) && \text{from Lemma 5(ii),} \\ L_{j, n}^{[6]} &= \frac{\widetilde{\mathcal{M}}_{a_j, n}^{(D)}(k_n/n) - K_{a_j}(X_{n-k_n, n})}{1 + a_j K_{a_j}(X_{n-k_n, n})} \\ &= \frac{1}{\sqrt{k_n}} \xi_{j, n}^{(3)} + \frac{\gamma}{1 - a_j \gamma} \left(1 + \frac{A(n/k_n)}{\gamma(1 - a_j \gamma - \rho)} (1 + o_P(1)) \right) && \text{from Lemma 6,} \end{aligned}$$

with $\xi_n^{(1)} \xrightarrow{d} \mathcal{N}(0, 1)$ and $\xi_n^{(3)} \xrightarrow{d} \mathcal{N}_J(0, C(\gamma, a_1, \dots, a_J))$ two independent random variables. Straightforward simplifications yield

$$L_{j, n}^{[5(ii)]} + L_{j, n}^{[6]} + a_j L_{j, n}^{[5(ii)]} L_{j, n}^{[6]} = \frac{(1 - a_j \gamma)}{\sqrt{k_n}} \xi_n^{(3)} + o_P(A(n/k_n)).$$

Replacing in (34) yields

$$\frac{\widetilde{\mathcal{M}}_{a_j, n}^{(D)}(k_n/n) - \mathcal{M}_{a_j}(k_n/n)}{1 + a_j \mathcal{M}_{a_j}(k_n/n)} = \left(\frac{\gamma}{\sqrt{k_n}} \xi_n^{(1)} + \frac{(1 - a_j \gamma)}{\sqrt{k_n}} \xi_{j, n}^{(3)} \right) (1 + o_P(1)) + o_P(A(n/k_n)).$$

The result follows from the independence of $\xi_n^{(1)}$ and $\xi_n^{(3)}$. \blacksquare

Proof of Theorem 2. The following decomposition holds for all $j \in \{1, \dots, J\}$:

$$\frac{\widetilde{\mathcal{M}}_{a_j, n}^{(1)}(k_n/n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(k_n/n)}{1 + a_j \mathcal{M}_{a_j}(k_n/n)} = L_{j, n}^{[3]} + \left(L_{j, n}^{[4(i)]} + L_{j, n}^{[5(ii)]} + a_j L_{j, n}^{[4(i)]} L_{j, n}^{[5(ii)]} \right) \left(1 + a_j L_{j, n}^{[3]} \right) \quad (35)$$

where each of the three terms in the previous expansion is studied in a dedicated lemma:

$$\begin{aligned} L_{j, n}^{[3]} &= K_{a_j} \left(\frac{X_{n-k_n, n}}{U(n/k_n)} \right) = \frac{\gamma}{\sqrt{k_n}} \xi_n^{(1)} (1 + o_P(1)) && \text{from Lemma 3,} \\ L_{j, n}^{[4(i)]} &= \frac{\hat{\gamma}_n}{1 - a_j \hat{\gamma}_n} \\ &= \frac{\gamma}{1 - a \gamma} + \frac{\mu}{(1 - a_j \gamma)^2} A(n/k_n) (1 + o_P(1)) + \frac{\sigma}{(1 - a_j \gamma)^2} \frac{\xi_n^{(2)}}{\sqrt{k_n}} (1 + o_P(1)) && \text{from Lemma 4(i),} \\ L_{j, n}^{[5(ii)]} &= \frac{K_{a_j}(U(n/k_n)) - \mathcal{M}_{a_j}(k_n/n)}{1 + a_j \mathcal{M}_{a_j}(k_n/n)} = -\gamma - \frac{1 - a_j \gamma}{1 - a_j \gamma - \rho} A(k_n/n) (1 + o(1)) && \text{from Lemma 5(ii)} \end{aligned}$$

with $\xi_n^{(1)} \xrightarrow{d} \mathcal{N}(0, 1)$ and $\xi_n^{(2)} \xrightarrow{d} \mathcal{N}(0, 1)$. Remarking that

