WORKING PAPERS

N° 1727

March 2026

"Extreme value inference for heterogeneous heavy-tailed data: A derandomization theory"

Abdelaati Daouia, Joseph Hachem and Gilles Stupfler

Toulouse School of Economics

Extreme value inference for heterogeneous heavy-tailed data: A derandomization theory

Abdelaati Daouia[a], Joseph Hachem[a] & Gilles Stupfler[b]

[a] Toulouse School of Economics, Université de Toulouse Capitole, Toulouse, France
[b] Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France

Abstract

A major mathematical difficulty in studying extreme value parameter estimators defined as empirical mean excesses is their reliance on high order statistics above a random threshold. Based on simple yet novel derandomization arguments, we provide sufficient conditions for deriving the joint asymptotic distribution of so-called tail empirical excesses and Expected Shortfall with the underlying threshold level. This high-level result allows for a strong degree of heterogeneity in the data-generating process as well as serial dependence. When the observations are independent and their average distribution is heavy-tailed, we obtain asymptotic normality results for the Hill estimator of the extreme value index, the Weissman estimator of extreme quantiles, and two estimators of Expected Shortfall above an extreme level, under substantially weaker, yet easily verifiable and interpretable conditions than those prevailing in the recent literature. In particular, we establish precise closed-form expressions for the asymptotic bias and variance of each estimator. Our assumptions hold in a wide range of models where existing results may not apply, including scenarios of contaminated samples, pooled samples from several populations, heterogeneous location-scale models and the situation where observed covariate information is ignored. We discuss practical consequences of our results on simulated data and two real data applications to cyber risk and financial risk management.

MSC 2020 subject classifications: 62G30, 62G32
Keywords: Derandomization, Expected Shortfall, Extreme quantile, Heavy tails, Heterogeneity, Hill estimator

1 Background and main contributions

1.1 Derandomization for tail mean excesses and Expected Shortfall

Let, for $n \geq 1$, $X_1^{(n)}, \ldots, X_n^{(n)}$ be (almost surely finite) random variables, and denote by $X_{1:n}^{(n)} \leq X_{2:n}^{(n)} \leq \cdots \leq X_{n:n}^{(n)}$ their order statistics. We do not assume that the $X_i^{(n)}$, $1 \leq i \leq n$, have the same distribution, and we do not assume for the moment that they are independent either. Let $k = k(n)$ be a sequence of integers tending to infinity such that $k/n \to 0$ and f be a smooth nondecreasing function. The original motivation for this work is the analysis of the asymptotic behavior of the statistic

$$\widehat{e}_{f,n}(k) = \frac{1}{k}\sum_{i=1}^{k} f(X_{n-i+1:n}^{(n)}) - f(X_{n-k:n}^{(n)}),$$

which encompasses several well-known classes of extreme value estimators. [This is different from quantities of the form $\frac{1}{k}\sum_{i=1}^{k} f(X_{n-i+1:n}^{(n)}/X_{n-k:n}^{(n)})\mathbb{1}\{X_{n-k:n}^{(n)} > 0\}$, called *residual estimators* and studied for independent and identically distributed data in Segers [2001].] In the special situation where the $X_i^{(n)}$ have the same distribution as a random variable X with quantile function $q(\cdot)$, the quantity $\widehat{e}_{f,n}(k)$ is a natural estimator of $\mathbb{E}(f(X)-f(q(1-k/n)) \mid X > q(1-k/n))$. The latter quantity is the mean excess value of $f(X)$ when $X > q(1-k/n)$, which motivates the name *mean f-excess*, and we will call $\widehat{e}_{f,n}(k)$ the *empirical mean f-excess* above the order statistic $X^{(n)}_{n-k:n}$ throughout. Prominent among these quantities are the ones obtained with $f = \log$ and $f = \mathrm{Id}$, respectively

$$\widehat{e}_{\log,n}(k) = \frac{1}{k}\sum_{i=1}^{k} \log X^{(n)}_{n-i+1:n} - \log X^{(n)}_{n-k:n} \quad \text{and} \quad \widehat{e}_{\mathrm{Id},n}(k) = \frac{1}{k}\sum_{i=1}^{k} X^{(n)}_{n-i+1:n} - X^{(n)}_{n-k:n}.$$

The first quantity is reminiscent of the renowned Hill [1975] estimator of the extreme value index $\gamma > 0$ of a heavy-tailed random variable X; here we take heavy-tailed to mean that the survival function $\overline{F}$ of X is regularly varying with index $-1/\gamma$, that is $\overline{F}(tx)/\overline{F}(t) \to x^{-1/\gamma}$ for all $x > 0$, as $t \to \infty$. The second quantity above is an empirical version of the mean residual life above a threshold level $X^{(n)}_{n-k:n}$ (and is close to the tail array sum of Rootzén et al. [1998] obtained with the identity function), which leads to

$$\widehat{\mathrm{ES}}_{\mathrm{Id},n}(k) = \widehat{e}_{\mathrm{Id},n}(k) + X^{(n)}_{n-k:n} = \frac{1}{k}\sum_{i=1}^{k} X^{(n)}_{n-i+1:n}$$

that defines an empirical version of the Expected Shortfall above level $X^{(n)}_{n-k:n}$. The Expected Shortfall, which is identical to the Conditional Tail Expectation for continuous loss variables, is a very important risk measure that is used in capital requirement calculations by the Canadian financial and actuarial sectors [International Monetary Fund, 2014], as well as for guaranteeing the sustainability of life insurance annuities in the USA [Organisation for Economic Co-operation and Development, 2016]. On the regulation side, the Basel Committee on Banking Supervision has recommended to use this risk measure rather than quantiles, or equivalently Value-at-Risk (VaR), in internal market risk models [Basel Committee on Banking Supervision, 2013]. Key to the good behavior of the Expected Shortfall in practical applications is the fact that it is a coherent risk measure which is jointly elicitable with VaR [Fissler and Ziegel, 2016]. More generally, for $f = f_p : x \mapsto x^p$, with $p > 0$, the quantity $\widehat{e}_{f_p,n}(k) + (X^{(n)}_{n-k:n})^p$ is an empirical Conditional Tail Moment of order p [El Methni et al., 2014].

The major theoretical difficulty in the study of random averages such as $\widehat{e}_{f,n}(k)$ is their reliance on order statistics of the sample. Writing

$$\widehat{e}_{f,n}(k) = \widehat{e}_{f,n}(k, X^{(n)}_{n-k:n}) = \frac{1}{k}\sum_{i=1}^{n} \left(f(X^{(n)}_i) - f(X^{(n)}_{n-k:n}) \right) \mathbb{1}\{X^{(n)}_i > X^{(n)}_{n-k:n}\}$$

and assuming that the intermediate order statistic $X^{(n)}_{n-k:n}$ consistently estimates some deterministic counterpart u_n, our first main high-level result, which is Theorem 1.1 below, provides a simple but very wide framework in which the convergence of $\widehat{e}_{f,n}(k, X^{(n)}_{n-k:n})$ is easy to obtain, jointly with $X^{(n)}_{n-k:n}$, from what we shall call the derandomized pseudo-estimator

$$\widehat{e}_{f,n}(k, u_n) = \frac{1}{k}\sum_{i=1}^{n} (f(X^{(n)}_i) - f(u_n)) \mathbb{1}\{X^{(n)}_i > u_n\}.$$

To the best of our knowledge, this is the first work to show such a general limit result for a generic function f and for both heterogeneous and homogeneous data. Throughout we set

$$\widehat{\overline{F}}_n(t) = 1 - \widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X^{(n)}_i > t\} \quad \text{and} \quad \overline{F}_n(t) = 1 - F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(X^{(n)}_i > t).$$

We also define

$$e_{f,n}(k, u_n) = \mathbb{E}(\widehat{e}_{f,n}(k, u_n)) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{E}((f(X^{(n)}_i) - f(u_n)) \mathbb{1}\{X^{(n)}_i > u_n\})$$

which is the natural population counterpart of $\widehat{e}_{f,n}(k,u_n)$, and we denote the f-Expected Shortfall estimator above level $X^{(n)}_{n-k:n}$ by $\widehat{\mathrm{ES}}_{f,n}(k,X^{(n)}_{n-k:n}) = \widehat{e}_{f,n}(k,X^{(n)}_{n-k:n}) + f(X^{(n)}_{n-k:n})$ (*i.e.* the conditional expectation of $f(Z)$ given that it exceeds its $(1-k/n)-$quantile when Z has distribution function $\widehat{F}_n$) whose natural population counterpart at level u_n is $\mathrm{ES}_{f,n}(k,u_n) = e_{f,n}(k,u_n) + f(u_n)$.

Theorem 1.1 (Derandomization theorem for tail empirical f-excesses)**.** *Pick $d \geq 1$ and assume that, in a neighborhood of infinity, $f_1,\ldots,f_d$ are increasing, continuously differentiable functions having regularly varying derivatives. Let $k = k(n) \to \infty$ be a sequence of integers with $k/n \to 0$, and assume that there is a positive sequence (u_n) tending to infinity such that, for any sequence (ε_n) converging to 0,*

$$\sqrt{k}\left(\frac{\widehat{e}_{f_1,n}(k,u_n)}{e_{f_1,n}(k,u_n)} - 1, \ldots, \frac{\widehat{e}_{f_d,n}(k,u_n)}{e_{f_d,n}(k,u_n)} - 1, \frac{\widehat{\overline{F}}_n((1+\varepsilon_n)u_n)}{\overline{F}_n((1+\varepsilon_n)u_n)} - 1\right) \xrightarrow{\mathrm{d}} (\Theta_1,\ldots,\Theta_d,\Psi)$$

where $(\Theta_1,\ldots,\Theta_d,\Psi)$ is a nondegenerate random vector whose marginals have continuous distribution functions. Suppose moreover that there are constants $c_1,\ldots,c_d, C_1 > 0$ and $C_2 \in \mathbb{R}$ such that

$$\forall j \in \{1,\ldots,d\}, \ \lim_{n\to\infty} \frac{e_{f_j,n}(k,u_n)}{u_n f_j'(u_n)} = c_j$$
$$\text{and} \ \ \forall t \in \mathbb{R}, \ \lim_{n\to\infty} \sqrt{k}\left(\frac{k/n}{\overline{F}_n((1+t/\sqrt{k})u_n)} - 1\right) = C_1 t + C_2.$$

Then

$$\forall j \in \{1,\ldots,d\}, \ \widehat{e}_{f_j,n}(k,X^{(n)}_{n-k:n}) = \widehat{e}_{f_j,n}(k,u_n) - u_n f_j'(u_n)\left(\frac{X^{(n)}_{n-k:n}}{u_n} - 1\right) + \mathrm{o}_{\mathbb{P}}\left(\frac{u_n f_j'(u_n)}{\sqrt{k}}\right)$$

and as a consequence

$$\sqrt{k}\left(\frac{\widehat{e}_{f_1,n}(k,X^{(n)}_{n-k:n})}{e_{f_1,n}(k,u_n)} - 1, \ldots, \frac{\widehat{e}_{f_d,n}(k,X^{(n)}_{n-k:n})}{e_{f_d,n}(k,u_n)} - 1, \frac{X^{(n)}_{n-k:n}}{u_n} - 1\right)$$
$$\xrightarrow{\mathrm{d}} \left(\Theta_1 - \frac{\Psi - C_2}{c_1 C_1}, \ldots, \Theta_d - \frac{\Psi - C_2}{c_d C_1}, \frac{\Psi - C_2}{C_1}\right).$$

Especially, for $d = 2$, $f = f_1$, $g = f_2$, if furthermore $g(x) \to \infty$ as $x \to \infty$ and $b - 1$ denotes the index of regular variation of g' with $b \geq 0$, then

$$\sqrt{k}\left(\frac{\widehat{e}_{f,n}(k,X^{(n)}_{n-k:n})}{e_{f,n}(k,u_n)} - 1, \frac{\widehat{\mathrm{ES}}_{g,n}(k,X^{(n)}_{n-k:n})}{\mathrm{ES}_{g,n}(k,u_n)} - 1, \frac{X^{(n)}_{n-k:n}}{u_n} - 1\right) \xrightarrow{\mathrm{d}} \left(\Theta_1 - \frac{\Psi - C_2}{c_1 C_1}, \frac{b c_2 \Theta_2}{1 + b c_2}, \frac{\Psi - C_2}{C_1}\right).$$

The remarkable feature of Theorem 1.1 is that its validity does not require any assumption on the dependence structure or the stationarity of $X^{(n)}_1,\ldots,X^{(n)}_n$. It is therefore useful in a very wide range of settings; a weaker derandomization trick was used by Hsing [1991] for estimating the extreme value index γ in a dependent stationary case.

To keep the article to manageable proportions, we specialize the discussion to a general model of independent and heterogeneous data, extending very recent work of Einmahl and He [2023b] in different directions. Although our applications will focus on cases where the derivatives of the f_j are power functions, we chose to keep Theorem 1.1 as general as possible to cover recent pieces of work such as Zhao et al. [2021] and Mao et al. [2023] where considering excess-type measures of regularly varying functions of the data has been of interest.

1.2 Inference for heterogeneous independent data

We consider independent random variables $X_1^{(n)}, \ldots, X_n^{(n)}$ whose average distribution behaves, as $n \to \infty$, as a Pareto-type distribution. The monograph by Koenker [2005] offers a broad overview of central quantile estimation, showing that considering independent but not necessarily identically distributed data is standard in this setting. Much less is known when the target is an extreme quantile of the data, whose order $\tau = \tau_n \to 1$ as $n \to \infty$. A pioneering theoretical and methodological contribution in this setting is Einmahl et al. [2016], where all observations are required to share the same extreme value index. This restriction was relaxed by de Haan and Zhou [2021] who allow different extreme value indices for the observations $X_i^{(n)}$, but require a gradually smooth change of the marginal distributions in i. Einmahl and He [2023a] consider more generally heterogeneous data, but only give consistency of the Hill estimator for the extreme value index of the average distribution in a possibly serially dependent setup. The latest contribution we are aware of is Einmahl and He [2023b] who, thanks to a powerful functional central limit theorem for the tail empirical process, derive asymptotic normality results for intermediate order statistics (Corollary 2.2 therein), the Hill estimator (Theorem 2.2 therein) and the Weissman estimator of extreme quantiles (Theorem 4.1 therein) for the average distribution. Due to their reliance on a Gaussian approximation to the tail empirical process, the conditions they impose can be considered strong: in particular, they require a certain boundedness condition on the average survival function and the average density function, in addition to a smoothness condition.

Our derandomization theory in Section 2 is more general in that it results in the joint asymptotic normality of the mean f-excess estimator $\widehat{e}_{f,n}(k, X_{n-k:n}^{(n)})$, a tail empirical g-Expected Shortfall $\widehat{\text{ES}}_{g,n}(k, X_{n-k:n}^{(n)})$ and the intermediate empirical quantile $X_{n-k:n}^{(n)}$ under substantially weaker conditions, with the influence of heterogeneity upon their asymptotic dependence structure clearly identified (Theorem 2.1, Corollary 2.1 and Theorem 2.2), and without relying on Gaussian approximations to the tail empirical process. In particular, we present novel asymptotic normality results for the Hill estimator jointly with its corresponding random threshold $X_{n-k:n}^{(n)}$ (Corollary 2.2) and for the Weissman estimator (Corollary 2.3). We considerably extend the results of Einmahl and He [2023b] by deriving the precise asymptotic bias of each estimator under milder L^1-type conditions which, we argue, are more suitable for analyzing the structure of mean f-excess estimators. Our conditions are valid in natural settings, including contaminated data, where those of Einmahl and He [2023b] cannot hold.

Another substantial contribution of this article is that it is the first work unravelling the problem of extreme value estimation of the tail Expected Shortfall for heterogeneous data with possibly very different distributions. We present novel asymptotic normality results for two intermediate (empirical and quantile-based) Expected Shortfall estimators jointly with the Hill estimator (Corollaries 2.4 and 2.5), and for two extrapolated Expected Shortfall estimators (Corollary 2.6). As with the Hill and Weissman estimators, the asymptotic variance of the Expected Shortfall estimators can be substantially reduced compared to the i.i.d. case due to heterogeneity. This finding is supported by finite-sample evidence from simulated data, and illustrated using two real data examples: the PRC (Privacy Rights Clearinghouse) database and daily loss returns for the American Express Company (AXP) from 20 March 2018 to 18 October 2024. Our case studies highlight the value of accounting for heterogeneity in extreme value analysis from an inferential point of view.

