



HAL
open science

Inference for extremal regression with dependent heavy-tailed data

Abdelaati Daouia, Gilles Stupfler, Antoine Usseglio-Carleve

► **To cite this version:**

Abdelaati Daouia, Gilles Stupfler, Antoine Usseglio-Carleve. Inference for extremal regression with dependent heavy-tailed data. 2022. hal-03612202v1

HAL Id: hal-03612202

<https://hal.science/hal-03612202v1>

Preprint submitted on 17 Mar 2022 (v1), last revised 23 Aug 2023 (v3)

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.

Inference for extremal regression with dependent heavy-tailed data

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

^a Toulouse School of Economics, University of Toulouse Capitole, France

^b Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France

^c Avignon Université, LMA EA 2151, 84000 Avignon, France

Abstract. Nonparametric inference on tail conditional quantiles and their least squares analogs, expectiles, remains limited to i.i.d. data. Expectiles are themselves quantiles of a transformation of the underlying distribution. We develop a fully operational kernel-based inferential theory for extreme conditional quantiles and expectiles in the challenging framework of α -mixing, conditional heavy-tailed data whose tail index may vary with covariate values. This extreme value problem requires a dedicated treatment to deal with data sparsity in the far tail of the response, in addition to handling difficulties inherent to mixing, smoothing, and sparsity associated to covariate localization. We prove the pointwise asymptotic normality of our estimators and obtain optimal rates of convergence reminiscent of those found in the i.i.d. regression setting, but which had not been established in the conditional extreme value literature so far. Our mathematical assumptions are satisfied in location-scale models with possible temporal misspecification, nonlinear regression models, and autoregressive models, among others. We propose full bias and variance reduction procedures, and simple but effective data-based rules for selecting tuning hyperparameters. Our inference strategy is shown to perform well in finite samples and is showcased in applications to stock returns and tornado loss data.

Keywords. Conditional quantiles, conditional expectiles, extreme value analysis, heavy tails, inference, mixing, nonparametric regression.

MSC2020 classification: 62G05, 62G08, 62G15, 62G20, 62G30, 62G32.

1 Introduction

1.1 Background and motivation

Quantile regression is a well-established statistical tool for the assessment of the impact of a vector of covariates $\mathbf{X} \in \mathbb{R}^p$ upon a response variable Y . It fully describes the conditional distribution of Y given \mathbf{X} by considering the family of conditional quantiles

$$q(\tau|\mathbf{x}) = \arg \min_{\theta \in \mathbb{R}} \mathbb{E}([\varrho_\tau(Y - \theta) - \varrho_\tau(Y)]|\mathbf{X} = \mathbf{x}), \quad \tau \in (0, 1),$$

where $\varrho_\tau(y) = |\tau - \mathbb{1}_{\{y \leq 0\}}||y|$ denotes the quantile check function (throughout, $\mathbb{1}_A$ stands for the indicator function of A). However, regression quantile estimators at the tails typically suffer from instability and inconsistency due to data sparseness, especially when the underlying conditional distributions are heavy-tailed. This class of distributions is ubiquitous in numerous fields, such as insurance (large losses due to large claims), financial

economics (large drops in market or stock indices, potentially leading to systemic risk), public health (high pollution levels) and natural sciences (large earthquake magnitudes, flood intensity, extreme rainfall). Existing approaches to extremal quantile regression in the challenging heavy-tailed setting fall in two main categories: linear quantile regression approaches, such as those of Chernozhukov and Fernández-Val (2011) and Wang et al. (2012), and at the opposite, nonparametric approaches such as those of Daouia et al. (2011) and Daouia et al. (2013). Yet, a major hurdle for the application of tail quantile regression in practice is that the theory in this literature generally hinges on the assumption of independent and identically distributed (i.i.d.) data. To the best of our knowledge, only Chernozhukov and Fernández-Val (2011) provide feasible inference tools for geometrically fast strong mixing (or α -mixing) data, but they assume that conditional quantiles $q(\tau|\mathbf{x})$ are linear in \mathbf{x} , they impose a constant extreme value index, and they appeal to self-normalized quantile regression statistics rather than to conceptually simpler and more intuitive asymptotic Gaussian confidence intervals. More broadly, the problem of inference on nonlinear extremal quantile regression remains untouched under serial dependence, even though heavy tails and dependence across time and space are commonplace in applications in insurance, finance and the natural sciences, see for example (Beirlant et al., 2004, Chapter 1.3) and (Resnick, 2007, p.1).

There are applications where certain features of quantile regression are disadvantages. In actuarial and financial risk management, robustness of quantiles may constitute a weakness as they are not sensitive and hence not alert to the severity of extreme losses. Their failure to satisfy the fundamental axiomatic coherence property (introduced in Artzner et al., 1999) in general is also a serious drawback. A better alternative in these respects is expectile regression (Newey and Powell, 1987), which focuses on the conditional expectiles

$$e(\tau|\mathbf{x}) = \arg \min_{\theta \in \mathbb{R}} \mathbb{E}([\eta_\tau(Y - \theta) - \eta_\tau(Y)] | \mathbf{X} = \mathbf{x}), \quad \tau \in (0, 1), \quad (1.1)$$

