



HAL
open science

Residual-based extreme value estimation for heavy tails

Marco Oesting, Gilles Stupfler

► **To cite this version:**

Marco Oesting, Gilles Stupfler. Residual-based extreme value estimation for heavy tails. 2025. ⟨hal-04705855v2⟩

HAL Id: hal-04705855

<https://hal.science/hal-04705855v2>

Preprint submitted on 24 Oct 2025

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

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



HAL Authorization

Residual-based extreme value estimation for heavy tails

Marco Oesting^a & Gilles Stupfler^b

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

^b Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France

Abstract

We consider the estimation of, and inference about, conditional extreme quantiles and related quantities based on non-i.i.d. random variables, termed residuals throughout. This is motivated by the recent use of location-scale regression models to support several extremal regression methods, where the extreme value step is carried out using the residuals of the model, rather than the unobserved i.i.d. errors, following a preliminary model estimation step. We show that residual-based versions of a certain large class of extreme value estimators are asymptotically normal under a high-level condition on the number of residuals considered to be too far from their i.i.d. counterparts. This condition is shown to be substantially weaker than analogous conditions in the recent literature, and we discuss how our theory applies to a wide range of examples containing location-scale parametric regression models, multivariate autoregressive time series, and noisy data models. Finite-sample simulation studies and real data examples illustrate the performance of our estimators and of the asymptotic confidence intervals derived from our theoretical results.

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

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

1 Introduction

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 \mathbf{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 socio-economic 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 $\mathbf{X} = \mathbf{x}$ at a level τ close to 1, which will be denoted in the sequel as $q_{Y|\mathbf{X}}(\tau|\mathbf{x})$. The literature has mainly attacked the extreme quantile regression problem using two types 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 dimensions.

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 $Y = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$. Here m and $\sigma > 0$ are two measurable functions of a covariate $\mathbf{X} \in \mathbb{R}^p$, and the unobserved error ε is independent of \mathbf{X} . This ensures that $q_{Y|\mathbf{X}}(\tau|\mathbf{x}) = m(\mathbf{x}) + \sigma(\mathbf{x})q(\tau)$, where $q(\tau)$ denotes the quantile of ε at level τ . Given data (Y_i, \mathbf{X}_i) , $1 \leq i \leq n$ from the model, the estimation of the location and scale functions by \hat{m}_n and $\hat{\sigma}_n$, say (whose exact construction depends on the model considered), is viewed as a preliminary step before estimating unconditional extreme quantiles $q(\tau)$ of ε by plugging the residuals $\hat{\varepsilon}_i^{(n)} = (Y_i - \hat{m}_n(\mathbf{X}_i))/\hat{\sigma}_n(\mathbf{X}_i)$ instead of the unobserved errors $\varepsilon_i = (Y_i - m(\mathbf{X}_i))/\sigma(\mathbf{X}_i)$ into an existing extreme quantile estimator (see Chapter 4 in de Haan and Ferreira, 2006, for a review of extreme quantile estimation). The intuition of McNeil and Frey (2000) is that since the whole sample of data can be used to estimate m and σ , while extreme quantile estimation 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 if it were 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 for the resulting conditional extreme quantile estimators that are reminiscent of those found in standard i.i.d. theory. The crucial point, however, is that these residuals will typically not be i.i.d. even if the unobserved errors are. Two interesting methodological problems are therefore the following:

- (i) Prove that residual-based versions of extreme value estimators indeed share the asymptotic behavior of their unachievable counterparts based on i.i.d. data.
- (ii) Construct asymptotic confidence intervals deduced from the theory of residual-based estimators which have finite-sample coverage close to the nominal level.

The first problem is 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. The second problem has been left largely untouched, although Martins-Filho et al. (2018) report fairly poor coverage probabilities when m and σ are estimated by nonparametric regression, and Girard et al. (2021) find the asymptotic Gaussian confidence intervals for the companion notion of conditional extreme expectiles to be unreliable.

Contribution of the paper Assume for now that the available data is a sample $(\hat{\varepsilon}_1^{(n)}, \dots, \hat{\varepsilon}_n^{(n)})$, viewed as approximating a sample of unobserved i.i.d. random variables $(\varepsilon_1, \dots, \varepsilon_n)$. The random variables $\hat{\varepsilon}_1^{(n)}, \dots, \hat{\varepsilon}_n^{(n)}$ are not assumed to be i.i.d. and may change with n . We call the $\hat{\varepsilon}_i^{(n)}$ the

residuals and the ε_i the *errors*, by analogy with the regression problem, and we consider the estimation of extreme value quantities linked to the distribution of $\varepsilon \equiv \varepsilon_1$ when it is heavy-tailed (*i.e.* Paretian-tailed). This construction is relevant to location-scale regression analysis, but also to other problems, such as when the data points are recorded subject to measurement error or corrupted. This then fits into the wider context of the extreme value analysis of dependent and nonstationary data.

Our contribution is essentially twofold. First, we show that one can substantially weaken the assumptions of Girard et al. (2021) and move away from uniform-type conditions when proving the asymptotic normality of residual-based extreme value estimators such as the Hill and Weissman estimators. This is done by exploiting 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 is nothing but another integral over the tail of this same empirical distribution function. This reduces our problem to the question of finding appropriate conditions under which the tail of the empirical distribution function of the residuals is sufficiently close to that of $(\varepsilon_1, \dots, \varepsilon_n)$, which prompts us to introduce a condition on the number of residuals that do not correctly predict the unobserved errors. Our main theoretical results then state that, if this number is small enough, then residual-based intermediate order statistics, Hill and Weissman estimators have the same asymptotic distributions as their counterparts based on i.i.d. data. This leads to a theoretical framework that can handle a wide range of regression, time series and noisy data models without having to impose unnecessary technical restrictions as in Girard et al. (2021), while being able to handle the joint estimation of intermediate and extreme quantiles and their weighted integrals in a multivariate setting.

Second, we show that the asymptotic results we obtain translate into reasonable results at the finite-sample level in the models given as examples where our theory applies. We shall indeed observe that, at least in simple models such as location-scale regression and ARMA time series, the quality of point estimates of the extreme value index given by the residual-based Hill estimator is essentially identical to that of the point estimates given by its unfeasible counterpart based on the unobserved errors, when the model structure is specified correctly. We will also illustrate that, in regression settings, over-specifying the regression model used in order to filter the data is not detrimental to the quality of the estimation procedure, and that, in situations where the tail is very heavy and for extreme value index estimation specifically, under-specifying the model still improves upon not filtering the data at all. This translates into conditional extreme quantile estimates whose accuracy is excellent when the model structure is correctly captured. Our theory also yields feasible and sensible asymptotic Gaussian confidence intervals for these quantities, in the sense that their coverage rates are generally close to those of the asymptotic Gaussian confidence intervals constructed using the unobservable errors, except of course in models whose structure is difficult to recover. This approach has lower statistical uncertainty than that of purely nonparametric techniques, whose applicability is restricted to low-dimensional settings, without being limited to models having a linear structure.

The outline of this article is the following. In Section 2, we provide a detailed exposition of our setting and modeling assumptions. Our main results are stated in Section 3. We discuss several examples of models that we cover in Section 4. A simulation study and real data analysis can be found in Section 5. Mathematical proofs and a complete set of finite-sample simulation results can be found in an online Supplementary Material document.

2 Model setting and assumptions

Let, for $1 \leq j \leq d$, the $n_j = n_j(n)$ be sequences of integers tending to infinity as $n \rightarrow \infty$. The available data is assumed to be made of samples of *residuals* $(\hat{\varepsilon}_{1,j}^{(n_j)}, \dots, \hat{\varepsilon}_{n_j,j}^{(n_j)})$, viewed as approximating samples of unobserved i.i.d. random variables $(\varepsilon_{1,j}, \dots, \varepsilon_{n_j,j})$ called *errors*, in a sense that will be made clear in Section 3. 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 $\bar{F}_j = 1 - F_j : t \mapsto \mathbb{P}(\varepsilon_j > t)$ respectively denote the distribution and survival functions of $\varepsilon_j \equiv \varepsilon_{1,j}$, and $q_j : \tau \mapsto \inf\{t \in \mathbb{R} \mid F_j(t) \geq \tau\}$ be the quantile function of ε_j . 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 $\bar{F}_j(tx)/\bar{F}_j(t) \rightarrow x^{-1/\gamma_j}$ as $t \rightarrow \infty$ for any $x > 0$. This assumption is generally appropriate to the modeling 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).

We seek to estimate extreme quantiles $q_j(1 - p_{n,j})$ of the ε_j , with the exceedance probability $p_{n,j} \downarrow 0$ such that $n_j(1 - p_{n,j})$ is bounded. If the $\varepsilon_{t,j}$ were observed, then one could use the Weissman extreme quantile estimator (Weissman, 1978), constructed as follows. Let $\varepsilon_{1:n_j,j} \leq \varepsilon_{2:n_j,j} \leq \dots \leq \varepsilon_{n_j:n_j,j}$ be the order statistics of the sample $(\varepsilon_{1,j}, \dots, \varepsilon_{n_j,j})$, and let $k_j = k_j(n) \rightarrow \infty$ with $k_j/n_j \rightarrow 0$ be an intermediate sequence of integers. The Weissman estimator is

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

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

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

where $\tilde{\bar{F}}_{n,j}$ denotes the empirical survival function constructed upon $(\varepsilon_{1,j}, \dots, \varepsilon_{n_j,j})$. The premise of this paper is that the $\varepsilon_{t,j}$ are not available but the residuals $\hat{\varepsilon}_{t,j}^{(n_j)}$ are, which suggests to introduce their order statistics $\hat{\varepsilon}_{1:n_j,j}^{(n_j)} \leq \hat{\varepsilon}_{2:n_j,j}^{(n_j)} \leq \dots \leq \hat{\varepsilon}_{n_j:n_j,j}^{(n_j)}$ and to use instead the residual-based estimators

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

and $\hat{\gamma}_{n,j} \equiv \hat{\gamma}_{n,j}(k_j) = \frac{1}{k_j} \sum_{i=1}^{k_j} \log \hat{\varepsilon}_{n_j - i + 1 : n_j, j}^{(n_j)} - \log \hat{\varepsilon}_{n_j - k_j : n_j, j}^{(n_j)} = \frac{n_j}{k_j} \int_{\hat{\varepsilon}_{n_j - k_j : n_j, j}^{(n_j)}}^{\infty} \frac{\hat{\bar{F}}_{n,j}(t)}{t} dt.$

Here $\hat{\bar{F}}_{n,j}$ denotes the empirical survival function constructed upon the residuals $(\hat{\varepsilon}_{1,j}^{(n_j)}, \dots, \hat{\varepsilon}_{n_j,j}^{(n_j)})$. Our objective is to analyze the asymptotic properties of $\hat{q}_{n,j}^*(1 - p_{n,j})$ and $\hat{\gamma}_{n,j}$ by relating them to those of the unfeasible $\tilde{q}_{n,j}^*(1 - p_{n,j})$ and $\tilde{\gamma}_{n,j}$, known to be asymptotically normal under classical conditions (see de Haan and Ferreira, 2006, Theorem 3.2.5 p.74 and Theorem 4.3.8 p.138). This will push us to discuss

the joint estimation of weighted integrals of tail quantiles, by comparing the estimator

$$\frac{n_j}{k_j} \int_1^\infty \widehat{F}_{n,j}^{\varepsilon_{n_j-k_j:n_j,j}}(x) \mu_j(dx) \quad \text{with the unfeasible} \quad \frac{n_j}{k_j} \int_1^\infty \widetilde{F}_{n,j}^{\varepsilon_{n_j-k_j:n_j,j}}(x) \mu_j(dx),$$

where the μ_j are weighting measures putting finite mass on compact sets.