$$\begin{aligned} L_{j,n}^{[4(i)]} + L_{j,n}^{[5(ii)]} + a_j L_{j,n}^{[4(i)]} L_{j,n}^{[5(ii)]} &= \frac{\sigma}{1 - a_j \gamma} \frac{\xi_n^{(2)}}{\sqrt{k_n}} (1 + o_P(1)) \\ &+ \frac{\mu}{1 - a_j \gamma} A(n/k_n) (1 + o_P(1)) - \frac{A(n/k_n)}{1 - a_j \gamma - \rho} (1 + o(1)) \end{aligned}$$

and replacing in (35) yield

$$\begin{aligned} \frac{\widetilde{\mathcal{M}}_{a_j,n}^{(1)}(k_n/n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(k_n/n)}{1 + a_j \mathcal{M}_{a_j}(k_n/n)} &= \frac{\gamma}{\sqrt{k_n}} \xi_n^{(1)} (1 + o_P(1)) + \frac{\sigma}{1 - a_j \gamma} \frac{\xi_n^{(2)}}{\sqrt{k_n}} (1 + o_P(1)) \\ &+ \frac{\mu}{1 - a_j \gamma} A(n/k_n) (1 + o_P(1)) - \frac{A(n/k_n)}{1 - a_j \gamma - \rho} (1 + o(1)). \end{aligned}$$

As a conclusion,

$$\sqrt{k_n} \left(\frac{\widetilde{\mathcal{M}}_{a_j,n}^{(1)}(k_n/n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(k_n/n)}{1 + a_j \mathcal{M}_{a_j}(k_n/n)} \right) = \gamma \xi_n^{(1)} + \frac{\sigma}{1 - a_j \gamma} \xi_n^{(2)} + \lambda \left(\frac{\mu}{1 - a_j \gamma} - \frac{1}{1 - a_j \gamma - \rho} \right) + o_P(1),$$

and the result is thus proved. \blacksquare

Proof of Theorem 3. The following decomposition holds for all $j \in \{1, \dots, J\}$:

$$\begin{aligned} &\sqrt{k_n} \left(\frac{\widetilde{\mathcal{M}}_{a_j,n}^{(1, \text{RB})}(k_n/n; \widehat{\gamma}_n^{(\text{RB})}) - \mathcal{M}_{a_j}(k_n/n)}{1 + a_j \mathcal{M}_{a_j}(k_n/n)} \right) \\ &= T_{j,n} + \sqrt{k_n} \varphi_{a_j}(\widehat{\gamma}_n^{(\text{RB})}, \hat{\rho}_n, \hat{\beta}_n) \left(\frac{n}{k_n} \right)^{\hat{\rho}_n} \left(1 + a_j \frac{T_{j,n}}{\sqrt{k_n}} \right), \end{aligned} \quad (36)$$

with $\varphi_{a_j}(\widehat{\gamma}_n^{(\text{RB})}, \hat{\rho}_n, \hat{\beta}_n) = \hat{\beta}_n \widehat{\gamma}_n^{(\text{RB})} / (1 - a_j \widehat{\gamma}_n^{(\text{RB})} - \hat{\rho}_n)$ and where $T_{j,n}$ is studied in Theorem 2:

$$T_{j,n} = \sqrt{k_n} \left(\frac{\widetilde{\mathcal{M}}_{a_j,n}^{(1)}(k_n/n; \widehat{\gamma}_n^{(\text{RB})}) - \mathcal{M}_{a_j}(k_n/n)}{1 + a_j \mathcal{M}_{a_j}(k_n/n)} \right) = \gamma \xi^{(2)} + \frac{\sigma}{1 - a_j \gamma} \xi^{(3)} - \frac{\lambda}{1 - a_j \gamma - \rho} + o_P(1).$$

Note that, here, $\hat{\gamma}_n = \widehat{\gamma}_n^{(\text{RB})}$ with $\mu = 0$, and that $\xi^{(2)}$ and $\xi^{(3)}$ are two standard Gaussian random variables. Lemma 7 entails

