Four contemporary problems in extreme value analysis

Jonathan El Methni, Stéphane Girard, Juliette Legrand, Gilles Stupfler, Antoine Usseglio-Carleve

Four contemporary problems in extreme value analysis

Jonathan El Methni[a], Stéphane Girard[a], Juliette Legrand[b],
Gilles Stupfler[c] & Antoine Usseglio-Carleve[d]

[a] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

[b] Univ Brest, CNRS UMR 6205, Laboratoire de Mathématiques de Bretagne Atlantique, France

[c] Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France

[d] Avignon Université, LMA UPR 2151, 84000 Avignon, France

Abstract. This article gives a summary of recent results on extreme value analysis that were presented at the Journées MAS 2024 in Poitiers. We first set the general background and motivation for these results, and we then discuss partial solutions to four contemporary problems in extreme value analysis: the construction of bias-reduced estimators of the Expected Shortfall (or Tail-Value-at-Risk) at extreme levels, extremal regression and inference in the presence of dependent data, multivariate inference about extreme quantiles using an analogue of the classical ANOVA method in regression, and improved estimation of tail risk measures using nonparametric resampling of multivariate Generalized Pareto distributions.

Keywords: Bias reduction, Extremal regression, Heavy-tailed distributions, Multivariate inference, Sampling methods, Tail risk measures, Time series

1 Background

Extreme value theory is a subfield of probability and statistics whose objective is to model, estimate and infer the characteristics of rare events. Applications of extreme value theory can be found in, among others, insurance and finance (see *e.g.* p.9 in Embrechts et al., 1997), the natural sciences (see Sections 1.3.1, 1.3.2 and 1.3.5 in Beirlant et al., 2004) as well as niche fields such as the analysis of teletraffic data (see Section 8 in Resnick, 2007). A particular field of application of extreme value theory is risk management, which is concerned with the analysis of rare events with high potential impact on the environment and the wider economy, from natural disaster risk (Bousquet and Bernardara, 2021), asset and investment risk that may result in systemic risk (Cai et al., 2015), as well as more recent areas of concern such as cyber risk (Farkas et al., 2021). A standard univariate risk measure is the quantile $q(\tau)$ of a random risk variable X, for a suitably chosen level $\tau \in (0,1)$; in finance, one typically takes X to be the negative of a net financial result, in which case high quantiles give a picture of high losses, and the quantile is then called Value-at-Risk or VaR. A typical goal of extreme value analysis, in such contexts, is the estimation of extreme quantiles of a univariate random variable of interest, such as the daily log-return of a stock market index, the magnitude of earthquakes in a given region, or the size of data packets transferred in a computer network.

The basic problem in extreme value analysis is the estimation of extreme quantiles which may lie beyond the range of the data. To put it differently, given an i.i.d. sample of data $(X_1, \ldots, X_n)$ of size n, we seek an estimator of an extreme quantile $q(1-p_n)$, with $p_n \downarrow 0$ and $np_n = \mathrm{O}(1)$. In univariate extreme value theory, this is the focus of the so-called Peaks-over-Threshold approach which rests upon the assumption, motivated by the Pickands-Balkema-de Haan theorem, that exceedances of the random variable of interest X above a high threshold can be modeled to an acceptable degree of approximation by a Generalized Pareto distribution. Equivalently, if F is the distribution function of X, q is its quantile function, and $U : t \mapsto q(1-1/t)$ is its tail quantile function, the basic assumption underlying the Peaks-over-Threshold method is that there exist $\gamma \in \mathbb{R}$, called the extreme value index of X, and a positive scale function a such that

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

or, equivalently, if $\sigma(t) = a(1/(1-F(t)))$,

$$\forall x > 0 \text{ with } 1+\gamma x > 0,\ \lim_{t\to\infty} \mathbb{P}\left(\frac{X-t}{\sigma(t)} > x \,|\, X > t\right) = (1+\gamma x)^{-1/\gamma}.$$

See Theorem 1.1.6 in de Haan and Ferreira (2006). These two equivalent conditions play a fundamental role in the construction of semiparametric extreme quantile estimators of X; they can be rewritten as the extrapolation formula $U(tx) \approx U(t) + a(t)\int_1^x s^{\gamma-1}\mathrm{d}s$ when t is large, on the one hand, and as

$$\frac{X-t}{\sigma(t)} \,|\, X > t \xrightarrow{\mathrm{d}} \text{Generalized Pareto}(\gamma)$$

as $t \to \infty$, on the other hand. The use of this extrapolation formula in statistical practice rests upon the construction of estimators for the location parameter $U(t)$, the scale parameter $a(t)$ and the shape parameter γ.

In the heavy-tailed setting $\gamma > 0$, which gathers many applications of extreme value theory to insurance, finance and climate science (Beirlant et al., 2004; Embrechts et al., 1997; Resnick, 2007), the extrapolation formula in the Pickands-Balkema-de Haan theorem takes the simpler form

$$\forall x > 0,\ \lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma} \Leftrightarrow \forall x > 0,\ \lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\gamma}.$$

See Theorem 1.2.1 and Corollary 1.2.10 in de Haan and Ferreira (2006). The latter approximation then suggests the heuristic extrapolation formula

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

This extrapolation formula entails that an extreme quantile $q(1-p_n)$ of a heavy-tailed distribution may be consistently estimated provided one can consistently estimate a quantile at a much lower (but still high) level τ and the extreme value index γ. A typical choice for $\tau = \tau_n$ is $\tau_n = 1 - k/n$, where $k = k(n)$ is a sequence of integers such that $k \to \infty$ and $k/n \to 0$; with this choice of τ, the empirical quantile $\widehat{q}_n(1-k/n) = X_{n-k:n}$, where $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ are the order statistics from the sample $(X_1, \ldots, X_n)$, is a consistent estimator of $q(1-k/n)$ under mild conditions. A standard choice for the estimator of γ, which turns out to be the Maximum Likelihood Estimator in a pure Pareto model, is the Hill estimator (Hill, 1975), namely

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

This suggests the classical Weissman-Hill extreme quantile estimator (Weissman, 1978):

$$\widehat{q}_n^{\star}(1-p_n) = \left(\frac{k}{np_n}\right)^{\widehat{\gamma}_n(k)} X_{n-k:n} = d_n^{\widehat{\gamma}_n(k)} X_{n-k:n}, \tag{2}$$

where $d_n = k/(np_n)$ is interpreted as an extrapolation factor. The rationale behind this construction is that one may use the heavy right tail assumption to extrapolate intermediate quantiles to properly extreme levels by using the shape of the tail of the underlying distribution. The key to the asymptotic analysis of $\widehat{q}_n^{\star}(1-p_n)$ is the identity

$$\log\left(\frac{\widehat{q}_n^{\star}(1-p_n)}{q(1-p_n)}\right) = \log(d_n)(\widehat{\gamma}_n(k)-\gamma) + \log\left(\frac{\widehat{q}_n(1-k/n)}{q(1-k/n)}\right) + \log\left(d_n^{\gamma}\frac{q(1-k/n)}{q(1-p_n)}\right). \tag{3}$$

As a consequence, the asymptotic behavior of $\widehat{q}_n^{\star}(1-p_n)$ is determined by the weak convergence properties of the random pair $\sqrt{k}(\widehat{\gamma}_n(k)-\gamma, \log(\widehat{q}_n(1-k/n)/q(1-k/n)))$ and by the asymptotics of the final, nonrandom bias term in (3). Obtaining this asymptotic behavior formally can be done under the following classical extreme value assumptions.

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

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

$\mathcal{K}(\lambda)$ Under $\mathcal{C}_2(\gamma, \rho, A)$, the sequence $k = k(n)$ satisfies $k \to \infty$, $k/n \to 0$ and $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$.

Assumption $\mathcal{C}_2(\gamma, \rho, A)$ is very mild, and is satisfied by every continuous heavy-tailed probability distribution that is commonly used in statistical modeling, see a non-exhaustive list in Table 4 in Daouia et al. (2024b). It quantifies, through the function A, the bias incurred in extreme value procedures by using the Pareto distribution as a model rather than the correct underlying distribution. It can be shown that $|A|$ is necessarily regularly varying in a neighborhood of infinity with index ρ by viewing any function $\overline{F}$ satisfying $\mathcal{C}_2(\gamma, \rho, A)$ within the wider framework of extended regular variation and by applying Theorem B.2.1 in de Haan and Ferreira (2006). When $\rho < 0$, this motivates the assumption of a further, simpler parametric form:

$$A(y) \sim \beta\gamma y^{\rho}, \text{ as } y \to \infty, \tag{4}$$

with $\rho < 0$ and $\beta \neq 0$. This situation arises for instance in the Hall-Welsh class of heavy-tailed distributions (Hall and Welsh, 1985). Condition $\mathcal{K}(\lambda)$ imposes a value of k that not only is intermediate, meaning that the quantile $q(1 - k/n)$ can be consistently estimated by the nonparametric estimator $X_{n-k:n}$, but also that makes it possible to control the bias in the estimation of γ by $\widehat{\gamma}_n(k)$. Then, under conditions $\mathcal{C}_2(\gamma, \rho, A)$ and $\mathcal{K}(\lambda)$,

$$\sqrt{k}\left(\widehat{\gamma}_n(k) - \gamma, \log\left(\frac{\widehat{q}_n(1-k/n)}{q(1-k/n)}\right)\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\frac{\lambda}{1-\rho}, 0\right), \begin{pmatrix}\gamma^2 & 0\\ 0 & \gamma^2\end{pmatrix}\right).$$

This is readily shown by, for example, following the proof of Theorem 3.2.5 in de Haan and Ferreira (2006) using tail empirical quantile processes, see p.76 therein. Moreover, the final, nonrandom term in (3) is, when $\rho < 0$, a $\mathrm{O}(1/\sqrt{k})$ if $k/(np_n) \to \infty$, see the top equation on p.139 of de Haan and Ferreira (2006). As a conclusion, when also $\sqrt{k}/\log(k/(np_n)) \to \infty$, then

$$\frac{\sqrt{k}}{\log(d_n)} \log\left(\frac{\widehat{q}^{\star}_n(1-p_n)}{q(1-p_n)}\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

This asymptotic normality result then yields asymptotic Gaussian confidence intervals whose coverage probability is fairly close to the nominal level in reasonable situations (see Figures B1–B3 in Buitendag et al., 2020), although refined confidence intervals exist (Buitendag et al., 2020; Gardes and Maistre, 2023).