The paper is further organized as follows. Section 2 presents in detail our inferential procedure for heterogeneous independent data, including model assumptions in Section 2.1, the results for a general mean f-excess in Section 2.2, those for extreme value index and extreme quantile estimation in Section 2.3, and for extreme Expected Shortfall estimation in Section 2.4. Section 3 discusses four representative models covered by our framework: contaminated samples in Section 3.1, pooling across multiple populations in Section 3.2, the heterogeneous scales model in Section 3.3, and omission of observed covariate information in Section 3.4. Section 4 highlights practical consequences of our theory using simulated data. Section 5 outlines the conclusions of our case studies on cyber risk and financial risk. All the necessary mathematical proofs, practical implementation guidelines and further details about the simulation and the real data analyses are given in an online supplementary file.

2 Extreme value inference for heterogeneous data

2.1 Model and assumptions

In the heterogeneous setup, the $X_i^{(n)}$ do not have a common distribution. It was established in Stigler [1974, Theorem 6] that a reasonable condition for linear functions of order statistics to be asymptotically normal is that the average distribution function F_n of the $X_i^{(n)}$ approaches a fixed distribution function F as $n \to \infty$. This suggests that, in our setting where we focus solely on functions of the top order statistics in the data, a minimal condition for the asymptotic normality of quantities such as

$$\widehat{e}_{f,n}(k, X_{n-k:n}^{(n)}) = \frac{1}{k}\sum_{i=1}^{k} f(X_{n-i+1:n}^{(n)}) - f(X_{n-k:n}^{(n)}) = \frac{n}{k}\int_{X_{n-k:n}^{(n)}}^{\infty} \widehat{\overline{F}}_n(x) f'(x)\mathrm{d}x$$

is that F_n approaches the distribution function of a heavy-tailed random variable X in the intermediate tail as $n \to \infty$. This motivates the two following fundamental assumptions:

$\mathcal{H}_1(\mathbb{P}_X, \gamma)$ The $X_i^{(n)}$, $1 \le i \le n$, are independent for each n, and there exist a distribution $\mathbb{P}_X$ of a heavy-tailed random variable X and $\gamma > 0$ such that

$$\forall x > 0,\ \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{P}(X_i^{(n)} > u_n x)}{\mathbb{P}(X > u_n)} = x^{-1/\gamma} \ \text{ whenever } u_n \to \infty \text{ with } n\mathbb{P}(X > u_n) \to \infty.$$

$\mathcal{I}(\mathbb{P}_X, \gamma, a)$ Assumption $\mathcal{H}_1(\mathbb{P}_X, \gamma)$ holds and $a \ge 0$ is such that, for some $\varepsilon > 0$,

$$\lim_{n\to\infty} \int_1^{\infty} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{P}(X_i^{(n)} > u_n x)}{\mathbb{P}(X > u_n)} - x^{-1/\gamma} \right| x^{a-1+\varepsilon}\mathrm{d}x = 0 \ \text{ whenever } u_n \to \infty \text{ with } n\mathbb{P}(X > u_n) \to \infty.$$

Remark 2.1 (On assumptions $\mathcal{H}_1(\mathbb{P}_X, \gamma)$ and $\mathcal{I}(\mathbb{P}_X, \gamma, a)$)**.** It is immediate that condition $\mathcal{H}_1(\mathbb{P}_X, \gamma)$ is equivalent to assuming that X is heavy-tailed with extreme value index $\gamma > 0$ and

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{P}(X_i^{(n)} > u_n)}{\mathbb{P}(X > u_n)} = 1 \ \text{ whenever } u_n \to \infty \text{ with } n\mathbb{P}(X > u_n) \to \infty.$$

This makes $\mathcal{H}_1(\mathbb{P}_X, \gamma)$ equivalent to the (heavy tail) Assumption 2.1 of Einmahl and He [2023b], because any nonincreasing right-continuous function with left limits which is regularly varying with negative index coincides with a survival function in a neighborhood of infinity. There is typically a natural choice of X, which is nonetheless not unique. However, its survival function $x \mapsto \mathbb{P}(X > x)$ is uniquely determined up to asymptotic equivalence as $x \to \infty$, and thus so are its extreme quantiles.

Condition $\mathcal{I}(\mathbb{P}_X, \gamma, a)$ is equivalent to assuming that there exist a distribution $\mathbb{P}_X$ of a heavy-tailed random variable X and $\gamma > 0$ such that, whenever (u_n) tends to infinity with $n\mathbb{P}(X > u_n) \to \infty$,

$$\exists \varepsilon > 0,\ \lim_{n\to\infty} \int_1^{\infty} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{P}(X_i^{(n)} > u_n x)}{\mathbb{P}(X > u_n)} - x^{-1/\gamma} \right| \nu_{a,\varepsilon}(\mathrm{d}x) = 0,$$

where $\nu_{a,\varepsilon}(\mathrm{d}x) = x^{a-1+\varepsilon}\mathbb{1}\{x \ge 1\}\mathrm{d}x + \delta_1(\mathrm{d}x)$, with δ_1 being the Dirac probability mass at 1. This relates assumption $\mathcal{I}(\mathbb{P}_X, \gamma, a)$ to the weighted L^1-convergence of a sequence of functions measuring the distance between the right tail of the average distribution of the observations and that of $\mathbb{P}_X$. Lemma A.4(i) in the Supplementary Material document shows that this condition is substantially weaker than the Stability Assumption 2.2 of Einmahl and He [2023b]. We note, furthermore, that unlike in this Stability Assumption, we do not require the existence of a bounded probability density function relative to the average distribution function F_n. We provide in Section 3 natural and important examples where $\mathcal{I}(\mathbb{P}_X, \gamma, a)$ is satisfied but Assumption 2.2 of Einmahl and He [2023b] is not.

In the specific setting $a = 0$, we also introduce a weaker version of $\mathcal{I}(\mathbb{P}_X, \gamma, 0)$ under which our results will still hold if slightly stronger assumptions on f and X are made. This will in particular cover the asymptotic analysis of the Hill estimator of γ, obtained with $f : x \mapsto \log x$ which is slowly varying, *i.e.* regularly varying with index $a = 0$.

$\mathcal{J}(\mathbb{P}_X, \gamma)$ Assumption $\mathcal{H}_1(\mathbb{P}_X, \gamma)$ holds and, for some $\varepsilon > 0$,

$$\lim_{n\to\infty} \int_1^\infty \left| \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{P}(X_i^{(n)} > u_n x)}{\mathbb{P}(X > u_n)} - x^{-1/\gamma} \right| x^{-1}(1 + (\log x)^{1+\varepsilon}) \mathrm{d}x = 0$$

whenever (u_n) is a sequence tending to infinity with $n\mathbb{P}(X > u_n) \to \infty$.

Remark 2.2 (On assumption $\mathcal{J}(\mathbb{P}_X, \gamma)$)**.** Similarly to $\mathcal{I}(\mathbb{P}_X, \gamma, a)$, the integral condition part of $\mathcal{J}(\mathbb{P}_X, \gamma)$ is equivalent to assuming that there exists $\varepsilon > 0$ such that, for any sequence (u_n) tending to infinity with $n\mathbb{P}(X > u_n) \to \infty$,

$$\lim_{n\to\infty} \int_1^\infty \left| \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{P}(X_i^{(n)} > u_n x)}{\mathbb{P}(X > u_n x)} - 1 \right| x^{-1/\gamma - 1}(1 + (\log x)^{1+\varepsilon}) \mathrm{d}x = 0.$$

The fact that $x^{-\varepsilon}(1 + (\log x)^{1+\varepsilon})$ is bounded in $x \geq 1$ makes this condition weaker than $\mathcal{I}(\mathbb{P}_X, \gamma, 0)$.

We also need an assumption, similar to Assumption 2.4 of Einmahl and He [2023b], in order to quantify the asymptotic variance of the estimators $X_{n-k:n}^{(n)}$ and $\widehat{e}_{f,n}(k, X_{n-k:n}^{(n)})$.

$\mathcal{H}_2(\mathbb{P}_X, \gamma, R)$ Assumption $\mathcal{H}_1(\mathbb{P}_X, \gamma)$ holds and there is a function R on $[0, \infty)^2$ such that, for any sequence (u_n) tending to infinity with $n\mathbb{P}(X > u_n) \to \infty$, and for all $(x, y) \in (0, \infty]^2$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{P}(X_i^{(n)} > u_n x)\, \mathbb{P}(X_i^{(n)} > u_n y)}{\mathbb{P}(X > u_n)} = R(x^{-1/\gamma}, y^{-1/\gamma}).$$

The function R is a special case of tail copula, see *e.g.* Schmidt and Stadtmüller [2006] for a list of properties, and as such it is nonnegative, symmetric, nondecreasing in each argument, satisfies $R(0, v) = R(u, 0) = 0$ for any $u, v \geq 0$, and is 1-homogeneous, and hence continuous on $[0, \infty)^2$. The specific value $R(1, 1) \in [0, 1]$, which is zero if and only if R is identically 0, provides a measure of heterogeneity among the observations and plays an essential role in asymptotic variance-covariance terms. It has been termed "spurious tail dependence coefficient" in Einmahl and He [2023b] because, as pointed out in Giraitis et al. [2024], violation of the i.i.d. property can lead to spurious detection of correlation in the sense that standard errors obtained under heterogeneity may not coincide with standard errors obtained with i.i.d. draws from the limiting distribution of X.

Remark 2.3 (On $\mathcal{H}_2(\mathbb{P}_X, \gamma, R)$ and the case $R \equiv 0$)**.** A simple sufficient condition for $\mathcal{H}_2(\mathbb{P}_X, \gamma, R)$ to hold with $R \equiv 0$, as established in Lemma A.5, is that $\mathcal{H}_1(\mathbb{P}_X, \gamma)$ holds and

$$\lim_{n\to\infty} \max_{1\leq i\leq n} \mathbb{P}(X_i^{(n)} > u_n) = 0 \quad \text{whenever } u_n \to \infty \text{ with } n\mathbb{P}(X > u_n) \to \infty.$$

This occurs in numerous settings, including when the extremes of the $X_i^{(n)}$ are heteroskedastic as in Einmahl et al. [2016], or more generally when there exists an envelope function ϕ converging to 0 at infinity such that $\mathbb{P}(X_i^{(n)} > x) \leq \phi(x)$ for all x, all $n \geq 1$ and all $i \in \{1, \ldots, n\}$. It is generally reasonably easy to check whether this last "bounded heterogeneity" criterion, applies or not in a given model, as discussed in Section 3, and we shall see that $R \equiv 0$ makes our asymptotic normality results essentially similar to those found when the $X_i^{(n)}$ are identically distributed.

We complete our set of assumptions with a second-order condition on the extreme value behavior of the average limiting distribution in assumption $\mathcal{H}_1(\mathbb{P}_X, \gamma)$. Such a condition is of course already necessary in the i.i.d. setting, see Section 3.2 in de Haan and Ferreira [2006].

$\mathcal{C}_2(\gamma,\rho,A)$ There exist $\gamma>0$, $\rho\le 0$ and a measurable auxiliary function A having constant sign and converging to 0 at infinity such that the random variable X satisfies

$$\forall x>0,\ \lim_{t\to\infty}\frac{1}{A(1/\mathbb{P}(X>t))}\left(\frac{\mathbb{P}(X>tx)}{\mathbb{P}(X>t)}-x^{-1/\gamma}\right)=\frac{x^{-1/\gamma}}{\gamma^2}\int_1^x s^{\rho/\gamma-1}\,\mathrm{d}s.$$

By Theorem 2.3.9 on p.48 in de Haan and Ferreira [2006], an equivalent version of this condition is

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

where $U:t\mapsto\inf\{x\in\mathbb{R}\,|\,1/\mathbb{P}(X>x)\ge t\}=q(1-1/t)$ is the tail quantile function of X. Throughout we shall use the notation $q(1-k/n)$ and $U(n/k)$ interchangeably.

2.2 Asymptotics of tail empirical mean f-excesses and Expected Shortfall

We first derive a joint asymptotic normality result for a finite number of tail empirical f-excesses together with an intermediate quantile. This will form the basis for our subsequent asymptotic results.

Theorem 2.1 (Joint convergence of tail empirical mean excesses)**.** *Pick $d\ge 1$ and assume that, in a neighborhood of infinity, $f_1,\ldots,f_d$ are increasing, continuously differentiable functions having regularly varying derivatives with index $a_1-1,\ldots,a_d-1$, respectively, for some $a_1,\ldots,a_d\ge 0$. Assume also that $\mathcal{H}_2(\mathbb{P}_X,\gamma,R)$ and $\mathcal{I}(\mathbb{P}_X,\gamma,2\max_{1\le j\le d}a_j)$ hold with $2\max_{1\le j\le d}a_j\gamma<1$, and that X satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$. Suppose finally that $k=k(n)$ tends to infinity with $k/n\to 0$, $\sqrt{k}A(n/k)=\mathrm{O}(1)$ and there is $\mu\in\mathbb{R}$ such that the following bias condition holds:*

$$\mathcal{B}(\mu)\qquad \forall t\in\mathbb{R},\ \lim_{n\to\infty}\sqrt{k}\left(\frac{1}{n}\sum_{i=1}^n\frac{\mathbb{P}(X_i^{(n)}>(1+t/\sqrt{k})U(n/k))}{\mathbb{P}(X>(1+t/\sqrt{k})U(n/k))}-1\right)=\mu.$$

Then

$$\sqrt{k}\left(\frac{\widehat{e}_{f_1,n}(k,X_{n-k:n}^{(n)})}{e_{f_1,n}(k,U(n/k))}-1,\ldots,\frac{\widehat{e}_{f_d,n}(k,X_{n-k:n}^{(n)})}{e_{f_d,n}(k,U(n/k))}-1,\frac{X_{n-k:n}^{(n)}}{U(n/k)}-1\right)$$
$$\xrightarrow{\mathrm{d}}\mathcal{N}\left((-(1-a_1\gamma)\mu,\ldots,-(1-a_d\gamma)\mu,\gamma\mu),\boldsymbol{V}-\boldsymbol{M}\right)$$

where $\boldsymbol{V}=\boldsymbol{V}(\gamma,a_1,\ldots,a_d)$ is the $(d+1)\times(d+1)$ symmetric positive semidefinite matrix with elements

$$V_{j,\ell}=\frac{1}{1-(a_j+a_\ell)\gamma}+a_ja_\ell\gamma^2,\ V_{j,d+1}=a_j\gamma^2\ \text{ and }\ V_{d+1,d+1}=\gamma^2$$

for each $j\in\{1,\ldots,d\}$, and, with the notation $I_1(a)=I_1(\gamma,a,R)=\int_1^\infty x^{a-1}R(x^{-1/\gamma},1)\,\mathrm{d}x$ and $I_2(a,b)=I_2(\gamma,a,b,R)=\int_{[1,\infty)^2}x^{a-1}y^{b-1}R(x^{-1/\gamma},y^{-1/\gamma})\,\mathrm{d}x\,\mathrm{d}y$, $\boldsymbol{M}=\boldsymbol{M}(\gamma,a_1,\ldots,a_d,R)$ is the $(d+1)\times(d+1)$ symmetric matrix with elements

$$M_{j,\ell}=(1-a_j\gamma)(1-a_\ell\gamma)\left(\frac{I_2(a_j,a_\ell)}{\gamma^2}-\frac{I_1(a_j)+I_1(a_\ell)}{\gamma}+R(1,1)\right),$$
$$M_{j,d+1}=\gamma(1-a_j\gamma)\left(\frac{I_1(a_j)}{\gamma}-R(1,1)\right)\ \text{ and }\ M_{d+1,d+1}=\gamma^2R(1,1),$$

for each $j,\ell\in\{1,\ldots,d\}$. If actually $a_1=\cdots=a_d=0$ and $|xf_j'(tx)/f_j'(t)-1|$ converges to 0 for each $j\in\{1,\ldots,d\}$ as $t\to\infty$, uniformly in $x\ge 1$, then condition $\mathcal{I}(\mathbb{P}_X,\gamma,0)$ can be replaced by the weaker version $\mathcal{J}(\mathbb{P}_X,\gamma)$.

As a corollary, we derive the joint asymptotic normality of the mean f-excess estimator $\widehat{e}_{f,n}(k, X_{n-k:n}^{(n)})$, a tail empirical g-Expected Shortfall $\widehat{\mathrm{ES}}_{g,n}(k, X_{n-k:n}^{(n)})$, and the intermediate empirical quantile $X_{n-k:n}^{(n)}$.

