WORKING PAPERS

N° 1565

August 2024

“A unified theory of extreme Expected Shortfall inference”

Abdelaati Daouia, Gilles Stupfler & Antoine Usseglio-Carleve

Toulouse School of Economics

A unified theory of extreme Expected Shortfall inference

Abdelaati Daouia[a], Gilles Stupfler[b] & Antoine Usseglio-Carleve[c]

[a] Toulouse School of Economics, University of Toulouse Capitole, France
ORCID: 0000-0003-2621-8860

[b] Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France
ORCID: 0000-0003-2497-9412

[c] Avignon Université, LMA UPR 2151, 84000 Avignon, France
ORCID: 0000-0002-8148-3758

Abstract. The use of the Expected Shortfall as a solution for various deficiencies of quantiles has gained substantial traction over the last 20 years. Its inference at extreme levels is a difficult problem in statistics, with existing approaches typically being limited to heavy-tailed distributions having a finite second tail moment. This constitutes a substantial restriction in areas like finance and environmental science, where the random variable of interest may have a much heavier tail or, at the opposite, may be light-tailed or short-tailed. Under a wider semiparametric extreme value framework, we develop comprehensive asymptotic theory for extreme Expected Shortfall estimation in the general class of distributions with finite first tail moment. By relying on the moment estimators of the scale and shape extreme value parameters, we construct refined asymptotic confidence intervals whose finite-sample coverage is found to be close to the nominal level on simulated data. We illustrate the usefulness of our construction on two sets of financial loss returns and flood insurance claims data.

AMS Subject Classifications: 62G05, 62G20, 62G30, 62G32.

Keywords: Expected Shortfall, extrapolation, inference, extreme value moment estimator, second-order regular variation, stable distribution, weak convergence.

1 Introduction

Tail risk assessment is concerned with the analysis of rare events that carry potential serious impacts on healthcare systems, the environment or the economy, including, but not limited to, geohazards and disaster risk, asset/investment risk, systemic risk, climate risk and cyber risk. The most common univariate risk measure is the quantile $q(\tau)$ of a random risk variable

X, for a suitable $\tau \in (0,1)$; in finance, this measure is generally called Value-at-Risk or VaR. Two important issues with the quantile risk measure lie in its inability to give any idea of the shape of the distribution of X beyond the level $q(\tau)$, and its failure to constitute a coherent risk measure in a financial and actuarial sense, see Artzner et al. (1999) and Acerbi (2002).

A better alternative to $q(\tau)$ in these respects is the Expected Shortfall at level τ, namely

$$\mathrm{ES}(\tau) := \frac{1}{1-\tau}\int_{\tau}^{1} q(t)\mathrm{d}t.$$

When the distribution of X is continuous, this is also the τ-Conditional Value-at-Risk or Conditional Tail Expectation $\mathrm{CTE}(\tau) = \mathbb{E}(X \,|\, X > q(\tau))$ (Rockafellar and Uryasev, 2002). The Expected Shortfall risk measure takes all values of the risk variable X beyond $q(\tau)$ into account. Being a spectral risk measure with positive and nonincreasing risk spectrum, it is coherent and also comonotonically additive, see Theorem 4.47 p.180 and Remark 4.85 p.199 in Föllmer and Schied (2004). It is preferred by practitioners concerned with exposure to catastrophic financial events. It is also favored by major regulators, including the EU, UK, Australia and Canada, which will be requiring the use of ES(97.5%), rather than VaR(99%), in alternative internal models from 1 January 2025. In the EU, this is codified by Article 325ba(1) of Regulation (EU) No 2019/876, itself a revision of the Capital Requirements Regulation (EU) No 575/2013, implementing the Basel Committee on Banking Supervision rules.

The estimation of the Expected Shortfall has so far mostly focused on its empirical counterpart. To simplify the presentation, we consider in this introduction the setting when $n(1-\tau)$ is an integer and F is continuous. Then the empirical estimator of $\mathrm{ES}(\tau)$ is

$$\frac{1}{1-\tau}\int_{\tau}^{1} \widehat{q}_n(t)\mathrm{d}t \quad \text{or equivalently} \quad \frac{1}{n(1-\tau)}\sum_{i=1}^{n(1-\tau)} X_{n-i+1:n},$$

where $\widehat{q}_n(t) = X_{\lceil nt \rceil:n}$ is the empirical quantile function of a sample $(X_1,\ldots,X_n)$ from X, with $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ being the order statistics in ascending order. This is an $L-$statistic, whose asymptotic behavior has been considered by Brazauskas et al. (2008) for fixed τ, and by El Methni et al. (2014, 2018), El Methni and Stupfler (2017) and Goegebeur et al. (2022) at extreme levels $\tau_n \uparrow 1$, under the assumption that $X_+ = \max(X,0)$ has a finite variance. This restricts appreciably the range of potential applications; for example, the analysis of the French commercial losses fire insurance dataset of El Methni and Stupfler (2018) suggests that the underlying distribution there is integrable but has an infinite second moment. Apart from El Methni et al. (2018), who consider a general extreme value setting in nonparametric regression, existing results also assume that the distribution of the risk variable is heavy-tailed; this may constitute a substantial restriction for environmental and climate science applications (Beirlant et al., 2004) but also in finance (Daouia et al., 2024a).

Obtaining a general result about the asymptotic behavior of the empirical Expected Shortfall at extreme levels under minimal restrictions on the tail behavior of X is the original motivation for this article. We work under the second-order version of the general extreme value assumption about the quantile function $q(\cdot)$, which essentially conveys that the right tail of the distribution of X can be approximated by that of a Generalized Pareto distribution. This makes it possible to construct extrapolated estimators of the Expected Shortfall using extreme value estimators of the scale and shape parameters of the approximating Generalized Pareto distribution. Depending on the class of estimators that is used, the extreme value assumption about $q(\cdot)$ may not be sufficient; for instance, if the moment estimators of Dekkers et al. (1989) are employed, then an analogous extreme value assumption about $\log q(\cdot)$ should hold, see pp.103-104 in de Haan and Ferreira (2006). This requires extra technical assumptions about the extreme value parameters of X, whose necessity for practical purposes is unclear.

Lifting these restrictions is the second motivation for this paper, whose contribution is to discuss the estimation of the extreme Expected Shortfall. We first treat the case of an intermediate, in-sample level $\tau_n \uparrow 1$ by examining the asymptotic behavior of two different classes of estimators: the empirical version and a semiparametric estimator based on an asymptotic equivalent of $\mathrm{ES}(\tau) - q(\tau)$ as $\tau \uparrow 1$. In particular, we obtain the convergence of the centered and renormalized empirical Expected Shortfall to a stable distribution when the right tail of X has an infinite second moment. We then get the asymptotic distributions of extrapolated versions of these Expected Shortfall estimators using the moment estimators of Dekkers et al. (1989); a useful tool for that purpose, of independent interest, is a new uniform second-order inequality on the log-quantile function of X that holds under no restrictions whatsoever on the extreme value parameters of X. This inequality is slightly weaker than the classical inequality used in existing theory, but it is enough to obtain asymptotic expansions of the bias incurred when using the moment estimators of Dekkers et al. (1989). We find that the quality of the asymptotic approximations to the finite-sample distributions of the extrapolated Expected Shortfall estimators can be disappointing, which motivates our construction of corrected asymptotic confidence intervals by building upon a fine understanding of the asymptotic behavior of the estimators and the uncertainty introduced by plugging in parameter estimates into asymptotic variances and correlations. Our intervals are found to perform reasonably well on simulations and real data examples.

This article is organized as follows. We introduce our extreme value framework in Section 2. The asymptotic properties of our extreme Expected Shortfall estimators are examined in Section 3. Inference procedures are considered in Section 4. A finite-sample simulation study and two real data analyses are reported in Section 5. Our methods are implemented in the R package `Expectrem`, available at `https://github.com/AntoineUC/Expectrem`.

2 Statistical framework and motivation

Let $F : x \mapsto \mathbb{P}(X \le x)$ denote the distribution function of the random variable of interest X. The associated quantile function is $q : \tau \mapsto \inf\{x \in \mathbb{R} \,|\, F(x) \ge \tau\}$, and the tail quantile function is $U : t \mapsto q(1 - 1/t)$, for $t > 1$. A crucial condition throughout will be the following second-order (extended) regular variation condition on U, which imposes that the right tail of the distribution of X can be approximated by a Generalized Pareto distribution tail at a known rate. This kind of condition cannot be avoided if no precise model structure is assumed on F and a semiparametric procedure to recover the right tail behavior of X is sought: see, for example, Sections 2.3 and 2.4 and Chapter 3 in de Haan and Ferreira (2006), Section 5.6 in Beirlant et al. (2004), and Chapter 2 in Falk et al. (2011) from Section 2.2 onwards.

$\mathcal{C}_2(\gamma, a, \rho, A)$ There are $\gamma \in \mathbb{R}$, a scale function $a(\cdot) > 0$, a second-order parameter $\rho \le 0$ and an auxiliary function $A(\cdot)$ having constant sign and converging to 0 at infinity such that

$$\forall x > 0, \ \lim_{t\to\infty} \frac{1}{A(t)} \left(\frac{U(tx) - U(t)}{a(t)} - \int_1^x s^{\gamma-1} \mathrm{d}s \right) = \int_1^x s^{\gamma-1} \left(\int_1^s u^{\rho-1} \mathrm{d}u \right) \mathrm{d}s.$$

Table 1 gives a non-exhaustive list of distributions satisfying this mild assumption, with corresponding values of γ, $a(\cdot)$, ρ and $A(\cdot)$, found by applying Lemmas A.1–A.3 in the Appendix. It follows from assumption $\mathcal{C}_2(\gamma, a, \rho, A)$ that for any $x > 0$, $U(tx) \approx U(t) + a(t) \int_1^x s^{\gamma-1} \mathrm{d}s$ as $t \to \infty$. This implies that extreme quantiles of X can be recovered from so-called intermediate quantiles (colloquially speaking, "extreme, but not too much") if the scale function $a(\cdot)$ and shape parameter γ of the Generalized Pareto model can be estimated. A popular estimation procedure for these two parameters is the pair of moment estimators developed by Dekkers et al. (1989), which we introduce now. Let $X_1, \ldots, X_n$ have the same distribution as X, and $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ be the associated order statistics. Let also (τ_n) be an intermediate sequence of probabilities, that is, $\tau_n \uparrow 1$ and $n(1 - \tau_n) \to \infty$. Set

$$M_n^{(1)} = \frac{1}{\lfloor n(1-\tau_n) \rfloor} \sum_{i=1}^{\lfloor n(1-\tau_n) \rfloor} \log \frac{X_{n-i+1:n}}{X_{\lceil n\tau_n \rceil:n}} \text{ and } M_n^{(2)} = \frac{1}{\lfloor n(1-\tau_n) \rfloor} \sum_{i=1}^{\lfloor n(1-\tau_n) \rfloor} \log^2 \frac{X_{n-i+1:n}}{X_{\lceil n\tau_n \rceil:n}}.$$

Then the moment estimators of the scale parameter $a(t)$ at $t = (1 - \tau_n)^{-1}$ and of the shape parameter γ are respectively

$$\widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1}) = X_{\lceil n\tau_n \rceil:n} M_n^{(1)} (1 - \widehat{\gamma}_{n,-}^{\mathrm{Mom}})$$

$$\text{and } \widehat{\gamma}_n^{\mathrm{Mom}} = M_n^{(1)} + \widehat{\gamma}_{n,-}^{\mathrm{Mom}}, \text{ with } \widehat{\gamma}_{n,-}^{\mathrm{Mom}} = 1 - \frac{1}{2} \left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right)^{-1}.$$

A semiparametric estimator of an extreme quantile $q(\tau_n')$, where (τ_n') is another sequence of probabilities such that $(1 - \tau_n')/(1 - \tau_n) \to 0$, follows as

$$\widehat{q}_n^{\star}(\tau_n') = X_{\lceil n\tau_n \rceil:n} + \widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1}) \int_1^{\{(1-\tau_n')/(1-\tau_n)\}^{-1}} s^{\widehat{\gamma}_n^{\mathrm{Mom}} - 1} \mathrm{d}s.$$

See Theorem 4.3.1 on p.134 in de Haan and Ferreira (2006) and a discussion of the asymptotic properties of this estimator in Section 4.3.2 on p.140 therein.

When attempting to transpose that construction to the estimation of the Expected Shortfall at extreme levels, it is natural to write first

$$\begin{aligned}\frac{\mathrm{ES}(\tau)-q(\tau)}{a((1-\tau)^{-1})} &= \int_1^\infty \frac{U((1-\tau)^{-1}x)-U((1-\tau)^{-1})}{a((1-\tau)^{-1})}\frac{\mathrm{d}x}{x^2} \\ &\approx \int_1^\infty \left(\int_1^x s^{\gamma-1}\mathrm{d}s\right)\frac{\mathrm{d}x}{x^2} = \frac{1}{1-\gamma} \text{ as } \tau\uparrow 1 \qquad (1)\end{aligned}$$

when $\gamma<1$. Besides, by Theorem 2.3.3 on p.44 in de Haan and Ferreira (2006), the function $a(\cdot)$ is regularly varying with index γ, so that

$$\begin{aligned}&\frac{\mathrm{ES}(\tau_n')-q(\tau_n')}{a((1-\tau_n)^{-1})} \approx \frac{1}{1-\gamma}\times\frac{a((1-\tau_n')^{-1})}{a((1-\tau_n)^{-1})} \approx \frac{1}{1-\gamma}\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma}, \\ &\frac{\mathrm{ES}(\tau_n)-q(\tau_n)}{a((1-\tau_n)^{-1})} \approx \frac{1}{1-\gamma} \text{ and } \frac{q(\tau_n')-q(\tau_n)}{a((1-\tau_n)^{-1})} \approx \int_1^{\{(1-\tau_n')/(1-\tau_n)\}^{-1}} s^{\gamma-1}\mathrm{d}s.\end{aligned}$$

Combining these three approximations provides

$$\frac{\mathrm{ES}(\tau_n')-\mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} \approx \frac{1}{1-\gamma}\int_1^{\{(1-\tau_n')/(1-\tau_n)\}^{-1}} s^{\gamma-1}\mathrm{d}s. \qquad (2)$$

Equations (1) and (2) suggest the following class of extrapolated estimators of $\mathrm{ES}(\tau_n')$:

$$\begin{aligned}&\overline{\mathrm{ES}}_n(\tau_n)+\widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1})\times\frac{1}{1-\widehat{\gamma}_n^{\mathrm{Mom}}}\int_1^{(1-\tau_n)/(1-\tau_n')} s^{\widehat{\gamma}_n^{\mathrm{Mom}}-1}\mathrm{d}s, \\ &\text{with } \overline{\mathrm{ES}}_n(\tau_n)=\widehat{\mathrm{ES}}_n(\tau_n)=\frac{1}{\lfloor n(1-\tau_n)\rfloor}\sum_{i=1}^{\lfloor n(1-\tau_n)\rfloor} X_{n-i+1:n} \qquad (3) \\ &\text{or } \overline{\mathrm{ES}}_n(\tau_n)=\widetilde{\mathrm{ES}}_n(\tau_n)=X_{\lceil n\tau_n\rceil:n}+\widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1})\frac{1}{1-\widehat{\gamma}_n^{\mathrm{Mom}}}. \qquad (4)\end{aligned}$$

There are, in our view, two main difficulties related to the asymptotic behavior of these extrapolated estimators. The first one is that the asymptotic theory about $\widehat{\mathrm{ES}}_n(\tau_n)$ has so far been restricted to the range $\gamma\in(0,1/2)$, which may be a strong limitation depending on the application at hand. The second one is that while U satisfies condition $\mathcal{C}_2(\gamma,a,\rho,A)$, the currently available asymptotic theory for the moment estimators of $a((1-\tau_n)^{-1})$ and γ, and the extrapolated estimators derived from their use, requires the somewhat unnatural assumption $\gamma\neq\rho$ so that a similar condition holds on $\log U(\cdot)$, see for instance the discussions on pp.103 and 140 in de Haan and Ferreira (2006) and Theorem 3.5.4 on p.104 therein. We shall provide solutions to these two difficulties in the next section, by studying the asymptotic distribution of $\widehat{\mathrm{ES}}_n(\tau_n)$ in the general setting $\gamma<1$, and by obtaining directly a slightly weaker, but sufficient, uniform second-order inequality on $\log U(\cdot)$ under no restriction on the pair (γ,ρ).

3 Unified theory of extreme Expected Shortfall estimation

Let $X_1, \ldots, X_n$ be independent and identically distributed copies of X, whose tail quantile function U satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$. The objective is to estimate an Expected Shortfall $\mathrm{ES}(\tau)$ of X above an extreme level $q(\tau)$, where $\tau = \tau_n \to 1$ as $n \to \infty$. As is typical in extreme value statistics, we start by the case of an intermediate level τ_n. In such a case the target quantity can be estimated nonparametrically. Intermediate quantile estimators are then extrapolated to properly extreme levels τ_n', typically satisfying $n(1-\tau_n') = \mathrm{O}(1)$, using a semiparametric extrapolation formula warranted by condition $\mathcal{C}_2(\gamma, a, \rho, A)$ along with estimators of the scale function $a(\cdot)$ and the shape parameter γ. An important tool will be the following asymptotic expansion of $\mathrm{ES}(\tau)$ at high levels τ; this shall especially be useful in controlling the bias of extreme value estimators of the Expected Shortfall. Throughout we let $x_+ = \max(x,0)$ and $x_- = \min(x,0)$ for any $x \in \mathbb{R}$ so that $x = x_+ + x_-$ and $|x| = x_+ - x_-$.

Proposition 1 (Second-order expansion of $\mathrm{ES}(\tau)$)**.** *Assume that condition $\mathcal{C}_2(\gamma, a, \rho, A)$ holds.*

(i) Then we have, for any $\delta, \varepsilon > 0$ sufficiently small, that for t large enough and all $x > 1$,

$$\left| \frac{U(tx) - U(t)}{a(t)} - \int_1^x s^{\gamma-1} \mathrm{d}s - A(t) \int_1^x s^{\gamma-1} \left(\int_1^s u^{\rho-1} \mathrm{d}u \right) \mathrm{d}s \right| \le \varepsilon |A(t)| x^{\gamma_+ + \delta}.$$

(ii) If moreover $\gamma < 1$, we have, as $\tau \uparrow 1$,

$$\begin{aligned}\frac{\mathrm{ES}(\tau) - q(\tau)}{a((1-\tau)^{-1})} &= \frac{\mathrm{CTE}(\tau) - q(\tau)}{a((1-\tau)^{-1})} + \mathrm{o}(|A((1-\tau)^{-1})|) \\ &= \frac{1}{1-\gamma} + A((1-\tau)^{-1}) \left(\frac{1}{(1-\gamma)(1-\gamma-\rho)} + \mathrm{o}(1) \right).\end{aligned}$$

(iii) If moreover $\gamma \in (0,1)$, then, as $\tau \uparrow 1$,

$$\begin{aligned}\frac{\mathrm{ES}(\tau) - q(\tau)}{q(\tau)} &= \frac{\gamma}{1-\gamma} + \left(\frac{a((1-\tau)^{-1})}{q(\tau)} - \gamma \right) \left(\frac{1}{1-\gamma} + \mathrm{o}(1) \right) \\ &\quad + A((1-\tau)^{-1}) \left(\frac{\gamma}{(1-\gamma)(1-\gamma-\rho)} + \mathrm{o}(1) \right).\end{aligned}$$

Proposition 1(i) is a convenient complement to the second-order inequality in Theorem 2.3.6 on p.46 in de Haan and Ferreira (2006) with a weaker upper bound that is sufficient for our purposes. Proposition 1(ii) is an extension of Lemma 3(ii) in El Methni and Stupfler (2017), which is nothing but a weaker version of Proposition 1(iii), to the general extreme value framework we consider.

3.1 At intermediate levels

Let $\tau \in (0,1)$ and let $\widehat{q}_n(t) = X_{\lceil nt \rceil:n}$ denote the empirical quantile function of the sample $(X_1, \ldots, X_n)$. When there are no ties in the sample (this is true with probability 1 if F is continuous) and if moreover $n(1-\tau)$ is an integer, the empirical estimator of $\mathrm{ES}(\tau)$ satisfies

$$\frac{1}{1-\tau}\int_\tau^1 \widehat{q}_n(t)\mathrm{d}t = \frac{1}{\lfloor n(1-\tau)\rfloor}\sum_{i=1}^{\lfloor n(1-\tau)\rfloor} X_{n-i+1:n} = \frac{\sum_{i=1}^n X_i \mathbb{1}_{\{X_i > X_{\lceil n\tau \rceil:n}\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > X_{\lceil n\tau \rceil:n}\}}}.$$

The latter estimator is the obvious empirical counterpart of the Conditional Tail Expectation $\mathrm{CTE}(\tau) = \mathbb{E}(X \,|\, X > q(\tau))$. It turns out that in general, in the case when $\tau = \tau_n \to 1$ and $n(1-\tau_n) \to \infty$, as $n \to \infty$, corresponding to an Expected Shortfall above an extreme but in-sample level, the three estimators in the above equation have the same asymptotic behavior (see Proposition A.1 for a rigorous statement), and we therefore work throughout with the estimator $\widehat{\mathrm{ES}}_n(\tau_n)$ defined in (3). Our first result in this section shows that at such an intermediate level τ_n, the estimator $\widehat{\mathrm{ES}}_n(\tau_n)$ is consistent, and provides its asymptotic behavior for any $\gamma < 1$. This is not a simple corollary of existing results, as Gaussian approximations to the tail empirical quantile process from Theorem 2.4.2 on p.51 in de Haan and Ferreira (2006) are not sufficient in order to handle the case $1/2 \le \gamma < 1$. To the best of our knowledge, no result on $\widehat{\mathrm{ES}}_n(\tau_n)$ has been shown in this challenging case.

Theorem 1 (Weak convergence of the empirical intermediate Expected Shortfall)**.** *Suppose that X satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$ with $\gamma < 1$. Let $\tau_n \uparrow 1$ be such that $n(1-\tau_n) \to \infty$.*

- *Assume $\gamma < 1/2$. Then, if $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$, one has*

$$\sqrt{n(1-\tau_n)}\frac{\widehat{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \frac{2}{(1-\gamma)(1-2\gamma)}\right).$$

- *If $\gamma = 1/2$, we have, provided $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1})/\sqrt{\log(n(1-\tau_n))} = \mathrm{O}(1)$,*

$$\frac{\sqrt{n(1-\tau_n)}}{\sqrt{\log(n(1-\tau_n))}}\frac{\widehat{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} \xrightarrow{\mathrm{d}} \mathcal{N}(0, 4).$$

- *Assume $1/2 < \gamma < 1$. Then, if $(n(1-\tau_n))^{1-\gamma}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$, one has*

$$(n(1-\tau_n))^{1-\gamma}\frac{\widehat{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} \xrightarrow{\mathrm{d}} \frac{1}{\gamma}\left\{-\frac{\Gamma(2-1/\gamma)}{1/\gamma - 1}\cos\left(\frac{\pi}{2\gamma}\right)\right\}^\gamma Z_{1/\gamma},$$

where, for any $\alpha \in (1,2)$, Z_α has a unit right-skewed stable distribution with Fourier transform

$$\mathbb{E}\left(e^{itZ_\alpha}\right) = \exp\left(-|t|^\alpha\left\{1 - i\tan\left(\frac{\pi\alpha}{2}\right)\mathrm{sign}(t)\right\}\right).$$

Remark 1 (Compatibility with existing results in the Fréchet domain of attraction with finite second moment)**.** When $0 < \gamma < 1$, $a(t)/U(t) \to \gamma$ as $t \to \infty$ and $\mathrm{ES}(\tau)/q(\tau) \to 1/(1-\gamma)$ as $\tau \uparrow 1$, see Proposition 1(iii). It follows from Theorem 1 that when $0 < \gamma < 1/2$ and $\sqrt{n(1-\tau_n)} A((1-\tau_n)^{-1}) = \mathrm{O}(1)$,

$$\sqrt{n(1-\tau_n)} \left(\frac{\widehat{\mathrm{ES}}_n(\tau_n)}{\mathrm{ES}(\tau_n)} - 1 \right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \frac{2\gamma^2(1-\gamma)}{1-2\gamma}\right).$$

This is Theorem 2 in El Methni and Stupfler (2017) for the distortion function being the identity function; see also Theorem 1 in El Methni et al. (2014) in the regression setting.

An alternative class of estimators of $\mathrm{ES}(\tau_n)$, motivated by Equation (4), is the quantile-based semiparametric version

$$\widetilde{\mathrm{ES}}_n(\tau_n) = X_{\lceil n\tau_n \rceil : n} + \widehat{a}_n((1-\tau_n)^{-1}) \frac{1}{1-\widehat{\gamma}_n}$$

where $\widehat{a}_n((1-\tau_n)^{-1})$ and $\widehat{\gamma}_n$ are consistent estimators of the scale parameter $a((1-\tau_n)^{-1})$ and the shape parameter γ, respectively. Examples include the semiparametric class of moment estimators (Dekkers et al., 1989), Generalized Pareto maximum likelihood estimators (Smith, 1987; Drees et al., 2004) and probability-weighted moment estimators (Hosking et al., 1985; Diebolt et al., 2007). These estimators typically converge at the rate $1/\sqrt{n(1-\tau_n)}$, see Chapter 3 in de Haan and Ferreira (2006). The next result discusses the asymptotic properties of $\widetilde{\mathrm{ES}}_n(\tau_n)$ depending on those of the random vector $(X_{\lceil n\tau_n \rceil : n}, \widehat{a}_n((1-\tau_n)^{-1}), \widehat{\gamma}_n)$.

Theorem 2 (Weak convergence of $\widetilde{\mathrm{ES}}_n(\tau_n)$ at intermediate levels)**.** *Suppose that X satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$ with $\gamma < 1$. Let $\tau_n \uparrow 1$ be such that $n(1-\tau_n) \to \infty$ and $\sqrt{n(1-\tau_n)} A((1-\tau_n)^{-1}) \to \lambda \in \mathbb{R}$. Assume that*

$$\sqrt{n(1-\tau_n)} \left(\frac{X_{\lceil n\tau_n \rceil : n} - q(\tau_n)}{a((1-\tau_n)^{-1})}, \frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} - 1, \widehat{\gamma}_n - \gamma \right) \xrightarrow{\mathrm{d}} (N_{\mathrm{loc}}, N_{\mathrm{scale}}, N_{\mathrm{shape}})$$

where the random vector $(N_{\mathrm{loc}}, N_{\mathrm{scale}}, N_{\mathrm{shape}})$ has a nondegenerate distribution. Then

$$\sqrt{n(1-\tau_n)} \left(\frac{\widetilde{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})}, \frac{X_{\lceil n\tau_n \rceil : n} - q(\tau_n)}{a((1-\tau_n)^{-1})}, \frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} - 1, \widehat{\gamma}_n - \gamma \right)$$
$$\xrightarrow{\mathrm{d}} \left(N_{\mathrm{loc}} + \frac{1}{1-\gamma} N_{\mathrm{scale}} + \frac{1}{(1-\gamma)^2} N_{\mathrm{shape}} - \frac{\lambda}{(1-\gamma)(1-\gamma-\rho)}, N_{\mathrm{loc}}, N_{\mathrm{scale}}, N_{\mathrm{shape}} \right).$$

3.2 At extreme levels

Let (τ_n') be another sequence of probabilities such that $(1-\tau_n')/(1-\tau_n) \to 0$. Typical examples include $\tau_n' = 1 - 1/n$, with the associated extreme quantile $q(\tau_n') \equiv q(1-1/n)$ having the order of magnitude of the sample maximum $X_{n:n}$ in the data. Equation (2) suggests the following, broad class of extrapolated estimators of $\mathrm{ES}(\tau_n')$:

$$\overline{\mathrm{ES}}_n^\star(\tau_n') = \overline{\mathrm{ES}}_n(\tau_n) + \widehat{a}_n((1-\tau_n)^{-1}) \times \frac{1}{1-\widehat{\gamma}_n}\int_1^{(1-\tau_n)/(1-\tau_n')} s^{\widehat{\gamma}_n - 1}\mathrm{d}s.$$

Here $\overline{\mathrm{ES}}_n(\tau_n)$ is a (relatively) consistent estimator of $\mathrm{ES}(\tau_n)$, such as $\widehat{\mathrm{ES}}_n(\tau_n)$ or $\widetilde{\mathrm{ES}}_n(\tau_n)$, and $\widehat{a}_n((1-\tau_n)^{-1})$ and $\widehat{\gamma}_n$ are consistent estimators of $a((1-\tau_n)^{-1})$ and γ, respectively. The objective of the next result in this section is to give a high-level result about the convergence of $\overline{\mathrm{ES}}_n^\star(\tau_n')$, its corollary when $\overline{\mathrm{ES}}_n(\tau_n) = \widehat{\mathrm{ES}}_n(\tau_n)$, resulting in an estimator $\overline{\mathrm{ES}}_n^\star(\tau_n') = \widehat{\mathrm{ES}}_n^\star(\tau_n')$, and its corollary when $\overline{\mathrm{ES}}_n(\tau_n) = \widetilde{\mathrm{ES}}_n(\tau_n)$, resulting in an estimator $\overline{\mathrm{ES}}_n^\star(\tau_n') = \widetilde{\mathrm{ES}}_n^\star(\tau_n')$.

Theorem 3 (Asymptotic expansion of the extrapolated estimator $\overline{\mathrm{ES}}_n^\star(\tau_n')$)**.** *Suppose that X satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$ with $\gamma < 1$, and $\rho < 0$ if $\gamma \geq 0$. Let $\tau_n, \tau_n' \uparrow 1$ be such that $n(1-\tau_n) \to \infty$ and $1/d_n := (1-\tau_n')/(1-\tau_n) \to 0$. Assume that*

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} - 1\right) = \mathrm{O}_{\mathbb{P}}(1) \quad \textit{and} \quad \sqrt{n(1-\tau_n)}(\widehat{\gamma}_n - \gamma) = \mathrm{O}_{\mathbb{P}}(1).$$

If $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda \in \mathbb{R}$ and $\log(d_n)/\sqrt{n(1-\tau_n)} \to 0$, then

$$\begin{aligned}\frac{\overline{\mathrm{ES}}_n^\star(\tau_n') - \mathrm{ES}(\tau_n')}{a((1-\tau_n)^{-1})} \stackrel{\mathrm{d}}{=} & \frac{\overline{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} - \frac{\gamma_-}{1-\gamma_-}\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s\left(\frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} - 1\right) \\ & + \left(\frac{1}{1-\gamma} - \frac{\gamma_-}{(1-\gamma_-)^2}\right)\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s\,(\widehat{\gamma}_n - \gamma) \\ & - \frac{\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s}{\sqrt{n(1-\tau_n)}}\left(\lambda\frac{\gamma_-(1-2\gamma_- - \rho)}{(1-\gamma_-)(1-\gamma_- - \rho)(\gamma_- + \rho)} + \mathrm{o}_{\mathbb{P}}(1)\right).\end{aligned}$$

An inspection of the proof reveals that the assumption that $\rho < 0$ when $\gamma \geq 0$ is necessary in order to control the bias term incurred in the approximation $a((1-\tau_n')^{-1})/a((1-\tau_n)^{-1}) \approx ((1-\tau_n')/(1-\tau_n))^{-\gamma}$ made when constructing the extrapolated estimator $\overline{\mathrm{ES}}_n^\star(\tau_n')$.

When using $\overline{\mathrm{ES}}_n(\tau_n) = \widehat{\mathrm{ES}}_n(\tau_n)$ for $\gamma < 1/2$, or $\overline{\mathrm{ES}}_n(\tau_n) = \widetilde{\mathrm{ES}}_n(\tau_n)$ when $\gamma < 1$, it follows from Theorems 1 and 2 that $(\overline{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n))/a((1-\tau_n)^{-1}) = \mathrm{O}_{\mathbb{P}}(1/\sqrt{n(1-\tau_n)})$. We may then draw the following useful corollary.

Corollary 1 (Weak convergence of the extrapolated estimator $\overline{\mathrm{ES}}_n^\star(\tau_n')$)**.** *With the notation and under the conditions of Theorem 3, if actually*

$$\sqrt{n(1-\tau_n)}\left(\frac{\overline{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})}, \frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} - 1, \widehat{\gamma}_n - \gamma\right) \stackrel{\mathrm{d}}{\longrightarrow} (Z_{\mathrm{loc}}, Z_{\mathrm{scale}}, Z_{\mathrm{shape}})$$

where the trivariate random vector $(Z_{\mathrm{loc}}, Z_{\mathrm{scale}}, Z_{\mathrm{shape}})$ has a nondegenerate distribution, then

$$\begin{aligned}&\sqrt{n(1-\tau_n)}\frac{\overline{\mathrm{ES}}_n^\star(\tau_n') - \mathrm{ES}(\tau_n')}{a((1-\tau_n)^{-1})\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s} \\ &\stackrel{\mathrm{d}}{\longrightarrow} \gamma_-^2 Z_{\mathrm{loc}} - \frac{\gamma_-}{1-\gamma_-}Z_{\mathrm{scale}} + \left(\frac{1}{1-\gamma} - \frac{\gamma_-}{(1-\gamma_-)^2}\right)Z_{\mathrm{shape}} - \lambda\frac{\gamma_-(1-2\gamma_- - \rho)}{(1-\gamma_-)(1-\gamma_- - \rho)(\gamma_- + \rho)}.\end{aligned}$$

The following corollary will be useful to deal with the specific case $\gamma \geq 0$.

Corollary 2 (Weak convergence of the extrapolated estimator $\overline{\mathrm{ES}}_n^\star(\tau_n')$, light or heavy tails)**.** *With the notation and under the conditions of Theorem 3 with $\gamma \geq 0$, if actually*

$$\sqrt{n(1-\tau_n)}\frac{\overline{\mathrm{ES}}_n(\tau_n)-\mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} = \mathrm{O}_{\mathbb{P}}(1),\ \sqrt{n(1-\tau_n)}\left(\frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})}-1\right) = \mathrm{O}_{\mathbb{P}}(1)$$

and $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_n-\gamma) \stackrel{\mathrm{d}}{\longrightarrow} Z$ where Z has a nondegenerate distribution, then

$$\sqrt{n(1-\tau_n)}\frac{\overline{\mathrm{ES}}_n^\star(\tau_n')-\mathrm{ES}(\tau_n')}{a((1-\tau_n)^{-1})\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s} \stackrel{\mathrm{d}}{\longrightarrow} \frac{1}{1-\gamma}Z.$$

Theorem 3 reveals that the interplay between the estimation of $\mathrm{ES}(\tau_n)$, at the intermediate level, and the estimation of the scale and shape parameters of the approximating Generalized Pareto distribution, is nonetheless potentially more complex than in the estimation of extreme quantiles because, unlike $X_{\lceil n\tau_n\rceil:n}$, $\widehat{\mathrm{ES}}_n(\tau_n)$ will converge to $\mathrm{ES}(\tau_n)$ at a rate slower than $1/\sqrt{n(1-\tau_n)}$ when $\gamma \in [1/2,1)$. We make this fully explicit in the case when $\widehat{a}_n((1-\tau_n)^{-1}) = \widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1})$ and $\widehat{\gamma}_n = \widehat{\gamma}_n^{\mathrm{Mom}}$. A crucial ingredient for that is the following second-order inequality controlling the asymptotic behavior of $\log U$, under condition $\mathcal{C}_2(\gamma,a,\rho,A)$.

Proposition 2 (On regular variation properties of $\log U$)**.** *Assume that $0 < U(\infty) = q(1) \leq \infty$, and that condition $\mathcal{C}_2(\gamma,a,\rho,A)$ holds. Then we have, for any $\delta,\varepsilon > 0$ sufficiently small, that for t large enough and all $x > 1$,*

$$\left|\frac{\log U(tx)-\log U(t)}{a(t)/U(t)} - \int_1^x s^{\gamma_- -1}\mathrm{d}s + \left(\frac{a(t)}{U(t)}-\gamma_+\right)\int_1^x s^{\gamma_- -1}\left(\int_1^s u^{-|\gamma|-1}\mathrm{d}u\right)\mathrm{d}s \right.$$
$$\left. -A(t)x^{-\gamma_+}\int_1^x s^{\gamma-1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s\right| \leq \varepsilon\left(\left|\frac{a(t)}{U(t)}-\gamma_+\right| + |A(t)|\right)x^\delta.$$

Proposition 2 is valid regardless of the values of γ and ρ and features the functions a and A themselves rather than modifications of these functions, unlike the inequality stated at the top of p.104 in de Haan and Ferreira (2006). It is a weaker inequality since it is restricted to the range $x > 1$ and its right-hand side does not feature the multiplicative term $x^{\gamma_-+\rho}$, but it shall be sufficient for our purpose of obtaining the asymptotic distribution of extrapolated Expected Shortfall estimators based on the moment estimators $\widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1})$ and $\widehat{\gamma}_n^{\mathrm{Mom}}$.

Remark 2 (Compatibility with existing results about $\log U$)**.** According to the discussion on p.103 in de Haan and Ferreira (2006), if condition $\mathcal{C}_2(\gamma,a,\rho,A)$ holds with $0 < U(\infty) = q(1) \leq \infty$, $\gamma \neq \rho$ and $\rho < 0$ when $\gamma > 0$, then $\log U$ satisfies a similar second-order condition:

$$\forall x > 0,\ \lim_{t\to\infty}\frac{1}{B(t)}\left(\frac{\log U(tx)-\log U(t)}{a(t)/U(t)} - \int_1^x s^{\gamma_- -1}\mathrm{d}s\right) = \int_1^x s^{\gamma_- -1}\left(\int_1^s u^{\rho'-1}\mathrm{d}u\right)\mathrm{d}s. \quad (5)$$

According to Lemma B.3.16 p.398 in de Haan and Ferreira (2006), the quantities ρ' and B are defined as

$$\rho' = \begin{cases} \rho & \text{if } \ \gamma < \rho \le 0 \ \text{ or } \ 0 < -\rho \le \gamma \ \text{ or } \ (0 < \gamma < -\rho \text{ and } l = 0), \\ -|\gamma| & \text{if } \ \rho < \gamma \le 0 \ \text{ or } \ (0 < \gamma < -\rho \text{ and } l \neq 0), \end{cases}$$

$$\text{and} \ \ B(t) = \begin{cases} A(t) & \text{if } \ \gamma < \rho \le 0, \\ \dfrac{\rho}{\gamma+\rho} A(t) & \text{if } \ 0 < -\rho < \gamma \ \text{ or } \ (0 < \gamma < -\rho \text{ and } l = 0), \\ \gamma_+ - \dfrac{a(t)}{U(t)} & \text{if } \ \rho < \gamma \le 0 \ \text{ or } \ \gamma = -\rho \ \text{ or } \ (0 < \gamma < -\rho \text{ and } l \neq 0), \end{cases}$$

where for $\gamma + \rho < 0$ the quantity l is defined as $l = \lim_{t\to\infty} U(t) - a(t)/\gamma$ (see Theorem 2.1 in Fraga Alves et al. (2007) for the existence of this limit). Theorem 2.1 in Fraga Alves et al. (2007) also shows that actually, when $\gamma \neq \rho$,

$$\lim_{t\to\infty} \frac{1}{A(t)} \left(\frac{a(t)}{U(t)} - \gamma_+ \right) = \begin{cases} 0 & \text{if } \ \gamma < \rho \le 0, \\ \dfrac{\gamma}{\gamma+\rho} & \text{if } \ 0 < -\rho < \gamma \ \text{ or } \ (0 < \gamma < -\rho \text{ and } l = 0), \\ \pm\infty & \text{if } \ \rho < \gamma \le 0 \ \text{ or } \ \gamma = -\rho \ \text{ or } \ (0 < \gamma < -\rho \text{ and } l \neq 0). \end{cases}$$

This means that the second-order condition (5) on $\log U$ is an immediate consequence of Proposition 2 when either $A(t)$ or $a(t)/U(t) - \gamma_+$ dominates the other, that is, in any of the four situations $\gamma < \rho \le 0$, $\rho < \gamma \le 0$, $\gamma = -\rho$ and $(0 < \gamma < -\rho$ and $l \neq 0)$. When this is not the case, that is, when $0 < -\rho < \gamma$ or $(0 < \gamma < -\rho$ and $l = 0)$, then Proposition 2 and the above result from Fraga Alves et al. (2007) yield, as $t \to \infty$,

$$\begin{aligned} &\frac{\log U(tx) - \log U(t)}{a(t)/U(t)} - \int_1^x s^{\gamma_- - 1} \mathrm{d}s \\ &= A(t) \left\{ -\frac{\gamma}{\gamma+\rho} \int_1^x s^{-1} \left(\int_1^s u^{-\gamma-1} \mathrm{d}u \right) \mathrm{d}s + x^{-\gamma} \int_1^x s^{\gamma-1} \left(\int_1^s u^{\rho-1} \mathrm{d}u \right) \mathrm{d}s + \mathrm{o}(1) \right\} \end{aligned}$$

pointwise in x. A straightforward calculation then shows that

$$\frac{\log U(tx) - \log U(t)}{a(t)/U(t)} - \int_1^x s^{\gamma_- - 1} \mathrm{d}s = \frac{\rho}{\gamma+\rho} A(t) \int_1^x s^{-1} \left(\int_1^s u^{\rho-1} \mathrm{d}u \right) \mathrm{d}s + \mathrm{o}(|A(t)|)$$

which, again, is nothing but the second-order condition (5). Proposition 2 therefore extends this second-order condition on $\log U$ by avoiding the restrictions $\gamma \neq \rho$ and $\rho < 0$ if $\gamma > 0$.