The methodology for extreme quantile inference based on univariate i.i.d. data is therefore well-established. In applications, however, it is often of interest to go beyond extreme quantile estimation, since the quantile risk measure $q(\tau)$ is unable to give any idea of the shape of the distribution of X beyond the level $q(\tau)$, and fails to constitute a coherent risk measure in a financial and actuarial sense, see Artzner et al. (1999) and Acerbi (2002). An alternative to the quantile that deals with these issues is the Expected Shortfall at level τ, namely

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

When the distribution of X is continuous, this is nothing but the τ-Conditional Value-at-Risk or Conditional Tail Expectation (Rockafellar and Uryasev, 2002):

$$\mathrm{CTE}(\tau) = \mathbb{E}(X \,|\, X > q(\tau)). \tag{5}$$

The Expected Shortfall indeed takes the values of X beyond $q(\tau)$ into account and, being a spectral risk measure with positive and nonincreasing risk spectrum, it is coherent and comonotonically additive, see Theorem 4.47 p.180 and Remark 4.85 p.199 in Föllmer and Schied (2004). Major regulators, such as the EU, UK, Australia and Canada, will require the use of ES(97.5%), rather than VaR(99%), in alternative internal models from 1 January 2025. The EU has formally codified this requirement in Article 325ba(1) of Regulation (EU) No 2019/876, itself a revision of the Capital Requirements Regulation (EU) No 575/2013, implementing the Basel Committee on Banking Supervision rules. The question of estimating the Expected Shortfall above extreme levels τ is therefore of high practical importance. Another practical problem in recent applications is to construct reliable estimation and inference methodologies when the structure of the data is more complex than univariate and i.i.d., *i.e.* when it features serial dependence, spatial heterogeneity, and/or is multivariate. In particular, the problem of multivariate inference about extreme quantiles has, perhaps quite surprisingly, been left largely untouched until a series of papers from 2016 to 2019 (Dematteo and Clémençon, 2016; Kinsvater et al., 2016; Hoga, 2018; Stupfler, 2019) which developed theoretical tools in order to obtain the joint weak convergence of several Hill estimators.

The purpose of this article is to give a brief account of four papers, presented on 28th August 2024 at the "Extrêmes" parallel session of talks of the Journées MAS 2024 in Poitiers, France, which discuss solutions to these contemporary problems. We start in Section 2 with the development of a bias reduction methodology for the estimation of the CTE above extreme levels in the presence of potentially very heavy tails, based upon Allouche et al. (2024a). Section 3 deals with nonparametric extremal regression upon serially dependent and spatially heterogeneous data, based upon Daouia et al. (2023). Section 4 then considers the problem of testing for equality of extreme quantiles by constructing a test statistic reminiscent of the traditional ANOVA statistic in linear regression, based upon a recent paper (Girard et al., 2024). Finally, Section 5 adopts a resolutely multivariate viewpoint and introduces a data augmentation methodology using resampling with the goal of improving the accuracy of existing tail Expected Shortfall estimators. This last section is based upon Legrand et al. (2023) and a recent preprint (Madhar et al., 2024).

2 Estimation of extreme risk measures for heavy-tailed data

The existence of the CTE (5) requires the random variable X to be integrable while the asymptotic normality of its empirical counterpart assumes the existence of the second moment, see (Brazauskas et al., 2008, Theorem 3.1). In heavy-tailed models, the integrability assumption may not hold as soon as $\gamma \geq 1$ and the finite second moment condition may be violated when $\gamma \geq 1/2$. This can be the case in insurance, see for instance El Methni and Stupfler (2018) where the extreme value index is estimated on commercial fire losses at $\widehat{\gamma}_n \simeq 0.7$ and in Beirlant et al. (2004, Example 1.2) where $\widehat{\gamma}_n \simeq 0.8$ on Norwegian fire data in 1976 (studied hereafter for the year 1990).

As an alternative, we consider the CTE for Box-Cox transforms (Box and Cox, 1964) of X:

$$\mathcal{M}_a(\tau) = \mathbb{E}(K_a(X) \mid X > q(\tau)), \text{ with } K_a(x) = \int_1^x u^{a-1}\mathrm{d}u, \tag{6}$$

and $a \in \mathbb{R}$. It can then be shown that $\mathcal{M}_a(\tau)$ exists for τ close to 1 provided that $a\gamma < 1$, and, moreover, from Allouche et al. (2024a, Lemma 5(i)):

$$\lim_{\tau\to 1} \frac{\mathcal{M}_a(\tau) - K_a(q(\tau))}{1 + aK_a(q(\tau))} = \frac{\gamma}{1-a\gamma}. \tag{7}$$

Clearly, considering $\mathcal{M}_a(\tau)$ permits to recover various risk measures since $\mathrm{CTE}(\tau) = \mathcal{M}_1(\tau) + 1$, the conditional tail variance (Valdez, 2005) is given by $\mathrm{CTV}(\tau) = 2(\mathcal{M}_2(\tau) - \mathcal{M}_1(\tau)) - \mathcal{M}_1^2(\tau)$ and more generally, the conditional tail moment introduced in El Methni et al. (2014) defined when $a \neq 0$ by $\mathrm{CTM}_a(\tau) = \mathbb{E}\left(X^a | X > q(\tau)\right)$ can be rewritten as $\mathrm{CTM}_a(\tau) = 1 + a\mathcal{M}_a(\tau)$.

Here, we propose non-parametric estimators of $\mathcal{M}_a(\tau_n)$ in the challenging situation where $\tau_n \to 1$ as $n \to \infty$ and when X is heavy-tailed. A direct estimator of $\mathcal{M}_a$ at an intermediate level is obtained by considering the empirical counterpart of (6):

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

where $k \to \infty$ with $k/n \to 0$ as $n \to \infty$. Similarly, one can introduce the following (direct) estimator of $\mathcal{M}_a(1-p_n)$ at an extreme level:

$$\widehat{\mathcal{M}}_{a,n}^{(\mathrm{D})}(1 - p_n; \widehat{\gamma}_n(k)) = \frac{1}{k}\sum_{i=1}^{k} K_a\left(X_{n-i+1:n}\, d_n^{\widehat{\gamma}_n(k)}\right), \tag{8}$$

with $p_n \to 0$ as $n \to \infty$. Here, $\widehat{\gamma}_n(k)$ is the Hill estimator (1), but bias-corrected versions (see for instance Caeiro et al. (2005)) can be considered. It is also possible to derive from (7) that

$$\mathcal{M}_a(\tau) \simeq \frac{1}{1-a\gamma}\left(K_a(q(\tau)) + \gamma\right), \tag{9}$$

when τ is close to one, and to build an indirect estimator of $\mathcal{M}_a$ at an intermediate level by estimating the intermediate quantile $q(1-k/n)$ by the associated order statistic:

$$\widetilde{\mathcal{M}}_{a,n}^{(\mathrm{I})}(1 - k/n; \widehat{\gamma}_n(k)) = \frac{1}{1-a\widehat{\gamma}_n(k)}\left(K_a(X_{n-k:n}) + \widehat{\gamma}_n(k)\right).$$

Note that $\widetilde{\mathcal{M}}_{0,n}^{(\mathrm{I})}(1-k/n;\widehat{\gamma}_n(k)) = \widetilde{\mathcal{M}}_{0,n}^{(\mathrm{D})}(1-k/n)$ if $\widehat{\gamma}_n$ is the Hill estimator. This estimator can also be used to build an indirect estimator of $\mathcal{M}_a$ at an extreme level:

$$\widehat{\mathcal{M}}_{a,n}^{(\mathrm{I})}(1-p_n;\widehat{\gamma}_n(k)) = \frac{1}{1-a\widehat{\gamma}_n(k)}\left(K_a\left(\widehat{q}_n^{\star}(1-p_n)\right)+\widehat{\gamma}_n(k)\right), \tag{10}$$

where $\widehat{q}_n^{\star}(1-p_n)$ is the Weissman estimator (2) of the extreme quantile $q(1-p_n)$.

In the sequel, we focus on the estimators (8) and (10) of $\mathcal{M}_a$ at an extreme level. The asymptotic normality of the direct and indirect estimators $\widehat{\mathcal{M}}_{a,n}^{(\mathrm{D})}(1-p_n;\widehat{\gamma}_n(k))$ and $\widehat{\mathcal{M}}_{a,n}^{(\mathrm{I})}(1-p_n;\widehat{\gamma}_n(k))$ is established in Allouche et al. (2024a, Theorem 4 and Theorem 6), under the conditions $a\gamma<1/2$ and $a\gamma<1$ respectively. It appears that both estimators are asymptotically biased under the assumptions $\mathcal{C}_2(\gamma,\rho,A)$ and $\mathcal{K}(\lambda)$. In particular, $\widehat{\mathcal{M}}_{a,n}^{(\mathrm{D})}(1-p_n;\widehat{\gamma}_n(k))$ has an asymptotic bias involving two components: a first one inherited from the estimator of the extreme value index γ and a second one due to the extrapolation method. The same observations have been made in the case of the Weissman estimator dedicated to extreme quantiles in Allouche et al. (2023). The asymptotic bias can thus be removed by using a reduced-bias estimator $\hat{\gamma}_n^{(\mathrm{RB})}$ of γ and by estimating and removing the bias term due to the second-order terms. More specifically, assume that $\sqrt{k}(\hat{\gamma}_n^{(\mathrm{RB})}-\gamma) \xrightarrow{\mathrm{d}} \mathcal{N}(0,\sigma^2)$ for some $\sigma>0$ under $\mathcal{C}_2(\gamma,\rho,A)$ and $\mathcal{K}(\lambda)$. Then, under the additional assumption (4), the reduced-bias version of the direct estimator of $\mathcal{M}_a$ at an extreme level is given by

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

where $\hat{\rho}_n$ and $\hat{\beta}_n$ are estimators of the second-order parameters ρ and β respectively such that $(\log(n/k))(\hat{\rho}_n-\rho)=\mathrm{o}_{\mathbb{P}}(1)$ and $\hat{\beta}_n \xrightarrow{\mathbb{P}} \beta$. In contrast, the asymptotic bias of $\widehat{\mathcal{M}}_{a,n}^{(\mathrm{I})}(1-p_n;\widehat{\gamma}_n(k))$ only stems from the asymptotic bias of $\widehat{\gamma}_n$ and can be removed by considering a reduced-bias estimator of γ. It is then proved under the above mentioned assumptions that both

$$\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widehat{\mathcal{M}}_{a,n}^{(\mathrm{D,RB})}\left(1-p_n;\hat{\gamma}_n^{(\mathrm{RB})}\right)-\mathcal{M}_a(1-p_n)}{1+a\mathcal{M}_a(1-p_n)}\right)$$
$$\text{and } \frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widehat{\mathcal{M}}_{a,n}^{(\mathrm{I})}\left(1-p_n;\hat{\gamma}_n^{(\mathrm{RB})}\right)-\mathcal{M}_a(1-p_n)}{1+a\mathcal{M}_a(1-p_n)}\right)$$

are asymptotically $\mathcal{N}(0,\sigma^2)$ distributed, see Allouche et al. (2024a, Theorem 5 and Corollary 1).

As mentioned, we consider fire insurance claims of a Norwegian insurance company for the year 1990. This data set was already analyzed in Beirlant et al. (2001) and consists of $n=628$ claims in units of 1000 kr (Norwegian crown) larger than the threshold 500,000 kr. The Hill plot $k \mapsto \widehat{\gamma}_n(k)$ presented in Figure 1 (left panel) features a nice stability region on the range $k\in[150,300]$ indicating a heavy-tail behaviour of the data. Two procedures were used to select the threshold, (Allouche et al., 2024a, Algorithm 1) and Beirlant et al. (1999), giving respectively $k^{\star}=279$ and $k^{\dagger}=290$. Both lead to the same estimation for the extreme value index: $\widehat{\gamma}_n(k^{\dagger})=\widehat{\gamma}_n(k^{\star})=0.62$. Moreover, the log quantile-quantile plot of the pairs $(\log((n+1)/i),\log X_{n-i+1:n})$ for $i\in\{1,\ldots,k^{\star}\}$ drawn on Figure 1 (right panel) appears to be approximately linear. This constitutes empirical evidence that the heavy-tail assumption makes sense and that $k^{\star}=279$ is a reasonable choice for the threshold. Besides, the estimated second-order parameter provided by the package `evt0` of the `R` software (Manjunath et al., 2013) is $\hat{\rho}_n=-1.85$, corresponding to a relatively low bias situation.

