Residual-based extreme value estimation for heavy tails

Marco Oesting, Gilles Stupfler

Residual-based extreme value estimation for heavy tails

Gilles Stupfler[a] & Marco Oesting[b]

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

[b] Stuttgart Center for Simulation Science (SC SimTech) & Institute for Stochastics and Applications, University of Stuttgart, D-70563 Stuttgart, Germany

Abstract

Motivated by the use of location-scale regression models to support several recent extremal regression methods, we consider the estimation of, and inference about, conditional extreme quantiles and related quantities based on standardized residuals obtained following a preliminary model estimation step. We show that residual-based versions of extreme value estimators are asymptotically normal, just as their unachievable counterparts based on unobserved regression errors would be, under a high-level condition which essentially requires the number of residuals that do not correctly predict the unobserved errors to be asymptotically smaller than the multiplicative inverse of the rate of convergence of the extreme value procedure. This condition is shown to be substantially weaker than corresponding conditions obtained in related recent work, and we discuss how our theory applies to a wide range of examples containing location-scale parametric regression models and autoregressive time series.

MSC 2020 subject classifications: 62G08, 62G30, 62G32

Keywords: Extreme quantile estimation, heavy tails, location-scale regression, residuals, tail dependence

1 Introduction

1.1 Background

Extremal regression is used in a variety of applications of statistics as a way to model the right tail of a univariate random variable $Y \in \mathbb{R}$ given valuable information represented by a covariate $\boldsymbol{X}$. One may, for example, be interested in forecasting potential high financial losses given global market performance (Chernozhukov et al., 2018), understanding wage inequalities given socioeconomic factors (D'Haultfoeuille et al., 2018), or estimating extreme levels of a climate variable given auxiliary weather information (Russell et al., 2016). This is typically done by estimating extreme quantiles of Y given $\boldsymbol{X} = \boldsymbol{x}$ at a level τ close to 1, which will be denoted in the sequel as $q_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x})$. The literature has mainly attacked the extreme quantile regression problem using two broad classes of solutions: on the one hand, techniques based on linear models and their natural extensions, such as the generalized linear models of Davison and Smith (1990) and Chavez-Demoulin and Davison (2005) or extremal linear quantile regression estimators, of which a nice review is Chernozhukov et al. (2018). On the other hand, one can construct nonparametric extremal regression estimators (Daouia et al., 2011, 2013,

2023), and pair them with methods such as random forests using the negative Generalized Pareto log-likelihood as loss function (Farkas et al., 2021, 2024; Gnecco et al., 2024) in order to cope with the curse of dimensionality. Statistical theory for the former approach is only well-developed in extremal linear quantile regression models; nonparametric theory tends to be better established, but the validity of nonparametric asymptotic inference has, to the best of our knowledge, only been checked by Daouia et al. (2023) for the extreme kernel quantile regression estimator in low-dimensional settings.

A third approach, initiated by McNeil and Frey (2000) and recently adopted in several papers focusing on the estimation of location-scale equivariant functionals (Martins-Filho et al., 2018; Hoga, 2019; Ahmad et al., 2020; Girard et al., 2021; Davison et al., 2023), is to work in heteroskedastic regression models of the form $Y = m(\boldsymbol{X}) + \sigma(\boldsymbol{X})\varepsilon$. Here m and $\sigma > 0$ are two measurable functions of a covariate $\boldsymbol{X} \in \mathbb{R}^p$, and the unobserved error ε is independent of $\boldsymbol{X}$. This ensures that $q_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x}) = m(\boldsymbol{x}) + \sigma(\boldsymbol{x})q(\tau)$, where $q(\tau) = q_\varepsilon(\tau)$ denotes the quantile of ε at level τ. Given data $(Y_i, \boldsymbol{X}_i)$, $1 \leq i \leq n$ from the model, the commonly adopted point of view to estimate $q_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x})$ is that the estimation of the location and scale functions by $\widehat{m}_n$ and $\widehat{\sigma}_n$, say (whose exact construction depends on the model considered), is a preliminary step before estimating unconditional extreme quantiles $q(\tau)$ of ε by plugging the residuals $\widehat{\varepsilon}_i^{(n)} = (Y_i - \widehat{m}_n(\boldsymbol{X}_i))/\widehat{\sigma}_n(\boldsymbol{X}_i)$ instead of the unobserved errors $\varepsilon_i = (Y_i - m(\boldsymbol{X}_i))/\sigma(\boldsymbol{X}_i)$ into an existing extreme quantile estimator (see Chapter 4 in de Haan and Ferreira, 2006, for a review of such estimators). The key intuition of McNeil and Frey (2000) is that since the whole sample of data can be used to estimate m and σ, while the extreme quantile estimator requires the use of a much smaller sample of the largest observations, the extreme quantile estimator based on residuals should have the same asymptotic behavior as its counterpart based on the unobservable errors. In doing so, one may carry out extremal regression in models whose dimension p is not low, while obtaining rates of convergence and asymptotic distributions reminiscent of i.i.d. theory.

The crucial point is that these residuals will typically not be independent or identically distributed even if the unobserved errors are. An interesting methodological problem is therefore to prove that residual-based versions of extreme value estimators indeed share the asymptotic behavior of their unachievable counterparts based on i.i.d. data. This has been tackled by Girard et al. (2021), who construct a Gaussian approximation to the tail process of residuals when both the uniform absolute and relative differences between the residuals and the corresponding errors converge to 0 at an appropriate rate depending on the quantile level. However, this turns out to impose undesirable restrictions in heavy-tailed autoregressive time series models where the covariate contains lags of the response variable.

1.2 Contribution of the paper

Assume for now that the available data is a sample $(\widehat{\varepsilon}_1^{(n)}, \ldots, \widehat{\varepsilon}_n^{(n)})$, viewed as approximating a sample of unobserved i.i.d. and heavy-tailed random variables $(\varepsilon_1, \ldots, \varepsilon_n)$. The random variables $\widehat{\varepsilon}_1^{(n)}, \ldots, \widehat{\varepsilon}_n^{(n)}$ are not assumed to be i.i.d. and may change with n. We call them *residuals*, by analogy with the motivating regression problem, and we consider the estimation of extreme value quantities linked to the distribution of $\varepsilon \equiv \varepsilon_1$ when this random variable is heavy-tailed (*i.e.* Paretian-tailed).

Our contribution is to show that one can substantially weaken the assumptions of Girard et al. (2021) and move away from uniform-type conditions. This is done by taking a different route to the proof

of the asymptotic normality of standard extreme value estimators such as the Hill estimator of the extreme value index (Hill, 1975) and Weissman estimator of extreme quantiles (Weissman, 1978) that relies neither on the validity of the Rényi representation of transformed high order statistics nor on a Gaussian approximation to the tail empirical process. We instead exploit the fact that the asymptotic behavior of intermediate quantile estimators is closely linked to that of the corresponding empirical distribution function, and that the Hill estimator, being a weighted integral of the tail empirical process, is nothing but another integral over the tail of this same empirical distribution function. This reduces our problem to the problem of finding appropriate conditions under which the tail of the empirical distribution function of the residuals is sufficiently close to that of $(\varepsilon_1, \ldots, \varepsilon_n)$, which prompts us to introducing a class of conditions that, roughly speaking, count the number of residuals that do not correctly predict the unobserved errors and impose that this number is smaller than the multiplicative inverse of the rate of convergence of the extreme value procedure under consideration if it were based on the ideal sample of data $(\varepsilon_1, \ldots, \varepsilon_n)$. We demonstrate how this leads to a theoretical framework that can handle a wider range of regression and time series models without having to impose unnecessary technical restrictions as in Girard et al. (2021), while being able to handle the joint estimation of not only intermediate and extreme quantiles, but also weighted integrals of tail quantiles, including for example the Expected Shortfall, in a multivariate setting.

The outline of the remainder of this article is the following. In Section 2, we provide a detailed explanation of our setting and main assumptions. Our main results are stated in Section 3. We discuss how our high-level conditions compare with those of earlier literature and several examples of models that we cover in Section 4. The proofs of our results are deferred to an Appendix.

2 Statistical framework

2.1 Residual-based estimation of extreme quantiles

Let, for $1 \leq j \leq d$, the $n_j = n_j(n)$ be sequences of integers tending to infinity as $n \to \infty$. The available data is assumed to be made of samples $(\widehat{\varepsilon}_{1,j}^{(n_j)}, \ldots, \widehat{\varepsilon}_{n_j,j}^{(n_j)})$, viewed as approximating samples of unobserved i.i.d. random variables $(\varepsilon_{1,j}, \ldots, \varepsilon_{n_j,j})$ in a sense that will be made clear in our theoretical results below. The random variables $\widehat{\varepsilon}_{1,j}^{(n_j)}, \ldots, \widehat{\varepsilon}_{n_j,j}^{(n_j)}$ are not assumed to be i.i.d. and may change with n_j. They will be called *residuals*, by analogy with the motivating regression problem whereby a d-dimensional response variable $\boldsymbol{Y}_t = (Y_{t,1}, \ldots, Y_{t,d})^\top$ is assumed to satisfy the system of equations

$$Y_{t,j} = m_j(\boldsymbol{X}_{t,j}) + \sigma_j(\boldsymbol{X}_{t,j})\varepsilon_{t,j}, \ 1 \leq j \leq d, \ t \in \mathbb{Z},$$

where, for each $j \in \{1, \ldots, d\}$, $(\boldsymbol{X}_{t,j})_{t\in\mathbb{Z}}$ is an $\mathbb{R}^p$-valued time series such that for any t, $\boldsymbol{X}_{t,j}$ is independent of the error $\varepsilon_{t,j}$, and the measurable functions $m_j : \mathbb{R}^p \to \mathbb{R}$ and $\sigma_j : \mathbb{R}^p \to (0, \infty)$ are unknown. In this regression model, one would typically construct $\widehat{\varepsilon}_{t,j}^{(n_j)} = (Y_{t,j} - \widehat{m}_{n,j}(\boldsymbol{X}_{t,j}))/\widehat{\sigma}_{n,j}(\boldsymbol{X}_{t,j})$ following the estimation of m_j and σ_j by $\widehat{m}_{n,j}$ and $\widehat{\sigma}_{n,j}$, respectively. Our general construction is of course relevant to location-scale regression analysis, but also to other problems, such as any context where the data points are recorded subject to measurement error or corrupted. We allow for different

sample sizes n_j so as to be able to consider the situation, common in economics, environmental science and finance, where the d samples of data have not been recorded over the same time period.

Let $F_j : t \mapsto \mathbb{P}(\varepsilon_j \leq t)$ and $\overline{F}_j = 1 - F_j : t \mapsto \mathbb{P}(\varepsilon_j > t)$ respectively denote the distribution and survival functions of $\varepsilon_j \equiv \varepsilon_{1,j}$, and $q_j : \tau \mapsto \inf\{t \in \mathbb{R} \,|\, F_j(t) \geq \tau\}$ be the quantile of ε_j at level $\tau \in (0,1)$. We assume that each one of the ε_j has a heavy-tailed (or Paretian-tailed) distribution with extreme value index $\gamma_j > 0$, in the sense that $\overline{F}_j(tx)/\overline{F}_j(t) \to x^{-1/\gamma_j}$ as $t \to \infty$ for any $x > 0$. This assumption is generally appropriate to the modelling of actuarial and financial data, see *e.g.* p.9 in Embrechts et al. (1997) and p.1 in Resnick (2007), and is also often used in environmental science, for instance when modeling extreme return levels of rainfall and river flows (Kinsvater and Fried, 2017). The basic question in extreme value analysis is to estimate extreme quantiles of the ε_j, that is, quantiles of the form $q_j(1 - p_{n,j})$, with an exceedance probability $p_{n,j} \uparrow 1$ such that $n_j(1 - p_{n,j})$ is bounded. If the $\varepsilon_{t,j}$ were observed, then one could use the classical Weissman extreme quantile estimator (Weissman, 1978), whose construction we briefly recall now. Let $\varepsilon_{1:n_j,j} \leq \varepsilon_{2:n_j,j} \leq \cdots \leq \varepsilon_{n_j:n_j,j}$ be the order statistics of the sample $(\varepsilon_{1,j}, \ldots, \varepsilon_{n_j,j})$, and let $k_j = k_j(n) \to \infty$ with $k_j/n_j \to 0$ be an intermediate sequence of integers. The Weissman extreme quantile estimator is

$$\widetilde{q}^{\star}_{n,j}(1 - p_{n,j}) = \left(\frac{k_j}{n_j p_{n,j}}\right)^{\widetilde{\gamma}_{n,j}} \widetilde{q}_{n,j}(1 - k_j/n_j)$$

where $\widetilde{q}_{n,j}(1 - k_j/n_j) = \varepsilon_{n_j - k_j : n_j, j}$ is the empirical quantile estimator of $q_j(1 - k_j/n_j)$ and $\widetilde{\gamma}_{n,j}$ is any consistent estimator of γ_j, such as the Hill estimator (Hill, 1975), that is,

$$\widetilde{\gamma}_{n,j} \equiv \widetilde{\gamma}_{n,j}(k_j) = \frac{1}{k_j} \sum_{i=1}^{k_j} \log \varepsilon_{n_j - i + 1 : n_j, j} - \log \varepsilon_{n_j - k_j : n_j, j} = \frac{n_j}{k_j} \int_{\varepsilon_{n_j - k_j : n_j, j}}^{\infty} \widetilde{\overline{F}}_{n,j}(t) \frac{\mathrm{d}t}{t},$$