Our first main modeling assumption for doing so, introduced to deal with the convergence of each extreme value estimator, is a second-order regular variation condition under which the right tail of the distribution of each ε_j is assumed to be close to a pure Pareto tail at an identified rate.

$\mathcal{C}_2(\boldsymbol{\gamma}, \boldsymbol{\rho}, \mathbf{A})$ For any $j \in \{1, \dots, d\}$, 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

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

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 Table 4 in Daouia et al. (2024b, Appendix C).

Our second modeling condition is a pairwise tail dependence condition specifically geared towards the obtention of the joint asymptotic normality of extreme value estimators. Recall that for any pair (j, ℓ) there exists, by Sklar's theorem, a (possibly not unique) *copula* function $C_{j,\ell}$, *i.e.* a bivariate 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 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 second condition is then as follows.

$\mathcal{J}(\mathbf{R})$ For all $j, \ell \in \{1, \dots, d\}$, there is a function $R_{j,\ell}$ such that $\lim_{s \rightarrow \infty} s \overline{C}_{j,\ell}(x_j/s, x_\ell/s) = R_{j,\ell}(x_j, x_\ell)$ 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 $\overline{C}_{j,\ell}$ is not uniquely determined because \overline{F}_j and/or \overline{F}_ℓ is not continuous, the limit should hold for one choice of $\overline{C}_{j,\ell}$.]

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). It is weaker than the classical multivariate regular variation assumption on $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_d)^\top$ (see Chapters 6 and 9 in Resnick, 2007).

3 Main results

3.1 Asymptotic behavior of high order statistics of residuals

The residual-based Weissman estimator extrapolates a single high order statistic $\widehat{q}_{n,j}(1 - k_j/n_j) = \widehat{\varepsilon}_{n_j-k_j:n_j,j}^{(n_j)}$, so it is reasonable to first consider the joint asymptotic behavior of intermediate quantile estimators $\widehat{q}_{n,j}(\tau_{n,j})$, where $\tau_{n,j} \rightarrow 1$ and $n_j(1 - \tau_{n,j}) \rightarrow \infty$. The crucial point is that if

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

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

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

with $\lfloor \cdot \rfloor$ denoting the floor function. It follows that, even though the residuals are not i.i.d., the estimator $\widehat{q}_{n,j}(\tau_{n,j})$ should have the same asymptotic distribution as its unfeasible counterpart $\widetilde{q}_{n,j}(\tau_{n,j})$, whose asymptotic behavior is well-understood, provided $\widehat{F}_{n,j}(x)$ is close enough to $\widetilde{F}_{n,j}(x)$. How close they should be, according to our 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 $(C_i^{(n)})_{1 \leq i \leq n}$ of positive random variables is said to be a u_n -small array if it satisfies $\max_{1 \leq i \leq n} C_i^{(n)} / u_n \xrightarrow{\mathbb{P}} 0$ and, for every positive nonrandom sequence (δ_n) converging to 0,

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

Loosely speaking, the (random) columns of a u_n -small array 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 high-level assumption on the residuals and errors, that will be at the heart of our theory. In this condition, $\tau_n \uparrow 1$ is such that $n(1 - \tau_n) \rightarrow \infty$ and μ is a measure on \mathbb{R} equipped with its Borel σ -algebra.

$\mathcal{H}(n, \tau_n, \mu)$ There is a $\sqrt{n(1 - \tau_n)}$ -small array $(C_i^{(n)})_{1 \leq i \leq n}$ such that $C_i^{(n)}$ is independent of ε_i for each $i \in \{1, \dots, n\}$ and $n \geq 1$, and the residuals $\widehat{\varepsilon}_i^{(n)}$ fulfill the convergence

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

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

We shall typically use this condition with a measure μ putting finite mass on compact Borel subsets of \mathbb{R} (e.g. a Radon measure), which guarantees that the above integral is a finite random variable (measurability follows from the Fubini-Tonelli theorem because μ is then necessarily σ -finite).

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 $\mathcal{C}_2(\gamma, \rho, \mathbf{A})$ and $\mathcal{J}(\mathbf{R})$ hold. Assume also that, for any $1 \leq j \leq d$:*

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

Then

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

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

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

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 different intermediate levels.

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

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

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

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

where $\mathbf{M}' = \mathbf{M}'(\boldsymbol{\theta})$ 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 consequence of Corollary 3.1 is that the residual-based order statistic $\widehat{q}_n(\tau_n) = \widehat{\varepsilon}_{[n\tau_n]:n}^{(n)}$ estimating the intermediate quantile $q(\tau_n)$ has the same asymptotic distribution as the intermediate empirical quantile $\widetilde{q}_n(\tau_n) = \varepsilon_{[n\tau_n]:n}$ from the sample of errors, that is,

$$\sqrt{n(1-\tau_n)} \left(\frac{\widehat{\varepsilon}_{[n\tau_n]:n}^{(n)}}{q(\tau_n)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2) \quad \text{and} \quad \sqrt{n(1-\tau_n)} \left(\frac{\varepsilon_{[n\tau_n]:n}}{q(\tau_n)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2)$$

when $\tau_n \uparrow 1$, $n(1-\tau_n) \rightarrow \infty$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathcal{O}(1)$, subject to the $\widehat{\varepsilon}_i^{(n)}$ being close enough to the ε_i according to condition $\mathcal{H}(n, \tau_n, \text{Dir}_1)$. The asymptotic normality result for $\varepsilon_{[n\tau_n]:n}$ is standard and follows from Theorem 2.4.8 p.52 in de Haan and Ferreira (2006).

Remark 3.1 (Interpretation of the high-level condition on the residuals). Condition $\mathcal{H}(n, \tau_n, \text{Dir}_1)$ guarantees that the number of residuals that do not correctly predict the unobserved errors is small enough for $\widehat{F}_n(x)$ to be closer to $\widetilde{F}_n(x)$ than $1/\sqrt{n(1-\tau_n)}$ when x lies in a neighborhood of $q(\tau_n)$. We shall show in Section 3.3 below that this condition is substantially weaker than existing conditions in the literature that require residuals and errors to be uniformly close. Importantly, the fact that exceedances of $|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|$ are counted above a random threshold $C_i^{(n)}$, which can take the past history of the process (ε_i) into account, is crucial (as pointed out in Section 4.3) in order to handle heavy-tailed time series models, where the natural covariate contains lags of the response variable, and in which one should not expect $\widehat{\varepsilon}_i^{(n)}$ and ε_i to always be close to each other depending on this past history.

3.2 Extreme value index estimation and Weissman estimator

When $p_{n,j} \downarrow 0$ such that $n_j(1 - p_{n,j})$ is bounded, the target quantiles may lie beyond the range of the data and the estimators $\hat{q}_{n,j}(1 - p_{n,j})$ will not be consistent. This motivates considering the residual-based version $\hat{q}_{n,j}^*(1 - p_{n,j})$ of the Weissman estimator. The main difficulty is then to understand the asymptotic behavior of the residual-based Hill estimator $\hat{\gamma}_{n,j}(k_j)$. This is done by writing

$$\begin{aligned} \hat{\gamma}_{n,j}(k_j) &= \frac{n_j}{k_j} \int_{\hat{\varepsilon}_{n_j - k_j; n_j, j}^{(n_j)}}^{\infty} \frac{\hat{F}_{n,j}(t)}{t} dt = \frac{n_j}{k_j} \int_1^{\infty} \frac{\hat{F}_{n,j}(x \hat{\varepsilon}_{n_j - k_j; n_j, j}^{(n_j)})}{x} dx \\ \text{and } \tilde{\gamma}_{n,j}(k_j) &= \frac{n_j}{k_j} \int_{\varepsilon_{n_j - k_j; n_j, j}}^{\infty} \frac{\tilde{F}_{n,j}(t)}{t} dt = \frac{n_j}{k_j} \int_1^{\infty} \frac{\tilde{F}_{n,j}(x \varepsilon_{n_j - k_j; n_j, j})}{x} dx \end{aligned}$$

and then by showing that if $\hat{F}_{n,j}(t)$ and $\tilde{F}_{n,j}(t)$ are sufficiently close over the right tail of the distribution of ε_j , then $\hat{\gamma}_{n,j}(k_j)$ is close to $\tilde{\gamma}_{n,j}(k_j)$, whose asymptotic behavior is well-known. We make this rigorous below by imposing condition $\mathcal{H}(n_j, \tau_{n,j}, x^{-1} \mathbb{1}\{x \geq 1\} dx)$ on the residuals $(\hat{\varepsilon}_{1,j}^{(n_j)}, \dots, \hat{\varepsilon}_{n_j,j}^{(n_j)})$.

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

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

Then

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

where the matrix $\mathbf{V} = \mathbf{V}(\mathbf{b}, \mathbf{c}, \boldsymbol{\gamma}, \mathbf{R})$ is symmetric positive semidefinite of size $2d \times 2d$ and described in block form by 2×2 blocks $\mathbf{V}_{j,\ell}$ ($1 \leq j, \ell \leq d$) having entries

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

It follows that Theorem 4 in Stupfler (2019) and Theorem 1 in Daouia et al. (2024a), on the joint asymptotic normality of several Hill estimators, remain valid for the residual-based versions. The following corollary on the joint asymptotic normality of residual-based Weissman estimators ensues.

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

for any j . Then

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

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

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

With our worked-out examples in mind, we state the following extension of Corollary 3.1, 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 $\varepsilon_1, \dots, \varepsilon_n$ satisfy condition $\mathcal{C}_2(\gamma, \rho, A)$. Suppose also that:*

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

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

$$\left(\sqrt{k_1} \{ \hat{\gamma}_n(k_1) - \gamma \}, \left(\sqrt{k_j} \left\{ \frac{\hat{q}_n(1-k_j/n)}{q(1-k_j/n)} - 1 \right\} \right)_{1 \leq j \leq d} \right) \xrightarrow{d} \mathcal{N} \left(\left(\frac{\lambda}{1-\rho}, 0, \dots, 0 \right), \gamma^2 \boldsymbol{\Sigma} \right)$$

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

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

for any $j, \ell \in \{1, \dots, d\}$.

(ii) If moreover $\rho < 0$, and $p_n \downarrow 0$ is such that $k/(np_n) \rightarrow \infty$ and $\sqrt{k}/\log(k/(np_n)) \rightarrow \infty$, then

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

In other words, under condition $\mathcal{H}(n, 1-k/n, x^{-1} \mathbb{1}\{x \geq 1\} dx)$, the estimators $\hat{\gamma}_n(k)$, $\hat{q}_n(1-k/n)$ and $\hat{q}_n^*(1-p_n)$ have the same asymptotic behavior as their traditional versions based on the sample $(\varepsilon_1, \dots, \varepsilon_n)$, as can be seen by comparing our results with those of Section 3.2 and Theorem 4.3.8 p.138 in de Haan and Ferreira (2006). Let us make clear that, as a comparison between Theorem 3.1 and Theorem 3.2 suggests, assumption $\mathcal{H}(n, 1-k/n, \mu)$, with $\mu(dx) = x^{-1} \mathbb{1}\{x \geq 1\} dx$, is indeed stronger than assumption $\mathcal{H}(n, 1-k/n, \text{Dir}_1)$. This is rigorously proven in Appendix A, see Lemma A.9.