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

$$\widehat{\mathcal{M}}_{a,n}^{(\mathrm{EMP})}(1-1/\lceil n\xi\rceil) = \frac{1}{\lceil n(1-\xi)\rceil}\sum_{i=1}^{\lceil n(1-\xi)\rceil}K_a(X_i^{\mathrm{test}}) = \frac{1}{\lceil n(1-\xi)\rceil}\sum_{i=1}^{\lceil n(1-\xi)\rceil}K_a(X_{n-i+1:n}) \tag{11}$$

	(EMP)	(D)	(D, RB)	(I)	(I, RB)
$\widehat{\mathrm{CTE}}_n(1-p_n;1)$		14724	7267	15624	4963
$\widehat{\mathrm{CTE}}_n(1-p_n;1/2)$		16606	7852	17056	5307
$\widehat{\mathrm{CTE}}_n^{(\mathrm{EMP})}(1-p_n)$	8732				

Table 1: Norwegian fire data. Estimations of $\mathrm{CTE}(1-p_n)$ (in 1000 kr) obtained from (12), (13) and (14) at the risk level $1-p_n = 1-1/\lceil n\xi\rceil$ with $\xi = 0.9$.

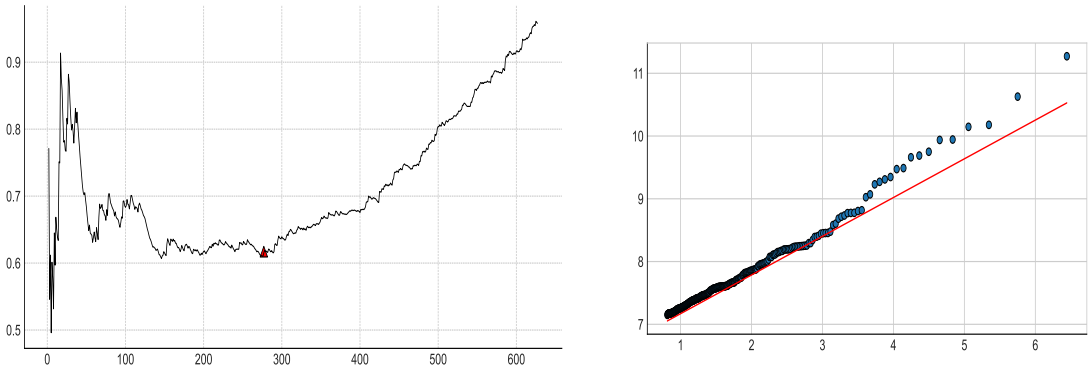


Figure 1: Norwegian fire data. Left panel: Hill estimator $\widehat{\gamma}_n(k)$ as a function of k and pair $(k^\star, \widehat{\gamma}_n(k^\star))$ selected by Allouche et al. (2024a, Algorithm 1) emphasized by a red triangle. Right panel: Log quantile-quantile plot (horizontally: $\log((n+1)/i)$, vertically: $\log(X_{n-i+1:n})$ for $i \in \{1,\ldots,k^\star\}$). The approximating line with slope $\widehat{\gamma}_n(k^\star)$ is superimposed in red.

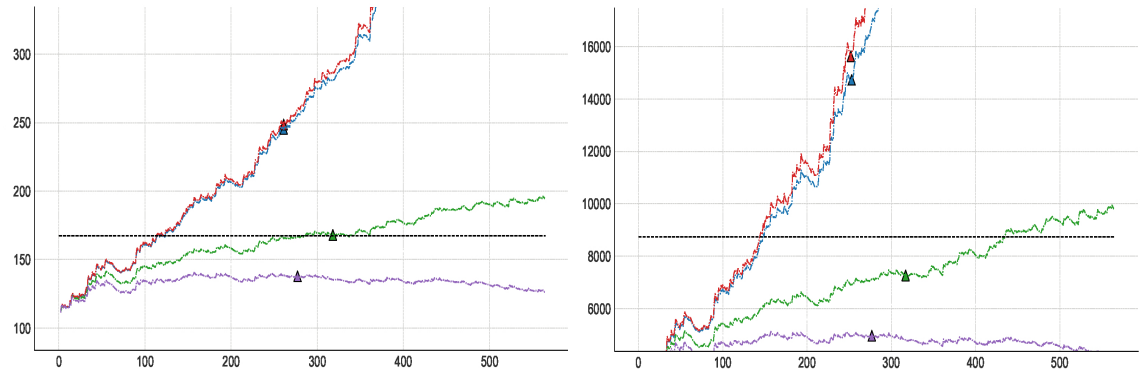


Figure 2: Norwegian fire data. Left panel: estimation of $\mathcal{M}_{1/2}(1-p_n)$ and right panel: estimation of $\mathcal{M}_1(1-p_n)$, for $p_n = 1/\lceil n\xi\rceil$, associated with $\widehat{\mathcal{M}}^{(\mathrm{EMP})}$ (black), $\widehat{\mathcal{M}}^{(\mathrm{D})}$ (blue), $\widehat{\mathcal{M}}^{(\mathrm{D,RB})}$ (green), $\widehat{\mathcal{M}}^{(\mathrm{I})}$ (red) and $\widehat{\mathcal{M}}^{(\mathrm{I,RB})}$ (purple) estimators as functions of $k \in \{2,\ldots,\lceil n\xi\rceil\}$ with $\xi = 0.9$. Estimations at the selected level $k^\star$ are emphasized by a triangle.

is the direct estimator of $\mathcal{M}_a(1-p_n)$ with $p_n = 1/\lceil n\xi \rceil$ computed on $\mathcal{X}^{\text{test}}$. Our first goal is to compute $\mathcal{M}_a(1-p_n)$ with $a \in \{1/2, 1\}$ at the extreme risk level $p_n = 1/\lceil n\xi \rceil$ on $\mathcal{X}^{\text{train}}$ using the four above competitors and to compare the results with the empirical estimation computed on $\mathcal{X}^{\text{test}}$ thanks to (11). As it can be seen on Figure 2, for respectively $a = 1/2$ (left panel) and $a = 1$ (right panel), both estimators $\widehat{\mathcal{M}}^{(\text{I})}$ and $\widehat{\mathcal{M}}^{(\text{D})}$ tend to be largely above the empirical estimate $\widehat{\mathcal{M}}^{(\text{EMP})}$, while the two reduced-bias ones are only slightly below the empirical estimate and benefit from a nice stable behaviour in a large neighborhood around their respective estimated values. The direct reduced-bias estimator seems to provide the closest estimations to the empirical ones (11) with respectively $\widehat{\mathcal{M}}_{1/2,n}^{(\text{D,RB})} = 168.0$ versus $\widehat{\mathcal{M}}_{1/2,n}^{(\text{EMP})} = 167.2$ and $\widehat{\mathcal{M}}_{1,n}^{(\text{D,RB})} = 7265.9$ versus $\widehat{\mathcal{M}}_{1,n}^{(\text{EMP})} = 8731.0$. The good performance of the direct reduced-bias estimator here is probably due to the low estimated bias ($\hat{\rho}_n = -1.85$) and the large estimated extreme-value index ($\widehat{\gamma}_n > 0.6$), which is a behavior already observed on simulated data, see (Allouche et al., 2024a, Section 4). Our second objective is to recover from the previous results some estimates of the CTE at the extreme level $p_n = 1/\lceil n\xi \rceil$. Since $\text{CTE}(1-p_n) = 1 + \mathcal{M}_1(1-p_n)$, one can straightforwardly obtain

$$\widehat{\text{CTE}}_n(1-p_n;1) = 1 + \widehat{\mathcal{M}}_{1,n}(1-p_n), \tag{12}$$

to get four estimators of the CTE with $\widehat{\mathcal{M}}_{1,n} \in \{\widehat{\mathcal{M}}_{1,n}^{(\text{D})}, \widehat{\mathcal{M}}_{1,n}^{(\text{D,RB})}, \widehat{\mathcal{M}}_{1,n}^{(\text{I})}, \widehat{\mathcal{M}}_{1,n}^{(\text{I,RB})}\}$. Moreover, from (7), one has

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

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

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

Taking advantage of the estimated extreme-value index $\widehat{\gamma}_n(k^\star)$ on $\mathcal{X}^{\text{train}}$, one obtains four additional estimators of the CTE:

$$\widehat{\text{CTE}}_n(1-p_n;1/2) = \Psi\left(\widehat{\mathcal{M}}_{1/2,n}(1-p_n), \widehat{\gamma}_n(k^\star)\right). \tag{13}$$

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

$$\widehat{\text{CTE}}_n^{(\text{EMP})}(1-1/\lceil n\xi \rceil) = \frac{1}{\lceil n(1-\xi) \rceil}\sum_{i=1}^{\lceil n(1-\xi) \rceil} X_i^{\text{test}} = \frac{1}{\lceil n(1-\xi) \rceil}\sum_{i=1}^{\lceil n(1-\xi) \rceil} X_{n-i+1:n}. \tag{14}$$

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

To conclude, we have introduced reduced-bias estimators of the CTE at intermediate and extreme levels for Box-Cox transformed heavy-tailed random variables. The use of the Box-Cox transform permits to define and estimate the CTE even for distributions with large extreme-value indices, the asymptotic normality of all estimators being established in Allouche et al. (2024a). The Box-Cox transform also allows to derive estimators for other risk measures such as the CTM, see El Methni et al. (2014). Finally, the bias reduction leads to estimators with stable sample paths that are easy to use in practice. These properties are illustrated both on simulated (Allouche et al., 2024a, Section 4) and real data thanks to Python code freely available for download. We also refer to Allouche et al. (2024b) for a comparison with neural network approaches. The work presented in this section is univariate in nature and is therefore unable to account for covariate-induced heterogeneity in order to produce a finer picture of extreme value behavior. Doing so is the focus of the next section.

3 Extremal regression with dependent heavy-tailed data

Consider a dataset which records, for each tornado that has occurred in the United States between 1 January 2010 and 31 December 2019, the associated monetary loss (`loss`), its starting and ending latitude and longitude (`slat`, `slon`, `elat` and `elon`), and the length and width of the area it traveled

over (`len` and `wid`). To reduce heterogeneity due to the strength and duration of a tornado, we consider the loss per surface unit Y (in USD) as a function of a tornado's average geographical location $\boldsymbol{X} = (X_1, X_2)$, that is, $Y = \texttt{loss}/(\texttt{len} \times \texttt{wid})$, $X_1 = (\texttt{slon} + \texttt{elon})/2$, and $X_2 = (\texttt{slat} + \texttt{elat})/2$. One then obtains a sample $(Y_t, \boldsymbol{X}_t)$ of size $n = 6{,}360$, including high-impact events such as the tornado which caused a 2.8 billion USD total loss across its path in Joplin, Missouri, on 22 May 2011. A scatterplot of the dataset and a smoothed estimate of the frequency of tornadoes across the eastern half of the United States are given in Figure 3 (a) and (b).