where $\widetilde{\overline{F}}_{n,j}$ denotes the empirical survival function constructed upon the $\varepsilon_{t,j}$. The Hill estimator is also the maximum likelihood estimator in a purely Pareto model and is arguably the most popular semiparametric estimator in the analysis of heavy tails; the Hill and Weissman estimators are known to be asymptotically normal under classical conditions, see Theorem 3.2.5 p.74 and Theorem 4.3.8 p.138 in de Haan and Ferreira (2006). The premise of this paper is that the $\varepsilon_{t,j}$ are not available but the residuals $\widehat{\varepsilon}^{(n_j)}_{t,j}$ are, which suggests to introduce their order statistics $\widehat{\varepsilon}^{(n_j)}_{1:n_j,j} \leq \widehat{\varepsilon}^{(n_j)}_{2:n_j,j} \leq \cdots \leq \widehat{\varepsilon}^{(n_j)}_{n_j:n_j,j}$ and to use instead the residual-based estimators

$$\widehat{q}^{\star}_{n,j}(1 - p_{n,j}) = \left(\frac{k_j}{n_j p_{n,j}}\right)^{\widehat{\gamma}_{n,j}} \widehat{q}_{n,j}(1 - k_j/n_j), \quad \text{with} \quad \widehat{q}_{n,j}(1 - k_j/n_j) = \widehat{\varepsilon}^{(n_j)}_{n_j - k_j : n_j, j}$$

$$\text{and} \quad \widehat{\gamma}_{n,j} \equiv \widehat{\gamma}_{n,j}(k_j) = \frac{1}{k_j} \sum_{i=1}^{k_j} \log \widehat{\varepsilon}^{(n_j)}_{n_j - i + 1 : n_j, j} - \log \widehat{\varepsilon}^{(n_j)}_{n_j - k_j : n_j, j} = \frac{n_j}{k_j} \int_{\widehat{\varepsilon}^{(n_j)}_{n_j - i + 1 : n_j, j}}^{\infty} \widehat{\overline{F}}_{n,j}(t) \frac{\mathrm{d}t}{t}.$$

Here $\widehat{\overline{F}}_{n,j}$ denotes the empirical survival function constructed upon the residuals $\widehat{\varepsilon}^{(n_j)}_{t,j}$. Our objective is to analyze the asymptotic properties of $\widehat{q}^{\star}_{n,j}(1 - p_{n,j})$ and $\widehat{\gamma}_{n,j}$ by relating them to those of the unfeasible estimators $\widetilde{q}^{\star}_{n,j}(1 - p_{n,j})$ and $\widetilde{\gamma}_{n,j}$. This will more generally push us to discuss the joint

estimation of weighted integrals of tail quantiles, of the form

$$\frac{n_j}{k_j}\int_1^\infty \overline{F}_j(xq_j(1-k_j/n_j))\mu_j(\mathrm{d}x)$$

where the μ_j are weighting measures. This contains the Hill estimator, of course, but also the so-called Expected Shortfall risk measure at extreme levels. This will be done by comparing the estimators

$$\frac{n_j}{k_j}\int_1^\infty \widehat{\overline{F}}_{n,j}(x\widehat{\varepsilon}^{(n_j)}_{n_j-i+1:n_j,j})\mu_j(\mathrm{d}x) \text{ with the unfeasible } \frac{n_j}{k_j}\int_1^\infty \widetilde{\overline{F}}_{n,j}(x\varepsilon_{n_j-i+1:n_j,j})\mu_j(\mathrm{d}x).$$

We now explain the main modeling conditions that we shall require for doing so.

2.2 Extreme value model

Let $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_d)^\top$. There are two main modeling assumptions in our work. The first one, introduced in order to deal with the marginal convergence of each extreme value estimator, is the following second-order regular variation condition on the $\overline{F}_j$, under which the right tail of the distribution of each ε_j is assumed to be close to pure Pareto tails at a rate which is precisely quantified.

$\mathcal{C}_2(\boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{A})$ For any $j \in \{1, \ldots, d\}$, the marginal distribution of ε_j satisfies the second-order regular variation condition $\mathcal{C}_2(\gamma_j, \rho_j, A_j)$, namely: there are $\gamma_j > 0$, a second-order parameter $\rho_j \leq 0$ and an auxiliary function A_j having constant sign and converging to 0 at infinity such that

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

This classical condition is satisfied in any of the reasonable statistical models featuring heavy tails, such as the Fréchet, Student and Inverse-Gamma models; see Appendix C in Daouia et al. (2024b) for a table of standard distributions satisfying this assumption.

The second condition we require is a pairwise tail dependence condition specifically geared towards the obtention of joint asymptotic results about extreme value estimators. Recall that for any (j, ℓ) there exists, by Sklar's theorem, a (possibly not unique) copula function $C_{j,\ell}$, *i.e.* a bivariate cumulative distribution function with uniform marginal distributions, such that $\mathbb{P}(\varepsilon_j \leq x_j, \varepsilon_\ell \leq x_\ell) = C_{j,\ell}(F_j(x_j), F_\ell(x_\ell))$ for any (x_j, x_ℓ). The associated survival copula is then defined as $\overline{C}_{j,\ell}(u, v) = u + v - 1 + C(1-u, 1-v)$ on $[0,1]^2$, so that $\mathbb{P}(\varepsilon_j > x_j, \varepsilon_\ell > x_\ell) = \overline{C}_{j,\ell}(\overline{F}_j(x_j), \overline{F}_\ell(x_\ell))$. Our joint dependence condition is as follows.

$\mathcal{J}(\boldsymbol{R})$ For all $j, \ell \in \{1, \ldots, d\}$, there is a function $R_{j,\ell}$ such that $\lim_{s\to\infty} s\,\overline{C}_{j,\ell}(x_j/s, x_\ell/s) = R_{j,\ell}(x_j, x_\ell)$ for any $(x_j, x_\ell) \in [0,\infty)^2$.

[For $j = \ell$, $R_{j,j}(x, y) = \min(x, y)$; when $j \neq \ell$, if $C_{j,\ell}$ is not uniquely determined because $\overline{F}_j$ and/or $\overline{F}_\ell$ is not continuous, this limit is required to be valid for one fixed choice of $C_{j,\ell}$.]

This condition imposes the existence of a limiting dependence structure in the joint right tail of ε_j and ε_ℓ, given by the *tail copula* $R_{j,\ell}$ (see Schmidt and Stadtmüller (2006)). This is a very mild condition

that is weaker than assuming a multivariate regular variation structure on $\boldsymbol{\varepsilon}$, as one would classically do in multivariate extreme value theory (see Chapters 6 and 9 in Resnick, 2007).

3 Main results

3.1 Asymptotic behavior of high order statistics of residuals

Since the residual-based Weissman extreme quantile estimator extrapolates a single high order statistic $\widehat{q}_{n,j}(1-k_j/n_j) = \widehat{\varepsilon}^{(n_j)}_{n_j-k_j:n_j,j}$, it is reasonable to first consider the joint asymptotic behavior of such intermediate quantile estimators in order to gain valuable insight about the more difficult motivating problem we have set. It should be clear that this simpler problem is nonetheless a difficult probabilistic question, as the residuals are not considered i.i.d. The crucial point is that, as a consequence of the definition of the quantile as a generalized inverse function, if

$$\widehat{\overline{F}}_{n,j}(t) = \frac{1}{n_j}\sum_{i=1}^{n_j} \mathbb{1}\{\widehat{\varepsilon}^{(n_j)}_{i,j} > t\} \quad \text{and} \quad \widetilde{\overline{F}}_{n,j}(t) = \frac{1}{n_j}\sum_{i=1}^{n_j} \mathbb{1}\{\varepsilon_{i,j} > t\}$$

respectively denote the empirical survival functions of the errors and residuals in sample j, then

$$\begin{aligned} \widehat{q}_{n,j}(\tau_{n,j}) = \widehat{\varepsilon}^{(n_j)}_{n_j-\lfloor n_j(1-\tau_{n,j})\rfloor:n_j,j} \le x \quad &\Longleftrightarrow \quad \widehat{\overline{F}}_{n,j}(x) \le 1-\tau_{n,j} \\ \text{and } \widetilde{q}_{n,j}(\tau_{n,j}) = \varepsilon_{n_j-\lfloor n_j(1-\tau_{n,j})\rfloor:n_j,j} \le x \quad &\Longleftrightarrow \quad \widetilde{\overline{F}}_{n,j}(x) \le 1-\tau_{n,j} \end{aligned}$$

with $\lfloor\cdot\rfloor$ denoting the floor function. It follows that, intuitively, $\widehat{q}_{n,j}(\tau_{n,j})$ should have the same asymptotic distribution as $\widetilde{q}_{n,j}(\tau_{n,j})$, whose asymptotic behavior is well-understood (see *e.g.* Theorem 2.4.8 p.52 in de Haan and Ferreira, 2006), provided $\widehat{\overline{F}}_{n,j}(x)$ is close to $\widetilde{\overline{F}}_{n,j}(x)$ in a suitable sense. How close they may be, according to our theoretical results, is based upon the following definition.

Definition 3.1 (u_n-small random array)**.** Let (u_n) be a nonrandom positive sequence tending to ∞. Then an array $(C_i^{(n)})_{1\le i\le n}$ of positive random variables is said to be a u_n-*small array* if they satisfy $\max_{1\le i\le n} C_i^{(n)}/u_n \xrightarrow{\mathbb{P}} 0$ and, for every positive nonrandom sequence (δ_n) converging to 0,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(C_i^{(n)}\mathbb{1}\{C_i^{(n)} \le \delta_n u_n\}) = 0.$$

Loosely speaking, the columns of a u_n-small array of random variables are small on average and their maximal element (for a given n) grows slower than u_n. This makes it possible to introduce the following assumption on a sample of residuals $(\widehat{\varepsilon}_1^{(n)},\dots,\widehat{\varepsilon}_n^{(n)})$ and the corresponding errors $(\varepsilon_1,\dots,\varepsilon_n)$ that will be at the heart of our main results. In this condition, $\tau_n \uparrow 1$ is such that $n(1-\tau_n)\to\infty$, and μ is a measure putting finite mass on compact Borel subsets of $\mathbb{R}$ (for example, a Radon measure).

$\mathcal{H}(\tau_n,\mu)$ There is a $\sqrt{n(1-\tau_n)}-$small array $(C_i^{(n)})_{1\le i\le n}$ such that $C_i^{(n)}$ is independent of ε_i for each

$i \in \{1, \ldots, n\}$ and the residuals $\widehat{\varepsilon}_i^{(n)}$ fulfill the convergence

$$\int_0^\infty \frac{1}{\sqrt{n(1-\tau_n)}} \sum_{i=1}^n \mathbb{1}\left\{ \frac{\sqrt{n(1-\tau_n)}|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|}{xq(\tau_n) + |\varepsilon_i|} > C_i^{(n)} \right\} \mu(\mathrm{d}x) \xrightarrow{\mathbb{P}} 0$$

where q is the common quantile function of the heavy-tailed errors ε_i.

The fact that μ assigns finite mass to compact Borel subsets of $\mathbb{R}$ guarantees that the integral appearing in condition $\mathcal{H}(\tau_n, \mu)$ is indeed a finite random variable because, with probability 1, the integrand is 0 for x large enough (measurability follows from the Fubini-Tonelli theorem, which applies because μ is then necessarily $\sigma-$finite). Using this condition, we may now state our first main result on the joint asymptotic normality of a finite number of high order statistics of residuals. Let, here and throughout, Dir_1 denote the Dirac mass at 1.

Theorem 3.1 (Joint asymptotic normality of high order statistics of residuals)**.** *Suppose that conditions $\mathcal{C}_2(\boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{A})$ and $\mathcal{J}(\boldsymbol{R})$ hold. Assume also that, for any $1 \leq j \leq d$:*

- *The $n_j = n_j(n)$ satisfy $n_1/n_j \to b_j \in (0, \infty)$ (with $b_1 = 1$), and the $\tau_{n,j} \uparrow 1$ satisfy $n_j(1-\tau_{n,j}) \to \infty$, $(1-\tau_{n,1})/(1-\tau_{n,j}) \to \theta_j \in (0, \infty)$ (with $\theta_1 = 1$) and $\sqrt{n_j(1-\tau_{n,j})}A_j((1-\tau_{n,j})^{-1}) = \mathrm{O}(1)$.*
- *The residuals $(\widehat{\varepsilon}_{1,j}^{(n_j)}, \ldots, \widehat{\varepsilon}_{n_j,j}^{(n_j)})$ and i.i.d. errors $(\varepsilon_{1,j}, \ldots, \varepsilon_{n_j,j})$ satisfy condition $\mathcal{H}(\tau_{n,j}, \mathrm{Dir}_1)$.*

Then

$$\left(\sqrt{n_j(1-\tau_{n,j})} \left\{ \frac{\widehat{q}_{n,j}(\tau_{n,j})}{q_j(\tau_{n,j})} - 1 \right\} \right)_{1 \leq j \leq d} \xrightarrow{\mathrm{d}} \mathcal{N}(\mathbf{0}_d, \boldsymbol{M})$$

where $\boldsymbol{M} = \boldsymbol{M}(\boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{R})$ is the positive semidefinite $d \times d$ symmetric matrix having (j, ℓ)th entry

$$\boldsymbol{M}_{j,\ell} = \gamma_j \gamma_\ell \frac{\sqrt{b_j b_\ell}}{\max(b_j, b_\ell)} R_{j,\ell}\left(\sqrt{\theta_\ell/\theta_j}, \sqrt{\theta_j/\theta_\ell} \right).$$

We may in particular draw from Theorem 3.1 a corollary about the joint convergence of order statistics of a single sequence of residuals at several intermediate levels.