$$\begin{aligned} (36) &= T_{j,n} + \left(\varphi_{a_j}(\gamma, \rho, \beta) \sqrt{k_n} \left(\frac{n}{k_n} \right)^\rho + o_P(1) \right) \left(1 + o_P \left(\frac{1}{\sqrt{k_n}} \right) \right), \\ &= T_{j,n} + \varphi_{a_j}(\gamma, \rho, \beta) \sqrt{k_n} \left(\frac{n}{k_n} \right)^\rho (1 + o_P(1)) + o_P(1), \\ &= T_{j,n} + \frac{\sqrt{k_n} A(n/k_n)}{1 - a_j \gamma - \rho} (1 + o_P(1)) + o_P(1), \\ &= T_{j,n} + \frac{\lambda}{1 - a_j \gamma - \rho} + o_P(1) \\ &= \gamma \xi^{(2)} + \frac{\sigma}{1 - a_j \gamma} \xi^{(3)} + o_P(1), \end{aligned}$$

and the conclusion follows. \blacksquare

Proof of Theorem 4. The following decomposition holds for all $j \in \{1, \dots, J\}$:

$$\frac{\widehat{\mathcal{M}}_{a_j, n}^{(D)}(p_n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(p_n)}{1 + a_j \mathcal{M}_{a_j}(p_n)} = L_{j, n} + L'_{j, n} + a_j L_{j, n} L'_{j, n},$$

with

$$\begin{aligned} L_{j, n} &= L_{j, n}^{[2]} + L_{j, n}^{[3]} + L_{j, n}^{[4(ii)]} + a_j \left(L_{j, n}^{[2]} L_{j, n}^{[3]} + L_{j, n}^{[2]} L_{j, n}^{[4(ii)]} + L_{j, n}^{[3]} L_{j, n}^{[4(ii)]} \right) + a_j^2 L_{j, n}^{[2]} L_{j, n}^{[3]} L_{j, n}^{[4(ii)]}, \\ L'_{j, n} &= L_{j, n}^{[5(ii)]} + L_{j, n}^{[6]} + a_j L_{j, n}^{[5(ii)]} L_{j, n}^{[6]}. \end{aligned}$$

First, each of the three terms appearing in L_n is studied in the associated lemma:

$$\begin{aligned} L_{j, n}^{[2]} &= K_{a_j} \left(\frac{\ell(n/k_n)}{\ell(1/p_n)} \right) = \frac{1}{\rho} A(n/k_n) (1 + o(1)) \xrightarrow{P} 0 && \text{from Lemma 2,} \\ L_{j, n}^{[3]} &= K_{a_j} \left(\frac{X_{n-k_n, n}}{U(n/k_n)} \right) = \frac{\gamma}{\sqrt{k_n}} \xi_n^{(1)} (1 + o_P(1)) \xrightarrow{P} 0 && \text{from Lemma 3,} \\ L_{j, n}^{[4(ii)]} &= K_{a_j} (d_n^{\hat{\gamma}_n - \gamma}) \\ &= \frac{\log(d_n)}{\sqrt{k_n}} \left(\mu \sqrt{k_n} A(n/k_n) + \sigma \xi_n^{(2)} \right) + O_P \left(\frac{(\log d_n)^2}{k_n} \right) \xrightarrow{P} 0 && \text{from Lemma 4(ii),} \end{aligned}$$

with $\xi_n^{(1)} \xrightarrow{d} \mathcal{N}(0, 1)$ and $\xi_n^{(2)} \xrightarrow{d} \mathcal{N}(0, 1)$. Consequently, straightforward simplifications yield

$$L_{j, n} = L_{j, n}^{[2]} (1 + o_P(1)) + L_{j, n}^{[3]} (1 + o_P(1)) + L_{j, n}^{[4(ii)]} (1 + o_P(1)) \xrightarrow{P} 0, \quad (37)$$

and therefore

$$\frac{\widehat{\mathcal{M}}_{a_j, n}^{(D)}(p_n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(p_n)}{1 + a_j \mathcal{M}_{a_j}(p_n)} = L_{j, n} + L'_{j, n} (1 + o_P(1)). \quad (38)$$