We are now ready to provide the asymptotic distribution of the estimator $\widehat{\mathrm{ES}}_n^\star(\tau_n')$. Define

$$B_1(\gamma,\rho) = \frac{1+\gamma+\gamma^2-\rho\gamma^2}{(1-\gamma)(1+\gamma)^2(1-\rho)^2}\mathbb{1}_{\{0\le\gamma<1\}} + \frac{\rho(1-\gamma)}{(1-\gamma-\rho)(1-2\gamma-\rho)(\gamma+\rho)}\mathbb{1}_{\{\gamma<0\}},$$

$$B_2(\gamma) = \frac{\gamma}{(1-\gamma)(1+\gamma)^2}\mathbb{1}_{\{0\le\gamma<1\}} - \frac{\gamma(1-2\gamma-2\gamma^2+5\gamma^3)}{(1-\gamma)^3(1-2\gamma)(1-3\gamma)}\mathbb{1}_{\{\gamma<0\}}$$

and $V(\gamma) = \dfrac{\gamma^2+1}{(1-\gamma)^2}\mathbb{1}_{\{0\le\gamma<1\}} + \dfrac{(1-\gamma)^2(1-3\gamma+4\gamma^2)}{(1-2\gamma)(1-3\gamma)(1-4\gamma)}\mathbb{1}_{\{\gamma<0\}}.$

Theorem 4 (Weak convergence of the moment-based estimator $\widehat{\mathrm{ES}}_n^\star(\tau_n')$)**.** *Suppose that X satisfies condition $\mathcal{C}_2(\gamma,a,\rho,A)$ with $0<U(\infty)=q(1)\le\infty$, $\gamma<1$, and $\rho<0$ if $\gamma\ge 0$. Let $\tau_n,\tau_n'\uparrow 1$ be such that $n(1-\tau_n)\to\infty$ and $n(1-\tau_n')=\mathrm{O}(1)$. Assume also that $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1})\to\lambda\in\mathbb{R}$, $\sqrt{n(1-\tau_n)}(a((1-\tau_n)^{-1})/q(\tau_n)-\gamma_+)\to\mu\in\mathbb{R}$ and $\log((1-\tau_n)/(1-\tau_n'))/\sqrt{n(1-\tau_n)}\to 0$. Take $\widehat{a}_n((1-\tau_n)^{-1})=\widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1})$ and $\widehat{\gamma}_n=\widehat{\gamma}_n^{\mathrm{Mom}}$. Then*

$$\sqrt{n(1-\tau_n)}\frac{\widehat{\mathrm{ES}}_n^\star(\tau_n')-\mathrm{ES}(\tau_n')}{a((1-\tau_n)^{-1})\int_1^{(1-\tau_n)/(1-\tau_n')}s^{\gamma-1}\log(s)\mathrm{d}s}\xrightarrow{\mathrm{d}}\mathcal{N}(\lambda B_1(\gamma,\rho)+\mu B_2(\gamma),V(\gamma)).$$

When $\gamma>0$ and $n(1-\tau_n')=\mathrm{O}(1)$, which covers for instance the standard case when $\tau_n'=1-1/n$ represents a quantile level in the neighborhood of the maximal observation in the sample, the asymptotic behavior of $\widehat{\mathrm{ES}}_n^\star(\tau_n')$ is dominated by that of the extreme value index estimator $\widehat{\gamma}_n$. This may not be the case if this condition on τ_n' is violated when $\gamma\in[1/2,1)$, because then $\widehat{\mathrm{ES}}_n(\tau_n)$ converges to $\mathrm{ES}(\tau_n)$ at a rate slower than $1/\sqrt{n(1-\tau_n)}$.

Our final main result examines the asymptotic distribution of $\widetilde{\mathrm{ES}}_n^\star(\tau_n')$. Define

$$B_3(\gamma) = \frac{\gamma}{(1-\gamma)(1+\gamma)^2}\mathbb{1}_{\{0\le\gamma<1\}} - \frac{\gamma(1-3\gamma^2)}{(1-\gamma)(1-2\gamma)(1-3\gamma)}\mathbb{1}_{\{\gamma<0\}}.$$

Theorem 5 (Weak convergence of the class of estimators $\widetilde{\mathrm{ES}}_n^\star(\tau_n')$)**.** *Suppose that X satisfies condition $\mathcal{C}_2(\gamma,a,\rho,A)$ with $0<U(\infty)=q(1)\le\infty$, $\gamma<1$, and $\rho<0$ if $\gamma\ge0$. Let $\tau_n,\tau_n'\uparrow 1$ be such that $n(1-\tau_n)\to\infty$ and $(1-\tau_n')/(1-\tau_n)\to 0$. Assume also that $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1})\to\lambda\in\mathbb{R}$, $\sqrt{n(1-\tau_n)}(a((1-\tau_n)^{-1})/q(\tau_n)-\gamma_+)\to\mu\in\mathbb{R}$ and $\log((1-\tau_n)/(1-\tau_n'))/\sqrt{n(1-\tau_n)}\to 0$. Take $\widehat{a}_n((1-\tau_n)^{-1})=\widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1})$ and $\widehat{\gamma}_n=\widehat{\gamma}_n^{\mathrm{Mom}}$. Then*

$$\sqrt{n(1-\tau_n)}\frac{\widetilde{\mathrm{ES}}_n^\star(\tau_n')-\mathrm{ES}(\tau_n')}{a((1-\tau_n)^{-1})\int_1^{(1-\tau_n)/(1-\tau_n')}s^{\gamma-1}\log(s)\mathrm{d}s}\xrightarrow{\mathrm{d}}\mathcal{N}(\lambda B_1(\gamma,\rho)+\mu B_3(\gamma),V(\gamma)).$$

These asymptotic results and their proofs are the basis for the construction of accurate and computationally straightforward confidence intervals about $\mathrm{ES}(\tau_n')$, as discussed below.

4 Corrected asymptotic inference

We first highlight various defects of the standard asymptotic confidence intervals for the extreme Expected Shortfall derived from our results, and we build upon our insight to design corrected versions of these intervals whose coverage is close to the nominal level in the general case $\gamma < 1$. Our methods are implemented in the freely available R package `Expectrem`, see Section B.1 for a description of the relevant commands. We rely on the moment estimators $\widehat{a}_n((1-\tau_n)^{-1}) = \widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1})$ of the scale and $\widehat{\gamma}_n = \widehat{\gamma}_n^{\mathrm{Mom}}$ of the shape extreme value parameters, respectively; we shall use the fact that if $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to 0$ and $\sqrt{n(1-\tau_n)}(a((1-\tau_n)^{-1})/q(\tau_n) - \gamma_+) \to 0$, then

$$\sqrt{n(1-\tau_n)}\left(\frac{X_{\lceil n\tau_n\rceil:n} - q(\tau_n)}{a((1-\tau_n)^{-1})}, \frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} - 1, \widehat{\gamma}_n - \gamma\right) \xrightarrow{\mathrm{d}} (N_{\mathrm{loc}}, N_{\mathrm{scale}}, N_{\mathrm{shape}}),$$

where $(N_{\mathrm{loc}}, N_{\mathrm{scale}}, N_{\mathrm{shape}})$ is trivariate Gaussian centered and has covariance matrix

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\gamma) = \begin{pmatrix} 1 & \gamma & 0 \\ \gamma & v_1(\gamma) & c(\gamma) \\ 0 & c(\gamma) & v_2(\gamma) \end{pmatrix},$$

with

$$v_1(\gamma) = (\gamma^2+2)\mathbb{1}_{\{\gamma\geq 0\}} + \frac{2-16\gamma+51\gamma^2-69\gamma^3+50\gamma^4-24\gamma^5}{(1-2\gamma)(1-3\gamma)(1-4\gamma)}\mathbb{1}_{\{\gamma<0\}},$$

$$v_2(\gamma) = (\gamma^2+1)\mathbb{1}_{\{\gamma\geq 0\}} + \frac{(1-\gamma)^2(1-2\gamma)(1-\gamma+6\gamma^2)}{(1-3\gamma)(1-4\gamma)}\mathbb{1}_{\{\gamma<0\}},$$

$$\text{and } c(\gamma) = -(1-\gamma)\mathbb{1}_{\{\gamma\geq 0\}} - \frac{(1-\gamma)^2(1-4\gamma+12\gamma^2)}{(1-3\gamma)(1-4\gamma)}\mathbb{1}_{\{\gamma<0\}}.$$

See the proof of Theorem 5 in the case $\gamma < 0$, and Corollary 4.2.2 on p.133 in de Haan and Ferreira (2006) when $\gamma \geq 0$. Throughout this section we let $Y_1, \ldots, Y_n$ be independent and identically unit Pareto distributed (*i.e.* with distribution function $y \mapsto 1 - 1/y$ for $y \geq 1$) and we recall that $(X_1, \ldots, X_n) \stackrel{\mathrm{d}}{=} (U(Y_1), \ldots, U(Y_n))$.

We consider the case where the probability level $\tau_n' \uparrow 1$ is such that $n(1-\tau_n')$ is bounded, which is the typical scenario in extreme value practice. The case of an intermediate level $\tau_n = 1 - k_n/n$, where $k_n \to \infty$ and $k_n/n \to 0$, is discussed in Section B.2 of the Appendix. We first construct confidence intervals based on $\widehat{\mathrm{ES}}_n^{\star}(\tau_n')$, the extrapolated version of the empirical Expected Shortfall estimator. For a heavy-tailed distribution with finite variance, an extensively studied competitor is the Weissman-type (after Weissman, 1978) estimator

$$\widehat{\mathrm{ES}}_n^{\mathrm{W}}(\tau_n') = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_n^{\mathrm{H}}} \widehat{\mathrm{ES}}_n(\tau_n) = \left(\frac{k_n}{n(1-\tau_n')}\right)^{\widehat{\gamma}_n^{\mathrm{H}}} \widehat{\mathrm{ES}}_n(1-k_n/n).$$

Here $\widehat{\gamma}_n^{\rm H} = M_n^{(1)}$ is the Hill (1975) estimator of γ. For $\gamma \in (0,1/2)$, the asymptotic behavior of this estimator is dominated by that of $\widehat{\gamma}_n^{\rm H}$ (see El Methni and Stupfler, 2017), so that

$$\frac{\sqrt{k_n}}{\log(k_n/(n(1-\tau_n')))} \log \frac{\widehat{\rm ES}_n^{\rm W}(\tau_n')}{{\rm ES}(\tau_n')} \stackrel{\rm d}{\longrightarrow} \mathcal{N}(0,\gamma^2).$$

The associated asymptotic confidence interval is

$$\widehat{\rm I}_1^\star(\alpha) = \left[\widehat{\rm ES}_n^{\rm W}(\tau_n') \exp\left(\pm \widehat{\gamma}_n^{\rm H} \frac{\log(k_n/(n(1-\tau_n')))}{\sqrt{k_n}} z_{1-\alpha/2}\right)\right]$$

where $z_{1-\alpha/2}$ is the quantile of level $1-\alpha/2$ of the standard Gaussian distribution. Lifting the restriction $\gamma \in (0,1/2)$ can be done using the extrapolated semiparametric estimator $\widehat{\rm ES}_n^\star(\tau_n')$ instead of $\widehat{\rm ES}_n^{\rm W}(\tau_n')$, with $\widehat{a}_n(n/k_n) = \widehat{a}_n^{\rm Mom}(n/k_n)$ and $\widehat{\gamma}_n = \widehat{\gamma}_n^{\rm Mom}$. Applying Theorem 4 (with the notation therein) suggests, when $\gamma < 1$, another confidence interval for ${\rm ES}(\tau_n')$:

$$\widehat{\rm I}_2^\star(\alpha) = \left[\widehat{\rm ES}_n^\star(\tau_n') \pm \frac{\widehat{a}_n(n/k_n) \int_1^{k_n/(n(1-\tau_n'))} s^{\widehat{\gamma}_n - 1} \log(s) {\rm d}s}{\sqrt{k_n}} \sqrt{V(\widehat{\gamma}_n)} z_{1-\alpha/2}\right].$$

This confidence interval is theoretically valid for $\gamma < 1$. In the proof of Theorem 4, the limiting distribution of $\widehat{\rm ES}_n^\star(\tau_n')$ is obtained, when $\gamma \in [0,1)$, by neglecting the finite-sample uncertainty in $\widehat{\rm ES}_n(1-k_n/n)$ and $\widehat{a}_n(n/k_n)$. To construct an accurate approximation to the distribution of $\widehat{\rm ES}_n^\star(\tau_n')$, we propose a sampling approach. Let $d_n = k_n/(n(1-\tau_n'))$ and note that up to the additive bias term ${\rm ES}(\tau_n') - {\rm ES}(1-k_n/n) - \frac{a(n/k_n)}{1-\gamma}\int_1^{d_n} s^{\gamma-1}{\rm d}s$ that we neglect,

$$\begin{aligned}\frac{\widehat{\rm ES}_n^\star(\tau_n') - {\rm ES}(\tau_n')}{\widehat{a}_n(n/k_n)} &\approx \frac{\widehat{\rm ES}_n(1-k_n/n) - {\rm ES}(1-k_n/n)}{\widehat{a}_n(n/k_n)} - \frac{\int_1^{d_n} s^{\gamma-1}{\rm d}s}{1-\gamma}\left(\frac{a(n/k_n)}{\widehat{a}_n(n/k_n)} - 1\right) \\ &\quad + \left(\frac{\int_1^{d_n} s^{\gamma-1}{\rm d}s}{(1-\gamma)^2} + \frac{\int_1^{d_n}\log(s) s^{\gamma-1}{\rm d}s}{1-\gamma}\right)(\widehat{\gamma}_n - \gamma).\end{aligned}$$

We obtain an asymptotic representation of each term of the right-hand side in the above approximation. Up to a negligible bias term,

$$\begin{aligned}\frac{\widehat{\rm ES}_n(1-k_n/n) - {\rm ES}(1-k_n/n)}{a(n/k_n)} &\approx \frac{\widehat{\rm ES}_n(1-k_n/n) - X_{n-k_n:n}}{a(n/k_n)} - \frac{1}{1-\gamma} + \frac{X_{n-k_n:n} - U(n/k_n)}{a(n/k_n)} \\ &\stackrel{\rm d}{\approx} \frac{a(Y_{n-k_n:n})}{a(n/k_n)} \times \frac{1}{k_n}\sum_{i=1}^{k_n}\left(\frac{U(Y_{n-i+1:n}) - U(Y_{n-k_n:n})}{a(Y_{n-k_n:n})} - \frac{1}{1-\gamma}\right) \\ &\quad + \frac{1}{1-\gamma}\left(\frac{a(Y_{n-k_n:n})}{a(n/k_n)} - 1\right) + \frac{U(Y_{n-k_n:n}) - U(n/k_n)}{a(n/k_n)}.\end{aligned}$$

Using condition $\mathcal{C}_2(\gamma, a, \rho, A)$, the Rényi representation (see (A.15) and (A.16) in the Appendix) and the fact that a is regularly varying with index γ, we get, with $D_r(x) = \int_1^x s^{r-1}\mathrm{d}s$,

$$\begin{aligned}\frac{\widehat{\mathrm{ES}}_n(1-k_n/n) - \mathrm{ES}(1-k_n/n)}{a(n/k_n)} &\overset{\mathrm{d}}{\approx} \left(\frac{k_n}{n}Y_{n-k_n:n}\right)^{\gamma} \times \frac{1}{k_n}\sum_{i=1}^{k_n}\left(D_\gamma(Y_{n-i+1:n}/Y_{n-k_n:n}) - \frac{1}{1-\gamma}\right) \\ &+ \frac{1}{1-\gamma}\left(\left(\frac{k_n}{n}Y_{n-k_n:n}\right)^{\gamma} - 1\right) + D_\gamma\left(\frac{k_n}{n}Y_{n-k_n:n}\right) \\ &\overset{\mathrm{d}}{=} \left(\frac{k_n}{n}\exp\left(\sum_{i=k_n+1}^{n}\frac{\log(Y_i)}{i}\right)\right)^{\gamma}\left(\frac{1}{k_n}\sum_{i=1}^{k_n}D_\gamma(Y_i) - \frac{1}{1-\gamma}\right) \\ &+ \frac{1}{1-\gamma}D_\gamma\left(\frac{k_n}{n}\exp\left(\sum_{i=k_n+1}^{n}\frac{\log(Y_i)}{i}\right)\right) =: -G_n^{(1)}(Y_1,\dots,Y_n,\gamma).\end{aligned}$$

Besides

$$\frac{\widehat{a}_n(n/k_n)}{a(n/k_n)} = \frac{X_{n-k_n:n} - U(n/k_n)}{a(n/k_n)}M_n^{(1)}(1-\widehat{\gamma}_{n,-}) + \left\{\frac{U(n/k_n)}{a(n/k_n)}M_n^{(1)}\right\}(1-\widehat{\gamma}_{n,-})$$

and as such, it follows from (A.58), (A.60), (A.61) in the proof of Lemma A.8 (see the Appendix) and the convergence $M_n^{(1)}(1-\widehat{\gamma}_{n,-}) \xrightarrow{\mathbb{P}} \gamma_+$ that

$$\begin{aligned}\frac{\widehat{a}_n(n/k_n)}{a(n/k_n)} &\overset{\mathrm{d}}{\approx} \gamma_+ D_\gamma\left(\frac{k_n}{n}\exp\left(\sum_{i=k_n+1}^{n}\frac{\log(Y_i)}{i}\right)\right) \\ &+ \frac{1}{2}\left(\frac{1}{k_n}\sum_{i=1}^{k_n}D_{\gamma_-}(Y_i)\right)\left(1 - \frac{\left(\frac{1}{k_n}\sum_{i=1}^{k_n}D_{\gamma_-}(Y_i)\right)^2}{\frac{1}{k_n}\sum_{i=1}^{k_n}(D_{\gamma_-}(Y_i))^2}\right)^{-1} =: G_n^{(2)}(Y_1,\dots,Y_n,\gamma).\end{aligned}$$

Finally,

$$\begin{aligned}\widehat{\gamma}_n - \gamma &= \frac{a(n/k_n)}{U(n/k_n)}\left\{\frac{U(n/k_n)}{a(n/k_n)}M_n^{(1)}\right\} - \gamma_+ + \left\{1 - \frac{1}{2}\left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1} - \gamma_-\right\} \\ &\approx \gamma_+\left(\left\{\frac{U(n/k_n)}{a(n/k_n)}M_n^{(1)}\right\} - 1\right) + \left\{1 - \frac{1}{2}\left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1} - \gamma_-\right\}\end{aligned}$$

so using again the Rényi representation and Equations (A.60) and (A.61) in the proof of Lemma A.8 (see the Appendix) leads to

$$\begin{aligned}&\left(\frac{\widehat{\mathrm{ES}}_n(1-k_n/n) - \mathrm{ES}(1-k_n/n)}{\widehat{a}_n(n/k_n)}, \frac{a(n/k_n)}{\widehat{a}_n(n/k_n)} - 1, \widehat{\gamma}_n - \gamma\right) \\ &\overset{\mathrm{d}}{\approx} \left(-G_n(Y_1,\dots,Y_n,\gamma), \frac{1}{G_n^{(2)}(Y_1,\dots,Y_n,\gamma)} - 1, H_n(Y_1,\dots,Y_n,\gamma)\right)\end{aligned}$$

with

$$H_n(Y_1,\dots,Y_n,\gamma) := \gamma_+\left(\frac{1}{k_n}\sum_{i=1}^{k_n} D_{\gamma_-}(Y_i)-1\right)+1-\frac{1}{2}\left(1-\frac{\left(\frac{1}{k_n}\sum_{i=1}^{k_n} D_{\gamma_-}(Y_i)\right)^2}{\frac{1}{k_n}\sum_{i=1}^{k_n} D_{\gamma_-}(Y_i)^2}\right)^{-1}-\gamma_-.$$

Conclude that

$$\frac{\widehat{\mathrm{ES}}_n^\star(\tau_n')-\mathrm{ES}(\tau_n')}{\widehat{a}_n(n/k_n)} \stackrel{\mathrm{d}}{\approx} -G_n(Y_1,\dots,Y_n,\gamma)-\frac{\int_1^{d_n} s^{\gamma-1}\mathrm{d}s}{1-\gamma}\left(\frac{1}{G_n^{(2)}(Y_1,\dots,Y_n,\gamma)}-1\right)$$
$$+\left(\frac{\int_1^{d_n} s^{\gamma-1}\mathrm{d}s}{(1-\gamma)^2}+\frac{\int_1^{d_n}\log(s)s^{\gamma-1}\mathrm{d}s}{1-\gamma}\right)H_n(Y_1,\dots,Y_n,\gamma) =: -E_n(Y_1,\dots,Y_n,\gamma).$$

The random variable $E_n(Y_1,\dots,Y_n,\gamma)$ is very easy to simulate and its distribution can thus be tabulated; this gives rise to the confidence interval

$$\widehat{\mathrm{I}}_3^\star(\alpha)=\left[\widehat{\mathrm{ES}}_n^\star(\tau_n')+\widehat{a}_n(n/k_n)e_{n,\alpha/2}(\widehat{\gamma}_n),\ \widehat{\mathrm{ES}}_n^\star(\tau_n')+\widehat{a}_n(n/k_n)e_{n,1-\alpha/2}(\widehat{\gamma}_n)\right]$$

where $e_{n,\tau}(\gamma)$ is the τth quantile of $E_n(Y_1,\dots,Y_n,\gamma)$. Our experience is that the performance of this interval is adversely affected by ignoring the finite-sample uncertainty in the plug-in step of replacing γ by $\widehat{\gamma}_n$ when the tail of X is very heavy. Since $\sqrt{k_n}(\widehat{\gamma}_n-\gamma)\stackrel{\mathrm{d}}{\approx}\mathcal{N}(0,v_2(\gamma))$, we propose to deal with this issue of uncertainty quantification by computing directly the quantiles of $E_n(Y_1,\dots,Y_n,\widetilde{\gamma}_n)$ for $\widetilde{\gamma}_n=\widehat{\gamma}_n+Z\sqrt{v_2(\widehat{\gamma}_n)/k_n}$, where $Z\sim\mathcal{N}(0,1)$ is independent from the data; in addition, we retain only those values of $\widetilde{\gamma}_n$ that are smaller than 1, *i.e.* we resample given $\widetilde{\gamma}_n<1$. This gives rise to an alternative interval $\widehat{\mathrm{I}}_4^\star(\alpha)$. Algorithm 1 gives all the details concerning the actual computation of both $\widehat{\mathrm{I}}_3^\star(\alpha)$ and $\widehat{\mathrm{I}}_4^\star(\alpha)$ (here and throughout the function Φ denotes the standard normal distribution function).

We turn to extrapolated versions of the semiparametric Expected Shortfall estimator. Again, in the heavy-tailed setting $\gamma\in(0,1)$, one may consider the Weissman-type estimator

$$\widetilde{\mathrm{ES}}_n^{\mathrm{W}}(\tau_n')=\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_n^{\mathrm{H}}}\widetilde{\mathrm{ES}}_n^{\mathrm{H}}(\tau_n)=\left(\frac{k_n}{n(1-\tau_n')}\right)^{\widehat{\gamma}_n^{\mathrm{H}}}\frac{X_{n-k_n:n}}{1-\widehat{\gamma}_n^{\mathrm{H}}}$$

built upon the intermediate estimator $\widetilde{\mathrm{ES}}_n^{\mathrm{H}}(\tau_n)=X_{n-k_n:n}/(1-\widehat{\gamma}_n^{\mathrm{H}})$ motivated by the approximation $\mathrm{ES}(\tau)\approx q(\tau)/(1-\tau)$ as $\tau\uparrow 1$, see Proposition 1(iii). It is also readily seen that the asymptotic behavior of $\widehat{\gamma}_n^{\mathrm{H}}$ dominates, leading to a confidence interval analogous to $\widehat{\mathrm{I}}_1^\star(\alpha)$:

$$\widetilde{\mathrm{I}}_1^\star(\alpha)=\left[\widetilde{\mathrm{ES}}_n^{\mathrm{W}}(\tau_n')\exp\left(\pm\widehat{\gamma}_n^{\mathrm{H}}\frac{\log(k_n/(n(1-\tau_n')))}{\sqrt{k_n}}z_{1-\alpha/2}\right)\right].$$

Just as in the construction of $\widehat{\mathrm{I}}_2^\star(\alpha)$ as opposed to $\widehat{\mathrm{I}}_1^\star(\alpha)$, one may construct a more widely applicable confidence interval $\widetilde{\mathrm{I}}_2^\star(\alpha)$ by using the estimator $\widetilde{\mathrm{ES}}_n^\star(\tau_n')$ instead of $\widetilde{\mathrm{ES}}_n^{\mathrm{W}}(\tau_n')$. Theorem 5 suggests the interval

$$\widetilde{\mathrm{I}}_2^\star(\alpha)=\left[\widetilde{\mathrm{ES}}_n^\star(\tau_n')\pm\frac{\widehat{a}_n(n/k_n)\int_1^{k_n/(n(1-\tau_n'))}s^{\widehat{\gamma}_n-1}\log(s)\mathrm{d}s}{\sqrt{k_n}}\sqrt{V(\widehat{\gamma}_n)}z_{1-\alpha/2}\right],$$

Algorithm 1 Confidence intervals for $\mathrm{ES}(\tau_n')$ - Extrapolated empirical estimator

Require: $N \geq 1$, $\alpha \in (0,1)$, $\widehat{\mathrm{ES}}_n(1-k_n/n)$, $\widehat{a}_n(n/k_n) = \widehat{a}_n^{\mathrm{Mom}}(n/k_n)$, $\widehat{\gamma}_n = \widehat{\gamma}_n^{\mathrm{Mom}}(k_n)$

Ensure: $\widehat{\gamma}_n < 1$ and $1-\tau_n' < k_n/n$

Compute $\widehat{\mathrm{ES}}_n^\star(\tau_n') = \widehat{\mathrm{ES}}_n(1-k_n/n) + \widehat{a}_n(n/k_n) \times \frac{1}{1-\widehat{\gamma}_n}\int_1^{k_n/(n(1-\tau_n'))} s^{\widehat{\gamma}_n - 1}\mathrm{d}s$

Simulate N replications $U_1, \ldots, U_N$ from a uniform distribution on $[0,1]$

for $i \in \{1, \ldots, N\}$ **do**

Calculate $\widetilde{\gamma}_{n,i} = \widehat{\gamma}_n + \sqrt{\frac{v_2(\widehat{\gamma}_n)}{k_n}}\Phi^{-1}\left(U_i \Phi\left((1-\widehat{\gamma}_n)\sqrt{\frac{k_n}{v_2(\widehat{\gamma}_n)}}\right)\right)$

Simulate n replications $Y_1, \ldots, Y_n$ from a unit Pareto distribution

Compute $E_i = E_n(Y_1, \ldots, Y_n, \widehat{\gamma}_n)$ and $\widetilde{E}_i = E_n(Y_1, \ldots, Y_n, \widetilde{\gamma}_{n,i})$

end for

Compute $\begin{cases} E_{\mathrm{up}} = E_{\mathrm{up}}(\alpha) = E_{\lfloor N(1-\alpha/2)\rfloor:N} \\ E_{\mathrm{down}} = E_{\mathrm{down}}(\alpha) = E_{\lfloor N\alpha/2\rfloor:N} \end{cases}$ and $\begin{cases} \widetilde{E}_{\mathrm{up}} = \widetilde{E}_{\mathrm{up}}(\alpha) = \widetilde{E}_{\lfloor N(1-\alpha/2)\rfloor:N} \\ \widetilde{E}_{\mathrm{down}} = \widetilde{E}_{\mathrm{down}}(\alpha) = \widetilde{E}_{\lfloor N\alpha/2\rfloor:N} \end{cases}$

return $\begin{cases} \widehat{\mathrm{I}}_3^\star(\alpha) = \left[\widehat{\mathrm{ES}}_n^\star(\tau_n') + \widehat{a}_n(n/k_n)E_{\mathrm{down}}, \widehat{\mathrm{ES}}_n^\star(\tau_n') + \widehat{a}_n(n/k_n)E_{\mathrm{up}}\right] \\ \widehat{\mathrm{I}}_4^\star(\alpha) = \left[\widehat{\mathrm{ES}}_n^\star(\tau_n') + \widehat{a}_n(n/k_n)\widetilde{E}_{\mathrm{down}}, \widehat{\mathrm{ES}}_n^\star(\tau_n') + \widehat{a}_n(n/k_n)\widetilde{E}_{\mathrm{up}}\right] \end{cases}$

---

which closely resembles $\widehat{\mathrm{I}}_2^\star(\alpha)$, the only difference being that $\widehat{\mathrm{ES}}_n^\star(\tau_n')$ is replaced by $\widetilde{\mathrm{ES}}_n^\star(\tau_n')$. Like $\widehat{\mathrm{I}}_2^\star(\alpha)$, this confidence interval is theoretically valid for $\gamma < 1$, but its finite-sample performance suffers when the underlying distribution is heavy-tailed because it neglects the statistical uncertainty in $\widetilde{\mathrm{ES}}_n(1-k_n/n)$ and $\widehat{a}_n(n/k_n)$. We devise a workaround based on viewing $\widetilde{\mathrm{ES}}_n^\star(\tau_n')$ as a quadratic form of a Gaussian random vector. Set $J_{n,1}(\gamma) = \int_1^{k_n/(n(1-\tau_n'))} s^{\gamma-1}\mathrm{d}s$ and $J_{n,2}(\gamma) = \int_1^{k_n/(n(1-\tau_n'))} s^{\gamma-1}\log(s)\mathrm{d}s$, and set

$$\widetilde{Z}_n^\star = \sqrt{k_n}\frac{\widetilde{\mathrm{ES}}_n^\star(\tau_n') - \mathrm{ES}(\tau_n')}{\widehat{a}_n(n/k_n)} = \sqrt{k_n}\frac{\widetilde{\mathrm{ES}}_n^\star(\tau_n') - \mathrm{ES}(\tau_n')}{a(n/k_n)}\frac{a(n/k_n)}{\widehat{a}_n(n/k_n)}.$$

Since

$$\begin{aligned}\frac{\widetilde{\mathrm{ES}}_n^\star(\tau_n') - \mathrm{ES}(\tau_n')}{a(n/k_n)} &\approx \frac{\widetilde{\mathrm{ES}}_n(1-k_n/n) - \mathrm{ES}(1-k_n/n)}{a(n/k_n)} \\ &\quad + \frac{\int_1^{d_n} s^{\gamma-1}\mathrm{d}s}{1-\gamma}\left(\frac{\widehat{a}_n(n/k_n)}{a(n/k_n)} - 1\right) + \left(\frac{\int_1^{d_n} s^{\gamma-1}\mathrm{d}s}{(1-\gamma)^2} + \frac{\int_1^{d_n}\log(s)s^{\gamma-1}\mathrm{d}s}{1-\gamma}\right)(\widehat{\gamma}_n - \gamma),\end{aligned}$$

recalling Theorem 2 suggests the approximation

$$\begin{aligned}\widetilde{Z}_n^\star &\overset{\mathrm{d}}{\approx} \left(N_{\mathrm{loc}} + \frac{1+J_{n,1}(\gamma)}{1-\gamma}N_{\mathrm{scale}} + \frac{1+J_{n,1}(\gamma)+(1-\gamma)J_{n,2}(\gamma)}{(1-\gamma)^2}N_{\mathrm{shape}}\right)\left(1 - \frac{N_{\mathrm{scale}}}{\sqrt{k_n}}\right) \\ &= \mathbf{u}^\star(\gamma)^\top\mathbf{N} + \frac{\mathbf{N}^\top\mathbf{S}^\star(\gamma)\mathbf{N}}{\sqrt{k_n}},\end{aligned}$$

where $\mathbf{N} = (N_{\text{loc}}, N_{\text{scale}}, N_{\text{shape}})^\top$ follows a trivariate centered normal distribution with covariance matrix $\mathbf{\Sigma}$,

$$\mathbf{u}^\star(\gamma) = \begin{pmatrix} 1 \\ \frac{1+J_{n,1}(\gamma)}{1-\gamma} \\ \frac{1+J_{n,1}(\gamma)+(1-\gamma)J_{n,2}(\gamma)}{(1-\gamma)^2} \end{pmatrix}$$

$$\text{and} \quad \mathbf{S}^\star(\gamma) = \begin{pmatrix} 0 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & -\frac{1+J_{n,1}(\gamma)}{1-\gamma} & -\frac{1+J_{n,1}(\gamma)+(1-\gamma)J_{n,2}(\gamma)}{2(1-\gamma)^2} \\ 0 & -\frac{1+J_{n,1}(\gamma)+(1-\gamma)J_{n,2}(\gamma)}{2(1-\gamma)^2} & 0 \end{pmatrix}.$$

Straightforward calculations show that, if $\mathbf{Z} = (Z_1, \ldots, Z_p)$ is a $p-$dimensional Gaussian random vector made of independent centered unit Gaussian random variables and $\mathbf{M}$ is a $p \times p$ symmetric matrix, one has $\mathbb{E}(\mathbf{Z}^\top \mathbf{M}\mathbf{Z}) = \operatorname{tr}(\mathbf{M})$ and $\mathbb{E}((\mathbf{Z}^\top \mathbf{M}\mathbf{Z})^2) = 2\operatorname{tr}(\mathbf{M}^2) + (\operatorname{tr}(\mathbf{M}))^2$. The mean $m^\star(\gamma)$ and standard deviation $s^\star(\gamma)$ of $\mathbf{u}^\star(\gamma)^\top \mathbf{N} + \mathbf{N}^\top \mathbf{S}^\star(\gamma)\mathbf{N}/\sqrt{k_n}$ are then

$$m^\star(\gamma) = \frac{\operatorname{tr}(\mathbf{S}^\star(\gamma)\mathbf{\Sigma}(\gamma))}{\sqrt{k_n}} \quad \text{and} \quad s^\star(\gamma) = \sqrt{\mathbf{u}^\star(\gamma)^\top \mathbf{\Sigma}(\gamma)\mathbf{u}^\star(\gamma) + 2\frac{\operatorname{tr}(\mathbf{S}^\star(\gamma)\mathbf{\Sigma}(\gamma)\mathbf{S}^\star(\gamma)\mathbf{\Sigma}(\gamma))}{k_n}}.$$

Approximating $\widetilde{Z}_n^\star$ by a Gaussian random variable with mean $m^\star(\widehat{\gamma}_n)$ and standard deviation $s^\star(\widehat{\gamma}_n)$ suggests the confidence interval

$$\widetilde{\mathrm{I}}_3^\star(\alpha) = \left[\widetilde{\mathrm{ES}}_n^\star(\tau_n') - \frac{\widehat{a}_n(n/k_n)}{\sqrt{k_n}} m^\star(\widehat{\gamma}_n) \pm \frac{\widehat{a}_n(n/k_n)}{\sqrt{k_n}} s^\star(\widehat{\gamma}_n) z_{1-\alpha/2}\right].$$

Similarly to $\widehat{\mathrm{I}}_3^\star(\alpha)$, the finite-sample performance of this interval is compromised because the statistical uncertainty of the estimator $\widehat{\gamma}_n$ plugged into $m^\star$ and $s^\star$ is not accounted for. To take this uncertainty into account, we analytically derive a correction term that should be added due to this plug-in step. Let $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}^\top$ be the Cholesky decomposition of $\mathbf{\Sigma}$, where

$$\mathbf{\Lambda} = \mathbf{\Lambda}(\gamma) = \begin{pmatrix} 1 & 0 & 0 \\ \gamma & \sqrt{v_1(\gamma) - \gamma^2} & 0 \\ 0 & \frac{c(\gamma)}{\sqrt{v_1(\gamma)-\gamma^2}} & \sqrt{v_2(\gamma) - \frac{c^2(\gamma)}{v_1(\gamma)-\gamma^2}} \end{pmatrix}$$

and note that $\mathbf{N} = \mathbf{\Lambda}\mathbf{Z}$ where $\mathbf{Z}$ is made of three independent centered unit Gaussian random variables. Recalling that $\sqrt{k_n}(\widehat{\gamma}_n - \gamma) \approx N_{\text{shape}}$, a Taylor expansion yields

$$\begin{aligned} \widetilde{Z}_n^\star &\overset{\mathrm{d}}{=} \mathbf{u}^\star(\gamma)^\top \mathbf{\Lambda}(\gamma)\mathbf{Z} + \frac{\mathbf{Z}^\top \mathbf{\Lambda}(\gamma)^\top \mathbf{S}^\star(\gamma)\mathbf{\Lambda}(\gamma)\mathbf{Z}}{\sqrt{k_n}} + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right) \\ &\overset{\mathrm{d}}{=} \mathbf{u}^\star(\widehat{\gamma}_n)^\top \mathbf{\Lambda}(\widehat{\gamma}_n)\mathbf{Z} + \frac{\mathbf{Z}^\top \mathbf{\Lambda}(\widehat{\gamma}_n)^\top \mathbf{S}^\star(\widehat{\gamma}_n)\mathbf{\Lambda}(\widehat{\gamma}_n)\mathbf{Z}}{\sqrt{k_n}} - \boldsymbol{\theta}^\star(\widehat{\gamma}_n)^\top \frac{N_{\text{shape}}}{\sqrt{k_n}}\mathbf{Z} + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right) \end{aligned}$$

where $\boldsymbol{\theta}^\star(\gamma) = \frac{\mathrm{d}\mathbf{u}^\star}{\mathrm{d}\gamma}(\gamma)^\top \mathbf{\Lambda}(\gamma) + \mathbf{u}^\star(\gamma)^\top \frac{\mathrm{d}\mathbf{\Lambda}}{\mathrm{d}\gamma}(\gamma)$.

Since $\mathbf{N} = \mathbf{\Lambda Z}$, one has $N_{\text{shape}} = \Lambda_{23}(\gamma)Z_2 + \Lambda_{33}(\gamma)Z_3 \overset{\text{d}}{=} \Lambda_{23}(\widehat{\gamma}_n)Z_2 + \Lambda_{33}(\widehat{\gamma}_n)Z_3 + \text{o}_{\mathbb{P}}(1)$. Then

$$\widetilde{Z}_n^\star \overset{\text{d}}{=} \mathbf{w}^\star(\widehat{\gamma}_n)^\top \mathbf{Z} + \frac{\mathbf{Z}^\top \mathbf{W}^\star(\widehat{\gamma}_n)\mathbf{Z}}{\sqrt{k_n}} + \text{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right),$$

where $\mathbf{w}^\star(\gamma) = \mathbf{\Lambda}^\top(\gamma)\mathbf{u}^\star(\gamma)$, and

$$\begin{aligned}\mathbf{W}^\star(\gamma) &= \mathbf{\Lambda}(\gamma)^\top \mathbf{S}^\star(\gamma)\mathbf{\Lambda}(\gamma)\\ &\quad - \frac{1}{2}\begin{pmatrix} 0 & \theta_1^\star(\gamma)\Lambda_{23}(\gamma) & \theta_1^\star(\gamma)\Lambda_{33}(\gamma) \\ \theta_1^\star(\gamma)\Lambda_{23}(\gamma) & 2\theta_2^\star(\gamma)\Lambda_{23}(\gamma) & \theta_2^\star(\gamma)\Lambda_{33}(\gamma) + \theta_3^\star(\gamma)\Lambda_{23}(\gamma) \\ \theta_1^\star(\gamma)\Lambda_{33}(\gamma) & \theta_2^\star(\gamma)\Lambda_{33}(\gamma) + \theta_3^\star(\gamma)\Lambda_{23}(\gamma) & 2\theta_3^\star(\gamma)\Lambda_{33}(\gamma) \end{pmatrix}.\end{aligned}$$

We omit the explicit expressions of $\mathbf{w}^\star(\gamma)$ and $\mathbf{W}^\star(\gamma)$ for the sake of brevity. As in the construction of $\widetilde{\text{I}}_3^\star(\alpha)$, we then approximate the distribution of the random variable $\mathbf{w}^\star(\gamma_0)^\top \mathbf{Z} + \mathbf{Z}^\top \mathbf{W}^\star(\gamma_0)\mathbf{Z}/\sqrt{k_n}$ (for any fixed $\gamma_0 < 1$) by a Gaussian distribution with mean $\text{tr}(\mathbf{W}^\star(\gamma_0))$ and variance $\|\mathbf{w}^\star(\gamma_0)\|_2^2 + 2\,\text{tr}((\mathbf{W}^\star(\gamma_0))^2)/k_n$. This suggests our final confidence interval

$$\widetilde{\text{I}}_4^\star(\alpha) = \left[\widetilde{\text{ES}}_n^\star(\tau_n') + \frac{\widehat{a}_n(n/k_n)}{\sqrt{k_n}}\left(-\,\text{tr}(\mathbf{W}^\star(\widehat{\gamma}_n)) \pm \sqrt{\|\mathbf{w}^\star(\widehat{\gamma}_n)\|_2^2 + 2\frac{\text{tr}((\mathbf{W}^\star(\widehat{\gamma}_n))^2)}{k_n}}\, z_{1-\alpha/2}\right)\right].$$

We next examine and compare the finite-sample performance of these eight intervals.