Corollary 3.1 (Joint asymptotic normality of multiple high residuals from a single sample)**.** *Assume that the i.i.d. errors $\varepsilon_1, \ldots, \varepsilon_n$ satisfy condition $\mathcal{C}_2(\gamma, \rho, A)$. Suppose also that:*

- *One has $\tau_n \uparrow 1$, $n(1-\tau_n) \to \infty$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$.*
- *The residuals $(\widehat{\varepsilon}_1^{(n)}, \ldots, \widehat{\varepsilon}_n^{(n)})$ and corresponding errors $(\varepsilon_1, \ldots, \varepsilon_n)$ satisfy assumption $\mathcal{H}(\tau_n, \mathrm{Dir}_1)$.*

If the $\tau_{n,j}$, for $1 \leq j \leq d$, are such that $\tau_{n,1} \equiv \tau_n$ and $(1-\tau_{n,1})/(1-\tau_{n,j}) \to \theta_j \in (0, \infty)$, then

$$\left(\sqrt{n(1-\tau_{n,j})} \left\{ \frac{\widehat{q}_n(\tau_{n,j})}{q(\tau_{n,j})} - 1 \right\} \right)_{1 \leq j \leq d} \xrightarrow{\mathrm{d}} \mathcal{N}(\mathbf{0}_d, \gamma^2 \boldsymbol{M}')$$

where $\boldsymbol{M}'$ is the positive semidefinite $d \times d$ symmetric matrix having (j, ℓ)th entry $\sqrt{\min(\theta_j, \theta_\ell)/\max(\theta_j, \theta_\ell)}$.

An important conclusion from Corollary 3.1 is that the residual-based order statistic $\widehat{q}_n(\tau_n) = \widehat{\varepsilon}^{(n)}_{\lceil n\tau_n\rceil:n}$ estimating the intermediate quantile $q(\tau_n)$ has the same asymptotic distribution as the intermediate empirical quantile $\widetilde{q}_n(\tau_n) = \varepsilon_{\lceil n\tau_n\rceil:n}$ from the error distribution, that is,

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\varepsilon}^{(n)}_{\lceil n\tau_n\rceil:n}}{q(\tau_n)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0,\gamma^2)$$

when $\tau_n \uparrow 1$, $n(1-\tau_n) \to \infty$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$, subject to the $\widehat{\varepsilon}^{(n)}_i$ being close enough to the ε_i in the sense of the high-level condition $\mathcal{H}(\tau_n, \mathrm{Dir}_1)$. The corresponding result for the intermediate order statistic from the error distribution, namely

$$\sqrt{n(1-\tau_n)}\left(\frac{\varepsilon_{\lceil n\tau_n\rceil:n}}{q(\tau_n)} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0,\gamma^2)$$

is a standard result following from Theorem 2.4.8 p.52 in de Haan and Ferreira (2006).

3.2 Extreme value index estimation and Weissman estimator

The original question we considered was to estimate extreme quantiles of the form $q_j(1-p_{n,j})$, with an exceedance probability $p_{n,j} \uparrow 1$ such that $n_j(1-p_{n,j})$ is bounded. This corresponds to the estimation of quantiles which may lie beyond the range of the data. In this situation, the empirical quantile estimators $\widehat{q}_{n,j}(1-p_{n,j})$ and their unfeasible counterparts $\widetilde{q}_{n,j}(1-p_{n,j})$ will not be consistent, because they are unable to extrapolate beyond this range. This motivates considering the residual-based version $\widehat{q}^{\star}_{n,j}(1-p_{n,j})$ of the Weissman extreme quantile estimator instead. The key to the joint asymptotic analysis of these estimators, for $1 \le j \le d$, is the identity

$$\begin{aligned}\log\frac{\widehat{q}^{\star}_{n,j}(1-p_{n,j})}{q_j(1-p_{n,j})} &= \log\left(\frac{k_j}{n_j p_{n,j}}\right)(\widehat{\gamma}_{n,j}(k_j)-\gamma_j) + \log\frac{\widehat{q}_{n,j}(1-k_j/n_j)}{q_j(1-k_j/n_j)} \\ &\quad + \log\left(\left(\frac{k_j}{n_j p_{n,j}}\right)^{\gamma_j}\frac{q_j(1-k_j/n_j)}{q_j(1-p_{n,j})}\right). \qquad (1)\end{aligned}$$

Under condition $\mathcal{C}_2(\boldsymbol{\gamma},\boldsymbol{\rho},\boldsymbol{A})$, and provided $\rho_j < 0$, the final, nonrandom term above is a $\mathrm{O}(1/\sqrt{k_j})$ when the k_j are intermediate sequences such that $k_j = k_j(n) \to \infty$, $k_j/n_j \to 0$, $\sqrt{k_j}A_j(n_j/k_j) = \mathrm{O}(1)$ and $k_j/(n_j p_{n,j}) \to \infty$, see the top equation on p.139 of de Haan and Ferreira (2006).

It remains to analyze the joint asymptotic behavior of $\sqrt{k_j}(\widehat{\gamma}_{n,j}(k_j)-\gamma_j, \log(\widehat{q}_{n,j}(1-k_j/n_j)/q_j(1-k_j/n_j)))$ or equivalently, by the delta-method, of $\sqrt{k_j}(\widehat{\gamma}_{n,j}(k_j)-\gamma_j, \widehat{q}_{n,j}(1-k_j/n_j)/q_j(1-k_j/n_j)-1)$. The main difficulty lies in understanding the asymptotic behavior of the residual-based Hill estimator $\widehat{\gamma}_{n,j}(k_j)$. This is done by writing

$$\widehat{\gamma}_{n,j}(k_j) = \frac{n_j}{k_j}\int_{\widehat{\varepsilon}^{(n_j)}_{n_j-k_j:n_j,j}}^{\infty} \widehat{\overline{F}}_{n,j}(t)\frac{\mathrm{d}t}{t} \quad \text{and} \quad \widetilde{\gamma}_{n,j}(k_j) = \frac{n_j}{k_j}\int_{\varepsilon_{n_j-k_j:n_j,j}}^{\infty} \widetilde{\overline{F}}_{n,j}(t)\frac{\mathrm{d}t}{t}$$

and then by showing that if $\widehat{\overline{F}}_{n,j}(t)$ and $\widetilde{\overline{F}}_{n,j}(t)$ are sufficiently close on average over the right tail of

the distribution of ε_j, then $\widehat{\gamma}_{n,j}(k_j)$ should be close to $\widetilde{\gamma}_{n,j}(k_j)$, whose asymptotic behavior is well-known (see Theorem 3.2.5 p.74 in de Haan and Ferreira, 2006). To make this intuition rigorous, we impose condition $\mathcal{H}(\tau_{n,j}, x^{-1}\mathbb{1}\{x \geq 1\}\mathrm{d}x)$ on the residuals $\widehat{\varepsilon}_{i,j}^{(n_j)}$, which makes it possible to obtain the following main result.

Theorem 3.2 (Joint asymptotic normality of residual-based Hill estimators and order statistics)**.** *Suppose that conditions $\mathcal{C}_2(\boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{A})$ and $\mathcal{J}(\boldsymbol{R})$ hold. Assume also that, for any $1 \leq j \leq d$:*

- *The $n_j = n_j(n)$ and $k_j = k_j(n)$ satisfy $n_1/n_j \to b_j \in (0, \infty)$ and $k_1/k_j \to c_j \in (0, \infty)$ (with $b_1 = c_1 = 1$), as well as $k_j \to \infty$, $k_j/n_j \to 0$ and $\sqrt{k_j}A_j(n_j/k_j) \to \lambda_j \in \mathbb{R}$.*
- *The residuals $(\widehat{\varepsilon}_{1,j}^{(n_j)}, \ldots, \widehat{\varepsilon}_{n_j,j}^{(n_j)})$ and corresponding i.i.d. errors $(\varepsilon_{1,j}, \ldots, \varepsilon_{n_j,j})$ satisfy condition $\mathcal{H}(1 - k_j/n_j, x^{-1}\mathbb{1}\{x \geq 1\}\mathrm{d}x)$.*

Then

$$\left(\sqrt{k_j}\{\widehat{\gamma}_{n,j}(k_j) - \gamma_j\},\ \sqrt{k_j}\left\{ \frac{\widehat{q}_{n,j}(1 - k_j/n_j)}{q_j(1 - k_j/n_j)} - 1 \right\} \right)_{1 \leq j \leq d} \xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\frac{\lambda_j}{1 - \rho_j}, 0 \right)_{1 \leq j \leq d}, \boldsymbol{V} \right)$$

where the matrix $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{\gamma}, \boldsymbol{R})$ is symmetric positive semidefinite of size $2d \times 2d$ and described in block form by 2×2 blocks $\boldsymbol{V}_{j,\ell}$ having entries

$$\begin{aligned}
[\boldsymbol{V}_{j,\ell}]_{1,1} = [\boldsymbol{V}_{j,\ell}]_{2,2} &= \gamma_j \gamma_\ell \frac{b_j b_\ell}{\max(b_j, b_\ell)} \frac{1}{\sqrt{c_j c_\ell}} R_{j,\ell}(c_\ell/b_\ell, c_j/b_j), \\
[\boldsymbol{V}_{j,\ell}]_{1,2} &= \gamma_j \gamma_\ell \frac{b_j b_\ell}{\max(b_j, b_\ell)} \frac{1}{\sqrt{c_j c_\ell}} \left(\int_0^1 R_{j,\ell}(u(c_\ell/b_\ell), c_j/b_j) \frac{\mathrm{d}u}{u} - R_{j,\ell}(c_\ell/b_\ell, c_j/b_j) \right), \\
\text{and } [\boldsymbol{V}_{j,\ell}]_{2,1} &= \gamma_j \gamma_\ell \frac{b_j b_\ell}{\max(b_j, b_\ell)} \frac{1}{\sqrt{c_j c_\ell}} \left(\int_0^1 R_{j,\ell}(c_\ell/b_\ell, v(c_j/b_j)) \frac{\mathrm{d}v}{v} - R_{j,\ell}(c_\ell/b_\ell, c_j/b_j) \right).
\end{aligned}$$

This result shows that earlier results on the joint asymptotic normality of several Hill estimators, such as Corollary 3.4 in Dematteo and Clémençon (2016), Theorem 4 in Stupfler (2019) and Theorem 1 in Daouia et al. (2024a) remain valid for their residual-based versions. The following corollary on the joint asymptotic normality of Weissman residual-based extreme quantile estimators is then essentially an immediate consequence of Equation (1).

Corollary 3.2 (Joint asymptotic normality of residual-based Weissman-Hill estimators)**.** *Work under the conditions of Theorem 3.2 with $\rho_j < 0$, $k_j/(n_j p_{n,j}) \to \infty$ and $\sqrt{k_j}/\log(k_j/(n_j p_{n,j})) \to \infty$ for any j. Then*

$$\begin{aligned}
\left(\frac{\sqrt{k_j}}{\log(k_j/(n_j p_{n,j}))} \left\{ \frac{\widehat{q}^{\star}_{n,j}(1 - p_{n,j})}{q_j(1 - p_{n,j})} - 1 \right\} \right)_{1 \leq j \leq d} &= \left(\sqrt{k_j}\{\widehat{\gamma}_{n,j}(k_j) - \gamma_j\} \right)_{1 \leq j \leq d} + \mathrm{o}_{\mathbb{P}}(1) \\
&\xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\frac{\lambda_1}{1 - \rho_1}, \ldots, \frac{\lambda_d}{1 - \rho_d} \right), \boldsymbol{V}_\gamma \right)
\end{aligned}$$

where $\boldsymbol{V}_\gamma = \boldsymbol{V}_\gamma(\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{\gamma}, \boldsymbol{R})$ *is the positive semidefinite* $d \times d$ *symmetric matrix defined elementwise as*

$$[\boldsymbol{V}_\gamma]_{j,\ell} = [\boldsymbol{V}_{j,\ell}]_{1,1} = \gamma_j \gamma_\ell \frac{b_j b_\ell}{\max(b_j, b_\ell)} \frac{1}{\sqrt{c_j c_\ell}} R_{j,\ell}(c_\ell/b_\ell, c_j/b_j).$$

For the sake of convenience, and to make it easier to discuss the scope of our results, we state the following simpler corollary about the convergence of Hill and Weissman residual-based estimators for a single sample.