It follows from our general asymptotic theory that, under a suitable high-level assumption, the residual-based Hill and Weissman estimators have the same asymptotics as their counterparts based on i.i.d. data. The key high-level condition to be checked as part of our main results in a given model is $\mathcal{H}(n, \tau_n, \mu)$, where μ is an appropriate weighting measure. The next section discusses simple sufficient conditions ensuring that this assumption is indeed valid.

3.3 Checking the high-level condition $\mathcal{H}(n, \tau_n, \mu)$

The main difficulty in checking the high-level condition $\mathcal{H}(n, \tau_n, \mu)$ is to construct the $\sqrt{n(1 - \tau_n)}$ -small array of random variables $C_i^{(n)}$ appearing in its statement. This section provides several sufficient conditions under which such an array is guaranteed to exist. We start with a result highlighting the relationship between condition $\mathcal{H}(n, \tau_n, \mu)$ and uniform-type conditions on the differences $|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|$.

Proposition 3.1 (Relationship between uniform convergence conditions and $\mathcal{H}(n, \tau_n, \mu)$). *Let $\tau_n \uparrow 1$ be such that $n(1 - \tau_n) \rightarrow \infty$. Assume that a sample of residuals $(\widehat{\varepsilon}_1^{(n)}, \dots, \widehat{\varepsilon}_n^{(n)})$ and the corresponding i.i.d. heavy-tailed errors $(\varepsilon_1, \dots, \varepsilon_n)$ satisfy*

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

where q is the quantile function of the ε_i . Then condition $\mathcal{H}(n, \tau_n, \mu)$ is satisfied for any measure μ that both puts finite mass on compact Borel subsets of \mathbb{R} and no mass on a neighborhood of 0.

Remark 3.2 (Comparison with a condition of Girard et al. (2021)). By Proposition 3.1, condition $\mathcal{H}(n, \tau_n, \mu)$ is weaker than earlier conditions of uniform-type in the residual-based extremal regression literature. For example, condition (2) in Girard et al. (2021), considered also by Ahmad et al. (2020) in the context of conditional extreme quantile estimation in a location-scale model, is

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

This is of course stronger than the condition of Proposition 3.1, due to the high quantile $q(\tau_n)$ appearing in the denominator of the latter, rather than the constant 1 in Girard et al. (2021). This additional term is helpful in handling relative discrepancies that may be large when ε_i is small.

Remark 3.3 (Comparison with a condition of Velthoen et al. (2019)). Residual-based extreme quantile regression has also been discussed in Velthoen et al. (2019) in the nonparametric homoskedastic model

$$q_{Y|X}(\tau|x) = r(x) + q(\tau), \quad \text{for } \tau > \tau_c \geq 0,$$

featuring a univariate covariate X and a smooth continuous baseline function r to be estimated above a fixed unknown quantile level τ_c . The main condition in their Theorem 2, covering the estimation of the conditional extreme value index γ of the unobserved and heavy-tailed variable ε , is $k^{\gamma+1} n^{-(\delta+\gamma)} \rightarrow 0$, where $k = k(n) = n(1 - \tau_n)$ is the effective number of residuals used as part of the estimation and $n^{-\delta}$ (where $\delta > 0$) represents the uniform error made in the estimation of r over an interval, and

hence quantifies the gap between the residuals they use and the true ε_i . Besides, when the second-order parameter ρ of ε is negative, it is a consequence of the discussion below Example 2.3.11 p.49 in de Haan and Ferreira (2006) that $q(\tau)$ is, as $\tau \uparrow 1$, asymptotically equivalent to a multiple of $(1 - \tau)^{-\gamma}$. In other words, the condition of Velthoen et al. (2019) can be interpreted as

$$n(1 - \tau_n) \max_{1 \leq i \leq n} \frac{|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|}{q_{\tau_n}} = k \max_{1 \leq i \leq n} \frac{|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|}{q_{1-k/n}} \xrightarrow{\mathbb{P}} 0.$$

This is also stronger than the uniform condition in Proposition 3.1.

The above two remarks show that the framework developed in the present paper is able to unify and extend the asymptotic results of Velthoen et al. (2019) and Girard et al. (2021). It should be noted that the latter two papers focus on obtaining Gaussian approximations to the tail empirical process of the residuals; our approach here is essentially different and uses directly the particular structure of the estimators we study in order to derive joint asymptotic normality results.

It may happen that one cannot check the uniform convergence condition of Proposition 3.1 unless strong restrictions are imposed. Typical examples can be found in location-scale models of the form $Y_i = m_i + \sigma_i \varepsilon_i$ where, for any i , the error ε_i is independent from the location and scale components m_i and $\sigma_i > 0$. Given estimators $\widehat{m}_i^{(n)}$ of m_i and $\widehat{\sigma}_i^{(n)} > 0$ of σ_i , the corresponding residuals $\widehat{\varepsilon}_i^{(n)}$ satisfy

$$\widehat{\varepsilon}_i^{(n)} - \varepsilon_i = \frac{Y_i - \widehat{m}_i^{(n)}}{\widehat{\sigma}_i^{(n)}} - \varepsilon_i = \frac{m_i - \widehat{m}_i^{(n)}}{\widehat{\sigma}_i^{(n)}} + \left(\frac{\sigma_i}{\widehat{\sigma}_i^{(n)}} - 1 \right) \varepsilon_i.$$

The obvious inequality

$$\sqrt{n(1 - \tau_n)} \frac{|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|}{q(\tau_n) + |\varepsilon_i|} \leq \frac{1}{\widehat{\sigma}_i^{(n)}} \left(\frac{\sqrt{n(1 - \tau_n)}}{q(\tau_n)} |\widehat{m}_i^{(n)} - m_i| + \sqrt{n(1 - \tau_n)} |\widehat{\sigma}_i^{(n)} - \sigma_i| \right)$$

suggests that Proposition 3.1 holds if $\max_{1 \leq i \leq n} (|\widehat{m}_i^{(n)} - m_i|/q(\tau_n) + |\widehat{\sigma}_i^{(n)} - \sigma_i|) = o_{\mathbb{P}}(1/\sqrt{n(1 - \tau_n)})$. This can be a restrictive assumption even in the linear regression model $Y_i = \mathbf{X}_i^\top \boldsymbol{\beta} + \varepsilon_i$ with centered and square-integrable noise, where the least squares estimator $\widehat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ would be \sqrt{n} -consistent, but

$$\frac{\sqrt{n(1 - \tau_n)}}{q(\tau_n)} \max_{1 \leq i \leq n} |\mathbf{X}_i^\top \widehat{\boldsymbol{\beta}}_n - \mathbf{X}_i^\top \boldsymbol{\beta}| \leq \frac{\sqrt{n(1 - \tau_n)}}{q(\tau_n)} \|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\| \max_{1 \leq i \leq n} \|\mathbf{X}_i\| = O_{\mathbb{P}} \left(\frac{\sqrt{1 - \tau_n}}{q(\tau_n)} \max_{1 \leq i \leq n} \|\mathbf{X}_i\| \right)$$

does not necessarily converge to 0 unless one puts the further restriction $\sqrt{1 - \tau_n} \max_{1 \leq i \leq n} \|\mathbf{X}_i\|/q(\tau_n) = o_{\mathbb{P}}(1)$ on τ_n . With this observation in mind, we give a simple second sufficient criterion for $\mathcal{H}(n, \tau_n, \mu)$, which often applies and provides a blueprint for the analysis of more difficult cases.

Proposition 3.2 (Sufficient condition for $\mathcal{H}(n, \tau_n, \mu)$ in location-scale models). *Let $\tau_n \uparrow 1$ be such that $n(1 - \tau_n) \rightarrow \infty$. Assume that a sample of residuals $(\widehat{\varepsilon}_1^{(n)}, \dots, \widehat{\varepsilon}_n^{(n)})$ and the corresponding i.i.d. heavy-tailed errors $(\varepsilon_1, \dots, \varepsilon_n)$ satisfy the set of equations*

$$\forall i \in \{1, \dots, n\}, \widehat{\varepsilon}_i^{(n)} - \varepsilon_i = \frac{m_i - \widehat{m}_i^{(n)}}{\widehat{\sigma}_i^{(n)}} + \left(\frac{\sigma_i}{\widehat{\sigma}_i^{(n)}} - 1 \right) \varepsilon_i$$

where, on an event whose probability tends to 1 as $n \rightarrow \infty$:

- There is $\sigma_0 > 0$ such that $\widehat{\sigma}_i^{(n)} \geq \sigma_0$ for any $i \in \{1, \dots, n\}$,
- There are positive random variables A_n and B_n , as well as positive integrable random variables $U_i^{(n)}$ and $V_i^{(n)}$ that are independent of ε_i for each $i \in \{1, \dots, n\}$, such that

$$\forall i \in \{1, \dots, n\}, \quad \frac{|\widehat{m}_i^{(n)} - m_i|}{q(\tau_n)} \leq A_n U_i^{(n)} \quad \text{and} \quad |\log \widehat{\sigma}_i^{(n)} - \log \sigma_i| \leq B_n V_i^{(n)}$$

where $A_n = O_{\mathbb{P}}(1/\sqrt{n})$, $B_n = O_{\mathbb{P}}(1/\sqrt{n})$,

$$\frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} (U_i^{(n)} + V_i^{(n)}) \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n \mathbb{E}(U_i^{(n)} + V_i^{(n)}) = O(1).$$

Then condition $\mathcal{H}(n, \tau_n, \mu)$ is satisfied for any measure μ that both puts finite mass on compact Borel subsets of \mathbb{R} and no mass on a neighborhood of 0.

The last important example we investigate is the class of (additive) noisy data models of the form $Y_i = W_i + Z_i$, $1 \leq i \leq n$, where the unobserved W_i may be random or nonrandom, and the Z_i are unobserved i.i.d. copies of a heavy-tailed random variable Z , and where the interest lies in estimating the extreme value features of the Z_i . The problem of estimating an extreme value parameter when the available data is recorded up to observational errors has been the focus of several articles over the last two decades or so, starting with Hall and Simar (2002) who considered the estimation of the boundary of a support, to, among others, Matsui et al. (2013) and Kim and Kokoszka (2020, 2023) who concentrated on the estimation of the extreme value index and extreme quantiles of heavy-tailed data. A case study of noisy data due to measurement errors in climate science is Marra et al. (2018). Even though this is not a location-scale regression model and no model estimation step takes place, our results in fact also apply if one views directly the Z_i as the unobserved variables and the Y_i as the “residuals”, namely, we set $\varepsilon_i = Z_i$ and $\widehat{\varepsilon}_i^{(n)} = Y_i$. Our final result of this section investigates a sufficient, non-uniform condition on the $W_i = Y_i - Z_i$ for our results to hold in this setting.