5 Numerical experiments

5.1 Finite-sample simulation study

We consider the Kumaraswamy, Reverse-Burr, Gumbel, Exponential, Pareto and Fréchet distributions, whose distribution functions and pertaining extreme value quantities are provided in Table 1. In each setting, we simulate $N = 10{,}000$ replications of an i.i.d. sample of size $n = 1{,}000$ from the chosen distribution. We estimate and infer the quantity $\text{ES}(\tau_n') = \text{ES}(1 - 1/n) = \text{ES}(0.999)$ using the estimators $\widehat{\text{ES}}_n^\star(\tau_n')$ and $\widetilde{\text{ES}}_n^\star(\tau_n')$, and the confidence intervals $\widehat{\text{I}}_1^\star(0.95)$, $\widehat{\text{I}}_2^\star(0.95)$, $\widehat{\text{I}}_3^\star(0.95)$, $\widehat{\text{I}}_4^\star(0.95)$, $\widetilde{\text{I}}_1^\star(0.95)$, $\widetilde{\text{I}}_2^\star(0.95)$, $\widetilde{\text{I}}_3^\star(0.95)$ and $\widetilde{\text{I}}_4^\star(0.95)$, at the nominal confidence level 0.95, and with base intermediate level $\tau_n = 1 - k_n/n$ with $k_n = 200$. The true values of ES(0.999), computed using Table C.1, and the empirical coverage probabilities of the competing intervals are provided in Table 2. We carried out a similar inferential exercise for the intermediate value $\text{ES}(\tau_n) = \text{ES}(1 - k_n/n) = \text{ES}(0.8)$; since this is arguably less relevant for statistical practice than inference at extreme levels near or beyond the limits of the sample, we defer the results of the analysis at the intermediate level to Section C of the Appendix.

The interval $\widehat{\mathrm{I}}_4^\star(0.95)$ appears to have coverage probability closest to the nominal level overall when the underlying distribution has a finite variance; note that $\widetilde{\mathrm{I}}_4^\star(0.95)$ performs worse when $\gamma < 0$, due to the fact that its construction neglects various sources of bias that are nonetheless substantial in that setting. By contrast, $\widetilde{\mathrm{I}}_4^\star(0.95)$ seems to offer the best trade-off between coverage and interval length for infinite-variance distributions, whereas $\widehat{\mathrm{I}}_4^\star(0.95)$ has slightly better coverage at the expense of being much wider. The naive Gaussian intervals $\widehat{\mathrm{I}}_1^\star(0.95)$, $\widehat{\mathrm{I}}_2^\star(0.95)$, $\widetilde{\mathrm{I}}_1^\star(0.95)$ and $\widetilde{\mathrm{I}}_2^\star(0.95)$ do not perform well even when the conditions for their theoretical validity are satisfied. The interpretation here is that the Gaussian asymptotic behavior of $\widehat{\mathrm{ES}}_n^\star$ and $\widetilde{\mathrm{ES}}_n^\star$, giving rise to the naive Gaussian confidence intervals, is obtained by neglecting the uncertainty in $\widehat{\mathrm{ES}}_n$, $\widetilde{\mathrm{ES}}_n$ and $\widehat{a}_n$, even though it is substantial in finite-sample settings.

5.2 Real data application 1: The OpenFEMA dataset

The OpenFEMA dataset, provided by the US government[1], contains a historical database of records of residential flood insurance claims in the USA updated at a monthly frequency. Our goal is to infer the average value of an extremely high claim, defined here as happening with probability 0.5%. To alleviate concerns linked to possible non-stationarity of the data, we consider the claims consecutive to floods caused by a stream, river, or lake overflow that occurred in 2012 (sample 1), on the one hand, and 2017 (sample 2), on the other hand. We further stratify along floods rated A in the database (each corresponding to a so-called Special Flood with no Base Flood Elevation on the insurance rate map) and those rated B (each corresponding to a so-called Moderate Flood from primary water source). A scatterplot of the locations of the claims, superimposed on a map of the USA, is given in Figure 1.

After data cleaning, this yields four samples of flood insurance claims, expressed in thousands of USD, denoted by A-2012 ($n = 555$), B-2012 ($n = 402$), A-2017 ($n = 1{,}131$) and B-2017 ($n = 925$). The large cluster of claims in New Jersey and New York in 2012 is linked to Hurricane Sandy in late October 2012. The two large clusters of 2017 claims are related to Hurricane Harvey in southeast Texas and Louisiana in August 2017, and to Hurricane Irma in Florida in September 2017. We give in Figure 2 moment extreme value index estimates and extrapolated Expected Shortfall estimates at level 0.995, along with the asymptotic confidence intervals $\widehat{\mathrm{I}}_4^\star(0.95)$ and $\widetilde{\mathrm{I}}_4^\star(0.95)$ constructed in Section 4 at the nominal confidence level 0.95. The analysis of sample A-2012 reveals a heavy right tail, with the estimates $\widehat{\mathrm{ES}}_n^\star$ and $\widetilde{\mathrm{ES}}_n^\star$ agreeing with the Weissman-Hill estimates $\widehat{\mathrm{ES}}_n^{\mathrm{W}}$ and $\widetilde{\mathrm{ES}}_n^{\mathrm{W}}$. The analysis of sample A-2017 is less conclusive, with the extreme value index estimate not significantly different from 0 at the 95% asymptotic confidence level, although the point estimate is always positive. In this

[1] Freely available at `https://www.fema.gov/openfema-data-page/fima-nfip-redacted-claims-v2`

situation, it is better to avoid assuming that the right tail is heavy, and therefore to use the estimates $\widehat{\mathrm{ES}}_n^\star$ and $\widetilde{\mathrm{ES}}_n^\star$ instead of the Weissman-Hill estimates. Samples B-2012 and B-2017 are pretty clearly light-tailed ($\gamma = 0$), and we observe that for such samples the Weissman-Hill estimates very quickly drift away from the estimates $\widehat{\mathrm{ES}}_n^\star$ and $\widetilde{\mathrm{ES}}_n^\star$. From an applied standpoint, it is reasonable to conclude that while there are clear differences between the tails along zone types, it is not obvious that the extreme Expected Shortfall changes across years within a zone even if the uncertainty about estimates is affected by changing sample sizes: a rough point estimate of ES(0.995) is 500,000 USD for zone A, and 400,000 USD for zone B.

5.3 Real data application 2: Daily financial loss returns

We examine the extreme value behavior of two series of daily loss returns (*i.e.* negative log-returns) for the CAC 40 and FTSE 100 indices, from 9 December 2008 to 12 July 2016 for the former and from 6 July 2005 to 10 June 2013 for the latter, covering the subprime financial crisis. These time series may feature substantial serial dependence. To handle this dependence in a dynamic setting, it is usual practice to apply a filter such as an ARMA-GARCH model. After goodness-of-fit checks, the residuals from the fitted model may be treated as independent observations and extreme value theory may be applied; this kind of procedure has been used since at least McNeil and Frey (2000). To the best of our knowledge, most of the recent methodological contributions in this framework assume that the residuals are heavy-tailed: see, among others, He and Einmahl (2016), Girard et al. (2021) and Kaibuchi et al. (2022).

However, Figure 3 provides empirical evidence of the fact that the residuals may have light-tailed or short-tailed distributions. Following the methodology of Girard et al. (2022), we estimate an ARMA(1,1)-GARCH(1,1) model on successive rolling windows of length $n =$ 1,500, for each dataset. The corresponding residuals form 438 (resp. 503) samples of size n for the CAC 40 (resp. FTSE 100) data. We estimate the extreme value index γ of each sample of residuals using the estimator $\widehat{\gamma}_n^{\mathrm{Mom}}$. As illustrated in the first two series of panels in Figure 3, the heavy right tail model $\gamma > 0$ can reasonably be excluded for these samples, and depending on the rolling window chosen, the appropriate model seems to be either light-tailed or short-tailed, although it is often difficult to firmly rule out a light-tailed model. This is especially clear when considering the final rolling window in each dataset, the moment plot of which is indicated in cyan in the leftmost panels (A) and (E) in Figure 3. It is reasonable to conclude from the associated asymptotic 95% confidence intervals in panels (B) and (F) that the residuals are, respectively, light-tailed for the CAC 40 and short-tailed for the FTSE 100 over this window. We arrived at the same conclusion using a more parsimonious GARCH(1,1) model and/or a different sample size n for the rolling window. Although this variety of tail behaviors does not seem to have been appreciated before in the statistical literature, it is of utmost interest to practitioners concerned with the accuracy of daily forecasts; as we showed in our simulation study, even when the data is heavy-tailed, there is no substantial drawback in using the general estimators $\widehat{\mathrm{ES}}_n^\star$ and $\widetilde{\mathrm{ES}}_n^\star$ we provide, whereas the Weissman-Hill estimators $\widehat{\mathrm{ES}}_n^{\mathrm{W}}$ and $\widetilde{\mathrm{ES}}_n^{\mathrm{W}}$ perform poorly outside of the heavy right tail model.

The resulting residual-based estimates $\widehat{\mathrm{ES}}_n^\star(1-1/n)$ and $\widetilde{\mathrm{ES}}_n^\star(1-1/n)$ obtained over the last rolling window are represented in panels (C) for CAC 40 and (G) for FTSE 100, along with their 95% confidence intervals $\widehat{\mathrm{I}}_4^\star(0.95)$ and $\widetilde{\mathrm{I}}_4^\star(0.95)$, respectively. They point towards very similar estimates and confidence intervals, with clearly identified stable regions around $k = 200$ resulting in pointwise $\mathrm{ES}(1-1/n)$ estimates around 5.3 for CAC 40 residuals and 3.9 for FTSE 100 residuals. Dynamic predictions of the extreme ES for the next day given past information can then easily be obtained for the raw data using the ARMA(1,1)-GARCH(1,1) structure. We do so by adapting the procedure described in Daouia et al. (2024b) and we provide, in the rightmost panels of Figure 3, the point forecast $\widetilde{\mathrm{ES}}_n^\star(1-1/n)$ in (D) and $\widehat{\mathrm{ES}}_n^\star(1-1/n)$ in (H) of $\mathrm{ES}(1-1/n)$ for the daily loss returns over the observation period, along with their 95% asymptotic confidence intervals and the realization of the future observation.

Acknowledgments

This research was supported by the French National Research Agency under the grants ANR-19-CE40-0013 (ExtremReg project), ANR-23-CE40-0009 (EXSTA project), ANR-17-EURE-0010 (EUR CHESS) and ANR-11-LABX-0020-01 (Centre Henri Lebesgue). A. Daouia and G. Stupfler acknowledge financial support from the TSE-HEC ACPR Chair "Regulation and systemic risks". G. Stupfler acknowledges further support from the Chair Stress Test, RISK Management and Financial Steering of the Foundation Ecole Polytechnique.

References

Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

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

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.

Daouia, A., Padoan, S. A., and Stupfler, G. (2024a). Extreme expectile estimation for short-tailed data. *Journal of Econometrics*, 241(2):105770.

Daouia, A., Stupfler, G., and Usseglio-Carleve, A. (2024b). Bias-reduced and variance-corrected asymptotic Gaussian inference about extreme expectiles. *Statistics & Computing*, 34(130).

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.

Dekkers, A. L. M., Einmahl, J. H. J., and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 17(4):1833–1855.

Diebolt, J., Guillou, A., and Rached, I. (2007). Approximation of the distribution of excesses through a generalized probability-weighted moments method. *Journal of Statistical Planning and Inference*, 137(3):841–857.

Drees, H., Ferreira, A., and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *Annals of Applied Probability*, 14(3):1179–1201.

El Methni, J., Gardes, L., and Girard, S. (2014). Nonparametric estimation of extreme risks from conditional heavy-tailed distributions. *Scandinavian Journal of Statistics*, 41(4):988–1012.

El Methni, J., Gardes, L., and Girard, S. (2018). Kernel estimation of extreme regression risk measures. *Electronic Journal of Statistics*, 12(1):359–398.

El Methni, J. and Stupfler, G. (2017). Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions. *Statistica Sinica*, 27(2):907–930.

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.

Falk, M., Hüsler, J., and Reiss, R.-D. (2011). *Laws of Small Numbers: Extremes and Rare Events (third, revised and extended edition)*. Birkhäuser.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications – Volume II (second edition)*. Wiley, New York.

Föllmer, H. and Schied, A. (2004). *Stochastic Finance – An Introduction in Discrete Time (second edition)*. de Gruyter, Berlin.

Fraga Alves, M. I., Gomes, M. I., de Haan, L., and Neves, C. (2007). A note on second order conditions in extreme value theory: Linking general and heavy tail conditions. *REVSTAT*, 5(3):285–304.

Fung, T. and Seneta, E. (2018). Quantile function expansion using regularly varying functions. *Methodology and Computing in Applied Probability*, 20(4):1091–1103.

Geluk, J. L. and de Haan, L. (2000). Stable probability distributions and their domains of attraction: A direct approach. *Probability and Mathematical Statistics*, 20(1):169–188.

Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2021). Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. *Annals of Statistics*, 49(6):3358–3382.

Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2022). On automatic bias reduction for extreme expectile estimation. *Statistics & Computing*, 32(4):64.

Goegebeur, Y., Guillou, A., Pedersen, T., and Qin, J. (2022). Extreme-value based estimation of the conditional tail moment with application to reinsurance rating. *Insurance: Mathematics and Economics*, 107:102–122.

He, Y. and Einmahl, J. H. J. (2016). Estimation of extreme depth-based quantile regions. *Journal of the Royal Statistical Society Series B*, 79(2):449–461.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5):1163–1174.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261.

Kaibuchi, H., Kawasaki, Y., and Stupfler, G. (2022). GARCH-UGH: a bias-reduced approach for dynamic extreme Value-at-Risk estimation in financial time series. *Quantitative Finance*, 22(7):1277–1294.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3–4):271–300.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional Value-at-Risk for general loss distributions. *Journal of Banking* & *Finance*, 26(7):1443–1471.

Smith, R. L. (1987). Estimating tails of probability distributions. *Annals of Statistics*, 15(3):1174–1207.

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.

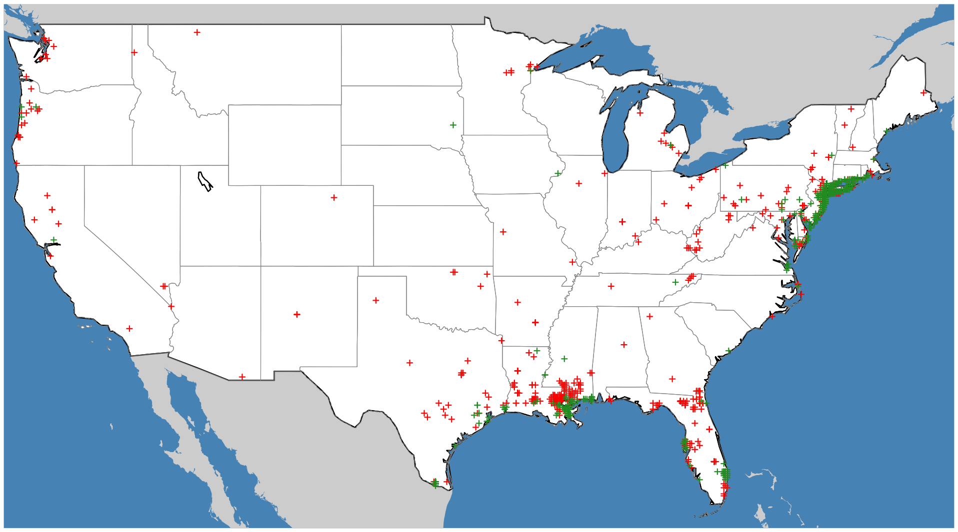
Distribution	$\overline{F}(t) = 1 - F(t) = \mathbb{P}(X > t)$	γ	$a(t)$	ρ	$A(t)$
Pareto	$t^{-\alpha},\ t > 1$	$1/\alpha$	$\frac{1}{\alpha} t^{1/\alpha}$	$-\infty$	0
Generalized Pareto	$(1 + t/\alpha)^{-\alpha},\ t > 0$	$1/\alpha$	$t^{1/\alpha}$	$-\infty$	0
Burr ($\alpha \neq \beta$)	$(1 + t^{\beta/\alpha})^{-1/\beta},\ t > 0$	α	$\alpha t^{\alpha} - \frac{\alpha}{\beta}(\alpha - \beta) t^{\alpha-\beta}$	$-\beta$	$(\alpha - \beta) t^{-\beta}$
Dagum ($\alpha \neq 1$)	$1 - (1 + t^{-\alpha})^{-\beta},\ t > 0$	$1/\alpha$	$\frac{\beta^{1/\alpha}}{\alpha} t^{1/\alpha} - \left(\frac{1}{\alpha} - 1\right)\left(\frac{1}{\beta} + 1\right) \frac{\beta^{1/\alpha}}{2\alpha} t^{1/\alpha - 1}$	-1	$\frac{1}{2}\left(\frac{1}{\alpha} - 1\right)\left(\frac{1}{\beta} + 1\right) t^{-1}$
Dagum ($\alpha = 1$, $\beta \neq 1$)	$1 - (1 + t^{-1})^{-\beta},\ t > 0$	1	$\beta t - \frac{\beta}{12}\left(\frac{1}{\beta} + 1\right)\left(\frac{1}{\beta} - 1\right) t^{-1}$	-2	$\frac{1}{6}\left(\frac{1}{\beta} + 1\right)\left(\frac{1}{\beta} - 1\right) t^{-2}$
Fréchet ($\alpha \neq 1$)	$1 - \exp(-t^{-\alpha}),\ t > 0$	$1/\alpha$	$\frac{1}{\alpha} t^{1/\alpha} - \frac{1}{2\alpha}\left(\frac{1}{\alpha} - 1\right) t^{1/\alpha - 1}$	-1	$\frac{1}{2}\left(\frac{1}{\alpha} - 1\right) t^{-1}$
Fréchet ($\alpha = 1$)	$1 - \exp(-t^{-1}),\ t > 0$	1	$t + \frac{1}{12} t^{-1}$	-2	$-\frac{1}{6} t^{-2}$
Inverse-Gamma	$\int_0^{1/t} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}\,\mathrm{d}x,\ t > 0$	$1/\alpha$	$\frac{1}{\alpha(\Gamma(\alpha+1))^{1/\alpha}} t^{1/\alpha} + \frac{(\Gamma(\alpha+1))^{1/\alpha}}{2\alpha(\alpha+1)(\alpha+2)} t^{-1/\alpha}$	$-2/\alpha$	$-\frac{(\Gamma(\alpha+1))^{2/\alpha}}{\alpha(\alpha+1)(\alpha+2)} t^{-2/\alpha}$
Cauchy	$1/2 - \arctan(t)/\pi$	1	$\frac{1}{\pi} t + \frac{\pi}{3} t^{-1}$	-2	$-\frac{2\pi^2}{3} t^{-2}$
Student	$\int_t^{\infty} \frac{\Gamma((\alpha+1)/2)}{\sqrt{\alpha\pi}\Gamma(\alpha/2)} \left(1 + \frac{x^2}{\alpha}\right)^{-(\alpha+1)/2} \mathrm{d}x$	$1/\alpha$	$\frac{C_\alpha}{\alpha} t^{1/\alpha} + \frac{\alpha+1}{2(\alpha+2)} C_\alpha^{-1} t^{-1/\alpha}$	$-2/\alpha$	$-\frac{\alpha+1}{\alpha+2} C_\alpha^{-2} t^{-2/\alpha}$
Exponential	$\exp(-t),\ t > 0$	0	1	$-\infty$	0
Gamma ($\alpha \neq 1$)	$\int_t^{\infty} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}\,\mathrm{d}x,\ t > 0$	0	$1 + \frac{\alpha-1}{\log(t)} - (\alpha-1)^2 \frac{\log\log(t)}{\log^2(t)} + (\alpha-1)\frac{\log\Gamma(\alpha)+\alpha-2}{\log^2(t)}$	0	$-\frac{\alpha-1}{\log^2(t)}$
Weibull ($\beta \neq 1$)	$\exp(-t^{\beta}),\ t > 0$	0	$\frac{1}{\beta} \log(t)^{1/\beta - 1}$	0	$\left(\frac{1}{\beta} - 1\right) \frac{1}{\log(t)}$
Gumbel	$1 - \exp(-\exp(-t))$	0	$1 + \frac{1}{2} t^{-1}$	-1	$-\frac{1}{2} t^{-1}$
Logistic	$1/(1 + \exp(t))$	0	$1 + t^{-1}$	-1	$-t^{-1}$
Normal	$1 - \Phi(t) = \int_t^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\mathrm{d}x$	0	$\frac{1}{\sqrt{2\log(t)}}\left(1 + \frac{\log\log(t)}{4\log(t)} + \left(\frac{1}{2}\log(4\pi) - 1\right)\frac{1}{2\log(t)}\right)$	0	$-\frac{1}{2\log(t)}$
Arcsine	$1 - 2\arcsin(\sqrt{t})/\pi,\ t \in [0,1]$	-2	$\frac{\pi^2}{2} t^{-2} - \frac{\pi^4}{12} t^{-4}$	-2	$\frac{\pi^2}{3} t^{-2}$
Beta ($\alpha \neq 1$)	$\int_t^1 \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x,\ t \in [0,1]$	$-1/\beta$	$\frac{(\beta B(\alpha,\beta))^{1/\beta}}{\beta} t^{-1/\beta} - \frac{2(1-\alpha)(\beta B(\alpha,\beta))^{2/\beta}}{\beta(\beta+1)} t^{-2/\beta}$	$-1/\beta$	$\frac{2(1-\alpha)(\beta B(\alpha,\beta))^{1/\beta}}{\beta(\beta+1)} t^{-1/\beta}$
Beta ($\alpha = 1$)	$(1-t)^{\beta},\ t \in [0,1]$	$-1/\beta$	$\frac{1}{\beta} t^{-1/\beta}$	$-\infty$	0
Kumaraswamy ($\alpha \neq 1$)	$(1 - t^{\alpha})^{\beta},\ t \in [0,1]$	$-1/\beta$	$\frac{1}{\alpha\beta} t^{-1/\beta} - \frac{1}{\alpha\beta}\left(\frac{1}{\alpha} - 1\right) t^{-2/\beta}$	$-1/\beta$	$\frac{1}{\beta}\left(\frac{1}{\alpha} - 1\right) t^{-1/\beta}$
Reverse-Burr	$(1 + (1-t)^{-\beta/\alpha})^{-1/\beta},\ t < 1$	$-\alpha$	$\alpha t^{-\alpha} + \frac{\alpha}{\beta}(\alpha + \beta) t^{-\beta-\alpha}$	$-\beta$	$-(\alpha + \beta) t^{-\beta}$
Reverse-Fréchet	$1 - \exp(-(1-t)^{\alpha}),\ t < 1$	$-1/\alpha$	$\frac{1}{\alpha} t^{-1/\alpha} + \frac{1}{2\alpha}\left(\frac{1}{\alpha} + 1\right) t^{-1/\alpha - 1}$	-1	$-\frac{1}{2}\left(\frac{1}{\alpha} + 1\right) t^{-1}$
Triangular	$1 - \frac{t^2}{2},\ t \in [0,1]$ and $\frac{(2-t)^2}{2},\ t \in [1,2]$	$-1/2$	$(2t)^{-1/2}$	$-\infty$	0

Table 1: A list of standard, unit-scale continuous distributions satisfying $\mathcal{C}_2(\gamma, a, \rho, A)$, with the associated values of γ, ρ and the functions a and A. The parameters α and β are positive, $\Gamma : x \in (0,\infty) \mapsto \Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\mathrm{d}t$ denotes Euler's Gamma function, and $B : (x,y) \in (0,\infty)^2 \mapsto B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\mathrm{d}t$ denotes Euler's Beta function. When, for all $x > 0$, $(U(tx) - U(t))/a(t) = (x^\gamma - 1)/\gamma$ for t large enough (*i.e.* when the Generalized Pareto model for the right tail is actually exact), we set $\rho = -\infty$ and $A(t) \equiv 0$. For the Student distribution with α degrees of freedom, we set $C_\alpha = \left(\frac{\Gamma((\alpha+1)/2)}{\sqrt{\alpha\pi}\Gamma(\alpha/2)} \alpha^{(\alpha-1)/2}\right)^{1/\alpha}$. [The Burr distribution with equal parameters $\alpha = \beta$ and the Dagum distribution with parameters $\alpha = \beta = 1$ are equal to the Generalized Pareto distribution with unit extreme value index. The Cauchy distribution is the Student distribution with 1 degree of freedom, the exponential distribution is the Gamma distribution with unit shape parameter, and the Arcsine distribution is the Beta(1/2, 1/2) distribution.]

Distribution	γ	$\mathrm{ES}(\tau_n')$	$\widehat{\mathrm{I}}_1^\star(0.95)$	$\widehat{\mathrm{I}}_2^\star(0.95)$	$\widehat{\mathrm{I}}_3^\star(0.95)$	$\widehat{\mathrm{I}}_4^\star(0.95)$	$\widetilde{\mathrm{I}}_1^\star(0.95)$	$\widetilde{\mathrm{I}}_2^\star(0.95)$	$\widetilde{\mathrm{I}}_3^\star(0.95)$	$\widetilde{\mathrm{I}}_4^\star(0.95)$
Kumaraswamy $\alpha = \beta = 2$	-0.5	0.989	[1.404,1.661] (0)	[0.947,1.014] (0.838)	[0.952,1.025] (0.920)	[0.950,1.028] (0.941)	[1.413,1.670] (0)	[0.946,1.012] (0.823)	[0.945,1.011] (0.806)	[0.948,1.014] (0.841)
Reverse-Burr $\alpha = 1/4,\ \beta = 3$	-0.25	0.858	[1.950,3.100] (0)	[0.757,0.990] (0.889)	[0.778,1.018] (0.880)	[0.765,1.040] (0.939)	[2.039,3.242] (0)	[0.754,0.987] (0.882)	[0.751,0.984] (0.887)	[0.760,1.002] (0.892)
Kumaraswamy $\alpha = 1,\ \beta = 10$	-0.1	0.544	[1.263,2.202] (0)	[0.457,0.692] (0.904)	[0.472,0.725] (0.910)	[0.453,0.755] (0.965)	[1.342,2.338] (0)	[0.456,0.691] (0.901)	[0.444,0.697] (0.926)	[0.455,0.728] (0.935)
Gumbel	0	7.908	[21.69,43.24] (0)	[6.189,10.50] (0.878)	[6.381,11.17] (0.909)	[5.989,11.77] (0.954)	[24.13,48.11] (0)	[6.171,10.48] (0.875)	[5.867,10.71] (0.907)	[6.064,11.34] (0.923)
Exponential $\lambda = 1$	0	7.908	[18.20,34.03] (0)	[6.305,10.92] (0.907)	[6.461,11.68] (0.937)	[6.021,12.37] (0.976)	[19.63,36.71] (0)	[6.299,10.91] (0.904)	[5.921,11.20] (0.945)	[6.131,11.89] (0.957)
Pareto $\alpha = 10$	0.1	2.217	[2.057,2.382] (0.891)	[1.831,2.524] (0.876)	[1.850,2.645] (0.925)	[1.785,2.771] (0.957)	[2.058,2.383] (0.891)	[1.831,2.524] (0.876)	[1.766,2.577] (0.895)	[1.797,2.687] (0.935)
Pareto $\alpha = 4$	0.25	7.498	[6.215,8.970] (0.880)	[4.837,9.580] (0.850)	[4.638,10.64] (0.918)	[3.914,11.70] (0.950)	[6.222,8.979] (0.879)	[4.840,9.580] (0.850)	[4.138,10.20] (0.891)	[4.341,11.23] (0.935)
Fréchet $\alpha = 2$	0.5	63.24	[48.67,105.6] (0.822)	[28.60,92.07] (0.806)	[17.81,113.1] (0.902)	[-0.397,140.8] (0.946)	[49.80,108.2] (0.802)	[28.66,92.18] (0.807)	[12.40,107.4] (0.883)	[13.33,131.9] (0.936)
Pareto $\alpha = 5/3$	0.6	157.7	[99.86,241.1] (0.805)	[59.53,231.1] (0.772)	[21.98,298.2] (0.891)	[-58.56,416.8] (0.946)	[100.7,243.0] (0.812)	[59.72,231.5] (0.774)	[7.192,283.3] (0.869)	[4.642,375.7] (0.933)

Table 2: Inference about $\mathrm{ES}(\tau_n') = \mathrm{ES}(0.999)$ – For each interval among $\widehat{\mathrm{I}}_1^\star(0.95)$, $\widehat{\mathrm{I}}_2^\star(0.95)$, $\widehat{\mathrm{I}}_3^\star(0.95)$, $\widehat{\mathrm{I}}_4^\star(0.95)$, $\widetilde{\mathrm{I}}_1^\star(0.95)$, $\widetilde{\mathrm{I}}_2^\star(0.95)$, $\widetilde{\mathrm{I}}_3^\star(0.95)$ and $\widetilde{\mathrm{I}}_4^\star(0.95)$, and in each tested case, we report between square brackets the median values (over the $N = 10{,}000$ replications) of its lower and upper bounds; the empirical coverage probability is indicated between round brackets. The sample size is $n = 1{,}000$.

Stream, river, or lake overflows in 2012



Stream, river, or lake overflows in 2017

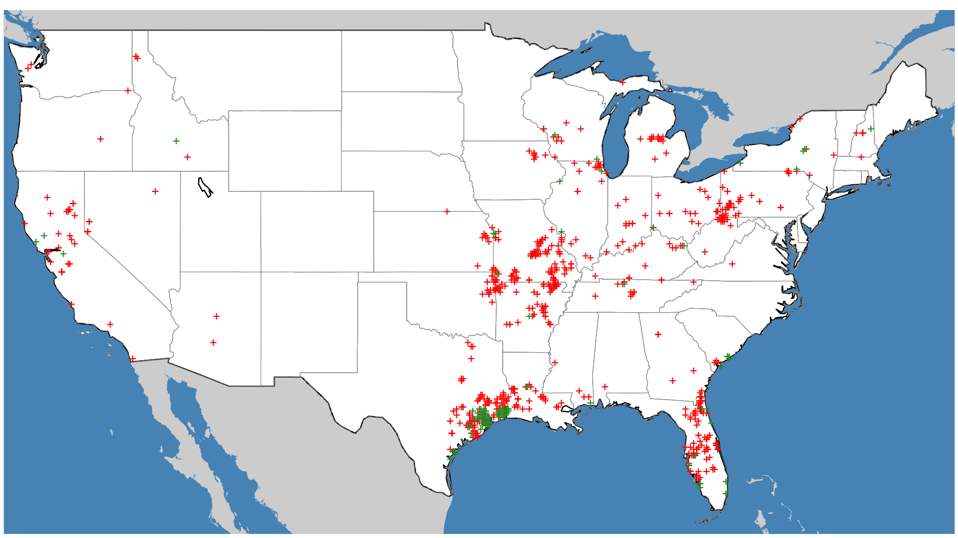


Figure 1: OpenFEMA Flood Insurance data – Locations of flood claims consecutive to a stream, river, or lake overflow in 2012 (top) and 2017 (bottom) in the United States. A red cross denotes a claim in an area rated A (Special Flood with no Base Flood Elevation) and a green cross denotes a claim in an area rated B (Moderate Flood from primary water source).

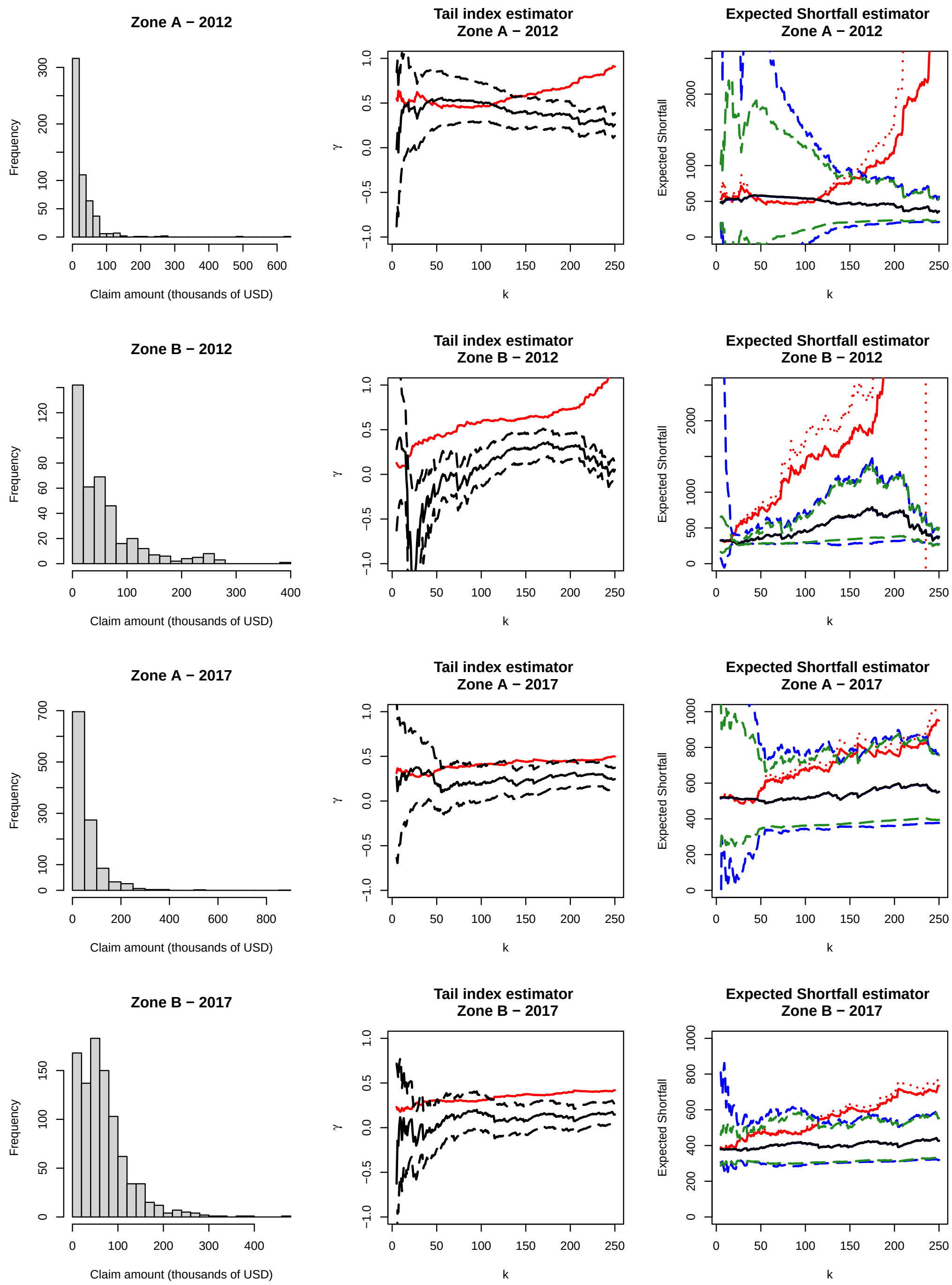


Figure 2: OpenFEMA Flood Insurance data – Left panels: Histograms of the claims. Middle panels: Extreme value index estimates $\widehat{\gamma}_n^{\text{Mom}}$ (solid black) with 95% asymptotic Gaussian confidence intervals (dashed black), and Hill estimates (solid red) as functions of k. Right panels: Expected Shortfall estimates of ES(0.995) (solid black) with 95% asymptotic confidence intervals $\widehat{\mathrm{I}}_4^\star(0.95)$ (dashed blue) and $\widetilde{\mathrm{I}}_4^\star(0.95)$ (dashed green), and Weissman-Hill Expected Shortfall estimates $\widehat{\mathrm{ES}}_n^{\mathrm{W}}(0.995)$ (solid red) and $\widetilde{\mathrm{ES}}_n^{\mathrm{W}}(0.995)$ (dotted red) as functions of k.

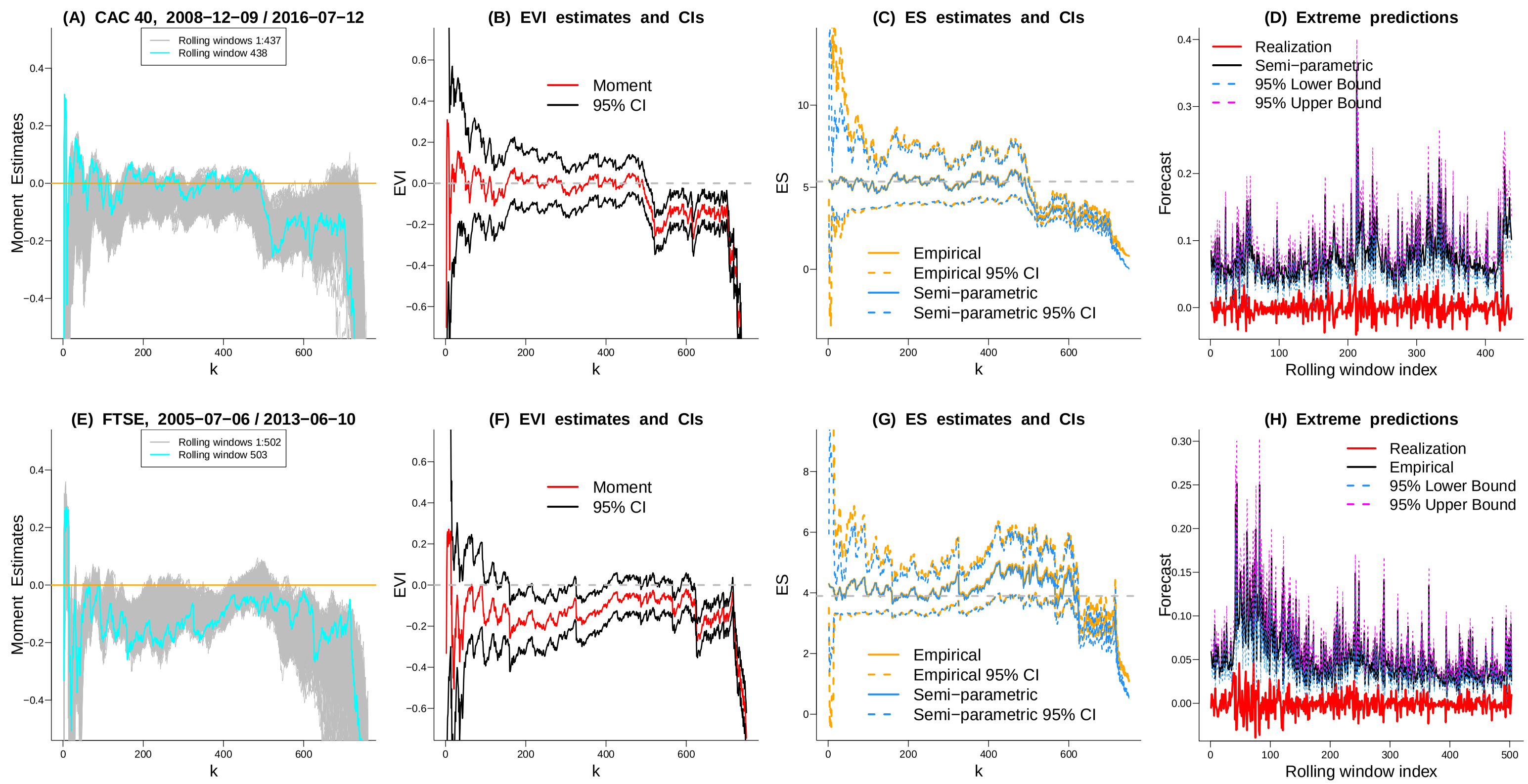


Figure 3: Financial daily loss returns data – Top panels: CAC 40 data from 9 December 2008 to 12 July 2016, bottom panels: FTSE 100 data from 6 July 2005 to 10 June 2013. Panels (A) and (E): Extreme value index estimates $\widehat{\gamma}_n^{\text{Mom}}$ for the residuals obtained from the ARMA-GARCH model over successive rolling windows of length $n = 1{,}500$. Panels (B) and (F): Over the highlighted rolling window (26 August 2010-12 July 2016 for CAC 40, 2 July 2007-10 June 2013 for FTSE 100), estimate $\widehat{\gamma}_n^{\text{Mom}}$ (red curve) with 95% asymptotic Gaussian confidence intervals (black curves). Panels (C) and (G): Over this same window, estimate $\widehat{\text{ES}}_n^{\star}(\tau_n')$ (solid orange) of $\text{ES}(\tau_n')$ for the sample of residuals at level $\tau_n' = 1 - 1/1{,}500$ with 95% asymptotic confidence intervals $\widehat{\text{I}}_4^{\star}(0.95)$ (dashed orange), along with the semiparametric quantile-based versions $\widetilde{\text{ES}}_n^{\star}(\tau_n')$ and $\widetilde{\text{I}}_4^{\star}(0.95)$ (blue curves). Panels (D) and (H): Next-day forecasts (black curve) $\widetilde{\text{ES}}_n^{\star}(\tau_n')$ in panel (D) and $\widehat{\text{ES}}_n^{\star}(\tau_n')$ in panel (H) of $\text{ES}(\tau_n')$ for the daily loss returns computed sequentially over each rolling window, with 95% lower (blue curve) and upper (magenta curve) asymptotic confidence bounds, along with the realization of the future observation (red curve).