where $\eta_\tau(y) = |\tau - \mathbb{1}_{\{y \leq 0\}}| y^2$ is an asymmetric quadratic loss function, in the same way $\varrho_\tau(y)$ defines an asymmetric absolute loss function. Expectiles similarly fully describe the conditional distribution of Y given \mathbf{X} , by extending the mean as quantiles extend the median. They induce the only coherent risk measure that is also elicitable (Ziegel, 2016), meaning that they come endowed with a natural methodology for backtesting and forecast verification. Besides, the τ th expectile is in fact precisely the τ th quantile of a transformed conditional distribution function defined by Jones (1994), in the sense that

$$e(\tau|\mathbf{x}) = \inf\{y \in \mathbb{R} | E(y|\mathbf{x}) \geq \tau\} \text{ with } E(y|\mathbf{x}) = \frac{\mathbb{E}[|Y - y| \mathbb{1}_{\{Y \leq y\}} | \mathbf{X} = \mathbf{x}]}{\mathbb{E}[|Y - y| | \mathbf{X} = \mathbf{x}]}. \quad (1.2)$$

Regression expectiles thus make an efficient use of the data, since they rely on the distance to all observations and not only on their probability, and they benefit from a transparent financial meaning in terms of their acceptance sets and the gain-loss ratio (Bellini and Di Bernardino, 2017). These properties and others (see *e.g.* Ehm et al., 2016, for an interpretation in investment problems with fixed costs and differential tax rates) have sparked much renewed interest in expectiles over the last decade for their unquestionable potential in defining an efficient instrument of risk protection. Only Usseglio-Carleve (2018) and Girard et al. (2021a, 2022) have initiated the development of extremal expectile regression. The pioneering contribution of Usseglio-Carleve (2018) is limited to elliptical heavy-tailed distributions, and the nonparametric approach of Girard et al. (2022) hinges on the i.i.d. assumption. The approach of Girard et al. (2021a) can handle time series location-scale models, but is highly sensitive to model misspecification, makes the strong assumption of a constant tail index, and their bootstrap scheme is difficult to calibrate.

1.2 Contribution and outline of the paper

We propose a general and fully operational nonparametric inferential theory for conditional tail quantiles and expectiles under the α -mixing dependence framework, which is one of the weakest and widely used serial dependence conditions in the literature. The tail index of the conditional heavy-tailed distributions is allowed to depend on covariate values. Our overarching strategy is that a quantile can be estimated by inverting an estimator of the associated distribution function. This is of course also relevant to expectile estimation in view of (1.2), and it is in fact very beneficial theoretically as the investigation of properties of nonparametric tail conditional expectile estimators directly via (1.1) would be quite technically involved (see Daouia et al., 2018, in the unconditional setting). When the level τ is intermediate (“extreme, but not too much”) in the sense that $\tau = \tau_n \uparrow 1$ slowly enough as the sample size $n \rightarrow \infty$, so that the population quantile or expectile curve at $\mathbf{X} = \mathbf{x}$ remains inside the data cloud, an asymptotically normal estimator is then obtained by inverting a kernel smoother of the associated conditional distribution function. The intermediate level τ_n is assumed to satisfy $nh_n^p(1 - \tau_n) \rightarrow \infty$ as $n \rightarrow \infty$, where $h_n > 0$ is the bandwidth sequence featuring in the kernel estimator. At more extreme levels $\tau = \tau'_n > \tau_n$ such that $nh_n^p(1 - \tau'_n) = O(1)$, the intermediate regression estimators are extrapolated following a technique of Weissman (1978) by making use of the heavy right tail assumption. The resulting extrapolated estimator hinges upon a conditional tail index estimator that can be constructed based on quantiles or expectiles.

Our asymptotic theory is obtained, on the one hand, under traditional assumptions in nonparametric statistics on the marginal density of \mathbf{X} and kernel function, and on the other hand, under reasonable regularity assumptions on the variation of the conditional distribution across the covariate space to control locally the oscillation of marginal conditional extremes and the degree of clustering in joint conditional extremes and covariate values. Under these assumptions, we derive the asymptotic normality of our estimators at rates of convergence reminiscent of the optimal rates obtained in classical nonparametric statistics under twice differentiability of the estimated function, which had not been found so far in the conditional extreme value literature. Our conditions are shown to be satisfied in a number of important regression models, including location-scale models with possible temporal misspecification, nonlinear regression models and autoregressive models. Surprisingly perhaps, the asymptotic distribution of the intermediate estimators is exactly the same as in the nonparametric i.i.d. setting (Daouia et al., 2011; Girard et al., 2022), which is however not the case for empirical intermediate quantiles and expectiles in the absence of covariates (Drees, 2003; Davison et al., 2021). The simple structure of the Weissman-type estimators at extreme levels, meanwhile, makes it possible to investigate very accurate bias-corrected versions and precise approximations to the actual variance of the estimators in finite samples, thus enabling the construction of asymptotic Gaussian confidence intervals. This had been observed to be very difficult for unconditional tail expectile inference (see a partial solution in Padoan and Stupfler, 2022). The bias correction methodology revolves upon identifying bias sources due to the extrapolation procedure, while the variance correction relies on evaluating the correlation between intermediate estimators and tail index estimators, and (for expectiles) eliminating variance distortions incurred as part of the linearizations necessary to obtain our asymptotic results. Of course, inherent to any extreme value kernel estimators are the choices of the bandwidth h_n which controls their smoothness and of the sample fraction $1 - \tau_n$ of the top local observations needed for tail extrapolation. We address this issue by proposing innovative rules-of-thumb, with the resulting confidence intervals achieving excellent coverage already for moderately large sample sizes, in both regression modes.