Proposition 3.3 (Sufficient condition for $\mathcal{H}(n, \tau_n, \mu)$ in noisy data models). *Let $\tau_n \uparrow 1$ be such that $n(1 - \tau_n) \rightarrow \infty$. Fix a measure μ that both puts finite mass on compact Borel subsets of \mathbb{R} and no mass on a neighborhood of 0. Assume that a sample of observations (Y_1, \dots, Y_n) satisfies*

$$\forall \delta > 0, \quad \lim_{n \rightarrow \infty} \int_0^\infty \frac{n}{\sqrt{n(1 - \tau_n)}} \max_{1 \leq i \leq n} \mathbb{P} \left(|Y_i - Z_i| > \delta \frac{1}{\sqrt{n(1 - \tau_n)}} (xq(\tau_n) + |Z_i|) \right) \mu(dx) = 0$$

where (Z_1, \dots, Z_n) are i.i.d. and heavy-tailed with quantile function q . Then condition $\mathcal{H}(n, \tau_n, \mu)$ is satisfied with $\varepsilon_i = Z_i$ and $\widehat{\varepsilon}_i^{(n)} = Y_i$.

The objective of the next section is to show how these criteria can be used in order to prove the asymptotic normality of residual-based extreme value estimators in a list of examples relevant to statistical practice.

4 Applications to noisy data problems and extremal regression

We start by drawing consequences of our results in the context where i.i.d. data is recorded up to errors. We then discuss well-known regression models in which our theory on the residual-based estimators $\hat{q}_n(1-k/n)$, $\hat{\gamma}_n(k)$ and $\hat{q}_n^*(1-p_n)$ applies; the sufficient criteria of Section 3.3 will be at the heart of our arguments. We also discuss statistical consequences on the motivating problem of extremal regression.

4.1 Noisy data

Suppose that the available data points Y_i come from the noisy data model $Y_i = W_i + Z_i$, $1 \leq i \leq n$, where the unobserved W_i may be random or nonrandom, and the Z_i are unobserved i.i.d. copies of a heavy-tailed random variable Z . The Y_i need not be independent even though the Z_i are. Viewing the extreme quantile $q(1-p_n)$ of Z or its extreme value index γ as targets and the Y_i as the available data points altered by the errors W_i gives rise to the estimators

$$\hat{q}_n^*(1-p_n) = \left(\frac{k}{np_n}\right)^{\hat{\gamma}_n} \hat{q}_n(1-k/n) \quad \text{with} \quad \hat{q}_n(1-k/n) = Y_{n-k:n} \quad \text{and} \quad \hat{\gamma}_n(k) = \frac{1}{k} \sum_{i=1}^k \log \frac{Y_{n-i+1:n}}{Y_{n-k:n}}.$$

Our first result leverages Proposition 3.1 in order to obtain their asymptotic normality.

Corollary 4.1 (Noisy data I). *Write the available data as $Y_i = W_i + Z_i$, for $i \geq 1$, where the Z_i are i.i.d. copies of a random variable Z satisfying condition $\mathcal{C}_2(\gamma, \rho, A)$ with common quantile function q . Let $k = k(n) \rightarrow \infty$ satisfy $k/n \rightarrow 0$ and $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$. Assume finally that*

$$\frac{\sqrt{k}}{q(1-k/n)} \max_{1 \leq i \leq n} |W_i| \xrightarrow{\mathbb{P}} 0.$$

(i) *Then Corollary 3.3(i) applies, that is, $(\sqrt{k}(\hat{\gamma}_n(k) - \gamma), \sqrt{k}(\hat{q}_n(1-k/n)/q(1-k/n) - 1))$ converges weakly to a pair of independent Gaussian random variables having respective means $\lambda/(1-\rho)$ and 0, and the same variance γ^2 .*

(ii) *If $\rho < 0$ and $p_n \downarrow 0$ is such that $k/(np_n) \rightarrow \infty$ and $\sqrt{k}/\log(k/(np_n)) \rightarrow \infty$, then Corollary 3.3(ii) applies, that is, $(\sqrt{k}/\log(k/(np_n)))(\hat{q}_n^*(1-p_n)/q(1-p_n) - 1)$ has the same asymptotic distribution as $\sqrt{k}(\hat{\gamma}_n(k) - \gamma)$.*

Note the implicit assumption $\sqrt{k}/q(1-k/n) \rightarrow 0$ as long as one of the W_i is nonzero. Corollary 4.1(i) contains the (informally) well-known result that the Hill estimator stays asymptotically normal when the data is shifted by a positive constant provided that $\gamma \neq -\rho$ and $\sqrt{k}/q(1-k/n) \rightarrow 0$, because then the shifted random variables still satisfy the second-order regular variation condition with auxiliary function proportional to $\max(|A(t)|, 1/q(1-1/t))$: see the discussion in the last paragraph of Section 3.2 p.83 in de Haan and Ferreira (2006). It in fact contains the asymptotic normality of intermediate empirical quantiles, the Hill estimator and the Weissman estimator in many more settings:

- With $W_i = w_i + c_i Z_i$, where the w_i and c_i are nonrandom, Corollary 4.1 holds if $\sqrt{k}/q(1-k/n) \rightarrow 0$, the shifting sequence (w_i) is bounded, and the random variable $\sup_{i \geq 1} |c_i Z_i|$ is finite with probability

1. The latter condition is for example satisfied when the Z_i are Fréchet distributed with extreme value index γ , *i.e.* $\mathbb{P}(Z_i \leq x) = \exp(-x^{-1/\gamma})$ for $x > 0$, and the series $\sum_{j \geq 1} |c_j|^{1/\gamma}$ is summable.

- With $W_i = 10^{-l} \lfloor 10^l Z_i \rfloor - Z_i$, namely, when the data is observed on a lattice, Corollary 4.1(i) extends Corollary 2.4 of Matsui et al. (2013) to random variables Z_i which are not Pareto distributed. We in fact allow for the more general setting $W_i = 10^{-l_i} \lfloor 10^{l_i} Z_i \rfloor - Z_i$, under the conditions that $\sqrt{k}/q(1 - k/n) \rightarrow 0$ and the l_i are bounded from below, meaning that our result applies as long as the data Z_i are rounded off to a precision that cannot get arbitrarily poor as more data gets recorded.

- With W_i being i.i.d. unit Gaussian centered, recall that $\max_{1 \leq i \leq n} |W_i|/\sqrt{2 \log n} \xrightarrow{\mathbb{P}} 1$. If moreover $\rho < 0$, then $q(\tau)$ is, as $\tau \uparrow 1$, asymptotically proportional to $(1 - \tau)^{-\gamma}$, and thus $\sqrt{k}/q(1 - k/n) = O(k^{\gamma+1/2}/n^\gamma)$. Conclude that if $k = O(n^\alpha)$ with $\alpha < 2\gamma/(2\gamma + 1)$, then $\sqrt{k} \max_{1 \leq i \leq n} |W_i|/q(1 - k/n) \xrightarrow{\mathbb{P}} 0$: the Hill and Weissman estimators of i.i.d. data corrupted by i.i.d. Gaussian errors keep their asymptotic distributions. This is the additive error setting of Section 3.5.2 in Marra et al. (2018).

The condition on the W_i in Corollary 4.1 is easy to check but is of uniform type. Under the assumption that the $|W_i|$ have the same distribution, we may provide a weaker sufficient condition that fully exploits the averaging structure in condition $\mathcal{H}(n, \tau_n, \mu)$ in the context of extreme quantile estimation.

Corollary 4.2 (Noisy data II). *Write the available data Y_i as $Y_i = W_i + Z_i$, for $i \geq 1$, where the $|W_i|$ have the same distribution as a heavy-tailed random variable $|W|$, and the Z_i are i.i.d. copies of a random variable Z that satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$ with common quantile function q . Let $k = k(n) \rightarrow \infty$ satisfy $k/n \rightarrow 0$ and $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$. Assume finally that*

$$\forall \delta > 0, \quad \frac{n}{\sqrt{k}} \mathbb{P} \left(|W| > \delta \frac{q(1 - k/n)}{\sqrt{k}} \right) \sim \sqrt{k} \frac{\mathbb{P}(|W| > \delta q(1 - k/n)/\sqrt{k})}{\mathbb{P}(Z > q(1 - k/n))} \rightarrow 0.$$

(i) *Then Corollary 3.3(i) applies.*

(ii) *If $\rho < 0$ and $p_n \downarrow 0$ is such that $k/(np_n) \rightarrow \infty$ and $\sqrt{k}/\log(k/(np_n)) \rightarrow \infty$, then Corollary 3.3(ii) applies.*

In Corollary 4.2 note, again, the implicit assumption $\sqrt{k}/q(1 - k/n) \rightarrow 0$. The asymptotic normality of the Hill and Weissman estimators therefore stays valid when the data Z_i is corrupted by additive noise W_i as long as the (heavy) tail of the noise variable W is, in an appropriate sense, lighter than that of Z . This can be compared with Corollary 2.2 in Kim and Kokoszka (2023): their condition is, with our notation, $\mathbb{P}(|W| > x) = o(x^{-\kappa})$ as $x \rightarrow \infty$ for some $\kappa > (1 + \max(-\rho, \gamma))/\gamma$. This is a different (and typically weaker) assumption than ours, but a crucial difference is that Corollary 2.2 in Kim and Kokoszka (2023) requires the W_i to be i.i.d. copies of W and that W is independent of Z , which allows them to use powerful results on convolutions of second-order regularly varying distributions. Moreover, their result is limited to the setting when Z is nonnegative and $\gamma < 1$. In addition to avoiding this restriction on γ and the sign of Z , we do not make any assumption whatsoever on the dependence within the sequence (W_i) or between W and Z and therefore, as one should expect, we have to put somewhat stronger tail restrictions on W . This, we believe, is a more realistic setting, as it makes it possible to handle the lattice data example, or more generally corruption of data points depending on their values, where the independence assumption is not reasonable.

4.2 Location-scale parametric regression

Suppose that observations $Y_{i,j}$, $1 \leq j \leq d$ and $1 \leq i \leq n_j$, are generated from univariate location-scale regression models of the form $Y_{i,j} = m_j(\mathbf{X}_{i,j}, \boldsymbol{\beta}_j) + \sigma_j(\mathbf{X}_{i,j}, \boldsymbol{\theta}_j)\varepsilon_{i,j}$, for $1 \leq j \leq d$ and $1 \leq i \leq n_j$, where the covariate value $\mathbf{X}_{i,j} \in \mathbb{R}^{p_j}$ is observed, and is either nonrandom, or random and independent of the unobserved error $\varepsilon_{i,j}$, the parameter vectors $\boldsymbol{\beta}_j \in \mathbb{R}^{r_j}$, $\boldsymbol{\theta}_j \in \mathbb{R}^{s_j}$ are unknown, and the families of measurable functions $m_j(\cdot, \boldsymbol{\beta}_j)$ and $\sigma_j(\cdot, \boldsymbol{\theta}_j) > 0$ are known. This model is considered for example in Carroll and Ruppert (1982) for $j = 1$ and a linear function $m = m_1$, namely $m(\mathbf{x}, \boldsymbol{\beta}) = \mathbf{x}^\top \boldsymbol{\beta}$.