Second, both terms appearing in $L'_{j, n}$ are studied in a dedicated lemma:

$$\begin{aligned} L_{j, n}^{[5(ii)]} &= \frac{K_{a_j}(U(1/p_n)) - \mathcal{M}_{a_j}(p_n)}{1 + a_j \mathcal{M}_{a_j}(p_n)} = -\gamma - \frac{1 - a_j \gamma}{1 - a_j \gamma - \rho} A(1/p_n) (1 + o(1)) && \text{from Lemma 5(ii),} \\ L_{j, n}^{[6]} &= \frac{\widehat{\mathcal{M}}_{a_j, n}^{(D)}(k_n/n) - K_{a_j}(X_{n-k_n, n})}{1 + a_j K_{a_j}(X_{n-k_n, n})} \\ &= \frac{1}{\sqrt{k_n}} \xi_{j, n}^{(3)} + \frac{\gamma}{1 - a_j \gamma} \left(1 + \frac{A(n/k_n)}{\gamma(1 - a_j \gamma - \rho)} (1 + o_P(1)) \right), && \text{from Lemma 6,} \end{aligned}$$

with $\xi_n^{(3)} = (\xi_{j, n}^{(3)})_{j=1, \dots, J} \xrightarrow{d} \mathcal{N}_J(0, C(a_1, \dots, a_J, \gamma))$. As a consequence,

$$L'_{j, n} = \frac{(1 - a_j \gamma)}{\sqrt{k_n}} \xi_{j, n}^{(3)} (1 + o_P(1)) + \frac{A(n/k_n)}{1 - a_j \gamma - \rho} (1 + o_P(1)) - \frac{A(1/p_n)}{1 - a_j \gamma - \rho} (1 + o_P(1)), \quad (39)$$

and replacing (37) and (39) in (38) yields

$$\begin{aligned} \frac{\widehat{\mathcal{M}}_{a_j, n}^{(D)}(p_n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(p_n)}{1 + a_j \mathcal{M}_{a_j}(p_n)} &= \left(\frac{\gamma}{\sqrt{k_n}} \xi_n^{(1)} + \frac{\sigma \log d_n}{\sqrt{k_n}} \xi_n^{(2)} + \frac{(1 - a_j \gamma)}{\sqrt{k_n}} \xi_{j, n}^{(3)} \right) (1 + o_P(1)) \\ &+ \left(\mu \log(d_n) + \frac{1}{1 - a_j \gamma - \rho} + \frac{1}{\rho} \right) A(n/k_n) (1 + o_P(1)) \\ &- \frac{1}{1 - a_j \gamma - \rho} A(1/p_n) (1 + o_P(1)). \end{aligned}$$

Remark that, since $|A|$ is regularly-varying with index $\rho < 0$, Potter bounds [8, Theorem 1.5.6] show that there exists $\varepsilon \in (0, -\rho)$ such that

$$\left| \frac{A(1/p_n)}{A(n/k_n)} \right| \leq (1 + \varepsilon) d_n^{\rho + \varepsilon},$$

for n large enough, and thus $A(1/p_n)/A(n/k_n) \rightarrow 0$ as $n \rightarrow \infty$. The above expansion can be simplified as

$$\frac{\widehat{\mathcal{M}}_{a_j, n}^{(D)}(p_n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(p_n)}{1 + a_j \mathcal{M}_{a_j}(p_n)} = \left(\frac{\gamma}{\sqrt{k_n}} \xi_n^{(1)} + \frac{\sigma \log d_n}{\sqrt{k_n}} \xi_n^{(2)} + \frac{(1 - a_j \gamma)}{\sqrt{k_n}} \xi_{j, n}^{(3)} \right) (1 + o_P(1)) \quad (40)$$

$$+ \left(\mu \log(d_n) + \frac{1 - a_j \gamma}{\rho(1 - a_j \gamma - \rho)} \right) A(n/k_n) (1 + o_P(1)), \quad (41)$$

leading to

$$\frac{\sqrt{k_n}}{\log(d_n)} \left(\frac{\widehat{\mathcal{M}}_{a_j, n}^{(D)}(p_n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(p_n)}{1 + a_j \mathcal{M}_{a_j}(p_n)} \right) = \sigma \xi_n^{(2)} + \lambda \left(\mu + \frac{1 - a_j \gamma}{\rho(1 - a_j \gamma - \rho)} \frac{1}{\log(d_n)} \right) + o_P(1).$$