The outline of our paper is the following. Sections 2 and 3 focus on nonparametric extremal quantile and expectile regression, respectively. Section 4 investigates examples of regression models where our assumptions are satisfied. Section 5 develops a fully operational inferential methodology, showcased in a simulation study in Section 6 and on real data in Section 7. Our methods and data have been incorporated into the R package `Expectrem`¹. Throughout we denote by $x_+ = \max(x, 0)$ and $x_- = \max(-x, 0)$ the positive and negative parts of a real number x . For a function f on \mathbb{R}^p , $\nabla f(\mathbf{x})$, $Jf(\mathbf{x})$ and $Hf(\mathbf{x})$ stand respectively for its gradient vector, Jacobian matrix, and Hessian matrix at the point \mathbf{x} . For a function $f = f(\mathbf{x}, y)$ on $\mathbb{R}^p \times \mathbb{R}^q$, let $\nabla_{\mathbf{x}}f$ and $H_{\mathbf{x}}f$ be its partial gradient vector and Jacobian matrix with respect to \mathbf{x} (*i.e.* its first p components). The symbols $\mathbf{0}_p$ and $\mathbf{1}_p$ denote vectors in \mathbb{R}^p with all components equal to 0 and 1 respectively.

2 Nonparametric extremal quantile regression

2.1 Framework and estimation method

Let $((\mathbf{X}_t, Y_t))_{t \geq 1}$ be a strictly stationary sequence of copies of a random vector $(\mathbf{X}, Y) \in \mathbb{R}^{p+1}$. Let $F(\cdot|\mathbf{x})$ denote the distribution function of Y given $\mathbf{X} = \mathbf{x}$, that is, $F(y|\mathbf{x}) = \mathbb{P}(Y \leq y | \mathbf{X} = \mathbf{x})$. Assume that \mathbf{X} has a probability density function (p.d.f.) g on \mathbb{R}^p and fix $\mathbf{x} \in \mathbb{R}^p$ with $g(\mathbf{x}) > 0$. Consider the following kernel estimator of $F(\cdot|\mathbf{x})$:

$$\hat{F}_n(y|\mathbf{x}) = \frac{1}{\hat{g}_n(\mathbf{x})} \times \frac{1}{nh_n^p} \sum_{t=1}^n \mathbb{1}_{\{Y_t \leq y\}} K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_n}\right)$$

with $\hat{g}_n(\mathbf{x}) = \frac{1}{nh_n^p} \sum_{t=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_n}\right)$.

Hereafter K is a kernel p.d.f. on \mathbb{R}^p and $h_n \rightarrow 0$ is a (positive) bandwidth sequence, with \hat{g}_n being the associated classical Parzen-Rosenblatt estimator of g . A conditional quantile $q(\tau|\mathbf{x}) \equiv \inf \{y \in \mathbb{R} | F(y|\mathbf{x}) \geq \tau\}$ can then be estimated by its empirical counterpart

$$\hat{q}_n(\tau|\mathbf{x}) = \inf \left\{ y \in \mathbb{R} | \hat{F}_n(y|\mathbf{x}) \geq \tau \right\}.$$

Our main objective is to estimate extreme conditional quantiles at a high level $\tau = \tau_n$ converging to 1 at any possible rate, as $n \rightarrow \infty$. The standard roadmap for solving such an extreme value problem is to consider first the so-called intermediate (“extreme, but not too much”) levels τ_n , for which $\hat{q}_n(\tau_n|\mathbf{x})$ is a (relatively) consistent estimator of $q(\tau_n|\mathbf{x})$. Then, for the estimation of properly extreme quantiles $q(\tau'_n|\mathbf{x})$ without any restriction on the rate of convergence of $\tau'_n > \tau_n$ to 1, one extrapolates such intermediate estimators to the right place using the shape of the tail of the underlying conditional distribution.

In the challenging maximum domain of attraction of conditional heavy-tailed distributions, the expressions of the intermediate and extrapolated estimators crucially depend on estimators of the tail index $\gamma(\mathbf{x}) > 0$ of the survival function $\bar{F}(\cdot|\mathbf{x}) = 1 - F(\cdot|\mathbf{x})$ of Y at $\mathbf{X} = \mathbf{x}$, which in its most general formulation is assumed to be regularly varying at infinity with index $-1/\gamma(\mathbf{x})$. The quantity $\gamma(\mathbf{x})$ controls the tail heaviness of Y given $\mathbf{X} = \mathbf{x}$: the higher $\gamma(\mathbf{x})$, the heavier the conditional upper tail of the response variable Y . Second-order refinements of this assumption are inevitable whenever one is interested in convergence rates or limiting distributions involving tail index estimators, see Beirlant

¹Available on GitHub at <https://github.com/AntoineUC/Expectrem>

et al. (2004) and de Haan and Ferreira (2006). Our conditional second-order regularly varying tails assumption has the following general formulation:

Condition $\mathcal{C}_2(\gamma(\mathbf{x}), \rho(\mathbf{x}), A(\cdot|\mathbf{x}))$ There exist $\gamma(\mathbf{x}) > 0$, $\rho(\mathbf{x}) \leq 0$ and a positive or negative measurable function $A(\cdot|\mathbf{x})$ converging to 0 at infinity such that for any $y > 0$,

$$\lim_{s \rightarrow \infty} \frac{1}{A(1/\bar{F}(s|\mathbf{x})|\mathbf{x})} \left(\frac{\bar{F}(sy|\mathbf{x})}{\bar{F}(s|\mathbf{x})} - y^{-1/\gamma(\mathbf{x})} \right) = \begin{cases} y^{-1/\gamma(\mathbf{x})} \frac{y^{\rho(\mathbf{x})/\gamma(\mathbf{x})} - 1}{\gamma(\mathbf{x})\rho(\mathbf{x})} & \text{if } \rho(\mathbf{x}) < 0, \\ y^{-1/\gamma(\mathbf{x})} \frac{\log(y)}{\gamma^2(\mathbf{x})} & \text{if } \rho(\mathbf{x}) = 0. \end{cases}$$

This standard condition, wherein $A(\cdot|\mathbf{x})$ is regularly varying with index $\rho(\mathbf{x})$ (see de Haan and Ferreira, 2006, Theorem 2.3.3 p.44), controls the proximity between the extremes of the underlying conditional distribution and those of the ideal Pareto distribution with extreme value index $\gamma(\mathbf{x})$. We therefore make the fundamental modeling assumption that

Condition \mathcal{M} $((\mathbf{X}_t, Y_t))_{t \geq 1}$ is a stationary α -mixing sequence of copies of a random vector (\mathbf{X}, Y) satisfying the second-order regularly varying tails assumption $\mathcal{C}_2(\gamma(\mathbf{x}), \rho(\mathbf{x}), A(\cdot|\mathbf{x}))$.

The α -mixing (or strong mixing) assumption is one of the weakest dependence conditions in the literature. More restrictive assumptions that will be employed below to relax certain regularity conditions involve the β -, ρ -, ϕ - and ψ -mixing coefficients, see Section A.1 for definitions and relationships between the different kinds of mixing. We make the following assumption on the α -mixing coefficients $(\alpha(n))_{n \geq 1}$ of the data $((\mathbf{X}_t, Y_t))_{t \geq 1}$:

Condition $\mathcal{A}(l_n, r_n)$ There exist sequences (l_n) and (r_n) such that $l_n \rightarrow \infty$, $r_n \rightarrow \infty$, $l_n/r_n \rightarrow 0$, $r_n/n \rightarrow 0$ and $n\alpha(l_n)/r_n \rightarrow 0$ as $n \rightarrow \infty$.

The sequences (l_n) and (r_n) formalize the intuitive notion of small-block and big-block sequences used to develop a “big blocks separated by small blocks” argument of the kind successfully used in the time series literature so as to derive asymptotic normality results.

Our extreme value analysis also requires reasonable regularity assumptions on the kernel function and the probabilistic behavior of the covariates. Let $\|\cdot\|$ denote the Euclidean norm on \mathbb{R}^p and denote by $B(\mathbf{x}, r)$ the open Euclidean ball with center \mathbf{x} and radius $r > 0$. Full details for the rationale behind our conditions, their interpretation and position with respect to assumptions in the literature are given in Section A.2.

Condition \mathcal{K} The p.d.f. K is bounded with a support contained in the unit closed Euclidean ball.

Condition \mathcal{L}_g The p.d.f. g satisfies $g(\mathbf{x}) > 0$ and is Lipschitz continuous at \mathbf{x} : there exist $c, r > 0$ such that for any $\mathbf{x}' \in B(\mathbf{x}, r)$, $|g(\mathbf{x}) - g(\mathbf{x}')| \leq c\|\mathbf{x} - \mathbf{x}'\|$.

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

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

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

Assumptions \mathcal{K} , \mathcal{L}_g and \mathcal{B}_p (whose first half is trivially true if t_0 can be chosen equal to 1) are in particular imposed to control the asymptotic behavior of $\hat{g}_n(\mathbf{x})$. To control the variation in conditional extreme value behavior across the covariate space, we introduce a Lipschitz-type assumption on the log-conditional survival function at extreme levels.

Condition \mathcal{L}_ω There exists $r > 0$ such that

$$\limsup_{y \rightarrow \infty} \sup_{\substack{\mathbf{x}' \in B(\mathbf{x}, r) \\ \mathbf{x}' \neq \mathbf{x}}} \frac{1}{\|\mathbf{x}' - \mathbf{x}\|} \left| \frac{1}{\log(y)} \log \frac{\bar{F}(y|\mathbf{x}')}{\bar{F}(y|\mathbf{x})} \right| < \infty.$$

We also impose an anti-clustering condition that translates into assuming that a joint conditional extreme value of (Y_1, Y_{t+1}) cannot be much more likely than a marginal conditional extreme of Y_1 , uniformly across time and locally uniformly across the covariate space. Let

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

Condition \mathcal{B}_Ω There exist $h, z > 0$ such that $\Omega_h(z|\mathbf{x}) < \infty$.

The aforementioned conditions will ensure the pointwise asymptotic normality of our estimators at rates of convergence that have hitherto been standard in the conditional extreme value framework. Achieving optimal rates of convergence requires, similarly to classical nonparametric estimation, stronger regularity conditions: when estimating, for instance, the p.d.f. g in \mathbb{R}^p with $\hat{g}_n(\mathbf{x})$ using a symmetric kernel p.d.f. K , it is well-known that the optimal rate of convergence $n^{-2/(p+4)}$ is obtained by solving the bias-variance tradeoff if g is twice differentiable at \mathbf{x} . This motivates the following additional assumptions.