Suppose further that $\widehat{\boldsymbol{\beta}}_{n,j}$ and $\widehat{\boldsymbol{\theta}}_{n,j}$ are $\sqrt{n_j}$ -consistent estimators of $\boldsymbol{\beta}_j$ and $\boldsymbol{\theta}_j$, respectively. If the functions σ_j are bounded from below and the functions $m_j(\mathbf{x}_j, \cdot)$ and $\log \sigma_j(\mathbf{x}_j, \cdot)$ satisfy an appropriate Lipschitz condition, one may apply Proposition 3.2 to the residuals $\widehat{\varepsilon}_{i,j}^{(n_j)} = (Y_{i,j} - m_j(\mathbf{X}_{i,j}, \widehat{\boldsymbol{\beta}}_{n,j}))/\sigma_j(\mathbf{X}_{i,j}, \widehat{\boldsymbol{\theta}}_{n,j})$ and, using the theory in Section 3.2, obtain convergence results for the corresponding residual-based extreme value estimators. A general result follows.

Corollary 4.3 (Residual-based extreme value estimation – Location-scale parametric regression model).

Assume that $Y_{i,j} = m_j(\mathbf{X}_{i,j}, \boldsymbol{\beta}_j) + \sigma_j(\mathbf{X}_{i,j}, \boldsymbol{\theta}_j)\varepsilon_{i,j}$, for $1 \leq j \leq d$ and $1 \leq i \leq n_j$, where the $n_j = n_j(n)$ satisfy $n_1/n_j \rightarrow b_j \in (0, \infty)$ (with $b_1 = 1$), the parameter vectors $\boldsymbol{\beta}_j \in \mathbb{R}^{r_j}$ and $\boldsymbol{\theta}_j \in \mathbb{R}^{s_j}$, the covariate vector $\mathbf{X}_{i,j} \in \mathbb{R}^{p_j}$ is either nonrandom, or random and independent from $\varepsilon_{i,j}$, the dispersion function σ_j is positive, and the $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,d})$ are independent copies of a random vector $\boldsymbol{\varepsilon}$ satisfying conditions $\mathcal{C}_2(\boldsymbol{\gamma}, \boldsymbol{\rho}, \mathbf{A})$ and $\mathcal{J}(\mathbf{R})$. Suppose that there are subsets \mathcal{X}_j of \mathbb{R}^{p_j} and neighborhoods V_j, W_j of $\mathbf{0}$ in \mathbb{R}^{r_j} and \mathbb{R}^{s_j} , respectively, with $\inf_{\mathbf{x}_j \in \mathcal{X}_j} \inf_{\mathbf{h}_j \in W_j} \sigma_j(\mathbf{x}_j, \boldsymbol{\theta}_j + \mathbf{h}_j) > 0$ and, for all $\mathbf{x}_j \in \mathcal{X}_j$,

$$M_j(\mathbf{x}_j) = \sup_{\mathbf{h} \in V_j} \frac{|m_j(\mathbf{x}_j, \boldsymbol{\beta}_j + \mathbf{h}) - m_j(\mathbf{x}_j, \boldsymbol{\beta}_j)|}{\|\mathbf{h}\|} < \infty$$

$$\text{and } S_j(\mathbf{x}_j) = \sup_{\mathbf{h} \in W_j} \frac{|\log \sigma_j(\mathbf{x}_j, \boldsymbol{\theta}_j + \mathbf{h}) - \log \sigma_j(\mathbf{x}_j, \boldsymbol{\theta}_j)|}{\|\mathbf{h}\|} < \infty.$$

Assume moreover that, for $1 \leq j \leq d$, one of the following two conditions is satisfied:

- The $\mathbf{X}_{i,j}$, for $1 \leq i \leq n_j$, are independent copies of a random vector \mathbf{X}_j with support \mathcal{X}_j , such that $T_j(\mathbf{X}_j) = M_j(\mathbf{X}_j) + S_j(\mathbf{X}_j)$ has a finite moment of order $2 + \delta$ for some $\delta > 0$.
- The $\mathbf{X}_{i,j} = \mathbf{x}_{i,j}$ are nonrandom, all belong to \mathcal{X}_j , and the $T_j(\mathbf{x}_{i,j}) = M_j(\mathbf{x}_{i,j}) + S_j(\mathbf{x}_{i,j})$ satisfy

$$\frac{1}{\sqrt{n_j}} \max_{1 \leq i \leq n_j} T_j(\mathbf{x}_{i,j}) \rightarrow 0 \quad \text{and} \quad \frac{1}{n_j} \sum_{i=1}^{n_j} T_j(\mathbf{x}_{i,j}) = O(1).$$

Finally, assume that $\sqrt{n_j}(\widehat{\boldsymbol{\beta}}_{n,j} - \boldsymbol{\beta}_j) = O_{\mathbb{P}}(1)$ and $\sqrt{n_j}(\widehat{\boldsymbol{\theta}}_{n,j} - \boldsymbol{\theta}_j) = O_{\mathbb{P}}(1)$, and define $\widehat{\varepsilon}_{i,j}^{(n_j)} = (Y_{i,j} - m_j(\mathbf{X}_{i,j}, \widehat{\boldsymbol{\beta}}_{n,j}))/\sigma_j(\mathbf{X}_{i,j}, \widehat{\boldsymbol{\theta}}_{n,j})$. Let the $k_j = k_j(n) \rightarrow \infty$ satisfy $k_1/k_j \rightarrow c_j \in (0, \infty)$ (with $c_1 = 1$), as well as $k_j/n_j \rightarrow 0$ and $\sqrt{k_j}A_j(n_j/k_j) \rightarrow \lambda_j \in \mathbb{R}$.

(i) Then Theorem 3.2 applies.

(ii) If the $\rho_j < 0$ and $p_{n,j} \downarrow 0$ are such that $k_j/(n_j p_{n,j}) \rightarrow \infty$ and $\sqrt{k_j}/\log(k_j/(n_j p_{n,j})) \rightarrow \infty$ for any $j \in \{1, \dots, d\}$, then Corollary 3.2 applies.

With Corollary 4.3 at our disposal, we may obtain joint asymptotic results on the estimators $\widehat{q}_{Y_j|\mathbf{X}_j}^*(1 - p_{n,j}|\mathbf{x}_j) \equiv \widehat{q}_{Y_j|\mathbf{X}_j,n}^*(1 - p_{n,j}|\mathbf{x}_j) = m_j(\mathbf{x}_j, \widehat{\boldsymbol{\beta}}_{n,j}) + \sigma_j(\mathbf{x}_j, \widehat{\boldsymbol{\theta}}_{n,j})\widehat{q}_{n,j}^*(1 - p_{n,j})$, which target the extreme conditional quantiles $q_{Y_j|\mathbf{X}_j}(1 - p_{n,j}|\mathbf{x}_j) = m_j(\mathbf{x}_j, \boldsymbol{\beta}_j) + \sigma_j(\mathbf{x}_j, \boldsymbol{\theta}_j)q_j(1 - p_{n,j})$.

Corollary 4.4 (Extremal regression – Location-scale parametric regression model). *Work under the conditions of Corollary 4.3(ii). Then for $\mathbf{x}_j \in \mathcal{X}_j$, $1 \leq j \leq d$, and with the notation of Corollary 3.2,*

$$\left(\frac{\sqrt{k_j}}{\log(k_j/(np_{n,j}))} \left\{ \frac{\widehat{q}_{Y_j|\mathbf{X}_j}^*(1 - p_{n,j}|\mathbf{x}_j)}{q_{Y_j|\mathbf{X}_j}(1 - p_{n,j}|\mathbf{x}_j)} - 1 \right\} \right)_{1 \leq j \leq d} \xrightarrow{d} \mathcal{N} \left(\left(\frac{\lambda_1}{1 - \rho_1}, \dots, \frac{\lambda_d}{1 - \rho_d} \right), \mathbf{V}_\gamma \right).$$

As an illustration, we discuss how the assumptions of Corollaries 4.3 and 4.4 can be satisfied in the class of models $Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \exp(\theta \phi(\mathbf{x}_i^\top \boldsymbol{\beta})) \varepsilon_i$, considered in Carroll and Ruppert (1982), where the \mathbf{x}_i are nonrandom, the ε_i are i.i.d. with a symmetric distribution, ϕ is a known smooth and Lipschitz function, and $\boldsymbol{\beta} \in \mathbb{R}^r$, $\theta \in \mathbb{R}$ are unknown parameters. It follows from Carroll and Ruppert (1982) that, in this class of models, and provided a preliminary \sqrt{n} -consistent estimator of $\boldsymbol{\beta}$ can be constructed, robust weighted and \sqrt{n} -consistent estimators of $\boldsymbol{\beta}$ and θ can be found under the assumptions that

$$\frac{1}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\exp(2\theta \phi(\mathbf{x}_i^\top \boldsymbol{\beta}))} \rightarrow \mathbf{S} \text{ (symmetric positive definite), } \frac{1}{n} \sum_{i=1}^n \{\|\mathbf{x}_i\|^2 + |\phi(\mathbf{x}_i^\top \boldsymbol{\beta})|^2\} = O(1),$$

$$\max_{1 \leq i \leq n} \{\|\mathbf{x}_i\| + |\phi(\mathbf{x}_i^\top \boldsymbol{\beta})|\} = o(\sqrt{n}), \quad \inf_{i \geq 1} \exp(\theta \phi(\mathbf{x}_i^\top \boldsymbol{\beta})) > 0 \quad \text{and} \quad \inf_{n \geq 1} \frac{1}{n} \sum_{i=1}^n \{\phi(\mathbf{x}_i^\top \boldsymbol{\beta})\}^2 > 0,$$

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

$$|t\phi(\mathbf{x}^\top \mathbf{b}) - \theta\phi(\mathbf{x}^\top \boldsymbol{\beta})| \leq |t|\phi(\mathbf{x}^\top \mathbf{b}) - \phi(\mathbf{x}^\top \boldsymbol{\beta})| + |t - \theta|\phi(\mathbf{x}^\top \boldsymbol{\beta})|,$$

it follows from the Lipschitz property of ϕ that one can choose $S(\mathbf{x}) = C(\|\mathbf{x}\| + |\phi(\mathbf{x}^\top \boldsymbol{\beta})|)$ for a suitable positive constant C , and the assumptions of Corollary 4.3 on the $T(\mathbf{x}_i) = M(\mathbf{x}_i) + S(\mathbf{x}_i)$ are then automatically fulfilled under the above conditions ensuring the existence of \sqrt{n} -consistent estimators of $\boldsymbol{\beta}$ and θ . A preliminary \sqrt{n} -consistent estimator of $\boldsymbol{\beta}$, necessary for their construction, can be obtained through L^1 -(median) regression, see Knight (1999), under the assumptions that the ε_i have a probability density function w.r.t. the Lebesgue measure which is continuous and positive at 0 and the (suitably normalized) limiting design matrix formed by the covariates is positive definite.

The construction of the residuals in Corollary 4.3 requires observing the covariate values. It thus does not apply when part of the covariate values are unobserved, as is the case in typical time series models, for which a dedicated treatment is necessary.