Corollary 3.3 (Marginal asymptotic normality of the residual-based Hill and Weissman estimators)**.** *Assume that the i.i.d. errors* $\varepsilon_1, \ldots, \varepsilon_n$ *satisfy condition* $\mathcal{C}_2(\gamma, \rho, A)$*. Suppose also that:*

- *One has* $k = k(n) \to \infty$ *with* $k/n \to 0$ *and* $\sqrt{k} A(n/k) \to \lambda \in \mathbb{R}$*.*
- *The residuals* $(\widehat{\varepsilon}_1^{(n)}, \ldots, \widehat{\varepsilon}_n^{(n)})$ *and errors* $(\varepsilon_1, \ldots, \varepsilon_n)$ *satisfy assumption* $\mathcal{H}(1 - k/n, x^{-1}\mathbb{1}\{x \geq 1\}\mathrm{d}x)$*.*

If the k_j*, for* $1 \leq j \leq d$*, are such that* $k_1 \equiv k$ *and* $k_1/k_j \to c_j \in (0, \infty)$*, then*

$$\begin{aligned}&\left(\sqrt{k_1}\{\widehat{\gamma}_n(k_1) - \gamma\},\ \sqrt{k_1}\left\{\frac{\widehat{q}_n(1 - k_1/n)}{q(1 - k_1/n)} - 1\right\}, \ldots, \sqrt{k_d}\left\{\frac{\widehat{q}_n(1 - k_d/n)}{q(1 - k_d/n)} - 1\right\}\right) \\ &\xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\frac{\lambda}{1 - \rho}, 0, \ldots, 0\right), \gamma^2 \boldsymbol{\Sigma}\right)\end{aligned}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{c})$ *is the positive semidefinite* $(d+1) \times (d+1)$ *symmetric matrix defined elementwise as*

$$\boldsymbol{\Sigma}_{1,1} = 1,\ \boldsymbol{\Sigma}_{1,\ell+1} = \frac{\log(c_\ell)}{\sqrt{c_\ell}} \mathbb{1}\{c_\ell > 1\} \text{ and } \boldsymbol{\Sigma}_{j+1,\ell+1} = \sqrt{\min(c_j, c_\ell)/\max(c_j, c_\ell)}$$

for any $j, \ell \in \{1, \ldots, d\}$*. If moreover* $\rho < 0$*, and* $p_n \uparrow 1$ *is such that* $k/(np_n) \to \infty$ *and* $\sqrt{k}/\log(k/(np_n)) \to \infty$*, then*

$$\frac{\sqrt{k}}{\log(k/(np_n))}\left(\frac{\widehat{q}_n^\star(1 - p_n)}{q(1 - p_n)} - 1\right) = \sqrt{k}(\widehat{\gamma}_n(k) - \gamma) + \mathrm{o}_{\mathbb{P}}(1) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\frac{\lambda}{1 - \rho}, \gamma^2\right).$$

In other words, under condition $\mathcal{H}(1 - k/n, x^{-1}\mathbb{1}\{x \geq 1\}\mathrm{d}x)$, the random pair of residual-based estimators $(\sqrt{k}\{\widehat{\gamma}_n(k) - \gamma\}, \sqrt{k}\{\widehat{q}_n(1 - k/n)/q(1 - k/n) - 1\})$ and the Weissman-Hill residual-based estimator $\widehat{q}_n^\star(1 - p_n)$ have the same asymptotic behavior than their traditional versions based on the sample $\varepsilon_1, \ldots, \varepsilon_n$, as can be seen from the results of Section 3.2 and Theorem 4.3.8 p.138 in de Haan and Ferreira (2006).

3.3 Expected Shortfall estimation above extreme levels

In risk management applications, the choice of risk measure is paramount. Part of the motivation for the choice of a given risk measure lies in its axiomatic properties. Artzner et al. (1999) argue that it is reasonable to ask a risk measure to be coherent, meaning that it should be translation equivariant, positive homogeneous, monotonic and subadditive. While quantiles, also called Value-at-Risk in the

literature, are not subadditive in general, the Expected Shortfall (also called Conditional-Value-at-Risk, or mean residual life) is so, and it has for this reason been the subject of substantial attention in actuarial and financial risk assessment. It now tends to be preferred to the quantile risk measure by practitioners concerned with exposure to a catastrophic event, and by major regulators such as the EU, the UK, Australia and Canada, which will be requiring the use of ES(0.975), rather than $q(0.99)$, in alternative internal models from 1 January 2025: the EU has formalized 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 latest Basel Committee on Banking Supervision rules.

The Expected Shortfall of a random variable ε, with survival function $\overline{F}$, above its τth quantile $q(\tau)$, is defined as

$$\mathrm{ES}(\tau) = \mathbb{E}(\varepsilon \,|\, \varepsilon > q(\tau)) = q(\tau) + \frac{1}{\overline{F}(q(\tau))} \int_{q(\tau)}^{\infty} \overline{F}(t)\,\mathrm{d}t.$$

If a sample $(\varepsilon_1, \ldots, \varepsilon_n)$ from ε is available, and if $\varepsilon_{1:n} \leq \varepsilon_{2:n} \leq \cdots \leq \varepsilon_{n:n}$ are the corresponding order statistics, then a natural estimator of $\mathrm{ES}(1-k/n)$ is

$$\widetilde{\mathrm{ES}}_n(1-k/n) = \frac{1}{k}\sum_{i=1}^{k} \varepsilon_{n-i+1:n} = \varepsilon_{n-k:n} + \frac{n}{k}\int_{\varepsilon_{n-k:n}}^{\infty} \widetilde{\overline{F}}_n(t)\,\mathrm{d}t.$$

Up to the additional order statistic, this estimator has exactly the same form as the Hill estimator, only with the weighting measure $x^{-1}\mathbb{1}\{x \geq 1\}\mathrm{d}x$ replaced by the restriction $\mathbb{1}\{x \geq 1\}\mathrm{d}x$ of the Lebesgue measure on $[1, \infty)$. Moreover, when ε has a heavy-tailed distribution with extreme value index $\gamma \in (0, 1)$, it is straightforward to obtain the convergence

$$\frac{\mathrm{ES}(\tau)}{q(\tau)} \to \frac{1}{1-\gamma} \text{ as } \tau \uparrow 1. \tag{2}$$

This follows, for instance, from Proposition B.1.10 p.369 in de Haan and Ferreira (2006). This means that, just like extreme quantiles, the Expected Shortfall at extreme levels satisfies an extrapolation relationship:

$$\frac{\mathrm{ES}(\tau')}{\mathrm{ES}(\tau)} \approx \left(\frac{1-\tau'}{1-\tau}\right)^{-\gamma} \text{ when } \tau' > \tau \uparrow 1.$$

Our general theory therefore also applies to the estimation of the Expected Shortfall at extreme levels. Let $\mathrm{ES}_j(\tau)$ be the Expected Shortfall of ε_j above its quantile $q_j(\tau)$ and

$$\widehat{\mathrm{ES}}_{n,j}(1-k_j/n_j) = \frac{1}{k_j}\sum_{i=1}^{k_j} \widehat{\varepsilon}^{(n_j)}_{n_j-i+1:n_j,j} = \widehat{\varepsilon}^{(n_j)}_{n_j-k_j:n_j,j} + \frac{n_j}{k_j}\int_{\widehat{\varepsilon}^{(n_j)}_{n_j-k_j:n_j,j}}^{\infty} \widehat{\overline{F}}_{n,j}(t)\,\mathrm{d}t$$

be its residual-based estimator. The following result examines the joint convergence of $\widehat{\gamma}_{n,j}(k_j)$, $\widehat{q}_{n,j}(1-k_j/n_j)$ and $\widehat{\mathrm{ES}}_{n,j}(1-k_j/n_j)$ when the k_j are intermediate sequences as a starting point, before obtaining the convergence of a Weissman-type estimator of the Expected Shortfall above an extreme level.

Theorem 3.3 (Joint asymptotic normality of residual-based Hill estimators, order statistics and empirical Expected Shortfall estimators)**.** *Assume that $\mathcal{C}_2(\boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{A})$ and $\mathcal{J}(\boldsymbol{R})$ hold. Assume also that, for any $1 \leq j \leq d$, $\gamma_j < 1/2$, and that:*

• *The* $n_j = n_j(n)$ *and* $k_j = k_j(n)$ *satisfy* $n_1/n_j \to b_j \in (0,\infty)$ *and* $k_1/k_j \to c_j \in (0,\infty)$ *(with* $b_1 = c_1 = 1$*), as well as* $k_j \to \infty$, $k_j/n_j \to 0$ *and* $\sqrt{k_j}A_j(n_j/k_j) \to \lambda_j \in \mathbb{R}$.

• *The residuals* $(\widehat{\varepsilon}_{1,j}^{(n_j)}, \ldots, \widehat{\varepsilon}_{n_j,j}^{(n_j)})$ *and corresponding i.i.d. errors* $(\varepsilon_{1,j}, \ldots, \varepsilon_{n_j,j})$ *satisfy condition* $\mathcal{H}(1 - k_j/n_j, \mathbb{1}\{x \geq 1\}\mathrm{d}x)$.

Then

$$\left(\sqrt{k_j}\{\widehat{\gamma}_{n,j}(k_j) - \gamma_j\},\ \sqrt{k_j}\left\{ \frac{\widehat{q}_{n,j}(1-k_j/n_j)}{q_j(1-k_j/n_j)} - 1 \right\},\ \sqrt{k_j}\left\{ \frac{\widehat{\mathrm{ES}}_{n,j}(1-k_j/n_j)}{\mathrm{ES}_j(1-k_j/n_j)} - 1 \right\} \right)_{1\leq j\leq d}$$
$$\xrightarrow{\mathrm{d}} \mathcal{N}\left(\left(\frac{\lambda_j}{1-\rho_j}, 0, 0 \right)_{1\leq j\leq d}, \boldsymbol{S} \right)$$

where the matrix $\boldsymbol{S} = \boldsymbol{S}(\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{\gamma}, \boldsymbol{R})$ *is symmetric positive semidefinite of size* $3d \times 3d$ *and described in block form by* 3×3 *blocks* $\boldsymbol{S}_{j,\ell}$ *having entries* $[\boldsymbol{S}_{j,\ell}]_{1,1} = [\boldsymbol{V}_{j,\ell}]_{1,1}$, $[\boldsymbol{S}_{j,\ell}]_{1,2} = [\boldsymbol{V}_{j,\ell}]_{1,2}$, $[\boldsymbol{S}_{j,\ell}]_{2,1} = [\boldsymbol{V}_{j,\ell}]_{2,1}$, $[\boldsymbol{S}_{j,\ell}]_{2,2} = [\boldsymbol{V}_{j,\ell}]_{2,2}$, *with the notation of Theorem 3.2, and*

$$[\boldsymbol{S}_{j,\ell}]_{1,3} = \gamma_j\gamma_\ell(1-\gamma_\ell)\frac{b_jb_\ell}{\max(b_j,b_\ell)}\frac{1}{\sqrt{c_jc_\ell}}\left(\int_0^1\int_0^1 R_{j,\ell}(u(c_\ell/b_\ell), v(c_j/b_j))\frac{\mathrm{d}u}{u}\frac{\mathrm{d}v}{v^{\gamma_\ell+1}} - \int_0^1 R_{j,\ell}(c_\ell/b_\ell, v(c_j/b_j))\frac{\mathrm{d}v}{v^{\gamma_\ell+1}} \right),$$
$$[\boldsymbol{S}_{j,\ell}]_{2,3} = \gamma_j\gamma_\ell(1-\gamma_\ell)\frac{b_jb_\ell}{\max(b_j,b_\ell)}\frac{1}{\sqrt{c_jc_\ell}}\int_0^1 R_{j,\ell}(c_\ell/b_\ell, v(c_j/b_j))\frac{\mathrm{d}v}{v^{\gamma_\ell+1}},$$
$$[\boldsymbol{S}_{j,\ell}]_{3,1} = \gamma_j(1-\gamma_j)\gamma_\ell\frac{b_jb_\ell}{\max(b_j,b_\ell)}\frac{1}{\sqrt{c_jc_\ell}}\left(\int_0^1\int_0^1 R_{j,\ell}(u(c_\ell/b_\ell), v(c_j/b_j))\frac{\mathrm{d}u}{u^{\gamma_j+1}}\frac{\mathrm{d}v}{v} - \int_0^1 R_{j,\ell}(u(c_\ell/b_\ell), c_j/b_j)\frac{\mathrm{d}u}{u^{\gamma_j+1}} \right),$$
$$[\boldsymbol{S}_{j,\ell}]_{3,2} = \gamma_j(1-\gamma_j)\gamma_\ell\frac{b_jb_\ell}{\max(b_j,b_\ell)}\frac{1}{\sqrt{c_jc_\ell}}\int_0^1 R_{j,\ell}(u(c_\ell/b_\ell), c_j/b_j)\frac{\mathrm{d}u}{u^{\gamma_j+1}},$$
$$\text{and } [\boldsymbol{S}_{j,\ell}]_{3,3} = \gamma_j(1-\gamma_j)\gamma_\ell(1-\gamma_\ell)\frac{b_jb_\ell}{\max(b_j,b_\ell)}\frac{1}{\sqrt{c_jc_\ell}}\int_0^1\int_0^1 R_{j,\ell}(u(c_\ell/b_\ell), v(c_j/b_j))\frac{\mathrm{d}u}{u^{\gamma_j+1}}\frac{\mathrm{d}v}{v^{\gamma_\ell+1}}.$$

An alternative estimator is a residual-based version of the so-called AE estimator of El Methni and Stupfler (2017). This estimator relies on convergence (2) and is defined, in our setting, as

$$\widetilde{\mathrm{ES}}_{n,j}(1-k_j/n_j) = \frac{\widehat{q}_{n,j}(1-k_j/n_j)}{1-\widehat{\gamma}_{n,j}(k_j)}.$$

The joint convergence of these estimators follows from Theorem 3.2.