Appendix to the paper “A unified theory of extreme Expected Shortfall inference”

Abdelaati Daouia, Gilles Stupfler & Antoine Usseglio-Carleve

This appendix contains all necessary auxiliary results, their proofs and the proofs of our main results, and provides extra finite-sample results about our simulation study. Unless specified otherwise, we denote throughout by F and $\overline{F} = 1 - F$ the distribution function and survival function of the random variable of interest X. The associated quantile function is $q : \tau \mapsto \inf\{x \in \mathbb{R} \,|\, F(x) \geq \tau\}$, and the tail quantile function is $U : t \mapsto q(1 - 1/t)$, for $t > 1$.

A Auxiliary results and proofs

Our first result gives simple sufficient conditions under which condition $\mathcal{C}_2(\gamma, a, \rho, A)$ holds. They are formulated in terms of asymptotic expansions of the tail quantile function.

Lemma A.1 (Sufficient conditions for $\mathcal{C}_2(\gamma, a, \rho, A)$)**.** *Let U be the tail quantile function of the random variable X.*

- *(Heavy tails) Assume that*

$$U(t) = t^{\alpha}(C + Dt^{-\beta} + \mathrm{o}(t^{-\beta})) \text{ as } t \to \infty$$

with $\alpha, \beta, C > 0$ and $D \neq 0$.

1. *If $\alpha \neq \beta$, then condition $\mathcal{C}_2(\gamma, a, \rho, A)$ holds with $\gamma = \alpha$, $\rho = -\beta$,*

$$A(t) = -\frac{\beta(\alpha - \beta)}{\alpha} \frac{D}{C} t^{-\beta} \text{ and } a(t) = \alpha C t^{\alpha} \left(1 - \frac{1}{\beta} A(t)\right) = \alpha C t^{\alpha} \left(1 + \frac{\alpha - \beta}{\alpha} \frac{D}{C} t^{-\beta}\right).$$

2. *If $\alpha = \beta$ and moreover*

$$U(t) = t^{\alpha}(C + Dt^{-\alpha} + D' t^{-\alpha - \alpha'} + \mathrm{o}(t^{-\alpha - \alpha'})) \text{ as } t \to \infty$$

with $\alpha' > 0$ and $D' \neq 0$, then condition $\mathcal{C}_2(\gamma, a, \rho, A)$ holds with $\gamma = \alpha$, $\rho = -\alpha - \alpha'$,

$$A(t) = \frac{\alpha'(\alpha + \alpha')}{\alpha} \frac{D'}{C} t^{-\alpha - \alpha'} \text{ and } a(t) = \alpha C t^{\alpha} \left(1 - \frac{\alpha'}{\alpha} \frac{D'}{C} t^{-\alpha - \alpha'}\right).$$

- *(Light tails) Assume that*

$$U(t) = \log^{\alpha}(t)(C + Dt^{-\beta} \log^{\delta}(t) + \mathrm{o}(t^{-\beta} \log^{\delta}(t))) \text{ as } t \to \infty$$

with $\alpha, \beta, C > 0$ and $D, \delta \in \mathbb{R}$.

1. If $\alpha = 1$ and $D \neq 0$, then condition $\mathcal{C}_2(\gamma, a, \rho, A)$ holds with $\gamma = 0$, $\rho = -\beta$,

$$A(t) = \beta^2 \frac{D}{C} t^{-\beta} \log^{1+\delta}(t) \quad \textit{and} \quad a(t) = C\left(1 - \frac{1}{\beta}A(t)\right) = C\left(1 - \beta\frac{D}{C}t^{-\beta}\log^{1+\delta}(t)\right).$$

2. If $\alpha \neq 1$, then condition $\mathcal{C}_2(\gamma, a, \rho, A)$ holds with $\gamma = \rho = 0$,

$$A(t) = \frac{\alpha - 1}{\log(t)} \quad \textit{and} \quad a(t) = \alpha C \log^{\alpha-1}(t).$$

- *(Short tails) Assume that $U(\infty) < +\infty$ and*

$$U(\infty) - U(t) = t^{-\alpha}(C + Dt^{-\beta} + \mathrm{o}(t^{-\beta})) \textit{ as } t \to \infty$$

with $\alpha, \beta, C > 0$ and $D \neq 0$. Then condition $\mathcal{C}_2(\gamma, a, \rho, A)$ holds with $\gamma = -\alpha$, $\rho = -\beta$,

$$A(t) = -\frac{\beta(\alpha+\beta)}{\alpha}\frac{D}{C}t^{-\beta} \quad \textit{and} \quad a(t) = \alpha C t^{-\alpha}\left(1 - \frac{1}{\beta}A(t)\right) = \alpha C t^{-\alpha}\left(1 + \frac{\alpha+\beta}{\alpha}\frac{D}{C}t^{-\beta}\right).$$

Proof of Lemma A.1. We prove each statement separately.

(Heavy tails) Assume that $U(t) = t^{\alpha}(C + Dt^{-\beta} + \mathrm{o}(t^{-\beta}))$ as $t \to \infty$, with $\alpha, \beta, C > 0$ and $D \neq 0$. If $\alpha \neq \beta$, set

$$A(t) = -\frac{\beta(\alpha-\beta)}{\alpha}\frac{D}{C}t^{-\beta} \quad \text{and} \quad a(t) = \alpha C t^{\alpha}\left(1 + \frac{\alpha-\beta}{\alpha}\frac{D}{C}t^{-\beta}\right).$$

Then

$$\begin{aligned}\frac{U(tx) - U(t)}{a(t)} - \frac{x^{\alpha}-1}{\alpha} &= -\frac{\alpha-\beta}{\alpha}\frac{D}{C}t^{-\beta}\frac{x^{\alpha}-1}{\alpha} + \frac{\alpha-\beta}{\alpha}\frac{D}{C}t^{-\beta}\frac{x^{\alpha-\beta}-1}{\alpha-\beta} + \mathrm{o}(t^{-\beta}) \\ &= A(t)\int_1^x s^{\alpha-1}\left(\int_1^s u^{-\beta-1}\mathrm{d}u\right)\mathrm{d}s + \mathrm{o}(|A(t)|)\end{aligned}$$

as announced. If in fact $\alpha = \beta$ and

$$U(t) = t^{\alpha}(C + Dt^{-\alpha} + D't^{-\alpha-\alpha'} + \mathrm{o}(t^{-\alpha-\alpha'})) = D + t^{\alpha}(C + D't^{-\alpha-\alpha'} + \mathrm{o}(t^{-\alpha-\alpha'}))$$

as $t \to \infty$, then $U(t) - D = t^{\alpha}(C + D't^{-\alpha-\alpha'} + \mathrm{o}(t^{-\alpha-\alpha'}))$ as $t \to \infty$ and the above calculation applies with D replaced by D' and β replaced by $\alpha + \alpha'$.

(Light tails) Assume that $U(t) = \log^{\alpha}(t)(C + Dt^{-\beta}\log^{\delta}(t) + \mathrm{o}(t^{-\beta}\log^{\delta}(t)))$ as $t \to \infty$ with $\alpha, \beta, C > 0$ and $D, \delta \in \mathbb{R}$. If $\alpha = 1$, then for any $x > 0$,

$$\begin{aligned}U(tx) - U(t) &= C\log(x) + D(x^{-\beta}-1)t^{-\beta}\log^{1+\delta}(t) + \mathrm{o}(t^{-\beta}\log^{1+\delta}(t)) \\ &= C\left(1 - \beta\frac{D}{C}t^{-\beta}\log^{1+\delta}(t)\right)\log(x) \\ &\quad + C \times \beta^2\frac{D}{C}t^{-\beta}\log^{1+\delta}(t)\int_1^x s^{-1}\left(\int_1^s u^{-\beta-1}\mathrm{d}u\right)\mathrm{d}s + \mathrm{o}(t^{-\beta}\log^{1+\delta}(t)).\end{aligned}$$

This proves the result in this case. If in fact $\alpha \neq 1$, then for any $x > 0$,

$$U(tx) = C\log^{\alpha}(t)\left(1 + \alpha\frac{\log(x)}{\log(t)} + \frac{\alpha(\alpha-1)}{2}\frac{\log^2(x)}{\log^2(t)} + \mathrm{o}\left(\frac{1}{\log^2(t)}\right)\right)$$

and therefore

$$\begin{aligned}\frac{U(tx)-U(t)}{\alpha C\log^{\alpha-1}(t)} &= \log(x) + \frac{\alpha-1}{\log(t)} \times \frac{1}{2}\log^2(x) + \mathrm{o}\left(\frac{1}{\log(t)}\right) \\ &= \log(x) + \frac{\alpha-1}{\log(t)}\int_1^x s^{-1}\left(\int_1^s u^{-1}\mathrm{d}u\right)\mathrm{d}s + \mathrm{o}\left(\frac{1}{\log(t)}\right)\end{aligned}$$

as required.

(Short tails) Assume that $U(\infty) < +\infty$ and $U(\infty) - U(t) = t^{-\alpha}(C + Dt^{-\beta} + \mathrm{o}(t^{-\beta}))$ as $t \to \infty$ with $\alpha, \beta, C > 0$ and $D \neq 0$. Set

$$A(t) = -\frac{\beta(\alpha+\beta)}{\alpha}\frac{D}{C}t^{-\beta} \quad \text{and} \quad a(t) = \alpha C t^{-\alpha}\left(1 - \frac{1}{\beta}A(t)\right) = \alpha C t^{-\alpha}\left(1 + \frac{\alpha+\beta}{\alpha}\frac{D}{C}t^{-\beta}\right).$$

Then

$$\begin{aligned}\frac{U(tx)-U(t)}{a(t)} - \frac{x^{-\alpha}-1}{-\alpha} &= -\frac{\alpha+\beta}{\alpha}\frac{D}{C}t^{-\beta}\frac{x^{-\alpha}-1}{-\alpha} + \frac{\alpha+\beta}{\alpha}\frac{D}{C}t^{-\beta}\frac{x^{-\alpha-\beta}-1}{-\alpha-\beta} + \mathrm{o}(t^{-\beta}) \\ &= A(t)\int_1^x s^{-\alpha-1}\left(\int_1^s u^{-\beta-1}\mathrm{d}u\right)\mathrm{d}s + \mathrm{o}(|A(t)|)\end{aligned}$$

as announced. □

The next two lemmas are dedicated to checking condition $\mathcal{C}_2(\gamma, a, \rho, A)$ for the Gamma distribution with unit scale and for the standard Gaussian distribution. This is done by computing asymptotic expansions of the tail quantile functions of these distributions that are of independent interest; in particular, we obtain approximations that are more precise than those proven in Sections 4.1 and 4.3 in Fung and Seneta (2018).

Lemma A.2 (Asymptotic expansion of the Gamma quantile function)**.** *Let U be the tail quantile function of the Gamma distribution with shape parameter $\alpha \neq 1$, whose distribution function is*

$$F(x) = \int_0^x \frac{1}{\Gamma(\alpha)}t^{\alpha-1}e^{-t}\,\mathrm{d}t \quad \textit{for} \quad x > 0.$$

Then

$$\begin{aligned}U(t) = \log(t)\Bigg(1 &+ (\alpha-1)\frac{\log\log(t)}{\log(t)} - \frac{\log\Gamma(\alpha)}{\log(t)} + (\alpha-1)^2\frac{\log\log(t)}{\log^2(t)} \\ &- \frac{(\alpha-1)(\log\Gamma(\alpha)-1)}{\log^2(t)} + \mathrm{o}\left(\frac{1}{\log^2(t)}\right)\Bigg) \quad \textit{as } t \to \infty.\end{aligned}$$

In particular, the Gamma distribution satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$, with $\gamma = \rho = 0$,

$$a(t) = 1 + \frac{\alpha-1}{\log(t)} - (\alpha-1)^2\frac{\log\log(t)}{\log^2(t)} + (\alpha-1)\frac{\log\Gamma(\alpha)+\alpha-2}{\log^2(t)} \quad \textit{and} \quad A(t) = -\frac{\alpha-1}{\log^2(t)}.$$

Proof of Lemma A.2. The key is to note that for any $\beta > 0$ and any $x > 0$,

$$\int_x^\infty t^{\beta-1}e^{-t}\,\mathrm{d}t = x^{\beta-1}e^{-x} + (\beta-1)\int_x^\infty t^{\beta-2}e^{-t}\,\mathrm{d}t$$

by an integration by parts, and since

$$0 \le \int_x^\infty t^{\beta-2}e^{-t}\,\mathrm{d}t \le \frac{1}{x}\int_x^\infty t^{\beta-1}e^{-t}\,\mathrm{d}t,$$

one has

$$\forall \beta > 0,\ \int_x^\infty t^{\beta-1}e^{-t}\,\mathrm{d}t = x^{\beta-1}e^{-x}\left(1 + \mathrm{O}\left(\frac{1}{x}\right)\right) \text{ as } x \to \infty.$$

Consequently

$$\begin{aligned}1 - F(x) &= \int_x^\infty \frac{1}{\Gamma(\alpha)} t^{\alpha-1}e^{-t}\,\mathrm{d}t \\ &= \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} + \frac{(\alpha-1)x^{\alpha-2}e^{-x}}{\Gamma(\alpha)} + \frac{(\alpha-1)(\alpha-2)}{\Gamma(\alpha)}\int_x^\infty t^{\alpha-3}e^{-t}\,\mathrm{d}t \\ &= \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}\left(1 + \frac{\alpha-1}{x} + \frac{(\alpha-1)(\alpha-2)}{x^2} + \mathrm{O}\left(\frac{1}{x^3}\right)\right) \text{ as } x \to \infty.\end{aligned}$$

The Gamma distribution is unbounded to the right, so $U(t) \to \infty$ as $t \to \infty$, and has a continuous distribution function, so $1 - F(U(t)) = 1/t$ for any $t > 1$. Plugging $x = U(t)$ in the above asymptotic expansion leads in particular to

$$\log(t) = U(t) - (\alpha-1)\log(U(t)) + \log\Gamma(\alpha) - \frac{\alpha-1}{U(t)} + \mathrm{O}\left(\frac{1}{(U(t))^2}\right) \text{ as } t \to \infty. \tag{A.1}$$

We repeatedly use (A.1) in order to get the successive terms appearing in the desired asymptotic expansion of $U(t)$.

Asymptotic equivalent of $U(t)$: The leading term in the right-hand side of (A.1) is obviously $U(t)$, so $U(t) \sim \log(t)$.

First term in the asymptotic expansion of $U(t)$: Write $U(t) = \log(t)(1 + \varepsilon_1(t))$, with $\varepsilon_1(t) \to 0$ as $t \to \infty$. Then (A.1) yields

$$0 = \log(t)\varepsilon_1(t) - (\alpha-1)\log\log(t)(1 + \mathrm{o}(1))$$

and therefore $\varepsilon_1(t) \sim (\alpha-1)\frac{\log\log(t)}{\log(t)}$.

Second term in the asymptotic expansion of $U(t)$: Write

$$U(t) = \log(t)\left(1 + (\alpha-1)\frac{\log\log(t)}{\log(t)}(1 + \varepsilon_2(t))\right),$$

with $\varepsilon_2(t) \to 0$ as $t \to \infty$. We find, using (A.1), that

$$0 = (\alpha-1)\log\log(t)\varepsilon_2(t) + \log\Gamma(\alpha) + \mathrm{o}(1)$$

and therefore $\varepsilon_2(t) \sim -\frac{\log\Gamma(\alpha)}{\alpha-1}\frac{1}{\log\log(t)}$.

Third term in the asymptotic expansion of $U(t)$: Write

$$U(t) = \log(t)\left(1 + (\alpha-1)\frac{\log\log(t)}{\log(t)} - \frac{\log\Gamma(\alpha)}{\log(t)}(1+\varepsilon_3(t))\right),$$

with $\varepsilon_3(t) \to 0$ as $t \to \infty$. Then $1/U(t) \sim 1/\log(t) = \mathrm{o}(\log\log(t)/\log(t))$ and

$$\log(U(t)) = \log\log(t) + (\alpha-1)\frac{\log\log(t)}{\log(t)} + \mathrm{o}\left(\frac{\log\log(t)}{\log(t)}\right).$$

Plugging these two relationships into (A.1) results in

$$0 = \log\Gamma(\alpha)\varepsilon_3(t) + (\alpha-1)^2\frac{\log\log(t)}{\log(t)}(1+\mathrm{o}(1))$$

and therefore $\varepsilon_3(t) \sim -\frac{(\alpha-1)^2}{\log\Gamma(\alpha)}\frac{\log\log(t)}{\log(t)}$.

Fourth term in the asymptotic expansion of $U(t)$: Write finally

$$U(t) = \log(t)\left(1 + (\alpha-1)\frac{\log\log(t)}{\log(t)} - \frac{\log\Gamma(\alpha)}{\log(t)} + (\alpha-1)^2\frac{\log\log(t)}{\log^2(t)}(1+\varepsilon_4(t))\right),$$

with $\varepsilon_4(t) \to 0$ as $t \to \infty$. One has $1/U(t) \sim 1/\log(t)$ and

$$\log(U(t)) = \log\log(t) + (\alpha-1)\frac{\log\log(t)}{\log(t)} - \frac{\log\Gamma(\alpha)}{\log(t)} + \mathrm{o}\left(\frac{1}{\log(t)}\right).$$

Plugging these two relationships into (A.1) results in

$$0 = (\alpha-1)\log\log(t)\varepsilon_4(t) + \log\Gamma(\alpha) - 1 + \mathrm{o}(1)$$

and therefore $\varepsilon_4(t) \sim -\frac{\log\Gamma(\alpha)-1}{\alpha-1}\frac{1}{\log\log(t)}$. Hence the desired asymptotic expansion of U.

To prove that U satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$, note that since

$$1 - F(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}\left(1 + \frac{\alpha-1}{x} + \frac{(\alpha-1)(\alpha-2)}{x^2} + \mathrm{O}\left(\frac{1}{x^3}\right)\right) \text{ as } x \to \infty,$$

one may write the following stronger version of (A.1):

$$\begin{aligned}\log(t) &= U(t) - (\alpha-1)\log(U(t)) + \log\Gamma(\alpha)\\ &\quad - \log\left(1 + \frac{\alpha-1}{U(t)} + \frac{(\alpha-1)(\alpha-2)}{(U(t))^2} + \mathrm{o}\left(\frac{1}{(U(t))^2}\right)\right) \text{ as } t \to \infty.\end{aligned}$$

In particular, for any $x > 0$,

$$\begin{aligned}U(tx) - U(t) &= \log(x) + (\alpha-1)\log\left(\frac{U(tx)}{U(t)}\right)\\ &\quad + \log\left(\frac{1 + (\alpha-1)/U(tx) + (\alpha-1)(\alpha-2)/(U(tx))^2 + \mathrm{o}(1/\log^2(t))}{1 + (\alpha-1)/U(t) + (\alpha-1)(\alpha-2)/(U(t))^2 + \mathrm{o}(1/\log^2(t))}\right). \quad \text{(A.2)}\end{aligned}$$

Since $U(tx)/U(t) \to 1$, this yields, first of all, the convergence $U(tx) - U(t) \to \log(x)$ as $t \to \infty$. Then straightforward calculations based on the asymptotic expansion of $U(t)$ we have previously obtained entail

$$\begin{aligned}\frac{U(tx)}{\log(tx)} - \frac{U(t)}{\log(t)} &= (\alpha-1)\left(\frac{\log\log(tx)}{\log(tx)} - \frac{\log\log(t)}{\log(t)}\right) - \left(\frac{\log\Gamma(\alpha)}{\log(tx)} - \frac{\log\Gamma(\alpha)}{\log(t)}\right) + \mathrm{o}\left(\frac{1}{\log^2(t)}\right) \\ &= \left(-(\alpha-1)\frac{\log\log(t)}{\log^2(t)} + \frac{\log\Gamma(\alpha)+\alpha-1}{\log^2(t)}\right)\log(x) + \mathrm{o}\left(\frac{1}{\log^2(t)}\right) \qquad \text{(A.3)}\end{aligned}$$

and, using the fact that $U(t) \sim \log(t)$,

$$\begin{aligned}&\frac{1+(\alpha-1)/U(tx)+(\alpha-1)(\alpha-2)/(U(tx))^2+\mathrm{o}(1/\log^2(t))}{1+(\alpha-1)/U(t)+(\alpha-1)(\alpha-2)/(U(t))^2+\mathrm{o}(1/\log^2(t))} - 1 \\ &= (\alpha-1)\left(\frac{1}{U(tx)} - \frac{1}{U(t)}\right) + \mathrm{o}\left(\frac{1}{\log^2(t)}\right) \\ &= -(\alpha-1)\frac{\log(x)}{\log^2(t)} + \mathrm{o}\left(\frac{1}{\log^2(t)}\right). \qquad \text{(A.4)}\end{aligned}$$

Combine (A.2), (A.3) and (A.4) and the fact that $U(t)/\log(t) = 1 + \mathrm{O}(\log\log(t)/\log(t))$ to get

$$\begin{aligned}U(tx) - U(t) &= \log(x) + (\alpha-1)\log\left(1 + \frac{\log(x)}{\log(t)}\right) \\ &\quad + (\alpha-1)\log\left(1 + \frac{\log(t)}{U(t)}\left(\frac{U(tx)}{\log(tx)} - \frac{U(t)}{\log(t)}\right)\right) \\ &\quad + \log\left(\frac{1+(\alpha-1)/U(tx)+(\alpha-1)(\alpha-2)/(U(tx))^2+\mathrm{o}(1/\log^2(t))}{1+(\alpha-1)/U(t)+(\alpha-1)(\alpha-2)/(U(t))^2+\mathrm{o}(1/\log^2(t))}\right) \\ &= \log(x)\left(1 + \frac{\alpha-1}{\log(t)} - (\alpha-1)^2\frac{\log\log(t)}{\log^2(t)} + (\alpha-1)\frac{\log\Gamma(\alpha)+\alpha-2}{\log^2(t)}\right) \\ &\quad - \frac{\alpha-1}{\log^2(t)} \times \frac{1}{2}\log^2(x) + \mathrm{o}\left(\frac{1}{\log^2(t)}\right).\end{aligned}$$

The conclusion directly follows. □

Lemma A.3 (Asymptotic expansion of the Gaussian quantile function)**.** *Let U be the tail quantile function of the standard Gaussian distribution. Then*

$$\begin{aligned}U(t) = \sqrt{2\log(t)}\Bigg(1 &- \frac{\log\log(t)}{4\log(t)} - \frac{\log(4\pi)}{4\log(t)} - \frac{\log^2(\log(t))}{32\log^2(t)} - \frac{1}{8}\left(\frac{1}{2}\log(4\pi) - 1\right)\frac{\log\log(t)}{\log^2(t)} \\ &- \frac{1}{4}\left(\frac{1}{8}\log^2(4\pi) - \frac{1}{2}\log(4\pi) + 1\right)\frac{1}{\log^2(t)} + \mathrm{o}\left(\frac{1}{\log^2(t)}\right)\Bigg) \quad \text{as } t \to \infty.\end{aligned}$$

In particular, the standard Gaussian distribution satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$, with $\gamma = \rho = 0$,

$$a(t) = \frac{1}{\sqrt{2\log(t)}}\left(1 + \frac{\log\log(t)}{4\log(t)} + \left(\frac{1}{2}\log(4\pi) - 1\right)\frac{1}{2\log(t)}\right) \quad \text{and} \quad A(t) = -\frac{1}{2\log(t)}.$$

Proof of Lemma A.3. Let Φ be the distribution function of the standard Gaussian distribution. The starting point is to integrate by parts twice in order to get

$$1-\Phi(x)=\int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-u^2/2}\mathrm{d}u=\frac{1}{\sqrt{2\pi}}\frac{e^{-x^2/2}}{x}\left(1-\frac{1}{x^2}\right)+\int_x^\infty \frac{3}{\sqrt{2\pi}}e^{-u^2/2}\frac{\mathrm{d}u}{u^4}$$
$$=\frac{1}{\sqrt{2\pi}}\frac{e^{-x^2/2}}{x}\left(1-\frac{1}{x^2}+\mathrm{O}\left(\frac{1}{x^4}\right)\right) \text{ as } x\to\infty.$$

Consequently

$$-\log(1-\Phi(x))=\frac{x^2}{2}+\log(x)+\frac{1}{2}\log(2\pi)+\frac{1}{x^2}+\mathrm{O}\left(\frac{1}{x^4}\right) \text{ as } x\to\infty.$$

The standard Gaussian distribution is unbounded to the right, so $U(t)\to\infty$ as $t\to\infty$, and Φ is continuous, so $1-\Phi(U(t))=1/t$ for any $t>1$. Plugging $x=U(t)$ in the above asymptotic expansion leads to

$$\log(t)=\frac{(U(t))^2}{2}+\log(U(t))+\frac{1}{2}\log(2\pi)+\frac{1}{(U(t))^2}+\mathrm{O}\left(\frac{1}{(U(t))^4}\right) \text{ as } t\to\infty. \tag{A.5}$$

As in the proof of Lemma A.2, we repeatedly use (A.5) in order to get the successive terms appearing in the desired asymptotic expansion of $U(t)$.

Asymptotic equivalent of $U(t)$: The leading term in the right-hand side of (A.5) is obviously $(U(t))^2/2$, which yields $U(t)\sim\sqrt{2\log(t)}$.

First term in the asymptotic expansion of $U(t)$: Writing $U(t)=\sqrt{2\log(t)}(1+\varepsilon_1(t))$, with $\varepsilon_1(t)\to 0$ as $t\to\infty$, one obtains from (A.5) that

$$0=2\log(t)\varepsilon_1(t)(1+\mathrm{o}(1))+\frac{1}{2}\log\log(t)+\mathrm{O}(1)$$

and therefore $\varepsilon_1(t)\sim-\frac{\log\log(t)}{4\log(t)}$.

Second term in the asymptotic expansion of $U(t)$: Writing

$$U(t)=\sqrt{2\log(t)}\left(1-\frac{\log\log(t)}{4\log(t)}(1+\varepsilon_2(t))\right),$$

with $\varepsilon_2(t)\to 0$ as $t\to\infty$, we get using (A.5)

$$0=-\frac{1}{2}\log\log(t)\varepsilon_2(t)+\frac{1}{2}\log(4\pi)+\mathrm{o}(1)$$

and therefore $\varepsilon_2(t)\sim\frac{\log(4\pi)}{\log\log(t)}$.

Third term in the asymptotic expansion of $U(t)$: Write

$$U(t)=\sqrt{2\log(t)}\left(1-\frac{\log\log(t)}{4\log(t)}-\frac{\log(4\pi)}{4\log(t)}(1+\varepsilon_3(t))\right),$$

with $\varepsilon_3(t) \to 0$ as $t \to \infty$. Then

$$\frac{(U(t))^2}{2} - \log(t) = -\frac{1}{2}\log\log(t) - \frac{1}{2}\log(4\pi)(1+\varepsilon_3(t)) + \frac{\log^2(\log(t))}{16\log(t)}(1+\mathrm{o}(1))$$

and

$$\log(U(t)) = \frac{1}{2}\log\log(t) + \frac{1}{2}\log(2) + \mathrm{o}\left(\frac{\log^2(\log(t))}{\log(t)}\right).$$

Plugging these two asymptotic expansions into (A.5) results in

$$0 = -\frac{1}{2}\log(4\pi)\varepsilon_3(t) + \frac{\log^2(\log(t))}{16\log(t)}(1+\mathrm{o}(1))$$

and therefore $\varepsilon_3(t) \sim \frac{\log^2(\log(t))}{8\log(4\pi)\log(t)}$.

Fourth term in the asymptotic expansion of $U(t)$: Write

$$U(t) = \sqrt{2\log(t)}\left(1 - \frac{\log\log(t)}{4\log(t)} - \frac{\log(4\pi)}{4\log(t)} - \frac{\log^2(\log(t))}{32\log^2(t)}(1+\varepsilon_4(t))\right),$$

with $\varepsilon_4(t) \to 0$ as $t \to \infty$. One has

$$\frac{(U(t))^2}{2} - \log(t) = -\frac{1}{2}\log\log(t) - \frac{1}{2}\log(4\pi) - \frac{\log^2(\log(t))}{16\log(t)}\varepsilon_4(t) + \frac{\log(4\pi)\log\log(t)}{8\log(t)}(1+\mathrm{o}(1))$$

and

$$\log(U(t)) = \frac{1}{2}\log\log(t) + \frac{1}{2}\log(2) - \frac{\log\log(t)}{4\log(t)}(1+\mathrm{o}(1)).$$

Plugging these two asymptotic expansions into (A.5) results in

$$0 = -\frac{\log^2(\log(t))}{16\log(t)}\varepsilon_4(t) + \frac{1}{4}\left(\frac{1}{2}\log(4\pi) - 1\right)\frac{\log\log(t)}{\log(t)}$$

and therefore $\varepsilon_4(t) \sim 4\left(\frac{1}{2}\log(4\pi) - 1\right)\frac{1}{\log\log(t)}$.

Fifth term in the asymptotic expansion of $U(t)$: Write finally

$$U(t) = \sqrt{2\log(t)}\left(1 - \frac{\log\log(t)}{4\log(t)} - \frac{\log(4\pi)}{4\log(t)} - \frac{\log^2(\log(t))}{32\log^2(t)} \right.$$
$$\left. -\frac{1}{8}\left(\frac{1}{2}\log(4\pi) - 1\right)\frac{\log\log(t)}{\log^2(t)}(1+\varepsilon_5(t))\right)$$

with $\varepsilon_5(t) \to 0$ as $t \to \infty$. One has

$$\frac{(U(t))^2}{2} - \log(t) = -\frac{1}{2}\log\log(t) - \frac{1}{2}\log(4\pi) - \frac{1}{4}\left(\frac{1}{2}\log(4\pi) - 1\right)\frac{\log\log(t)}{\log^2(t)}\varepsilon_5(t)$$
$$+ \frac{\log\log(t)}{4\log(t)} + \frac{\log^2(4\pi)}{16\log(t)}(1+\mathrm{o}(1))$$

and

$$\log(U(t)) = \frac{1}{2}\log\log(t) + \frac{1}{2}\log(2) - \frac{\log\log(t)}{4\log(t)} - \frac{\log(4\pi)}{4\log(t)}(1+\mathrm{o}(1)).$$

Plugging the two above asymptotic expansions along with the asymptotic equivalent $1/(U(t))^2 \sim 1/(2\log(t))$ into (A.5) yields

$$0 = -\frac{1}{4}\left(\frac{1}{2}\log(4\pi) - 1\right)\frac{\log\log(t)}{\log(t)}\varepsilon_5(t) + \frac{1}{2}\left(\frac{1}{8}\log^2(4\pi) - \frac{1}{2}\log(4\pi) + 1\right)\frac{1}{\log(t)}(1+\mathrm{o}(1))$$

and therefore

$$\frac{1}{8}\left(\frac{1}{2}\log(4\pi) - 1\right)\varepsilon_5(t) \sim \frac{1}{4}\left(\frac{1}{8}\log^2(4\pi) - \frac{1}{2}\log(4\pi) + 1\right)\frac{1}{\log\log(t)}.$$

This results in the desired asymptotic expansion of U.

To prove that U satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$, note that the asymptotic expansion we have just shown gives, for any $x > 0$,

$$\begin{aligned}
U(tx) &= \sqrt{2\log(t)}\left(1 + \frac{\log(x)}{2\log(t)} - \frac{\log^2(x)}{8\log^2(t)} + \mathrm{o}\left(\frac{1}{\log^2(t)}\right)\right) \\
&\quad\times\left(1 - \frac{\log\log(t)}{4\log(t)}\left\{1 - \frac{\log(x)}{\log(t)}\right\} - \frac{\log(x)}{4\log^2(t)} - \frac{\log(4\pi)}{4\log(t)}\left\{1 - \frac{\log(x)}{\log(t)}\right\} - \frac{\log^2(\log(t))}{32\log^2(t)}\right. \\
&\quad\left. -\frac{1}{8}\left(\frac{1}{2}\log(4\pi) - 1\right)\frac{\log\log(t)}{\log^2(t)} - \frac{1}{4}\left(\frac{1}{8}\log^2(4\pi) - \frac{1}{2}\log(4\pi) + 1\right)\frac{1}{\log^2(t)}(1+\mathrm{o}(1))\right) \\
&= U(t) + \sqrt{2\log(t)}\left(\frac{\log(x)}{2\log(t)} + \frac{\log\log(t)}{8\log^2(t)}\log(x) + \left(\frac{1}{2}\log(4\pi) - 1\right)\frac{\log(x)}{4\log^2(t)}\right. \\
&\quad\left. -\frac{\log^2(x)}{8\log^2(t)} + \mathrm{o}\left(\frac{1}{\log^2(t)}\right)\right)
\end{aligned}$$

as $t \to \infty$. Consequently

$$\begin{aligned}
U(tx) - U(t) = \frac{1}{\sqrt{2\log(t)}}&\left(\left\{1 + \frac{\log\log(t)}{4\log(t)} + \left(\frac{1}{2}\log(4\pi) - 1\right)\frac{1}{2\log(t)}\right\}\log(x)\right. \\
&\left. -\frac{1}{2\log(t)} \times \frac{1}{2}\log^2(x) + \mathrm{o}\left(\frac{1}{\log(t)}\right)\right)
\end{aligned}$$

from which the result follows immediately. □

We next introduce notation linked to the second-order condition $\mathcal{C}_2(\gamma, a, \rho, A)$ that shall be used several times in subsequent proofs. If this condition holds, define

$$a_\star(t) = \begin{cases} a(t)\left(1 - \dfrac{1}{\rho}A(t)\right), & \rho < 0, \\ a(t)\left(1 - \dfrac{1}{\gamma}A(t)\right), & \gamma \neq 0,\ \rho = 0, \\ a(t), & \gamma = 0,\ \rho = 0, \end{cases} \quad \text{and } A_\star(t) = \begin{cases} \dfrac{1}{\rho}A(t), & \rho < 0, \\ A(t), & \rho = 0 \end{cases} \tag{A.6}$$

as well as

$$\Psi_{\gamma,\rho}(x) = \begin{cases} \dfrac{x^{\gamma+\rho}-1}{\gamma+\rho}, & \gamma+\rho \neq 0,\ \rho<0, \\ \log(x), & \gamma+\rho=0,\ \rho<0, \\ \dfrac{1}{\gamma}x^{\gamma}\log(x), & \gamma\neq 0,\ \rho=0, \\ \dfrac{1}{2}\log^2(x), & \gamma=0,\ \rho=0. \end{cases} \tag{A.7}$$

Then, up to replacing $a_\star$ and $A_\star$ by suitable functions $a_{\star,0}$ and $A_{\star,0}$ such that $a_{\star,0}(t) = a_\star(t)(1+\mathrm{o}(|A(t)|))$ and $A_{\star,0}(t) = A_\star(t)(1+\mathrm{o}(1))$ as $t\to\infty$, Theorem 2.3.6 p.46 in de Haan and Ferreira (2006) guarantees, for any $\delta,\varepsilon>0$, that t, tx can be chosen so large that

$$\left| \frac{U(tx)-U(t)}{a_\star(t)} - \int_1^x s^{\gamma-1}\mathrm{d}s - A_\star(t)\Psi_{\gamma,\rho}(x) \right| \leq \varepsilon |A(t)| x^{\gamma+\rho}\max(x^\delta, x^{-\delta}). \tag{A.8}$$

Note also that $a_\star$ is asymptotically equivalent to a in a neighborhood of infinity and $a_\star(t)/U(t) - \gamma_+ = \mathrm{O}(|a(t)/U(t)-\gamma_+|) + \mathrm{O}(|A(t)|)$ as $t\to\infty$. We now have the material necessary to prove Proposition 1.

Proof of Proposition 1. (i) One has

$$\begin{aligned} \frac{U(tx)-U(t)}{a(t)} - \int_1^x s^{\gamma-1}\mathrm{d}s &= \frac{U(tx)-U(t)}{a_\star(t)} - \int_1^x s^{\gamma-1}\mathrm{d}s + \left(\frac{a_\star(t)}{a(t)}-1\right)\int_1^x s^{\gamma-1}\mathrm{d}s \\ &+ \left(\frac{a_\star(t)}{a(t)}-1\right)\left(\frac{U(tx)-U(t)}{a_\star(t)} - \int_1^x s^{\gamma-1}\mathrm{d}s\right). \end{aligned} \tag{A.9}$$

Fix $\delta,\varepsilon>0$ sufficiently small. Inequality (A.8) yields, for t large enough and all $x>1$,

$$\left| \left(\frac{a_\star(t)}{a(t)}-1\right)\left(\frac{U(tx)-U(t)}{a_\star(t)} - \int_1^x s^{\gamma-1}\mathrm{d}s\right) \right| \leq \frac{\varepsilon}{3}|A(t)|x^{\gamma_+ +\delta}. \tag{A.10}$$

Note also the relationship $a_\star(t)/a(t) - 1 = c(\gamma,\rho)A(t)(1+\mathrm{o}(|A(t)|))$ as $t\to\infty$, where $c(\gamma,\rho) = -1/\rho$ if $\rho<0$, $-1/\gamma$ if $\rho=0$ and $\gamma\neq 0$, and 0 if $\gamma=0$ and $\rho=0$. Then, for t large enough and all $x>1$,

$$\left| \left(\frac{a_\star(t)}{a(t)}-1\right)\int_1^x s^{\gamma-1}\mathrm{d}s - c(\gamma,\rho)A(t)\int_1^x s^{\gamma-1}\mathrm{d}s \right| \leq \frac{\varepsilon}{3}|A(t)|x^{\gamma_+ +\delta}. \tag{A.11}$$

Set finally $d(\rho) = 1/\rho$ if $\rho<0$ and 1 if $\rho=0$, and combine inequality (A.8) again with (A.9), (A.10) and (A.11) to obtain that for t large enough and all $x>1$,

$$\left| \frac{U(tx)-U(t)}{a(t)} - \int_1^x s^{\gamma-1}\mathrm{d}s - c(\gamma,\rho)A(t)\int_1^x s^{\gamma-1}\mathrm{d}s - d(\rho)A(t)\Psi_{\gamma,\rho}(x) \right| \leq \varepsilon|A(t)|x^{\gamma_+ +\delta}.$$

It is then sufficient to prove the identity

$$c(\gamma,\rho)\int_1^x s^{\gamma-1}\mathrm{d}s + d(\rho)\Psi_{\gamma,\rho}(x) = \int_1^x s^{\gamma-1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s. \tag{A.12}$$

We prove this identity by exhaustion. For this we recall that

$$\int_1^x s^{\gamma-1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s = \begin{cases} \dfrac{1}{\rho}\left(\dfrac{x^{\gamma+\rho}-1}{\gamma+\rho} - \dfrac{x^{\gamma}-1}{\gamma}\right) & \text{if } \gamma \neq 0,\ \rho < 0,\ \gamma+\rho \neq 0, \\ \dfrac{1}{\rho}\left(\log(x) - \dfrac{x^{\gamma}-1}{\gamma}\right) & \text{if } \gamma \neq 0,\ \rho < 0,\ \gamma+\rho = 0, \\ \dfrac{1}{\rho}\left(\dfrac{x^{\rho}-1}{\rho} - \log(x)\right) & \text{if } \gamma = 0,\ \rho < 0, \\ \dfrac{1}{\gamma}\left(x^{\gamma}\log(x) - \dfrac{x^{\gamma}-1}{\gamma}\right) & \text{if } \gamma \neq 0,\ \rho = 0, \\ \dfrac{1}{2}\log^2(x) & \text{if } \gamma = 0,\ \rho = 0. \end{cases}$$

If $\gamma \neq 0$, $\rho < 0$ and $\gamma + \rho \neq 0$, the left-hand side of (A.12) is

$$\frac{1}{\rho}\left(\frac{x^{\gamma+\rho}-1}{\gamma+\rho} - \frac{x^{\gamma}-1}{\gamma}\right) = \int_1^x s^{\gamma-1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s,$$

and if instead $\gamma + \rho = 0$, the left-hand side of (A.12) is

$$\frac{1}{\rho}\left(\log(x) - \frac{x^{\gamma}-1}{\gamma}\right) = \int_1^x s^{\gamma-1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s.$$

When $\gamma = 0$ and $\rho < 0$, the left-hand side of (A.12) is

$$\frac{1}{\rho}\left(\frac{x^{\rho}-1}{\rho} - \log(x)\right) = \int_1^x s^{\gamma-1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s.$$

If $\gamma \neq 0$ and $\rho = 0$, the left-hand side of (A.12) is

$$\frac{1}{\gamma}\left(x^{\gamma}\log(x) - \frac{x^{\gamma}-1}{\gamma}\right) = \int_1^x s^{\gamma-1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s,$$

and when $\gamma = 0$ and $\rho = 0$, it is

$$\frac{1}{2}\log^2(x) = \int_1^x s^{\gamma-1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s.$$

Equation (A.12) follows and the proof of (i) is complete.