4.3 (V)AR(I)MA time series

ARIMA model Consider the ARIMA(p, d, q) model $P(L) \circ (\text{Id} - L)^d(Z_t) = Q(L)(\varepsilon_t)$, for $t \in \mathbb{Z}$, where $L(Z_t) = Z_{t-1}$ is the lag operator, d is a nonnegative integer, the coefficients $\phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q \in \mathbb{R}$ of the polynomials $P(z) = 1 - \sum_{j=1}^p \phi_j z^j$ and $Q(z) = 1 - \sum_{j=1}^q \psi_j z^j$ are such that P and Q

have no common root and no root inside the closed unit disk of the complex plane, and (ε_t) is an i.i.d. sequence of heavy-tailed innovations with extreme value index γ , such that $\varepsilon \equiv \varepsilon_0$ satisfies $\mathbb{P}(\varepsilon > x)/\mathbb{P}(|\varepsilon| > x) \rightarrow l \in (0, 1]$ as $x \rightarrow \infty$ (i.e. the left tail of ε cannot dominate its right tail). This reduces to a causal and invertible ARMA(p, q) model on the differenced time series $Y_t = (\text{Id} - L)^d(Z_t)$.

Assume that observations $(Z_{1-d-p}, \dots, Z_0, Z_1, \dots, Z_n)$ are available from the ARIMA(p, d, q) model, i.e. differenced data $(Y_{1-p}, \dots, Y_0, Y_1, \dots, Y_n)$ are available from the ARMA(p, q) model. It was shown by Zhu and Ling (2012) that one may construct a self-weighted local absolute deviations estimator of the model parameters that is \sqrt{n} -consistent if the noise variable has zero median and a bounded and continuous probability density function w.r.t. Lebesgue measure that is also positive at 0. In the infinite variance situation $\gamma > 1/2$, Davis (1996) discusses least absolute deviations estimators and local least absolute deviations estimators, as well as a general class of M-estimators, that are even n^γ -consistent under a further tail balancing condition. Assume then that an estimator $(\hat{\phi}_{n,1}, \dots, \hat{\phi}_{n,p}, \hat{\psi}_{n,1}, \dots, \hat{\psi}_{n,q})$ is given such that $n^{\max(1/2, \gamma)}|\hat{\phi}_{n,i} - \phi_{n,i}| = O_{\mathbb{P}}(1)$ and $n^{\max(1/2, \gamma)}|\hat{\psi}_{n,j} - \psi_{n,j}| = O_{\mathbb{P}}(1)$ for each $i \in \{1, \dots, p\}$ and $j \in \{1, \dots, q\}$. The residuals from the model are recursively found as

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

with $\hat{\varepsilon}_{1-q}^{(n)}, \dots, \hat{\varepsilon}_0^{(n)}$ arbitrary choices of the residuals at times $1 - q, \dots, 0$. Here Proposition 3.2 does not apply, not least because the random variables Y_t and ε_t (which would be key components of the quantities $U_i^{(n)}$ and $V_i^{(n)}$ in Proposition 3.2) are not necessarily integrable under our assumptions. One may, however, show directly the existence of an appropriate $\sqrt{n(1 - \tau_n)}$ -small array for the differences $|\hat{\varepsilon}_t^{(n)} - \varepsilon_t|$ by providing first a fine upper bound for these differences. This leads to the following result.

Theorem 4.1 (Residual-based extreme value estimation – ARIMA model). *Assume that the finite differences $Y_t = (\text{Id} - L)^d(Z_t)$ of a time series (Z_t) satisfy the causal and invertible ARMA(p, q) equations $Y_t = \sum_{j=1}^p \phi_j Y_{t-j} - \sum_{j=1}^q \psi_j \varepsilon_{t-j} + \varepsilon_t$, for $t \in \mathbb{Z}$, where $\phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q \in \mathbb{R}$ are such 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 the ε_t are independent copies of a random variable ε satisfying condition $C_2(\gamma, \rho, A)$ and $\mathbb{P}(\varepsilon > x)/\mathbb{P}(|\varepsilon| > x) \rightarrow l \in (0, 1]$ as $x \rightarrow \infty$.*

Suppose that $\hat{\phi}_{n,1}, \dots, \hat{\phi}_{n,p}, \hat{\psi}_{n,1}, \dots, \hat{\psi}_{n,q}$ are $n^{\max(1/2, \gamma)}$ -consistent estimators of $\phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q$, and $\hat{\varepsilon}_t^{(n)} = Y_t - \sum_{j=1}^p \hat{\phi}_{n,j} Y_{t-j} + \sum_{j=1}^q \hat{\psi}_{n,j} \hat{\varepsilon}_{t-j}^{(n)}$ are the pertaining residuals with initial values $\hat{\varepsilon}_{1-q}^{(n)} \equiv \hat{\varepsilon}_{1-q}, \dots, \hat{\varepsilon}_0^{(n)} \equiv \hat{\varepsilon}_0$ that have a finite moment of order $\min(1, 1/(2\gamma))$ and are determined independently of the sequence (Y_t) . Finally, let $k = k(n) \rightarrow \infty$ satisfy $k/n \rightarrow 0$ and $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$.

(i) *Then Corollary 3.3(i) applies.*

(ii) *If $\rho < 0$ and $p_n \downarrow 0$ with $k/(np_n) \rightarrow \infty$ and $\sqrt{k}/\log(k/(np_n)) \rightarrow \infty$, then Corollary 3.3(ii) applies.*

As in Section 4.2, we illustrate the implications of this result by considering the asymptotic behavior of the estimator $\hat{q}_{Y_{n+1}}^*(1 - p_n | \mathcal{F}_n) = \sum_{j=1}^p \hat{\phi}_{n,j} Y_{n+1-j} - \sum_{j=1}^q \hat{\psi}_{n,j} \hat{\varepsilon}_{n+1-j}^{(n)} + \hat{q}_{n,\varepsilon}^*(1 - p_n)$ of the dynamic extreme quantile $q_{Y_{n+1}}(1 - p_n | \mathcal{F}_n) = \sum_{j=1}^p \phi_j Y_{n+1-j} - \sum_{j=1}^q \psi_j \varepsilon_{n+1-j} + q_\varepsilon(1 - p_n)$ in an ARMA(p, q) model, where $\mathcal{F}_n = \sigma(Y_n, Y_{n-1}, \dots)$ is the past σ -algebra at time n .

Corollary 4.5 (Dynamic extremal regression – ARMA model). *Under the conditions of Theorem 4.1(ii),*

$$\frac{\sqrt{k}}{\log(k/(np_n))} \left(\frac{\widehat{q}_{Y_{n+1}}^*(1-p_n|\mathcal{F}_n)}{q_{Y_{n+1}}(1-p_n|\mathcal{F}_n)} - 1 \right) \xrightarrow{d} \mathcal{N} \left(\frac{\lambda}{1-\rho}, \gamma^2 \right).$$

Theorem 4.1 constitutes an improvement upon currently available results because it does not impose any restriction on k apart from the standard conditions $k/n \rightarrow 0$ and $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$ contrary to Proposition 3.3 in Resnick and Stărică (1997) and Theorem 3.3 in Girard et al. (2021), or on the left tail behavior of the noise variable (see Theorem 2.1 in Ling and Peng, 2004), and it does not require a Hall class-type assumption on the ε_t (see condition (3.1) in Hill, 2015). We note that since SARIMA observations can be turned into ARMA observations after suitable differencing (Brockwell and Davis, 1991, p.324), it is immediate that Theorem 4.1 carries over to SARIMA models as long as one finds estimators of the SARIMA parameters that converge sufficiently fast.

VARMA model This extends to the multivariate setting as follows. For identifiability reasons, consider the VARMA “final equations form” (Lütkepohl, 2005, Definition 12.1) $\mathbf{Y}_t = \sum_{j=1}^p \phi_j \mathbf{Y}_{t-j} - \sum_{j=1}^q \mathbf{M}_j \boldsymbol{\varepsilon}_{t-j} + \boldsymbol{\varepsilon}_t$, for $t \in \mathbb{Z}$, where \mathbf{Y}_t and $\boldsymbol{\varepsilon}_t$ are d -dimensional, $\phi_1, \dots, \phi_p \in \mathbb{R}$ and $\mathbf{M}_1, \dots, \mathbf{M}_q$ are $d \times d$ square matrices. We assume that the matrix-valued polynomial $P(z) = (1 - \phi_1 z - \dots - \phi_p z^p) \mathbf{I}_d$ is left-coprime with $Q(z) = \mathbf{I}_d - z \mathbf{M}_1 - \dots - z^q \mathbf{M}_q$ (i.e. they do not have any nontrivial common factor for left multiplication), that the polynomials $\det P(z)$ and $\det Q(z)$ have no root inside the closed unit disk of the complex plane, and that $(\boldsymbol{\varepsilon}_t)$ is an i.i.d. sequence of centered innovations, so that a stationary solution to the VARMA equations indeed exists and has a stable and invertible representation.

For the sake of simplicity, we focus on the setting where the $\boldsymbol{\varepsilon}_t$ are i.i.d. copies of $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_d)^\top$ having a finite covariance matrix and heavy-tailed components satisfying the condition $\mathbb{P}(\varepsilon_i > x) / \mathbb{P}(|\varepsilon_i| > x) \rightarrow l_i \in (0, 1]$ as $x \rightarrow \infty$ for any $i \in \{1, \dots, d\}$. Given data points $(\mathbf{Y}_{1-p}, \dots, \mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_n)$ generated by the VARMA equations, it follows that, for a future data point $\mathbf{Y}_{n+1} = (Y_{n+1,1}, \dots, Y_{n+1,d})^\top$,

$$\begin{pmatrix} q_{Y_{n+1,1}}(\alpha|\mathcal{F}_n) \\ \vdots \\ q_{Y_{n+1,d}}(\alpha|\mathcal{F}_n) \end{pmatrix} = \sum_{j=1}^p \phi_j \mathbf{Y}_{n+1-j} - \sum_{j=1}^q \mathbf{M}_j \boldsymbol{\varepsilon}_{n+1-j} + \begin{pmatrix} q_1(\alpha) \\ \vdots \\ q_d(\alpha) \end{pmatrix}$$

if \mathcal{F}_n is the σ -algebra generated by the past of the VARMA process at time n and q_i (resp. $q_{Y_{n+1,i}}(\cdot|\mathcal{F}_n)$) denotes the quantile function of ε_i (resp. the dynamic quantile function of $Y_{n+1,i}$ given its past).

Under reasonable structural assumptions, a \sqrt{n} -consistent estimator $(\widehat{\phi}_{n,1}, \dots, \widehat{\phi}_{n,p}, \widehat{\mathbf{M}}_{n,1}, \dots, \widehat{\mathbf{M}}_{n,q})$ of the model parameters can be constructed by a quasi-maximum Gaussian likelihood method, see Theorem 4.3.1 in Hannan and Deistler (2012). This is asymptotically equivalent to Whittle estimation under further regularity conditions, see Theorem 5 and Corollary 3 in Dunsmuir and Hannan (1976). This induces a definition of the vector of residuals at time t as

$$\widehat{\boldsymbol{\varepsilon}}_t^{(n)} = \mathbf{Y}_t - \sum_{j=1}^p \widehat{\phi}_{n,j} \mathbf{Y}_{t-j} + \sum_{j=1}^q \widehat{\mathbf{M}}_{n,j} \widehat{\boldsymbol{\varepsilon}}_{t-j}^{(n)}, \text{ for all } 1 \leq t \leq n,$$

with $\widehat{\varepsilon}_{1-q}^{(n)}, \dots, \widehat{\varepsilon}_0^{(n)}$ being user-specified initial values of the residuals at times $1 - q, \dots, 0$. Adapting the proof of Theorem 4.1 then leads to the following result which, to the best of our knowledge, is the first one considering the problem of extreme value estimation within the VARMA model.