Corollary 3.4 (Joint asymptotic normality of residual-based Hill estimators, order statistics and quantile-based Expected Shortfall estimators)**.** *Work under the conditions of Theorem 3.2, with* $\gamma_j < 1$

for any j. Then

$$\left(\sqrt{k_j}\{\widehat{\gamma}_{n,j}(k_j)-\gamma_j\},\ \sqrt{k_j}\left\{\frac{\widehat{q}_{n,j}(1-k_j/n_j)}{q_j(1-k_j/n_j)}-1\right\},\ \sqrt{k_j}\left\{\frac{\widetilde{\mathrm{ES}}_{n,j}(1-k_j/n_j)}{\mathrm{ES}_j(1-k_j/n_j)}-1\right\}\right)_{1\le j\le d}$$
$$\xrightarrow{\mathrm{d}}\mathcal{N}\left(\left(\frac{\lambda_j}{1-\rho_j},0,-\frac{\lambda_j\gamma_j\rho_j}{(1-\gamma_j)(1-\rho_j)(1-\gamma_j-\rho_j)}\right)_{1\le j\le d},\boldsymbol{V}_{\mathrm{ES}}\right)$$

where the matrix $\boldsymbol{V}_{\mathrm{ES}}=\boldsymbol{V}_{\mathrm{ES}}(\boldsymbol{b},\boldsymbol{c},\boldsymbol{\gamma},\boldsymbol{R})$ is symmetric positive semidefinite of size $3d\times 3d$ and described in block form by 3×3 blocks $\boldsymbol{V}_{\mathrm{ES},j,\ell}$ having entries

$$[\boldsymbol{V}_{\mathrm{ES},j,\ell}]_{1,1}=[\boldsymbol{V}_{j,\ell}]_{1,1},\ [\boldsymbol{V}_{\mathrm{ES},j,\ell}]_{1,2}=[\boldsymbol{V}_{j,\ell}]_{1,2},\ [\boldsymbol{V}_{\mathrm{ES},j,\ell}]_{2,1}=[\boldsymbol{V}_{j,\ell}]_{2,1},\ [\boldsymbol{V}_{\mathrm{ES},j,\ell}]_{2,2}=[\boldsymbol{V}_{j,\ell}]_{2,2},$$
$$[\boldsymbol{V}_{\mathrm{ES},j,\ell}]_{1,3}=\frac{1}{1-\gamma_\ell}[\boldsymbol{V}_{j,\ell}]_{1,1}+[\boldsymbol{V}_{j,\ell}]_{1,2},$$
$$[\boldsymbol{V}_{\mathrm{ES},j,\ell}]_{3,1}=\frac{1}{1-\gamma_j}[\boldsymbol{V}_{j,\ell}]_{1,1}+[\boldsymbol{V}_{j,\ell}]_{2,1},$$
$$[\boldsymbol{V}_{\mathrm{ES},j,\ell}]_{2,3}=\frac{1}{1-\gamma_\ell}[\boldsymbol{V}_{j,\ell}]_{2,1}+[\boldsymbol{V}_{j,\ell}]_{2,2},$$
$$[\boldsymbol{V}_{\mathrm{ES},j,\ell}]_{3,2}=\frac{1}{1-\gamma_j}[\boldsymbol{V}_{j,\ell}]_{1,2}+[\boldsymbol{V}_{j,\ell}]_{2,2},$$
$$\text{and } [\boldsymbol{V}_{\mathrm{ES},j,\ell}]_{3,3}=\frac{1}{(1-\gamma_j)(1-\gamma_\ell)}[\boldsymbol{V}_{j,\ell}]_{1,1}+\frac{1}{1-\gamma_j}[\boldsymbol{V}_{j,\ell}]_{1,2}+\frac{1}{1-\gamma_\ell}[\boldsymbol{V}_{j,\ell}]_{2,1}+[\boldsymbol{V}_{j,\ell}]_{2,2}$$

with the notation of Theorem 3.2.

These two estimators of the Expected Shortfall at intermediate levels give rise to classes of Weissman-type extrapolated estimators:

$$\widehat{\mathrm{ES}}^{\star}_{n,j}(1-p_{n,j})=\left(\frac{k_j}{n_jp_{n,j}}\right)^{\widehat{\gamma}_{n,j}}\widehat{\mathrm{ES}}_{n,j}(1-k_j/n_j),$$

based on the intermediate estimator $\widehat{\mathrm{ES}}_{n,j}(1-k_j/n_j)$, and

$$\widetilde{\mathrm{ES}}^{\star}_{n,j}(1-p_{n,j})=\left(\frac{k_j}{n_jp_{n,j}}\right)^{\widehat{\gamma}_{n,j}}\widetilde{\mathrm{ES}}_{n,j}(1-k_j/n_j),$$

based on the intermediate estimator $\widetilde{\mathrm{ES}}_{n,j}(1-k_j/n_j)$. The following result, which is obtained by applying Theorem 3.3 and Corollary 3.4, examines their asymptotic behavior.

Corollary 3.5 (Joint asymptotic normality of Weissman-type Expected Shortfall estimators)**.** *Work under the conditions of Theorem 3.3 (resp. Corollary 3.4) with $\rho_j<0$, $k_j/(n_jp_{n,j})\to\infty$ and $\sqrt{k_j}/\log(k_j/(n_jp_{n,j}))\to\infty$ for any j. Then*

$$\left(\frac{\sqrt{k_j}}{\log(k_j/(n_jp_{n,j}))}\left\{\frac{\overline{\mathrm{ES}}^{\star}_{n,j}(1-p_{n,j})}{\mathrm{ES}_j(1-p_{n,j})}-1\right\}\right)_{1\le j\le d}=\left(\sqrt{k_j}\{\widehat{\gamma}_{n,j}(k_j)-\gamma_j\}\right)_{1\le j\le d}+\mathrm{o}_{\mathbb{P}}(1)$$
$$\xrightarrow{\mathrm{d}}\mathcal{N}\left(\left(\frac{\lambda_1}{1-\rho_1},\dots,\frac{\lambda_d}{1-\rho_d}\right),\boldsymbol{V}_{\gamma}\right)$$

where $\overline{\mathrm{ES}}^{\star}_{n,j} = \widehat{\mathrm{ES}}^{\star}_{n,j}$ *(resp.* $\overline{\mathrm{ES}}^{\star}_{n,j} = \widetilde{\mathrm{ES}}^{\star}_{n,j}$*), with the notation of Corollary 3.2.*

We conclude with the following corollary in the univariate setting which, regarding the particular estimator $\widehat{\mathrm{ES}}_n$, was noted in Theorem 2 in El Methni and Stupfler (2017) and Corollary 5 in Stupfler (2019) for i.i.d. data. The result concerning $\widetilde{\mathrm{ES}}_n$ is, to the best of our knowledge, new even for i.i.d. data.

Corollary 3.6 (Marginal asymptotic normality of the residual-based tail Expected Shortfall estimators)**.** *Assume that the i.i.d. errors* $\varepsilon_1, \ldots, \varepsilon_n$ *satisfy condition* $\mathcal{C}_2(\gamma, \rho, A)$*. Suppose also that:*

- *One has* $k = k(n) \to \infty$ *with* $k/n \to 0$ *and* $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$*.*
- *The residuals* $(\widehat{\varepsilon}_1^{(n)}, \ldots, \widehat{\varepsilon}_n^{(n)})$ *and errors* $(\varepsilon_1, \ldots, \varepsilon_n)$ *satisfy assumption* $\mathcal{H}(1-k/n, x^{-1}\mathbb{1}\{x \geq 1\}\mathrm{d}x)$*.*

Then, if $\gamma < 1$*, one has*

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

If moreover assumption $\mathcal{H}(1-k/n, \mathbb{1}\{x \geq 1\}\mathrm{d}x)$ *holds and* $\gamma < 1/2$*, then*

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

Finally, if $\rho < 0$ *and* $p_n \uparrow 1$ *is such that* $k/(np_n) \to \infty$ *and* $\sqrt{k}/\log(k/(np_n)) \to \infty$*, then*

$$\frac{\sqrt{k}}{\log(k/(np_n))}\left(\frac{\overline{\mathrm{ES}}^{\star}_n(1-p_n)}{\mathrm{ES}(1-p_n)} - 1\right) = \sqrt{k}(\widehat{\gamma}_n(k) - \gamma) + \mathrm{o}_{\mathbb{P}}(1) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\right)$$

where $\overline{\mathrm{ES}}^{\star}_n = \widetilde{\mathrm{ES}}^{\star}_n$ *if* $\gamma < 1$ *and assumption* $\mathcal{H}(1-k/n, x^{-1}\mathbb{1}\{x \geq 1\}\mathrm{d}x)$ *holds, or* $\overline{\mathrm{ES}}^{\star}_n = \widehat{\mathrm{ES}}^{\star}_n$ *if* $\gamma < 1/2$ *and assumption* $\mathcal{H}(1-k/n, \mathbb{1}\{x \geq 1\}\mathrm{d}x)$ *holds.*

4 Examples of models covered by our framework

4.1 Location-scale parametric regression

Suppose that the observations are generated from a location-scale regression model of the form

$$Y_i = m(\boldsymbol{X}_i) + \sigma(\boldsymbol{X}_i)\varepsilon_i,\ 1 \leq i \leq n,$$

where the covariate value $\boldsymbol{X}_i \in \mathbb{R}^p$ is observed, and is either nonrandom, or random and independent from ε_i. Assume that a suitable model identifiability condition (depending on the structure of the model and/or the assumptions on ε_i) holds and that the functions m and $\sigma > 0$ can hence be estimated by $\widehat{m}_n$ and $\widehat{\sigma}_n > 0$, so that the residuals from the model are

$$\widehat{\varepsilon}_i^{(n)} = \frac{Y_i - \widehat{m}_n(\boldsymbol{X}_i)}{\widehat{\sigma}_n(\boldsymbol{X}_i)} = \frac{m(\boldsymbol{X}_i) - \widehat{m}_n(\boldsymbol{X}_i)}{\widehat{\sigma}_n(\boldsymbol{X}_i)} + \frac{\sigma(\boldsymbol{X}_i)}{\widehat{\sigma}_n(\boldsymbol{X}_i)}\varepsilon_i.$$

As a consequence

$$\begin{aligned}\frac{|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|}{q(\tau_n) + |\varepsilon_i|} &\leq \frac{|\widehat{m}_n(\boldsymbol{X}_i) - m(\boldsymbol{X}_i)| + |\widehat{\sigma}_n(\boldsymbol{X}_i) - \sigma(\boldsymbol{X}_i)||\varepsilon_i|}{\widehat{\sigma}_n(\boldsymbol{X}_i)(q(\tau_n) + |\varepsilon_i|)} \\ &\leq \frac{1}{\widehat{\sigma}_n(\boldsymbol{X}_i)}\left(\frac{|\widehat{m}_n(\boldsymbol{X}_i) - m(\boldsymbol{X}_i)|}{q(\tau_n)} + |\widehat{\sigma}_n(\boldsymbol{X}_i) - \sigma(\boldsymbol{X}_i)|\right).\end{aligned}$$

Assume now that $m(\boldsymbol{x}) = m(\boldsymbol{x}, \boldsymbol{\beta})$ and $\sigma(\boldsymbol{x}) = \sigma(\boldsymbol{x}, \boldsymbol{\theta})$, for $\boldsymbol{\beta}, \boldsymbol{\theta} \in \mathbb{R}^r$, and that $\widehat{m}_n$, $\widehat{\sigma}_n$ are obtained as $\widehat{m}_n(\boldsymbol{x}) = m(\boldsymbol{x}, \widehat{\boldsymbol{\beta}}_n)$ and $\widehat{\sigma}_n(\boldsymbol{x}) = \sigma(\boldsymbol{x}, \widehat{\boldsymbol{\theta}}_n)$, where $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\theta}}_n$ are $\sqrt{n}-$consistent estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, respectively. Assume also that there is a neighborhood V of $\mathbf{0}$ in $\mathbb{R}^r$ such that

$$M(\boldsymbol{x}) = \sup_{\boldsymbol{h} \in V} \frac{|m(\boldsymbol{x}, \boldsymbol{\beta} + \boldsymbol{h}) - m(\boldsymbol{x}, \boldsymbol{\beta})|}{\|\boldsymbol{h}\|} < \infty \ \text{ and } \ S(\boldsymbol{x}) = \sup_{\boldsymbol{h} \in V} \frac{|\log \sigma(\boldsymbol{x}, \boldsymbol{\theta} + \boldsymbol{h}) - \log \sigma(\boldsymbol{x}, \boldsymbol{\theta})|}{\|\boldsymbol{h}\|} < \infty.$$

Assume finally that σ is bounded from below by a positive constant σ_0 on V. Then, with arbitrarily high probability as $n \to \infty$,

$$\frac{|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|}{q(\tau_n) + |\varepsilon_i|} \leq \frac{\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|}{\sigma_0 q(\tau_n)} M(\boldsymbol{X}_i) + \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| S(\boldsymbol{X}_i) \exp(\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| S(\boldsymbol{X}_i)),$$

where the inequality $|e^x - 1| \leq |x| e^{|x|}$ for all x, coming as a consequence of the mean value theorem, was used to obtain the second term. If moreover $\max_{1 \leq i \leq n} S(\boldsymbol{X}_i) = \mathrm{o}_{\mathbb{P}}(\sqrt{n})$, then, with arbitrarily high probability as $n \to \infty$,

$$\frac{|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|}{q(\tau_n) + |\varepsilon_i|} \leq \frac{\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|}{\sigma_0 q(\tau_n)} M(\boldsymbol{X}_i) + 2\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| S(\boldsymbol{X}_i).$$

As such

$$\sqrt{n(1-\tau_n)}\frac{|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|}{q(\tau_n) + |\varepsilon_i|} \leq \left(\frac{1}{\sigma_0} + 2\right)(\sqrt{n}\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\| + \sqrt{n}\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|) \times \sqrt{1-\tau_n}\left(\frac{M(\boldsymbol{X}_i)}{q(\tau_n)} + S(\boldsymbol{X}_i)\right).$$