(ii) Since X has the same distribution as $U(Y)$, where Y has a unit Pareto distribution (with distribution function $y \mapsto 1 - 1/y$, for $y > 1$), one may write

$$\mathrm{CTE}(\tau) - q(\tau) = \mathbb{E}(U(Y) - U(t) \,|\, U(Y) > U(t))$$

where throughout $t = t(\tau) = (1-\tau)^{-1} \to \infty$ as $\tau \uparrow 1$. Note also that condition $\mathcal{C}_2(\gamma, a, \rho, A)$ is equivalent to the following second-order condition on $\overline{F}$:

$$\lim_{s \uparrow q(1)} \frac{1}{A(1/\overline{F}(s))}\left(\frac{\overline{F}(s + x\, a(1/\overline{F}(s)))}{\overline{F}(s)} - Q_{\gamma}(x)\right) = G_{\gamma,\rho}(x), \quad \text{when } 1 + \gamma x > 0,$$

where Q_γ and $G_{\gamma,\rho}$ are functions that are continuous at 0 with $Q_\gamma(0) = 1$ and $G_{\gamma,\rho}(0) = 0$, see Theorem 2.3.8 p.48 in de Haan and Ferreira (2006). This equivalent second-order condition is known to be true locally uniformly in x, see Theorem B.3.19 p.401 in de Haan and Ferreira (2006). Taking $s = U(t)$ and $x = \varepsilon A(t)$ for $|\varepsilon|$ arbitrarily small, mimicking the first seven lines of the proof of this Theorem B.3.19 p.401 in de Haan and Ferreira (2006) leads to $t\overline{F}(U(t)) = 1 + \mathrm{o}(|A(t)|)$ and then to

$$\begin{aligned}\mathrm{CTE}(\tau) - q(\tau) &= \frac{\mathbb{E}((U(Y) - U(t))\mathbb{1}_{\{Y>1/\overline{F}(U(t))\}})}{\overline{F}(U(t))} \\ &= t\,\mathbb{E}((U(Y) - U(t))\mathbb{1}_{\{Y>1/\overline{F}(U(t))\}})(1 + \mathrm{o}(|A(t)|)).\end{aligned}$$

Moreover

$$\begin{aligned}&|\mathbb{E}((U(Y) - U(t))\mathbb{1}_{\{Y>1/\overline{F}(U(t))\}}) - \mathbb{E}((U(Y) - U(t))\mathbb{1}_{\{Y>t\}})| \\ &\leq \left|U(1/\overline{F}(U(t))) - U(t)\right| \times \frac{1}{t}|t\overline{F}(U(t)) - 1| = \mathrm{o}\left(\frac{a(t)}{t}|A(t)|\right),\end{aligned}$$

using the fact that the extended regular variation property satisfied by U holds in fact locally uniformly in $x > 0$, see Theorem B.2.18 p.383 in de Haan and Ferreira (2006), and taking $x = x(t) = 1/(t\overline{F}(U(t))) \to 1$ as $t \to \infty$. This entails

$$\begin{aligned}\mathrm{CTE}(\tau) - q(\tau) &= t\,\mathbb{E}((U(Y) - U(t))\mathbb{1}_{\{Y>t\}}) + \mathrm{o}(a(t)|A(t)|) \\ &= t\int_t^\infty (U(y) - U(t))\,\frac{\mathrm{d}y}{y^2} + \mathrm{o}(a(t)|A(t)|).\end{aligned}$$

Since

$$t\int_t^\infty (U(y) - U(t))\,\frac{\mathrm{d}y}{y^2} = \frac{1}{1-\tau}\int_\tau^1 (q(s) - q(\tau))\mathrm{d}s = \mathrm{ES}(\tau) - q(\tau)$$

this shows that

$$\frac{\mathrm{CTE}(\tau) - q(\tau)}{a((1-\tau)^{-1})} = \frac{\mathrm{ES}(\tau) - q(\tau)}{a((1-\tau)^{-1})} + \mathrm{o}(|A((1-\tau)^{-1})|).$$

The change of variables $y = tx$ then yields

$$\frac{\mathrm{ES}(\tau) - q(\tau)}{a((1-\tau)^{-1})} = \int_1^\infty \frac{U(tx) - U(t)}{a(t)}\,\frac{\mathrm{d}x}{x^2} + \mathrm{o}(|A(t)|).$$

Combining (i) with condition $\gamma < 1$, we obtain

$$\begin{aligned}\frac{\mathrm{ES}(\tau) - q(\tau)}{a((1-\tau)^{-1})} &= \int_1^\infty \left(\int_1^x s^{\gamma-1}\mathrm{d}s\right)\frac{\mathrm{d}x}{x^2} \\ &\quad + A(t)\int_1^\infty \left\{\int_1^x s^{\gamma-1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s\right\}\frac{\mathrm{d}x}{x^2} + \mathrm{o}(|A(t)|).\end{aligned}$$

Use finally the identity

$$\int_1^\infty \left\{\int_1^x s^{b-1}\left(\int_1^s u^{c-1}\mathrm{d}u\right)\mathrm{d}s\right\}\frac{\mathrm{d}x}{x^2} = \frac{1}{(1-b)(1-b-c)}$$

valid for $b < 1$ and $b + c < 1$ to conclude the proof of (ii).

(iii) This is a consequence of (ii) and the convergence $a(t)/U(t) \to \gamma_+ = \gamma$ as $t \to \infty$. $\square$

Throughout this appendix, for $1 < \alpha < 2$, we denote by Z_α a unit, right-skewed stable random variable having Fourier transform

$$\mathbb{E}\left(e^{itZ_\alpha}\right) = \exp\left(-|t|^\alpha \left\{1 - i \tan\left(\frac{\pi\alpha}{2}\right) \operatorname{sign}(t)\right\}\right).$$

We also recall the notation $\Gamma : x \in (0, \infty) \mapsto \Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \mathrm{d}t$ for Euler's Gamma function.

The following lemma is a central limit theorem for sums of Pareto random variables. It is the fundamental tool for the proof of Proposition A.1 below.

Lemma A.4 (Limiting behavior of sample averages of Pareto random variables)**.** *Let (Y_i) be a sequence of independent unit Pareto random variables.*

- *[Case $\gamma < 1/2$, finite second moment] When $\gamma < 1/2$, we have*

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^k Y_i^\gamma - \frac{1}{1-\gamma}\right) \xrightarrow{\mathrm{d}} N\left(0, \frac{\gamma^2}{(1-\gamma)^2(1-2\gamma)}\right).$$

- *[Case $\gamma = 1/2$, infinite second moment, phase transition] We have*

$$\frac{\sqrt{k}}{\sqrt{\log k}}\left(\frac{1}{k}\sum_{i=1}^k \sqrt{Y_i} - 2\right) \xrightarrow{\mathrm{d}} N(0,1).$$

- *[Case $1/2 < \gamma < 1$, infinite second moment] When $1/2 < \gamma < 1$, we have*

$$k^{1-\gamma}\left(\frac{1}{k}\sum_{i=1}^k Y_i^\gamma - \frac{1}{1-\gamma}\right) \xrightarrow{\mathrm{d}} \left\{-\frac{\Gamma(2-1/\gamma)}{1/\gamma - 1}\cos\left(\frac{\pi}{2\gamma}\right)\right\}^\gamma Z_{1/\gamma}.$$

Proof of Lemma A.4. First and foremost, for any $\alpha < 1$, $\mathbb{E}(Y^\alpha) = 1/(1-\alpha)$. The first convergence is then an obvious consequence of the standard central limit theorem.

To show the second and third convergences, we follow Section XVII.5 of Feller (1971), which provides a unified treatment of the two cases. Equivalent results can be obtained by applying Theorems 1 and 2 in Geluk and de Haan (2000). Let $\mu = \mu_\gamma$ be the truncated second moment function of Y^γ, defined by

$$\forall x > 1,\ \mu(x) := \int_{-x}^x y^2 \mathbb{P}(Y^\gamma \in \mathrm{d}y) = \frac{1}{\gamma}\int_1^x y^{1-1/\gamma}\mathrm{d}y$$
$$= \begin{cases} \dfrac{1}{2\gamma-1}(x^{2-1/\gamma} - 1) & \text{if } \gamma \in (1/2, 1), \\ 2\log x & \text{if } \gamma = 1/2. \end{cases}$$

Then the sequence

$$a_k := \begin{cases} (2\gamma-1)^{-\gamma} k^\gamma & \text{if } \gamma \in (1/2,1) \\ \sqrt{k \log k} & \text{if } \gamma = 1/2 \end{cases}$$

clearly satisfies $a_k \to \infty$ and $k\mu(a_k)/a_k^2 \to 1$. Furthermore, the random variable Y^γ is such that

$$\forall y > 1,\ \mathbb{P}(Y^\gamma \le -y) = 0 \ \text{ and } \ \mathbb{P}(Y^\gamma > y) = y^{-1/\gamma}.$$

The random variable Y^γ therefore satisfies the conditions of Theorem 2 in Section XVII.5 of Feller (1971); note that its balanced tails condition (5.18) implicitly allows that one tail is regularly varying and dominates the other, as can be seen by comparing this condition with the setup of (3.17) and (3.18) in Section XVII.3 of Feller (1971). By Theorem 3 in Section XVII.5 therein,

$$\frac{1}{a_k}\left(\sum_{i=1}^k \{Y_i^\gamma - \mathbb{E}(Y^\gamma)\}\right) \xrightarrow{\text{d}} W_{1/\gamma}, \tag{A.13}$$

where $W_{1/\gamma}$ has a stable distribution with Fourier transform

$$\mathbb{E}\left(e^{itW_{1/\gamma}}\right) = \exp\left(|t|^{1/\gamma}\frac{\Gamma(3-1/\gamma)}{(1/\gamma)(1/\gamma-1)}\left\{\cos\left(\frac{\pi}{2\gamma}\right) - i\sin\left(\frac{\pi}{2\gamma}\right)\text{sign}(t)\right\}\right).$$

[Note the typo below (3.19) in Section XVII.3 of Feller (1971), which should read "here the lower sign applies when $\zeta > 0$, the upper for $\zeta < 0$", as can be seen by comparing (3.19) with (3.11) therein.] The result is then immediate for $\gamma = 1/2$. When $\gamma \in (1/2, 1)$, the above Fourier transform is also

$$\mathbb{E}\left(e^{itW_{1/\gamma}}\right) = \exp\left(-|t|^{1/\gamma}\left\{-(2\gamma-1)\frac{\Gamma(2-1/\gamma)}{1/\gamma-1}\cos\left(\frac{\pi}{2\gamma}\right)\right\}\left\{1 - i\tan\left(\frac{\pi}{2\gamma}\right)\text{sign}(t)\right\}\right).$$

Rephrase then convergence (A.13) as

$$k^{1-\gamma}\left(\frac{1}{k}\sum_{i=1}^k Y_i^\gamma - \frac{1}{1-\gamma}\right) \xrightarrow{\text{d}} (2\gamma-1)^{-\gamma} W_{1/\gamma} \overset{\text{d}}{=} \left\{-\frac{\Gamma(2-1/\gamma)}{1/\gamma-1}\cos\left(\frac{\pi}{2\gamma}\right)\right\}^\gamma Z_{1/\gamma} \tag{A.14}$$

to conclude the proof. □

Proposition A.1 below gives the joint asymptotic behavior of the empirical counterpart of the Expected Shortfall and the empirical quantile at intermediate levels. It is the key to the proof of Theorem 1. Before proceeding to the statement and proof of Proposition A.1 it will be useful to recall a couple of results linked to Rényi's representation of order statistics from an independent unit exponential sample (see p.37 in de Haan and Ferreira, 2006), as this will be an important tool in some of our proofs below. Let $(E_1, \ldots, E_n)$ be independent unit exponential random variables: one has

$$(E_{i:n})_{1\le i\le n} \stackrel{\mathrm{d}}{=} \left(\sum_{j=1}^{i} \frac{E_j}{n-j+1}\right)_{1\le i\le n}.$$

Consequently

$$((E_{n-i+1:n} - E_{n-k:n})_{1\le i\le k}, E_{n-k:n}) \stackrel{\mathrm{d}}{=} \left(\left(\sum_{j=n-k+1}^{n-i+1} \frac{E_j}{n-j+1}\right)_{1\le i\le k}, \sum_{j=1}^{n-k} \frac{E_j}{n-j+1}\right). \tag{A.15}$$

It follows in particular that $(E_{n-i+1:n} - E_{n-k:n})_{1\le i\le k} \stackrel{\mathrm{d}}{=} (E_{k-i+1:k})_{1\le i\le k}$ and thus that, for any measurable function f on $(0,\infty)$,

$$\frac{1}{k}\sum_{i=1}^{k} f(E_{n-i+1:n} - E_{n-k:n}) \stackrel{\mathrm{d}}{=} \frac{1}{k}\sum_{i=1}^{k} f(E_i) \quad \text{and} \quad \frac{1}{k}\sum_{i=1}^{k} f(E_{n-i+1:n} - E_{n-k:n}) \perp\!\!\!\perp E_{n-k:n}.$$

This can be equivalently rephrased in the following way: if $(Y_1, \ldots, Y_n)$ are independent unit Pareto random variables, then for any measurable function f on $(1,\infty)$,

$$\frac{1}{k}\sum_{i=1}^{k} f(Y_{n-i+1:n}/Y_{n-k:n}) \stackrel{\mathrm{d}}{=} \frac{1}{k}\sum_{i=1}^{k} f(Y_i) \quad \text{and} \quad \frac{1}{k}\sum_{i=1}^{k} f(Y_{n-i+1:n}/Y_{n-k:n}) \perp\!\!\!\perp Y_{n-k:n}. \tag{A.16}$$

[This follows from the fact that if Y is unit Pareto, then $\log Y$ is unit exponential.] A particular consequence of that is the law of large numbers

$$\frac{1}{k}\sum_{i=1}^{k} f(Y_{n-i+1:n}/Y_{n-k:n}) \xrightarrow{\mathbb{P}} \mathbb{E}(f(Y)) = \int_1^\infty f(x)\frac{\mathrm{d}x}{x^2} \tag{A.17}$$

valid as soon as $x \mapsto x^{-2} f(x) \mathbb{1}_{\{x\ge 1\}}$ is integrable. This convergence will mostly be used with the following choices for f:

- For $f(x) = x^a\{\int_1^x s^{b-1}\mathrm{d}s\}$ with $a < 1$ and $a + b < 1$, yielding

$$\mathbb{E}\left(Y^a\left\{\int_1^Y s^{b-1}\mathrm{d}s\right\}\right) = \frac{1}{(1-a)(1-a-b)}. \tag{A.18}$$

- For $f(x) = x^a\{\int_1^x s^{b-1}(\int_1^s u^{c-1}\mathrm{d}u)\mathrm{d}s\}$ with $a < 1$, $a + b < 1$ and $a + b + c < 1$, yielding

$$\mathbb{E}\left(Y^a\left\{\int_1^Y s^{b-1}\left(\int_1^s u^{c-1}\mathrm{d}u\right)\mathrm{d}s\right\}\right) = \frac{1}{(1-a)(1-a-b)(1-a-b-c)}. \tag{A.19}$$

- For $f(x) = x^a\{\int_1^x s^{b-1}(\int_1^s u^{c-1}\mathrm{d}u)\mathrm{d}s\}(\int_1^x s^{d-1}\mathrm{d}s)$ with $a < 1$, $a+b < 1$, $a+b+c < 1$, $a+d < 1$, $a+b+d < 1$ and $a+b+c+d < 1$, yielding

$$\mathbb{E}\left(Y^a\left\{\int_1^Y s^{b-1}\left(\int_1^s u^{c-1}\mathrm{d}u\right)\mathrm{d}s\right\}\left(\int_1^Y s^{d-1}\mathrm{d}s\right)\right)$$
$$= \frac{(1-a-b-d)(1-a-b-c-d)+(1-a)(2-2a-2b-c-d)}{(1-a)(1-a-d)(1-a-b)(1-a-b-d)(1-a-b-c)(1-a-b-c-d)}. \quad \text{(A.20)}$$

We also note the following two lemmas (possibly known elsewhere) used in the proof of Proposition A.1. We provide concise proofs of these auxiliary results for the sake of completeness.

Lemma A.5. *Let f be a measurable positive function on $(0,\infty)$. Assume that there are $\alpha \in \mathbb{R}$, $\beta \le 0$ and a measurable function C having constant sign and converging to 0 at infinity such that the following property holds: for any $\delta, \varepsilon > 0$, there is $t_0 > 0$ such that for $t, tx \ge t_0$,*

$$\left|\frac{f(tx)}{f(t)} - x^\alpha - C(t)x^\alpha \int_1^x s^{\beta-1}\mathrm{d}s\right| \le \varepsilon|C(t)|x^{\alpha+\beta}\max(x^\delta, x^{-\delta}).$$

Then if $(Y_1,\ldots,Y_n)$ are independent unit Pareto random variables, and if $k = k(n)$ is a sequence of integers with $k\to\infty$ and $k/n \to 0$, one has

$$\frac{f(Y_{n-k:n})}{f(n/k)} - \left(\frac{k}{n}Y_{n-k:n}\right)^\alpha = \mathrm{o}_{\mathbb{P}}(|C(n/k)|).$$

Proof of Lemma A.5. It follows from Corollary 2.2.2 p.41 in de Haan and Ferreira (2006) that $\frac{k}{n}Y_{n-k:n} = 1 + \mathrm{o}_{\mathbb{P}}(1)$, and in particular that $Y_{n-k:n} \xrightarrow{\mathbb{P}} \infty$. Use the assumption on f with $t = n/k \to \infty$ and $tx = Y_{n-k:n}$ to obtain, for any $\delta > 0$,

$$\frac{f(Y_{n-k:n})}{f(n/k)} - \left(\frac{k}{n}Y_{n-k:n}\right)^\alpha = C(n/k)\left(\frac{k}{n}Y_{n-k:n}\right)^\alpha \int_1^{\frac{k}{n}Y_{n-k:n}} s^{\beta-1}\mathrm{d}s$$
$$+ \mathrm{o}_{\mathbb{P}}\left(|C(n/k)|\left(\frac{k}{n}Y_{n-k:n}\right)^{\alpha+\beta\pm\delta}\right).$$

Use again the fact that $\frac{k}{n}Y_{n-k:n} = 1 + \mathrm{o}_{\mathbb{P}}(1)$ to complete the proof. □

Lemma A.6 (Asymptotic (random) inversion lemma)**.** *Let $X_1,\ldots,X_n,\ldots$ be independent and identically distributed random copies of a random variable X satisfying condition $\mathcal{C}_2(\gamma,a,\rho,A)$. Let $k = k(n) \to \infty$ be a sequence of integers such that $k/n \to 0$ and $\sqrt{k}A(n/k) = \mathrm{O}(1)$. Define*

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \le x\}} \text{ and } \widehat{\overline{F}}_n(x) = 1 - \widehat{F}_n(x).$$

Then

$$\sqrt{k}\left(\frac{n}{k}\widehat{\overline{F}}_n(X_{n-k:n}) - 1\right) \xrightarrow{\mathbb{P}} 0.$$

[This sample analog of the identity $\overline{F}(q(\tau))/(1-\tau) = 1 + \mathrm{o}(|A((1-\tau)^{-1})|)$ (see the proof of Proposition 1(ii)) is obvious if F is continuous, since then $\widehat{\overline{F}}_n(X_{n-k:n}) = k/n$.]

Proof of Lemma A.6. Fix $t > 0$ and write

$$\sqrt{k}\left|\frac{n}{k}\widehat{\overline{F}}_n(X_{n-k:n}) - 1\right| \geq t \Leftrightarrow \sqrt{k}\left(\frac{n}{k}\widehat{\overline{F}}_n(X_{n-k:n}) - 1\right) \leq -t$$

because $\widehat{\overline{F}}_n(X_{n-k:n}) \leq k/n$. Then clearly

$$\sqrt{k}\left|\frac{n}{k}\widehat{\overline{F}}_n(X_{n-k:n}) - 1\right| \geq t \Leftrightarrow 1 - \frac{k}{n}\left(1 - \frac{t}{\sqrt{k}}\right) \leq \widehat{F}_n(X_{n-k:n}).$$

Since $\alpha \mapsto X_{\lceil n\alpha \rceil:n}$ is the left-continuous inverse of $\widehat{F}_n$, one has $\alpha \leq \widehat{F}_n(x) \Leftrightarrow X_{\lceil n\alpha \rceil:n} \leq x$ for any $\alpha \in (0,1)$ and $x \in \mathbb{R}$. We obtain

$$\sqrt{k}\left|\frac{n}{k}\widehat{\overline{F}}_n(X_{n-k:n}) - 1\right| \geq t \Leftrightarrow X_{n-\lfloor k(1-t/\sqrt{k})\rfloor:n} \leq X_{n-k:n}$$

or equivalently

$$\mathbb{P}\left(\sqrt{k}\left|\frac{n}{k}\widehat{\overline{F}}_n(X_{n-k:n}) - 1\right| \geq t\right) = \mathbb{P}\left(\sqrt{k}\frac{X_{n-k:n} - X_{n-\lfloor k(1-t/\sqrt{k})\rfloor:n}}{a_0(n/k)} \geq 0\right)$$

where a_0 is a function that is asymptotically equivalent to a in a neighborhood of infinity and such that Corollary 2.4.6 p.52 in de Haan and Ferreira (2006) applies. Using this theorem, one finds

$$\sqrt{k}\left(\frac{X_{n-k:n} - X_{n-\lfloor k(1-t/\sqrt{k})\rfloor:n}}{a_0(n/k)} + \int_1^{(1-t/\sqrt{k})^{-1}} z^{\gamma-1}\mathrm{d}z\right) \xrightarrow{\mathbb{P}} 0.$$

Consequently

$$\sqrt{k}\frac{X_{n-k:n} - X_{n-\lfloor k(1-t/\sqrt{k})\rfloor:n}}{a_0(n/k)} \xrightarrow{\mathbb{P}} -t < 0$$

and then

$$\mathbb{P}\left(\sqrt{k}\left|\frac{n}{k}\widehat{\overline{F}}_n(X_{n-k:n}) - 1\right| \geq t\right) = \mathbb{P}\left(\sqrt{k}\frac{X_{n-k:n} - X_{n-\lfloor k(1-t/\sqrt{k})\rfloor:n}}{a_0(n/k)} + t \geq t\right) \to 0$$

as required. □

Let here and throughout $X_1, \ldots, X_n$ be independent and identically distributed random copies of a random variable X satisfying condition $\mathcal{C}_2(\gamma, a, \rho, A)$. We are now ready to state and prove Proposition A.1 and, in passing, that the estimator $\widehat{\mathrm{ES}}_n(\tau_n)$ is asymptotically equivalent to the two, *a priori* different estimators

$$\widehat{\mathrm{ES}}_n^{(1)}(\tau_n) = \frac{\sum_{i=1}^n X_i \mathbb{1}_{\{X_i > X_{\lceil n\tau_n \rceil:n}\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > X_{\lceil n\tau_n \rceil:n}\}}} \text{ and } \widehat{\mathrm{ES}}_n^{(2)}(\tau_n) = \frac{1}{1-\tau_n}\int_{\tau_n}^1 \widehat{q}_n(t)\mathrm{d}t$$

when τ_n is intermediate.

Proposition A.1 (Joint weak convergence of $\widehat{\mathrm{ES}}_n(\tau_n)$ and the empirical quantile)**.** *Suppose that X satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$ with $\gamma < 1$. Let $\tau_n \uparrow 1$ with $n(1-\tau_n) \to \infty$. Let also*

$$v_n(\gamma) = \begin{cases} \sqrt{n(1-\tau_n)} & \text{if } 0 < \gamma < 1/2, \\ \sqrt{n(1-\tau_n)}/\sqrt{\log(n(1-\tau_n))} & \text{if } \gamma = 1/2, \\ (n(1-\tau_n))^{1-\gamma} & \text{if } 1/2 < \gamma < 1. \end{cases}$$

Then, if $v_n(\gamma) A((1-\tau_n)^{-1}) = \mathrm{O}(1)$, one has

$$v_n(\gamma) \left(\frac{\widehat{\mathrm{ES}}_n(\tau_n) - X_{\lceil n\tau_n \rceil:n}}{a((1-\tau_n)^{-1})} - \frac{1}{1-\gamma} - \frac{1}{(1-\gamma)(1-\gamma-\rho)} A((1-\tau_n)^{-1}),\ \frac{X_{\lceil n\tau_n \rceil:n} - q(\tau_n)}{a((1-\tau_n)^{-1})} \right)$$

$$\xrightarrow{\mathrm{d}} \begin{cases} \mathcal{N}\left(\mathbf{0}_2, \begin{pmatrix} \dfrac{1+\gamma+2\gamma^2}{(1-\gamma)(1-2\gamma)} & \dfrac{\gamma}{1-\gamma} \\ \dfrac{\gamma}{1-\gamma} & 1 \end{pmatrix} \right) & \text{if } \gamma < 1/2, \\ (\mathcal{N}(0,4),\ 0) & \text{if } \gamma = 1/2, \\ \left(\dfrac{1}{\gamma} \left\{ -\dfrac{\Gamma(2-1/\gamma)}{1/\gamma - 1} \cos\left(\dfrac{\pi}{2\gamma} \right) \right\}^{\gamma} Z_{1/\gamma},\ 0 \right) & \text{if } 1/2 < \gamma < 1. \end{cases}$$

If moreover $\sqrt{n(1-\tau_n)} A((1-\tau_n)^{-1}) = \mathrm{O}(1)$, then

$$\begin{aligned} v_n(\gamma) \left(\frac{\widehat{\mathrm{ES}}_n(\tau_n) - X_{\lceil n\tau_n \rceil:n}}{a((1-\tau_n)^{-1})} - \frac{1}{1-\gamma} \right) &= v_n(\gamma) \left(\frac{\widehat{\mathrm{ES}}_n^{(1)}(\tau_n) - X_{\lceil n\tau_n \rceil:n}}{a((1-\tau_n)^{-1})} - \frac{1}{1-\gamma} \right) + \mathrm{o}_{\mathbb{P}}(1) \\ &= v_n(\gamma) \left(\frac{\widehat{\mathrm{ES}}_n^{(2)}(\tau_n) - X_{\lceil n\tau_n \rceil:n}}{a((1-\tau_n)^{-1})} - \frac{1}{1-\gamma} \right) + \mathrm{o}_{\mathbb{P}}(1) \end{aligned}$$

and the above bivariate convergence result holds with $\widehat{\mathrm{ES}}_n$ replaced by either $\widehat{\mathrm{ES}}_n^{(1)}$ or $\widehat{\mathrm{ES}}_n^{(2)}$.

Proof of Proposition A.1. Set $k = n - \lceil n\tau_n \rceil = \lfloor n(1-\tau_n) \rfloor$, so that $k = k(n)$ is a sequence of integers asymptotically equivalent to $n(1-\tau_n)$; in particular, $k \to \infty$ and $k/n \to 0$. Note that $X_{\lceil n\tau_n \rceil:n} = X_{n-k:n}$, and therefore $\widehat{\mathrm{ES}}_n(\tau_n) = \widehat{\mathrm{ES}}_n(1-k/n)$. Besides, a combination of Theorem 2.3.3 p.44 in de Haan and Ferreira (2006) with Theorem B.2.18 p.383 in de Haan and Ferreira (2006) applied to the function $t \mapsto t^{-\gamma} a(t)$ yields, for any $\delta, \varepsilon > 0$ sufficiently small, that for t, tx large enough,

$$\left| \frac{a(tx)}{a(t)} - x^{\gamma} - A_0(t) x^{\gamma} \int_1^x s^{\rho-1} \mathrm{d}s \right| \leq \varepsilon |A_0(t)| x^{\gamma+\rho} \max(x^{\delta}, x^{-\delta}) \tag{A.21}$$

up to replacing the function A by a suitable, asymptotically equivalent function A_0. Since clearly

$$\frac{n(1-\tau_n)}{k} = 1 + \mathrm{O}\left(\frac{1}{n(1-\tau_n)} \right),$$

it follows from Proposition 1(i) that

$$\frac{q(\tau_n)-q(1-k/n)}{a((1-\tau_n)^{-1})} = \mathrm{O}\left(\frac{1}{n(1-\tau_n)}\right) + \mathrm{o}(|A((1-\tau_n)^{-1})|) = \mathrm{o}\left(\frac{1}{v_n(\gamma)}\right), \tag{A.22}$$

and from (A.21) that

$$\frac{a((1-\tau_n)^{-1})}{a(n/k)} - 1 = \mathrm{O}\left(\frac{1}{n(1-\tau_n)}\right) + \mathrm{o}(|A((1-\tau_n)^{-1})|) = \mathrm{o}\left(\frac{1}{v_n(\gamma)}\right). \tag{A.23}$$

We split the proof into two parts.

Step 1: Convergence properties of $\widehat{\mathrm{ES}}_n(\tau_n)$

From Equations (A.22) and (A.23) and the convergence $A((1-\tau_n)^{-1})/A(n/k) \to 1$ (coming as a consequence of the regular variation property of A) it follows that it suffices to work in the setting $\tau_n = 1-k/n$ with $k=k(n)$ a sequence of integers such that $k\to\infty$ and $k/n\to 0$, in which case

$$\widehat{\mathrm{ES}}_n(1-k/n) - X_{n-k:n} = \frac{1}{k}\sum_{i=1}^{k}(X_{n-i+1:n} - X_{n-k:n}).$$

With this in mind, we consider two cases.

Case $\gamma < 1/2$: Recall the notation of (A.6) and (A.7), and use the versions of $a_\star$ and $A_\star$ such that (A.8) holds. According to Theorem 2.4.2 and Corollary 2.4.6 pp.51-52 in de Haan and Ferreira (2006), up to enlarging the underlying probability space, one can construct a sequence of standard Brownian motions (W_n) such that for any $\varepsilon > 0$,

$$\sup_{0<s\le 1} \min(1, s^{\gamma+1/2+\varepsilon})\left|\sqrt{k}\left(\frac{X_{n-\lfloor ks\rfloor:n} - X_{n-k:n}}{a_\star(n/k)} - \int_1^{1/s} z^{\gamma-1}\mathrm{d}z\right)\right.$$
$$\left. - (s^{-\gamma-1}W_n(s) - W_n(1)) - \sqrt{k}A_\star(n/k)\Psi_{\gamma,\rho}(s^{-1})\right| \xrightarrow{\mathbb{P}} 0$$
$$\text{and } \left|\sqrt{k}\frac{X_{n-k:n} - q(1-k/n)}{a_\star(n/k)} - W_n(1)\right| \xrightarrow{\mathbb{P}} 0.$$

Then

$$\sqrt{k}\frac{X_{n-k:n} - q(1-k/n)}{a(n/k)} = W_n(1) + \mathrm{o}_{\mathbb{P}}(1)$$

because $a_\star$ and a are asymptotically equivalent in a neighborhood of infinity. Also

$$\int_0^1 \left(\sqrt{k}\left(\frac{X_{n-\lfloor ks\rfloor:n} - X_{n-k:n}}{a_\star(n/k)} - \int_1^{1/s} z^{\gamma-1}\mathrm{d}z\right) - \sqrt{k}A_\star(n/k)\Psi_{\gamma,\rho}(s^{-1})\right)\mathrm{d}s$$
$$= \int_0^1 (s^{-\gamma-1}W_n(s) - W_n(1))\mathrm{d}s + \mathrm{o}_{\mathbb{P}}(1).$$

In particular

$$\frac{1}{k}\sum_{i=1}^{k}\frac{X_{n-i+1:n}-X_{n-k:n}}{a_\star(n/k)}=\int_0^1\frac{X_{n-\lfloor ks\rfloor:n}-X_{n-k:n}}{a_\star(n/k)}\mathrm{d}s=\frac{1}{1-\gamma}+\mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right). \tag{A.24}$$

Let, as in the proof of Proposition 1(i), $c(\gamma,\rho)=-1/\rho$ if $\rho<0$, $-1/\gamma$ if $\rho=0$ and $\gamma\neq 0$, and 0 if $\gamma=0$ and $\rho=0$, as well as $d(\rho)=1/\rho$ if $\rho<0$ and 1 if $\rho=0$, so that $a_\star(t)=a(t)(1+c(\gamma,\rho)A(t)(1+\mathrm{o}(1)))$ and $A_\star(t)=d(\rho)A(t)(1+\mathrm{o}(1))$. Combining (A.12), (A.18) and (A.19), one obtains

$$\begin{aligned}\frac{c(\gamma,\rho)}{1-\gamma}+d(\rho)\int_0^1\Psi_{\gamma,\rho}(s^{-1})\mathrm{d}s&=\int_1^\infty\left(c(\gamma,\rho)\int_1^x s^{\gamma-1}\mathrm{d}s+d(\rho)\Psi_{\gamma,\rho}(x)\right)\frac{\mathrm{d}x}{x^2}\\&=\int_1^\infty\left\{\int_1^x s^{\gamma-1}\left(\int_1^s u^{\rho-1}\right)\mathrm{d}s\right\}\frac{\mathrm{d}x}{x^2}\\&=\frac{1}{(1-\gamma)(1-\gamma-\rho)}.\end{aligned} \tag{A.25}$$

Combining (A.24) and (A.25) results in

$$\begin{aligned}&\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k}\frac{X_{n-i+1:n}-X_{n-k:n}}{a(n/k)}-\frac{1}{1-\gamma}-\frac{1}{(1-\gamma)(1-\gamma-\rho)}A(n/k)\right)\\&=\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k}\frac{X_{n-i+1:n}-X_{n-k:n}}{a_\star(n/k)}-\frac{1}{1-\gamma}\right)(1+\mathrm{o}_{\mathbb{P}}(1))\\&+\frac{c(\gamma,\rho)}{1-\gamma}\sqrt{k}A(n/k)-\frac{1}{(1-\gamma)(1-\gamma-\rho)}\sqrt{k}A(n/k)+\mathrm{o}_{\mathbb{P}}(1)\\&=\int_0^1\left(\sqrt{k}\left(\frac{X_{n-\lfloor ks\rfloor:n}-X_{n-k:n}}{a_\star(n/k)}-\int_1^{1/s}z^{\gamma-1}\mathrm{d}z\right)-\sqrt{k}A_\star(n/k)\Psi_{\gamma,\rho}(s^{-1})\right)\mathrm{d}s+\mathrm{o}_{\mathbb{P}}(1)\\&=\int_0^1(s^{-\gamma-1}W_n(s)-W_n(1))\mathrm{d}s+\mathrm{o}_{\mathbb{P}}(1).\end{aligned}$$

Since $\mathrm{Cov}(W_n(s),W_n(t))=\min(s,t)$ for any $s,t>0$, it is readily checked that

$$\left(\int_0^1(s^{-\gamma-1}W_n(s)-W_n(1))\mathrm{d}s,\ W_n(1)\right)\stackrel{\mathrm{d}}{=}\mathcal{N}\left(\mathbf{0}_2,\begin{pmatrix}\dfrac{1+\gamma+2\gamma^2}{(1-\gamma)(1-2\gamma)} & \dfrac{\gamma}{1-\gamma}\\ \dfrac{\gamma}{1-\gamma} & 1\end{pmatrix}\right).$$

The conclusion follows in this case.

Case $1/2\le\gamma<1$: Let (Y_i) be a sequence of independent unit Pareto random variables. Recall from Corollary 2.2.2 p.41 in de Haan and Ferreira (2006) that $\frac{k}{n}Y_{n-k:n}=1+\mathrm{O}_{\mathbb{P}}(1/\sqrt{k})$, and in particular that $Y_{n-k:n}\stackrel{\mathbb{P}}{\longrightarrow}\infty$. Apply Proposition 1(i) to get

$$\frac{X_{n-k:n}-q(1-k/n)}{a(n/k)}\stackrel{\mathrm{d}}{=}\frac{U(Y_{n-k:n})-U(n/k)}{a(n/k)}=\frac{(\frac{k}{n}Y_{n-k:n})^\gamma-1}{\gamma}+\mathrm{o}_{\mathbb{P}}(|A(n/k)|). \tag{A.26}$$

By Corollary 2.2.2 p.41 in de Haan and Ferreira (2006) again, we obtain

$$v_n(\gamma)\frac{X_{n-k:n}-q(1-k/n)}{a(n/k)} = \mathrm{O}_{\mathbb{P}}\left(\frac{v_n(\gamma)}{\sqrt{k}}\right) + \mathrm{o}_{\mathbb{P}}(v_n(\gamma)|A(n/k)|) = \mathrm{o}_{\mathbb{P}}(1).$$

Besides,

$$\frac{1}{k}\sum_{i=1}^{k}\frac{X_{n-i+1:n}-X_{n-k:n}}{a(n/k)} \stackrel{\mathrm{d}}{=} \frac{a(Y_{n-k:n})}{a(n/k)} \times \frac{1}{k}\sum_{i=1}^{k}\frac{U(Y_{n-i+1:n})-U(Y_{n-k:n})}{a(Y_{n-k:n})}. \tag{A.27}$$

Pick $\delta > 0$ such that $\gamma+\delta<1$. By Proposition 1(i) again,

$$\begin{aligned}\frac{1}{k}\sum_{i=1}^{k}\frac{U(Y_{n-i+1:n})-U(Y_{n-k:n})}{a(Y_{n-k:n})} &= \frac{1}{k}\sum_{i=1}^{k}\frac{(Y_{n-i+1:n}/Y_{n-k:n})^{\gamma}-1}{\gamma} \\ &\quad + A(Y_{n-k:n})\times\frac{1}{k}\sum_{i=1}^{k}\int_1^{Y_{n-i+1:n}/Y_{n-k:n}} s^{\gamma-1}\left(\int_1^s u^{\rho-1}\right)\mathrm{d}s \\ &\quad + \mathrm{o}_{\mathbb{P}}(|A(Y_{n-k:n})|)\times\frac{1}{k}\sum_{i=1}^{k}\left(\frac{Y_{n-i+1:n}}{Y_{n-k:n}}\right)^{\gamma+\delta}. \end{aligned}\tag{A.28}$$

Using Lemma A.5 applied to the function a (which, by (A.21), satisfies the uniform inequality required in that Lemma), the fact that $\frac{k}{n}Y_{n-k:n} = 1+\mathrm{O}_{\mathbb{P}}(1/\sqrt{k})$, and the regular variation property of A, we obtain

$$\frac{a(Y_{n-k:n})}{a(n/k)} = 1+\mathrm{o}_{\mathbb{P}}\left(\frac{1}{v_n(\gamma)}\right) \text{ and } v_n(\gamma)A(Y_{n-k:n}) = v_n(\gamma)A(n/k)(1+\mathrm{o}_{\mathbb{P}}(1)) = \mathrm{O}_{\mathbb{P}}(1). \tag{A.29}$$

Combining (A.16), (A.18), (A.19), (A.27), (A.28) and (A.29), we find

$$\begin{aligned}\frac{1}{k}\sum_{i=1}^{k}\frac{X_{n-i+1:n}-X_{n-k:n}}{a(n/k)} &\stackrel{\mathrm{d}}{=} \left(\frac{1}{k}\sum_{i=1}^{k}\frac{Y_i^{\gamma}-1}{\gamma} + \frac{1}{(1-\gamma)(1-\gamma-\rho)}A(n/k) + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{v_n(\gamma)}\right)\right) \\ &\quad \times\left(1+\mathrm{o}_{\mathbb{P}}\left(\frac{1}{v_n(\gamma)}\right)\right).\end{aligned}$$

Apply Lemma A.4 to

$$\frac{1}{k}\sum_{i=1}^{k}\frac{Y_i^{\gamma}-1}{\gamma} - \frac{1}{1-\gamma} = \frac{1}{\gamma}\left(\frac{1}{k}\sum_{i=1}^{k}Y_i^{\gamma} - \frac{1}{1-\gamma}\right)$$

in order to complete the proof in this case.