Corollary 2.1 (Convergence of the tail empirical mean excess and Expected Shortfall I)**.** *Assume that, in a neighborhood of infinity, f and g are increasing, continuously differentiable functions having regularly varying derivatives with index $a-1$ and $b-1$, respectively, for some $a, b \geq 0$, and that $g(x) \to \infty$ as $x \to \infty$. Assume also that $\mathcal{H}_2(\mathbb{P}_X, \gamma, R)$ and $\mathcal{I}(\mathbb{P}_X, \gamma, 2\max(a,b))$ hold with $2\max(a,b)\gamma < 1$, and that X satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$. Suppose finally that $k = k(n)$ tends to infinity with $k/n \to 0$, $\sqrt{k}A(n/k) = \mathrm{O}(1)$ and there is $\mu \in \mathbb{R}$ such that condition $\mathcal{B}(\mu)$ holds. Then*

$$\sqrt{k}\left(\frac{\widehat{e}_{f,n}(k, X_{n-k:n}^{(n)})}{e_{f,n}(k, U(n/k))} - 1, \frac{\widehat{\mathrm{ES}}_{g,n}(k, X_{n-k:n}^{(n)})}{\mathrm{ES}_{g,n}(k, U(n/k))} - 1, \frac{X_{n-k:n}^{(n)}}{U(n/k)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left((-(1-a\gamma)\mu, 0, \gamma\mu), \mathfrak{V} - \mathfrak{M}\right)$$

where $\mathfrak{V} = \mathfrak{V}(\gamma, a, b)$ is the 3×3 symmetric positive semidefinite matrix with elements

$$\mathfrak{V}_{1,1} = \frac{1}{1-2a\gamma} + a^2\gamma^2, \ \mathfrak{V}_{1,2} = \frac{b\gamma(1 + a\gamma - a(a+b)\gamma^2)}{1-(a+b)\gamma}, \ \mathfrak{V}_{1,3} = a\gamma^2,$$
$$\mathfrak{V}_{2,2} = \frac{2b^2\gamma^2(1-b\gamma)}{1-2b\gamma}, \ \mathfrak{V}_{2,3} = b\gamma^2 \quad \textit{and} \quad \mathfrak{V}_{3,3} = \gamma^2,$$

and, with the notation of Theorem 2.1, $\mathfrak{M} = \mathfrak{M}(\gamma, a, b, R)$ is the 3×3 symmetric matrix with elements

$$\mathfrak{M}_{1,1} = (1-a\gamma)^2\left(\frac{I_2(a,a)}{\gamma^2} - \frac{2}{\gamma}I_1(a) + R(1,1)\right),$$
$$\mathfrak{M}_{1,2} = b(1-a\gamma)(1-b\gamma)\left(\frac{I_2(a,b)}{\gamma} - I_1(b)\right), \ \mathfrak{M}_{1,3} = \gamma(1-a\gamma)\left(\frac{1}{\gamma}I_1(a) - R(1,1)\right),$$
$$\mathfrak{M}_{2,2} = b^2(1-b\gamma)^2 I_2(b,b), \ \mathfrak{M}_{2,3} = b\gamma(1-b\gamma)I_1(b) \quad \textit{and} \quad \mathfrak{M}_{3,3} = \gamma^2 R(1,1).$$

If actually $a = b = 0$ and $|xf'(tx)/f'(t) - 1|$ and $|xg'(tx)/g'(t) - 1|$ converge to 0 as $t \to \infty$, uniformly in $x \geq 1$, then condition $\mathcal{I}(\mathbb{P}_X, \gamma, 0)$ can be replaced by the weaker version $\mathcal{J}(\mathbb{P}_X, \gamma)$.

Corollary 2.1 extends the asymptotic normality result of Theorem 1 in Stupfler [2019], valid in the i.i.d. setting only and restricted to the joint asymptotic normality of an average f-excess and an intermediate empirical quantile. In particular, we find back the asymptotic variance of Theorem 1 in Stupfler [2019] by taking $R \equiv 0$ in $\mathfrak{M}_{1,1}$.

Note further that, from the regular variation properties of $\overline{F}_n$ and f',

$$e_{f,n}(k, U(n/k)) = \frac{n}{k}\int_{U(n/k)}^{\infty} \overline{F}_n(x) f'(x)\,\mathrm{d}x \sim \frac{\gamma}{1-a\gamma}U(n/k)f'(U(n/k)).$$

In our third asymptotic normality result, which will prove particularly useful for the asymptotic analysis of the Hill estimator, we replace $e_{f,n}(k, U(n/k))$ with its natural asymptotic equivalent; for the sake of simplicity, and since this contains the applications of our theory here, we assume in this result that f' is equal to a power function in a neighborhood of infinity.

Theorem 2.2 (Convergence of the tail empirical mean excess and Expected Shortfall II)**.** *Work under the assumptions of Corollary 2.1, with the added requirement that $f'(x)$ is proportional to x^{a-1} for x sufficiently large. Assume further that there is $m = m(\gamma, a) \in \mathbb{R}$ such that*

$$\mathcal{IB}(\gamma, a, m) \qquad \lim_{n\to\infty} \sqrt{k}\int_1^{\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{P}(X_i^{(n)} > U(n/k)x)}{\mathbb{P}(X > U(n/k)x)} - 1\right)x^{-1/\gamma+a-1}\mathrm{d}x = m.$$

Suppose finally that there is $\lambda \in \mathbb{R}$ such that $\sqrt{k}A(n/k) \to \lambda$. Then

$$\sqrt{k}\left(\frac{\widehat{e}_{f,n}(k, X_{n-k:n}^{(n)})}{U(n/k)f'(U(n/k))} - \frac{\gamma}{1-a\gamma}, \frac{\widehat{\mathrm{ES}}_{g,n}(k, X_{n-k:n}^{(n)})}{\mathrm{ES}_{g,n}(k, U(n/k))} - 1, \frac{X_{n-k:n}^{(n)}}{U(n/k)} - 1\right)$$
$$\xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\frac{\lambda}{(1-a\gamma)(1-a\gamma-\rho)} + m - \gamma\mu, 0, \gamma\mu\right), \gamma^2(\boldsymbol{\mathcal{V}} - \boldsymbol{\mathcal{M}})\right)$$

where $\boldsymbol{\mathcal{V}} = \boldsymbol{\mathcal{V}}(\gamma, a, b)$ is the 3×3 symmetric positive semidefinite matrix with elements

$$\mathcal{V}_{1,1} = \frac{1}{(1-a\gamma)^2}\left(\frac{1}{1-2a\gamma} + a^2\gamma^2\right), \ \mathcal{V}_{1,2} = \frac{b(1+a\gamma-a(a+b)\gamma^2)}{(1-a\gamma)(1-(a+b)\gamma)}, \ \mathcal{V}_{1,3} = \frac{a\gamma}{1-a\gamma},$$
$$\mathcal{V}_{2,2} = \frac{2b^2(1-b\gamma)}{1-2b\gamma}, \ \mathcal{V}_{2,3} = b \ \text{ and } \ \mathcal{V}_{3,3} = 1,$$

and, with the notation of Theorem 2.1, $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{M}}(\gamma, a, b, R)$ is the 3×3 symmetric matrix with elements

$$\mathcal{M}_{1,1} = \frac{I_2(a,a)}{\gamma^2} - \frac{2}{\gamma}I_1(a) + R(1,1), \ \mathcal{M}_{1,2} = b(1-b\gamma)\left(\frac{I_2(a,b)}{\gamma^2} - \frac{I_1(b)}{\gamma}\right)$$
$$\mathcal{M}_{1,3} = \frac{I_1(a)}{\gamma} - R(1,1), \ \mathcal{M}_{2,2} = \frac{b^2(1-b\gamma)^2}{\gamma^2}I_2(b,b), \ \mathcal{M}_{2,3} = \frac{b(1-b\gamma)}{\gamma}I_1(b)$$
$$\text{and } \ \mathcal{M}_{3,3} = R(1,1).$$

If actually $a = b = 0$ and $|xg'(tx)/g'(t) - 1|$ converges to 0 as $t \to \infty$, uniformly in $x \geq 1$, then condition $\mathcal{I}(\mathbb{P}_X, \gamma, 0)$ can be replaced by the weaker version $\mathcal{J}(\mathbb{P}_X, \gamma)$ provided $\rho < 0$.

Remark 2.4 (On the bias conditions $\mathcal{B}(\mu)$ and $\mathcal{IB}(\gamma, a, m)$)**.** Condition $\mathcal{B}(\mu)$ (resp. $\mathcal{IB}(\gamma, a, m)$) complements $\mathcal{H}_1(\mathbb{P}_X, \gamma)$ (resp. $\mathcal{I}(\mathbb{P}_X, \gamma, a)$) by quantifying the bias incurred by treating the statistic $X_{n-k:n}^{(n)}$ (resp. $\widehat{e}_{f,n}(k, X_{n-k:n}^{(n)})$) as an estimator of the quantile $U(n/k)$ of X (resp. of the mean f-excess of X above $U(n/k)$), rather than as an estimator of the corresponding quantile of this average distribution (resp. of the quantity $e_{f,n}(k, U(n/k))$).

Remark 2.5 (On the variance-covariance structures)**.** In Theorem 2.1, Corollary 2.1 and Theorem 2.2, the matrices $\boldsymbol{V}$, $\mathfrak{V}$ and $\boldsymbol{\mathcal{V}}$ are the asymptotic covariance matrices of the estimators under the i.i.d. assumption. The matrices $\boldsymbol{M}$, $\mathfrak{M}$ and $\boldsymbol{\mathcal{M}}$ represent the influence of heterogeneity upon the asymptotic dependence structure of these estimators. Interestingly, in simple examples, such as in Corollary 2.2 below, one can easily show that their diagonal elements are nonnegative, in which case they stand for the gain in marginal uncertainty obtained by accounting for heterogeneity in the data.

2.3 Application 1: Extreme value index and extreme quantile estimation

Let us first specialize the discussion to the case $f = \log$, which corresponds to the Hill estimator

$$\widehat{\gamma}_n(k) := \widehat{e}_{\log,n}(k, X_{n-k:n}^{(n)}) = \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1:n}^{(n)} - \log X_{n-k:n}^{(n)}$$

of the extreme value index γ of the distribution $\mathbb{P}_X$. The joint convergence of $\widehat{\gamma}_n(k)$ and $X_{n-k:n}^{(n)}$ follows immediately from Theorem 2.2.

Corollary 2.2 (Convergence of the Hill estimator)**.** *Assume that $\mathcal{H}_2(\mathbb{P}_X, \gamma, R)$ and $\mathcal{I}(\mathbb{P}_X, \gamma, 0)$ hold, and that X satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$. Suppose that $k = k(n)$ tends to infinity with $k/n \to 0$ and $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$, and that there exist $\mu, m \in \mathbb{R}$ such that conditions $\mathcal{B}(\mu)$ and $\mathcal{IB}(\gamma, 0, m)$ hold. Then*

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma, \frac{X_{n-k:n}^{(n)}}{U(n/k)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left((\lambda b_{\mathrm{H}}(\rho) + m - \gamma\mu, \gamma\mu), \gamma^2(\mathrm{I}_2 - \boldsymbol{\mathcal{M}}_0)\right)$$

where $b_{\mathrm{H}}(\rho) = 1/(1-\rho)$ *and* $\boldsymbol{\mathcal{M}}_0 = \boldsymbol{\mathcal{M}}_0(R)$ *is the* 2×2 *symmetric matrix with elements*

$$[\mathcal{M}_0]_{1,1} = [\mathcal{M}_0]_{2,2} = R(1,1) \quad \text{and} \quad [\mathcal{M}_0]_{1,2} = \int_0^1 u^{-1} R(u,1)\,\mathrm{d}u - R(1,1).$$

When $\rho < 0$*, condition* $\mathcal{I}(\mathbb{P}_X, \gamma, 0)$ *can be replaced by the weaker version* $\mathcal{J}(\mathbb{P}_X, \gamma)$*.*

In Corollary 2.2, $\lambda b_{\mathrm{H}}(\rho)$ and γ^2 are the asymptotic bias and variance, respectively, of the Hill estimator for i.i.d. data. Interestingly, unlike in the i.i.d. setting, the Hill estimator and its corresponding random threshold $X_{n-k:n}^{(n)}$ need not be asymptotically independent when heterogeneity is present, and their common asymptotic variance $\gamma^2(1-R(1,1))$ does not exceed its counterpart γ^2 obtained in the homogeneous case when the data is made of random draws from X.

Now, based on both $\widehat{\gamma}_n(k)$ and $X_{n-k:n}^{(n)}$, define for an extreme level τ_n', that may converge to 1 at an arbitrarily fast rate as $n \to \infty$, the Weissman estimator

$$\widehat{q}_n^{\star}(\tau_n') \equiv \widehat{q}_n^{\star}(\tau_n'|k) := \left(\frac{k}{n(1-\tau_n')}\right)^{\widehat{\gamma}_n(k)} X_{n-k:n}^{(n)}$$

of the extreme quantile $q(\tau_n')$ of $\mathbb{P}_X$. Its construction is justified by the heavy right tail assumption on X, of which a consequence is the approximation

$$\frac{q(\tau')}{q(\tau)} \approx \left(\frac{1-\tau'}{1-\tau}\right)^{-\gamma} \quad \text{as } \tau, \tau' \uparrow 1. \tag{1}$$

The asymptotic normality of the Weissman estimator, highlighting the interplay between the Hill estimator and the order statistic at the level $1-k/n$, essentially follows from Corollary 2.2.

Corollary 2.3 (Convergence of the Weissman estimator)**.** *Under the conditions of Corollary 2.2, if moreover* $\rho < 0$ *and* $\tau_n' \to 1$ *is such that* $d_n := k/(n(1-\tau_n')) \to \ell \in (1, \infty]$ *and* $\sqrt{k}/\log(d_n) \to \infty$*, then*

$$\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widehat{q}_n^{\star}(\tau_n')}{q(\tau_n')} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\lambda b_{\mathrm{W}}(\ell,\rho) + m - \left(1 - \frac{1}{\log(\ell)}\right)\gamma\mu, \gamma^2\sigma_{\mathrm{W}}^2(\ell)(1-\pi_{\mathrm{W}}(\ell,R))\right)$$

$$\text{where} \quad b_{\mathrm{W}}(\ell,\rho) = \frac{1}{1-\rho} - \frac{1}{\rho}\frac{\ell^{\rho}-1}{\log(\ell)}, \quad \sigma_{\mathrm{W}}^2(\ell) = 1 + \frac{1}{\log^2(\ell)}$$

$$\text{and} \quad \pi_{\mathrm{W}}(\ell,R) = R(1,1) + \frac{2}{\log(\ell)}\frac{\int_0^1 u^{-1}R(u,1)\,\mathrm{d}u - R(1,1)}{1 + 1/\log^2(\ell)}.$$

[Here and throughout, when h *converges to a finite limit at infinity, we define* $h(\infty) = \lim_{x\to\infty} h(x)$*.]*

Analogously to what was observed in Corollary 2.2, the quantities $\lambda b_{\mathrm{W}}(\ell,\rho)$ and $\gamma^2\sigma_{\mathrm{W}}^2(\ell)$ are the asymptotic bias and variance, respectively, of the Weissman extreme quantile estimator for i.i.d. data. The quantity $\pi_{\mathrm{W}}(\ell,R)$ represents the variance reduction factor when heterogeneity is present because

$$\pi_{\mathrm{W}}(\ell,R) = \frac{1}{1+1/\log^2(\ell)}\left(\left(1 - \frac{1}{\log(\ell)}\right)^2 R(1,1) + \frac{2}{\log(\ell)}\int_0^1 u^{-1}R(u,1)\,\mathrm{d}u\right) \geq 0.$$

When $\ell = \infty$, corresponding to the case of an exceedance level $1-\tau_n'$ well beyond the intermediate level k/n, one has $1-\pi_{\mathrm{W}}(\ell,R) = 1-R(1,1)$, with heterogeneity then playing a simple but crucial role in the reduction of the asymptotic variance of the Weissman estimator compared to the i.i.d. setting.

Remark 2.6 (Comparison with existing results)**.** Corollaries 2.2 and 2.3 improve on the existing results of Einmahl et al. [2016] and de Haan and Zhou [2021], by removing any assumption on the extreme value indices of the observations $X_i^{(n)}$. Most importantly, they also improve on the recent results in Einmahl and He [2023b] for $X_{n-k:n}^{(n)}$ in their Corollary 2.2, for $\widehat{\gamma}_n(k)$ in their Theorem 2.2, and for $\widehat{q}_n^{\star}(\tau_n'|k)$ in their Theorem 4.1. The added value is twofold:

• We obtain the asymptotic normality jointly for $X^{(n)}_{n-k:n}$ and $\widehat{\gamma}_n(k)$ in Corollary 2.2 with the explicit underlying bias and variance-covariance terms, and by identifying the distinct sources contributing to the asymptotic bias, before we establish in Corollary 2.3 the asymptotic normality of $\widehat{q}^\star_n(\tau'_n|k)$ and provide closed-form expressions for the corresponding asymptotic bias and variance.