The fact that $\sqrt{n}\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|$ and $\sqrt{n}\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|$ are $\mathrm{O}_{\mathbb{P}}(1)$ suggests defining $C_i^{(n)}$ as

$$C_i^{(n)} = u_n\sqrt{1-\tau_n}\left(\frac{M(\boldsymbol{X}_i)}{q(\tau_n)} + S(\boldsymbol{X}_i)\right)$$

where (u_n) is a suitably chosen nonrandom sequence tending to infinity. By definition, for each $n \geq 1$ and $i \in \{1, \ldots, n\}$, $C_i^{(n)}$ is independent of ε_i, and

$$\max_{1\leq i\leq n} \frac{1}{C_i^{(n)}} \times \frac{\sqrt{n(1-\tau_n)}|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|}{q(\tau_n) + |\varepsilon_i|} \xrightarrow{\mathbb{P}} 0.$$

Checking that $(C_i^{(n)})_{1\leq i\leq n}$ is a $\sqrt{n(1-\tau_n)}-$array then essentially amounts to showing that one can choose (u_n) so that

$$\max_{1\leq i\leq n} M(\boldsymbol{X}_i) = \mathrm{o}_{\mathbb{P}}\left(q(\tau_n)\frac{\sqrt{n}}{u_n}\right) \quad \text{and} \quad \max_{1\leq i\leq n} S(\boldsymbol{X}_i) = \mathrm{o}_{\mathbb{P}}\left(\frac{\sqrt{n}}{u_n}\right)$$

as well as

$$\frac{1}{n}\sum_{i=1}^n M(\boldsymbol{X}_i) = \mathrm{o}_{\mathbb{P}}\left(\frac{q(\tau_n)}{u_n\sqrt{1-\tau_n}}\right) \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n M(\boldsymbol{X}_i) = \mathrm{o}_{\mathbb{P}}\left(\frac{1}{u_n\sqrt{1-\tau_n}}\right).$$

Such a choice of (u_n) can for instance obviously be made as soon as there is $\delta > 0$ such that

$$\max_{1\leq i\leq n} M(\boldsymbol{X}_i) = \mathrm{o}_{\mathbb{P}}(n^{1/2-\delta}) \quad \text{and} \quad \max_{1\leq i\leq n} S(\boldsymbol{X}_i) = \mathrm{o}_{\mathbb{P}}(n^{1/2-\delta}),$$

as well as

$$\frac{1}{n}\sum_{i=1}^n M(\boldsymbol{X}_i) = \mathrm{O}_{\mathbb{P}}(1) \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n S(\boldsymbol{X}_i) = \mathrm{O}_{\mathbb{P}}(1).$$

This can either be imposed as an assumption on the covariate values $\boldsymbol{X}_i$ (when they are nonrandom) or can follow from probabilistic considerations on the sequence $(\boldsymbol{X}_i)_{i\geq 1}$ (when it is random). We give a simple, non-optimal general result following this discussion.

Corollary 4.1 (Extremal regression in a parametric location-scale regression model)**.** *Assume that $Y_i = m(\boldsymbol{X}_i, \boldsymbol{\beta}) + \sigma(\boldsymbol{X}_i, \boldsymbol{\theta})\varepsilon_i$, for $i \geq 1$, where $\boldsymbol{\beta}, \boldsymbol{\theta} \in \mathbb{R}^r$ and $\boldsymbol{X}_i$ is either nonrandom, or random and independent from ε_i, with the ε_i being independent copies of a random variable ε satisfying condition $\mathcal{C}_2(\gamma, \rho, A)$. Assume that $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \mathrm{O}_{\mathbb{P}}(1)$ and $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = \mathrm{O}_{\mathbb{P}}(1)$, and define $\widehat{\varepsilon}_i^{(n)} = (Y_i - m(\boldsymbol{X}_i, \widehat{\boldsymbol{\beta}}_n))/\sigma(\boldsymbol{X}_i, \widehat{\boldsymbol{\theta}}_n)$. Suppose that there is a subset $\mathcal{X}$ of $\mathbb{R}^p$ and a neighborhood V of $\mathbf{0}$ in $\mathbb{R}^r$ with $\inf_{\boldsymbol{x}\in\mathcal{X}} \inf_{\boldsymbol{h}\in V} \sigma(\boldsymbol{x}, \boldsymbol{\theta} + \boldsymbol{h}) > 0$ and, for all $\boldsymbol{x} \in \mathcal{X}$,*

$$M(\boldsymbol{x}) = \sup_{\boldsymbol{h}\in V} \frac{|m(\boldsymbol{x}, \boldsymbol{\beta} + \boldsymbol{h}) - m(\boldsymbol{x}, \boldsymbol{\beta})|}{\|\boldsymbol{h}\|} < \infty \quad \text{and} \quad S(\boldsymbol{x}) = \sup_{\boldsymbol{h}\in V} \frac{|\log\sigma(\boldsymbol{x}, \boldsymbol{\theta} + \boldsymbol{h}) - \log\sigma(\boldsymbol{x}, \boldsymbol{\theta})|}{\|\boldsymbol{h}\|} < \infty.$$

Let $k = k(n) \to \infty$ satisfy $k/n \to 0$ and $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$.

- *In the random design setting, suppose that the $\boldsymbol{X}_i$ are independent copies of a random vector $\boldsymbol{X} \in \mathbb{R}^p$*

with support $\mathcal{X}$, such that $T(\boldsymbol{X}) = M(\boldsymbol{X}) + S(\boldsymbol{X})$ belongs to the max-domain of attraction of an extreme value distribution with extreme value index less than $1/2$. Then $(\sqrt{k}(\widehat{\gamma}_n(k) - \gamma), \sqrt{k}\log(\widehat{q}_n(1 - k/n)/q(1-k/n)))$ converges weakly to a pair of independent Gaussian random variables having respective means $\lambda/(1-\rho)$ and 0, and the same variance γ^2. If moreover $\rho < 0$, and $\tau'_n \uparrow 1$ is such that $k/(n(1-\tau'_n)) \to \infty$ and $\sqrt{k}/\log(k/(n(1-\tau'_n))) \to \infty$, then

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

• *In the nonrandom design setting, the above result stays true provided the $\boldsymbol{X}_i = \boldsymbol{x}_i$ all belong to the set $\mathcal{X}$ and there is $\delta > 0$ such that the $T(\boldsymbol{x}_i) = M(\boldsymbol{x}_i) + S(\boldsymbol{x}_i)$ satisfy*

$$\max_{1\le i\le n} T(\boldsymbol{x}_i) = \mathrm{o}(n^{1/2-\delta}) \quad \textit{and} \quad \frac{1}{n}\sum_{i=1}^{n} T(\boldsymbol{x}_i) = \mathrm{O}(1).$$

The conditions of Corollary 4.1 are quite general and, in particular, make it possible to handle a wide range of linear heteroskedastic models. We discuss, as an illustration, the case of the class of models $Y_i = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \exp(\theta h(\boldsymbol{x}_i^\top\boldsymbol{\beta}))\varepsilon_i$, considered in Carroll and Ruppert (1982), where the $\boldsymbol{x}_i$ are nonrandom, the ε_i are independent, identically distributed with a symmetric distribution, h is a known smooth and Lipschitz function, and $\boldsymbol{\beta} \in \mathbb{R}^r$, $\theta \in \mathbb{R}$ are unknown parameters to be estimated. This contains, for example, the models $Y_i = \boldsymbol{x}_i^\top\boldsymbol{\beta} + (1 + |\boldsymbol{x}_i^\top\boldsymbol{\beta}|)^\theta\varepsilon_i$ and $Y_i = \boldsymbol{x}_i^\top\boldsymbol{\beta} + (1 + (\boldsymbol{x}_i^\top\boldsymbol{\beta})^2)^\theta\varepsilon_i$. It follows from Carroll and Ruppert (1982) that, in this general class of models, and provided a preliminary $\sqrt{n}-$consistent estimator of $\boldsymbol{\beta}$ can be constructed, robust weighted estimators of $\boldsymbol{\beta}$ and θ can be found under the design assumptions that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top}{\exp(2\theta h(\boldsymbol{x}_i^\top\boldsymbol{\beta}))} \to \boldsymbol{S}, \text{ positive definite, and } \frac{1}{n}\sum_{i=1}^{n}\{\|\boldsymbol{x}_i\|^2 + |h(\boldsymbol{x}_i^\top\boldsymbol{\beta})|^2\} = \mathrm{O}(1),$$

as well as

$$\max_{1\le i\le n}\{\|\boldsymbol{x}_i\| + |h(\boldsymbol{x}_i^\top\boldsymbol{\beta})|\} = \mathrm{o}(n^{1/2}),\ \inf_{i\ge 1}\exp(\theta h(\boldsymbol{x}_i^\top\boldsymbol{\beta})) > 0 \ \text{ and } \ \inf_{n\ge 1}\frac{1}{n}\sum_{i=1}^{n}\{h(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}^2 > 0,$$

see Theorems 1 and 2 in Carroll and Ruppert (1982). In this setting, m and $\log\sigma$ are respectively $m(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{x}^\top\boldsymbol{\beta}$ and $\log\sigma(\boldsymbol{x}, \boldsymbol{\beta}, \theta) = \theta h(\boldsymbol{x}^\top\boldsymbol{\beta})$. Then $M(\boldsymbol{x}) = \|\boldsymbol{x}\|$, and since

$$|\widehat{\theta}h(\boldsymbol{x}^\top\widehat{\boldsymbol{\beta}}) - \theta h(\boldsymbol{x}^\top\boldsymbol{\beta})| \le |\widehat{\theta}||h(\boldsymbol{x}^\top\widehat{\boldsymbol{\beta}}) - h(\boldsymbol{x}^\top\widehat{\boldsymbol{\beta}})| + |\widehat{\theta} - \theta||h(\boldsymbol{x}^\top\boldsymbol{\beta})|,$$

it readily follows from the Lipschitz property of h that one can choose $S(\boldsymbol{x}) = C(\|\boldsymbol{x}\| + |h(\boldsymbol{x}^\top\boldsymbol{\beta})|)$ for a suitable positive constant C. Fulfilling the assumptions of Corollary 4.1 then only requires imposing $\max_{1\le i\le n}\{\|\boldsymbol{x}_i\| + |h(\boldsymbol{x}_i^\top\boldsymbol{\beta})|\} = \mathrm{o}(n^{1/2-\delta})$ for some $\delta > 0$, rather than the slightly weaker $\mathrm{o}(n^{1/2})$ as above. A preliminary $\sqrt{n}-$consistent estimator of $\boldsymbol{\beta}$, necessary for the construction of the robust weighted estimators of $\boldsymbol{\beta}$ and θ, can be, for instance, obtained through L^1-(median) regression, see Knight (1999), under the assumption that the ε_i have a probability density function w.r.t. the

Lebesgue measure which is positive at 0 and provided further mild conditions on the design hold.

It is important to note that the construction of the residuals in Corollary 4.1 requires observing the covariate values $\boldsymbol{X}_i$. This result therefore does not apply when part of the covariate values are unobserved, as would be the case in time series models such as ARMA and GARCH. The general construction of $C_i^{(n)}$ we have introduced here, obtained by bounding $|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|/(q(\tau_n) + |\varepsilon_i|)$ from above and then isolating terms that are independent of ε_i, will however inspire their construction in such models where covariate values need to be estimated, as we illustrate next in ARMA models.

4.2 ARMA time series

Consider the ARMA(p,q) model $Y_t = \sum_{j=1}^p \phi_j Y_{t-j} - \sum_{j=1}^q \psi_j \varepsilon_{t-j} + \varepsilon_t$, for $t \in \mathbb{Z}$, where $\phi_1, \ldots, \phi_p$ and $\psi_1, \ldots, \psi_q$ are unknown real-valued coefficients. It is assumed that the polynomials $P(z) = 1 - \sum_{j=1}^p \phi_j z^j$ and $Q(z) = 1 - \sum_{j=1}^q \psi_j z^j$ have no common root, and no root inside the closed unit disk of the complex plane, and that (ε_t) is an i.i.d. sequence of centered heavy-tailed innovations having a finite variance (*i.e.* $\gamma < 1/2$) and with extreme value index γ, such that $\varepsilon \equiv \varepsilon_0$ satisfies $\mathbb{P}(\varepsilon > x)/\mathbb{P}(|\varepsilon| > x) \to l \in (0,1]$ as $x \to \infty$ (meaning that the left tail of ε cannot dominate its right tail). We assume that observations $(Y_{1-p}, \ldots, Y_0, Y_1, \ldots, Y_n)$ are available from this model.