Step 2: Asymptotic equivalence between $\widehat{\mathrm{ES}}_n(\tau_n)$, $\widehat{\mathrm{ES}}_n^{(1)}(\tau_n)$ and $\widehat{\mathrm{ES}}_n^{(2)}(\tau_n)$

Recall that in this part of the result we assume moreover $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$. Clearly $\widehat{\mathrm{ES}}_n^{(1)}(\tau_n) = \widehat{\mathrm{ES}}_n^{(1)}(1-k/n)$, and thus, by Equation (A.23),

$$\frac{\widehat{\mathrm{ES}}_n^{(1)}(\tau_n) - X_{\lceil n\tau_n\rceil:n}}{a((1-\tau_n)^{-1})} = \frac{\widehat{\mathrm{ES}}_n^{(1)}(1-k/n) - X_{n-k:n}}{a(n/k)}\left(1+\mathrm{o}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right)\right).$$

Furthermore,

$$\widehat{\mathrm{ES}}_n^{(1)}(1-k/n) - X_{n-k:n} = \frac{\sum_{i=1}^n (X_i - X_{n-k:n})\mathbb{1}_{\{X_i > X_{n-k:n}\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > X_{n-k:n}\}}} = \frac{\sum_{i=1}^k (X_{n-i+1:n} - X_{n-k:n})}{\sum_{i=1}^n \mathbb{1}_{\{X_i > X_{n-k:n}\}}}$$

so that, by Lemma A.6,

$$\begin{aligned}\widehat{\mathrm{ES}}_n^{(1)}(1-k/n) - X_{n-k:n} &= \frac{\frac{1}{k}\sum_{i=1}^k (X_{n-i+1:n} - X_{n-k:n})}{\frac{n}{k}\widehat{\overline{F}}_n(X_{n-k:n})} \\ &= \frac{1}{k}\sum_{i=1}^k (X_{n-i+1:n} - X_{n-k:n})\left(1 + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)\right) \\ &= \left(\widehat{\mathrm{ES}}_n(\tau_n) - X_{\lceil n\tau_n\rceil:n}\right)\left(1 + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right)\right).\end{aligned}$$

Since obviously

$$\begin{aligned}\widehat{\mathrm{ES}}_n(\tau_n) - X_{\lceil n\tau_n\rceil:n} &= \frac{n}{k}\int_{1-k/n}^1 (\widehat{q}_n(t) - \widehat{q}_n(1-k/n))\mathrm{d}t \\ &= \frac{n(1-\tau_n)}{k} \times \frac{1}{1-\tau_n}\int_{\tau_n}^1 (\widehat{q}_n(t) - \widehat{q}_n(\tau_n))\mathrm{d}t \\ &= \left(\widehat{\mathrm{ES}}_n^{(2)}(\tau_n) - X_{\lceil n\tau_n\rceil:n}\right)\left(1 + \mathrm{o}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right)\right),\end{aligned}$$

one finds

$$\begin{aligned}\frac{\widehat{\mathrm{ES}}_n(\tau_n) - X_{\lceil n\tau_n\rceil:n}}{a((1-\tau_n)^{-1})} &= \frac{\widehat{\mathrm{ES}}_n^{(1)}(\tau_n) - X_{\lceil n\tau_n\rceil:n}}{a((1-\tau_n)^{-1})}\left(1 + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right)\right) \\ &= \frac{\widehat{\mathrm{ES}}_n^{(2)}(\tau_n) - X_{\lceil n\tau_n\rceil:n}}{a((1-\tau_n)^{-1})}\left(1 + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right)\right).\end{aligned}$$

Applying the convergence result obtained as a first step completes the proof because in any of the cases considered $v_n(\gamma) = \mathrm{O}(\sqrt{n(1-\tau_n)})$. □

Proof of Theorem 1. Write

$$\frac{\widehat{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} = \frac{\widehat{\mathrm{ES}}_n(\tau_n) - X_{\lceil n\tau_n\rceil:n}}{a((1-\tau_n)^{-1})} + \frac{X_{\lceil n\tau_n\rceil:n} - q(\tau_n)}{a((1-\tau_n)^{-1})} - \frac{\mathrm{ES}(\tau_n) - q(\tau_n)}{a((1-\tau_n)^{-1})}.$$

Set, as in Proposition A.1,

$$v_n(\gamma) = \begin{cases} \sqrt{n(1-\tau_n)} & \text{if } 0 < \gamma < 1/2, \\ \sqrt{n(1-\tau_n)}/\sqrt{\log(n(1-\tau_n))} & \text{if } \gamma = 1/2, \\ (n(1-\tau_n))^{1-\gamma} & \text{if } 1/2 < \gamma < 1. \end{cases}$$

Combine Proposition 1(ii) and Proposition A.1 to get

$$v_n(\gamma)\left(\frac{\widehat{\mathrm{ES}}_n(\tau_n)-X_{\lceil n\tau_n\rceil:n}}{a((1-\tau_n)^{-1})}-\frac{\mathrm{ES}(\tau_n)-q(\tau_n)}{a((1-\tau_n)^{-1})},\ \frac{X_{\lceil n\tau_n\rceil:n}-q(\tau_n)}{a((1-\tau_n)^{-1})}\right)$$

$$\xrightarrow{\mathrm{d}}\begin{cases}\mathcal{N}\left(\mathbf{0}_2,\begin{pmatrix}\dfrac{1+\gamma+2\gamma^2}{(1-\gamma)(1-2\gamma)} & \dfrac{\gamma}{1-\gamma}\\ \dfrac{\gamma}{1-\gamma} & 1\end{pmatrix}\right) & \text{if } \gamma<1/2,\\ (\mathcal{N}(0,4),\,0) & \text{if } \gamma=1/2,\\ \left(\dfrac{1}{\gamma}\left\{-\dfrac{\Gamma(2-1/\gamma)}{1/\gamma-1}\cos\left(\dfrac{\pi}{2\gamma}\right)\right\}^{\gamma}Z_{1/\gamma},\,0\right) & \text{if } 1/2<\gamma<1.\end{cases}$$

The result is now a simple consequence of this joint convergence property. □

Proof of Theorem 2. Note that

$$\frac{\widetilde{\mathrm{ES}}_n(\tau_n)-\mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})}=\frac{X_{\lceil n\tau_n\rceil:n}-q(\tau_n)}{a((1-\tau_n)^{-1})}+\frac{1}{1-\widehat{\gamma}_n}\left(\frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})}-1\right)+\frac{\widehat{\gamma}_n-\gamma}{(1-\widehat{\gamma}_n)(1-\gamma)}$$
$$-\left\{\frac{\mathrm{ES}(\tau_n)-q(\tau_n)}{a((1-\tau_n)^{-1})}-\frac{1}{1-\gamma}\right\}.$$

The result is then a simple consequence of Proposition 1(ii). □

Proof of Theorem 3. Recall the notation $d_n=(1-\tau_n)/(1-\tau_n')$ and start by writing

$$\begin{aligned}\frac{\overline{\mathrm{ES}}^{\star}_n(\tau_n')-\mathrm{ES}(\tau_n')}{a((1-\tau_n)^{-1})}&=\frac{\overline{\mathrm{ES}}_n(\tau_n)-\mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})}+\left(\frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})}-1\right)\frac{1}{1-\widehat{\gamma}_n}\int_1^{d_n}s^{\widehat{\gamma}_n-1}\mathrm{d}s\\&+\frac{\widehat{\gamma}_n-\gamma}{(1-\widehat{\gamma}_n)(1-\gamma)}\int_1^{d_n}s^{\widehat{\gamma}_n-1}\mathrm{d}s+\frac{1}{1-\gamma}\left(\int_1^{d_n}s^{\widehat{\gamma}_n-1}\mathrm{d}s-\int_1^{d_n}s^{\gamma-1}\mathrm{d}s\right)\\&-\left(\frac{\mathrm{ES}(\tau_n')-\mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})}-\frac{1}{1-\gamma}\int_1^{d_n}s^{\gamma-1}\mathrm{d}s\right).\end{aligned}\tag{A.30}$$

Now, for any $t\in\mathbb{R}$,

$$\left|\frac{\int_1^{d_n}s^{t-1}\mathrm{d}s}{\int_1^{d_n}s^{\gamma-1}\mathrm{d}s}-1\right|\le\sup_{1\le s\le d_n}|s^{t-\gamma}-1|\le|t-\gamma|\log(d_n)\exp(|t-\gamma|\log(d_n))$$

by the mean value theorem. Since $\log(d_n)/\sqrt{n(1-\tau_n)}\to 0$ and $\widehat{\gamma}_n$ is $\sqrt{n(1-\tau_n)}$−consistent, this yields

$$\frac{\int_1^{d_n}s^{\widehat{\gamma}_n-1}\mathrm{d}s}{\int_1^{d_n}s^{\gamma-1}\mathrm{d}s}\xrightarrow{\mathbb{P}}1.$$

Besides, a straightforward calculation gives

$$\int_1^{d_n}s^{\gamma-1}\log(s)\mathrm{d}s=\begin{cases}\dfrac{1}{\gamma}\left(d_n^{\gamma}\log(d_n)-\dfrac{d_n^{\gamma}-1}{\gamma}\right) & \text{if } \gamma\neq 0,\\ \dfrac{\log^2(d_n)}{2} & \text{if } \gamma=0.\end{cases}\tag{A.31}$$

As a consequence,

$$\frac{1}{\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s} \to \gamma_-^2 \quad \text{and} \quad \frac{\int_1^{d_n} s^{\widehat{\gamma}_n-1}\mathrm{d}s}{\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s} \xrightarrow{\mathbb{P}} -\gamma_-. \tag{A.32}$$

Furthermore, for any $s\in[1,d_n]$ and any $t\in\mathbb{R}$, a Taylor expansion yields

$$\begin{aligned} |s^{t-1}-s^{\gamma-1}-(t-\gamma)s^{\gamma-1}\log(s)| &\le \frac{(t-\gamma)^2\log^2(s)}{2}\max_{u\in[\gamma,t]} s^{u-1} \\ &\le \frac{(t-\gamma)^2}{2}(s^{\gamma-1}+s^{t-1})\log^2(s). \end{aligned}$$

It follows that

$$\begin{aligned} &\left|\int_1^{d_n} s^{\widehat{\gamma}_n-1}\mathrm{d}s - \int_1^{d_n} s^{\gamma-1}\mathrm{d}s - (\widehat{\gamma}_n-\gamma)\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s\right| \\ &\le \frac{(\widehat{\gamma}_n-\gamma)^2\log(d_n)}{2}\int_1^{d_n}(s^{\gamma-1}+s^{\widehat{\gamma}_n-1})\log(s)\mathrm{d}s. \end{aligned}$$

Since for any $t\in\mathbb{R}$,

$$\left|\frac{\int_1^{d_n} s^{t-1}\log(s)\mathrm{d}s}{\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s}-1\right| \le \sup_{1\le s\le d_n}|s^{t-\gamma}-1| \le |t-\gamma|\log(d_n)\exp(|t-\gamma|\log(d_n))$$

by the mean value theorem again, one finds

$$\frac{\int_1^{d_n} s^{\widehat{\gamma}_n-1}\log(s)\mathrm{d}s}{\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s} \xrightarrow{\mathbb{P}} 1$$

and thus

$$\frac{\sqrt{n(1-\tau_n)}}{\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s}\left|\int_1^{d_n} s^{\widehat{\gamma}_n-1}\mathrm{d}s - \int_1^{d_n} s^{\gamma-1}\mathrm{d}s - (\widehat{\gamma}_n-\gamma)\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s\right| \xrightarrow{\mathbb{P}} 0. \tag{A.33}$$

Conclude from (A.32) and (A.33) that

$$\begin{aligned} &\frac{\overline{\mathrm{ES}}_n(\tau_n)-\mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} + \left(\frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})}-1\right)\frac{1}{1-\widehat{\gamma}_n}\int_1^{d_n} s^{\widehat{\gamma}_n-1}\mathrm{d}s \\ &+ \frac{\widehat{\gamma}_n-\gamma}{(1-\widehat{\gamma}_n)(1-\gamma)}\int_1^{d_n} s^{\widehat{\gamma}_n-1}\mathrm{d}s + \frac{1}{1-\gamma}\left(\int_1^{d_n} s^{\widehat{\gamma}_n-1}\mathrm{d}s - \int_1^{d_n} s^{\gamma-1}\mathrm{d}s\right) \\ &\stackrel{\mathrm{d}}{=} \frac{\overline{\mathrm{ES}}_n(\tau_n)-\mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} - \frac{\gamma_-}{1-\gamma_-}\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s\left(\frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})}-1\right) \\ &+ \left(\frac{1}{1-\gamma}-\frac{\gamma_-}{(1-\gamma_-)^2}\right)\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s\,(\widehat{\gamma}_n-\gamma) + \mathrm{o}_{\mathbb{P}}\left(\frac{\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s}{\sqrt{n(1-\tau_n)}}\right). \end{aligned} \tag{A.34}$$

The final (bias) term in (A.30) is controlled by noting that, for any $\gamma\in\mathbb{R}$,

$$\frac{\gamma}{1-\gamma}\int_1^{d_n} s^{\gamma-1}\mathrm{d}s = \frac{1}{1-\gamma}\left(\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma}-1\right).$$

[Note that indeed both sides are 0 when $\gamma = 0$.] This makes it possible to write

$$\begin{aligned}
&\frac{\mathrm{ES}(\tau_n') - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} - \frac{1}{1-\gamma}\int_1^{d_n} s^{\gamma-1}\mathrm{d}s \\
&= \frac{a((1-\tau_n')^{-1})}{a((1-\tau_n)^{-1})}\left(\frac{\mathrm{ES}(\tau_n') - q(\tau_n')}{a((1-\tau_n')^{-1})} - \frac{1}{1-\gamma}\right) + \frac{1}{1-\gamma}\left(\frac{a((1-\tau_n')^{-1})}{a((1-\tau_n)^{-1})} - \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma}\right) \\
&- \left(\frac{\mathrm{ES}(\tau_n) - q(\tau_n)}{a((1-\tau_n)^{-1})} - \frac{1}{1-\gamma}\right) + \left(\frac{q(\tau_n') - q(\tau_n)}{a((1-\tau_n)^{-1})} - \int_1^{d_n} s^{\gamma-1}\mathrm{d}s\right). \qquad \text{(A.35)}
\end{aligned}$$

Recalling (A.21) and using the assumption $\rho < 0$ when $\gamma \geq 0$ along with (A.31), we obtain

$$\begin{aligned}
\frac{a((1-\tau_n')^{-1})}{a((1-\tau_n)^{-1})} - \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma} &= \begin{cases} A((1-\tau_n)^{-1})d_n^{\gamma}\left(-\dfrac{1}{\rho} + \mathrm{o}(1)\right) & \text{if } \gamma \geq 0,\ \rho < 0, \\ \mathrm{o}(|A((1-\tau_n)^{-1})|) & \text{if } \gamma < 0 \end{cases} \\
&= \mathrm{o}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s\right). \qquad \text{(A.36)}
\end{aligned}$$

Combining Proposition 1(ii) and the assumption $\rho < 0$ when $\gamma \geq 0$ again with the above equality, Potter bounds (see de Haan and Ferreira, 2006, Proposition B.1.9.5 p.367) on the function $|A|$ and (A.31), we find

$$\begin{aligned}
&\frac{a((1-\tau_n')^{-1})}{a((1-\tau_n)^{-1})}\left(\frac{\mathrm{ES}(\tau_n') - q(\tau_n')}{a((1-\tau_n')^{-1})} - \frac{1}{1-\gamma}\right) \\
&= \mathrm{O}\left(\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma}|A((1-\tau_n')^{-1})|\right) + \mathrm{o}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s\right) \\
&= \mathrm{o}(|A((1-\tau_n)^{-1})|d_n^{\gamma}) + \mathrm{o}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s\right) \\
&= \mathrm{o}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s\right). \qquad \text{(A.37)}
\end{aligned}$$

Besides, Proposition 1(ii) and (A.32) yield

$$\begin{aligned}
&\frac{\sqrt{n(1-\tau_n)}}{\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s}\left(\frac{\mathrm{ES}(\tau_n) - q(\tau_n)}{a((1-\tau_n)^{-1})} - \frac{1}{1-\gamma}\right) \\
&= \begin{cases} \mathrm{o}(\sqrt{n(1-\tau_n)}|A((1-\tau_n)^{-1})|) = \mathrm{o}(1) & \text{if } \gamma \geq 0,\ \rho < 0, \\ \lambda\dfrac{\gamma^2}{(1-\gamma)(1-\gamma-\rho)} + \mathrm{o}(1) & \text{if } \gamma < 0 \end{cases} \\
&= \lambda\frac{\gamma_-^2}{(1-\gamma_-)(1-\gamma_- - \rho)} + \mathrm{o}(1). \qquad \text{(A.38)}
\end{aligned}$$

It is, finally, shown on p.137 in de Haan and Ferreira (2006) that

$$\frac{\sqrt{n(1-\tau_n)}}{\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s}\left\{\frac{q(\tau_n')-q(\tau_n)}{a((1-\tau_n)^{-1})}-\int_1^{d_n} s^{\gamma-1}\mathrm{d}s\right\}=\lambda\frac{\gamma_-}{\gamma_-+\rho}+\mathrm{o}(1). \tag{A.39}$$

Combine (A.35), (A.36), (A.37), (A.38) and (A.39) to get

$$\begin{aligned}&\frac{\sqrt{n(1-\tau_n)}}{\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s}\left(\frac{\mathrm{ES}(\tau_n')-\mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})}-\frac{1}{1-\gamma}\int_1^{d_n} s^{\gamma-1}\mathrm{d}s\right)\\ &\to\lambda\frac{\gamma_-(1-2\gamma_--\rho)}{(1-\gamma_-)(1-\gamma_--\rho)(\gamma_-+\rho)}.\end{aligned}$$

Report this together with (A.34) into (A.30) to complete the proof. □

Recall that the Dekkers et al. (1989) moment estimators of the scale parameter $a(t)$ at $t = (1-\tau_n)^{-1}$ and of the shape parameter γ are respectively

$$\begin{aligned}&\widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1})=X_{\lceil n\tau_n\rceil:n}M_n^{(1)}(1-\widehat{\gamma}_{n,-}^{\mathrm{Mom}})\\ &\qquad\text{and } \widehat{\gamma}_n^{\mathrm{Mom}}=M_n^{(1)}+\widehat{\gamma}_{n,-}^{\mathrm{Mom}},\ \text{with } \widehat{\gamma}_{n,-}^{\mathrm{Mom}}=1-\frac{1}{2}\left(1-\frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1},\end{aligned}$$

where

$$M_n^{(1)}=\frac{1}{\lfloor n(1-\tau_n)\rfloor}\sum_{i=1}^{\lfloor n(1-\tau_n)\rfloor}\log\frac{X_{n-i+1:n}}{X_{\lceil n\tau_n\rceil:n}}\ \text{and } M_n^{(2)}=\frac{1}{\lfloor n(1-\tau_n)\rfloor}\sum_{i=1}^{\lfloor n(1-\tau_n)\rfloor}\log^2\frac{X_{n-i+1:n}}{X_{\lceil n\tau_n\rceil:n}}.$$

The main technical ingredient for the asymptotic analysis of these moment estimators is Proposition 2, whose proof relies on the following lemma.

Lemma A.7 (Uniform inequality on $\log U$)**.** *Assume that* $0 < U(\infty) = q(1) \le \infty$*, that condition* $\mathcal{C}_2(\gamma, a, \rho, A)$ *holds, and recall the definitions of* $a_\star$*,* $A_\star$ *and* $\Psi_{\gamma,\rho}$ *as in Equations* (A.6) *and* (A.7)*. Then, up to replacing the functions* $a_\star$ *and* $A_\star$ *by suitable functions* $a_{\star,0}$ *and* $A_{\star,0}$ *such that* $a_{\star,0}(t) = a_\star(t)(1 + \mathrm{o}(|A(t)|))$ *and* $A_{\star,0}(t) = A_\star(t)(1 + \mathrm{o}(1))$ *as* $t \to \infty$*, we have, for any* $\delta, \varepsilon > 0$ *sufficiently small, that for* t *large enough and all* $x > 1$*,*

$$\begin{aligned}&\left|\frac{\log U(tx)-\log U(t)}{a_\star(t)/U(t)}-\int_1^x s^{\gamma_--1}\mathrm{d}s+\left(\frac{a_\star(t)}{U(t)}-\gamma_+\right)\int_1^x s^{\gamma_--1}\left(\int_1^s u^{-|\gamma|-1}\mathrm{d}u\right)\mathrm{d}s\right.\\ &\left.-A_\star(t)x^{-\gamma_+}\Psi_{\gamma,\rho}(x)\right|\le\varepsilon\left(\left|\frac{a(t)}{U(t)}-\gamma_+\right|+|A(t)|\right)x^\delta.\end{aligned}$$

Proof of Lemma A.7. Work throughout with the versions of the functions $a_\star = a_{\star,0}$ and $A_\star = A_{\star,0}$ that make (A.8) hold true. Take t so large that $U(t) > 0$ and write, for any $x > 1$,

$$\frac{\log U(tx)-\log U(t)}{a_\star(t)/U(t)}=\frac{U(t)}{a_\star(t)}\log\left(1+\frac{a_\star(t)}{U(t)}\frac{U(tx)-U(t)}{a_\star(t)}\right).$$

We treat the two cases $\gamma \le 0$ and $\gamma > 0$ separately, and when considering $\gamma > 0$ we shall deal with the subcases $\rho < 0$ and $\rho = 0$ separately. Fix $\delta, \varepsilon > 0$ small enough; in particular, we impose that $\delta < -\rho$ in the case $\gamma > 0$ and $\rho < 0$. Fix also $\varepsilon \in (0, 1)$.

Case $\gamma \le 0$: It follows from (A.8) that

$$\forall \eta > 0,\ \sup_{x\ge 1} x^{-\eta} \left| \frac{U(tx) - U(t)}{a_\star(t)} - \int_1^x s^{\gamma-1} \mathrm{d}s \right| = \mathrm{O}(|A(t)|) \to 0 \text{ as } t \to \infty. \tag{A.40}$$

In particular,

$$\forall \eta > 0,\ \sup_{x\ge 1} x^{-\eta} \frac{U(tx) - U(t)}{a_\star(t)} = \mathrm{O}(1) \text{ as } t \to \infty. \tag{A.41}$$

These facts and the convergence $a_\star(t)/U(t) \to 0$ motivate writing the identity

$$\begin{aligned} &\frac{\log U(tx) - \log U(t)}{a_\star(t)/U(t)} - \int_1^x s^{\gamma-1} \mathrm{d}s \\ &= \frac{U(t)}{a_\star(t)} \left\{ \log\left(1 + \frac{a_\star(t)}{U(t)} \frac{U(tx) - U(t)}{a_\star(t)}\right) - \frac{a_\star(t)}{U(t)} \frac{U(tx) - U(t)}{a_\star(t)} \right\} \\ &+ \frac{U(tx) - U(t)}{a_\star(t)} - \int_1^x s^{\gamma-1} \mathrm{d}s. \end{aligned} \tag{A.42}$$

We concentrate on the first term. By Taylor's theorem,

$$\left| \log(1+h) - h + \frac{h^2}{2} \right| \le \frac{h^3}{3} \text{ for any } h > 0.$$

Combining this for $h = \frac{a_\star(t)}{U(t)} \frac{U(tx)-U(t)}{a_\star(t)} > 0$ with (A.41) results in

$$\begin{aligned} &\left| \frac{U(t)}{a_\star(t)} \left\{ \log\left(1 + \frac{a_\star(t)}{U(t)} \frac{U(tx) - U(t)}{a_\star(t)}\right) - \frac{a_\star(t)}{U(t)} \frac{U(tx) - U(t)}{a_\star(t)} \right\} + \frac{1}{2} \frac{a_\star(t)}{U(t)} \left(\frac{U(tx) - U(t)}{a_\star(t)} \right)^2 \right| \\ &\le \frac{\varepsilon}{2} \frac{a(t)}{U(t)} x^\delta \end{aligned}$$

for t large enough, uniformly in $x > 1$. Then, using (A.40) and (A.41) again, we find

$$\begin{aligned} &\left| \frac{U(t)}{a_\star(t)} \left\{ \log\left(1 + \frac{a_\star(t)}{U(t)} \frac{U(tx) - U(t)}{a_\star(t)}\right) - \frac{a_\star(t)}{U(t)} \frac{U(tx) - U(t)}{a_\star(t)} \right\} + \frac{1}{2} \frac{a_\star(t)}{U(t)} \left(\int_1^x s^{\gamma-1} \mathrm{d}s \right)^2 \right| \\ &\le \varepsilon \frac{a(t)}{U(t)} x^\delta \end{aligned} \tag{A.43}$$

for t large enough, uniformly in $x > 1$. Combining (A.8), (A.42) and (A.43), we get, for t large enough,

$$\begin{aligned} &\left| \frac{\log U(tx) - \log U(t)}{a_\star(t)/U(t)} - \int_1^x s^{\gamma-1} \mathrm{d}s + \frac{1}{2} \frac{a_\star(t)}{U(t)} \left(\int_1^x s^{\gamma-1} \mathrm{d}s \right)^2 - A_\star(t) \Psi_{\gamma,\rho}(x) \right| \\ &\le \varepsilon \left(\frac{a(t)}{U(t)} x^\delta + |A(t)| x^{\gamma+\rho+\delta} \right) \le \varepsilon \left(\frac{a(t)}{U(t)} + |A(t)| \right) x^\delta \text{ for all } x > 1. \end{aligned}$$

Since, when $\gamma \leq 0$,

$$\int_1^x s^{\gamma_- -1} \left(\int_1^s u^{-|\gamma|-1} \mathrm{d}u \right) \mathrm{d}s = \int_1^x s^{\gamma-1} \left(\int_1^s u^{\gamma-1} \mathrm{d}u \right) \mathrm{d}s = \frac{1}{2} \left(\int_1^x s^{\gamma-1} \mathrm{d}s \right)^2,$$

the desired result follows in this case.

Case $\gamma > 0$: One has

$$\frac{\log U(tx) - \log U(t)}{a_\star(t)/U(t)} - \log(x) = \left(\frac{U(t)}{a_\star(t)} - \frac{1}{\gamma} \right) (\log U(tx) - \log U(t)) \\ + \frac{1}{\gamma} \log \left(1 + x^{-\gamma} \frac{a_\star(t)}{U(t)} \frac{U(tx) - U(t)}{a_\star(t)} - (1 - x^{-\gamma}) \right). \quad \text{(A.44)}$$

We start by controlling the first term in (A.44). To this end we use the convergence $a(t)/U(t) \to \gamma$ as $t \to \infty$ in order to write

$$\frac{U(t)}{a_\star(t)} - \frac{1}{\gamma} = -\frac{1}{\gamma^2} \left(\frac{a_\star(t)}{U(t)} - \gamma \right) (1 + \eta(t))$$

where $\eta(t) \to 0$ as $t \to \infty$. This results in

$$\begin{aligned} &\left(\frac{U(t)}{a_\star(t)} - \frac{1}{\gamma} \right) (\log U(tx) - \log U(t)) \\ &= \gamma \left(\frac{U(t)}{a_\star(t)} - \frac{1}{\gamma} \right) \log(x) + \left(\frac{U(t)}{a_\star(t)} - \frac{1}{\gamma} \right) (\log U(tx) - \log U(t) - \gamma \log(x)) \\ &= -\frac{1}{\gamma} \log(x) \left(\frac{a_\star(t)}{U(t)} - \gamma \right) (1 + \eta(t)) \\ &- \frac{1}{\gamma^2} \left(\frac{a_\star(t)}{U(t)} - \gamma \right) (\log U(tx) - \log U(t) - \gamma \log(x))(1 + \eta(t)). \end{aligned} \quad \text{(A.45)}$$

The function U is regularly varying with index γ, so

$$\forall x > 0, \ \lim_{t \to \infty} \frac{\log U(tx) - \log U(t)}{\gamma} = \log(x).$$

This convergence makes it possible to apply Proposition B.2.17 p.382 in de Haan and Ferreira (2006) to the function $\log U$; combined with (A.45), this provides, for t large enough,

$$\left| \left(\frac{U(t)}{a_\star(t)} - \frac{1}{\gamma} \right) (\log U(tx) - \log U(t)) + \frac{1}{\gamma} \log(x) \left(\frac{a_\star(t)}{U(t)} - \gamma \right) \right| \leq \varepsilon \left(\left| \frac{a(t)}{U(t)} - \gamma \right| + |A(t)| \right) x^\delta \quad \text{(A.46)}$$

for all $x > 1$ (recall that $|a_\star(t)/U(t) - \gamma| = \mathrm{O}(|a(t)/U(t) - \gamma| + |A(t)|)$ by definition of $a_\star$).

We turn to the control of the second term in (A.44). When $\rho < 0$, note that (A.8) and (A.40) imply

$$
\begin{aligned}
x^{-\gamma}\frac{a_\star(t)}{U(t)}\frac{U(tx)-U(t)}{a_\star(t)} - (1-x^{-\gamma}) &= \left(\frac{a_\star(t)}{U(t)} - \gamma\right)\frac{x^{-\gamma}-1}{-\gamma} + \gamma x^{-\gamma} A_\star(t)\Psi_{\gamma,\rho}(x) \\
&\quad + x^{-\gamma}\left(\frac{a_\star(t)}{U(t)} - \gamma\right)\left(\frac{U(tx)-U(t)}{a_\star(t)} - \frac{x^{\gamma}-1}{\gamma}\right) \\
&\quad + \gamma x^{-\gamma}\left(\frac{U(tx)-U(t)}{a_\star(t)} - \frac{x^{\gamma}-1}{\gamma} - A_\star(t)\Psi_{\gamma,\rho}(x)\right) \\
&= \left(\frac{a_\star(t)}{U(t)} - \gamma\right)\frac{x^{-\gamma}-1}{-\gamma} + \gamma x^{-\gamma} A_\star(t)\Psi_{\gamma,\rho}(x) \\
&\quad + \mathrm{o}\left(\left|\frac{a(t)}{U(t)} - \gamma\right| + |A(t)|\right) \qquad \text{(A.47)}
\end{aligned}
$$

uniformly in $x > 1$ as $t \to \infty$. In particular

$$
\sup_{x>1}\left|x^{-\gamma}\frac{a_\star(t)}{U(t)}\frac{U(tx)-U(t)}{a_\star(t)} - (1-x^{-\gamma})\right| = \mathrm{O}\left(\left|\frac{a(t)}{U(t)} - \gamma\right| + |A(t)|\right) \to 0 \qquad \text{(A.48)}
$$

as $t \to \infty$. Combine (A.47) and (A.48) with a Taylor expansion of $z \mapsto \log(1+z)$ around $z = 0$ to obtain

$$
\begin{aligned}
&\frac{1}{\gamma}\log\left(1 + x^{-\gamma}\frac{a_\star(t)}{U(t)}\frac{U(tx)-U(t)}{a_\star(t)} - (1-x^{-\gamma})\right) \\
&= \frac{1}{\gamma}\left\{\left(\frac{a_\star(t)}{U(t)} - \gamma\right)\frac{x^{-\gamma}-1}{-\gamma} + \gamma x^{-\gamma} A_\star(t)\Psi_{\gamma,\rho}(x)\right\} + \mathrm{o}\left(\left|\frac{a(t)}{U(t)} - \gamma\right| + |A(t)|\right) \qquad \text{(A.49)}
\end{aligned}
$$

uniformly in $x > 1$ as $t \to \infty$. When $\rho = 0$, write instead the second term in (A.44) as

$$
\frac{1}{\gamma}\log\left(1 + x^{-\gamma}\frac{a_\star(t)}{U(t)}\frac{U(tx)-U(t)}{a_\star(t)} - (1-x^{-\gamma})\right) = \frac{\log((tx)^{-\gamma}U(tx)) - \log(t^{-\gamma}U(t))}{\gamma}.
$$

By Theorems 2.1 and 3.1 in Fraga Alves et al. (2007),

$$
\forall x > 0,\ \lim_{t\to\infty}\frac{1}{A(t)}\left(\frac{U(tx)}{U(t)} - x^{\gamma}\right) = x^{\gamma}\log(x). \qquad \text{(A.50)}
$$

Taking the logarithm in (A.50) produces the convergence

$$
\forall x > 0,\ \lim_{t\to\infty}\frac{\log((tx)^{-\gamma}U(tx)) - \log(t^{-\gamma}U(t))}{A(t)} = \log(x).
$$

Apply then Proposition B.2.17 p.382 in de Haan and Ferreira (2006) to obtain that

$$
\left|\log((tx)^{-\gamma}U(tx)) - \log(t^{-\gamma}U(t)) - A(t)\log(x)\right| \le \varepsilon|A(t)|x^{\delta}
$$

for all $x > 1$ when t is large enough. In other words

$$
\begin{aligned}
&\left|\frac{1}{\gamma}\log\left(1 + x^{-\gamma}\frac{a_\star(t)}{U(t)}\frac{U(tx)-U(t)}{a_\star(t)} - (1-x^{-\gamma})\right) - A(t)x^{-\gamma}\Psi_{\gamma,\rho}(x)\right| \\
&= \left|\frac{\log((tx)^{-\gamma}U(tx)) - \log(t^{-\gamma}U(t))}{\gamma} - A(t)x^{-\gamma}\Psi_{\gamma,\rho}(x)\right| \le \varepsilon|A(t)|x^{\delta} \qquad \text{(A.51)}
\end{aligned}
$$

for all $x > 1$ when t is large enough, because $\Psi_{\gamma,\rho}(x) = \frac{1}{\gamma}x^{\gamma}\log(x)$ when $\rho = 0$ and $\gamma > 0$. Furthermore, Theorem 2.1 in Fraga Alves et al. (2007) guarantees that

$$\begin{aligned}\frac{a_\star(t)}{U(t)} - \gamma &= \frac{a(t)}{U(t)}\left(1 - \frac{1}{\gamma}A(t) + \mathrm{o}(|A(t)|)\right) - \gamma \\ &= \frac{a(t)}{U(t)} - \gamma - A(t) + \mathrm{o}(|A(t)|) = \mathrm{o}(|A(t)|). \qquad \text{(A.52)}\end{aligned}$$

Conclude, from a combination of (A.44), (A.46) and (A.49) when $\rho < 0$, that

$$\Bigg|\frac{\log U(tx) - \log U(t)}{a_\star(t)/U(t)} - \log(x) + \frac{1}{\gamma}\left(\frac{a_\star(t)}{U(t)} - \gamma\right)\left(\log(x) - \frac{x^{-\gamma}-1}{-\gamma}\right) \\ - A_\star(t)x^{-\gamma}\Psi_{\gamma,\rho}(x)\Bigg| \le \varepsilon\left(\left|\frac{a(t)}{U(t)} - \gamma\right| + |A(t)|\right)x^{\delta}$$

when t is large enough uniformly in $x > 1$, and likewise from a combination of (A.44), (A.46), (A.51) and (A.52) when $\rho = 0$ (since then $A_\star = A$). The result follows because when $\gamma > 0$,

$$\int_1^x s^{\gamma_- - 1}\left(\int_1^s u^{-|\gamma|-1}\mathrm{d}u\right)\mathrm{d}s = \int_1^x s^{-1}\left(\int_1^s u^{-\gamma-1}\mathrm{d}u\right)\mathrm{d}s = \frac{1}{\gamma}\left(\log(x) - \frac{x^{-\gamma}-1}{-\gamma}\right).$$

The proof is complete. □

Proof of Proposition 2. We prove the result by suitably modifying the proof of Proposition 1(i). Fix $\delta, \varepsilon > 0$ such that the conclusion of Lemma A.7 holds. Write, with the notation of Lemma A.7,

$$\begin{aligned}&\frac{\log U(tx) - \log U(t)}{a(t)/U(t)} - \int_1^x s^{\gamma_- - 1}\mathrm{d}s \\ &= \frac{\log U(tx) - \log U(t)}{a_\star(t)/U(t)} - \int_1^x s^{\gamma_- - 1}\mathrm{d}s + \left(\frac{a_\star(t)}{a(t)} - 1\right)\int_1^x s^{\gamma_- - 1}\mathrm{d}s \\ &+ \left(\frac{a_\star(t)}{a(t)} - 1\right)\left(\frac{\log U(tx) - \log U(t)}{a_\star(t)/U(t)} - \int_1^x s^{\gamma_- - 1}\mathrm{d}s\right). \qquad \text{(A.53)}\end{aligned}$$

By Lemma A.7, one has, for t large enough and all $x > 1$,

$$\left|\left(\frac{a_\star(t)}{a(t)} - 1\right)\left(\frac{\log U(tx) - \log U(t)}{a_\star(t)/U(t)} - \int_1^x s^{\gamma_- - 1}\mathrm{d}s\right)\right| \le \frac{\varepsilon}{3}\left(\left|\frac{a(t)}{U(t)} - \gamma_+\right| + |A(t)|\right)x^{\delta}. \qquad \text{(A.54)}$$

Use Lemma A.7 again to obtain, for t large enough and all $x > 1$,

$$\Bigg|\frac{\log U(tx) - \log U(t)}{a_\star(t)/U(t)} - \int_1^x s^{\gamma_- - 1}\mathrm{d}s + \left(\frac{a_\star(t)}{U(t)} - \gamma_+\right)\int_1^x s^{\gamma_- - 1}\left(\int_1^s u^{-|\gamma|-1}\mathrm{d}u\right)\mathrm{d}s \\ - A_\star(t)x^{-\gamma_+}\Psi_{\gamma,\rho}(x)\Bigg| \le \frac{\varepsilon}{3}\left(\left|\frac{a(t)}{U(t)} - \gamma_+\right| + |A(t)|\right)x^{\delta}. \qquad \text{(A.55)}$$

Recall the convergence $a(t)/U(t) \to \gamma_+$ and, from the proof of Proposition 1(i), the notation $c(\gamma, \rho) = -1/\rho$ if $\rho < 0$, $-1/\gamma$ if $\rho = 0$ and $\gamma \neq 0$, and 0 if $\gamma = 0$ and $\rho = 0$, as well as the notation $d(\rho) = 1/\rho$ if $\rho < 0$ and 1 if $\rho = 0$, so that $a_\star(t)/a(t) - 1 = c(\gamma,\rho)A(t)(1 + \mathrm{o}(|A(t)|))$ and $A_\star(t) = d(\rho)A(t)(1 + \mathrm{o}(|A(t)|))$ as $t \to \infty$. Then

$$\frac{a_\star(t)}{U(t)} - \gamma_+ = \frac{a(t)}{U(t)} - \gamma_+ + c(\gamma,\rho)\gamma_+ A(t) + \mathrm{o}\left(\left|\frac{a(t)}{U(t)} - \gamma_+\right| + |A(t)|\right)$$

and, for t large enough and all $x > 1$,

$$\left|\left(\frac{a_\star(t)}{a(t)} - 1\right)\int_1^x s^{\gamma_- - 1}\mathrm{d}s - c(\gamma,\rho)A(t)\int_1^x s^{\gamma_- - 1}\mathrm{d}s\right| \le \frac{\varepsilon}{3}|A(t)|x^\delta. \tag{A.56}$$

Conclude that for t large enough and all $x > 1$,

$$\begin{aligned}&\left|\frac{\log U(tx) - \log U(t)}{a(t)/U(t)} - \int_1^x s^{\gamma_- - 1}\mathrm{d}s\right.\\ &+ \left(\frac{a(t)}{U(t)} - \gamma_+\right)\int_1^x s^{\gamma_- - 1}\left(\int_1^s u^{-|\gamma|-1}\mathrm{d}u\right)\mathrm{d}s - c(\gamma,\rho)A(t)\int_1^x s^{\gamma_- - 1}\mathrm{d}s\\ &\left.+ c(\gamma,\rho)\gamma_+ A(t)\int_1^x s^{\gamma_- - 1}\left(\int_1^s u^{-|\gamma|-1}\mathrm{d}u\right)\mathrm{d}s - d(\rho)A(t)x^{-\gamma_+}\Psi_{\gamma,\rho}(x)\right|\\ &\le \varepsilon\left(\left|\frac{a(t)}{U(t)} - \gamma_+\right| + |A(t)|\right)x^\delta\end{aligned}$$

by combining (A.53), (A.54), (A.55) and (A.56). To conclude it is then sufficient to show that

$$\begin{aligned}&c(\gamma,\rho)\left(\int_1^x s^{\gamma_- - 1}\mathrm{d}s - \gamma_+\int_1^x s^{\gamma_- - 1}\left(\int_1^s u^{-|\gamma|-1}\mathrm{d}u\right)\mathrm{d}s\right)\\ &+ d(\rho)x^{-\gamma_+}\Psi_{\gamma,\rho}(x) = x^{-\gamma_+}\int_1^x s^{\gamma-1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s.\end{aligned} \tag{A.57}$$

A simple calculation proves the identity

$$\int_1^x s^{\gamma_- - 1}\mathrm{d}s - \gamma_+\int_1^x s^{\gamma_- - 1}\left(\int_1^s u^{-|\gamma|-1}\mathrm{d}u\right)\mathrm{d}s = x^{-\gamma_+}\int_1^x s^{\gamma-1}\mathrm{d}s$$

so that (A.57) is in fact equivalent to (A.12), which was shown to be correct in the proof of Proposition 1(i). Conclude that (A.57) holds, thus completing the proof. □

We may now proceed to the statement of Lemma A.8, which is nothing but a joint convergence result for $(M_n^{(1)}, M_n^{(2)}, X_{\lceil n\tau_n\rceil:n})$. It is the key to the proof of the joint convergence Lemma A.9 about $\widehat{\mathrm{ES}}_n(\tau_n)$, the corresponding intermediate empirical quantile $X_{\lceil n\tau_n\rceil:n}$ and the moment estimators when $\gamma < 0$, which is an essential ingredient in the proofs of Theorems 4 and 5. Let us highlight that the present lemma is *not* a consequence of earlier results by Dekkers et al. (1989) and Chapter 3.5 in de Haan and Ferreira (2006), because it does not feature the restriction that $\gamma \neq \rho$.