The result is proved. \blacksquare

Proof of Theorem 5. The following decomposition holds for all $j \in \{1, \dots, J\}$:

$$\begin{aligned} & \frac{\sqrt{k_n}}{\log(d_n)} \left(\frac{\widehat{\mathcal{M}}_{a_j, n}^{(D, RB)}(p_n; \hat{\gamma}_n^{(RB)}) - \mathcal{M}_{a_j}(1/p_n)}{1 + a_j \mathcal{M}_{a_j}(1/p_n)} \right) \\ &= T_{j, n} + \frac{\sqrt{k_n}}{\log(d_n)} \varphi_{a_j}(\hat{\gamma}_n^{(RB)}, \hat{\rho}_n, \hat{\beta}_n) \left(\frac{n}{k_n} \right)^{\hat{\rho}_n} \left(1 + a T_{j, n} \frac{\log(d_n)}{\sqrt{k_n}} \right), \end{aligned} \quad (42)$$

with $\varphi_{a_j}(\hat{\gamma}_n^{(RB)}, \hat{\rho}_n, \hat{\beta}_n) = (a_j \hat{\gamma}_n^{(RB)} - 1) \hat{\beta}_n \hat{\gamma}_n^{(RB)} / (\hat{\rho}_n (1 - a_j \hat{\gamma}_n^{(RB)} - \hat{\rho}_n))$ and where $T_{j, n}$ is studied in Theorem 4:

$$T_{j, n} = \frac{\sqrt{k_n}}{\log(d_n)} \left(\frac{\widehat{\mathcal{M}}_{a_j, n}^{(D)}(p_n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(p_n)}{1 + a_j \mathcal{M}_{a_j}(p_n)} \right) = \sigma \xi^{(4)} + \frac{\lambda(1 - a_j \gamma)}{\rho(1 - a_j \gamma - \rho)} \frac{1}{\log(d_n)} + o_P(1).$$

Note that, here, $\hat{\gamma}_n = \hat{\gamma}_n^{(RB)}$ and thus $\mu = 0$, and that $\xi^{(4)}$ is a standard Gaussian random variable. Lemma 7 entails

$$\begin{aligned} (42) &= T_{j, n} + \left(\varphi_{a_j}(\gamma, \rho, \beta) \frac{\sqrt{k_n}}{\log(d_n)} \left(\frac{n}{k_n} \right)^\rho + o_P \left(\frac{1}{\log(d_n)} \right) \right) \left(1 + o_P \left(\frac{\log(d_n)}{\sqrt{k_n}} \right) \right), \\ &= T_{j, n} + \varphi_{a_j}(\gamma, \rho, \beta) \frac{\sqrt{k_n}}{\log(d_n)} \left(\frac{n}{k_n} \right)^\rho (1 + o_P(1)) + o_P \left(\frac{1}{\log(d_n)} \right), \\ &= T_{j, n} + \frac{\sqrt{k_n} A(n/k_n)}{\log(d_n)} \frac{a_j \gamma - 1}{\rho(1 - a_j \gamma - \rho)} (1 + o_P(1)) + o_P(1), \\ &= T_{j, n} + \frac{\lambda}{\log(d_n)} \frac{a_j \gamma - 1}{\rho(1 - a_j \gamma - \rho)} (1 + o_P(1)) + o_P(1), \\ &= \sigma \xi^{(4)} + o_P(1), \end{aligned}$$

and the result is thus proved. \blacksquare

Proof of Theorem 6. The following decomposition holds for all $j \in \{1, \dots, J\}$:

$$\frac{\widehat{\mathcal{M}}_{a_j,n}^{(1)}(p_n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(p_n)}{1 + a_j \mathcal{M}_{a_j}(p_n)} = L_{j,n}^{[4(iii)]} + L_{j,n}'' + a_j L_{j,n}^{[4(iii)]} L_{j,n}'', \quad (43)$$

where $L_{j,n}'' = L_{j,n}^{[4(i)]} + L_{j,n}^{[5(ii)]} + a_j L_{j,n}^{[4(i)]} L_{j,n}^{[5(ii)]}$. First, Lemma 4(iii) yields

$$L_{j,n}^{[4(iii)]} = K_{a_j} \left(\frac{\hat{q}_n(p_n)}{q(p_n)} \right) = \frac{\log(d_n)}{\sqrt{k_n}} (\mu \sqrt{k_n} A(n/k_n) + \sigma \xi_n^{(2)}) + O_P \left(\frac{(\log d_n)^2}{k_n} \right) \xrightarrow{P} 0$$

where $\xi_n^{(2)} \xrightarrow{d} \mathcal{N}(0, 1)$ as $n \rightarrow \infty$. Therefore, (43) can be simplified as

$$\frac{\widehat{\mathcal{M}}_{a_j,n}^{(1)}(p_n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(p_n)}{1 + a_j \mathcal{M}_{a_j}(p_n)} = L_{j,n}^{[4(iii)]} + L_{j,n}'' (1 + o_P(1)),$$