• Our conditions are weaker than those of Einmahl and He [2023b] in their own context which ignores the asymptotic bias of the Hill and Weissman estimators. While we assume similar conditions $\mathcal{H}_1(\mathbb{P}_X,\gamma)$ and $\mathcal{H}_2(\mathbb{P}_X,\gamma,R)$ to their Assumptions 2.1 and 2.4 respectively, as well as the same Assumption 2.3 about the intermediate sequence $k = k(n)$, we use a substantially weaker integral condition $\mathcal{I}(\mathbb{P}_X,\gamma,a)$ than their Stability Assumption 2.2, as proved in Lemma A.4. In particular, we do not impose any boundedness or smoothness condition on the average distribution, nor do we rely on Gaussian approximations as in Einmahl and He [2023b]. This is because the structure of mean f-excess estimators is inherently suited to L^1-based arguments rather than to more elaborate L^∞-approaches, of which Gaussian approximations are part. Finally, instead of their Condition (4), and of Condition (8) in their Theorem 4.1, we employ the usual second-order condition $\mathcal{C}_2(\gamma,\rho,A)$ (see, *e.g.*, Theorem 2.3.9 of de Haan and Ferreira [2006]), so that we are able to quantify asymptotic biases.

As a side note, while the results of Einmahl and He [2023b] target the estimation of tail (intermediate and extreme) quantiles of the average distribution F_n, Corollaries 2.2 and 2.3 focus on the tail quantities of the heavy-tailed model distribution $\mathbb{P}_X$ itself. This is advantageous in certain applications: for example, from a robustness point of view, the population distribution $\mathbb{P}_X$ is not subject to contamination issues as would be the case for the average distribution.

2.4 Application 2: Tail Expected Shortfall estimation

For $\tau \in (0,1)$, define the Expected Shortfall of a random variable X above level $q(\tau)$ as $\mathrm{ES}(\tau) = \mathbb{E}(X|X > q(\tau))$. This formulation as a conditional expectation immediately suggests the following empirical estimator of $\mathrm{ES}(1-k/n)$ based on the average distribution of the $X_i^{(n)}$ above level $X^{(n)}_{n-k:n}$:

$$\widehat{\mathrm{ES}}_n(1-k/n) = \widehat{\mathrm{ES}}_{\mathrm{Id},n}(k, X^{(n)}_{n-k:n}) = \frac{1}{k}\sum_{i=1}^{k} X^{(n)}_{n-i+1:n} = \widehat{e}_{\mathrm{Id},n}(k, X^{(n)}_{n-k:n}) + X^{(n)}_{n-k:n}.$$

Moreover, if the extreme value index γ of X is smaller than 1, one obtains, from Proposition B.1.10 on p.369 in de Haan and Ferreira [2006], that

$$\frac{\mathrm{ES}(\tau)}{q(\tau)} = 1 + \frac{\mathbb{E}(X - q(\tau)\,|\,X > q(\tau))}{q(\tau)} = 1 + \int_1^\infty \frac{\overline{F}(xq(\tau))}{\overline{F}(q(\tau))}\,\mathrm{d}x \to 1 + \int_1^\infty x^{-1/\gamma}\,\mathrm{d}x = \frac{1}{1-\gamma} \tag{2}$$

as $\tau \uparrow 1$. In other words, the Expected Shortfall at an extreme level is proportional to its corresponding extreme quantile, which means that it satisfies an asymptotic extrapolation relationship analogous to (1) that justifies the construction of the extrapolated Expected Shortfall estimator

$$\widehat{\mathrm{ES}}^\star_n(\tau'_n) \equiv \widehat{\mathrm{ES}}^\star_n(\tau'_n|k) = \left(\frac{k}{n(1-\tau'_n)}\right)^{\widehat{\gamma}_n(k)} \widehat{\mathrm{ES}}_n(1-k/n)$$

with $\tau'_n \to 1$. Equation (2) also suggests the following alternative, quantile-based estimators

$$\widetilde{\mathrm{ES}}_n(1-k/n) = \frac{X^{(n)}_{n-k:n}}{1-\widehat{\gamma}_n(k)} \text{ and } \widetilde{\mathrm{ES}}^\star_n(\tau'_n) \equiv \widetilde{\mathrm{ES}}^\star_n(\tau'_n|k) = \frac{\widehat{q}^\star_n(\tau'_n)}{1-\widehat{\gamma}_n(k)}.$$

We next consider the asymptotic normality of the estimators $\widehat{\mathrm{ES}}_n(1-k/n)$, $\widetilde{\mathrm{ES}}_n(1-k/n)$, $\widehat{\mathrm{ES}}^\star_n(\tau'_n)$ and $\widetilde{\mathrm{ES}}^\star_n(\tau'_n)$. We first focus on $\widehat{\mathrm{ES}}_n(1-k/n)$ and prove its joint asymptotic normality with $\widehat{\gamma}_n(k)$.

Corollary 2.4 (Joint convergence of the empirical Expected Shortfall and Hill estimators)**.** *Assume that $\mathcal{H}_2(\mathbb{P}_X, \gamma, R)$ and $\mathcal{I}(\mathbb{P}_X, \gamma, 2)$ hold with $0 < \gamma < 1/2$, and that X satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$. Suppose that $k = k(n)$ tends to infinity with $k/n \to 0$ and $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$, and that there exist $\mu, m_0, m_1 \in \mathbb{R}$ such that conditions $\mathcal{B}(\mu)$, $\mathcal{IB}(\gamma, 0, m_0)$ and $\mathcal{IB}(\gamma, 1, m_1)$ hold. Then*

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma, \frac{\widehat{\mathrm{ES}}_n(1-k/n)}{\mathrm{ES}(1-k/n)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left((\lambda b_{\mathrm{H}}(\rho) + m_0 - \gamma\mu, (1-\gamma)m_1), \gamma^2(\boldsymbol{V}_{\mathrm{E}} - \boldsymbol{M}_{\mathrm{E}})\right)$$

where $\boldsymbol{V}_{\mathrm{E}} = \boldsymbol{V}_{\mathrm{E}}(\gamma)$ is the 2×2 symmetric positive semidefinite matrix with elements

$$[V_{\mathrm{E}}]_{1,1} = 1, \ [V_{\mathrm{E}}]_{1,2} = \frac{1}{1-\gamma} \ \textit{and} \ [V_{\mathrm{E}}]_{2,2} = \frac{2(1-\gamma)}{1-2\gamma},$$

and $\boldsymbol{M}_{\mathrm{E}} = \boldsymbol{M}_{\mathrm{E}}(\gamma, R)$ is the 2×2 symmetric matrix with elements

$$[M_{\mathrm{E}}]_{1,1} = R(1,1), \ [M_{\mathrm{E}}]_{1,2} = (1-\gamma)\left(\int_{(0,1]^2} u^{-\gamma-1}v^{-1}R(u,v)\,\mathrm{d}u\,\mathrm{d}v - \int_0^1 u^{-\gamma-1}R(u,1)\,\mathrm{d}u\right)$$

$$\textit{and} \ [M_{\mathrm{E}}]_{2,2} = (1-\gamma)^2\int_{(0,1]^2} u^{-\gamma-1}v^{-\gamma-1}R(u,v)\,\mathrm{d}u\,\mathrm{d}v.$$

We turn to the calculation of the joint limiting distribution of $\widetilde{\mathrm{ES}}_n(1-k/n)$ and $\widehat{\gamma}_n(k)$. This is essentially a consequence of Corollary 2.2 combined with the delta-method.

Corollary 2.5 (Joint convergence of the quantile-based Expected Shortfall and Hill estimators)**.** *Work under the conditions of Corollary 2.2 and assume further that $0 < \gamma < 1$. Then*

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma, \frac{\widetilde{\mathrm{ES}}_n(1-k/n)}{\mathrm{ES}(1-k/n)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\lambda b_{\mathrm{H}}(\rho) + m - \gamma\mu, \lambda b_{\mathrm{QB}}(\gamma,\rho) + \frac{m-\gamma^2\mu}{1-\gamma}\right), \gamma^2(\boldsymbol{V}_{\mathrm{QB}} - \boldsymbol{M}_{\mathrm{QB}})\right)$$

where $b_{\mathrm{QB}}(\gamma,\rho) = -\gamma\rho/((1-\gamma)(1-\rho)(1-\gamma-\rho))$, $\boldsymbol{V}_{\mathrm{QB}} = \boldsymbol{V}_{\mathrm{QB}}(\gamma)$ is the 2×2 symmetric positive semidefinite matrix with elements

$$[V_{\mathrm{QB}}]_{1,1} = 1, \ [V_{\mathrm{QB}}]_{1,2} = \frac{1}{1-\gamma} \ \textit{and} \ [V_{\mathrm{QB}}]_{2,2} = 1 + \frac{1}{(1-\gamma)^2},$$

and $\boldsymbol{M}_{\mathrm{QB}} = \boldsymbol{M}_{\mathrm{QB}}(\gamma, R)$ is the 2×2 symmetric matrix with elements

$$[M_{\mathrm{QB}}]_{1,1} = R(1,1), \ [M_{\mathrm{QB}}]_{1,2} = \frac{\gamma}{1-\gamma}R(1,1) + \int_0^1 u^{-1}R(u,1)\,\mathrm{d}u$$

$$\textit{and} \ [M_{\mathrm{QB}}]_{2,2} = \left(1 + \frac{1}{(1-\gamma)^2}\right)R(1,1) + \frac{2}{1-\gamma}\left(\int_0^1 u^{-1}R(u,1)\,\mathrm{d}u - R(1,1)\right).$$

It is now easy to get the asymptotic distribution of the extrapolated estimators $\widehat{\mathrm{ES}}^{\star}_n(\tau_n')$ and $\widetilde{\mathrm{ES}}^{\star}_n(\tau_n')$ since these estimators are functions of $\widehat{\gamma}_n(k)$ and, respectively, of $\widehat{\mathrm{ES}}_n(1-k/n)$ and $\widetilde{\mathrm{ES}}_n(1-k/n)$.

Corollary 2.6 (Convergence of the extrapolated Expected Shortfall estimators)**.** *Assume that $\mathcal{H}_2(\mathbb{P}_X, \gamma, R)$ holds, and that X satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$ with $\rho < 0$. Suppose that $k = k(n)$ tends to infinity with $k/n \to 0$ and $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$, and that condition $\mathcal{B}(\mu)$ holds. Let finally $\tau_n' \to 1$ be such that $d_n := k/(n(1-\tau_n')) \to \ell \in (1, \infty]$ and $\sqrt{k}/\log(d_n) \to \infty$.*

(i) Suppose that $\mathcal{I}(\mathbb{P}_X, \gamma, 2)$ holds with $0 < \gamma < 1/2$, and that conditions $\mathcal{IB}(\gamma, 0, m_0)$ and $\mathcal{IB}(\gamma, 1, m_1)$ hold as well. Then

$$\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widehat{\mathrm{ES}}^{\star}_n(\tau_n')}{\mathrm{ES}(\tau_n')} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\lambda b_{\mathrm{XE}}(\ell,\gamma,\rho) + m_0 - \gamma\mu + \frac{(1-\gamma)m_1}{\log(\ell)}, \gamma^2\sigma^2_{\mathrm{XE}}(\ell,\gamma)\left(1 - \pi_{\mathrm{XE}}(\ell,\gamma,R)\right)\right)$$

where

$$b_{\mathrm{XE}}(\ell,\gamma,\rho)=\frac{1}{1-\rho}-\frac{1-\gamma}{\rho(1-\gamma-\rho)}\frac{\ell^{\rho}-1}{\log(\ell)},$$

$$\sigma^2_{\mathrm{XE}}(\ell,\gamma)=1+\frac{2}{(1-\gamma)\log(\ell)}+\frac{2(1-\gamma)}{(1-2\gamma)\log^2(\ell)} \quad \textit{and} \quad \pi_{\mathrm{XE}}(\ell,\gamma,R)=\frac{\delta_{\mathrm{XE}}(\ell,\gamma,R)}{\sigma^2_{\mathrm{XE}}(\ell,\gamma)}$$

$$\textit{with } \delta_{\mathrm{XE}}(\ell,\gamma,R)=R(1,1)+\frac{(1-\gamma)^2}{\log^2(\ell)}\left(\int_{(0,1]^2}u^{-\gamma-1}v^{-\gamma-1}R(u,v)\,\mathrm{d}u\,\mathrm{d}v\right)$$
$$+2\frac{1-\gamma}{\log(\ell)}\left(\int_{(0,1]^2}u^{-\gamma-1}v^{-1}R(u,v)\,\mathrm{d}u\,\mathrm{d}v-\int_0^1 u^{-\gamma-1}R(u,1)\,\mathrm{d}u\right).$$

(ii) Suppose that $\mathcal{I}(\mathbb{P}_X,\gamma,0)$ holds with $0<\gamma<1$, and that condition $\mathcal{IB}(\gamma,0,m)$ holds as well. Then

$$\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widetilde{\mathrm{ES}}^{\star}_n(\tau_n')}{\mathrm{ES}(\tau_n')}-1\right)\xrightarrow{\mathrm{d}}\mathcal{N}\left(\lambda b_{\mathrm{XQB}}(\ell,\gamma,\rho)+m-\gamma\mu+\frac{m-\gamma^2\mu}{(1-\gamma)\log(\ell)},\gamma^2\sigma^2_{\mathrm{XQB}}(\ell,\gamma)(1-\pi_{\mathrm{XQB}}(\ell,\gamma,R))\right)$$

where

$$b_{\mathrm{XQB}}(\ell,\gamma,\rho)=\frac{1}{1-\rho}-\frac{1}{(1-\gamma-\rho)\log(\ell)}\left(\frac{\gamma\rho}{(1-\gamma)(1-\rho)}+\frac{(1-\gamma)(\ell^{\rho}-1)}{\rho}\right),$$

$$\sigma^2_{\mathrm{XQB}}(\ell,\gamma)=\left(1+\frac{1}{(1-\gamma)\log(\ell)}\right)^2+\frac{1}{\log^2(\ell)} \quad \textit{and} \quad \pi_{\mathrm{XQB}}(\ell,\gamma,R)=\frac{\delta_{\mathrm{XQB}}(\ell,\gamma,R)}{\sigma^2_{\mathrm{XQB}}(\ell,\gamma)}$$

$$\textit{with } \delta_{\mathrm{XQB}}(\ell,\gamma,R)=\left(\left(1+\frac{1}{(1-\gamma)\log(\ell)}\right)^2+\frac{1}{\log^2(\ell)}\right)R(1,1)$$
$$+\frac{2}{\log(\ell)}\left(1+\frac{1}{(1-\gamma)\log(\ell)}\right)\left(\int_0^1 u^{-1}R(u,1)\,\mathrm{d}u-R(1,1)\right).$$

Remark 2.7 (On the variance-covariance structures and variance reduction factors)**.** In Corollaries 2.4 and 2.5, the matrices $\boldsymbol{M}_{\mathrm{E}}$ and $\boldsymbol{M}_{\mathrm{QB}}$ represent the influence of heterogeneity upon the asymptotic covariance structure of the estimators, given by $\boldsymbol{V}_{\mathrm{E}}$ and $\boldsymbol{V}_{\mathrm{QB}}$ respectively under the i.i.d. assumption. Note that the reduction factors of the asymptotic variances are different (unless $R(1,1)=0$) for the two intermediate Expected Shortfall estimators $\widehat{\mathrm{ES}}_n(1-k/n)$ and $\widetilde{\mathrm{ES}}_n(1-k/n)$ and is given by $[M_{\mathrm{E}}]_{2,2}$ in Corollary 2.4 and $[M_{\mathrm{QB}}]_{2,2}$ in Corollary 2.5, respectively.

In Corollary 2.6, the quantities $\lambda b_{\mathrm{XE}}(\ell,\gamma,\rho)$ and $\gamma^2\sigma^2_{\mathrm{XE}}(\ell,\gamma)$ in part (i), respectively the quantities $\lambda b_{\mathrm{XQB}}(\ell,\gamma,\rho)$ and $\gamma^2\sigma^2_{\mathrm{XQB}}(\ell,\gamma)$ in part (ii), are the asymptotic bias and variance of the extrapolated Expected Shortfall estimator $\widehat{\mathrm{ES}}^{\star}_n(\tau_n')$, respectively $\widetilde{\mathrm{ES}}^{\star}_n(\tau_n')$, for i.i.d. data. These two estimators have variance reduction factors $\pi_{\mathrm{XE}}(\ell,\gamma,R)$ and $\pi_{\mathrm{XQB}}(\ell,\gamma,R)$, respectively, which both reduce to $R(1,1)$ in the case $\ell=\infty$, *i.e.*, for an extreme level τ_n' well beyond the intermediate level $1-k/n$.

3 Examples of models covered by our framework

We retain the notation of Corollaries 2.2–2.6 and draw consequences of our results in four models relevant to statistical practice. As in the previous section, we make the convention that, if h is a function converging to a finite limit at infinity, then $h(\infty)=\lim_{x\to\infty}h(x)$.