Lemma A.8 (Joint weak convergence of the building blocks for the moment estimator)**.** *Suppose that the tail quantile function U of X satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$ with $0 < U(\infty) = q(1) \leq \infty$. Let $k = k(n)$ be a sequence of integers with $k \to \infty$, $k/n \to 0$, $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$ and $\sqrt{k}(a(n/k)/U(n/k) - \gamma_+) \to \mu \in \mathbb{R}$. Let further (Y_i) be a sequence of independent copies of a unit Pareto random variable Y. Define*

$$\mathfrak{M}_n^{(1)} = \frac{1}{k}\sum_{i=1}^{k} \log \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})} \quad \textit{and} \quad \mathfrak{M}_n^{(2)} = \frac{1}{k}\sum_{i=1}^{k} \log^2 \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})}.$$

Then

$$\sqrt{k}\left(\frac{U(Y_{n-k:n})}{a(Y_{n-k:n})}\mathfrak{M}_n^{(1)} - \frac{1}{1-\gamma_-}, \left\{\frac{U(Y_{n-k:n})}{a(Y_{n-k:n})}\right\}^2 \mathfrak{M}_n^{(2)} - \frac{2}{(1-\gamma_-)(1-2\gamma_-)}, \frac{U(Y_{n-k:n}) - U(n/k)}{a(n/k)}\right)$$
$$\xrightarrow{\text{d}} \mathcal{N}\left(\lambda(B_A^{(1)}(\gamma,\rho), B_A^{(2)}(\gamma,\rho), 0) + \mu(B_{a/U}^{(1)}(\gamma), B_{a/U}^{(2)}(\gamma), 0), \boldsymbol{V}\right)$$

where the bias terms in the limiting distribution are

$$B_A^{(1)}(\gamma,\rho) = \frac{1}{(1+|\gamma|)(1-\gamma_- - \rho)} = \begin{cases} \dfrac{1}{(1+\gamma)(1-\rho)} & \textit{if } \gamma \geq 0, \\ \dfrac{1}{(1-\gamma)(1-\gamma-\rho)} & \textit{if } \gamma < 0, \end{cases}$$

$$\begin{aligned} B_A^{(2)}(\gamma,\rho) &= \frac{2(3 + 2\gamma_+ - 4\gamma_- - 2\rho - \gamma_+\rho)}{(1+\gamma_+)(1+|\gamma|)(1-2\gamma_-)(1-\gamma_- - \rho)(1-2\gamma_- - \rho)} \\ &= \begin{cases} \dfrac{2(3 + 2\gamma - 2\rho - \gamma\rho)}{(1+\gamma)^2(1-\rho)^2} & \textit{if } \gamma \geq 0, \\ \dfrac{2(3 - 4\gamma - 2\rho)}{(1-\gamma)(1-2\gamma)(1-\gamma-\rho)(1-2\gamma-\rho)} & \textit{if } \gamma < 0, \end{cases} \end{aligned}$$

$$B_{a/U}^{(1)}(\gamma) = -\frac{1}{(1-\gamma_-)(1-\gamma_- + |\gamma|)} = \begin{cases} -\dfrac{1}{1+\gamma} & \textit{if } \gamma \geq 0, \\ -\dfrac{1}{(1-\gamma)(1-2\gamma)} & \textit{if } \gamma < 0, \end{cases}$$

$$\begin{aligned} \textit{and } B_{a/U}^{(2)}(\gamma) &= -\frac{2(3+2\gamma_+)}{(1-\gamma_-)(1-\gamma_- + |\gamma|)(1-2\gamma_- + |\gamma|)} \\ &= \begin{cases} -\dfrac{2(3+2\gamma)}{(1+\gamma)^2} & \textit{if } \gamma \geq 0, \\ -\dfrac{6}{(1-\gamma)(1-2\gamma)(1-3\gamma)} & \textit{if } \gamma < 0, \end{cases} \end{aligned}$$

and the covariance matrix $\boldsymbol{V}$ is

$$\boldsymbol{V} = \begin{pmatrix} \dfrac{1}{(1-\gamma_-)^2(1-2\gamma_-)} & \dfrac{4}{(1-\gamma_-)^2(1-2\gamma_-)(1-3\gamma_-)} & 0 \\ \dfrac{4}{(1-\gamma_-)^2(1-2\gamma_-)(1-3\gamma_-)} & \dfrac{4(5-11\gamma_-)}{(1-\gamma_-)^2(1-2\gamma_-)^2(1-3\gamma_-)(1-4\gamma_-)} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Proof of Lemma A.8. Combine Corollary 2.2.2 p.41 in de Haan and Ferreira (2006), Proposition 1(i) and the assumption $\sqrt{k}A(n/k) = \mathrm{O}(1)$ to get

$$\sqrt{k}\frac{U(Y_{n-k:n}) - U(n/k)}{a(n/k)} = \sqrt{k}\int_1^{\frac{k}{n}Y_{n-k:n}} s^{\gamma-1}\mathrm{d}s + \mathrm{o}_{\mathbb{P}}(1) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1). \tag{A.58}$$

Write then

$$\frac{a(Y_{n-k:n})}{U(Y_{n-k:n})} - \frac{a(n/k)}{U(n/k)} = \frac{a(n/k)}{U(Y_{n-k:n})}\left(\frac{a(Y_{n-k:n})}{a(n/k)} - 1 - \frac{a(n/k)}{U(n/k)}\frac{U(Y_{n-k:n}) - U(n/k)}{a(n/k)}\right).$$

It follows from the fact that $\frac{k}{n}Y_{n-k:n} = 1 + \mathrm{o}_{\mathbb{P}}(1)$ and the regular variation property of U with index γ_+ (use Corollary 1.2.10.1 on p.23 in de Haan and Ferreira (2006) when $\gamma > 0$, Lemma 1.2.9.3 on p.22 in de Haan and Ferreira (2006) when $\gamma = 0$, and the convergence of $U(t)$ to $U(\infty) \in (0,\infty)$ when $\gamma < 0$) that $a(n/k)/U(Y_{n-k:n}) \xrightarrow{\mathbb{P}} \gamma_+$. Recalling (A.21) and combining Lemma A.5 for the function a with (A.58) yields

$$\frac{a(Y_{n-k:n})}{U(Y_{n-k:n})} - \frac{a(n/k)}{U(n/k)} = (\gamma_+ + \mathrm{o}_{\mathbb{P}}(1))\left(\left(\frac{k}{n}Y_{n-k:n}\right)^{\gamma} - 1 - \gamma_+\int_1^{\frac{k}{n}Y_{n-k:n}} s^{\gamma-1}\mathrm{d}s + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)\right).$$

Consequently, from (A.58) again,

$$\sqrt{k}\left(\frac{a(Y_{n-k:n})}{U(Y_{n-k:n})} - \gamma_+\right) = \sqrt{k}\left(\frac{a(n/k)}{U(n/k)} - \gamma_+\right) + \mathrm{o}_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} \mu. \tag{A.59}$$

Now, combining Proposition 2 with (A.17), (A.59) and the regular variation property of A, we find

$$\begin{aligned} &\sqrt{k}\left(\frac{U(Y_{n-k:n})}{a(Y_{n-k:n})}\mathfrak{M}_n^{(1)} - \frac{1}{1-\gamma_-}\right) \\ &= \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k}\int_1^{Y_{n-i+1:n}/Y_{n-k:n}} s^{\gamma_- -1}\mathrm{d}s - \frac{1}{1-\gamma_-}\right) \\ &- \sqrt{k}\left(\frac{a(n/k)}{U(n/k)} - \gamma_+\right)\mathbb{E}\left(\int_1^Y s^{\gamma_- -1}\left(\int_1^s u^{-|\gamma|-1}\mathrm{d}u\right)\mathrm{d}s\right) \\ &+ \sqrt{k}A(n/k)\mathbb{E}\left(Y^{-\gamma_+}\int_1^Y s^{\gamma-1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s\right) + \mathrm{o}_{\mathbb{P}}(1). \end{aligned} \tag{A.60}$$

Likewise, writing

$$\begin{aligned}
&\left(\frac{\log U(tx)-\log U(t)}{a(t)/U(t)}\right)^2-\left(\int_1^x s^{\gamma_--1}\mathrm{d}s\right)^2\\
&=2\left(\int_1^x s^{\gamma_--1}\mathrm{d}s\right)\left(\frac{\log U(tx)-\log U(t)}{a(t)/U(t)}-\int_1^x s^{\gamma_--1}\mathrm{d}s\right)\\
&+\left(\frac{\log U(tx)-\log U(t)}{a(t)/U(t)}-\int_1^x s^{\gamma_--1}\mathrm{d}s\right)^2
\end{aligned}$$

we find

$$\begin{aligned}
&\sqrt{k}\left(\left\{\frac{U(Y_{n-k:n})}{a(Y_{n-k:n})}\right\}^2\mathfrak{M}_n^{(2)}-\frac{2}{(1-\gamma_-)(1-2\gamma_-)}\right)\\
&=\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^k\left(\int_1^{Y_{n-i+1:n}/Y_{n-k:n}} s^{\gamma_--1}\mathrm{d}s\right)^2-\frac{2}{(1-\gamma_-)(1-2\gamma_-)}\right)\\
&-2\sqrt{k}\left(\frac{a(n/k)}{U(n/k)}-\gamma_+\right)\mathbb{E}\left(\left(\int_1^Y s^{\gamma_--1}\mathrm{d}s\right)\left\{\int_1^Y s^{\gamma_--1}\left(\int_1^s u^{-|\gamma|-1}\mathrm{d}u\right)\mathrm{d}s\right\}\right)\\
&+2\sqrt{k}A(n/k)\mathbb{E}\left(\left(\int_1^Y s^{\gamma_--1}\mathrm{d}s\right)Y^{-\gamma_+}\left\{\int_1^Y s^{\gamma-1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s\right\}\right)+\mathrm{o}_{\mathbb{P}}(1). \qquad \text{(A.61)}
\end{aligned}$$

Obviously

$$\begin{aligned}
&\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^k\int_1^{Y_{n-i+1:n}/Y_{n-k:n}} s^{\gamma_--1}\mathrm{d}s-\frac{1}{1-\gamma_-},\right.\\
&\left.\qquad\frac{1}{k}\sum_{i=1}^k\left(\int_1^{Y_{n-i+1:n}/Y_{n-k:n}} s^{\gamma_--1}\mathrm{d}s\right)^2-\frac{2}{(1-\gamma_-)(1-2\gamma_-)},\int_1^{\frac{k}{n}Y_{n-k:n}} s^{\gamma-1}\mathrm{d}s\right)\\
&\xrightarrow{\mathrm{d}}\mathcal{N}(\mathbf{0}_3,\boldsymbol{V}) \qquad \text{(A.62)}
\end{aligned}$$

by using (A.16) combined with the Cramér-Wold device, the fact that $\int_1^Y s^{\gamma_--1}\mathrm{d}s=\log Y$ is unit exponential when $\gamma\geq 0$, and the straightforward identity

$$\forall\alpha,\beta>0 \text{ with } (\alpha+\beta)\gamma<1,\ \operatorname{Cov}(Y^{\alpha\gamma},Y^{\beta\gamma})=\frac{\alpha\beta\gamma^2}{(1-\alpha\gamma)(1-\beta\gamma)(1-(\alpha+\beta)\gamma)},$$

as well as the standard central limit theorem, and Corollary 2.2.2 p.41 in de Haan and Ferreira (2006) for the asymptotic normality of the third component. From (A.62), it follows that it suffices to compute the bias terms in order to complete the proof. Using (A.19) twice, we get

$$\mathbb{E}\left(\int_1^Y s^{\gamma_--1}\left(\int_1^s u^{-|\gamma|-1}\mathrm{d}u\right)\mathrm{d}s\right)=\frac{1}{(1-\gamma_-)(1-\gamma_-+|\gamma|)}$$

and

$$\begin{aligned}
\mathbb{E}\left(Y^{-\gamma_+}\int_1^Y s^{\gamma-1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s\right)&=\frac{1}{(1+\gamma_+)(1+\gamma_+-\gamma)(1+\gamma_+-\gamma-\rho)}\\
&=\frac{1}{(1+|\gamma|)(1-\gamma_--\rho)}.
\end{aligned}$$

Using (A.20), we find after straightforward calculations that

$$\mathbb{E}\left(\left(\int_1^Y s^{\gamma_- -1}\mathrm{d}s\right)\left\{\int_1^Y s^{\gamma_- -1}\left(\int_1^s u^{-|\gamma|-1}\mathrm{d}u\right)\mathrm{d}s\right\}\right)$$
$$= \frac{3-7\gamma_- + 6\gamma_-^2 + 2|\gamma|}{(1-\gamma_-)^2(1-2\gamma_-)(1-\gamma_- + |\gamma|)(1-2\gamma_- + |\gamma|)}$$
$$= \frac{3+2\gamma_+}{(1-\gamma_-)(1-\gamma_- + |\gamma|)(1-2\gamma_- + |\gamma|)}$$

and

$$\mathbb{E}\left(\left(\int_1^Y s^{\gamma_- -1}\mathrm{d}s\right)Y^{-\gamma_+}\left\{\int_1^Y s^{\gamma -1}\left(\int_1^s u^{\rho-1}\mathrm{d}u\right)\mathrm{d}s\right\}\right)$$
$$= \frac{3+2\gamma_+ - 7\gamma_- + 4\gamma_-^2 - 2\rho - \gamma_+\rho + 2\gamma_-\rho}{(1+\gamma_+)^2(1-\gamma_-)^2(1-2\gamma_-)(1-\gamma_- -\rho)(1-2\gamma_- -\rho)}$$
$$= \frac{3+2\gamma_+ - 4\gamma_- - 2\rho - \gamma_+\rho}{(1+\gamma_+)(1+|\gamma|)(1-2\gamma_-)(1-\gamma_- -\rho)(1-2\gamma_- -\rho)}.$$

The proof is complete. □

Lemma A.9 (Joint weak convergence of $\widehat{\mathrm{ES}}_n(\tau_n)$ and the building blocks for the moment estimator in the short tail case)**.** *Suppose that X satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$ with $\gamma < 0$ and $0 < U(\infty) = q(1) < \infty$. Let $\tau_n \uparrow 1$ be such that $n(1-\tau_n) \to \infty$, $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda \in \mathbb{R}$ and $\sqrt{n(1-\tau_n)}\,a((1-\tau_n)^{-1})/q(\tau_n) \to \mu \in \mathbb{R}$.*

(i) One has

$$\sqrt{n(1-\tau_n)}\frac{\widehat{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})}$$
$$= \sqrt{n(1-\tau_n)}\left(\frac{X_{\lceil n\tau_n\rceil:n}M_n^{(1)}}{a((1-\tau_n)^{-1})} - \frac{1}{1-\gamma}\right) + \sqrt{n(1-\tau_n)}\frac{X_{\lceil n\tau_n\rceil:n} - q(\tau_n)}{a((1-\tau_n)^{-1})}$$
$$- \frac{\lambda}{(1-\gamma)(1-\gamma-\rho)} + \frac{\mu}{(1-\gamma)(1-2\gamma)} + \mathrm{o}_{\mathbb{P}}(1).$$

(ii) The following joint weak convergence result holds:

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})}, \frac{X_{\lceil n\tau_n\rceil:n}M_n^{(1)}}{a((1-\tau_n)^{-1})} - \frac{1}{1-\gamma}, \widehat{\gamma}_{n,-}^{\mathrm{Mom}} - \gamma, \frac{X_{\lceil n\tau_n\rceil:n} - q(\tau_n)}{a((1-\tau_n)^{-1})}\right)$$
$$\stackrel{\mathrm{d}}{\longrightarrow} \mathcal{N}\left(\left(0, \frac{\lambda}{(1-\gamma)(1-\gamma-\rho)} - \frac{\mu}{(1-\gamma)(1-2\gamma)}, \frac{\lambda(1-\gamma)(1-2\gamma)}{(1-\gamma-\rho)(1-2\gamma-\rho)} - \frac{\mu(1-\gamma)}{1-3\gamma}, 0\right), \mathbf{C}\right)$$

where $\mathbf{C} = \mathbf{C}(\gamma)$ *is the* 4×4 *positive semidefinite symmetric matrix whose entries are*

$$\begin{aligned}
C_{11} &= \frac{2}{(1-\gamma)(1-2\gamma)}, \ C_{12} = \frac{1+2\gamma}{(1-\gamma)(1-2\gamma)}, \ C_{13} = \frac{2\gamma}{1-3\gamma}, \ C_{14} = \frac{1}{1-\gamma}, \\
C_{22} &= \frac{1+\gamma+2\gamma^2}{(1-\gamma)(1-2\gamma)}, \ C_{23} = \frac{2\gamma}{1-3\gamma}, \ C_{24} = \frac{\gamma}{1-\gamma}, \\
C_{33} &= \frac{(1-\gamma)^2(1-2\gamma)(1-\gamma+6\gamma^2)}{(1-3\gamma)(1-4\gamma)}, \ C_{34} = 0 \ \ \textit{and} \ \ C_{44} = 1.
\end{aligned}$$

This results in particular corrects Corollary 3.5.6 p.108 in de Haan and Ferreira (2006), in which the reported asymptotic variance of $\widehat{\gamma}_{n,-}^{\mathrm{Mom}} - \gamma$ is unfortunately incorrect.

Proof of Lemma A.9. (i) From Propositions 1(i) and 2, and with the notation therein, we have, for any $\delta, \varepsilon > 0$ sufficiently small (in particular, we choose $\delta < 1$), that for t large enough and all $x > 1$,

$$\left| \frac{U(tx)-U(t)}{a(t)} - \frac{\log U(tx) - \log U(t)}{a(t)/U(t)} - \frac{a(t)}{U(t)} \frac{1}{2} \left(\frac{x^\gamma - 1}{\gamma} \right)^2 \right| \le \varepsilon \left(\frac{a(t)}{U(t)} + |A(t)| \right) x^\delta. \quad \text{(A.63)}$$

Besides, and as in the proof of Proposition A.1, it suffices to work in the case when $k = n(1-\tau_n)$ is a sequence of integers, with then

$$\frac{\widehat{\mathrm{ES}}_n(\tau_n) - X_{\lceil n\tau_n \rceil:n}}{a((1-\tau_n)^{-1})} \ \text{ replaced by } \ \frac{1}{k} \sum_{i=1}^{k} \frac{X_{n-i+1:n} - X_{n-k:n}}{a(n/k)}.$$

Let then (Y_i) be a sequence of independent unit Pareto random variables. Obviously

$$\begin{aligned}
&\sqrt{k} \left(\frac{1}{k} \sum_{i=1}^{k} \frac{X_{n-i+1:n} - X_{n-k:n}}{a(n/k)}, \frac{X_{n-k:n} M_n^{(1)}}{a(n/k)} \right) \\
&\stackrel{\mathrm{d}}{=} \sqrt{k} \frac{a(Y_{n-k:n})}{a(n/k)} \left(\frac{1}{k} \sum_{i=1}^{k} \frac{U(Y_{n-i+1:n}) - U(Y_{n-k:n})}{a(Y_{n-k:n})}, \frac{1}{k} \sum_{i=1}^{k} \frac{\log U(Y_{n-i+1:n}) - \log U(Y_{n-k:n})}{(a/U)(Y_{n-k:n})} \right)
\end{aligned}$$

so that

$$\begin{aligned}
&\sqrt{k} \left(\frac{1}{k} \sum_{i=1}^{k} \frac{X_{n-i+1:n} - X_{n-k:n}}{a(n/k)} - \frac{X_{n-k:n} M_n^{(1)}}{a(n/k)} \right) \\
&\stackrel{\mathrm{d}}{=} \frac{a(Y_{n-k:n})}{a(n/k)} \left(\sqrt{k} \frac{a(Y_{n-k:n})}{U(Y_{n-k:n})} \times \frac{1}{k} \sum_{i=1}^{k} \frac{1}{2} \left(\frac{(Y_{n-i+1:n}/Y_{n-k:n})^\gamma - 1}{\gamma} \right)^2 + \mathrm{o}_{\mathbb{P}}(1) \right) \\
&\stackrel{\mathrm{d}}{=} \frac{\mu}{(1-\gamma)(1-2\gamma)} + \mathrm{o}_{\mathbb{P}}(1)
\end{aligned}$$

by Equations (A.17), (A.18) and the uniform inequality (A.63), along with the convergence $\frac{k}{n} Y_{n-k:n} \stackrel{\mathbb{P}}{\longrightarrow} 1$ combined with the regular variation properties of a, U and A (see de Haan and Ferreira, 2006, Lemma 1.2.9 and Corollary 1.2.10 pp.22-23). This yields

$$\frac{\widehat{\mathrm{ES}}_n(\tau_n) - X_{\lceil n\tau_n\rceil:n}}{a((1-\tau_n)^{-1})} - \frac{X_{\lceil n\tau_n\rceil:n} M_n^{(1)}}{a((1-\tau_n)^{-1})} = \frac{1}{\sqrt{n(1-\tau_n)}}\left(\frac{\mu}{(1-\gamma)(1-2\gamma)} + \mathrm{o}_{\mathbb{P}}(1)\right). \tag{A.64}$$

To conclude the proof of (i), write

$$\begin{aligned}
&\sqrt{n(1-\tau_n)}\frac{\widehat{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} \\
&= \sqrt{n(1-\tau_n)}\left(\frac{\widehat{\mathrm{ES}}_n(\tau_n) - X_{\lceil n\tau_n\rceil:n}}{a((1-\tau_n)^{-1})} - \frac{1}{1-\gamma}\right) + \sqrt{n(1-\tau_n)}\frac{X_{\lceil n\tau_n\rceil:n} - q(\tau_n)}{a((1-\tau_n)^{-1})} \\
&- \sqrt{n(1-\tau_n)}\left(\frac{\mathrm{ES}(\tau_n) - q(\tau_n)}{a((1-\tau_n)^{-1})} - \frac{1}{1-\gamma}\right) \\
&= \sqrt{n(1-\tau_n)}\left(\frac{X_{\lceil n\tau_n\rceil:n} M_n^{(1)}}{a((1-\tau_n)^{-1})} - \frac{1}{1-\gamma}\right) + \sqrt{n(1-\tau_n)}\frac{X_{\lceil n\tau_n\rceil:n} - q(\tau_n)}{a((1-\tau_n)^{-1})} \\
&- \frac{\lambda}{(1-\gamma)(1-\gamma-\rho)} + \frac{\mu}{(1-\gamma)(1-2\gamma)} + \mathrm{o}_{\mathbb{P}}(1)
\end{aligned}$$

by (A.64) and Proposition 1(ii).

(ii) Using (i), straightforward calculations show that it suffices to prove that, again when $k = n(1-\tau_n)$ is a sequence of integers,

$$\begin{aligned}
&\sqrt{k}\left(\frac{X_{n-k:n}}{a(n/k)}M_n^{(1)} - \frac{1}{1-\gamma}, \left\{\frac{X_{n-k:n}}{a(n/k)}\right\}^2 M_n^{(2)} - \frac{2}{(1-\gamma)(1-2\gamma)}, \frac{X_{n-k:n} - q(1-k/n)}{a(n/k)}\right) \\
&\xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\frac{\lambda}{(1-\gamma)(1-\gamma-\rho)} - \frac{\mu}{(1-\gamma)(1-2\gamma)},\right.\right. \\
&\quad \left.\frac{2(3-4\gamma-2\rho)\lambda}{(1-\gamma)(1-2\gamma)(1-\gamma-\rho)(1-2\gamma-\rho)} - \frac{6\mu}{(1-\gamma)(1-2\gamma)(1-3\gamma)}, 0\right), \\
&\quad \left.\begin{pmatrix} \frac{1+\gamma+2\gamma^2}{(1-\gamma)(1-2\gamma)} & \frac{4(1+\gamma^2-3\gamma^3)}{(1-\gamma)^2(1-2\gamma)(1-3\gamma)} & \frac{\gamma}{1-\gamma} \\ \frac{4(1+\gamma^2-3\gamma^3)}{(1-\gamma)^2(1-2\gamma)(1-3\gamma)} & \frac{4(5-\gamma+2\gamma^2-24\gamma^3)}{(1-\gamma)^2(1-2\gamma)(1-3\gamma)(1-4\gamma)} & \frac{4\gamma}{(1-\gamma)(1-2\gamma)} \\ \frac{\gamma}{1-\gamma} & \frac{4\gamma}{(1-\gamma)(1-2\gamma)} & 1 \end{pmatrix}\right).
\end{aligned} \tag{A.65}$$

Indeed, if this convergence is true, then a simple linearization of $\widehat{\gamma}_{n,-}^{\mathrm{Mom}} - \gamma$ yields

$$\begin{aligned}
\widehat{\gamma}_{n,-}^{\mathrm{Mom}} - \gamma = (1-\gamma)^2(1-2\gamma)&\left\{-2\left(\frac{X_{n-k:n}}{a(n/k)}M_n^{(1)} - \frac{1}{1-\gamma}\right)\right. \\
&\left.+\frac{1-2\gamma}{2}\left(\left\{\frac{X_{n-k:n}}{a(n/k)}\right\}^2 M_n^{(2)} - \frac{2}{(1-\gamma)(1-2\gamma)}\right)\right\} + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)
\end{aligned}$$

and straightforward matrix algebra gives the desired convergence. With the notation of Lemma A.8, the random vector on the left-hand side of (A.65) has the same distribution as

$$\sqrt{k}\left(\frac{U(Y_{n-k:n})}{a(n/k)}\mathfrak{M}_n^{(1)}-\frac{1}{1-\gamma},\left\{\frac{U(Y_{n-k:n})}{a(n/k)}\right\}^2\mathfrak{M}_n^{(2)}-\frac{2}{(1-\gamma)(1-2\gamma)},\frac{U(Y_{n-k:n})-U(n/k)}{a(n/k)}\right).$$

Now, by Lemma A.5 applied to the function a (which satisfies the required second-order inequality, see (A.21)), Lemma A.8 and (A.58),

$$\begin{aligned}
&\sqrt{k}\left(\frac{U(Y_{n-k:n})}{a(n/k)}\mathfrak{M}_n^{(1)}-\frac{1}{1-\gamma}\right)\\
&=\sqrt{k}\left(\frac{U(Y_{n-k:n})}{a(Y_{n-k:n})}\mathfrak{M}_n^{(1)}-\frac{1}{1-\gamma}\right)+\sqrt{k}\left(\frac{a(Y_{n-k:n})}{a(n/k)}-1\right)\frac{U(Y_{n-k:n})}{a(Y_{n-k:n})}\mathfrak{M}_n^{(1)}\\
&=\sqrt{k}\left(\frac{U(Y_{n-k:n})}{a(Y_{n-k:n})}\mathfrak{M}_n^{(1)}-\frac{1}{1-\gamma}\right)+\frac{\gamma}{1-\gamma}\times\sqrt{k}\frac{U(Y_{n-k:n})-U(n/k)}{a(n/k)}+\mathrm{o}_{\mathbb{P}}(1).
\end{aligned}$$

Likewise,

$$\begin{aligned}
&\sqrt{k}\left(\left\{\frac{U(Y_{n-k:n})}{a(n/k)}\right\}^2\mathfrak{M}_n^{(2)}-\frac{2}{(1-\gamma)(1-2\gamma)}\right)\\
&=\sqrt{k}\left(\left\{\frac{U(Y_{n-k:n})}{a(Y_{n-k:n})}\right\}^2\mathfrak{M}_n^{(2)}-\frac{2}{(1-\gamma)(1-2\gamma)}\right)\\
&+\sqrt{k}\left(\frac{a(Y_{n-k:n})}{a(n/k)}-1\right)\left(\frac{a(Y_{n-k:n})}{a(n/k)}+1\right)\left\{\frac{U(Y_{n-k:n})}{a(Y_{n-k:n})}\right\}^2\mathfrak{M}_n^{(2)}\\
&=\sqrt{k}\left(\left\{\frac{U(Y_{n-k:n})}{a(Y_{n-k:n})}\right\}^2\mathfrak{M}_n^{(2)}-\frac{2}{(1-\gamma)(1-2\gamma)}\right)\\
&+\frac{4\gamma}{(1-\gamma)(1-2\gamma)}\times\sqrt{k}\frac{U(Y_{n-k:n})-U(n/k)}{a(n/k)}+\mathrm{o}_{\mathbb{P}}(1).
\end{aligned}$$

Convergence (A.65) now follows immediately from Lemma A.8 and straightforward, if burdensome, calculations. □

Our final auxiliary result will be useful when establishing the asymptotic properties of moment-based extrapolated Expected Shortfall estimators. This lemma can be seen as a complement to Theorem 3.5.4 p.104 in de Haan and Ferreira (2006), because it provides a unified representation of the bias term of $\widehat{\gamma}_n^{\mathrm{Mom}}$.

Lemma A.10 (Convergence of $\widehat{\gamma}_n^{\mathrm{Mom}}$)**.** *Suppose that X satisfies condition $\mathcal{C}_2(\gamma,a,\rho,A)$ with $0<U(\infty)=q(1)\le\infty$. Let $\tau_n\uparrow 1$ be such that $n(1-\tau_n)\to\infty$, $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1})\to\lambda\in\mathbb{R}$ and $\sqrt{n(1-\tau_n)}(a((1-\tau_n)^{-1})/q(\tau_n)-\gamma_+)\to\mu\in\mathbb{R}$. Then $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_n^{\mathrm{Mom}}-\gamma)$ is*

asymptotically Gaussian with mean $\lambda\, b_1(\gamma,\rho) + \mu\, b_2(\gamma)$, *where*

$$b_1(\gamma,\rho) = \begin{cases} \dfrac{1+\gamma+\gamma^2-\rho\gamma^2}{(1+\gamma)^2(1-\rho)^2} & \textit{if } \gamma \geq 0, \\ \dfrac{(1-\gamma)(1-2\gamma)}{(1-\gamma-\rho)(1-2\gamma-\rho)} & \textit{if } \gamma < 0, \end{cases}$$

$$\textit{and } b_2(\gamma) = \begin{cases} \dfrac{\gamma}{(1+\gamma)^2} & \textit{if } \gamma \geq 0, \\ -\dfrac{\gamma(1+\gamma)}{(1-\gamma)(1-3\gamma)} & \textit{if } \gamma < 0, \end{cases}$$

and variance

$$\begin{cases} \gamma^2+1 & \textit{if } \gamma \geq 0, \\ \dfrac{(1-\gamma)^2(1-2\gamma)(1-\gamma+6\gamma^2)}{(1-3\gamma)(1-4\gamma)} & \textit{if } \gamma < 0. \end{cases}$$

Proof of Lemma A.10. Recall that $\widehat{\gamma}_n^{\mathrm{Mom}} = M_n^{(1)} + \widehat{\gamma}_{n,-}^{\mathrm{Mom}}$. Set $k = \lfloor n(1-\tau_n) \rfloor$: in particular $k \to \infty$, $k/n \to 0$, $\sqrt{k}A(n/k) \to \lambda$ (because of the regular variation property of A) and $\sqrt{k}(a(n/k)/U(n/k) - \gamma_+) \to \mu$ (using Equations (A.22) and (A.23)). Note that

$$(M_n^{(1)} - \gamma_+, \widehat{\gamma}_{n,-}^{\mathrm{Mom}} - \gamma_-) \stackrel{\mathrm{d}}{=} \left(\frac{a(Y_{n-k:n})}{U(Y_{n-k:n})} \overline{\mathfrak{M}}_n^{(1)} - \gamma_+, 1 - \frac{1}{2} \left(1 - \frac{(\overline{\mathfrak{M}}_n^{(1)})^2}{\overline{\mathfrak{M}}_n^{(2)}} \right)^{-1} - \gamma_- \right)$$

where, letting the Y_i be independent copies of a unit Pareto random variable Y,

$$\overline{\mathfrak{M}}_n^{(j)} = \left\{ \frac{U(Y_{n-k:n})}{a(Y_{n-k:n})} \right\}^j \times \frac{1}{k} \sum_{i=1}^{k} \log^j \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})}.$$

Recalling (A.59), a straightforward linearization then yields

$$\begin{aligned} &(M_n^{(1)} - \gamma_+, \widehat{\gamma}_{n,-}^{\mathrm{Mom}} - \gamma_-) \stackrel{\mathrm{d}}{=} \left(\gamma_+ \left(\overline{\mathfrak{M}}_n^{(1)} - \frac{1}{1-\gamma_-} \right) + \frac{1}{\sqrt{k}} \frac{\mu}{1-\gamma_-}, \right. \\ &\left. (1-\gamma_-)^2(1-2\gamma_-) \left\{ -2 \left(\overline{\mathfrak{M}}_n^{(1)} - \frac{1}{1-\gamma_-} \right) + \frac{1-2\gamma_-}{2} \left(\overline{\mathfrak{M}}_n^{(2)} - \frac{2}{(1-\gamma_-)(1-2\gamma_-)} \right) \right\} \right) \\ &+ \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}} \right). \end{aligned} \tag{A.66}$$

Conclude that

$$\sqrt{n(1-\tau_n)}(\widehat{\gamma}_n^{\mathrm{Mom}} - \gamma) = \frac{\mu}{1-\gamma} + \sqrt{n(1-\tau_n)}(\widehat{\gamma}_{n,-}^{\mathrm{Mom}} - \gamma) + \mathrm{o}_{\mathbb{P}}(1) \tag{A.67}$$

when $\gamma < 0$, so that the result is an immediate consequence of Lemma A.9 in this case. When $\gamma \geq 0$, Equation (A.66) yields

$$\sqrt{n(1-\tau_n)}(\widehat{\gamma}_n^{\mathrm{Mom}} - \gamma) \stackrel{\mathrm{d}}{=} (\gamma - 2)\left(\overline{\mathfrak{M}}_n^{(1)} - 1 \right) + \frac{1}{2}\left(\overline{\mathfrak{M}}_n^{(2)} - 2 \right) + \mu + \mathrm{o}_{\mathbb{P}}(1)$$

and the result follows from Lemma A.8 and straightforward calculations. □

Proof of Theorem 4. We consider three cases: $\gamma < 0$, then $0 \le \gamma < 1/2$ and finally $1/2 \le \gamma < 1$. As a preliminary step we note that when $\gamma \ge 0$ and $\rho < 0$ one has, according to Theorem 4.2.1 p.131 in de Haan and Ferreira (2006),

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} - 1\right) = \mathrm{O}_{\mathbb{P}}(1). \tag{A.68}$$

Case $\gamma < 0$: Recall that

$$\widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1}) = X_{\lceil n\tau_n\rceil:n} M_n^{(1)} (1-\widehat{\gamma}_{n,-}^{\mathrm{Mom}}).$$

It immediately follows from Lemma A.9 that

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} - 1\right) = (1-\gamma)\times\sqrt{n(1-\tau_n)}\left(\frac{X_{\lceil n\tau_n\rceil:n} M_n^{(1)}}{a((1-\tau_n)^{-1})} - \frac{1}{1-\gamma}\right)$$
$$- \frac{1}{1-\gamma}\times\sqrt{n(1-\tau_n)}(\widehat{\gamma}_{n,-}^{\mathrm{Mom}} - \gamma) + \mathrm{o}_{\mathbb{P}}(1). \tag{A.69}$$

Recalling (A.67) and applying Lemma A.9 then yields, after straightforward matrix algebra,

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})}, \frac{\widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} - 1, \widehat{\gamma}_n^{\mathrm{Mom}} - \gamma\right) \stackrel{\mathrm{d}}{\longrightarrow} (Z_{\mathrm{loc}}, Z_{\mathrm{scale}}, Z_{\mathrm{shape}})$$

where the vector $(Z_{\mathrm{loc}}, Z_{\mathrm{scale}}, Z_{\mathrm{shape}})$ is trivariate Gaussian with mean vector defined as

$$\begin{aligned}
\mathbb{E}(Z_{\mathrm{loc}}) &= 0,\\
\mathbb{E}(Z_{\mathrm{scale}}) &= -\lambda\frac{\rho}{(1-\gamma-\rho)(1-2\gamma-\rho)} + \mu\frac{\gamma}{(1-2\gamma)(1-3\gamma)},\\
\text{and } \mathbb{E}(Z_{\mathrm{shape}}) &= \lambda\frac{(1-\gamma)(1-2\gamma)}{(1-\gamma-\rho)(1-2\gamma-\rho)} - \mu\frac{\gamma(1+\gamma)}{(1-\gamma)(1-3\gamma)},
\end{aligned}$$

and with covariance matrix

$$\begin{pmatrix}
\frac{2}{(1-\gamma)(1-2\gamma)} & \frac{1-4\gamma-\gamma^2+6\gamma^3}{(1-\gamma)(1-2\gamma)(1-3\gamma)} & \frac{2\gamma}{1-3\gamma}\\
\frac{1-4\gamma-\gamma^2+6\gamma^3}{(1-\gamma)(1-2\gamma)(1-3\gamma)} & \frac{2-16\gamma+51\gamma^2-69\gamma^3+50\gamma^4-24\gamma^5}{(1-2\gamma)(1-3\gamma)(1-4\gamma)} & -\frac{(1-\gamma)^2(1-4\gamma+12\gamma^2)}{(1-3\gamma)(1-4\gamma)}\\
\frac{2\gamma}{1-3\gamma} & -\frac{(1-\gamma)^2(1-4\gamma+12\gamma^2)}{(1-3\gamma)(1-4\gamma)} & \frac{(1-\gamma)^2(1-2\gamma)(1-\gamma+6\gamma^2)}{(1-3\gamma)(1-4\gamma)}
\end{pmatrix}.$$

The result now follows from a direct application of Corollary 1.

Case $0 \le \gamma < 1/2$: In this setting, the result is an immediate consequence of Corollary 2, whose assumptions are fulfilled thanks to Theorem 1, (A.68) and Lemma A.10.

Case $1/2 \le \gamma < 1$: Let, as in the statement of Theorem 3, $d_n = (1-\tau_n)/(1-\tau_n')$. By Equation (A.31),

$$\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s = \frac{1}{\gamma}d_n^{\gamma}\log(d_n)(1+\mathrm{o}(1))$$

and since $1/d_n = \mathrm{O}(1/(n(1-\tau_n))) \to 0$, one obviously has

$$\frac{1}{\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s} = \mathrm{o}(d_n^{-\gamma}) = \mathrm{o}((n(1-\tau_n))^{-\gamma}).$$

Consequently $\sqrt{n(1-\tau_n)}/\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s \to 0$. Combining Theorems 1 and 3 with (A.68) and Lemma A.10, we get

$$\sqrt{n(1-\tau_n)}\frac{\widehat{\mathrm{ES}}_n^\star(\tau_n') - \mathrm{ES}(\tau_n')}{a((1-\tau_n)^{-1})\int_1^{d_n} s^{\gamma-1}\log(s)\mathrm{d}s} = \frac{1}{1-\gamma}\sqrt{n(1-\tau_n)}(\widehat{\gamma}_n - \gamma) + \mathrm{o}_{\mathbb{P}}(1).$$

The conclusion again follows from Lemma A.10. □

Proof of Theorem 5. We split the proof into two parts: we first consider the case $\gamma < 0$ and then the case $0 \le \gamma < 1$. As a starting point we note that in general, if $\widetilde{\mathrm{ES}}_n(\tau_n)$ is defined as

$$\widetilde{\mathrm{ES}}_n(\tau_n) = X_{\lceil n\tau_n\rceil:n} + \widehat{a}_n((1-\tau_n)^{-1})\frac{1}{1-\widehat{\gamma}_n},$$

where it is assumed that

$$\sqrt{n(1-\tau_n)}\left(\frac{X_{\lceil n\tau_n\rceil:n} - q(\tau_n)}{a((1-\tau_n)^{-1})}, \frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} - 1, \widehat{\gamma}_n - \gamma\right) \xrightarrow{\mathrm{d}} (N_{\mathrm{loc}}, N_{\mathrm{scale}}, N_{\mathrm{shape}}),$$

in which the trivariate random vector $(N_{\mathrm{loc}}, N_{\mathrm{scale}}, N_{\mathrm{shape}})$ has a nondegenerate distribution, then a simple combination of Theorem 2 and Corollary 1 entails

$$\sqrt{n(1-\tau_n)}\frac{\widetilde{\mathrm{ES}}_n^\star(\tau_n') - \mathrm{ES}(\tau_n')}{a((1-\tau_n)^{-1})\int_1^{(1-\tau_n)/(1-\tau_n')} s^{\gamma-1}\log(s)\mathrm{d}s}$$
$$\xrightarrow{\mathrm{d}} \gamma_-^2 N_{\mathrm{loc}} - \gamma_- N_{\mathrm{scale}} + \frac{1}{1-\gamma_+}N_{\mathrm{shape}} - \lambda\frac{\gamma_-}{\gamma_- + \rho}. \tag{A.70}$$

Case $\gamma < 0$: Use Lemma A.9 in conjunction with (A.67) and (A.69) to obtain that

$$\sqrt{n(1-\tau_n)}\left(\frac{X_{\lceil n\tau_n\rceil:n} - q(\tau_n)}{a((1-\tau_n)^{-1})}, \frac{\widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} - 1, \widehat{\gamma}_n^{\mathrm{Mom}} - \gamma\right) \xrightarrow{\mathrm{d}} (N_{\mathrm{loc}}, N_{\mathrm{scale}}, N_{\mathrm{shape}})$$

where the vector $(N_{\mathrm{loc}}, N_{\mathrm{scale}}, N_{\mathrm{shape}})$ is trivariate Gaussian with mean vector defined as

$$\begin{aligned}
\mathbb{E}(N_{\mathrm{loc}}) &= 0,\\
\mathbb{E}(N_{\mathrm{scale}}) &= -\lambda\frac{\rho}{(1-\gamma-\rho)(1-2\gamma-\rho)} + \mu\frac{\gamma}{(1-2\gamma)(1-3\gamma)},\\
\text{and } \mathbb{E}(N_{\mathrm{shape}}) &= \lambda\frac{(1-\gamma)(1-2\gamma)}{(1-\gamma-\rho)(1-2\gamma-\rho)} - \mu\frac{\gamma(1+\gamma)}{(1-\gamma)(1-3\gamma)},
\end{aligned}$$

and with covariance matrix

$$\begin{pmatrix} 1 & \gamma & 0 \\ \gamma & \frac{2-16\gamma+51\gamma^2-69\gamma^3+50\gamma^4-24\gamma^5}{(1-2\gamma)(1-3\gamma)(1-4\gamma)} & -\frac{(1-\gamma)^2(1-4\gamma+12\gamma^2)}{(1-3\gamma)(1-4\gamma)} \\ 0 & -\frac{(1-\gamma)^2(1-4\gamma+12\gamma^2)}{(1-3\gamma)(1-4\gamma)} & \frac{(1-\gamma)^2(1-2\gamma)(1-\gamma+6\gamma^2)}{(1-3\gamma)(1-4\gamma)} \end{pmatrix}.$$

Combine this convergence with (A.70) to obtain the result after straightforward calculations.