It is clear that the data feature considerable spatial heterogeneity as well as serial dependence. In such a situation, how to construct confidence intervals about extreme conditional quantiles is not obvious. One possible solution is to use extremal linear quantile regression, as worked out by Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011). The linear extreme quantile model is, however, a strong model specification, for which goodness-of-fit testing procedures are not easy to develop and, in fact, do not exist as far as we are aware. Assuming more general forms of regression models, meanwhile, is possible but requires prior extra knowledge about the phenomenon of interest. At the opposite, nonparametric approaches to extremal regression such as those of Daouia et al. (2011) and Daouia et al. (2013) are limited to i.i.d. data. They constitute nonetheless a reasonable starting point which we use to sketch out a solution to the problem of nonparametric extremal regression and inference for conditional extreme quantiles.

Assume that $((Y_t, \boldsymbol{X}_t))_{t\geq 1}$ is a strictly stationary sequence of copies of a random pair $(Y, \boldsymbol{X}) \in \mathbb{R} \times \mathbb{R}^p$. Let $F(\cdot|\boldsymbol{x})$ denote the distribution function of Y given $\boldsymbol{X} = \boldsymbol{x}$. Assume that $\boldsymbol{X}$ has a probability density function (p.d.f.) g on $\mathbb{R}^p$ and fix $\boldsymbol{x} \in \mathbb{R}^p$ with $g(\boldsymbol{x}) > 0$, and a kernel p.d.f. K on $\mathbb{R}^p$. Consider the Nadaraya-Watson estimator of $F(\cdot|\boldsymbol{x})$:

$$\widehat{F}_n(y|\boldsymbol{x}) = \frac{1}{nh_n^p\, \widehat{g}_n(\boldsymbol{x})} \sum_{t=1}^{n} \mathbb{1}_{\{Y_t \leq y\}} K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_t}{h_n}\right) \text{ with } \widehat{g}_n(\boldsymbol{x}) = \frac{1}{nh_n^p} \sum_{t=1}^{n} K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_t}{h_n}\right).$$

In this section, $h_n \to 0$ is a (positive) bandwidth sequence. A conditional quantile $q(\tau|\boldsymbol{x}) \equiv \inf\{y \in \mathbb{R} \,|\, F(y|\boldsymbol{x}) \geq \tau\}$ of $(Y, \boldsymbol{X})$ can then be estimated by its empirical counterpart $\widehat{q}_n(\tau|\boldsymbol{x}) = \inf\{y \in \mathbb{R} \,|\, \widehat{F}_n(y|\boldsymbol{x}) \geq \tau\}$. Just as in the unconditional setting, the idea is to show first that $\widehat{q}_n(\tau_n|\boldsymbol{x})$ is a (relatively) consistent estimator of $q(\tau_n|\boldsymbol{x})$ at intermediate levels τ_n, and then to extrapolate such intermediate estimators using the shape of the tail of the conditional distribution. This is done using a kernel Hill-type estimator of $\gamma(\boldsymbol{x})$ introduced in Daouia et al. (2011):

$$\widehat{\gamma}_{\tau_n}^{(J)}(\boldsymbol{x}) = \frac{1}{\log(J!)} \sum_{j=1}^{J} \log\left(\frac{\widehat{q}_n(1-(1-\tau_n)/j|\boldsymbol{x})}{\widehat{q}_n(\tau_n|\boldsymbol{x})}\right), \text{ for a fixed } J \geq 2.$$

This results in the conditional extreme quantile estimator

$$\widehat{q}^{\,W}_{n,\tau_n}(1-p_n|\boldsymbol{x}) = \left(\frac{p_n}{1-\tau_n}\right)^{-\widehat{\gamma}_{\tau_n}^{(J)}(\boldsymbol{x})} \widehat{q}_n(\tau_n|\boldsymbol{x}).$$

We provide here a brief account of the asymptotic properties of $\widehat{q}_n(\tau_n|\boldsymbol{x})$, $\widehat{\gamma}_{\tau_n}^{(J)}(\boldsymbol{x})$ and $\widehat{q}^{\,W}_{n,\tau_n}(1-p_n|\boldsymbol{x})$; full details can be found in Daouia et al. (2023). The main model assumption in that article is

Condition $\mathcal{M}$ $((Y_t, \boldsymbol{X}_t))_{t\geq 1}$ is a stationary $\alpha-$mixing sequence of copies of a random vector $(Y, \boldsymbol{X})$ whose conditional distribution at $\boldsymbol{X} = \boldsymbol{x}$ satisfies condition $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$.

This condition is used in conjunction with a technical condition in order to develop a "big block-small block" argument. This technique splits $\widehat{F}_n(y|\boldsymbol{x})$ into large blocks of equal size, and then small block-sized holes (whose size nonetheless increases as $n \to \infty$) are created within each large block in order to take advantage of the asymptotic independence condition provided by the mixing assumption. If the size of the small block is negligible relative to that of the large block, then this should not alter the asymptotic behavior of the blocks, and the asymptotics of $\widehat{F}_n(y|\boldsymbol{x})$ (and therefore of $\widehat{q}_n(\tau_n|\boldsymbol{x})$) then follow from viewing large blocks as (essentially) independent random variables.

One also requires anti-clustering assumptions that prevent the covariates from clustering in small balls and conditional extremes from repeating too often. Assume, on the one hand, that

Condition $\mathcal{B}_p$ There exists an integer $t_0 \geq 1$ such that

$$1 \leq t < t_0 \Rightarrow \lim_{r\to 0} r^{-p}\, \mathbb{P}(\boldsymbol{X}_1 \in B(\boldsymbol{x}, r), \boldsymbol{X}_{t+1} \in B(\boldsymbol{x}, r)) = 0$$

$$\text{and} \quad \limsup_{r\to 0} \sup_{t\geq t_0} r^{-2p}\, \mathbb{P}(\boldsymbol{X}_1 \in B(\boldsymbol{x}, r), \boldsymbol{X}_{t+1} \in B(\boldsymbol{x}, r)) < \infty.$$

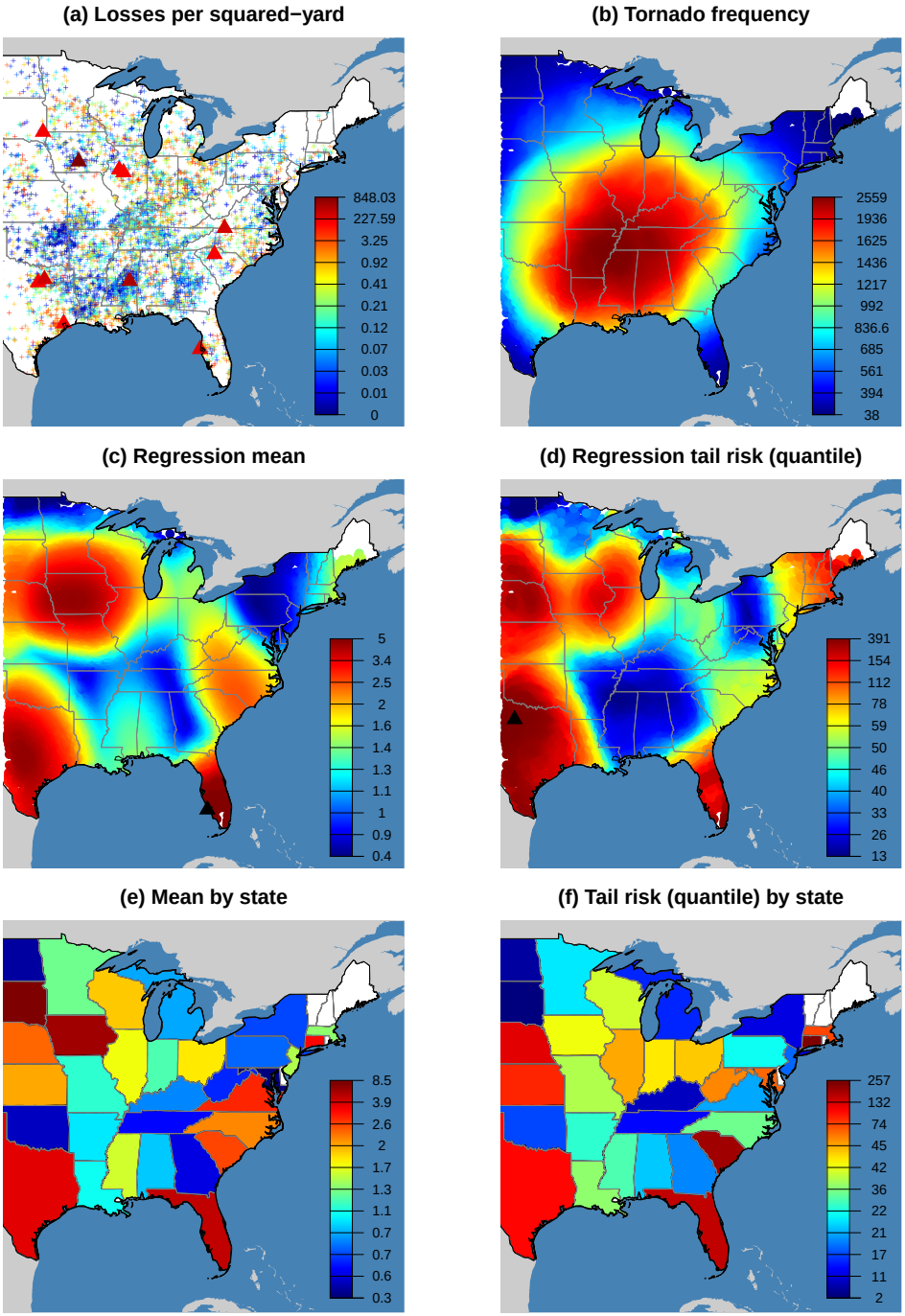


Figure 3: Tornado losses data. Top row, left: Data across the eastern half of the US, right: Local number of observations. Middle row, left: Estimated conditional mean of losses per squared yard, right: Extrapolated conditional quantile estimate of those losses at level $1 - p = 0.995$. Cities with the highest estimated conditional average loss and extreme loss are marked with a black triangle in the left and right panels, respectively. Bottom row, left: Unconditional statewide estimation, using the sample average, right: Using the bias-reduced extreme quantile estimator of Gomes and Pestana (2007) at level $1 - p = 0.995$. Losses (all panels except (b), in USD) and tornado frequency (panel (b)) are indicated by a color scheme, ranging from dark blue (lowest) to dark red (highest).

On the other hand, it is assumed that a joint conditional extreme value of (Y_1, Y_{t+1}) cannot be much more likely than a marginal conditional extreme of Y_1, uniformly across time and locally uniformly across the covariate space, namely

Condition $\mathcal{B}_\Omega$ There exist $h, z > 0$ such that

$$\sup_{t\geq 1} \sup_{y,y'\geq z} \sup_{\boldsymbol{x}',\boldsymbol{x}''\in B(\boldsymbol{x},h)} \frac{\mathbb{P}(Y_1 > y, Y_{t+1} > y' | \boldsymbol{X}_1 = \boldsymbol{x}', \boldsymbol{X}_{t+1} = \boldsymbol{x}'')}{\sqrt{\overline{F}(y|\boldsymbol{x}')\overline{F}(y'|\boldsymbol{x}'')}} = \mathrm{O}(1).$$

This should be considered as a weak requirement compared with the existence of a (conditional) tail copula, as assumed in *e.g.* Drees (2003) and Davison et al. (2023) in the unconditional setting.