Condition \mathcal{KS} The p.d.f. K is bounded and symmetric (*i.e.* $K(\mathbf{u}) = K(-\mathbf{u})$) with a support contained in the unit closed Euclidean ball.

Condition \mathcal{D}_g The p.d.f. g satisfies $g(\mathbf{x}) > 0$, is continuously differentiable in a neighborhood of \mathbf{x} and its gradient is Lipschitz continuous at \mathbf{x} .

Condition \mathcal{D}_ω For y large enough, the function $\bar{F}(y|\cdot)$ is differentiable at \mathbf{x} , the function $y \mapsto \nabla_{\mathbf{x}} \log \bar{F}(y|\mathbf{x}) / \log(y)$ has a limit $\boldsymbol{\mu}(\mathbf{x}) \in \mathbb{R}^p$ as $y \rightarrow \infty$, and there exists $r > 0$ with

$$\limsup_{y \rightarrow \infty} \sup_{\substack{\mathbf{x}' \in B(\mathbf{x}, r) \\ \mathbf{x}' \neq \mathbf{x}}} \frac{1}{\|\mathbf{x}' - \mathbf{x}\|^2} \left| \frac{1}{\log(y)} \log \frac{\bar{F}(y|\mathbf{x}')}{\bar{F}(y|\mathbf{x})} - (\mathbf{x}' - \mathbf{x})^\top \frac{\nabla_{\mathbf{x}} \log \bar{F}(y|\mathbf{x})}{\log(y)} \right| < \infty.$$

In condition \mathcal{D}_ω , the assumption that $\nabla_{\mathbf{x}} \log \bar{F}(y|\mathbf{x}) / \log(y)$ converges as $y \rightarrow \infty$ is motivated by the fact that, in the setup of conditional heavy tails,

$$\frac{\log \bar{F}(y|\mathbf{x})}{\log(y)} = -\frac{1}{\gamma(\mathbf{x})} + \frac{\log L(y|\mathbf{x})}{\log(y)}$$

where $L(\cdot|\mathbf{x})$ is a slowly varying function. In particular, $\log L(y|\mathbf{x}) / \log(y) \rightarrow 0$ as $y \rightarrow \infty$ (see de Haan and Ferreira, 2006, Proposition B.1.9.1 p.36). The assumption asks for this convergence to hold also when taking the gradient with respect to \mathbf{x} , *i.e.* the function $L(\cdot|\mathbf{x})$ does not vary too wildly in \mathbf{x} when y is large. The finite limit of $\nabla_{\mathbf{x}} \log \bar{F}(y|\mathbf{x}) / \log(y)$ as $y \rightarrow \infty$ will then be $\boldsymbol{\mu}(\mathbf{x}) = \nabla \gamma(\mathbf{x}) / \gamma^2(\mathbf{x}) \in \mathbb{R}^p$. The intuition is that while condition \mathcal{L}_ω is a Lipschitz-type requirement on the log-conditional survival function at extreme levels, condition \mathcal{D}_ω is essentially an appropriate analog about its gradient with respect to \mathbf{x} .

2.2 Intermediate quantile estimation

Let $\tau = \tau_n \uparrow 1$ as $n \rightarrow \infty$. We start by showing that the estimator $\hat{q}_n(\tau_n|\mathbf{x})$ is asymptotically normal if $\tau_n \uparrow 1$ slowly enough, *i.e.* $nh_n^p(1 - \tau_n) \rightarrow \infty$ so that the target quantity

$q(\tau_n|\mathbf{x})$ lies inside the sample. Hereafter, under condition \mathcal{KS} and \mathcal{D}_ω , we let

$$\Upsilon_K : (\Delta, \mathbf{x}) \in \mathbb{R} \times \mathbb{R}^p \mapsto \frac{\gamma(\mathbf{x})\Delta}{2} \int_{\mathbb{R}^p} K(\mathbf{u})(\mathbf{u}^\top \boldsymbol{\mu}(\mathbf{x}))^2 d\mathbf{u}.$$

Theorem 1. *Assume that conditions \mathcal{M} , $\mathcal{A}(l_n, r_n)$, \mathcal{K} , \mathcal{L}_g , \mathcal{L}_ω , \mathcal{B}_p and \mathcal{B}_Ω hold with $\sum_{j=1}^\infty j^\eta \alpha(j) < \infty$ for some $\eta > 1$. Let $\tau_n \uparrow 1$, fix $J \geq 1$, pick distinct $c_j \in (0, 1]$ and let $\tau_{n,j}$ be such that $1 - \tau_{n,j} = c_j(1 - \tau_n)(1 + o(1))$ as $n \rightarrow \infty$ (for $1 \leq j \leq J$). Assume further that $h_n \rightarrow 0$ is such that $nh_n^p(1 - \tau_n) \rightarrow \infty$, $nh_n^{p+2}(1 - \tau_n) \log^2(1 - \tau_n) \rightarrow 0$ and $\sqrt{nh_n^p(1 - \tau_n)}A((1 - \tau_n)^{-1}|\mathbf{x}) = O(1)$, and that there is $\delta > 0$ such that $r_n(r_n/\sqrt{nh_n^p(1 - \tau_n)})^\delta \rightarrow 0$. Then*