3.1 Contaminated samples

Let N be a positive integer. Assume that the available data is made of $n-N$ independent observations from a heavy-tailed distribution and N independent observations from a contaminating distribution, that is, we consider the model

(M1) The available data points $X_i^{(n)}$, $1 \le i \le n$, are independent for each n and

$$\mathbb{P}(X_i^{(n)} > x) = \begin{cases} \overline{F}(x) \text{ if } 1 \le i \le n-N, \\ \overline{G}(x) \text{ if } i \in \{n-N+1, \ldots, n\}, \end{cases}$$

where $\overline{F}$ is the survival function of a heavy-tailed random variable X with index $\gamma > 0$.

This is the so-called stratified sampling example of Stigler [1976]. In this setting, we have

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{P}(X_i^{(n)} > x)}{\overline{F}(x)} = \frac{n-N}{n} + \frac{N}{n\overline{F}(x)}\overline{G}(x).$$

When $N = N(n)$ stays bounded as n grows, the second term on the right-hand side converges to 0 along any sequence (u_n) such that $n\overline{F}(u_n) \to \infty$, so the intuition dictates that our assumptions will be satisfied, with the approximating distribution being the distribution of X, provided extra reasonable integrability assumptions on $\overline{G}$ hold. This motivates the following two conditions.

(I_a) There exists $\delta > 0$ such that $x \mapsto x^{a-1+\delta}\overline{G}(x)$ is integrable in a neighborhood of infinity.

(J_0) There exists $\delta > 0$ such that $x \mapsto x^{-1}(\log x)^{1+\delta}\overline{G}(x)$ is integrable in a neighborhood of infinity.

/Assume now that the number of contaminating observations diverges, that is, $N = N(n) \to \infty$. In this case, we should intuitively require a tradeoff between the degree of contamination, represented by N/n, and the relative difference in tail heaviness between the target and contaminating distributions, represented by $\overline{G}(u_n)/\overline{F}(u_n)$. This is the rationale behind the next condition (I'_a) and its variant (J'_a).

(I'_a) The survival function $\overline{G}$ is regularly varying with index $-1/\xi$ such that $\xi > \gamma$ and $a\xi < 1$, and $N = N(n) \to \infty$ is such that $N = \mathrm{o}(n^{\gamma/\xi - \delta})$ for some $\delta > 0$.

(J'_a) The survival function $\overline{G}$ is such that $x^{1/\xi}\overline{G}(x)$ converges to a finite positive limit as $x \to \infty$, where $\xi > \gamma$ and $a\xi < 1$, and $N = N(n) \to \infty$ is such that $N = \mathrm{O}(n^{\gamma/\xi})$.

We may now state the following result relative to the asymptotic behavior of the Hill, Weissman and tail Expected Shortfall estimators in model (M1).

Theorem 3.1 (Contaminated data – Hill, Weissman and tail Expected Shortfall estimators)**.** *Assume that the data is generated from model (M1). Suppose that X satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$, and let $k = k(n) \to \infty$ be such that $k/n \to 0$ and $\sqrt{k}A(n/k) \to \lambda$, as well as $\tau'_n \to 1$ be such that $d_n := k/(n(1-\tau'_n)) \to \ell \in (1, \infty]$ and $\sqrt{k}/\log(d_n) \to \infty$. Let $N = N(n)$ be such that $N\overline{G}(U(n/k))/\sqrt{k} \to c \in [0, \infty)$ (if N stays bounded, then $c = 0$ and every term proportional to c below should be read as 0).*

(i) If N is bounded and (I_0) holds, or if $N \to \infty$ and (I'_0) holds, then

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma, \frac{X_{n-k:n}^{(n)}}{U(n/k)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left((\lambda b_{\mathrm{H}}(\rho) + c(\xi - \gamma), c\gamma), \gamma^2 \mathrm{I}_2\right).$$

If moreover $\rho < 0$, then this convergence remains valid when N is bounded under the weaker assumption (J_0) instead of (I_0), or when $N \to \infty$ under the weaker assumption (J'_0) instead of (I'_0), and in any case,

$$\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widehat{q}_n^{\star}(\tau'_n)}{q(\tau'_n)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\lambda b_{\mathrm{W}}(\ell, \rho) + c\left(\xi - \gamma + \frac{\gamma}{\log(\ell)}\right), \gamma^2 \sigma_{\mathrm{W}}^2(\ell)\right).$$

(ii) If furthermore $\gamma < 1$ then

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma, \frac{\widetilde{\mathrm{ES}}_n(1-k/n)}{\mathrm{ES}(1-k/n)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\lambda b_{\mathrm{H}}(\rho) + c(\xi - \gamma), \lambda b_{\mathrm{QB}}(\gamma, \rho) + c\frac{\xi - \gamma^2}{1-\gamma}\right), \gamma^2 \boldsymbol{V}_{\mathrm{QB}}\right).$$

If moreover $\rho < 0$, then this convergence also remains valid when N is bounded under the weaker assumption (J_0) instead of (I_0), or when $N \to \infty$ under the weaker assumption (J_0') instead of (I_0'), and in any case,

$$\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widetilde{\mathrm{ES}}_n^{\star}(\tau_n')}{\mathrm{ES}(\tau_n')}-1\right)\xrightarrow{\mathrm{d}}\mathcal{N}\left(\lambda b_{\mathrm{XQB}}(\ell,\gamma,\rho)+c\left(\xi-\gamma+\frac{\xi-\gamma^2}{(1-\gamma)\log(\ell)}\right),\gamma^2\sigma^2_{\mathrm{XQB}}(\ell,\gamma)\right).$$

(iii) If $\gamma < 1/2$ then, if either N is bounded and (I_2) holds, or if $N \to \infty$ and (I_2') holds, one has

$$\sqrt{k}\left(\widehat{\gamma}_n(k)-\gamma,\frac{\widehat{\mathrm{ES}}_n(1-k/n)}{\mathrm{ES}(1-k/n)}-1\right)\xrightarrow{\mathrm{d}}\mathcal{N}\left(\left(\lambda b_{\mathrm{H}}(\rho)+c(\xi-\gamma),(1-\gamma)c\frac{\xi}{1-\xi}\right),\gamma^2\boldsymbol{V}_{\mathrm{E}}\right).$$

If moreover $\rho < 0$, then condition (I_2') can be weakened to (J_2') if $N \to \infty$, and in any case,

$$\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widehat{\mathrm{ES}}_n^{\star}(\tau_n')}{\mathrm{ES}(\tau_n')}-1\right)\xrightarrow{\mathrm{d}}\mathcal{N}\left(\lambda b_{\mathrm{XE}}(\ell,\gamma,\rho)+c\left(\xi-\gamma+\frac{(1-\gamma)\xi}{(1-\xi)\log(\ell)}\right),\gamma^2\sigma^2_{\mathrm{XE}}(\ell,\gamma)\right).$$

Given Theorem 3.1, and for bounded N, the asymptotic properties of the Hill, Weissman and tail Expected Shortfall estimators should mirror those obtained with an uncontaminated sample from F. This is true under very weak assumptions on the contaminating distribution. In particular, we do not assume that the function $\overline{G}$ is regularly varying, and for $a = 0$, corresponding for example to the asymptotic normality of the Hill estimator, we may even allow a super heavy-tailed contaminating distribution in the sense of Fraga Alves et al. [2009]: for instance, any log-Pareto distribution whose survival function is defined as $\overline{G}(x) = (\log x)^{-2-\eta}$ ($x > e$, $\eta > 0$) satisfies assumption (J_0). Crucially, when $\overline{G}(x)/\overline{F}(x)$ is unbounded and there is at least one contaminating observation, the function

$$x \mapsto \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{P}(X_i^{(n)} > x)}{\overline{F}(x)} = \frac{n-N}{n} + \frac{N}{n}\frac{\overline{G}(x)}{\overline{F}(x)}$$

is not bounded whatever the value of n is, meaning that the Stability Assumption 2.2 of Einmahl and He [2023b] can never hold in model (M1) as long as the tail of the contaminating distribution is heavier than the tail of the distribution function F.

In the case of growing N, the intuition behind assumption $N = \mathrm{o}(n^{\gamma/\xi-\delta})$ for some $\delta > 0$ in condition (I_a'), or $N = \mathrm{O}(n^{\gamma/\xi})$ in condition (J_a'), is that for any intermediate quantile $U(n/k)$ from the distribution function F, the average number of draws $N\overline{G}(U(n/k))$ from $\overline{G}$ above $U(n/k)$, where U denotes the tail quantile function related to the majority distribution function F, should be negligible with respect to the effective intermediate sample size k in the extreme value estimators, so that the number of these draws from the contaminating distribution is too small to contribute asymptotically. Condition $N\overline{G}(U(n/k))/\sqrt{k} \to c \in [0,\infty)$ then strengthens this assumption and controls the finite-sample bias introduced in the estimation by contaminating the sample using N draws from $\overline{G}$.

Let us highlight that the situation $N/n \to p \in (0,1)$, corresponding to the case when the data points from the distribution function G asymptotically represent $100p\%$ of the sample, is essentially different and cannot be handled using the techniques used to prove Theorem 3.1. The next section examines this different setup in the context of pooling.

3.2 Pooling from several samples

We consider here a natural variant of the model in the previous section, which consists in assuming that the data points may be generated from several distributions, one of which dominates the others and is represented in the sample with positive probability. This is in contrast with the setting of Section 3.1, where the distribution having the heaviest tail is considered "rare" in the sample. More precisely, we consider the following model.

(M2) The available data points $X_i^{(n)}$, $1 \le i \le n$, are independent for each n and there is $D \ge 2$ such that

$$\mathbb{P}(X_i^{(n)} > x) = \begin{cases} \overline{G}(x) \text{ if } 1 \le i \le n_0, \\ \overline{G}_j(x) \text{ if } i \in \{\sum_{l=0}^{j-1} n_l + 1, \ldots, \sum_{l=0}^{j} n_l\} \text{ for some } j \in \{1, \ldots, D-1\}, \end{cases}$$

where $n_0 = n_0(n) = N(n)$ and $n_j = n_j(n)$, $1 \le j \le D-1$, are positive sequences of integers such that $N/n \to p \in (0,1)$ and $N + \sum_{j=1}^{D-1} n_j = n$, the functions $\overline{G}$ and $\overline{G}_1$ are survival functions of heavy-tailed distributions with extreme value indices $\gamma > \xi_1$, respectively, and (if $D \ge 3$) the functions $\overline{G}_2, \ldots, \overline{G}_{D-1}$ are such that $\overline{G}_j/\overline{G}_1$ converges to 0 at infinity for each $j \in \{2, \ldots, D-1\}$.

We do not assume that the fractions n_j/n, $1 \le j \le D-1$, of the data coming from the dominated distributions converge as $n \to \infty$, and, when $D \ge 3$, we do not even assume that the data coming from the dominated distributions $G_2, \ldots, G_{D-1}$ is heavy-tailed. This is a substantially more difficult situation than if the data were drawn from a mixture of heavy-tailed distribution functions G, $G_1, \ldots, G_{D-1}$ with fixed weights $p, q_1, \ldots, q_{D-1} \in (0,1)$, which could be tackled by standard extreme value theory.

We have the following result about the asymptotic normality of the Hill, Weissman, and extrapolated Expected Shortfall estimators in this pooled data framework.

Theorem 3.2 (Data pooled from several samples – Hill, Weissman and tail Expected Shortfall estimators)**.** *Assume that the data is generated from model (M2), with the survival function $\overline{G}$ satisfying condition $\mathcal{C}_2(\gamma, \rho, A)$. Assume further that $N = \lfloor np \rfloor$ and $n_1/n \to q_1 \in (0,1)$, and that $k = k(n)$ tends to infinity with $k/n \to 0$ and $n_1 \overline{G}_1(V(np/k))/\sqrt{k} \to c \in [0, \infty)$, where V is the tail quantile function related to G, and that $\sqrt{k} A(n/k) \to \lambda$. Finally, let $\tau_n' \to 1$ be such that $d_n := k/(n(1-\tau_n')) \to \ell \in (1, \infty]$ and $\sqrt{k}/\log(d_n) \to \infty$, and let $U(t) = V(pt)$ for t large enough.*

(i) Then

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma, \frac{X_{n-k:n}^{(n)}}{U(n/k)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left((\lambda b_{\mathrm{H}}(\rho) + c(\xi_1 - \gamma), c\gamma), \gamma^2 \mathrm{I}_2\right).$$

If moreover $\rho < 0$, then one has

$$\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widehat{q}_n^\star(\tau_n')}{q(\tau_n')} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\lambda b_{\mathrm{W}}(\ell, \rho) + c\left(\xi_1 - \gamma + \frac{\gamma}{\log(\ell)}\right), \gamma^2 \sigma_{\mathrm{W}}^2(\ell)\right).$$

(ii) If $\gamma < 1$ then

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma, \frac{\widetilde{\mathrm{ES}}_n(1-k/n)}{\mathrm{ES}(1-k/n)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\lambda b_{\mathrm{H}}(\rho) + c(\xi_1 - \gamma), \lambda b_{\mathrm{QB}}(\gamma, \rho) + c\frac{\xi_1 - \gamma^2}{1-\gamma}\right), \gamma^2 \boldsymbol{V}_{\mathrm{QB}}\right).$$

If moreover $\rho < 0$, then one has

$$\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widetilde{\mathrm{ES}}_n^\star(\tau_n')}{\mathrm{ES}(\tau_n')} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\lambda b_{\mathrm{XQB}}(\ell, \gamma, \rho) + c\left(\xi_1 - \gamma + \frac{\xi_1 - \gamma^2}{(1-\gamma)\log(\ell)}\right), \gamma^2 \sigma_{\mathrm{XQB}}^2(\ell, \gamma)\right).$$

(iii) If $\gamma < 1/2$ then

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma, \frac{\widehat{\mathrm{ES}}_n(1-k/n)}{\mathrm{ES}(1-k/n)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\lambda b_{\mathrm{H}}(\rho) + c(\xi_1 - \gamma), (1-\gamma)c\frac{\xi_1}{1-\xi_1}\right), \gamma^2 \boldsymbol{V}_{\mathrm{E}}\right).$$

If moreover $\rho < 0$, then one has

$$\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widehat{\mathrm{ES}}_n^\star(\tau_n')}{\mathrm{ES}(\tau_n')} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\lambda b_{\mathrm{XE}}(\ell, \gamma, \rho) + c\left(\xi_1 - \gamma + \frac{(1-\gamma)\xi_1}{(1-\xi_1)\log(\ell)}\right), \gamma^2 \sigma_{\mathrm{XE}}^2(\ell, \gamma)\right).$$

Theorem 3.2(i) is, to the best of our knowledge, entirely new. It is a stronger version of Theorem 3 in Quintos et al. [2001], which is limited to $D = 2$ and only provides consistency of the Hill estimator. As the intuition suggests, the Hill estimator is, in model (M2), biased downwards, since the observations from $G_1, \ldots, G_{D-1}$ have a lighter tail than those from G. This stands in contrast to what is observed in Theorem 3.1, where the Hill estimator tends to be biased upwards when the contaminating observations have a heavier tail than those from the majority distribution. Note moreover that condition $n_1\overline{G}_1(U(n/k))/\sqrt{k} = n_1\overline{G}_1(V(np/k))/\sqrt{k} \to c$ is exactly the analogue of condition $N\overline{G}(U(n/k))/\sqrt{k} \to c$ in model (M1), since it counts the number of "contaminating" observations from the distribution function G_1 above the relevant quantile $U(n/k)$ of the distribution of X. This condition, which has the merit of unifying the bias terms in Theorems 3.1 and 3.2, is equivalent to the perhaps simpler version $n\overline{G}_1(V(n/k))/\sqrt{k} \to c' = p^{\gamma/\xi_1}c/q_1$.

3.3 Heterogeneous scales model

We consider here the leading heterogeneous scales example of Einmahl and He [2023b]:

(M3) The available data points $X_i^{(n)}$ are such that $X_i^{(n)} = x_0 + Q_\sigma(1 - \pi(i)/n)Z_i$ where the Z_i are i.i.d. random variables with common survival function S, Q_σ is a quantile function such that $Q_\sigma(0) > 0$ and $\pi = \pi_n$ is an unknown permutation of $\{1, \ldots, n\}$.

Einmahl and He [2023b] identify two main regimes: Z (resp. σ) may be heavy-tailed with σ (resp. Z) having a lighter tail. We therefore assume that there exists $\gamma > 0$ such that one of the following two conditions is satisfied:

(LS1a) The survival function S satisfies $\lim_{t\to\infty} t^{1/\gamma}S(t) = c \in (0, \infty)$ and $\int_0^1 Q_\sigma^{1/\gamma}(\alpha)\,\mathrm{d}\alpha < \infty$.