Second, both terms appearing in $L_{j,n}''$ are studied separately thanks to dedicated lemmas:

$$\begin{aligned} L_{j,n}^{[4(i)]} &= \frac{\hat{\gamma}_n}{1 - a_j \hat{\gamma}_n} \\ &= \frac{\gamma}{1 - a_j \gamma} + \frac{\mu}{(1 - a_j \gamma)^2} A(n/k_n) (1 + o_P(1)) + \frac{\sigma}{(1 - a_j \gamma)^2} \frac{\xi_n^{(2)}}{\sqrt{k_n}} (1 + o_P(1)) \quad \text{from Lemma 4(i),} \\ L_{j,n}^{[5(ii)]} &= \frac{K_{a_j}(U(1/p_n)) - \mathcal{M}_{a_j}(p_n)}{1 + a_j \mathcal{M}_{a_j}(p_n)} \\ &= -\gamma - \frac{1 - a_j \gamma}{1 - a_j \gamma - \rho} A(1/p_n) (1 + o(1)). \quad \text{from Lemma 5(ii).} \end{aligned}$$

Combining the two above expansions, it follows

$$L_{j,n}'' = \frac{\sigma}{1 - a_j \gamma} \frac{\xi_n^{(2)}}{\sqrt{k_n}} (1 + o_P(1)) + \frac{\mu}{1 - a_j \gamma} A(n/k_n) (1 + o_P(1)) - \frac{A(1/p_n)}{1 - a_j \gamma - \rho} (1 + o_P(1)).$$

As a consequence, one has

$$\begin{aligned} \frac{\widehat{\mathcal{M}}_{a_j,n}^{(1)}(p_n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(p_n)}{1 + a_j \mathcal{M}_{a_j}(p_n)} &= \frac{\sigma}{1 - a_j \gamma} \frac{\xi_n^{(2)}}{\sqrt{k_n}} (1 + o_P(1)) + \frac{\sigma \log(d_n)}{\sqrt{k_n}} \xi_n^{(2)} (1 + o_P(1)) \\ &\quad + \mu \log(d_n) A(n/k_n) \\ &\quad + \frac{\mu}{1 - a_j \gamma} A(n/k_n) (1 + o_P(1)) - \frac{A(1/p_n)}{1 - a_j \gamma - \rho} (1 + o_P(1)), \end{aligned}$$

and therefore

$$\frac{\sqrt{k_n}}{\log(d_n)} \left(\frac{\widehat{\mathcal{M}}_{a_j,n}^{(1)}(p_n; \hat{\gamma}_n) - \mathcal{M}_{a_j}(p_n)}{1 + a_j \mathcal{M}_{a_j}(p_n)} \right)_{j=1, \dots, J} = \sigma u_J \xi_n^{(2)} + \lambda \mu \left(1 + \frac{1}{1 - a_j \gamma} \frac{1}{\log(d_n)} \right)_{j=1, \dots, J} + o_P(1),$$

since $A(1/p_n)/A(n/k_n) \rightarrow 0$ as $n \rightarrow \infty$. The result is thus proved. \blacksquare

B Appendix: Calculations of Box-Cox transformed CTEs

Here, we provide some details on the calculations of \mathcal{M}_a given in Table 8 for $a > 0$. The Box-Cox transformed CTEs of Pareto and Student's t distributions are straightforwardly derived from (4).

We thus focus on GPD, Burr, Fisher-Snedecor, Fréchet, and Inverse gamma distributions. It is shown below that these five distributions can be encompassed in only two families of distributions. Since all the previous distributions have a density f with respect to the Lebesgue measure, (21) yields for $a > 0$:

$$\mathcal{M}_a(\alpha) = \frac{g_a(\alpha) - g_0(\alpha)}{a\alpha} = \frac{g_a(\alpha)/\alpha - 1}{a}, \text{ with } g_a(\alpha) = \int_{q(\alpha)}^{\infty} t^a f(t) dt. \quad (44)$$

B.1 Inverse generalized Gamma (IGG) distribution

The density of the IGG distribution is defined for all $t > 0$ by

$$f(t) = \frac{\theta}{\Gamma(\zeta)} t^{-\theta\zeta-1} \exp(-t^{-\theta}),$$

with $\zeta, \theta > 0$. As noted in [39], the IGG distribution includes several submodels as particular cases, such as the inverse exponential, inverse Weibull, inverse gamma, inverse chi-squared and Fréchet distributions.