This ARMA model is causal and invertible. The parameter vector $(\phi_1, \ldots, \phi_p, \psi_1, \ldots, \psi_q)$ can be estimated by $\sqrt{n}-$consistent estimators, such as an ordinary least squares estimator or a pseudo-maximum Gaussian likelihood estimator, see Theorem 10.8.2 in Brockwell and Davis (1991). We may therefore assume here that an estimator $(\widehat{\phi}_{n,1}, \ldots, \widehat{\phi}_{n,p}, \widehat{\psi}_{n,1}, \ldots, \widehat{\psi}_{n,q})$ is given such that

$$\forall i \in \{1, \ldots, p\},\ n^{1/2}|\widehat{\phi}_{n,i} - \phi_{n,i}| = \mathrm{O}_{\mathbb{P}}(1) \text{ and } \forall j \in \{1, \ldots, q\},\ n^{1/2}|\widehat{\psi}_{n,j} - \psi_{n,j}| = \mathrm{O}_{\mathbb{P}}(1).$$

The residuals from the model are

$$\widehat{\varepsilon}_t^{(n)} = Y_t - \sum_{j=1}^p \widehat{\phi}_{n,j} Y_{t-j} + \sum_{j=1}^q \widehat{\psi}_{n,j} \widehat{\varepsilon}_{t-j}^{(n)},\ \text{for all } 1 \le t \le n,$$

with $\widehat{\varepsilon}_{1-q}^{(n)}, \ldots, \widehat{\varepsilon}_0^{(n)}$ arbitrary choices of the residuals at times $1-q, \ldots, 0$. Typically in practice, these random variables $\widehat{\varepsilon}_{1-q}^{(n)} \equiv \widehat{\varepsilon}_{1-q}, \ldots, \widehat{\varepsilon}_0^{(n)} \equiv \widehat{\varepsilon}_0$ may be taken to be constant (for instance, equal to 0), although we shall only assume below that each of them is independent of the ARMA observations and has a finite moment of order 1.

The key step in order to get a good grasp of the difference $|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|$ is to express the ARMA equations

and the residuals in vector form. This can be done by noting that $\boldsymbol{Y}_t = \boldsymbol{A}\boldsymbol{Y}_{t-1} - \boldsymbol{B}\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t$ with

$$\boldsymbol{Y}_t = \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-r+1} \end{pmatrix} \text{ and } \boldsymbol{A} = \begin{pmatrix} \phi_1 & \cdots & \cdots & \cdots & \phi_r \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix},$$

$$\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-r+1} \end{pmatrix} \text{ and } \boldsymbol{B} = \begin{pmatrix} \psi_1 & \cdots & \cdots & \cdots & \psi_r \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix},$$

where $r = \max(p, q)$ and with the convention $\phi_j = 0$ for $p+1 \leq j \leq r$ and $\psi_j = 0$ for $q+1 \leq j \leq r$; likewise, $\widehat{\boldsymbol{\varepsilon}}_t^{(n)} = \boldsymbol{Y}_t - \widehat{\boldsymbol{A}}_n \boldsymbol{Y}_{t-1} + \widehat{\boldsymbol{B}}_n \widehat{\boldsymbol{\varepsilon}}_{t-1}^{(n)}$ where

$$\widehat{\boldsymbol{\varepsilon}}_t^{(n)} = \begin{pmatrix} \widehat{\varepsilon}_t^{(n)} \\ \widehat{\varepsilon}_{t-1}^{(n)} \\ \vdots \\ \widehat{\varepsilon}_{t-r+1}^{(n)} \end{pmatrix}, \ \widehat{\boldsymbol{A}}_n = \begin{pmatrix} \widehat{\phi}_{1,n} & \cdots & \cdots & \cdots & \widehat{\phi}_{r,n} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}, \ \widehat{\boldsymbol{B}}_n = \begin{pmatrix} \widehat{\psi}_{1,n} & \cdots & \cdots & \cdots & \widehat{\psi}_{r,n} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

with the convention $\widehat{\phi}_{j,n} = 0$ for $p+1 \leq j \leq r$ and $\widehat{\psi}_{j,n} = 0$, $\widehat{\varepsilon}_{1-j}^{(n)} = 0$ for $q+1 \leq j \leq r$. [A similar vector form representation is used in the proof of Theorem 3.3 in Girard et al. (2021), although theirs is incorrect when $p \neq q$. Their proof remains valid if the above correct representation is used.] Then the identity $\widehat{\boldsymbol{\varepsilon}}_t^{(n)} - \boldsymbol{\varepsilon}_t = (\boldsymbol{A} - \widehat{\boldsymbol{A}}_n)\boldsymbol{Y}_{t-1} - (\boldsymbol{B} - \widehat{\boldsymbol{B}}_n)\boldsymbol{\varepsilon}_{t-1} + \widehat{\boldsymbol{B}}_n(\widehat{\boldsymbol{\varepsilon}}_{t-1}^{(n)} - \boldsymbol{\varepsilon}_{t-1})$ yields recursively

$$\forall t \in \{1, \ldots, n\}, \ \widehat{\boldsymbol{\varepsilon}}_t^{(n)} - \boldsymbol{\varepsilon}_t = \sum_{j=1}^{t} (\boldsymbol{A} - \widehat{\boldsymbol{A}}_n)\widehat{\boldsymbol{A}}_n^{j-1}\boldsymbol{Y}_{t-j} - \sum_{j=1}^{t} (\boldsymbol{B} - \widehat{\boldsymbol{B}}_n)\widehat{\boldsymbol{B}}_n^{j-1}\boldsymbol{\varepsilon}_{t-j} + \widehat{\boldsymbol{B}}_n^t(\widehat{\varepsilon}_0 - \boldsymbol{\varepsilon}_0).$$

Let $\boldsymbol{w} \in \mathbb{R}^r$ be the vector having first entry 1 and all other entries equal to 0. Then

$$\begin{aligned} &\sqrt{n(1-\tau_n)}\frac{|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|}{q(\tau_n) + |\varepsilon_t|} \leq \frac{\sqrt{n(1-\tau_n)}}{q(\tau_n) + |\varepsilon_t|}\|\widehat{\boldsymbol{\varepsilon}}_t^{(n)} - \boldsymbol{\varepsilon}_t\| \\ &\leq \frac{\sqrt{n(1-\tau_n)}}{q(\tau_n) + |\varepsilon_t|}\left\{\|\widehat{\boldsymbol{A}}_n - \boldsymbol{A}\|\sum_{j=0}^{t-1}\|\widehat{\boldsymbol{A}}_n^j\|\|\boldsymbol{Y}_{t-j-1}\| + \|\widehat{\boldsymbol{B}}_n - \boldsymbol{B}\|\sum_{j=0}^{t-1}\|\widehat{\boldsymbol{B}}_n^j\|\|\boldsymbol{\varepsilon}_{t-j-1}\| + \|\widehat{\boldsymbol{B}}_n^t\|\|\widehat{\varepsilon}_0 - \varepsilon_0\|\right\} \end{aligned}$$

where $\|\cdot\|$ indifferently denotes the supremum norm on $\mathbb{R}^r$ and the induced operator norm $\|\boldsymbol{M}\| = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \|\boldsymbol{M}\boldsymbol{x}\|/\|\boldsymbol{x}\|$ on the space of $r \times r$ matrices. Based on this last inequality, and mimicking the construction in Section 4.1, the intuition behind the construction of a valid $\sqrt{n(1-\tau_n)}$−small array $(C_t^{(n)})$ here is that $C_t^{(n)}$ should be chosen to be only slightly larger than the right-hand side above (with arbitrarily high probability as $n \to \infty$), while being independent of ε_t. The main difficulty is

to handle the random matrix norms $\|\widehat{\boldsymbol{A}}_n^{j-1}\|$ and $\|\widehat{\boldsymbol{B}}_n^{j-1}\|$, which are not independent of ε_t. The key point is to note that the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are essentially companion matrices for the polynomials P and Q, with in fact

$$\det(\lambda \mathrm{I}_r - \boldsymbol{A}) = \lambda^r P(1/\lambda) \quad \text{and} \quad \det(\lambda \mathrm{I}_r - \boldsymbol{B}) = \lambda^r Q(1/\lambda).$$

Since P and Q have all their roots outside the closed unit disk, all the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{B}$ must have modulus less than 1, *i.e.* the spectral radii $\rho(\boldsymbol{A})$ of $\boldsymbol{A}$ and $\rho(\boldsymbol{B})$ of $\boldsymbol{B}$ are smaller than 1. Moreover, by Gelfand's spectral radius formula, $\|\boldsymbol{A}^N\|^{1/N} \to \rho(\boldsymbol{A})$ and $\|\boldsymbol{B}^N\|^{1/N} \to \rho(\boldsymbol{B})$ as $N \to \infty$. Conclude that there are $\delta > 0$ and $N_0 \geq 1$ such that $\max(\|\boldsymbol{A}^{N_0}\|, \|\boldsymbol{B}^{N_0}\|) \leq 1 - 2\delta$. Note further that any integer j can be written in the form $j = N_0\lfloor j/N_0 \rfloor + r$, with $r = j - N_0\lfloor j/N_0 \rfloor \in \{0, \ldots, N_0 - 1\}$, yielding the bound

$$\begin{aligned}
&\sqrt{n(1-\tau_n)}\frac{|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|}{q(\tau_n) + |\varepsilon_t|} \\
&\leq \frac{\sqrt{n(1-\tau_n)}}{q(\tau_n)}\Bigg\{ \|\widehat{\boldsymbol{A}}_n - \boldsymbol{A}\| \max(1, \|\widehat{\boldsymbol{A}}_n\|^{N_0-1}) \sum_{j=0}^{t-1} \|\widehat{\boldsymbol{A}}_n^{N_0}\|^{\lfloor j/N_0 \rfloor} \|\boldsymbol{Y}_{t-j-1}\| \\
&\qquad + \|\widehat{\boldsymbol{B}}_n - \boldsymbol{B}\| \max(1, \|\widehat{\boldsymbol{B}}_n\|^{N_0-1}) \sum_{j=0}^{t-1} \|\widehat{\boldsymbol{B}}_n^{N_0}\|^{\lfloor j/N_0 \rfloor} \|\boldsymbol{\varepsilon}_{t-j-1}\| \\
&\qquad + \max(1, \|\widehat{\boldsymbol{B}}_n\|^{N_0-1}) \|\widehat{\boldsymbol{B}}_n^{N_0}\|^{\lfloor t/N_0 \rfloor} \|\widehat{\boldsymbol{\varepsilon}}_0 - \boldsymbol{\varepsilon}_0\| \Bigg\}.
\end{aligned}$$

Now $\|\widehat{\boldsymbol{A}}_n^{N_0}\| \to \|\boldsymbol{A}^{N_0}\| \leq 1 - 2\delta$ and $\|\widehat{\boldsymbol{B}}_n^{N_0}\| \to \|\boldsymbol{B}^{N_0}\| \leq 1 - 2\delta$ in probability, and $\|\widehat{\boldsymbol{A}}_n - \boldsymbol{A}\| + \|\widehat{\boldsymbol{B}}_n - \boldsymbol{B}\| = \mathrm{O}_{\mathbb{P}}(n^{-1/2})$, thus suggesting to construct $C_t^{(n)}$ as

$$C_t^{(n)} = u_n \frac{\sqrt{n(1-\tau_n)}}{q(\tau_n)} \left\{ \frac{1}{n^{1/2}} \sum_{j=0}^{t-1} (1-\delta)^{\lfloor j/N_0 \rfloor} (\|\boldsymbol{Y}_{t-j-1}\| + \|\boldsymbol{\varepsilon}_{t-j-1}\|) + (1-\delta)^{\lfloor t/N_0 \rfloor} \|\widehat{\boldsymbol{\varepsilon}}_0 - \boldsymbol{\varepsilon}_0\| \right\}$$

where (u_n) is a nonrandom sequence that diverges to infinity slowly enough. The sequence $\sum_{j=0}^{t-1}(1-\delta)^{\lfloor j/N_0 \rfloor}$, $t \geq 1$, is bounded and increases to N_0/δ. Then, by construction, $C_t^{(n)}$ is independent of ε_t for each $n \geq 1$ and $t \in \{1, \ldots, n\}$ (since $\widehat{\boldsymbol{\varepsilon}}_0$ is chosen independently from the ARMA observations).