(LS2a) The tail quantile function $U_\sigma : t \mapsto Q_\sigma(1 - 1/t)$ satisfies $\lim_{t\to\infty} t^{-\gamma}U_\sigma(t) = c \in (0, \infty)$ and $0 < \mathbb{E}(\max(Z, 0)^{1/\gamma}) = \int_0^\infty S(v^\gamma)\,\mathrm{d}v < \infty$.

Under either of these two conditions, part of the proof of Theorem 3.3 below (see Proposition C.4) consists in showing that condition $\mathcal{H}_1(\mathbb{P}_X, \gamma)$ holds with

$$\mathbb{P}(X > x) = \int_0^1 S\left(\frac{x - x_0}{Q_\sigma(1-u)}\right)\mathrm{d}u.$$

Proving that this random variable X satisfies the second-order condition $\mathcal{C}_2(\gamma, \rho, A)$ requires further assumptions on σ and Z, that we introduce below.

(LS1b) Assumption (LS1a) holds, and there are $d \neq 0$ and $\rho < 0$ such that $\lim_{t\to\infty} t^{-\rho/\gamma}(t^{1/\gamma}S(t) - c) = d$ and $\int_0^1 Q_\sigma^{(1-\rho)/\gamma}(\alpha)\,\mathrm{d}\alpha < \infty$.

(LS2b) Assumption (LS2a) holds, and there are $d \neq 0$ and $\rho < 0$ such that $\lim_{t\to\infty} t^{-\rho}(t^{-\gamma}U_\sigma(t) - c) = d$; moreover, S is continuously differentiable with derivative $-s$, where $s \geq 0$ is ultimately nonincreasing and satisfies $\int_0^\infty v^{\gamma-\rho}s(v^\gamma)\,\mathrm{d}v < \infty$.

Equipped with these assumptions, we may state the following result about the asymptotic normality of the Hill, Weissman, and extrapolated Expected Shortfall estimators in the heterogeneous scales model.

Theorem 3.3 (Heterogeneous scales model – Hill, Weissman and tail Expected Shortfall estimators)**.** *Assume that the data is generated from model (M3) with $x_0 = 0$ and work under either assumption (LS1b) or (LS2b). Let $k = k(n) \to \infty$ be such that $k/n \to 0$ and $\sqrt{k}(n/k)^\rho \to \lambda$. Suppose finally that, if assumption (LS1b) holds,*

$$\lim_{n\to\infty} \sqrt{k}\int_{1-1/n}^1 Q_\sigma^{1/\gamma}(\alpha)\,\mathrm{d}\alpha = 0.$$

Define, as in Proposition C.4 (see the Supplementary Material document),

$$R(x,y) = 0 \quad \textit{and} \quad K = \frac{d\int_0^1 Q_\sigma^{(1-\rho)/\gamma}(\alpha)\,\mathrm{d}\alpha}{\left(c\int_0^1 Q_\sigma^{1/\gamma}(\alpha)\,\mathrm{d}\alpha\right)^{1-\rho}} \quad \textit{if (LS1b) holds,}$$

$$R(x,y) = \frac{\int_0^\infty S(x^{-\gamma}v^\gamma)S(y^{-\gamma}v^\gamma)\,\mathrm{d}v}{\int_0^\infty S(v^\gamma)\,\mathrm{d}v} \quad \textit{and} \quad K = \frac{d(1-\rho)}{c\gamma}\frac{\int_0^\infty v^{-\rho}S(v^\gamma)\,\mathrm{d}v}{\left(\int_0^\infty S(v^\gamma)\,\mathrm{d}v\right)^{1-\rho}} \quad \textit{if (LS2b) holds.}$$

Finally, let $\tau_n' \to 1$ be such that $d_n := k/(n(1-\tau_n')) \to \ell \in (1,\infty]$ and $\sqrt{k}/\log(d_n) \to \infty$.

(i) Then

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma, \frac{X_{n-k:n}^{(n)}}{U(n/k)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\gamma\rho K\lambda b_{\mathrm{H}}(\rho), 0\right), \gamma^2(\mathrm{I}_2 - \boldsymbol{\mathcal{M}}_0)\right)$$

$$\textit{and} \quad \frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widehat{q}_n^\star(\tau_n')}{q(\tau_n')} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\gamma\rho K\lambda b_{\mathrm{W}}(\ell,\rho), \gamma^2\sigma_{\mathrm{W}}^2(\ell)(1-\pi_{\mathrm{W}}(\ell,R))\right).$$

(ii) If $\gamma < 1$ then

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma, \frac{\widetilde{\mathrm{ES}}_n(1-k/n)}{\mathrm{ES}(1-k/n)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\gamma\rho K\lambda b_{\mathrm{H}}(\rho), \lambda b_{\mathrm{QB}}(\gamma,\rho)\right), \gamma^2(\boldsymbol{V}_{\mathrm{QB}} - \boldsymbol{M}_{\mathrm{QB}})\right)$$

$$\textit{and} \quad \frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widetilde{\mathrm{ES}}_n^\star(\tau_n')}{\mathrm{ES}(\tau_n')} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\gamma\rho K\lambda b_{\mathrm{XQB}}(\ell,\gamma,\rho), \gamma^2\sigma_{\mathrm{XQB}}^2(\ell,\gamma)(1-\pi_{\mathrm{XQB}}(\ell,\gamma,R))\right).$$

(iii) If $\gamma < 1/2$ then

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma, \frac{\widehat{\mathrm{ES}}_n(1-k/n)}{\mathrm{ES}(1-k/n)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\gamma\rho K\lambda b_{\mathrm{H}}(\rho), 0\right), \gamma^2(\boldsymbol{V}_{\mathrm{E}} - \boldsymbol{M}_{\mathrm{E}})\right)$$

$$\textit{and} \quad \frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widehat{\mathrm{ES}}_n^\star(\tau_n')}{\mathrm{ES}(\tau_n')} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\gamma\rho K\lambda b_{\mathrm{XE}}(\ell,\gamma,\rho), \gamma^2\sigma_{\mathrm{XE}}^2(\ell,\gamma)\left(1-\pi_{\mathrm{XE}}(\ell,\gamma,R)\right)\right).$$

Theorem 3.3(i) provides a sharper asymptotic theory than that of Einmahl and He [2023b], by establishing joint asymptotic normality for $X_{n-k:n}^{(n)}$ and $\widehat{\gamma}_n(k)$, with explicit bias and variance-covariance terms, as well as asymptotic normality for $\widehat{q}_n^\star(\tau_n'|k)$, together with closed-form expressions for the corresponding asymptotic bias and variance. These stronger results are achieved under milder conditions. Besides the Stability Assumption 2.2 of Einmahl and He [2023b], we dispense with the Lipschitz requirement on $t \mapsto \log U_\sigma(e^t)$ over $[0,\infty)$ and with the envelope-type conditions on the density g of Z used in their Theorems 3.1–3.2. Moreover, while our analysis relies on the standard second-order condition $\mathcal{C}_2(\gamma,\rho,A)$, for which we also provide sufficient conditions, the analogous requirements in Einmahl and He [2023b], namely Conditions (4) and (8) needed for their Theorems 2.2 and 4.1, were not checked in the heterogeneous scales framework. Theorem 3.3(ii) and (iii) provides asymptotic theory for extreme Expected Shortfall estimation, which was not discussed in Einmahl and He [2023b].

3.4 Ignoring observed covariate information

We finally consider the following model, related to Example 1 in Section 2 of Einmahl and He [2023b].

(M4) The available data points $X_i^{(n)}$, $1 \le i \le n$, satisfy $X_i^{(n)} = (\phi(Z_i))^{\gamma(i/n)}$, where ϕ is a positive, continuous and increasing function on $[1,\infty)$ that is regularly varying with index 1, the Z_i are i.i.d. unit Pareto, and $s \mapsto \gamma(s)$ is positive, decreasing and continuously differentiable on $[0,1]$ with $\gamma'(0) < 0$.

This model has a natural interpretation in a regression setting: if (Y_i, S_i), $1 \le i \le n$, are independent data points with $\mathbb{P}(Y_i > y \,|\, S_i = s) = 1/\phi^{-1}(y^{1/\gamma(s)})$, then $(X_1^{(n)}, \ldots, X_n^{(n)})$ have the same distribution as $(Y_1, \ldots, Y_n)$ given that the $S_i = i/n$ have been observed. Model (M4) can therefore be viewed as a model for data collected while ignoring observed covariate information, represented by the S_i. If $\phi(z) = z\log(z)$ and $\gamma(s) = 1/(1+s)$, this is just Example 1 in Einmahl and He [2023b].

In this model, the $X_i^{(n)}$ come from different distributions having different extreme value indices. Notice that we have for any x sufficiently large

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{P}(X_i^{(n)} > x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\phi^{-1}(x^{1/\gamma(i/n)})} \to \int_0^1 \frac{\mathrm{d}s}{\phi^{-1}(x^{1/\gamma(s)})} \text{ as } n \to \infty.$$

Proposition C.5 (see the Supplementary Material document) formally proves that $\mathcal{H}_1(\mathbb{P}_X, \gamma)$ is satisfied with $\mathbb{P}(X > x) = \int_0^1 \mathrm{d}s/\phi^{-1}(x^{1/\gamma(s)})$ for x large enough. This is the cornerstone for our next result.

Theorem 3.4 (Ignoring covariate information – Hill, intermediate quantile and Expected Shortfall estimators)**.** *Assume that the data is generated from model (M4) and work under the regularity conditions of Proposition C.5. Let $k = k(n)$ tend to infinity with $k/n \to 0$ and $\sqrt{k}/\log(n) \to \lambda$.*

(i) Then

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma(0), \frac{X_{n-k:n}^{(n)}}{U(n/k)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left((-\lambda\gamma(0), 0), \gamma^2(0)\mathrm{I}_2\right).$$

(ii) If $\gamma(0) < 1$, then

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma(0), \frac{\widetilde{\mathrm{ES}}_n(1-k/n)}{\mathrm{ES}(1-k/n)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left((-\lambda\gamma(0), 0), \gamma^2(0)\boldsymbol{V}_{\mathrm{QB}}(\gamma(0))\right).$$

(iii) If $\gamma(0) < 1/2$, then

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma(0), \frac{\widehat{\mathrm{ES}}_n(1-k/n)}{\mathrm{ES}(1-k/n)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left((-\lambda\gamma(0), 0), \gamma^2(0)\boldsymbol{V}_{\mathrm{E}}(\gamma(0))\right).$$

Note that a result about the estimation of extreme quantiles and the Expected Shortfall at extreme levels would not be a simple corollary of our high-level results, because the distribution of X in Theorem 3.4 has second-order parameter $\rho = 0$, which is not handled by Corollaries 2.3 and 2.6. In fact, thanks to the asymptotic expansion of $\mathbb{P}(X > x)$ stated in Proposition C.5, one can show by straightforward but tedious calculations that the tail quantile function U of X satisfies

$$U(t) = \left(-C\frac{\gamma(0)}{\gamma'(0)}\right)^{\gamma(0)}\left(\frac{t}{\log(t)}\right)^{\gamma(0)}\left(1 + \gamma(0)\frac{\log\log(t)}{\log(t)}(1 + \mathrm{o}(1))\right) \text{ as } t \to \infty.$$

This means that the bias terms in the Weissman-type approximations appearing in the proofs of Corollaries 2.3 and 2.6 can theoretically also be controlled in the current setting; however, the rate of convergence in the estimation of extreme quantiles and extreme Expected Shortfall within model (M4) would be constrained by the rate of convergence of the Hill estimator which, as can be seen in Theorem 3.4, is at most $1/\log(n)$. Statistically speaking, this discussion shows that ignoring observed covariate information tends to be very detrimental in extreme value settings, because one will identify only the value of the extreme value index at the point in the covariate space where it is maximal; furthermore, this will be done only at a very slow rate.

4 Simulation study

In this section, to illustrate how heterogeneity affects the asymptotic behavior of the Hill estimator, the extreme quantile estimator and the extreme Expected Shortfall estimators, we conduct Monte-Carlo experiments in the three settings investigated in Section 3 of contaminated samples, pooling from several populations, and the heterogeneous scales model. All the experiments have sample size $n = 1{,}000$ and employ 10,000 Monte-Carlo replications. To save space, we report here only the main conclusions from this simulation study. A complete description of the models employed with their parameters is given in Section D.1. A full set of conclusions about these experiments is given in Section D.2 (extreme value index estimation), Section D.3 (extreme quantile estimation) and Section D.4 (tail Expected Shortfall estimation). All supporting figures are deferred to Section D.5.

Extreme value index estimation and inference In the contamination model, where $n - N$ of the observations come from a Pareto, Fréchet or Student t-distribution, with extreme value index γ, the n data points include $N \in \{15, 31, 47\}$ contaminating observations from another Pareto, Fréchet or Student t-distribution with extreme value index $\xi = 2\gamma$. It can be seen from Figures D.1 to D.3 that the asymptotic bias of the Hill estimator increases with the degree of contamination in the data. The MSE estimates for the three contamination scenarios exhibit the same qualitative behavior with respect to k as in the i.i.d. case $N = 0$, and their values decrease (respectively, increase) uniformly in k as either N or γ decreases (respectively, increases). It can also be seen that the bias estimates are positive in the three cases of contamination, as expected in, for instance, the Pareto case from Theorem 3.1(i). The asymptotic normality stated in Theorem 3.1(i) is corroborated by Gaussian QQ-plots in the Pareto and Fréchet scenarios (where the value $k := k_{\mathrm{H}}$ is taken to minimize the estimated MSE). For the Student t-scenario, the alignment of the Gaussian QQ-plot is less obvious, but improves as γ increases, reflecting both the faster convergence of the Hill estimator in this case and the presence of a substantial asymptotic bias characteristic of the Student t-distribution, whose second-order parameter is $\rho = -2\gamma$.

In the pooling model, the bias varies with the proportion p of data points drawn from the distribution $\overline{F}$ with heaviest tail. We consider three pooling scenarios, with $p = 0.25, 0.5, 0.75$, resulting in $N = 250, 500, 750$, respectively. The benchmark i.i.d. case $p = 1$ (*i.e.* no pooling) is included. The corresponding results with Pareto, Fréchet, and Student t-distributions are shown in Figures D.4 to D.6. For the Pareto and Fréchet distributions, the bias and MSE estimates clearly show that the performance of the Hill estimator improves (respectively, deteriorates) uniformly in k as p increases, and likewise as γ decreases (respectively, as p decreases, and likewise as γ increases), across a broad range of thresholds. In contrast, for the real-valued Student t-distribution, this behavior persists only for small thresholds $k \leq 50$ when $\gamma \geq 1/2$; the interpretation is that for lower values of γ, the second-order bias inherent to Student t-observations becomes so pronounced that it is difficult to disentangle the respective contributions of each source of bias. As for the Gaussian QQ-plots, they support the asymptotic normality stated in Theorem 3.2(i), while also revealing the presence of an asymptotic bias. For the Pareto distribution, this bias remains quite important for the three values of $p < 1$, whereas for both the Fréchet and Student t-distributions, it becomes more attenuated when $p \leq 0.5$.

In the heterogeneous scales model, observations are generated from a Pareto, Fréchet or Student t-distribution on Z, respectively, with extreme value index ξ, multiplied by heterogeneous scales given by the tail quantile function $U_\sigma^{(\gamma,-\infty)}(t) = t^\gamma$ (*i.e.* the Pareto distribution with extreme value index γ), the tail quantile function $U_\sigma^{(\gamma,-1)}(t) = t^\gamma(1 - t^{-1}/2)$, for which $\rho = -1$, and $U_\sigma^{(\gamma,-1/2)}(t) = t^\gamma(1 - t^{-1/2}/2)$, in which case $\rho = -1/2$, respectively. We take $\gamma = 3\xi$. This yields scenarios with varying degree of difficulty and heterogeneity. The scenario corresponding to Pareto scales results in better bias and MSE estimates than the other two for all values of γ. In terms of Gaussian QQ-plots, both the Pareto (Figure D.7) and Fréchet (Figure D.8) cases provide evidence that the finite-sample distribution of the Hill estimator matches the asymptotic distribution in Theorem 3.3(i), with only minor biases. In contrast, for the Student t-case (Figure D.9), the fit again improves as γ increases.