Lemma 8 *The Box-Cox transformed CTE of level $\alpha \in (0, 1)$ associated with the IGG distribution exists for all $a, \zeta, \theta > 0$, such as $\zeta\theta > a$ and is given by*

$$\mathcal{M}_a(\alpha) = \frac{1}{a} \left(\frac{\Gamma_{\ell}(\zeta - a/\theta, q(\alpha)^{-\theta})}{\alpha\Gamma(\zeta)} - 1 \right),$$

where $\Gamma(\cdot)$ and $\Gamma_{\ell}(\cdot, \cdot)$ are respectively the gamma and lower incomplete gamma functions.

Proof. Let $a, \zeta, \theta > 0$, such as $\zeta\theta > a$ and consider the change of variable $t \mapsto x = t^{-\theta}$ in (44):

$$g_a(\alpha) = \frac{1}{\Gamma(\zeta)} \int_0^{q(\alpha)^{-\theta}} x^{\zeta-1-a/\theta} \exp(-x) dx.$$

The result follows. ■

The Box-Cox transformed CTE associated with Inverse gamma and Fréchet distributions given in Table 8 are obtained by plugging respectively $\theta = 1$ and $\zeta = 1$ in the above result.

B.2 Burr-Fisher-Snedecor (BFS) distribution

Up to our knowledge, there is no distribution in the literature encompassing both Burr and Fisher-Snedecor distributions. We thus propose to define such a distribution by considering the function defined for all $t > 0$ by

$$f(t) = c(z_1, z_2, \lambda, \zeta) t^{\zeta-1} (1 + \lambda t^{z_1})^{-z_2}, \text{ with } c(z_1, z_2, \lambda, \zeta) = \frac{z_1 \lambda^{\zeta/z_1}}{B(z_2 - \zeta/z_1, \zeta/z_1)}, \quad (45)$$

for some $z_1, z_2, \lambda > 0$, $0 < \zeta < z_1 z_2$ and where $B(\cdot, \cdot)$ denotes the beta function.

Lemma 9 *The function $t > 0 \mapsto f(t)$ defined in (45) is a density.*

Proof. Letting $z_1, z_2, \lambda > 0$ and $0 < \zeta < z_1 z_2$, we have

$$\int_0^\infty t^{\zeta-1} (1 + \lambda t^{z_1})^{-z_2} dt = \frac{1}{z_1} \int_0^\infty x^{\frac{\zeta}{z_1}-1} (1 + \lambda x)^{-z_2} dx \quad (46)$$

$$\begin{aligned} &= \frac{1}{z_1 \lambda^{\zeta/z_1}} \int_0^1 y^{z_2 - \zeta/z_1 - 1} (1 - y)^{\zeta/z_1 - 1} dy \\ &= \frac{1}{z_1 \lambda^{\zeta/z_1}} B(z_2 - \zeta/z_1, \zeta/z_1) = \frac{1}{c(z_1, z_2, \lambda, \zeta)}, \end{aligned} \quad (47)$$

thanks to the successive changes of variables $t \mapsto x = t^{z_1}$ in (46) and $x \mapsto y = 1/(1 + \lambda x)$ in (47). ■