More precisely, assume that $u_n/q(\tau_n) \to 0$. One has

$$\begin{aligned}
\frac{1}{\sqrt{n(1-\tau_n)}} \max_{1 \leq t \leq n} C_t^{(n)} &= \mathrm{O}_{\mathbb{P}}\left(\frac{u_n}{q(\tau_n)} \left\{ 1 + \frac{1}{\sqrt{n}} \max_{1-r \leq i \leq n-1} (|Y_i| + |\varepsilon_i|) \right\} \right) \\
&= \mathrm{O}_{\mathbb{P}}\left(\frac{u_n}{q(\tau_n)} \left\{ 1 + \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} (|Y_i| + |\varepsilon_i|) \right\} \right).
\end{aligned}$$

By causality of $(Y_t)_{t \in \mathbb{Z}}$, the Y_t have the linear representation $Y_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, and it is a consequence of the arguments in the proof of Theorem 3.1.1 in Brockwell and Davis (1991) that the coefficients a_j define a summable series and decay geometrically fast, *i.e.* $|a_j| \leq CR^j$ for real constants $C > 0$ and

$R \in (0,1)$. Then for any $1 \leq t \leq n$,

$$|Y_t| \leq \sum_{j=0}^{t-1} |a_j||\varepsilon_{t-j}| + \sum_{j=t}^{\infty} |a_j||\varepsilon_{t-j}| \leq \left(\sum_{j=0}^{\infty} |a_j| \right) \max_{1 \leq t \leq n} |\varepsilon_t| + CR \sum_{k=0}^{\infty} R^k |\varepsilon_{-k}|.$$

The last sum on the right-hand side is finite with probability 1 because ε has a finite first moment. Conclude that

$$\max_{1 \leq i \leq n} |Y_i| = \mathrm{O}_{\mathbb{P}} \left(\max_{1 \leq i \leq n} |\varepsilon_i| \right).$$

[The bound $|Y_0| \leq C \sum_{j=0}^{\infty} R^j |\varepsilon_{-j}|$ also implies that Y_0 is integrable.] Now by pairing Theorem 1.1.6 p.10 and Lemma 1.2.9 p.22 in de Haan and Ferreira (2006), one obtains

$$\frac{1}{\gamma} \left(\frac{\max_{1 \leq i \leq n} \varepsilon_i}{q(1-1/n)} - 1 \right) \xrightarrow{\mathrm{d}} \mathrm{GEV}(\gamma).$$

Then, using condition $\mathbb{P}(\varepsilon > x)/\mathbb{P}(|\varepsilon| > x) \to l \in (0,1]$ as $x \to \infty$, one gets, for any $t > 0$,

$$\begin{aligned} \log \mathbb{P} \left(\frac{\max_{1 \leq i \leq n} |\varepsilon_i|}{q(1-1/n)} \leq t \right) &= n \log \left\{ 1 - \left(\frac{1}{l} + \mathrm{o}(1) \right) \mathbb{P} \left(\frac{\varepsilon}{q(1-1/n)} > t \right) \right\} \\ &\sim -\frac{n}{l} \mathbb{P} \left(\frac{\varepsilon}{q(1-1/n)} > t \right) \\ &\sim \frac{1}{l} \log \mathbb{P} \left(\frac{\max_{1 \leq i \leq n} \varepsilon_i}{q(1-1/n)} \leq t \right) \text{ as } n \to \infty. \end{aligned}$$

Consequently

$$\frac{\max_{1 \leq i \leq n} |\varepsilon_i|}{q(1-1/n)} = \mathrm{O}_{\mathbb{P}}(1).$$

Finally, from Potter bounds applied to the function $t \mapsto q(1-1/t)$, see *e.g.* Proposition B.1.9.5 p.367 in de Haan and Ferreira (2006), one gets

$$\max_{1 \leq i \leq n} |\varepsilon_i| = \mathrm{O}_{\mathbb{P}}(n^{\gamma+\iota}) \text{ for any } \iota > 0.$$

This yields

$$\frac{1}{\sqrt{n(1-\tau_n)}} \max_{1 \leq t \leq n} C_t^{(n)} = \mathrm{O}_{\mathbb{P}} \left(\frac{u_n}{q(\tau_n)} \right) = \mathrm{o}_{\mathbb{P}}(1).$$

Also, for any positive nonrandom sequence (δ_n) converging to 0, obviously

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}(C_t^{(n)} \mathbb{1}\{C_t^{(n)} \leq \delta_n \sqrt{n(1-\tau_n)}\}) \leq \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}(C_t^{(n)})$$

and therefore

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}(C_t^{(n)}\mathbb{1}\{C_t^{(n)}\leq\delta_n\sqrt{n(1-\tau_n)}\})\\
&\leq u_n\frac{\sqrt{1-\tau_n}}{q(\tau_n)}\times\frac{1}{n}\sum_{t=1}^{n}\left\{\sum_{j=0}^{t-1}(1-\delta)^{\lfloor j/N_0\rfloor}\mathbb{E}(\|\boldsymbol{Y}_0\|+\|\boldsymbol{\varepsilon}_0\|)+\sqrt{n}(1-\delta)^{\lfloor t/N_0\rfloor}\mathbb{E}(\|\widehat{\boldsymbol{\varepsilon}}_0-\boldsymbol{\varepsilon}_0\|)\right\}\\
&\leq u_n\frac{\sqrt{1-\tau_n}}{q(\tau_n)}\left(\frac{N_0}{\delta}\mathbb{E}(\|\boldsymbol{Y}_0\|+\|\boldsymbol{\varepsilon}_0\|)+\frac{1}{\sqrt{n}}\frac{N_0}{\delta}\mathbb{E}(\|\widehat{\boldsymbol{\varepsilon}}_0-\boldsymbol{\varepsilon}_0\|)\right)=\mathrm{O}_{\mathbb{P}}\left(u_n\frac{\sqrt{1-\tau_n}}{q(\tau_n)}\right)=\mathrm{o}_{\mathbb{P}}(1)
\end{aligned}$$

from the stationarity and integrability of the sequence (Y_t). It follows that $(C_t^{(n)})$ is a $\sqrt{n(1-\tau_n)}-$small array.

Finally, recall that $\|\widehat{\boldsymbol{A}}_n^{N_0}\|\to\|\boldsymbol{A}^{N_0}\|\leq 1-2\delta$ and $\|\widehat{\boldsymbol{B}}_n^{N_0}\|\to\|\boldsymbol{B}^{N_0}\|\leq 1-2\delta$ in probability, to obtain that in particular $\|\widehat{\boldsymbol{A}}_n^{N_0}\|\leq 1-\delta$ and $\|\widehat{\boldsymbol{B}}_n^{N_0}\|\leq 1-\delta$ with arbitrarily high probability as $n\to\infty$, and therefore

$$\begin{aligned}
&\max_{1\leq t\leq n}\frac{1}{C_t^{(n)}}\frac{\sqrt{n(1-\tau_n)}|\widehat{\varepsilon}_t^{(n)}-\varepsilon_t|}{q(\tau_n)+|\varepsilon_t|}\\
&\leq\frac{1}{u_n}\max_{1\leq t\leq n}\left(n^{1/2}\|\widehat{\boldsymbol{A}}_n-\boldsymbol{A}\|\max(1,\|\widehat{\boldsymbol{A}}_n\|^{N_0-1})\frac{\sum_{j=0}^{t-1}\|\widehat{\boldsymbol{A}}_n^{N_0}\|^{\lfloor j/N_0\rfloor}\|\boldsymbol{Y}_{t-j-1}\|}{\sum_{j=0}^{t-1}(1-\delta)^{\lfloor j/N_0\rfloor}\|\boldsymbol{Y}_{t-j-1}\|}\right.\\
&\qquad+n^{1/2}\|\widehat{\boldsymbol{B}}_n-\boldsymbol{B}\|\max(1,\|\widehat{\boldsymbol{B}}_n\|^{N_0-1})\frac{\sum_{j=0}^{t-1}\|\widehat{\boldsymbol{B}}_n^{N_0}\|^{\lfloor j/N_0\rfloor}\|\boldsymbol{\varepsilon}_{t-j-1}\|}{\sum_{j=0}^{t-1}(1-\delta)^{\lfloor j/N_0\rfloor}\|\boldsymbol{\varepsilon}_{t-j-1}\|}\\
&\qquad\left.+\max(1,\|\widehat{\boldsymbol{B}}_n\|^{N_0-1})\left(\frac{\|\widehat{\boldsymbol{B}}_n^{N_0}\|}{1-\delta}\right)^{\lfloor t/N_0\rfloor}\right)\\
&=\mathrm{O}_{\mathbb{P}}\left(\frac{1}{u_n}\left\{n^{1/2}\|\widehat{\boldsymbol{A}}_n-\boldsymbol{A}\|+n^{1/2}\|\widehat{\boldsymbol{B}}_n-\boldsymbol{B}\|+1\right\}\right)=\mathrm{o}_{\mathbb{P}}(1).
\end{aligned}$$

This yields the following corollary about the intermediate quantile estimator $\widehat{q}_n(1-k/n)$, Hill estimator $\widehat{\gamma}_n(k)$, and Weissman extreme quantile estimator $\widehat{q}_n^{\star}(\tau_n')$ constructed upon the residuals $\widehat{\varepsilon}_1^{(n)},\ldots,\widehat{\varepsilon}_n^{(n)}$.

Corollary 4.2 (Dynamic extreme value estimation in the ARMA model)**.** *Work in the causal and invertible ARMA(p,q) model $Y_t=\sum_{j=1}^{p}\phi_jY_{t-j}-\sum_{j=1}^{q}\psi_j\varepsilon_{t-j}+\varepsilon_t$, for $t\in\mathbb{Z}$, where $\phi_1,\ldots,\phi_p$ and $\psi_1,\ldots,\psi_q$ are real-valued coefficients such that the polynomials $P(z)=1-\sum_{j=1}^{p}\phi_jz^j$ and $Q(z)=1-\sum_{j=1}^{q}\psi_jz^j$ have no common root, and no root inside the closed unit disk of the complex plane, and (ε_t) is an independent and identically distributed sequence such that $\varepsilon\equiv\varepsilon_0$ satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$ with $\gamma<1/2$, and $\mathbb{P}(\varepsilon>x)/\mathbb{P}(|\varepsilon|>x)\to l\in(0,1]$ as $x\to\infty$. Let $k=k(n)\to\infty$ satisfy $k/n\to 0$ and $\sqrt{k}A(n/k)\to\lambda\in\mathbb{R}$.*

• *If $\widehat{\phi}_{n,1},\ldots,\widehat{\phi}_{n,p},\widehat{\psi}_{n,1},\ldots,\widehat{\psi}_{n,q}$ are $n^{1/2}-$consistent estimators of $\phi_1,\ldots,\phi_p,\psi_1,\ldots,\psi_q$, and $\widehat{\varepsilon}_t^{(n)}=Y_t-\sum_{j=1}^{p}\widehat{\phi}_{n,j}Y_{t-j}+\sum_{j=1}^{q}\widehat{\psi}_{n,j}\widehat{\varepsilon}_{t-j}^{(n)}$ are the pertaining residuals with initial values $\widehat{\varepsilon}_{1-q}^{(n)}\equiv\widehat{\varepsilon}_{1-q},\ldots,\widehat{\varepsilon}_0^{(n)}\equiv\widehat{\varepsilon}_0$ that have a finite moment of order 1 and are determined independently of the sequence (Y_t), then $(\sqrt{k}(\widehat{\gamma}_n(k)-\gamma),\sqrt{k}\log(\widehat{q}_n(1-k/n)/q(1-k/n)))$ converges weakly to a pair of independent Gaussian random variables having respective means $\lambda/(1-\rho)$ and 0, and the same variance γ^2.*

• *If moreover* $\rho < 0$*, and* $\tau'_n \uparrow 1$ *is such that* $k/(n(1-\tau'_n)) \to \infty$ *and* $\sqrt{k}/\log(k/(n(1-\tau'_n))) \to \infty$*, then*

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

This example makes essential use of the randomness in the $C_t^{(n)}$ in order to handle the very large fluctuations of the heavy-tailed ARMA process; we did not manage to construct a nonrandom version $c_t^{(n)}$ of this sequence allowing us to check each condition without introducing further restrictions on γ or τ_n. Corollary 4.2 constitutes a substantial improvement upon results such as Theorem 3.3 in Girard et al. (2021) or Lemma 2 in Hill (2015) in the sense that it does not impose any restriction on k (or equivalently $\tau_n = 1 - k/n$) or on the second-order properties of ε_t (see condition (3.1) in Hill, 2015).

Acknowledgments

G. Stupfler acknowledges financial support from the French National Research Agency under the grants ANR-23-CE40-0009 (EXSTA project), ANR-19-CE40-0013 (ExtremReg project) and ANR-11-LABX-0020-01 (Centre Henri Lebesgue), as well as from the TSE-HEC ACPR Chair "Regulation and systemic risks", an AXA Research Fund Award on "Mitigating risk in the wake of the COVID-19 pandemic", and the Chair Stress Test, RISK Management and Financial Steering of the Foundation Ecole Polytechnique.

References

Ahmad, A., Deme, E. H., Diop, A., Girard, S., and Usseglio-Carleve, A. (2020). Estimation of extreme quantiles from heavy-tailed distributions in a location-dispersion regression model. *Electronic Journal of Statistics*, 14(2):4421–4456.

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

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods (second edition)*. Springer.

Carroll, R. J. and Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. *Annals of Statistics*, 10(2):429–441.

Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C*, 54(1):207–222.

Chernozhukov, V., Fernández-Val, I., and Kaji, T. (2018). Extremal quantile regression. In Koenker, R., Chernozhukov, V., He, X., and Peng, L., editors, *Handbook of Quantile Regression*, pages 333–362. Chapman and Hall/CRC.

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. *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 and 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.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B*, 52(3):393–442.

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.

D'Haultfoeuille, X., Maurel, A., and Zhang, Y. (2018). Extremal quantile regressions for selection models and the black–white wage gap. *Journal of Econometrics*, 203(1):129–142.

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

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

Farkas, S., Heranval, A., Lopez, O., and Thomas, M. (2024). Generalized Pareto regression trees for extreme event analysis. *Extremes*, 27(3):437–477.

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.

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

Gnecco, N., Terefe, E. M., and Engelke, S. (2024). Extremal random forests. *Journal of the American Statistical Association*, to appear.

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

Hill, J. B. (2015). Tail index estimation for a filtered dependent time series. *Statistica Sinica*, 25(2):609–629.

Hoga, Y. (2019). Confidence intervals for conditional tail risk measures in ARMA–GARCH models. *Journal of Business & Economic Statistics*, 37(4):613–624.

Kinsvater, P. and Fried, R. (2017). Conditional heavy-tail behavior with applications to precipitation and river flow extremes. *Stochastic Environmental Research and Risk Assessment*, 31(5):1155–1169.

Knight, K. (1999). Asymptotics for L_1-estimators of regression parameters under heteroscedasticity. *Canadian Journal of Statistics*, 27(3):497–507.

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

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

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

Russell, B. T., Cooley, D. S., Porter, W. C., Reich, B. J., and Heald, C. L. (2016). Data mining to investigate the meteorological drivers for extreme ground level ozone events. *The Annals of Applied Statistics*, 10(3):1673–1698.

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

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

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.