Extreme quantile estimation and inference We report the Monte-Carlo estimates of the relative bias and MSE for the Weissman estimator, that is, the bias and MSE of $(\widehat{q}_n^\star(\tau_n')/q(\tau_n')-1)$, computed from the 10,000 simulation replications over the range of all possible k values, for $\tau_n' = 1-1/n = 0.999$. We also represent Gaussian QQ-plots of $\widehat{q}_n^\star(\tau_n')$ on the log-scale, that is,

$$\frac{\sqrt{k}}{\gamma\sqrt{1-R(1,1)}\log(k)}\log\left(\frac{\widehat{q}_n^\star(\tau_n')}{q(\tau_n')}\right)$$

is graphed against the standard normal distribution for the value $k := k_{\mathrm{W}}$ that minimizes the estimated relative MSE. In all three models, the results for the Pareto and Fréchet distributions, shown in Figures D.10, D.11, D.13, D.14, D.16 and D.17, are qualitatively very similar to those previously obtained for the Hill estimator in Figures D.1, D.2, D.4, D.5, D.7 and D.8, respectively. Moreover, it can be seen that the optimal values k_{H} (Hill) and k_{W} (Weissman) are generally close for each fixed distribution and index γ. This is due to the fact that, with the notation of Corollary 2.3, one has $\ell = \infty$, meaning that the asymptotic behavior of the Hill estimator dominates the asymptotic behavior of the Weissman estimator. For the Student t-distribution, this pattern does not systematically hold, as can be seen from the pooling (Figures D.6 and D.15) and heterogeneous scales (Figures D.9 and D.18) models when $\gamma = 1/4$. This is likely due to the additional bias term introduced by the Weissman estimation, which, although theoretically negligible according to Corollary 2.3, exerts a significant finite-sample effect in this setting where $\rho = -2\gamma = -1/2$ is closer to zero.

Tail Expected Shortfall estimation and inference We examine the Monte-Carlo estimates of the relative bias and MSE for $\widehat{\mathrm{ES}}_n^\star(\tau_n')$ and $\widetilde{\mathrm{ES}}_n^\star(\tau_n')$, based on 10,000 simulation replications and across the full range of k values, for $\tau_n' = 1-1/n = 0.999$. We also construct the Gaussian QQ-plots of

$$\frac{\sqrt{k}}{\gamma\sqrt{1-R(1,1)}\log(k)}\log\left(\frac{\widehat{\mathrm{ES}}_n^\star(\tau_n')}{\mathrm{ES}(\tau_n')}\right) \quad\text{and}\quad \frac{\sqrt{k}}{\gamma\sqrt{1-R(1,1)}\log(k)}\log\left(\frac{\widetilde{\mathrm{ES}}_n^\star(\tau_n')}{\mathrm{ES}(\tau_n')}\right),$$

using the k-values $k := k_{\mathrm{XE}}$ and $k := k_{\mathrm{XQB}}$ that minimize the estimated relative MSE for $\widehat{\mathrm{ES}}_n^\star(\tau_n')$ and $\widetilde{\mathrm{ES}}_n^\star(\tau_n')$, respectively. The simulation design is similar, except that we take $\gamma < 1/2$ to comply with our assumptions. Results are displayed in Figures D.19 to D.36, and complemented by Tables D.1, D.2 and D.3 for additional results on bias and MSE at $k = k_{\mathrm{XE}}$ for $\widehat{\mathrm{ES}}_n^\star(\tau_n')$ and $k = k_{\mathrm{XQB}}$ for $\widetilde{\mathrm{ES}}_n^\star(\tau_n')$. Qualitatively, the conclusions are broadly consistent with those of the previous section regarding both bias and MSE. It is worth noting that, in the contamination model, the MSEs of $\widehat{\mathrm{ES}}_n^\star(\tau_n')$ tend, as expected, to be fairly high when $N > 0$ and $2\gamma > 1/2$. This effect is clearly visible in the middle panels of the second and third rows of Figures D.19, D.21 and D.23. The problem of inferring the tail Expected Shortfall appears to be much harder, as evidenced by the Gaussian QQ-plots. While the bias of $\widetilde{\mathrm{ES}}_n^\star(\tau_n')$ is clearly present, its variance seems to be incorrectly estimated overall. The situation is even worse for $\widehat{\mathrm{ES}}_n^\star(\tau_n')$, where the Gaussian approximation generally fails to hold in the contamination and pooling models when γ approaches 1/2. It should, however, be noted that $\widehat{\mathrm{ES}}_n^\star(\tau_n')$ performs better than $\widetilde{\mathrm{ES}}_n^\star(\tau_n')$ in terms of both bias and MSE across all distributions in the heterogeneous scales model and, to a smaller extent, in the pooling framework under the Student t-distribution.

5 Real data applications to risk assessment

This section first examines the cyber risk associated with personal data breaches in Subsection 5.1, and then turns to a second application on dynamic financial risk forecasting in Subsection 5.2. For brevity, we report only our main conclusions; full details, including our rationale for the tuning-parameter choices, are provided in Section E of the Supplementary Material.

Our applications require an estimate of the (rescaled) asymptotic variance $\sigma^2 = \gamma^2(1 - R(1,1))$ of the Hill estimator; see Corollary 2.2. The estimator of $R(1,1)$ proposed by Einmahl and He [2023b] in their Equation (5) cannot be used here, because splitting our datasets into two duplicate halves is not feasible. Instead, we use $\widehat{\sigma}_n^2 \equiv \widehat{\sigma}_n^2(J,k) := \left(\log \frac{k}{J}\right)^{-1} \sum_{i=J}^{k} \left(\widehat{\gamma}_n(i) - \widehat{\gamma}_n(k)\right)^2$ to construct the 95% confidence interval $\widehat{\mathrm{CI}}_{0.95}(\gamma) = \left[\widehat{\gamma}_n(k) \pm 1.96\sqrt{\widehat{\sigma}_n^2/k}\right]$ for γ. In our non-identically distributed setting, we argue heuristically and without formally establishing the consistency of $\widehat{\sigma}_n^2$ that, following the suggestion of Drees [2003] in the stationary time series case, one may choose the smallest value of J for which $\widehat{\sigma}_n^2$ is well defined and, in our heterogeneous context, does not exceed the Hill estimator's rescaled asymptotic variance estimate $\widehat{\gamma}_n^2(k)$ under the i.i.d. assumption.

5.1 Cyber risk: Privacy Rights Clearinghouse Database

Large-scale cyberattacks are a relatively recent, yet now common, manifestation of cybersecurity risk. They can lead to mass identity fraud and potentially disastrous outcomes for affected victims. Existing studies of cyber risk, such as Farkas et al. [2021] and Daouia et al. [2024], try to address the large heterogeneity in the data by stratifying the breaches into more homogeneous subsamples. Among these, only Daouia et al. [2024] conducts extreme value inference, but under the unrealistic assumption that the data are identically distributed.

We use the PRC (Privacy Rights Clearinghouse) database[1] which is the most comprehensive open scientific dataset for publicly reported data breaches across the United States between 2005 and 2019. It contains eight types of breach events and classifies the victim entities into eight categories. We employ the recently updated version of the database (Data Breach Chronology v1.5) and focus on the severity of cyberattacks through the available $n = 9{,}001$ breach sizes $X_1^{(n)}, \ldots, X_n^{(n)}$, with $X_i^{(n)}$ being the estimated number of records impacted by the breach $i = 1, \ldots, n$.[2] Figure 1 displays the counts of reported events. Clearly, the data is heterogeneous, overdispersed and highly right-skewed.

In Figure 2, panel (A) displays the Hill plot $k \mapsto \widehat{\gamma}_n(k)$, along with its stability region $k \in [1130, 1700]$, the 95% confidence interval $\widehat{\mathrm{CI}}_{0.95}(\gamma)$, and the benchmark interval based on the untenable i.i.d. modeling assumption. We find a point estimate $\widehat{\gamma}_n = \widehat{\gamma}_n(1700) = 2.021$, and a pointwise confidence interval $[1.952, 2.091]$ that is 28.10% narrower than the interval $[1.925, 2.118]$ based on the i.i.d. model. See panel (B) for the estimation of σ^2 by $\widehat{\sigma}_n^2 \equiv \widehat{\sigma}_n^2(48, 1700) = 2.113$, equivalent to an estimate $\widehat{R}_n(1,1) = 1 - \widehat{\sigma}_n^2/\widehat{\gamma}_n^2 = 0.483$ of the tail heterogeneity coefficient $R(1,1)$.

We turn to risk assessment and, because $\widehat{\gamma}_n > 1$, we focus on extreme quantile estimation. The tail probability level of interest, corresponding to an infrequent breach size expected to be exceeded once every T years, is $\tau_n' = 1 - 1/(T \times 600)$. This extreme level $\tau_n' \geq 0.998$ is well beyond the intermediate level $1 - k/n \approx 0.811$ whatever $T \geq 1$. This motivates the choice of $\ell = \infty$ in the asymptotic variance of the Weissman estimator $\widehat{q}_n^\star(\tau_n'|k)$ of $q(\tau_n')$, thus leading to the 95% confidence interval

$$\widehat{\mathrm{CI}}_{0.95}\big(q(\tau_n')\big) = \left[\widehat{q}_n^\star(\tau_n'|k) \exp\left(\pm 1.96 \log(d_n) \sqrt{\widehat{\sigma}_n^2/k}\right)\right],$$

with $d_n = k/(n(1 - \tau_n'))$. When $\tau_n' = 1 - 1/n$, corresponding to the full observation period of $T = 15$ years, panel (C) of Figure 2 indicates a point estimate $\widehat{q}_n^\star(\tau_n') = \widehat{q}_n^\star(\tau_n'|1700) = 34$ billion records affected, which vastly exceeds the historical maximum number of records affected (3 billion). The final 95% confidence interval $\widehat{\mathrm{CI}}_{0.95}\big(q(\tau_n')\big) = [20.34, 56.86]$ is 30.93% narrower than its counterpart $[16.63, 69.51]$ (both in billion records) based on the i.i.d. model. When varying T, as seen from panel (D), $T = 5$ years is the earliest return period at which the Weissman estimate surpasses this maximum breach size, yielding $\widehat{q}_n^*(\tau_n') \approx \widehat{q}_n^*(0.9996) = 3.68$ billion records. The lower confidence bound for $q(\tau_n')$ exceeds $X_{n:n}^{(n)}$ at $T = 6$, while the upper confidence bound surpasses $X_{n:n}^{(n)}$ already at $T = 4$.

[1] Available at https://privacyrights.org/data-breaches. See Eling et al. [2023] for a comparison with the other main cyber loss datasets considered in the literature, namely Advisen and SAS OpRisk.

[2] To prevent ties in the breach sizes from interfering with our inferential method, which assumes continuous data, we use a jittered version of the records by adding uniformly distributed perturbations on the interval $(0, 0.05)$.

In panels (E) to (H), we specifically consider the 4,337 breaches involving healthcare, medical providers, and medical insurance services (MED), which are the most frequent cyber events. They exhibit significantly lower, though still persistent, heterogeneity and tail heaviness: their Hill estimate drops to 1.433 with a corresponding 95% confidence interval of $[1.340, 1.525]$, which is 14.21% tighter than the i.i.d.-based interval. The percent change in the Hill asymptotic variance reduction factor is -45.34%, reflecting a marked decrease in data heterogeneity. In addition, the T-year cyber risk for the medical sector is at least 46 times lower than the global risk evaluated across all breach events. It first exceeds the sample maximum at $T = 10$ years, while its 95% lower confidence bound does so at $T = 15$ years. It is therefore relevant to differentiate this specific risk from the others and to account for the induced heterogeneity due to different breach types and information sources.

5.2 Financial risk: Filtering to handle serial dependence

Time series of daily financial returns often feature substantial serial dependence. To handle this dependence in a dynamic setting, it is usual practice to apply a time series filter, such as an ARMA-GARCH model, and then to treat the residuals from this model as i.i.d. observations to which extreme value inference may be applied. This was pioneered by McNeil and Frey [2000]. While the adequacy of the fitted model and the independence of residuals can generally be confirmed through goodness-of-fit checks, residuals typically do not have the same distribution, even in settings as simple as linear regression. Extreme value theory for heterogeneous data makes this two-stage procedure fully justified.

We consider daily loss returns (*i.e.* negative log-returns) for the five stocks of American Express Company (AXP), Bank of America Corporation (BAC), Berkshire Hathaway (BRKA), Wells Fargo & Company (WFC), and Walmart (WMT), from 15 July 2015 to 23 June 2025. For each observed time series, we consider successive rolling windows of length $n = 1{,}500$, which results in 1,000 samples of n daily loss returns data. Figures E.37 and E.38 provide evidence of stationarity across all rolling windows (by the KPSS test), of independence of the ARMA(1,1)-GARCH(1,1) residuals within each sample (by the Ljung-Box test) on both the residuals and their squares, and of goodness-of-fit of the ARMA-GARCH model (by the Lagrange Multiplier test and the sign bias test of Engle and Ng [1993]).

To evaluate the tail heaviness, heterogeneity of residuals, and risk exposure within each sample, we focus in the sequel on the AXP time series. Similar considerations evidently apply to the other stock returns. To simplify the presentation, we report results only for the successive 159 rolling windows, spanning 20 March 2018 to 18 October 2024. As shown in the top panels of Figure 3, exploratory analyses indicate that the plots of all extreme value estimators (Hill estimator, Drees-type estimator of the asymptotic variance σ^2, tail quantile estimator and tail Expected Shortfall estimators) display a common initial stable region around $k = 99$. Our analysis uses the tail probability level $\tau_n' = 1 - 1/n \approx 0.999$, which is well beyond the hence chosen intermediate level $1 - k/n \approx 0.934$. The red lines in the four top panels of Figure 3 show the resulting estimates based on the residuals of the last rolling window (1 November 2018 - 18 October 2024). The middle panels illustrate the application of our inferential procedure to this 159th estimation window: in particular, for the selected values $k = 99$ and $J = 1$, we found here $\widehat{\gamma}_n(k) = 0.302$, $\widehat{\sigma}_n^2 \equiv \widehat{\sigma}_n^2(J,k) = 0.040$ and $\widehat{R}_n(1,1) = 1 - \widehat{\sigma}_n^2/\widehat{\gamma}_n^2(k) = 0.561$.

The final dynamic results across successive rolling windows are presented in the bottom panels of Figure 3, along with the percentage reductions by which the accurate confidence intervals, justified by our asymptotic theory, are narrower than their benchmark i.i.d.-based counterparts. These reductions, related to the estimators $\widehat{\gamma}_n(k)$, $\widehat{q}_n^\star(\tau_n'|k)$, $\widehat{\mathrm{ES}}_n^\star(\tau_n'|k)$ and $\widetilde{\mathrm{ES}}_n^\star(\tau_n'|k)$ are nearly identical, ranging from 20.15% to 44.23% across successive windows. This highlights substantial heterogeneity in the ARMA-GARCH residuals over each window, as also evidenced by the strong variation in the $R(1,1)$ estimates, which fluctuate between 0.36 and 0.68. By contrast, the tail heaviness of residuals remains quite stable, with γ estimates barely evolving between 0.285 and 0.308. The targeted point forecasts and associated confidence bounds for both tail risk measures appear to track the data trends fairly closely.

Acknowledgments

This research was supported by the French National Research Agency under the grants ANR-23-CE40-0009 (EXSTA project), ANR-19-CE40-0013 (ExtremReg project), ANR-18-EURE-0023 (EUR MINT), 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. G. Stupfler acknowledges further support from the Chair Stress Test, RISK Management and Financial Steering of the Foundation Ecole Polytechnique.

References

Basel Committee on Banking Supervision. Fundamental review of the trading book: A revised market risk framework, 2013.

P. Billingsley. *Probability and Measure (Third Edition)*. John Wiley and Sons, 1995.

A. Daouia, G. Stupfler, and A. Usseglio-Carleve. Bias-reduced and variance-corrected asymptotic gaussian inference about extreme expectiles. *Statistics and Computing*, 34(4):130, 2024.

A. Daouia, G. Stupfler, and A. Usseglio-Carleve. Corrected inference about the extreme Expected Shortfall in the general Max-Domain of Attraction. *Information and Inference*, 14(3):iaaf023, 2025.

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

L. de Haan and C. Zhou. Trends in extreme value indices. *Journal of the American Statistical Association*, 116(535):31–51, 2021.

H. Drees. Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli*, 9 (4):617–657, 2003.

B. Edwards, S. Hofmeyr, and S. Forrest. Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity*, 2(1):3–14, 2016.

J. H. J. Einmahl and Y. He. Extreme value estimation for heterogeneous data. *Journal of Business & Economic Statistics*, 41(1):255–269, 2023a.

J. H. J. Einmahl and Y. He. Extreme value inference for heterogeneous power law data. *The Annals of Statistics*, 51(3):1331–1356, 2023b.

J. H. J. Einmahl, L. de Haan, and C. Zhou. Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B*, 78(1):31–51, 2016.

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

M. Eling and N. Loperfido. Data breaches: Goodness of fit, pricing, and risk measurement. *Insurance: Mathematics and Economics*, 75:126–136, 2017.

M. Eling, R. Ibragimov, and D. Ning. Time dynamics of cyber risk. *University of St. Gallen*, 2023.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

R. F. Engle and V. K. Ng. Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5):1749–1778, 1993.

S. Farkas, O. Lopez, and M. Thomas. Cyber claim analysis using Generalized Pareto regression trees with applications to insurance. *Insurance: Mathematics and Economics*, 98:92–105, 2021.