$$\sqrt{nh_n^p(1 - \tau_n)} \left(\frac{\hat{q}_n(\tau_{n,j}|\mathbf{x})}{q(\tau_{n,j}|\mathbf{x})} - 1 \right)_{1 \leq j \leq J} \xrightarrow{d} \mathcal{N} \left(\mathbf{0}_J, \frac{\int_{\mathbb{R}^p} K^2}{g(\mathbf{x})} \gamma^2(\mathbf{x}) \mathbf{M} \right),$$

where $\mathbf{M} = [1/\max(c_j, c_l)]_{1 \leq j, l \leq J}$. If conditions \mathcal{KS} , \mathcal{D}_g and \mathcal{D}_ω hold, then condition $nh_n^{p+2}(1 - \tau_n) \log^2(1 - \tau_n) \rightarrow 0$ can be replaced by the weaker bias assumption $\sqrt{nh_n^p(1 - \tau_n)} \times h_n^2 \log^2(1 - \tau_n) \rightarrow \Delta(\mathbf{x}) \in [0, \infty)$, in which case, provided $r_n h_n^p \rightarrow 0$, the asymptotic mean $\mathbf{0}_J$ of the above Gaussian limit is replaced by $\Upsilon_K(\Delta(\mathbf{x}), \mathbf{x}) \mathbf{1}_J$.

If $((\mathbf{X}_t, Y_t))_{t \geq 1}$ is moreover ρ -mixing, condition $\sum_{j=1}^\infty j^\eta \alpha(j) < \infty$ can be replaced by $\sum_{j=1}^\infty \rho(j) < \infty$. If $((\mathbf{X}_t, Y_t))_{t \geq 1}$ is in fact also ψ -mixing with $\sum_{j=1}^\infty \psi(j) < \infty$ (instead of $\sum_{j=1}^\infty j^\eta \alpha(j) < \infty$ for some $\eta > 1$, or $\sum_{j=1}^\infty \rho(j) < \infty$), all conditions on (l_n) and (r_n) (including $\mathcal{A}(l_n, r_n)$) as well as \mathcal{B}_p and \mathcal{B}_Ω can also be dropped.

It is remarkable that the asymptotic distribution in Theorem 1 is exactly the one obtained in the i.i.d. setting by Daouia et al. (2011) under an unnecessary regularity assumption on conditional tails. This is not true in the unconditional setting, see Drees (2003). The essential difference is that the kernel estimator only takes into account those pairs (\mathbf{X}_t, Y_t) such that \mathbf{X}_t are close enough to \mathbf{x} , and the mixing and stationarity assumptions ensure that such data points are far enough apart in time and hence asymptotically independent. This phenomenon has been observed in other contexts, such as nonparametric conditional Expected Shortfall estimation (Linton and Xiao, 2013, p.784).

Theorem 1 actually holds under weaker bias assumptions than those of Daouia et al. (2011), although this naturally comes at the cost of the reinforced regularity conditions \mathcal{D}_g and \mathcal{D}_ω . Indeed, when $((\mathbf{X}_t, Y_t))_{t \geq 1}$ is geometrically α -mixing, and if $A(t|\mathbf{x}) \propto t^{\rho(\mathbf{x})}$, as is the case in a wide range of heavy-tailed models used in practice (see e.g. Beirlant et al., 2004, Table 2.1 p.59), the optimal rate of convergence is $n^{\rho(\mathbf{x})/(1-(p+2)\rho(\mathbf{x}))}$ when \mathcal{K} , \mathcal{L}_g and \mathcal{L}_ω hold, while it is equal to $n^{2\rho(\mathbf{x})/(2-(p+4)\rho(\mathbf{x}))}$ if \mathcal{KS} , \mathcal{D}_g and \mathcal{D}_ω are satisfied (see Section A.3 for details). In the latter setting, note that $p = 0$ yields the optimal convergence rate $n^{\rho(\mathbf{x})/(1-2\rho(\mathbf{x}))}$ of unconditional extreme value estimators in heavy-tailed models (see de Haan and Ferreira, 2006, p.77), while the case $\rho(\mathbf{x}) \rightarrow -\infty$, corresponding to the ideal but unrealistic scenario where all the Y_t such that $\mathbf{X}_t \in B(\mathbf{x}, h_n)$ can be used, yields the optimal convergence rate $n^{-2/(p+4)}$, i.e. the optimal convergence rate of nonparametric estimators of a twice continuously differentiable *central* conditional quantile, see Chaudhuri (1991).

The bias component $\Upsilon_K(\Delta(\mathbf{x}), \mathbf{x})$ appearing under conditions \mathcal{D}_g and \mathcal{D}_ω is an analog of the bias component in kernel regression for the mean. It is linked to the gradient of the target extreme conditional quantile through the vector $\boldsymbol{\mu}(\mathbf{x}) = \nabla \gamma(\mathbf{x})/\gamma^2(\mathbf{x})$. This means that a key factor in extremal regression is the *variation in conditional extremes* and especially of tail heaviness across space, as opposed to standard regression where it

is also important to account for *changes of shape* in the underlying regression function through its second derivatives (see *e.g.* Wasserman, 2006, p.73). We finally note that the vector $\gamma(\mathbf{x})\boldsymbol{\mu}(\mathbf{x}) = \nabla\gamma(\mathbf{x})/\gamma(\mathbf{x}) = \nabla(\log\gamma)(\mathbf{x})$ is reminiscent of the *design bias* for classical regression in the terminology of Wasserman (2006): the higher and less variable γ around \mathbf{x} , the bigger and more stable the local number of extreme observations, and the easier the conditional extreme value estimation problem.