The Burr distribution with parameters (ζ, θ) corresponds to the particular case $\lambda = 1$, $z_1 = \zeta$ and $z_2 = \theta + 1$ leading to the normalizing constant $c(\zeta, \theta + 1, 1, \zeta) = \theta \zeta$. Quantiles are explicit and are given by $q(\alpha) = (\alpha^{-1/\theta} - 1)^{1/\zeta}$, for all $\alpha \in (0, 1)$. The Fisher-Snedecor distribution with parameters (ν_1, ν_2) is obtained with $z_1 = 1$, $z_2 = (\nu_1 + \nu_2)/2$, $\lambda = \nu_1/\nu_2$ and $\zeta = \nu_1/2$ leading to the normalizing constant $c(1, (\nu_1 + \nu_2)/2, \nu_1/\nu_2, \nu_1/2) = (\nu_1/\nu_2)^{\nu_1/2} / B(\nu_1/2, \nu_2/2)$. Moreover, the GPD with shape parameter $\xi > 0$ is a particular case of Fisher-Snedecor distribution with $\nu_1 = 2$ and $\nu_2 = 2/\xi$ yielding $c(1, 1 + 1/\xi, \xi, 1) = 1$ and the quantiles are explicit: $q(\alpha) = (\alpha^{-\xi} - 1)/\xi$ for all $\alpha \in (0, 1)$. The next result provides two equivalent expressions of $\mathcal{M}_a(\alpha)$, $a > 0$. The first one (48) yields a simple result in the case of the Burr distribution, while the second one (49) has been used in the case of the Fisher-Snedecor distribution, see Table 8.

Lemma 10 *The Box-Cox transformed CTE of level $\alpha \in (0, 1)$ associated with the BFS distribution exists for all $a, z_1, z_2, \lambda, \zeta > 0$ such as $0 < \zeta < z_1 z_2 - a$ and is given by*

$$\mathcal{M}_a(\alpha) = \frac{1}{a} \left(\frac{c(z_1, z_2, \lambda, \zeta) q(\alpha)^{\zeta+a-z_1 z_2}}{\alpha(z_1 z_2 - (\zeta + a)) \lambda^{z_2}} \left({}_2F_1 \left(z_2, z_2 - \frac{\zeta + a}{z_1}; z_2 + 1 - \frac{\zeta + a}{z_1}, -\frac{1}{\lambda q(\alpha)^{z_1}} \right) - 1 \right) \right) \quad (48)$$

$$\begin{aligned} &= \frac{1}{a} \left(\frac{c(z_1, z_2, \lambda, \zeta) q(\alpha)^{\zeta+a-z_1} (1 + \lambda q(\alpha)^{z_1})^{1-z_2}}{\alpha(z_1 z_2 - (\zeta + a)) \lambda} \right. \\ &\quad \left. \times {}_2F_1 \left(1 - \frac{\zeta + a}{z_1}, 1; z_2 + 1 - \frac{\zeta + a}{z_1}, -\frac{1}{\lambda q(\alpha)^{z_1}} \right) - 1 \right), \end{aligned} \quad (49)$$

where ${}_2F_1(\cdot, \cdot; \cdot, \cdot)$ is the hypergeometric function.

Proof. Let $a, z_1, z_2, \lambda > 0$ and $0 < \zeta < z_1 z_2 - a$. From (44), one has

$$g_a(\alpha) = \frac{c(z_1, z_2, \lambda, \zeta)}{z_1} \int_0^{q(\alpha)^{-z_1}} x^{z_2-1-(\zeta+a)/z_1} (x + \lambda)^{-z_2} dx \quad (50)$$

$$= \frac{c(z_1, z_2, \lambda, \zeta) \lambda^{-z_2} q(\alpha)^{\zeta+a-z_1 z_2}}{z_1} \int_0^1 y^{z_2-1-(\zeta+a)/z_1} \left(1 + y \frac{1}{\lambda q(\alpha)^{z_1}} \right)^{-z_2} dy, \quad (51)$$

where we considered the successive changes of variables $t \mapsto x = t^{-z_1}$ in (50), and $x \mapsto y = x q(\alpha)^{z_1}$ in (51). Introducing the Euler type integral representation of the hypergeometric function [1, Equation (15.3.1)] entails

$$g_a(\alpha) = \frac{c(z_1, z_2, \lambda, \zeta) \lambda^{-z_2} q(\alpha)^{\zeta+a-z_1 z_2}}{z_1 (z_2 - (\zeta + a)/z_1)} {}_2F_1 \left(z_2, z_2 - \frac{\zeta + a}{z_1}; z_2 + 1 - \frac{\zeta + a}{z_1}, -\frac{1}{\lambda q(\alpha)^{z_1}} \right),$$

which coincides with (48). Finally, Euler's hypergeometric transformation [1, Equation (15.3.3)] yields (49). ■