T. Fissler and J. F. Ziegel. Higher order elicitability and Osband's principle. *Annals of Statistics*, 44 (4):1680–1707, 2016.

I. Fraga Alves, L. de Haan, and C. Neves. A test procedure for detecting super-heavy tails. *Journal of Statistical Planning and Inference*, 139(2):213–227, 2009.

L. Giraitis, Y. Li, and P. C. B. Phillips. Robust inference on correlation under general heterogeneity. *Journal of Econometrics*, 240(1):105691, 2024.

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

Yi He, Liang Peng, Dabao Zhang, and Zifeng Zhao. Risk analysis via Generalized Pareto distributions. *Journal of Business & Economic Statistics*, 40(2):852–867, 2022.

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

Y. Hoga. Limit theory for forecasts of extreme distortion risk measures and expectiles. *Journal of Financial Econometrics*, 20(1):18–44, 2022.

T. Hsing. On tail index estimation using dependent data. *The Annals of Statistics*, 19(3):1547–1569, 1991.

International Monetary Fund. Canada: financial sector assessment program – insurance core principles – detailed assessment of observance. Washington, D.C., 2014.

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

R. Koenker. *Quantile Regression*. Cambridge University Press, 2005.

S. Li, L. Peng, and X. Song. Simultaneous Confidence Bands For Conditional Value-At-Risk And Expected Shortfall. *Econometric Theory*, 39(5):1009–1043, 2023.

Y. Lu, J. Zhang, and W. Zhu. Cyber risk modeling: a discrete multivariate count process approach. *Scandinavian Actuarial Journal*, 2024(6):625–655, 2024.

T. Mao, G. Stupfler, and F. Yang. Asymptotic properties of generalized shortfall risk measures for heavy-tailed risks. *Insurance: Mathematics and Economics*, 111:173–192, 2023.

C. Martins-Filho, F. Yao, and M. Torero. Nonparametric Estimation of Conditional Value-at-Risk and Expected Shortfall Based on Extreme Value Theory. *Econometric Theory*, 34(1):23–67, 2018.

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

Organisation for Economic Co-operation and Development. Life annuity products and their guarantees. OECD Publishing, Paris, 2016.

J. W. Pratt. On interchanging limits and integrals. *Annals of Mathematical Statistics*, 31(1):74–77, 1960.

C. Quintos, Z. Fan, and P. C. B. Phillips. Structural change tests in tail behavior and the Asian crisis. *Review of Economic Studies*, 68(3):633–663, 2001.

H. Rootzén, M. R. Leadbetter, and L. de Haan. On the distribution of tail array sums for strongly mixing stationary sequences. *The Annals of Applied Probability*, 8(3):868–885, 1998.

R. Schmidt and U. Stadtmüller. Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics*, 33(2):307–335, 2006.

J. Segers. Residual estimators. *Journal of Statistical Planning and Inference*, 98(1–2):15–27, 2001.

S. M. Stigler. Linear functions of order statistics with smooth weight functions. *Annals of Statistics*, 2(4):676–693, 1974.

S. M. Stigler. The effect of sample heterogeneity on linear functions of order statistics, with applications to robust estimation. *Journal of the American Statistical Association*, 71(356):956–960, 1976.

G. Stupfler. On a relationship between randomly and non-randomly thresholded empirical average excesses for heavy tails. *Extremes*, 22(4):749–769, 2019.

J. Velthoen, J.-J. Cai, G. Jongbloed, and M. Schmeits. Improving precipitation forecasts using extreme quantile regression. *Extremes*, 22(4):599–622, 2019.

Y. Zhao, T. Mao, and F. Yang. Estimation of the Haezendonck-Goovaerts risk measure for extreme risks. *Scandinavian Actuarial Journal*, 2021(7):599–622, 2021.

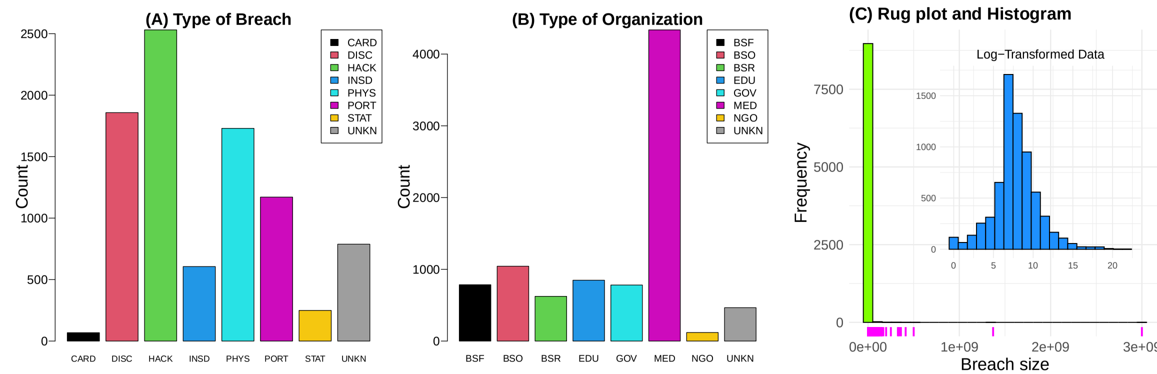

Figure 1: (A) Count of reported events following the "Type of Breach" labels that describe the nature of the data breach, and (B) following the "Type of Organization" labels that categorize the entity where the breach occurred. (C) Rug plot (magenta) and histogram (green) of the available 9,001 breach sizes, along with the histogram on the log-scale (blue).

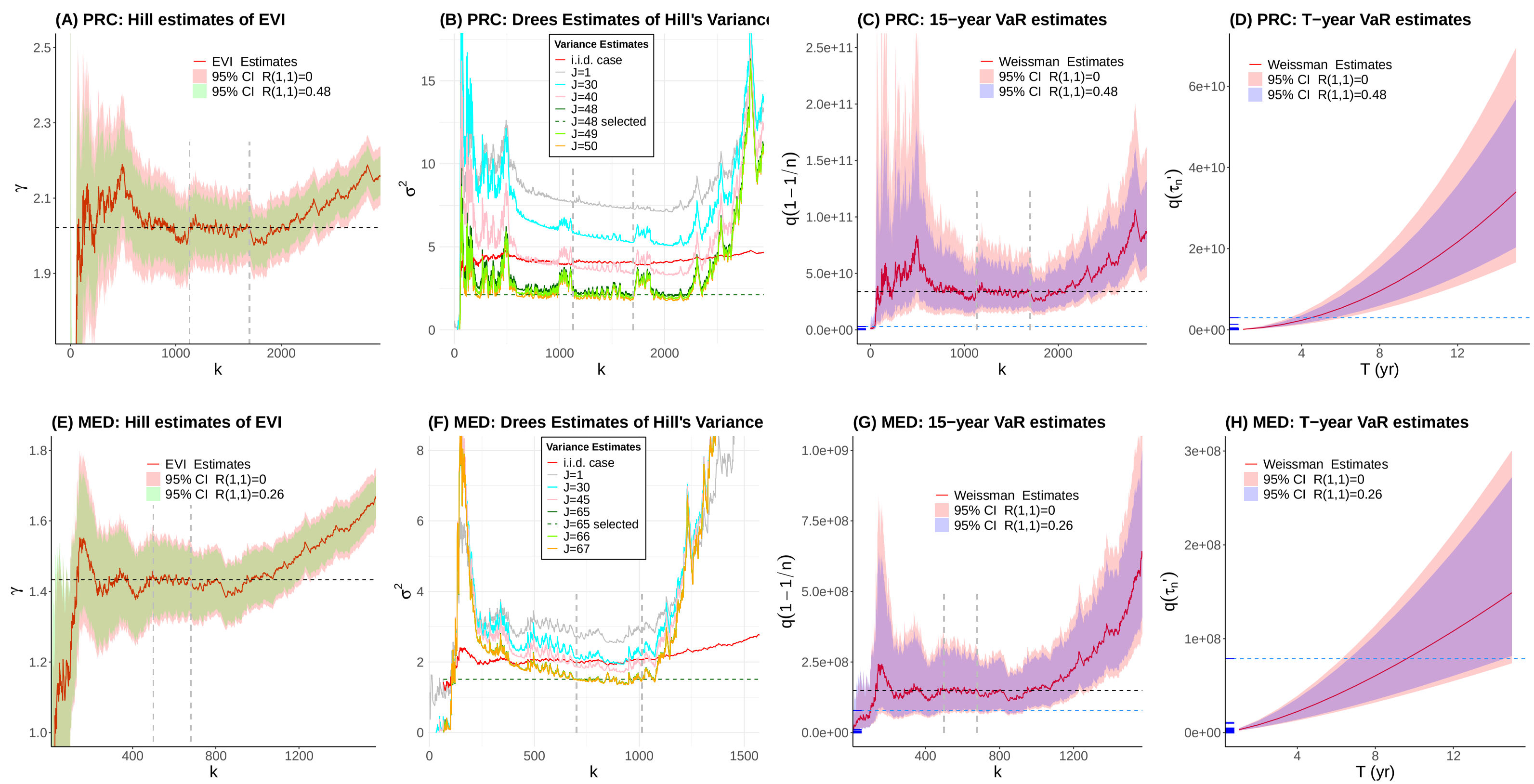

Figure 2: Results for the PRC dataset of 9,001 breach sizes (top panels) and for the MED cluster of 4,337 breach sizes (bottom panels). Panels (A) and (E): Hill plot $k \mapsto \widehat{\gamma}_n(k)$ (red line), 95% confidence interval (green shaded), benchmark i.i.d.-based confidence interval (red shaded), stability region of the plot (vertical dashed lines), and selected estimate $\widehat{\gamma}_n$ (horizontal dashed line). Panels (B) and (F): Drees plots $k \mapsto \widehat{\sigma}_n^2(J,k)$ for various numbers J, with the selected estimate $\widehat{\sigma}_n^2$ (horizontal dashed line), the associated stability region of the plot (vertical dashed lines) and the benchmark i.i.d.-based plot $k \mapsto (\widehat{\gamma}_n(k))^2$ (red line). Panels (C) and (G): Weissman plot $k \mapsto \widehat{q}_n^*(\tau_n'|k)$ for $\tau_n' = 1 - 1/n$ (red line), 95% confidence interval (blue shaded), i.i.d.-based confidence interval (red shaded), selected estimate $\widehat{q}_n^*(1-1/n)$ (horizontal dashed black line), rug plot of the data (blue) along the y-axis, and sample maximum (horizontal dashed blue line). Panels (D) and (H): Final return level $\widehat{q}_n^*(\tau_n')$ against $T = 1, \ldots, 15$ (red line), 95% confidence intervals (shaded areas), rug plot (blue) and sample maximum (dashed blue line).

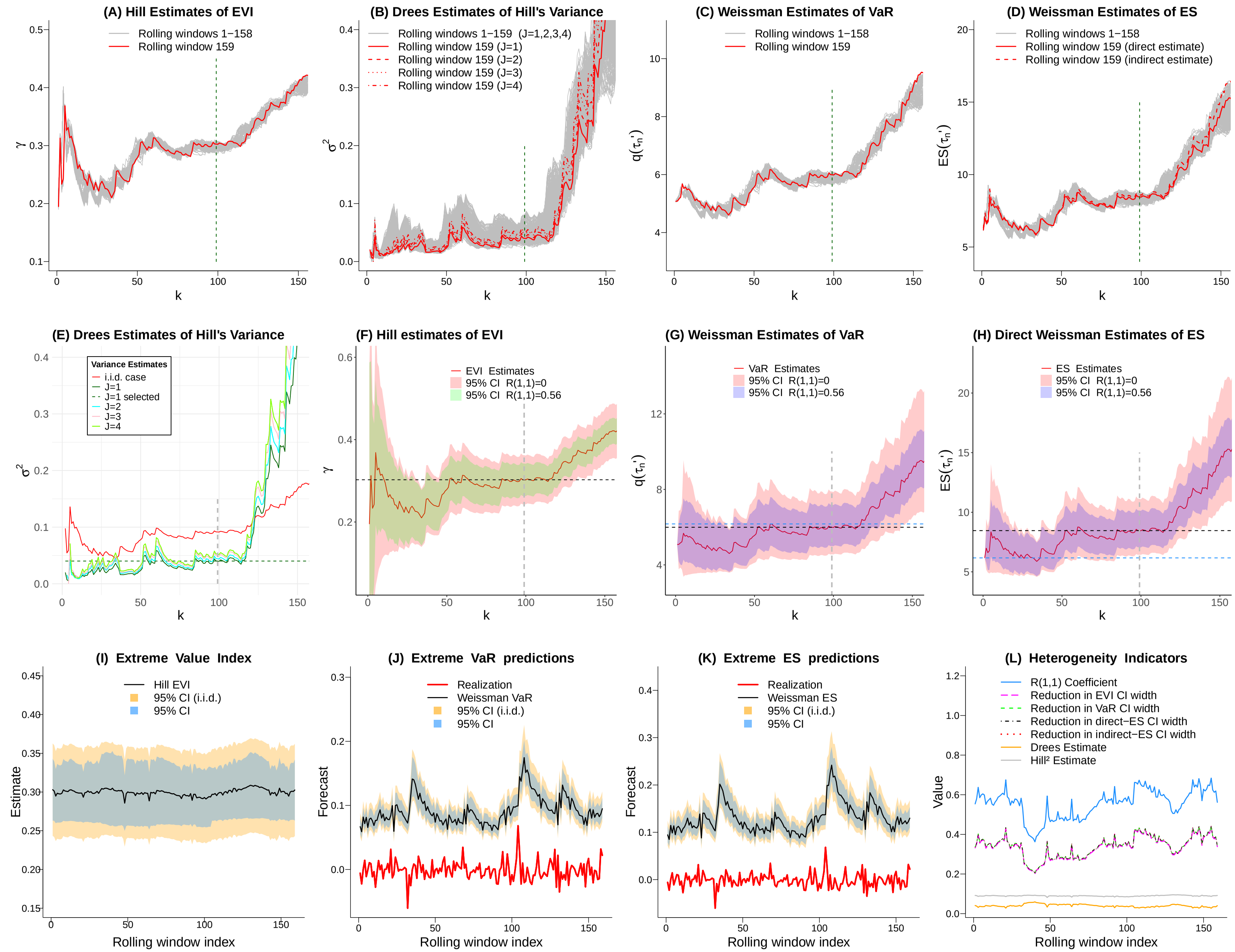

Figure 3: Results for AXP daily loss returns from 20 March 2018 to 18 October 2024. Top panels: overlay of estimate plots based on the residuals from 159 successive rolling windows of length $n = 1{,}500$, with $k \mapsto \widehat{\gamma}_n(k)$ in panel (A), $k \mapsto \widehat{\sigma}^2_n(J,k)$ in panel (B) for $J \in \{1,2,3,4\}$, $k \mapsto \widehat{q}^\star_n(\tau'_n|k)$ in panel (C), and both $\widehat{\mathrm{ES}}^\star_n(\tau'_n|k)$ and $\widetilde{\mathrm{ES}}^\star_n(\tau'_n|k)$ versus k in panel (D), for $\tau'_n = 1 - 1/n$. Middle panels: over the 159th and final window, asymptotic variance estimates $\widehat{\sigma}^2_n(J,k)$ and $(\widehat{\gamma}_n(k))^2$ in panel (E), Hill estimates $\widehat{\gamma}_n(k)$ in panel (F), VaR estimates $\widehat{q}^\star_n(\tau'_n|k)$ in panel (G) and direct ES estimates $\widehat{\mathrm{ES}}^\star_n(\tau'_n|k)$ in panel (H), with selected pointwise estimates (dashed black), 95% confidence intervals (shaded areas), and maximum residual (dashed blue). Bottom panels: over each window, the final Hill estimate (black) in panel (I) with 95% confidence interval (blue shaded) and i.i.d.-based interval (orange shaded), the next-day forecasts (black) $\widehat{q}^\star_n(\tau'_n|k)$ in panel (J) and $\widehat{\mathrm{ES}}^\star_n(\tau'_n|k)$ in panel (K) for the original loss returns with 95% confidence intervals (shaded areas) and the realization of the future observation (red), and the final estimates $\widehat{R}_n(1,1)$ (blue), $\widehat{\sigma}^2_n(J,k)$ (orange) and $\widehat{\gamma}^2_n(k)$ (gray) in panel (L) with the percentage reduction in the width of the confidence intervals (dot-dash lines) associated with $\widehat{\gamma}_n(k)$, $\widehat{q}^\star_n(\tau'_n|k)$, $\widehat{\mathrm{ES}}^\star_n(\tau'_n|k)$ and $\widetilde{\mathrm{ES}}^\star_n(\tau'_n|k)$.