Theorem 4.2 (Residual-based extreme value estimation – VARMA model). *Assume that the time series (\mathbf{Y}_t) satisfies the VARMA(p, q) equations $\mathbf{Y}_t = \sum_{j=1}^p \phi_j \mathbf{Y}_{t-j} - \sum_{j=1}^q \mathbf{M}_j \varepsilon_{t-j} + \varepsilon_t$, for $t \in \mathbb{Z}$, where $\phi_1, \dots, \phi_p \in \mathbb{R}$ and $\mathbf{M}_1, \dots, \mathbf{M}_q$ are $d \times d$ square matrices such that the matrix-valued polynomials $P(z) = (1 - \phi_1 z - \dots - \phi_p z^p) \mathbf{I}_d$ and $Q(z) = \mathbf{I}_d - z \mathbf{M}_1 - \dots - z^q \mathbf{M}_q$ are left-coprime and the polynomials $\det P(z)$ and $\det Q(z)$ have no root inside the closed unit disk of the complex plane, and (ε_t) is an i.i.d. sequence of copies of a centered random vector $\varepsilon = (\varepsilon_1, \dots, \varepsilon_d)$ which has a finite second moment and satisfies condition $\mathcal{C}_2(\boldsymbol{\gamma}, \boldsymbol{\rho}, \mathbf{A})$, with $\gamma_i < 1/2$ for any $i \in \{1, \dots, d\}$, and condition $\mathcal{J}\mathcal{C}(\mathbf{R})$, together with the condition $\mathbb{P}(\varepsilon_i > x) / \mathbb{P}(|\varepsilon_i| > x) \rightarrow l_i \in (0, 1]$ as $x \rightarrow \infty$ for any $i \in \{1, \dots, d\}$.*

Suppose that $\widehat{\phi}_{n,1}, \dots, \widehat{\phi}_{n,p}, \widehat{\mathbf{M}}_{n,1}, \dots, \widehat{\mathbf{M}}_{n,q}$ are \sqrt{n} -consistent estimators of $\phi_1, \dots, \phi_p, \mathbf{M}_1, \dots, \mathbf{M}_q$. Let $\widehat{\varepsilon}_t^{(n)} = \mathbf{Y}_t - \sum_{j=1}^p \widehat{\phi}_{n,j} \mathbf{Y}_{t-j} + \sum_{j=1}^q \widehat{\mathbf{M}}_{n,j} \widehat{\varepsilon}_{t-j}^{(n)}$ be the pertaining residuals, with initial values $\widehat{\varepsilon}_{1-q}^{(n)} \equiv \widehat{\varepsilon}_{1-q}, \dots, \widehat{\varepsilon}_0^{(n)} \equiv \widehat{\varepsilon}_0$ that have a finite first moment and are independent of the sequence (\mathbf{Y}_t) . Finally, let, for $1 \leq j \leq d$, the $k_j = k_j(n) \rightarrow \infty$ satisfy $k_1/k_j \rightarrow c_j \in (0, \infty)$, $k_j/n \rightarrow 0$ and $\sqrt{k_j} A_j(n/k_j) \rightarrow \lambda_j \in \mathbb{R}$, and set $n_j = n$.

(i) Then Theorem 3.2 applies.

(ii) If the $\rho_j < 0$ and $p_{n,j} \downarrow 0$ are such that $k_j/(np_{n,j}) \rightarrow \infty$ and $\sqrt{k_j}/\log(k_j/(np_{n,j})) \rightarrow \infty$ for any $j \in \{1, \dots, d\}$, then Corollary 3.2 applies.

As in Section 4.3, we illustrate the implications of this result on dynamic extremal regression, that is, we consider the estimators

$$\widehat{q}_{Y_{n+1,i}}^*(1 - p_{n,i} | \mathcal{F}_n) = \sum_{j=1}^p \widehat{\phi}_{n,j} Y_{n+1-j,i} - \sum_{j=1}^q \left[\widehat{\mathbf{M}}_{n,j} \widehat{\varepsilon}_{n+1-j}^{(n)} \right]_i + \widehat{q}_{n,i}^*(1 - p_{n,i}), \quad 1 \leq i \leq d,$$

of the dynamic quantiles $q_{Y_{n+1,i}}(1 - p_{n,i} | \mathcal{F}_n) = \sum_{j=1}^p \phi_j Y_{n+1-j,i} - \sum_{j=1}^q [\mathbf{M}_j \varepsilon_{n+1-j}]_i + q_i(1 - p_{n,i})$.

Corollary 4.6 (Dynamic extremal regression – VARMA model). *Under the conditions of Theorem 4.2(ii),*

$$\left(\frac{\sqrt{k_j}}{\log(k_j/(np_{n,j}))} \left\{ \frac{\widehat{q}_{Y_{n+1,j}}^*(1 - p_{n,j} | \mathcal{F}_n)}{q_{Y_{n+1,j}}(1 - p_{n,j} | \mathcal{F}_n)} - 1 \right\} \right)_{1 \leq j \leq d} \xrightarrow{d} \mathcal{N} \left(\left(\frac{\lambda_1}{1 - \rho_1}, \dots, \frac{\lambda_d}{1 - \rho_d} \right), \mathbf{V}_\gamma \right)$$

with the notation of Corollary 3.2, that is, \mathbf{V}_γ is the positive semidefinite $d \times d$ symmetric matrix defined elementwise as $[\mathbf{V}_\gamma]_{j,\ell} = \gamma_j \gamma_\ell R_{j,\ell}(c_\ell, c_j) / \sqrt{c_j c_\ell}$.

It is readily checked that Theorem 4.2 and Corollary 4.6 indeed contain Theorem 4.1 and Corollary 4.5 in the case $d = 1$ when the innovation has a finite second moment.

5 Numerical experiments

5.1 Finite-sample simulation study

We examine how our residual-based extreme value inference methods behave in three settings: noisy data, data from (possibly nonlinear) regression models and data from ARMA models. This is done by reporting Monte-Carlo estimates of the Root Mean Squared Errors (RMSEs) of the extreme value index estimators and of the empirical and Weissman-type quantile estimators. We also compare the performance of asymptotic confidence intervals for inference about the extreme value index and extreme quantiles, constructed according to our results in Section 4. For the sake of brevity, we only comment upon our results, obtained using $N = 1000$ replicated samples in the models we describe below. A complete set of results, along with full details of how our RMSE estimates and asymptotic confidence intervals are constructed, can be found in Appendix D.

Noisy data We consider two models where, for $1 \leq i \leq n$, $n \in \{500, 2000\}$, the i th data point Y_i is obtained by adding noise to Z_i , where the Z_i are unobserved, i.i.d. copies of a Fréchet random variable with extreme value index $\gamma \in \{0.2, 0.4\}$, *i.e.* $\mathbb{P}(Z \leq z) = \exp(-z^{-1/\gamma})$ for $z > 0$. The models are:

(C) $Y_i = (1 + \varphi(Z_i/\sigma))Z_i$, where φ is taken among $\varphi_1(x) = \exp(-x)$, $\varphi_2(x) = \exp(-x^2)$, $\varphi_3(x) = \min(\varphi_1(x), \varphi_2(x))$ or $\varphi_4(x) = \max(1 - x, 0)$, and σ is either $\sigma_1 = 1/2$ or $\sigma_2 = 2$.

(L) $Y_i = (10^{-l_i} [10^{l_i} Z_i] - Z_i) + Z_i$, where either $l_i = l \in \{0, 1, 2\}$, corresponding to the model of Matsui et al. (2013), or the l_i are drawn at random in $\{0, 1, 2\}$ with probabilities $(p_0, p_1, 1 - p_0 - p_1)$, according to the distributions $\mathbb{P}_1, \mathbb{P}_2$ and \mathbb{P}_3 having $(p_0, p_1) = (2/3, 1/6), (1/6, 2/3)$ and $(1/6, 1/6)$ respectively.

We consider the Hill, intermediate quantile and extreme quantile estimators of γ , $q(1 - k/n) = q_Z(1 - k/n)$ and $q(1 - p) = q_Z(1 - p)$ based on the observed data Y_i , that is,

$$\widehat{\gamma}_n(k) = \frac{1}{k} \sum_{i=1}^k \log Y_{n-i+1:n} - \log Y_{n-k:n}, \quad \widehat{q}_n(1 - k/n) = Y_{n-k:n} \quad \text{and} \quad \widehat{q}_n^*(1 - p) = \left(\frac{k}{np}\right)^{\widehat{\gamma}_n(k)} Y_{n-k:n}.$$

The extreme quantile level is $p = 1 - 10/n \in \{0.98, 0.995\}$. These estimators are compared to those that would be constructed if the data Z_i were available. We take $k = 75$ when $n = 500$ (*i.e.* 15% of the sample size) and $k = 100$ when $n = 2000$ (*i.e.* 5% of the sample size) for the sake of simplicity.

Tables D.1–D.4 report the results in model (C). A higher scale parameter entails a greater degree of data corruption, resulting in worse results overall. Inference-wise, results are typically best, and closest to those given by the raw data, for $\varphi = \varphi_4$. This is reasonable because this function is, among the φ_j , the one for which the $|Y_i - Z_i| = \varphi(Z_i/\sigma)Z_i$ are generally the smallest, see Figure D.1. Point estimation results tend to improve with increasing n , suggesting that noise matters less when sample size increases, and with increasing γ , because the $|Y_i - Z_i|/Z_i = \varphi(Z_i/\sigma)$ are smaller for larger Z_i .

Tables D.5–D.8 report the results in model (L). As expected, a stronger degree of truncation, represented by a lower value of l (when it is constant) or a higher probability of generating a low value of l (when it is random), generally yields worse results. Results tend to improve with increasing n , but not

necessarily with increasing γ , even though truncating larger values should be less consequential than truncating smaller values. This suggests that rather than the shape of the distribution, it is its scale that matters in ascertaining whether truncation is detrimental.

Location-scale regression We consider here homoskedastic regression models of the form $Y_i = m(X_i) + \varepsilon_i$, for $1 \leq i \leq n$, $n \in \{2000, 10000\}$, for different types of regression functions m , namely:

(R_a) The linear function $m(X) = 10 + 5X$,

(R_b) The non-linear function $m(X) = 10 + 5X + (2X - 1)_+ + (2X - 1)_+^2$, where $t_+ = \max(0, t)$,

(R_c) The non-linear, high-frequency function $m(X) = 10 + 5X + (2X - 1)_+ + (2X - 1)_+^2 + 10 \sin(25X)$.