Under these conditions, plus adaptations of classical conditions from the regression literature ensuring that regression estimators achieve an optimal pointwise rate of convergence, plus a transparent bias condition and a standard restriction on the mixing rate, one obtains the joint asymptotic normality of $\widehat{\gamma}^{(J)}_{\tau_n}(\boldsymbol{x})$ and $\widehat{q}_n(\tau_n|\boldsymbol{x})$, for $\tau_n \uparrow 1$, and $h_n \to 0$ such that $nh^p_n(1-\tau_n) \to \infty$. This ultimately results in the weak convergence

$$\frac{\sqrt{nh^p_n(1-\tau_n)}}{\log[(1-\tau_n)/p_n]} \left(\frac{\widehat{q}^W_{n,\tau_n}(1-p_n|\boldsymbol{x})}{q(1-p_n|\boldsymbol{x})} - 1 \right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})} \gamma^2(\boldsymbol{x}) \times \frac{J(J-1)(2J-1)}{6\log^2(J!)}\right).$$

It is remarkable that the asymptotic distribution of $\widehat{q}^W_{n,\tau_n}(1-p_n|\boldsymbol{x})$ is exactly the one obtained in the i.i.d. setting by Daouia et al. (2011) under a further unnecessary regularity assumption on conditional tails. This is not true in the unconditional setting, see Drees (2003). This phenomenon, which is very important as far as inference is concerned since it drastically simplifies the estimation of the asymptotic variance, has already been observed for nonparametric conditional Expected Shortfall estimation (Linton and Xiao, 2013, p.784).

The Nadaraya-Watson estimates of the conditional mean of losses per squared yard and a (bias-reduced version of the) estimate $\widehat{q}^W_{n,\tau_n}(1-p_n|\boldsymbol{x})$ are shown in Figure 3 (c) and (d), respectively. An important benefit of the nonparametric approach to extremal regression is its ability to provide a smooth representation of conditional extreme value behavior, without recourse to strong spatio-temporal model specifications. This is made clear by comparing the estimates with those obtained by calculating, in each state, the mean and quantile at level 99.5% from the univariate sample of losses in this state only, see Figure 3 (e) and (f). This approach yields estimates in the riskiest areas that are up to 67% lower than those of the regression method, and produces unrealistic discontinuities in the estimates, see the examples of Texas-Oklahoma-Kansas, Nebraska-South Dakota and Georgia-South Carolina, unlike the kernel regression method with may pool regional information across state boundaries. Inference can also be carried out at specific locations of interest, see Table 1 in Daouia et al. (2023), where corrected confidence intervals based on the asymptotic normality result for $\widehat{q}^W_{n,\tau_n}(1-p_n|\boldsymbol{x})$ are calculated, and compared to those based on purely empirical estimates.

A large part of the difficulty in carrying out inference using nonparametric extremal regression is that the extreme value index of Y given $\boldsymbol{X} = \boldsymbol{x}$ is allowed to depend on $\boldsymbol{x}$. If it were in fact constant with respect to $\boldsymbol{x}$, then under reasonable assumptions (such as those of Einmahl et al., 2016) the observations Y_t could be pooled in order to estimate their common extreme value index. It is, more generally, of interest to test for equality of extreme value features of several samples of data, as a preliminary step before improving estimation and inference via pooling as done, for instance, in Kinsvater et al. (2016) and Daouia et al. (2024a). This is the focus of the next section.

4 Analysis of variability for heavy-tailed extremes

Let us now consider $J > 1$ samples $E_j = \{X^{(j)}_i,\, i = 1, \ldots, n_j\}$, $j = 1, \ldots, J$, with independence between samples and possibly different sample sizes $n_j > 1$ such that $n_j/n \to \lambda_j > 0$ as $n \to \infty$, for $j = 1, \ldots, J$. We assume that the random variables in each E_j are i.i.d. with cumulative distribution function F_j and quantile function q_j satisfying condition $\mathcal{C}_2(\gamma_j, \rho_j, A_j)$. While the classical ANalysis Of VAriance (ANOVA) aims at testing the equality between the means (central quantities) of the J samples using a decomposition of the total variance into intra-class and inter-class variances (Scheffé, 1999), we are interested in testing if the samples have the same tail properties, and introduce an ANalysis Of Variability for EXtremes (ANOVEX). A first possibility, already proposed in the literature (see Mougeot and Tribouley (2010) or Worms and Worms (2015) for instance) may be to test the equality of the extreme value indices $\gamma_1, \ldots, \gamma_J$. However, two distributions may have the same extreme value index and a widely different scale factor. Since two heavy-tailed distributions

have the same shape and scale if and only if their quantiles are asymptotically equivalent (as the quantile level tends to 1), we thus propose to test the equality of the extreme quantiles $q_1, \ldots, q_J$. More precisely, our null hypothesis is defined by

(H_0) For all $(j, j') \in \{1, \ldots, J\}^2$ with $j \neq j'$, we have $q_{j'}(1-p)/q_j(1-p) \to 1$ as $p \to 0$.

Note that, as introduced in Section 2, any extreme quantile $q_{j,n}$ may be suitably estimated using the Weissman estimator $\widehat{q}^{\star}_{j,n}$. The ANOVEX approach is the following: consider $L \geq 1$ extreme levels

$$p_{\ell,n} = \tau_\ell / n, \tau_\ell > 0 \text{ for } \ell = 1, \ldots, L,$$

and focus on the decomposition of the following variability measure:

$$\Delta_n = \frac{1}{JL} \sum_{j=1}^{J} \sum_{\ell=1}^{L} \left(\log \widehat{q}^{\star}_{j,n}(1-p_{\ell,n}) - \mu_{p,n} \right)^2, \quad \text{with } \mu_{p,n} = \frac{1}{JL} \sum_{j=1}^{J} \sum_{\ell=1}^{L} \log \widehat{q}^{\star}_{j,n}(1-p_{\ell,n}).$$

Some straightforward calculations (Girard et al., 2024, Lemma 2.1) show that this total variability may be decomposed into variability measures $\Delta_{1,n}$ (due to the different samples) and $\Delta_{2,n}$ (due to the different quantile levels), i.e. $\Delta_n = \Delta_{1,n} + \Delta_{2,n}$ holds, where

$$\Delta_{1,n} = \frac{1}{JL} \sum_{\ell=1}^{L} \sum_{j=1}^{J} \left(\log \widehat{q}^{\star}_{j,n}(1-p_{\ell,n}) - \frac{1}{J} \sum_{j=1}^{J} \log \widehat{q}^{\star}_{j,n}(1-p_{\ell,n}) \right)^2,$$

$$\Delta_{2,n} = \frac{1}{L} \sum_{\ell=1}^{L} \left(\frac{1}{J} \sum_{j=1}^{J} \log \widehat{q}^{\star}_{j,n}(1-p_{\ell,n}) - \mu_{p,n} \right)^2.$$

Inspired by the ANOVA, we define the ANOVEX test statistic

$$T_n = \frac{J \mathrm{varlog}(\tau_{1:L})\, k}{S_n(k, \tau_{1:L})} \frac{\Delta_{1,n}}{\Delta_{2,n}},$$

where

$$\mathrm{varlog}(\tau_{1:L}) = \frac{1}{L} \sum_{\ell=1}^{L} \log^2(\tau_\ell) - \left(\frac{1}{L} \sum_{\ell=1}^{L} \log(\tau_\ell) \right)^2 \text{ and } S_n(k, \tau_{1:L}) = \frac{1}{L} \sum_{\ell=1}^{L} \log^2 \left(\frac{k}{\tau_\ell} \right).$$

If all the random variables are heavy-tailed, then the limit distribution of T_n is obtained in the following result.

Theorem 1. *Suppose that F_j fulfills condition $\mathcal{C}_2(\gamma_j, \rho_j, A_j)$ and $\mathcal{K}(0)$ for each $j = 1, \ldots, J$. Then, under (H_0), $T_n \xrightarrow{\mathrm{d}} \chi^2_{J-1}$ as $n \to \infty$.*

The ANOVEX test thus rejects (H_0) with asymptotic level $\alpha \in (0,1)$ if $T_n > \chi^2_{J-1,1-\alpha}$, where $\chi^2_{J-1,1-\alpha}$ denotes the quantile of level $1-\alpha$ of the chi-square distribution with $J-1$ degrees of freedom.

A natural question is obviously the power of such a test. For that purpose, we propose to consider three situations (all with $J = 2$ samples) where an approximation of the error rate is available. The first one is the case where the two samples are identically Pareto distributed. In such a case, the ANOVEX test is supposed to be rejected with a rate (or Type I error) α.

Proposition 1. *Consider two independent samples $E_j = \{X_1^{(j)}, \ldots, X_n^{(j)}\}$, $j = 1, 2$, of i.i.d. variables following the same Pareto distribution (denoted $\mathcal{P}(1/\gamma)$ in the sequel) with cumulative survival function (c.s.f.) $x^{-1/\gamma}$, $\gamma > 0$, $x > 1$. Assume that k is an intermediate sequence such that $(k \log(n))/n \to 0$ as $n \to 0$. Then, as $n \to \infty$,*

$$T_n \overset{\mathrm{d}}{=} \Gamma^2 \left(1 + \frac{1}{S_n(k, \tau_{1:L})} \right) \left(1 + \mathrm{O}_{\mathbb{P}} \left(\frac{1}{\sqrt{k}} \right) + \mathrm{O}_{\mathbb{P}} \left(\frac{k}{n \log(k)} \right) \right)$$

where Γ is a standard normal random variable.

Therefore, the probability $\mathbb{P}_{H_0}\left(T_n > \chi^2_{1,1-\alpha}\right)$ to wrongly reject (H_0) with asymptotic level $\alpha \in (0,1)$ is for large n approximately equal to

$$2\bar{\Phi}\left(\bar{\Phi}^{-1}(\alpha/2)\left(1+\frac{1}{S_n(k,\tau_{1:L})}\right)^{-1/2}\right),$$

where $\overline{\Phi}(\cdot)$ is the standard Gaussian survival function.

The second situation deals with two Pareto samples with a different scale parameter. In this context, the test is supposed to be rejected, and the error (called Type II error) is the probability to wrongly not reject (H_0). An approximation is obtained, using the following notation:

$$s_n(k,\tau_{1:L}) = \frac{1}{L}\sum_{\ell=1}^{L}\sqrt{1+\log^2\left(\frac{k}{\tau_\ell}\right)}.$$

Recall also the notation $x \vee y$ for the maximum of x and y and $u_n \ll v_n$ to denote the convergence $u_n/v_n \to 0$.