2.3 Extreme quantile estimation

Consider now a level τ'_n such that $nh_n^p(1 - \tau'_n) \rightarrow c < \infty$. Then, in a strip around \mathbf{x} , very few or no top observations Y_t will be close to the extreme value $q(\tau'_n|\mathbf{x})$, and the empirical estimator $\widehat{q}_n(\tau'_n|\mathbf{x})$ will therefore not be consistent. A solution is to employ the extrapolation methodology of Weissman (1978) from unconditional extreme quantile estimation: the conditional heavy tail assumption suggests the approximation $q(\tau'_n|\mathbf{x}) \approx ((1 - \tau'_n)/(1 - \tau_n))^{-\widehat{\gamma}(\mathbf{x})}q(\tau_n|\mathbf{x})$ for n large enough. Plugging in consistent estimators $\widehat{\gamma}(\mathbf{x})$ of $\gamma(\mathbf{x})$ and $\bar{q}_n(\tau_n|\mathbf{x})$ of $q(\tau_n|\mathbf{x})$ (such as $\widehat{q}_n(\tau_n|\mathbf{x})$) yields the following conditional Weissman estimator of $q(\tau'_n|\mathbf{x})$:

$$\widehat{q}_{n,\tau_n}^W(\tau'_n|\mathbf{x}) = \left(\frac{1 - \tau'_n}{1 - \tau_n}\right)^{-\widehat{\gamma}(\mathbf{x})} \bar{q}_n(\tau_n|\mathbf{x}).$$

The next result shows that the choice of $\widehat{\gamma}(\mathbf{x})$ is crucial since the extrapolated estimator $\widehat{q}_{n,\tau_n}^W(\tau'_n|\mathbf{x})$ inherits its asymptotic distribution.

Theorem 2. *Assume that condition \mathcal{M} holds with $\rho(\mathbf{x}) < 0$. Let $\tau_n, \tau'_n \uparrow 1$ be such that $(1 - \tau'_n)/(1 - \tau_n) \rightarrow 0$ and assume that $v_n(\bar{q}_n(\tau_n|\mathbf{x})/q(\tau_n|\mathbf{x}) - 1) = O_{\mathbb{P}}(1)$ and $v_n(\widehat{\gamma}(\mathbf{x}) - \gamma(\mathbf{x})) \xrightarrow{d} \Gamma$, where Γ is a nondegenerate distribution and $v_n \rightarrow \infty$. If $v_n A((1 - \tau_n)^{-1}|\mathbf{x}) = O(1)$ and $v_n/\log[(1 - \tau_n)/(1 - \tau'_n)] \rightarrow \infty$, then*

$$\frac{v_n}{\log[(1 - \tau_n)/(1 - \tau'_n)]} \left(\frac{\widehat{q}_{n,\tau_n}^W(\tau'_n|\mathbf{x})}{q(\tau'_n|\mathbf{x})} - 1 \right) \xrightarrow{d} \Gamma.$$

In our context, v_n is typically $\sqrt{nh_n^p(1 - \tau_n)}$ and $\bar{q}_n(\tau_n|\mathbf{x}) = \widehat{q}_n(\tau_n|\mathbf{x})$. As a conditional tail index estimator, we propose to use the refined Pickands-type kernel smoother from Daouia et al. (2011): for an integer $J \geq 2$,

$$\widehat{\gamma}_{\tau_n}^{(J)}(\mathbf{x}) = \frac{1}{\log(J!)} \sum_{j=1}^J \log \left(\frac{\widehat{q}_n(1 - (1 - \tau_n)/j|\mathbf{x})}{\widehat{q}_n(\tau_n|\mathbf{x})} \right).$$

This estimator has the theoretical advantage to only use a finite, but otherwise arbitrary, number of intermediate quantile estimators, and hence its asymptotic distribution can be deduced from Theorem 1. This would not be the case for a conditional version of the Hill estimator (see *e.g.* de Haan and Ferreira, 2006, Section 3.2) that would basically involve the very difficult study of conditional tail empirical quantile processes.

Theorem 3. *Work under the conditions of Theorem 1, and assume in addition that $\sqrt{nh_n^p(1 - \tau_n)}A((1 - \tau_n)^{-1}|\mathbf{x}) \rightarrow \lambda(\mathbf{x}) \in \mathbb{R}$. Then,*

$$\begin{aligned} & \sqrt{nh_n^p(1 - \tau_n)} \left(\widehat{\gamma}_{\tau_n}^{(J)}(\mathbf{x}) - \gamma(\mathbf{x}), \frac{\widehat{q}_n(\tau_n|\mathbf{x})}{q(\tau_n|\mathbf{x})} - 1 \right) \\ & \xrightarrow{d} \mathcal{N} \left(\left(\frac{\lambda(\mathbf{x})}{\log(J!)} \sum_{j=2}^J \frac{j^{\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})}, 0 \right), \frac{\int_{\mathbb{R}^p} K^2}{g(\mathbf{x})} \gamma^2(\mathbf{x}) \begin{pmatrix} \frac{J(J-1)(2J-1)}{6 \log^2(J!)} & 0 \\ 0 & 1 \end{pmatrix} \right). \end{aligned}$$

If conditions \mathcal{KS} , \mathcal{D}_g and \mathcal{D}_ω hold, under the weaker bias assumption $\sqrt{nh_n^p(1-\tau_n)} \times h_n^2 \log^2(1-\tau_n) \rightarrow \Delta(\mathbf{x}) \in [0, \infty)$ and if $r_n h_n^p \rightarrow 0$, the second component of the asymptotic mean of the above Gaussian limit is replaced by $\Upsilon_K(\Delta(\mathbf{x}), \mathbf{x})$.