In addition, we consider the heteroskedastic regression model $Y_i = m(X_i) + \sigma(X_i)\varepsilon_i$, where the location component is $m(X) = 10 + 5X$ and:

(R_d) The scale component is $\sigma(X) = \exp(1 + X)$,

(R_e) The scale component is $\sigma(X) = \exp(1 + X + 0.5 \sin(2\pi X))$.

The ε_i are i.i.d. replications of a noise variable ε having a symmetric Fréchet distribution, *i.e.* $\varepsilon = \delta Z$, with $\mathbb{P}(\delta = 1) = \mathbb{P}(\delta = -1) = 1/2$, Z is a Fréchet random variable with extreme value index $\gamma = 0.2$, and δ is independent of Z . The covariates X_i , independent of the ε_i , are i.i.d. and either uniformly distributed on $[0, 1]$ or Fréchet distributed with extreme value index $\gamma/2 = 0.1$ (and, thus, possess lighter tails than the errors). We take $p = 1 - 10/n \in \{0.995, 0.999\}$.

Following the methodology of Section 4.2, we compare the residual-based Hill estimator of γ , the residual-based empirical quantile estimator of the intermediate quantile $q(1 - k/n) = q_\varepsilon(1 - k/n)$, with $k = 100$ when $n = 2000$ (*i.e.* 5% of the sample size) and $k = 250$ when $n = 10000$ (*i.e.* 2.5% of the sample size), and the residual-based Weissman estimator of the extreme quantile $q(1 - p) = q_\varepsilon(1 - p)$. We also compare the performance of the asymptotic Gaussian confidence intervals for the conditional intermediate and extreme quantiles $q_{Y|X}(1 - k/n|x)$ and $q_{Y|X}(1 - p|x)$. In models (R_a), (R_b) and (R_c), the residuals are obtained after fitting the following models to the location component m :

- No model at all, *i.e.* the raw data Y_i is used;
- A linear fit with ordinary least squares through the R function `lm`;
- Locally polynomial regression models, *i.e.* a local linear and a local quadratic fit, obtained with the R function `loess` based on the S implementation by Chambers and Hastie (1992);
- Cubic smoothing splines, obtained using the R function `smooth.spline` based on the S implementation by Chambers and Hastie (1992).

In models (R_d) and (R_e), the procedure to get the residuals is the following:

1. Fit an ordinary least squares model to the Y_i given the X_i , and fetch the residuals $\hat{\varepsilon}_i^{(n)}$;

2. Motivated by the identity $\log(|Y - m(X)|) = \mathbb{E}(\log |\varepsilon|) + \log \sigma(X) + (\log |\varepsilon| - \mathbb{E}(\log |\varepsilon|))$, estimate $\sigma(x)$ (up to a multiplicative constant) by regressing the $\log(|\hat{\varepsilon}_i^{(n)}|)$ upon the X_i using either a linear fit, a locally polynomial regression model, or cubic smoothing splines as explained above;
3. Using the estimate $\hat{\sigma}(x)$ obtained at the previous step, fit a weighted least squares model with weights $w_i = 1/\hat{\sigma}^2(X_i)$ to the Y_i given the X_i , producing an estimate $\hat{m}(x)$, and fetch the standardized residuals $\hat{\varepsilon}_i^{(n)} = (Y_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$.

All of these residual-based methods are compared to what would be obtained if the ε_i were observed and the m and σ functions were known (the so-called ‘‘Oracle’’).

Tables D.9–D.18 report the results. We observe that (i) ignoring the influence of the covariate produces very poor estimates of γ , (ii) intermediate conditional quantiles are more difficult to infer than extreme conditional quantiles due to the mean component having a greater importance at intermediate levels, (iii) under-specification of the regression model is detrimental to finite-sample performance, while over-specification generally has no substantial influence, and (iv) in homoskedastic models, both the intermediate and extreme quantiles of the ε_i are easier to infer than their conditional counterparts. We also note that, in model (R_c), the RMSEs stay acceptable when the covariates are heavy-tailed but coverage rates sharply deteriorate. This may be due to increased difficulty in estimating the function m in this context. All in all, the presence of nonlinearities and/or high-frequency components make pure parametric fits and low-degree local polynomial fits unsuitable for producing residuals that may be used in a conditional extreme value analysis. That being said, in contrast to what was observed for extreme conditional expectile inference in Girard et al. (2021), asymptotic Gaussian inference appears to work fairly well for extreme conditional quantiles as long as model components are estimated correctly.

ARMA time series Here the n data points, with $n \in \{2000, 10000\}$, are generated by an ARMA(1, 1) time series model. Let the innovations ε_t follow an i.i.d. symmetric Fréchet sequence with extreme value index $\gamma \in \{0.2, 0.4\}$. We consider the following four data-generating processes:

$$(T_1) \quad Y_{t+1} = 1 + 0.3(Y_t - 1) + \varepsilon_{t+1} + 0.3\varepsilon_t,$$

$$(T_2) \quad Y_{t+1} = 1 + 0.3(Y_t - 1) + \varepsilon_{t+1} + 0.9\varepsilon_t,$$

$$(T_3) \quad Y_{t+1} = 1 + 0.9(Y_t - 1) + \varepsilon_{t+1} + 0.3\varepsilon_t,$$

$$(T_4) \quad Y_{t+1} = 1 + 0.9(Y_t - 1) + \varepsilon_{t+1} + 0.9\varepsilon_t.$$

They correspond to, respectively, a weakly dependent ARMA process, a weakly dependent ARMA process with large moving average component, a strongly dependent ARMA process with small moving average component, and a strongly dependent ARMA process with large moving average component.

Following Section 4.3, we compare the residual-based Hill estimator of γ , the residual-based empirical quantile estimator of the intermediate quantile $q(1 - k/n) = q_\varepsilon(1 - k/n)$, with $k = 100$ when $n = 2000$ and $k = 250$ when $n = 10000$, and the residual-based Weissman estimator of the extreme quantile $q(1 - p) = q_\varepsilon(1 - p)$, for $p = 1 - 10/n \in \{0.995, 0.999\}$, obtained after fitting the following models:

- No model at all, *i.e.* the raw time series is used;
- The AR(1), AR(2), MA(1) and MA(2) models, which are all under-specified;
- A correctly specified ARMA(1, 1) model;
- The ARMA(2, 1), ARMA(1, 2) and ARMA(2, 2) models, which are all over-specified.

We also compare the performance of the asymptotic Gaussian confidence intervals for the dynamic intermediate and extreme quantiles obtained through these models. The coefficients of the fitted AR(p) and ARMA(p, q) models are obtained via the Yule-Walker equations and Gaussian maximum likelihood, respectively, using the routines `ar.yw` and `arima` from the base R package `stats`. When convergence issues appear in the maximization of the likelihood we simply fit the null model, *i.e.* the model with a constant term corresponding to the ARMA(0, 0) model, fitted using the empirical mean. Fitted residuals are obtained from the `arima0` routine from the base R package `stats`, which uses backpropagation based on Kalman filters, see Gardner et al. (1980). Note that in the vast majority of cases, this provides virtually the same results as initializing the first residuals at 0 and calculating subsequent residuals using the fitted ARMA equations, but the Kalman filter approach has the advantage of providing numerically stable optimization of the likelihood and calculation of the residuals.

Results are reported in Tables D.19–D.22. As observed in the regression examples above, not accounting for serial dependence is very detrimental to finite-sample performance when estimating and inferring γ , and the intermediate conditional quantile inference problem is generally not easier than its counterpart at extreme levels, due to model structure playing a proportionally larger role in the value of the intermediate conditional quantile estimates. Over-specification of the model typically does not affect results substantially, with RMSEs and coverage rates staying close to those that would be achieved if the model were exactly known, with the caveat that convergence of the maximum likelihood algorithm can get more difficult as more coefficients are added. By contrast, under-specification has visible consequences, although it seems to matter more when the ARMA structure is far from independence, and when the innovations have a tail that is not too heavy. As expected, an under-specified AR (resp. MA) model performs well if the AR (resp. MA) coefficient is high and the MA (resp. AR) coefficient is low. Interestingly, though, not accounting for strong MA effects seems to be more consequential than not accounting for strong AR effects; our interpretation is that strong MA components at low lags are properly accounted for in MA models, but are poorly represented in AR models with low coefficients. Finally, as in the regression examples, we observe that asymptotic Gaussian inference for extreme conditional quantiles works fairly well when the model is correctly or over-specified. This stands in contrast to the extreme conditional expectile inference problem, see Girard et al. (2021).

Acknowledgments

Financial support from the French CNRS within the project “Extreme value analysis of time series through de-randomization techniques”, funded through the IEA program, is gratefully acknowledged. G. Stupfler acknowledges further financial support from the French *Agence Nationale de la Recherche*

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.
- Billingsley, P. (1995). *Probability and Measure (Third Edition)*. John Wiley and Sons.
- 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.
- Chambers, J. M. and Hastie, T. J. (1992). *Statistical Models in S*. Wadsworth & Brooks/Cole.
- 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.
- Davis, R. A. (1996). Gauss-Newton and M-estimation for ARMA processes with infinite variance. *Stochastic Processes and their Applications*, 63(1):75–95.
- 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.
- Dunsmuir, W. and Hannan, E. J. (1976). Vector linear time series models. *Advances in Applied Probability*, 8(2):339–364.
- 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.
- 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.
- Gardner, G., Harvey, A. C., and Phillips, G. D. A. (1980). Algorithm AS 154: An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering. *Journal of the Royal Statistical Society: Series C*, 29(3):311–322.
- 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*, 203(1):3059–3072.
- Hall, P. and Simar, L. (2002). Estimating a changepoint, boundary, or frontier in the presence of observation error. *Journal of the American Statistical Association*, 97(458):523–534.
- Hannan, E. J. and Deistler, M. (2012). *The Statistical Theory of Linear Systems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia.
- 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.
- Kim, M. and Kokoszka, P. (2020). Consistency of the Hill estimator for time series observed with measurement errors. *Journal of Nonparametric Statistics*, 41(3):421–435.

- Kim, M. and Kokoszka, P. (2023). Asymptotic and finite sample properties of Hill-type estimators in the presence of errors in observations. *Journal of Nonparametric Statistics*, 35(1):1–18.
- 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.
- Ling, S. and Peng, L. (2004). Hill’s estimator for the tail index of an ARMA model. *Journal of Statistical Planning and Inference*, 123(2):279–293.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer Science and Business Media.
- Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., and Morin, E. (2018). Metastatistical Extreme Value analysis of hourly rainfall from short records: Estimation of high quantiles and impact of measurement errors. *Advances in Water Resources*, 117:27–39.
- 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.
- Matsui, M., Mikosch, T., and Tafakori, L. (2013). Estimation of the tail index for lattice-valued sequences. *Extremes*, 16(4):429–455.
- 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.
- Resnick, S. and Stărică, C. (1997). Asymptotic behavior of Hill’s estimator for autoregressive data. *Communications in Statistics – Stochastic Models*, 13(4):703–721.
- 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.
- Velthoen, J., Cai, J.-J., Jongbloed, G., and Schmeits, M. (2019). Improving precipitation forecasts using extreme quantile regression. *Extremes*, 22(4):599–622.
- 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.
- Zhu, K. and Ling, S. (2012). The global weighted LAD estimators for finite/infinite variance ARMA(p, q) models. *Econometric Theory*, 28(5):1065–1086.