Proposition 2. *Consider two independent samples denoted by $E_1 = \{X_1^{(1)},\ldots,X_n^{(1)}\}$ and $E_2 = \{X_1^{(2)},\ldots,X_n^{(2)}\}$ where $X_i^{(1)} \sim \mathcal{P}(1/\gamma)$ and $X_i^{(2)} \overset{\mathrm{d}}{=} \lambda_n X_i^{(1)}$, $i = 1,\ldots,n$ and $\lambda_n \to 1$ as $n \to \infty$ (assumption $(H_{1,n})$). Assume that*

$$\frac{\log(k)}{k^{3/4}} \vee \sqrt{\frac{\log(k)}{n}} \ll \log(\lambda_n) \ll \frac{1}{\sqrt{k}}.$$

Then, as $n \to \infty$,

$$T_n \overset{\mathrm{d}}{=} \left(\frac{\log^2(\lambda_n)k}{2\gamma^2 S_n(k,\tau_{1:L})} - \frac{\sqrt{2k}\log(\lambda_n)s_n(k,\tau_{1:L})}{\gamma S_n(k,\tau_{1:L})}\Gamma + \frac{1+S_n(k,\tau_{1:L})}{S_n(k,\tau_{1:L})}\Gamma^2\right) \times \left(1+\mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)+\mathrm{O}_{\mathbb{P}}\left(\frac{k}{n\log(k)}\right)\right)$$

where Γ is a standard normal random variable.

Note that we assume that $\lambda_n \to 1$ (and in a sense $(H_{1,n}) \to (H_0)$) as $n \to \infty$. This situation is referred to as the contiguity concept (see *e.g.* Falk and Marohn, 1993) in the test literature. By assuming the slightly stronger condition

$$\frac{\log^2(k)}{k^{3/4}} \vee \sqrt{\frac{\log^3(k)}{n}} = \mathrm{o}(\log(\lambda_n)),$$

then the probability $\mathbb{P}_{H_{1,n}}\left(T_n \le \chi^2_{1,1-\alpha}\right)$ to (wrongly) not reject (H_0) with asymptotic level $\alpha \in (0,1)$ is, for large n, approximately equal to $\bar{\Phi}\left(\Omega_{1,n} - \sqrt{\Omega_{2,n}}\right) - \bar{\Phi}\left(\Omega_{1,n} + \sqrt{\Omega_{2,n}}\right)$, where

$$\Omega_{1,n} = \frac{\log(\lambda_n)\sqrt{k}s_n(k,\tau_{1:L})}{\sqrt{2}\gamma\left(1+S_n(k,\tau_{1:L})\right)},$$
$$\Omega_{2,n} = \frac{\log^2(\lambda_n)k}{2\gamma^2}\frac{s_n(k,\tau_{1:L})^2 - 1 - S_n(k,\tau_{1:L})}{\left(1+S_n(k,\tau_{1:L})\right)^2} + \frac{S_n(k,\tau_{1:L})}{1+S_n(k,\tau_{1:L})}\chi^2_{1,1-\alpha} > 0.$$

The third situation deals with two Pareto samples with a different shape parameter. Here also, the test is supposed to be rejected, and the Type II error is the probability to wrongly not reject (H_0). An accurate approximation of the Type II error may be obtained, using the notations:

$$\mathfrak{s}_n(k,\tau_{1:L}) = \frac{1}{L}\sum_{\ell=1}^{L}\log\left(\frac{n}{\tau_\ell}\right)\sqrt{1+\log^2\left(\frac{k}{\tau_\ell}\right)} \quad \text{and} \quad \mathrm{smlog}(n/\tau_{1:L}) = \frac{1}{L}\sum_{\ell=1}^{L}\log^2\left(\frac{n}{\tau_\ell}\right).$$

Proposition 3. *Consider two independent samples $E_1 = \{X_1^{(1)},\ldots,X_n^{(1)}\}$ and $E_2 = \{X_1^{(2)},\ldots,X_n^{(2)}\}$, where $X_i^{(1)} \sim \mathcal{P}(1/\gamma)$ and $X_i^{(2)} \overset{\mathrm{d}}{=} (X_i^{(1)})^{\theta_n}$, $i = 1,\ldots,n$ and $\theta_n \to 1$ as $n \to \infty$ (assumption $(H'_{1,n})$).*

Let us assume that (θ_n) *satisfies the same conditions as* (λ_n) *in Proposition 2. Then,*

$$T_n \overset{\mathrm{d}}{=} 2\left(\frac{k(1-\theta_n)^2}{(1+\theta_n)^2}\frac{\mathrm{smlog}(n/\tau_{1:L})}{S_n(k,\tau_{1:L})} + 2\frac{\sqrt{k}\sqrt{1+\theta_n^2}(1-\theta_n)}{(1+\theta_n)^2}\frac{\mathfrak{s}_n(k,\tau_{1:L})}{S_n(k,\tau_{1:L})}\Gamma + \frac{(1+\theta_n^2)}{(1+\theta_n)^2}\frac{1+S_n(k,\tau_{1:L})}{S_n(k,\tau_{1:L})}\Gamma^2\right)$$
$$\times\left(1+\mathrm{O}_{\mathbb{P}}\left(\frac{k}{n\log(k)}\right)+\mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)\right),\ \textit{as } n\to\infty,$$

where Γ *is a standard normal random variable.*

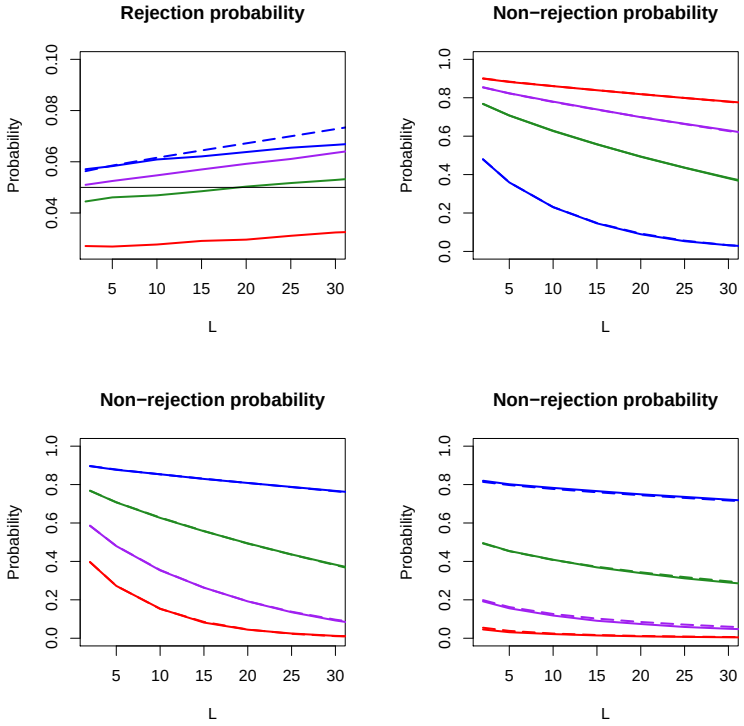


Figure 4: Empirical (solid curves) and approximated (dashed curves) Type I (top left panel) and Type II (top right and bottom panels) errors obtained for 10,000 replications with $J = 2$, $n = 1{,}000$ and $k = 100$ as functions of L. Top left panel: The two samples are identically Pareto (blue), Fréchet (purple), Burr with $\rho = -0.75$ (green) and GP (red) distributed. Top right panel: Setting of $(H_{1,n})$ with $\lambda_n = 1 + 2n^{-1/3} = 1.2$ and $\gamma = 0.15$ (blue curves), 0.25 (green curves), 0.35 (purple curves) and 0.5 (red curves). Bottom left panel: Setting of $(H_{1,n})$ with $\gamma = 0.25$ and $\lambda_n = 1.1$ (blue curves), $\lambda_n = 1.2$ (green curves), $\lambda_n = 1.3$ (purple curves) and $\lambda_n = 1.4$ (red curves). Bottom right panel: Setting of $(H'_{1,n})$ with $\gamma = 0.25$ and $\theta_n = 1 + n^{-1/3} = 1.1$ (green curves), $\theta_n = 1.2$ (brown curves), $\theta_n = 1.3$ (blue curves) and $\theta_n = 1.4$ (red curves).

Under a slightly stronger assumption on θ_n, then the probability $\mathbb{P}_{H'_{1,n}}\left(T_n \le \chi^2_{1,1-\alpha}\right)$ to (wrongly) not reject (H_0) with asymptotic level $\alpha \in (0,1)$ may be approximated, for n large enough, by $\bar{\Phi}\left(\Psi_{1,n} - \sqrt{\Psi_{2,n}}\right) - \bar{\Phi}\left(\Psi_{1,n} + \sqrt{\Psi_{2,n}}\right)$ where

$$\Psi_{1,n} = \frac{\sqrt{k}(\theta_n - 1)\mathfrak{s}_n(k,\tau_{1:L})}{\sqrt{1+\theta_n^2}\,(1+S_n(k,\tau_{1:L}))},$$
$$\Psi_{2,n} = \frac{(\theta_n-1)^2 k}{(1+\theta_n^2)}\frac{\mathfrak{s}_n(k,\tau_{1:L})^2 - (1+S_n(k,\tau_{1:L}))\,\mathrm{smlog}(n/\tau_{1:L})}{(1+S_n(k,\tau_{1:L}))^2} + \frac{(1+\theta_n)^2}{(1+\theta_n^2)}\frac{S_n(k,\tau_{1:L})}{1+S_n(k,\tau_{1:L})}\frac{\chi^2_{1,1-\alpha}}{2}$$
$$> 0.$$

The accuracy of these three approximations has been evaluated using a simulation study involving the Pareto (c.s.f. $x^{-1/\gamma}, x > 1$), Burr (c.s.f. $(1+x^{-\rho/\gamma})^{1/\rho}, \rho < 0, x > 0$), Fréchet (c.s.f. $1-\exp(-x^{-1/\gamma})$, $x > 0$) and Generalized Pareto (GP, c.s.f. $(1+\gamma x)^{-1/\gamma}$, $x > 0$) distributions. Figure 4 gives an overview of the obtained results. They show in particular a very good approximation of the considered Type II errors.

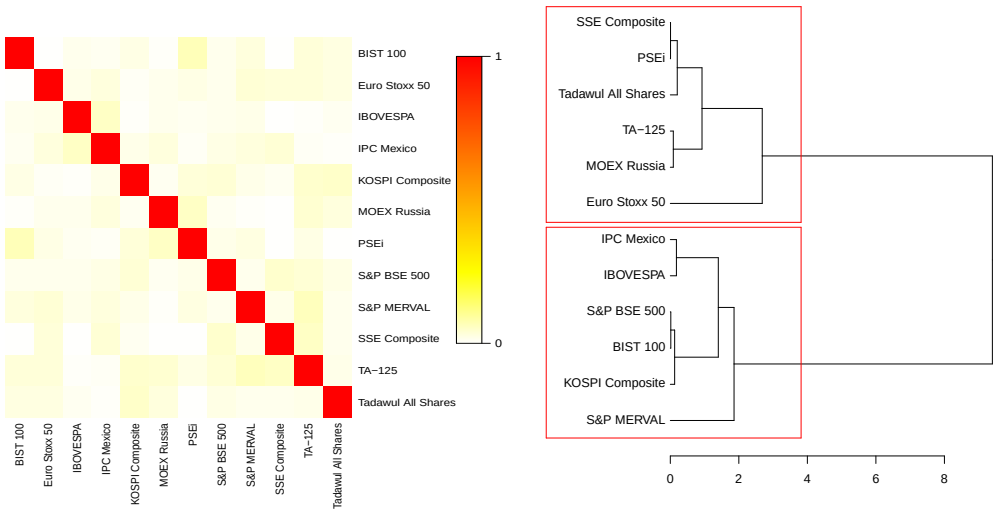


Figure 5: Multivariate stock market data. Left: Correlation matrix (absolute values) between 12 series of negative daily log-returns of market stock indices (pre-filtered *via* an ARMA(1,1)-GARCH(1,1) process). Right: Dendrogram of the market stock indices, where the dissimilarity measure is the test statistic T_n (and the linkage method is the complete-linkage).