Case $0 \le \gamma < 1$: It is a particular consequence of Lemma A.8 that

$$\sqrt{n(1-\tau_n)} \frac{X_{\lceil n\tau_n \rceil:n} - q(\tau_n)}{a((1-\tau_n)^{-1})} = \mathrm{O}_{\mathbb{P}}(1). \tag{A.71}$$

Combine (A.68), (A.71) and Lemma A.10 to get

$$\sqrt{n(1-\tau_n)} \frac{\widetilde{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} = \mathrm{O}_{\mathbb{P}}(1) \text{ and } \sqrt{n(1-\tau_n)} \left(\frac{\widehat{a}_n^{\mathrm{Mom}}((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} - 1 \right) = \mathrm{O}_{\mathbb{P}}(1).$$

Use this in conjunction with Lemma A.10 and Corollary 2 to obtain the desired result. □

B Further details about the construction and finite-sample behavior of the corrected inference procedures

B.1 The `Expectrem` R package

We have implemented our methods in an `R` package called `Expectrem`, freely available on GitHub[2]. The objective of this package is to provide methods for the estimation of certain expectation-type quantities, such as expectiles and the Expected Shortfall. We have added the following objects to the existing version of this `R` package:

- `momentindex`: For a dataset `X` given in input, this function returns the moment estimator $\widehat{\gamma}_n^{\mathrm{Mom}}$ using the k_n (argument `k`) top log-spacings. The asymptotic Gaussian confidence interval (with confidence level `ci.level`, default 0.95) derived from Theorem 3.5.4 p.104 in de Haan and Ferreira (2006) is also returned.

- `extES`: At a given probability level τ_n' (argument `tau`), and for a dataset `X` given in input, compute an extrapolated estimator of the Expected Shortfall at level τ_n' using the k_n top order statistics (argument `k`). Setting the argument `estim` to `"Hill"` returns the estimator $\widehat{\mathrm{ES}}_n^{\mathrm{W}}(\tau_n')$ if `method="direct"`, and $\widetilde{\mathrm{ES}}_n^{\mathrm{W}}(\tau_n')$ if `method="indirect"`. Setting the argument `estim` to `"Moment"` leads to $\widehat{\mathrm{ES}}_n^{\star}(\tau_n')$ if `method="direct"`, and to $\widetilde{\mathrm{ES}}_n^{\star}(\tau_n')$ if `method="indirect"`. The associated confidence intervals at an asymptotic confidence level α (argument `ci.level`, default 0.95), which are $\widehat{\mathrm{I}}_1^{\star}(\alpha)$, $\widetilde{\mathrm{I}}_1^{\star}(\alpha)$, $\widehat{\mathrm{I}}_4^{\star}(\alpha)$ and $\widetilde{\mathrm{I}}_4^{\star}(\alpha)$, respectively, are also returned.

- `ESexp`, `ESfrechet`, `ESgumbel`, `ESkumaraswamy`, `ESlnorm`, `ESlogis`, `ESnorm`, `ESpareto`, `ESrev_frechet`, `ESt`, `ESweibull`: These functions give the values of the Expected Shortfall (at a vector of levels specified by the argument `probs`) of the exponential, Fréchet,

[2] `https://github.com/AntoineUC/Expectrem`

Gumbel, Kumaraswamy, log-normal, logistic, Gaussian, Pareto, reverse-Fréchet, Student and Weibull distributions.

- `flood_data`: The OpenFEMA flood claim amounts dataset used in Section 5.2.

B.2 Inference at intermediate levels

We consider here the intermediate setting when $\tau_n = 1 - k_n/n$ with $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. In this situation, $\mathrm{ES}(\tau_n)$ may be estimated either by the empirical Expected Shortfall $\widehat{\mathrm{ES}}_n(\tau_n)$, or by its quantile-based semiparametric estimator $\widetilde{\mathrm{ES}}_n(\tau_n)$. To the best of our knowledge, the literature on $\widehat{\mathrm{ES}}_n(\tau_n)$ has previously focused on the case when X is heavy-tailed and has a finite variance, that is, $\gamma \in (0, 1/2)$. In this case, Remark 1 suggests the following (competitor) confidence interval at level $1 - \alpha$:

$$\widehat{\mathrm{I}}_1(\alpha) = \left[\widehat{\mathrm{ES}}_n(1 - k_n/n) \exp\left(\pm \frac{\widehat{\gamma}_n^{\mathrm{H}}}{\sqrt{k_n}} \sqrt{\frac{2(1 - \widehat{\gamma}_n^{\mathrm{H}})}{1 - 2\widehat{\gamma}_n^{\mathrm{H}}}} z_{1-\alpha/2}\right)\right],$$

where $z_{1-\alpha/2}$ is the quantile of level $1-\alpha/2$ of the standard normal distribution, and $\widehat{\gamma}_n^{\mathrm{H}} = M_n^{(1)}$ is the Hill (1975) estimator. This confidence interval is naturally restricted to the situation $\gamma \in (0, 1/2)$. A more general inference procedure can be obtained as follows: note that up to a negligible bias term,

$$\frac{\widehat{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} \approx \frac{1}{k_n} \sum_{i=1}^{k_n} \frac{X_{n-i+1:n} - X_{n-k_n:n}}{a(n/k_n)} - \frac{1}{1-\gamma} + \frac{X_{n-k_n:n} - q(1 - k_n/n)}{a(n/k_n)}.$$

Therefore

$$\begin{aligned} \frac{\widehat{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} &\overset{\mathrm{d}}{\approx} \frac{a(Y_{n-k_n:n})}{a(n/k_n)} \times \frac{1}{k_n} \sum_{i=1}^{k_n} \left(\frac{U(Y_{n-i+1:n}) - U(Y_{n-k_n:n})}{a(Y_{n-k_n:n})} - \frac{1}{1-\gamma} \right) \\ &\quad + \frac{1}{1-\gamma} \left(\frac{a(Y_{n-k_n:n})}{a(n/k_n)} - 1 \right) + \frac{U(Y_{n-k_n:n}) - U(n/k_n)}{a(n/k_n)}. \end{aligned}$$

Recall that Y is unit Pareto if and only if $\log(Y)$ is unit exponential. Using the regular variation property of the function a, condition $\mathcal{C}_2(\gamma, a, \rho, A)$ and the Rényi representation (see (A.15) and (A.16) in the Appendix), we get, with the notation $D_r(x) = \int_1^x s^{r-1}\mathrm{d}s$,

$$\begin{aligned}\frac{\widehat{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{a((1-\tau_n)^{-1})} &\overset{\mathrm{d}}{\approx} \left(\frac{k_n}{n}Y_{n-k_n:n}\right)^{\gamma} \times \frac{1}{k_n}\sum_{i=1}^{k_n}\left(D_\gamma(Y_{n-i+1:n}/Y_{n-k_n:n}) - \frac{1}{1-\gamma}\right) \\ &+ \frac{1}{1-\gamma}\left(\left(\frac{k_n}{n}Y_{n-k_n:n}\right)^{\gamma} - 1\right) + D_\gamma\left(\frac{k_n}{n}Y_{n-k_n:n}\right) \\ &\overset{\mathrm{d}}{=} \left(\frac{k_n}{n}\exp\left(\sum_{i=k_n+1}^{n}\frac{\log(Y_i)}{i}\right)\right)^{\gamma}\left(\frac{1}{k_n}\sum_{i=1}^{k_n}D_\gamma(Y_i) - \frac{1}{1-\gamma}\right) \\ &+ \frac{1}{1-\gamma}D_\gamma\left(\frac{k_n}{n}\exp\left(\sum_{i=k_n+1}^{n}\frac{\log(Y_i)}{i}\right)\right) =: -G_n^{(1)}(Y_1,\ldots,Y_n,\gamma).\end{aligned}$$

Besides

$$\frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} = \frac{X_{n-k_n:n} - U(n/k_n)}{a(n/k_n)}M_n^{(1)}(1-\widehat{\gamma}_{n,-}) + \left\{\frac{U(n/k_n)}{a(n/k_n)}M_n^{(1)}\right\}(1-\widehat{\gamma}_{n,-})$$

and as such, it follows from (A.58), (A.60), (A.61) and the convergence $M_n^{(1)}(1-\widehat{\gamma}_{n,-}) \xrightarrow{\mathbb{P}} \gamma_+$ that

$$\begin{aligned}\frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} &\overset{\mathrm{d}}{\approx} \gamma_+ D_\gamma\left(\frac{k_n}{n}Y_{n-k_n:n}\right) \\ &+ \frac{1}{2}\left(\frac{1}{k_n}\sum_{i=1}^{k_n}D_{\gamma_-}(Y_{n-i+1:n}/Y_{n-k_n:n})\right)\left(1 - \frac{\left(\frac{1}{k_n}\sum_{i=1}^{k_n}D_{\gamma_-}(Y_{n-i+1:n}/Y_{n-k_n:n})\right)^2}{\frac{1}{k_n}\sum_{i=1}^{k_n}(D_{\gamma_-}(Y_{n-i+1:n}/Y_{n-k_n:n}))^2}\right)^{-1}.\end{aligned}$$

Conclude by the Rényi representation again (with the same variables Y_i) that

$$\begin{aligned}\frac{\widehat{a}_n((1-\tau_n)^{-1})}{a((1-\tau_n)^{-1})} &\overset{\mathrm{d}}{\approx} \gamma_+ D_\gamma\left(\frac{k_n}{n}\exp\left(\sum_{i=k_n+1}^{n}\frac{\log(Y_i)}{i}\right)\right) \\ &+ \frac{1}{2}\left(\frac{1}{k_n}\sum_{i=1}^{k_n}D_{\gamma_-}(Y_i)\right)\left(1 - \frac{\left(\frac{1}{k_n}\sum_{i=1}^{k_n}D_{\gamma_-}(Y_i)\right)^2}{\frac{1}{k_n}\sum_{i=1}^{k_n}(D_{\gamma_-}(Y_i))^2}\right)^{-1} =: G_n^{(2)}(Y_1,\ldots,Y_n,\gamma)\end{aligned}$$

and therefore

$$\frac{\widehat{\mathrm{ES}}_n(\tau_n) - \mathrm{ES}(\tau_n)}{\widehat{a}_n((1-\tau_n)^{-1})} \overset{\mathrm{d}}{\approx} -\frac{G_n^{(1)}(Y_1,\ldots,Y_n,\gamma)}{G_n^{(2)}(Y_1,\ldots,Y_n,\gamma)} =: -G_n(Y_1,\ldots,Y_n,\gamma).$$

For any $\gamma < 1$, the random quantity $G_n(Y_1,\ldots,Y_n,\gamma)$ is straightforward to simulate, meaning that its distribution can be tabulated. This leads to the following confidence interval for $\mathrm{ES}(1-k_n/n)$:

$$\widehat{\mathrm{I}}_2(\alpha) = \left[\widehat{\mathrm{ES}}_n(1-k_n/n) + \widehat{a}_n(n/k_n)g_{n,\alpha/2}(\widehat{\gamma}_n),\, \widehat{\mathrm{ES}}_n(1-k_n/n) + \widehat{a}_n(n/k_n)g_{n,1-\alpha/2}(\widehat{\gamma}_n)\right]$$

where $g_{n,\tau}(\gamma)$ is the τth quantile of $G_n(Y_1,\ldots,Y_n,\gamma)$. Our experience with this interval is that it provides excellent results for $\gamma \leq 1/4$, but when $\gamma > 1/4$, its performance is adversely affected by the finite-sample uncertainty in the plug-in step of replacing γ by $\widehat{\gamma}_n$. To put it differently, for large positive values of γ, the distributions of $G_n(Y_1,\ldots,Y_n,\gamma)$ and $G_n(Y_1,\ldots,Y_n,\widehat{\gamma}_n)$ may look substantially different in finite samples. Since $\sqrt{k_n}(\widehat{\gamma}_n-\gamma) \overset{\mathrm{d}}{\approx} \mathcal{N}(0,v_2(\gamma))$, we propose to deal with this issue of uncertainty quantification by computing directly the quantiles of $G_n(Y_1,\ldots,Y_n,\widetilde{\gamma}_n)$ for $\widetilde{\gamma}_n = \widehat{\gamma}_n + Z\sqrt{v_2(\widehat{\gamma}_n)/k_n}$, where $Z \sim \mathcal{N}(0,1)$ is independent from the data; in addition, we retain only those values of $\widetilde{\gamma}_n$ that are smaller than 1, *i.e.* we resample given $\widetilde{\gamma}_n < 1$. This gives rise to an approximation for the distribution of $G_n(Y_1,\ldots,Y_n,\widehat{\gamma}_n)$ and, accordingly, to a different interval $\widehat{\mathrm{I}}_3(\alpha)$. A step-by-step description of the construction of both $\widehat{\mathrm{I}}_2(\alpha)$ and $\widehat{\mathrm{I}}_3(\alpha)$ can be found in Algorithm 2 (where Φ denotes the standard normal distribution function).

Algorithm 2 Confidence intervals for $\mathrm{ES}(1-k_n/n)$ - Empirical estimator

Require: $N \geq 1$, $\alpha \in (0,1)$, $\widehat{\mathrm{ES}}_n(1-k_n/n)$, $\widehat{a}_n(n/k_n) = \widehat{a}_n^{\mathrm{Mom}}(n/k_n)$, $\widehat{\gamma}_n = \widehat{\gamma}_n^{\mathrm{Mom}}(k_n)$

Ensure: $\widehat{\gamma}_n < 1$

Simulate N replications $U_1,\ldots,U_N$ of a uniform distribution on $[0,1]$

for $i \in \{1,\ldots,N\}$ **do**

Calculate $\widetilde{\gamma}_{n,i} = \widehat{\gamma}_n + \sqrt{\frac{v_2(\widehat{\gamma}_n)}{k_n}}\Phi^{-1}\left(U_i\Phi\left((1-\widehat{\gamma}_n)\sqrt{\frac{k_n}{v_2(\widehat{\gamma}_n)}}\right)\right)$

Simulate n replications $Y_1,\ldots,Y_n$ of a unit Pareto distribution

Compute $G_i = G_n(Y_1,\ldots,Y_n,\widehat{\gamma}_n)$ and $\widetilde{G}_i = G_n(Y_1,\ldots,Y_n,\widetilde{\gamma}_{n,i})$

end for

Compute $\begin{cases} G_{\mathrm{up}} = G_{\mathrm{up}}(\alpha) = G_{\lfloor N(1-\alpha/2)\rfloor:N} \\ G_{\mathrm{down}} = G_{\mathrm{down}}(\alpha) = G_{\lfloor N\alpha/2\rfloor:N} \end{cases}$ and $\begin{cases} \widetilde{G}_{\mathrm{up}} = \widetilde{G}_{\mathrm{up}}(\alpha) = \widetilde{G}_{\lfloor N(1-\alpha/2)\rfloor:N} \\ \widetilde{G}_{\mathrm{down}} = \widetilde{G}_{\mathrm{down}}(\alpha) = \widetilde{G}_{\lfloor N\alpha/2\rfloor:N} \end{cases}$

return $\begin{cases} \widehat{\mathrm{I}}_2(\alpha) = \left[\widehat{\mathrm{ES}}_n(1-k_n/n) + \widehat{a}_n(n/k_n)G_{\mathrm{down}}, \widehat{\mathrm{ES}}_n(1-k_n/n) + \widehat{a}_n(n/k_n)G_{\mathrm{up}}\right] \\ \widehat{\mathrm{I}}_3(\alpha) = \left[\widehat{\mathrm{ES}}_n(1-k_n/n) + \widehat{a}_n(n/k_n)\widetilde{G}_{\mathrm{down}}, \widehat{\mathrm{ES}}_n(1-k_n/n) + \widehat{a}_n(n/k_n)\widetilde{G}_{\mathrm{up}}\right] \end{cases}$

We turn to the quantile-based estimator $\widetilde{\mathrm{ES}}_n(\tau_n)$. We first note that in the heavy-tailed setting $\gamma \in (0,1)$, owing to the convergence $a(t)/U(t) \to \gamma$ as $t \to \infty$, the following simpler version of the quantile-based estimator has been considered by El Methni and Stupfler (2017):

$$\widetilde{\mathrm{ES}}_n^{\mathrm{H}}(1-k_n/n) = \frac{X_{n-k_n:n}}{1-\widehat{\gamma}_n^{\mathrm{H}}},$$

where $\widehat{\gamma}_n^{\mathrm{H}}$ is again the Hill estimator, see Section 3.2 therein. Since, by (A.16) with $f = \log$,

$X_{n-k_n:n}$ and $\widehat{\gamma}_n^{\mathrm{H}}$ are asymptotically independent, it is straightforward to prove that

$$\sqrt{k_n}\log\left(\frac{\widetilde{\mathrm{ES}}_n^{\mathrm{H}}(1-k_n/n)}{\mathrm{ES}(1-k_n/n)}\right) = \sqrt{k_n}\left(\frac{\widetilde{\mathrm{ES}}_n^{\mathrm{H}}(1-k_n/n)}{\mathrm{ES}(1-k_n/n)}-1\right)+\mathrm{o}_{\mathbb{P}}(1)$$
$$\stackrel{\mathrm{d}}{\longrightarrow}\mathcal{N}\left(0,\frac{\gamma^2}{(1-\gamma)^2}(1+(1-\gamma)^2)\right).$$

The following naive (competitor) confidence interval follows:

$$\widetilde{\mathrm{I}}_1(\alpha)=\left[\widetilde{\mathrm{ES}}_n^{\mathrm{H}}(1-k_n/n)\exp\left(\pm\frac{\widehat{\gamma}_n^{\mathrm{H}}}{1-\widehat{\gamma}_n^{\mathrm{H}}}\sqrt{1+(1-\widehat{\gamma}_n^{\mathrm{H}})^2}\frac{z_{1-\alpha/2}}{\sqrt{k_n}}\right)\right].$$

This interval is only valid for $\gamma\in(0,1)$. Using the semiparametric estimator $\widetilde{\mathrm{ES}}_n$ (with the moment estimators $\widehat{a}_n(n/k_n)=\widehat{a}_n^{\mathrm{Mom}}(n/k_n)$ of the scale and $\widehat{\gamma}_n=\widehat{\gamma}_n^{\mathrm{Mom}}$ of the shape extreme value parameters, respectively) instead of $\widetilde{\mathrm{ES}}_n^{\mathrm{H}}$ and applying Theorem 2 leads to the alternative confidence interval

$$\widetilde{\mathrm{I}}_2(\alpha)=\Bigg[\widetilde{\mathrm{ES}}_n(1-k_n/n)$$
$$\pm\frac{\widehat{a}_n(n/k_n)}{\sqrt{k_n}}\sqrt{\frac{(1+\widehat{\gamma}_n)(1-\widehat{\gamma}_n)^3+v_1(\widehat{\gamma}_n)(1-\widehat{\gamma}_n)^2+2c(\widehat{\gamma}_n)(1-\widehat{\gamma}_n)+v_2(\widehat{\gamma}_n)}{(1-\widehat{\gamma}_n)^4}}z_{1-\alpha/2}\Bigg].$$

This confidence interval is, in theory, valid for any $\gamma<1$ but its finite-sample performance is often disappointing, particularly when γ is close to 1. We correct this interval so as to push its finite-sample coverage close to the nominal level while keeping it Gaussian in nature. Set

$$\widetilde{Z}_n=\sqrt{k_n}\frac{\widetilde{\mathrm{ES}}_n(1-k_n/n)-\mathrm{ES}(1-k_n/n)}{\widehat{a}_n(n/k_n)}=\sqrt{k_n}\frac{\widetilde{\mathrm{ES}}_n(1-k_n/n)-\mathrm{ES}(1-k_n/n)}{a(n/k_n)}\frac{a(n/k_n)}{\widehat{a}_n(n/k_n)}.$$

Theorem 2 suggests the finer approximation

$$\widetilde{Z}_n\stackrel{\mathrm{d}}{\approx}\left(N_{\mathrm{loc}}+\frac{1}{1-\gamma}N_{\mathrm{scale}}+\frac{1}{(1-\gamma)^2}N_{\mathrm{shape}}\right)\left(1-\frac{N_{\mathrm{scale}}}{\sqrt{k_n}}\right)\stackrel{\mathrm{d}}{=}\mathbf{u}^\top\mathbf{N}+\frac{\mathbf{N}^\top\mathbf{S}\mathbf{N}}{\sqrt{k_n}},$$

where $\mathbf{N}=(N_{\mathrm{loc}},N_{\mathrm{scale}},N_{\mathrm{shape}})^\top$ follows a trivariate centered normal distribution with covariance matrix $\boldsymbol{\Sigma}$,

$$\mathbf{S}=\mathbf{S}(\gamma)=\begin{pmatrix}0 & -\frac{1}{2} & 0\\ -\frac{1}{2} & -\frac{1}{1-\gamma} & -\frac{1}{2(1-\gamma)^2}\\ 0 & -\frac{1}{2(1-\gamma)^2} & 0\end{pmatrix}\text{ and }\mathbf{u}=\mathbf{u}(\gamma)=\begin{pmatrix}1\\ \frac{1}{1-\gamma}\\ \frac{1}{(1-\gamma)^2}\end{pmatrix}.$$

Straightforward calculations show that, if $\mathbf{Z}=(Z_1,\ldots,Z_p)$ is a $p-$dimensional Gaussian random vector made of independent centered unit Gaussian random variables and $\mathbf{M}$ is a $p\times p$

symmetric matrix, one has $\mathbb{E}(\mathbf{Z}^\top \mathbf{M}\mathbf{Z}) = \operatorname{tr}(\mathbf{M})$ and $\mathbb{E}((\mathbf{Z}^\top \mathbf{M}\mathbf{Z})^2) = 2\operatorname{tr}(\mathbf{M}^2) + (\operatorname{tr}(\mathbf{M}))^2$. The mean $m(\gamma)$ and variance $s^2(\gamma)$ of the random variable $\mathbf{u}^\top \mathbf{N} + \mathbf{N}^\top \mathbf{S}\mathbf{N}/\sqrt{k_n}$ are then

$$m(\gamma) = \frac{\operatorname{tr}(\mathbf{S}\boldsymbol{\Sigma})}{\sqrt{k_n}} \quad \text{and} \quad s^2(\gamma) = \mathbf{u}^\top \boldsymbol{\Sigma}\mathbf{u} + 2\frac{\operatorname{tr}(\mathbf{S}\boldsymbol{\Sigma}\mathbf{S}\boldsymbol{\Sigma})}{k_n}.$$

Approximating $\widetilde{Z}_n$ by a Gaussian random variable with mean $m(\widehat{\gamma}_n)$ and variance $s^2(\widehat{\gamma}_n)$ suggests the confidence interval

$$\widetilde{\mathrm{I}}_3(\alpha) = \left[\widetilde{\mathrm{ES}}_n(1 - k_n/n) - \frac{\widehat{a}_n(n/k_n)}{\sqrt{k_n}} m(\widehat{\gamma}_n) \pm \frac{\widehat{a}_n(n/k_n)}{\sqrt{k_n}} s(\widehat{\gamma}_n) z_{1-\alpha/2}\right].$$

Similarly to $\widehat{\mathrm{I}}_2(\alpha)$, the finite-sample performance of this interval suffers from not taking into account the statistical uncertainty of the estimator $\widehat{\gamma}_n$ plugged into m and s^2. To take this uncertainty into account, we analytically derive the correction term that should be added due to this plug-in step. Let $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$ be the Cholesky decomposition of $\boldsymbol{\Sigma}$, where

$$\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\gamma) = \begin{pmatrix} 1 & 0 & 0 \\ \gamma & \sqrt{v_1(\gamma) - \gamma^2} & 0 \\ 0 & \frac{c(\gamma)}{\sqrt{v_1(\gamma)-\gamma^2}} & \sqrt{v_2(\gamma) - \frac{c^2(\gamma)}{v_1(\gamma)-\gamma^2}} \end{pmatrix}$$

and note that $\mathbf{N} = \boldsymbol{\Lambda}\mathbf{Z}$ where $\mathbf{Z}$ is a 3−dimensional Gaussian random vector made of independent centered unit Gaussian random variables. Recalling that $\sqrt{k_n}(\widehat{\gamma}_n - \gamma) \approx N_{\text{shape}}$, a Taylor expansion yields

$$\begin{aligned} \widetilde{Z}_n &\overset{\mathrm{d}}{=} \mathbf{u}(\gamma)^\top \boldsymbol{\Lambda}(\gamma)\mathbf{Z} + \frac{\mathbf{Z}^\top \boldsymbol{\Lambda}(\gamma)^\top \mathbf{S}(\gamma)\boldsymbol{\Lambda}(\gamma)\mathbf{Z}}{\sqrt{k_n}} + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right) \\ &\overset{\mathrm{d}}{=} \mathbf{u}(\widehat{\gamma}_n)^\top \boldsymbol{\Lambda}(\widehat{\gamma}_n)\mathbf{Z} + \frac{\mathbf{Z}^\top \boldsymbol{\Lambda}(\widehat{\gamma}_n)^\top \mathbf{S}(\widehat{\gamma}_n)\boldsymbol{\Lambda}(\widehat{\gamma}_n)\mathbf{Z}}{\sqrt{k_n}} - \boldsymbol{\theta}(\widehat{\gamma}_n)^\top \frac{N_{\text{shape}}}{\sqrt{k_n}}\mathbf{Z} + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right) \end{aligned}$$

where $\boldsymbol{\theta}(\gamma) = \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}\gamma}(\gamma)^\top \boldsymbol{\Lambda}(\gamma) + \mathbf{u}(\gamma)^\top \frac{\mathrm{d}\boldsymbol{\Lambda}}{\mathrm{d}\gamma}(\gamma)$.

Since $\mathbf{N} = \boldsymbol{\Lambda}\mathbf{Z}$, one has $N_{\text{shape}} = \Lambda_{23}(\gamma)Z_2 + \Lambda_{33}(\gamma)Z_3 \overset{\mathrm{d}}{=} \Lambda_{23}(\widehat{\gamma}_n)Z_2 + \Lambda_{33}(\widehat{\gamma}_n)Z_3 + \mathrm{o}_{\mathbb{P}}(1)$. Then

$$\widetilde{Z}_n \overset{\mathrm{d}}{=} \mathbf{w}(\widehat{\gamma}_n)^\top \mathbf{Z} + \frac{\mathbf{Z}^\top \mathbf{W}(\widehat{\gamma}_n)\mathbf{Z}}{\sqrt{k_n}} + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right),$$

where $\mathbf{w}(\gamma) = \boldsymbol{\Lambda}^\top(\gamma)\mathbf{u}(\gamma)$ and

$$\begin{aligned} \mathbf{W}(\gamma) &= \boldsymbol{\Lambda}(\gamma)^\top \mathbf{S}(\gamma)\boldsymbol{\Lambda}(\gamma) \\ &\quad - \frac{1}{2}\begin{pmatrix} 0 & \theta_1(\gamma)\Lambda_{23}(\gamma) & \theta_1(\gamma)\Lambda_{33}(\gamma) \\ \theta_1(\gamma)\Lambda_{23}(\gamma) & 2\theta_2(\gamma)\Lambda_{23}(\gamma) & \theta_2(\gamma)\Lambda_{33}(\gamma) + \theta_3(\gamma)\Lambda_{23}(\gamma) \\ \theta_1(\gamma)\Lambda_{33}(\gamma) & \theta_2(\gamma)\Lambda_{33}(\gamma) + \theta_3(\gamma)\Lambda_{23}(\gamma) & 2\theta_3(\gamma)\Lambda_{33}(\gamma) \end{pmatrix}. \end{aligned}$$

Explicit expressions can of course be given for $\mathbf{w}(\gamma)$ and $\mathbf{W}(\gamma)$; we omit the details for the sake of brevity. As in the construction of the confidence interval $\widetilde{\mathrm{I}}_3(\alpha)$, we then approximate the distribution of the random variable $\mathbf{w}(\gamma_0)^\top \mathbf{Z} + \mathbf{Z}^\top \mathbf{W}(\gamma_0)\mathbf{Z}/\sqrt{k_n}$ (for any fixed $\gamma_0 < 1$) by a Gaussian distribution with mean $\mathrm{tr}(\mathbf{W}(\gamma_0))$ and variance $\|\mathbf{w}(\gamma_0)\|_2^2 + 2\,\mathrm{tr}(\mathbf{W}^2(\gamma_0))/k_n$. This suggests our final confidence interval for $\mathrm{ES}(1-k_n/n)$:

$$\widetilde{\mathrm{I}}_4(\alpha) = \left[\widetilde{\mathrm{ES}}_n(1-k_n/n) + \frac{\widehat{a}_n(n/k_n)}{\sqrt{k_n}}\left(-\,\mathrm{tr}(\mathbf{W}(\widehat{\gamma}_n)) \pm \sqrt{\|\mathbf{w}(\widehat{\gamma}_n)\|_2^2 + 2\frac{\mathrm{tr}(\mathbf{W}^2(\widehat{\gamma}_n))}{k_n}}\right) z_{1-\alpha/2}\right].$$

We shall see that this interval has coverage probabilities close to the nominal level across the full range $\gamma < 1$ in a reasonably large class of models.

C Finite-sample results at intermediate levels

We check the quality of our estimators and related inference procedures at an intermediate level on simulated datasets. We consider the same distributions as in the main paper, that is:

- The Kumaraswamy distribution with $1 - F(t) = (1-t^\alpha)^\beta$ for $t \in [0,1]$, and the Reverse-Burr distribution with $1 - F(t) = (1 + (1-t)^{-\beta/\alpha})^{-1/\beta}$ for $t < 1$ (here $\alpha > 0$, $\beta > 0$), with respective extreme value indices $\gamma = -1/\beta < 0$ and $\gamma = -\alpha < 0$.

- The Gumbel distribution with $1 - F(t) = 1 - \exp(-\exp(-t))$, and the Exponential distribution $1 - F(t) = \exp(-t)$ for $t > 0$, both having extreme value index $\gamma = 0$.

- The Pareto distribution with $1 - F(t) = t^{-\alpha}$ for $t > 1$, and the Fréchet distribution with $1-F(t) = 1-\exp(-t^{-\alpha})$ for $t > 0$ (here $\alpha > 0$), both having extreme value index $\gamma = 1/\alpha > 0$.

In each setting, we simulate N = 10,000 replications of an i.i.d. sample of size n = 1,000 from the chosen distribution. We fix $k_n = 200$ and we estimate and infer the quantity $\mathrm{ES}(1-k_n/n) = \mathrm{ES}(0.8)$ using the estimators $\widehat{\mathrm{ES}}_n(1-k_n/n)$ and $\widetilde{\mathrm{ES}}_n(1-k_n/n)$, and the confidence intervals $\widehat{\mathrm{I}}_1(0.95)$, $\widehat{\mathrm{I}}_2(0.95)$, $\widehat{\mathrm{I}}_3(0.95)$, $\widetilde{\mathrm{I}}_1(0.95)$, $\widetilde{\mathrm{I}}_2(0.95)$, $\widetilde{\mathrm{I}}_3(0.95)$ and $\widetilde{\mathrm{I}}_4(0.95)$ at confidence level 0.95. The true values of ES(0.8), evaluated using Table C.1, and the empirical coverage probabilities of the competing intervals are provided in Table C.2.

At the intermediate level, it appears that the interval $\widehat{\mathrm{I}}_3(0.95)$ overall performs best among the confidence intervals based on $\widehat{\mathrm{ES}}_n$; in particular, its coverage probability is very close to the nominal level even for the infinite-variance Fréchet distribution with $\gamma = 1/2$. The conclusion is somewhat more complex regarding the intervals based on $\widetilde{\mathrm{ES}}_n$: the interval $\widetilde{\mathrm{I}}_2(0.95)$, which is purely based on the asymptotic normality of $\widetilde{\mathrm{ES}}_n$, seems to perform best for finite-variance distributions, while $\widetilde{\mathrm{I}}_4(0.95)$ should be preferred to $\widetilde{\mathrm{I}}_2(0.95)$ when the underlying distribution has an infinite variance. Overall, the confidence intervals $\widehat{\mathrm{I}}_1(0.95)$ and $\widetilde{\mathrm{I}}_1(0.95)$ behave very poorly when the assumption of a heavy right tail is violated, and the intervals $\widehat{\mathrm{I}}_1(0.95)$ and $\widehat{\mathrm{I}}_2(0.95)$ also have poor coverage when the variance of the underlying distribution is infinite.

Distribution	Density function $f(t)$	$\mathrm{ES}(\tau)$
Pareto ($\alpha > 1$)	$\alpha t^{-\alpha-1},\ t > 1$	$\frac{\alpha}{\alpha-1}(1-\tau)^{-1/\alpha}$
Burr ($0 < \alpha < 1$)	$\frac{1}{\alpha} t^{\beta/\alpha-1}\left(1+t^{\beta/\alpha}\right)^{-1/\beta-1},\ t > 0$	$\frac{\left((1-\tau)^{-\beta}-1\right)^{\alpha/\beta}}{1-\alpha}\, {}_2F_1\left(1, 1+\frac{1-\alpha}{\beta}, \frac{1}{1-\tau^{-\beta}}; -\frac{\alpha}{\beta}\right)$
Fréchet ($\alpha > 1$)	$\alpha t^{-1-\alpha}\exp(-t^{-\alpha}),\ t > 0$	$\frac{\gamma_{\log(1/\tau)}(1-1/\alpha)}{1-\tau}$
Student ($\alpha > 1$)	$\Phi'_\alpha(t) = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\sqrt{\alpha\pi}\,\Gamma\left(\frac{\alpha}{2}\right)}\left(1+\frac{t^2}{\alpha}\right)^{-(\alpha+1)/2}$	$\frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\sqrt{\alpha\pi}\,\Gamma\left(\frac{\alpha}{2}\right)}\frac{\alpha}{\alpha-1}\left(1+\frac{(\Phi_\alpha^{-1}(\tau))^2}{\alpha}\right)^{(1-\alpha)/2}$
Exponential	$\exp(-t),\ t > 0$	$1-\log(1-\tau)$
Weibull	$\beta t^{\beta-1}\exp(-t^{\beta}),\ t > 0$	$\frac{\Gamma_{\log(1/(1-\tau))}(1+1/\beta)}{1-\tau}$
Log-normal	$\frac{1}{t\sigma\sqrt{2\pi}}\exp\left(-\frac{(\log(t)-\mu)^2}{2\sigma^2}\right),\ t > 0$	$\exp\left(\mu+\frac{\sigma^2}{2}\right)\frac{\Phi\left(\sigma-\Phi^{-1}(\tau)\right)}{1-\tau}$
Gumbel	$\exp(-t)\exp(-\exp(-t))$	$\frac{c+\tau\log(-\log(\tau))+E_1(-\log(\tau))}{1-\tau}$
Laplace	$\frac{1}{2}\exp(-\lvert t\rvert)$	$\frac{\min\{\tau,1-\tau\}(1-\log(2\min\{\tau,1-\tau\}))}{1-\tau}$
Logistic	$\frac{\exp(-t)}{(1+\exp(-t))^2}$	$\left(\frac{\tau}{1-\tau}\log(\tau)+\log(1-\tau)\right)$
Normal	$\varphi(t) = \Phi'(t) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)$	$\frac{\varphi\left(\Phi^{-1}(\tau)\right)}{1-\tau}$
Kumaraswamy	$\alpha\beta t^{\alpha-1}\left(1-t^{\alpha}\right)^{\beta-1},\ t \in [0,1]$	$\beta\frac{B\left(\frac{1}{\alpha}+1,\beta\right)-B_{1-(1-\tau)^{1/\beta}}\left(\frac{1}{\alpha}+1,\beta\right)}{1-\tau}$
Reverse-Fréchet	$\alpha(1-t)^{\alpha-1}\exp\left(-(1-t)^{\alpha}\right),\ t < 1$	$1-\frac{\gamma_{-\log(\tau)}\left(\frac{1}{\alpha}\right)-\alpha\tau(-\log(\tau))^{1/\alpha}}{\alpha(1-\tau)}$
Triangular	$2t,\ t \in [0,1]$ and $(2-t),\ t \in [1,2]$	$\begin{cases}\frac{1-(2\tau)^{3/2}/3}{1-\tau} & \text{for } \tau \le 1/2,\\ 2\left(1-\frac{\sqrt{2(1-\tau)}}{3}\right) & \text{for } \tau \ge 1/2\end{cases}$

Table C.1: A list of standard, unit-scale continuous distributions and the associated values of the Expected Shortfall. The parameters α and β are positive, and for the log-normal distribution, we take $\mu \in \mathbb{R}$ and $\sigma > 0$. We write Φ for the standard Gaussian distribution function and $\varphi = \Phi'$ for the associated density function, Φ_α for the Student distribution function with α degrees of freedom, $x \mapsto {}_2F_1(a, b, c; x)$ for the ordinary hypergeometric function found as one of the two fundamental solutions of the differential equation $x(1-x)y''(x) + (c-(a+b+1)x)y'(x) - ab\,y(x) = 0$, $\gamma_x(\alpha) = \int_0^x t^{\alpha-1}e^{-t}\,\mathrm{d}t$ for the lower incomplete Gamma function, $\Gamma_x(\alpha) = \int_x^\infty t^{\alpha-1}e^{-t}\,\mathrm{d}t$ for the upper incomplete Gamma function, c for the Euler-Mascheroni constant, $E_1(x) = \int_x^\infty\{\exp(-t)/t\}\mathrm{d}t$ for the exponential integral and $B_x(\alpha,\beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t$ for the lower incomplete Beta function.

Distribution	γ	$\mathrm{ES}(\tau_n)$	$\widehat{\mathrm{I}}_1(0.95)$	$\widehat{\mathrm{I}}_2(0.95)$	$\widehat{\mathrm{I}}_3(0.95)$	$\widetilde{\mathrm{I}}_1(0.95)$	$\widetilde{\mathrm{I}}_2(0.95)$	$\widetilde{\mathrm{I}}_3(0.95)$	$\widetilde{\mathrm{I}}_4(0.95)$
Kumaraswamy $\alpha = \beta = 2$	−0.5	0.835	[0.815,0.856] (0.992)	[0.823,0.852] (0.942)	[0.823,0.852] (0.946)	[0.819,0.859] (0.988)	[0.819,0.849] (0.958)	[0.818,0.849] (0.958)	[0.818,0.849] (0.958)
Reversed Burr $\alpha = 1/4$, $\beta = 3$	−0.25	0.465	[0.427,0.506] (0.998)	[0.445,0.491] (0.940)	[0.444,0.491] (0.944)	[0.447,0.522] (0.941)	[0.438,0.485] (0.952)	[0.438,0.485] (0.953)	[0.438,0.485] (0.953)
Kumaraswamy $\alpha = 1$, $\beta = 10$	−0.1	0.226	[0.200,0.255] (0.999)	[0.214,0.242] (0.944)	[0.213,0.242] (0.950)	[0.216,0.264] (0.916)	[0.210,0.239] (0.955)	[0.210,0.239] (0.957)	[0.210,0.240] (0.959)
Gumbel	0	2.556	[2.023,3.198] (0.834)	[2.377,2.774] (0.946)	[2.369,2.778] (0.953)	[2.460,3.250] (0.830)	[2.331,2.744] (0.951)	[2.329,2.746] (0.953)	[2.335,2.753] (0.955)
Exponential $\lambda = 1$	0	2.609	[2.199,3.078] (0.996)	[2.433,2.820] (0.955)	[2.424,2.824] (0.960)	[2.488,3.156] (0.892)	[2.397,2.799] (0.955)	[2.396,2.801] (0.956)	[2.402,2.808] (0.960)
Pareto $\alpha = 10$	0.1	1.305	[1.278,1.332] (0.951)	[1.281,1.333] (0.945)	[1.280,1.334] (0.953)	[1.278,1.332] (0.957)	[1.278,1.331] (0.955)	[1.277,1.332] (0.956)	[1.278,1.333] (0.957)
Pareto $\alpha = 4$	0.25	1.994	[1.875,2.114] (0.949)	[1.878,2.112] (0.949)	[1.868,2.118] (0.957)	[1.880,2.110] (0.955)	[1.874,2.106] (0.950)	[1.873,2.108] (0.953)	[1.879,2.117] (0.957)
Fréchet $\alpha = 2$	0.5	4.395	[2.734,5.825] (0.217), NA: 77.6%	[3.205,5.051] (0.916)	[3.052,5.206] (0.938)	[3.772,5.311] (0.967)	[3.616,5.122] (0.938)	[3.608,5.142] (0.941)	[3.654,5.305] (0.958)
Pareto $\alpha = 5/3$	0.6	6.566	[3.119,8.161] (0.009), NA: 99.0%	[3.471,7.760] (0.871)	[3.083,8.423] (0.918)	[5.236,8.173] (0.950)	[4.817,8.059] (0.917)	[4.794,8.100] (0.921)	[4.862,8.644] (0.950)

Table C.2: Inference about $\mathrm{ES}(1 - k_n/n) = \mathrm{ES}(0.8)$ – For each interval among $\widehat{\mathrm{I}}_1(0.95)$, $\widehat{\mathrm{I}}_2(0.95)$, $\widehat{\mathrm{I}}_3(0.95)$, $\widetilde{\mathrm{I}}_1(0.95)$, $\widetilde{\mathrm{I}}_2(0.95)$, $\widetilde{\mathrm{I}}_3(0.95)$ and $\widetilde{\mathrm{I}}_4(0.95)$, and in each tested case, we report between square brackets the median values (over the $N = 10{,}000$ replications) of its lower and upper bounds; the empirical coverage probability is indicated between round brackets. The sample size is $n = 1{,}000$.