Finally, the test statistic T_n has been applied as a tail dissimilarity measure to the daily negative log-returns of $J = 12$ stock market indices (pre-filtered using an ARMA(1,1)-GARCH(1,1) model). The independence within and between the samples may be justified (see the correlation matrix in Figure 5), and this leads to the dendrogram of Figure 5. By hierarchically testing (following the hierarchy provided by the dendrogram) the tail equalities of the samples, it appears that these 12 indices may be split into two groups:

- A first one containing the indices from Central America (IPC Mexico) and South America (IBOVESPA, Brazil and S&P MERVAL, Argentina), along with the S&P BSE 500 (India), KOSPI Composite (South Korea) and BIST 100 (Turkey) indices,
- A second one containing the other Eurasian indices (Euro Stoxx 50, MOEX, Russia and TA-125, Israel), along with the SSE Composite (China), PSEi (Philippines) and Tadawul All Shares (Saudi Arabia).

It appears that these two groups cannot be merged, since the corresponding test is rejected. In such a situation, the data cannot be pooled in a unique sample, and a multivariate approach to extreme risk is mandatory, motivating the focus of the next section. We also refer to Girard et al. (2024) for an application to the analysis of daily precipitation in Germany.

5 Improving extreme risk estimation through simulation of multivariate extremes

As mentioned in Section 1, one of the main challenges in attempting to quantify risk measures at a high level is the inherent scarcity of available extreme events. Another consideration is that univariate risk measures can be improved by taking into account the potential asymptotic dependence structure exhibited by the risk variable of interest X, with other covariates. This section presents an alternative approach to the one employed in Section 2 for the estimation of risk measures, focusing exclusively on empirical estimators. In order to address the two aforementioned issues, *i.e.* the scarcity of data at extreme levels and the presence of extremal dependence structure between variables, we develop a stochastic generator of multivariate extreme events (see Legrand et al., 2023; Madhar et al., 2024).

Three risk measures are considered. For a multivariate random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$, the CTE at level τ of a target risk variable X_j, $j \in \{1, \ldots, d\}$, has been defined in Equation (5) and is the first risk measure of interest. However, this risk measure does not take into account the potential asymptotic dependence between X_j and the other components of $\boldsymbol{X}$. To address this specific aspect,

we consider, first, the Multivariate Marginal Expected Shortfall (MMES) of X_j at level τ, that is,

$$\mathrm{MMES}_j(\tau) = \mathbb{E}(X_j \mid \boldsymbol{X}_{-j} \geq \boldsymbol{v}^{\tau}_{-j}) \tag{15}$$

where $\boldsymbol{v}^{\tau}$ is the vector made of the marginal quantiles of X at level τ, and with the notation $\boldsymbol{v}_{-j}$ denoting the vector $\boldsymbol{v}$ deprived from its jth component. Through this formulation the aim is to capture the behavior of X_j when other components of $\boldsymbol{X}$ reach extreme levels. The third risk measure we consider can be used to quantify the risk of X_j when other components of $\boldsymbol{X}$ as well as X_j itself reach extreme levels simultaneously. Its definition is an extension of the Dependent Conditional Tail Expectation (DCTE) to the multivariate case:

$$\mathrm{DCTE}_j(\tau) = \mathbb{E}(X_j \mid \boldsymbol{X} \geq \boldsymbol{v}^{\tau}) \tag{16}$$

For the above mentioned risk measures, parametric and nonparametric estimators have been proposed but these are often limited to the bivariate case. In this study, we rather consider empirical estimation of the CTE, the MMES and the DCTE in a multivariate setting, *i.e.* for $d \geq 2$. In order to overcome the limitations of the amount of available data for carrying out empirical estimation, we develop a joint simulation algorithm of multivariate extreme events.

The simulation procedure is based on the stochastic representation of a standard multivariate generalized Pareto (MGP) vector. Assuming that the original data $\boldsymbol{X}$ can be marginally transformed to the exponential scale (using the probability integral transform) and denoting by $\boldsymbol{X}^E$ the transformed data, one can consider the conditional distribution

$$[\boldsymbol{X}^E - \boldsymbol{u} \mid \boldsymbol{X}^E \not\leq \boldsymbol{u}] \tag{17}$$

of the vector of exceedances, where $\boldsymbol{X}^E \not\leq \boldsymbol{0}$ means that at least one of the component of $\boldsymbol{X}^E$ is positive. Then, for $\boldsymbol{u}$ large enough, the conditional distribution of the vector of exceedances defined in (17) can be approximated by a vector $\boldsymbol{Z}$ having an MGP distribution (see Rootzén and Tajvidi, 2006). Rootzén et al. (2018) have shown that such vectors $\boldsymbol{Z}$, called standard MGP vectors, can be decomposed as

$$\boldsymbol{Z} = E + \boldsymbol{T} - \max(\boldsymbol{T}) \tag{18}$$

where E is a unit exponential random variable and $\boldsymbol{T}$ is a random vector independent of E. Parametric forms have been proposed for $\boldsymbol{T}$ (see, *e.g.*, Kiriliouk et al., 2019). We propose here a distribution-free stochastic generator of MGP vectors $\boldsymbol{Z}$ using nonparametric bootstrap.

The main idea of this generator is to consider the multivariate differences

$$\Delta^{j,k} = Z_j - Z_k = T_j - T_k, \text{ for all } j,k \in \{1,\ldots,d\}.$$

From these differences, Equation (18) can then be rewritten as

$$Z_j = E + \sum_{k=1,k\neq j}^{d} \Delta^{j,k} \prod_{\ell=1,\ell\neq k}^{d} \mathbb{1}_{\{\Delta^{\ell,k}<0\}}, \text{ for all } j \in \{1,\ldots,d\}. \tag{19}$$

It is then sufficient to generate realizations of the differences $\Delta^{j,k}$ for all $j,k \in \{1,\ldots,d\}$, and since $\Delta^{j,k} = \Delta^{1,k} - \Delta^{1,j}$ for all $j,k \in \{1,\ldots,d\}$, it is sufficient to generate realizations of the differences $\Delta^{1,k}$ for all $k \in \{2,\ldots,d\}$. Denoting by $\boldsymbol{\Delta} = (\Delta^{1,1},\ldots,\Delta^{1,d}) \in \mathbb{R}^d$, the algorithm generating MGP vectors is then simply based on a nonparametric resampling of the observed differences $\boldsymbol{\Delta}$, see Algorithm 1. Proposition 7.1 in (Madhar et al., 2024) guarantees that the samples simulated with Algorithm 1 are indeed distributed according to a standard MGP distribution.

Numerical experiments can be found in Legrand et al. (2023, Section 4.2) and in Madhar et al. (2024, Section 3.2). We illustrate the benefits of using this resampling algorithm using simulation evidence when the (3-dimensional) random vector $\boldsymbol{X} = (X_1, X_2, X_3)$ has marginal Student t-distribution with degrees of freedom $\nu = (2, 3, 2.5)$. This distribution is chosen so as to mimic financial returns, which are typically heavy-tailed. To ensure that the components of $\boldsymbol{X}$ are upper tail dependent, we generate them using a Gumbel copula with parameter $\theta \geq 1$ (see, e.g., Nelsen, 2006). Simply stated, the larger θ, the stronger the asymptotic dependence structure between the components of $\boldsymbol{X}$. For the sake of simplicity, we only consider the risk measures associated to the risk variable X_1. Then, we empirically estimate $\mathrm{CTE}(1-p)$, $\mathrm{MMES}_1(1-p)$ and $\mathrm{DCTE}_1(1-p)$ (see Equations (5), (15) and (16)) for different (small) values of level p but also different values of θ, corresponding to the degree of asymptotic dependence between the components of $\boldsymbol{X}$. Since benchmark values for each risk measure can be analytically computed, we can consider the relative

Algorithm 1 Nonparametric joint MGP simulation

Input: Observations $(\boldsymbol{Z}_i)_{1\leq i\leq n} = (Z_{i,1},\ldots,Z_{i,d})_{1\leq i\leq n}$ from a standard MGP distribution

Output: A standard MGP simulated sample $\left(\widetilde{\boldsymbol{Z}}_m\right)_{1\leq \ell\leq m} = \left(\widetilde{Z}_{\ell,1},\ldots,\widetilde{Z}_{\ell,d}\right)_{1\leq \ell\leq m}$

procedure

  Compute $\Delta_i^{1,k} \leftarrow Z_{i,1} - Z_{i,k}$, for $1\leq i\leq n$ and $1\leq k\leq d$

  Generate $E_1,\ldots,E_m \overset{\text{i.i.d.}}{\sim} \text{Exp}(1)$, and independent of $(\Delta_i^{1,k})_{1\leq i\leq n,1\leq k\leq d}$

  Generate a m-bootstrap sample $\widetilde{\boldsymbol{\Delta}} = \left(\widetilde{\boldsymbol{\Delta}}_\ell\right)_{\ell=1,\ldots,m}$ from $(\boldsymbol{\Delta}_i)_{1\leq i\leq n}$

  Compute $\widetilde{\Delta}_\ell^{r,s} \leftarrow \widetilde{\Delta}_\ell^{1,s} - \widetilde{\Delta}_\ell^{1,r}$, for $1\leq \ell\leq m$ and all $1\leq r,s\leq d$

  Compute $\widetilde{Z}_{\ell,j} \leftarrow E_\ell + \sum_{s=1,s\neq j}^d \widetilde{\Delta}_\ell^{j,s} \prod_{r=1,r\neq s}^d \mathbb{1}_{\{\widetilde{\Delta}_\ell^{r,s}<0\}}$ for all $1\leq \ell\leq m$ and $1\leq j\leq d$

end procedure

errors between the empirical estimates and the theoretical values. These are estimated using, first, the original sample without data augmentation (denoted by *Orig.*), then the data generated using Algorithm 1 only (denoted by *Simu.*) and the original sample and the simulated sample combined (denoted by *Ext.*).

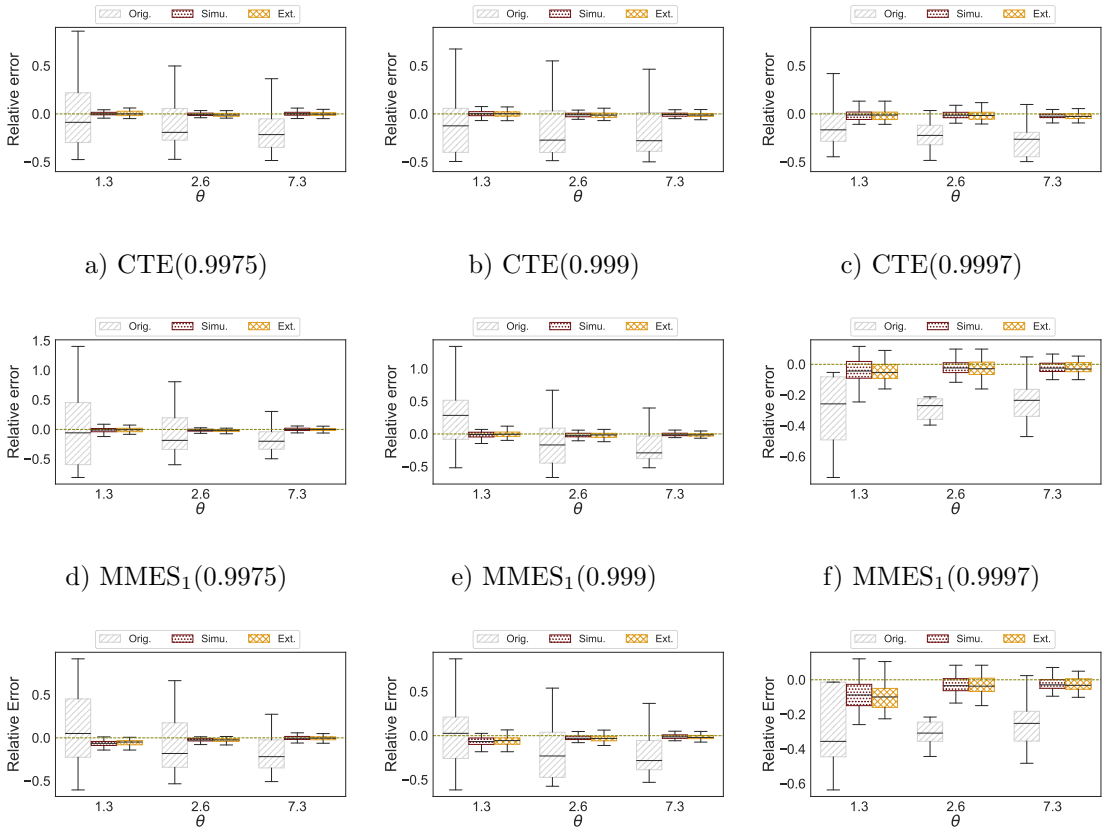
Figure 6 illustrates the distribution of the relative errors in each of the aforementioned cases considering 50 original simulated samples. It is noteworthy that the simulation of multivariate extremes using Algorithm 1 improves not only the estimation of the MMES and the DCTE but also the estimation of the CTE which is defined solely in terms of the variable X_1. This suggests that the estimation of univariate risk measures may be enhanced by the consideration of additional relevant variables (in particular, variables exhibiting asymptotic dependence with the risk variable of interest).

6 Discussion

As Section 5 illustrates, there is substantial potential in combining empirical estimation methods with resampling algorithms or generative methods based on neural networks (Allouche et al., 2022; Huster et al., 2021) in order to improve estimation accuracy. It would be interesting to pair the nonparametric stochastic generator of multivariate extreme data points with bias-reduced estimators in order to test for tail homogeneity, as a preliminary step before carrying out (or not) extremal regression. Another interesting avenue for research would be the adaptation of the resampling methodology to the extremal regression context, where finite-sample uncertainty in the estimation is often sizeable, in order to produce estimators that are much more accurate than what is currently available without having to resort to strong model specifications such as linear quantile regression models. It would, finally, be important to adapt the methods presented in this article to the presence of serial dependence so as to be able to handle real data applications in risk management such as the tornado losses problem, or the tail homogeneity assessment exercise for financial data without having to rely on the ARMA-GARCH filtering step (whose effect on the time dependence structure is not guaranteed should this model be incorrect).

Acknowledgements

The authors gratefully acknowledge support from the French National Research Agency under the grant ANR-23-CE40-0009 (EXSTA project). Gilles Stupfler acknowledges further financial support from the French National Research Agency under the grant ANR-11-LABX-0020-01 (Centre Henri Lebesgue), as well as from the TSE-HEC ACPR Chair "Regulation and systemic risks". Stéphane Girard and Gilles Stupfler acknowledge support from the Chair "Stress Test, Risk Management and Financial Steering", led by the French Ecole Polytechnique and its Foundation and sponsored by BNP Paribas.



a) CTE(0.9975) b) CTE(0.999) c) CTE(0.9997)

d) $\text{MMES}_1(0.9975)$ e) $\text{MMES}_1(0.999)$ f) $\text{MMES}_1(0.9997)$

g) $\text{DCTE}_1(0.9975)$ h) $\text{DCTE}_1(0.999)$ i) $\text{DCTE}_1(0.9997)$

Figure 6: Distribution of relative approximation errors of the estimations of risk measures on 50 original samples *(grey oblique lines)*, 50 simulated samples *(red dots)* and 50 extended samples *(yellow grid)* for the CTE (Figures a), b), c)), the MMES_1 (Figures d), e), f)) and the DCTE_1 (Figures g), h), i)) at varying levels $\tau = 1 - p \in \{0.9975, 0.999, 0.9997\}$ with respect to copula parameter $\theta \in \{1.3, 2.6, 7.3\}$.

References

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

Allouche, M., El Methni, J., and Girard, S. (2023). A refined Weissman estimator for extreme quantiles. *Extremes*, 26(3):545–572.

Allouche, M., El Methni, J., and Girard, S. (2024a). Reduced-bias estimation of the extreme conditional tail expectation for Box–Cox transforms of heavy-tailed distributions. *Journal of Statistical Planning and Inference*, 233:106189.

Allouche, M., Girard, S., and Gobet, E. (2022). EV-GAN: Simulation of extreme events with ReLU neural networks. *Journal of Machine Learning Research*, 23(150):1–39.

Allouche, M., Girard, S., and Gobet, E. (2024b). Learning extreme expected shortfall and conditional tail moments with neural networks. Application to cryptocurrency data. *Neural Networks*, to appear.

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

Beirlant, J., Dierckx, G., Goegebeur, Y., and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes*, 2(2):177–200.

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

Beirlant, J., Matthys, G., and Dierckx, G. (2001). Heavy-tailed distributions and rating. *ASTIN Bulletin: The Journal of the IAA*, 31(1):37–58.

Bousquet, N. and Bernardara, P. (2021). *Extreme Value Theory with Applications to Natural Hazards*. Springer Cham.

Box, G. and Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B*, 26(2):211–243.

Brazauskas, V., Jones, B., Puri, M., and Zitikis, R. (2008). Estimating conditional tail expectation with actuarial applications in view. *Journal of Statistical Planning and Inference*, 138(11):3590–3604.

Buitendag, S., Beirlant, J., and de Wet, T. (2020). Confidence intervals for extreme Pareto-type quantiles. *Scandinavian Journal of Statistics*, 47(1):36–55.

Caeiro, F., Gomes, M., and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *REVSTAT-Statistical Journal*, 3(2):113–136.

Cai, J.-J., Einmahl, J. H. J., de Haan, L., and Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *Journal of the Royal Statistical Society: Series B*, 77(2):417–442.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics*, 33(2):806–839.

Chernozhukov, V. and Fernández-Val, I. (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *Review of Economic Studies*, 78(2):559–589.

Daouia, A., Gardes, L., and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19(5B):2557–2589.

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

Daouia, A., Padoan, S. A., and Stupfler, G. (2024a). Optimal weighted pooling for inference about the tail index and extreme quantiles. *Bernoulli*, 30(2):1287–1312.

Daouia, A., Stupfler, G., and Usseglio-Carleve, A. (2023). Inference for extremal regression with dependent heavy-tailed data. *The Annals of Statistics*, 51(5):2040–2066.

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

Davison, A. C., Padoan, S. A., and Stupfler, G. (2023). Tail risk inference via expectiles in heavy-tailed time series. *Journal of Business and Economic Statistics*, 41(3):876–889.

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

Dematteo, A. and Clémençon, S. (2016). On tail index estimation based on multivariate data. *Journal of Nonparametric Statistics*, 28(1):152–176.

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

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

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

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

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin.

Falk, M. and Marohn, F. (1993). Asymptotically optimal tests for conditional distributions. *The Annals of Statistics*, 21(1):45–60.

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

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

Gardes, L. and Maistre, S. (2023). Nonparametric asymptotic confidence intervals for extreme quantiles. *Scandinavian Journal of Statistics*, 50(2):825–841.

Girard, S., Opitz, T., and Usseglio-Carleve, A. (2024). ANOVEX: ANalysis Of Variability for heavy-tailed EXtremes. *Electronic Journal of Statistics*, to appear.

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

Hall, P. and Welsh, A. (1985). Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13(1):331–341.

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

Hoga, Y. (2018). Detecting tail risk differences in multivariate time series. *Journal of Time Series Analysis*, 39(5):665–689.

Huster, T., Cohen, J. E. J., Lin, Z., Chan, K., Kamhoua, C., Leslie, N., Chiang, C. Y. J., and Sekar, V. (2021). Pareto GAN: Extending the representational power of GANs to heavy-tailed distributions. In *International Conference on Machine Learning*, pages 4523–4532. PMLR.

Kinsvater, P., Fried, R., and Lilienthal, J. (2016). Regional extreme value index estimation and a test of tail homogeneity. *Environmetrics*, 27(2):103–115.

Kiriliouk, A., Rootzén, H., Segers, J., and Wadsworth, J. L. (2019). Peaks over thresholds modeling with multivariate generalized Pareto distributions. *Technometrics*, 61(1):123–135.

Legrand, J., Ailliot, P., Naveau, P., and Raillard, N. (2023). Joint stochastic simulation of extreme coastal and offshore significant wave heights. *The Annals of Applied Statistics*, 17(4):3363–3383.

Linton, O. and Xiao, Z. (2013). Estimation of and inference about the Expected Shortfall for time series with infinite variance. *Econometric Theory*, 29(4):771–807.

Madhar, N., Legrand, J., and Thomas, M. (2024). Assessing extreme risk using stochastic simulation of extremes. *https://arxiv.org/abs/2406.08019*.

Manjunath, B., Caeiro, F., Gomes, M., and Alves, M. (2013). evt0: Mean of order p, peaks over random threshold, Hill and high quantile estimates. *R package version*.

Mougeot, M. and Tribouley, K. (2010). Procedure of test to compare the tail indices. *Annals of the Institute of Statistical Mathematics*, 62(2):383–412.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer.

Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer.

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

Rootzén, H., Segers, J., and Wadsworth, J. L. (2018). Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131.

Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 12(5):917–930.

Scheffé, H. (1999). *The analysis of variance*, volume 72. John Wiley & Sons.

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

Valdez, E. (2005). Tail conditional variance for elliptically contoured distributions. *Belgian Actuarial Bulletin*, 5(1):26–36.

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

Worms, J. and Worms, R. (2015). A test for comparing tail indices for heavy-tailed distributions via empirical likelihood. *Communications in Statistics-Theory and Methods*, 44(15):3289–3302.