The asymptotic variance of $\hat{\gamma}_{\tau_n}^{(J)}(\mathbf{x})$ reaches its minimum when $J = 9$, and is approximately $1.25 \times \gamma^2(\mathbf{x})$. We shall adopt this choice in our finite-sample experiments in Sections 6 and 7. Theorem 3 improves upon Corollary 2 in Daouia et al. (2011), by removing unnecessary assumptions about the structure of the conditional right tail, and most importantly by establishing the asymptotic normality of $\hat{\gamma}_{\tau_n}^{(J)}(\mathbf{x})$ under a substantially weaker bias assumption in conjunction with reasonable regularity conditions, see the discussion below Theorem 1. Surprisingly perhaps, the bias component $\Delta(\mathbf{x})$ does not appear in the limiting distribution of $\hat{\gamma}_{\tau_n}^{(J)}(\mathbf{x})$, which suggests that the local variation of conditional extremes is not as important in the estimation of the conditional shape parameter $\gamma(\mathbf{x})$ of the underlying Pareto approximating distribution as it is for its scale $q(\tau_n|\mathbf{x})$.

3 Nonparametric extremal expectile regression

3.1 Framework and estimation method

We enrich our framework to handle extreme conditional expectile estimation. Rewrite the transformed distribution of (1.2) as $E(\cdot|\mathbf{x}) = 1 - \varphi^{(1)}(y|\mathbf{x}) / (2\varphi^{(1)}(y|\mathbf{x}) + y - m(\mathbf{x}))$ where $\varphi^{(a)}(y|\mathbf{x}) = \mathbb{E}((Y - y)^a \mathbb{1}_{\{Y > y\}} | \mathbf{X} = \mathbf{x})$ stands for the conditional tail moment of order $a > 0$, and $m(\mathbf{x}) = \mathbb{E}(Y | \mathbf{X} = \mathbf{x})$ for the regression mean. Consider the following kernel smoother for $E(y|\mathbf{x})$,

$$\begin{aligned} \hat{E}_n(y|\mathbf{x}) &= 1 - \frac{\hat{\varphi}_n^{(1)}(y|\mathbf{x})}{2\hat{\varphi}_n^{(1)}(y|\mathbf{x}) + (y - \hat{m}_n(\mathbf{x}))}, \\ \text{with } \hat{\varphi}_n^{(1)}(y|\mathbf{x}) &= \frac{1}{\hat{g}_n(\mathbf{x})} \times \frac{1}{nh_n^p} \sum_{t=1}^n (Y_t - y) \mathbb{1}_{\{Y_t > y\}} K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_n}\right) \\ \text{and } \hat{m}_n(\mathbf{x}) &= \frac{1}{\hat{g}_n(\mathbf{x})} \times \frac{1}{nh_n^p} \sum_{t=1}^n Y_t K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_n}\right). \end{aligned}$$

The estimator \hat{m}_n is the Nadaraya-Watson estimator of the regression function m . The characterization of the conditional expectiles as $e(\tau|\mathbf{x}) \equiv \inf\{y \in \mathbb{R} | E(y|\mathbf{x}) \geq \tau\}$ (see Jones, 1994) implies that they can be estimated by their empirical counterparts

$$\hat{e}_n(\tau|\mathbf{x}) = \inf\left\{y \in \mathbb{R} | \hat{E}_n(y|\mathbf{x}) \geq \tau\right\}.$$

A straightforward calculation shows that this estimator is identical to the asymmetric least squares estimator obtained directly by smoothing up the loss function defining $e(\tau|\mathbf{x})$ in (1.1), that is, $\hat{e}_n(\tau|\mathbf{x}) = \arg \min_{\theta \in \mathbb{R}} \int_{\mathbb{R}} (\eta_\tau(y - \theta) - \eta_\tau(y)) d\hat{F}_n(y|\mathbf{x})$. Our goal is to estimate extreme conditional expectiles at a high asymmetry level $\tau = \tau_n$ converging to 1 at any possible rate, as $n \rightarrow \infty$. In addition to the conditions of Section 2.1, we make the following assumptions on conditional tail heaviness and regularity of conditional moments.

Condition \mathcal{H}_δ One has $\gamma(\mathbf{x}) < 1/(2 + \delta)$ and there exists $r > 0$ such that $\mathbb{E}(Y_-^{2+\delta} | \mathbf{X} = \cdot)$ is bounded on $B(\mathbf{x}, r)$.

Condition \mathcal{L}_m The response Y has a finite second moment given $\mathbf{X} = \mathbf{x}$, and the conditional mean functions $\mathbb{E}(Y | \mathbf{X} = \cdot)$ and $\mathbb{E}(Y^2 | \mathbf{X} = \cdot)$ are Lipschitz continuous at \mathbf{x} .

Condition \mathcal{B}_m There exists $r > 0$ such that

$$\sup_{t \geq 1} \sup_{\mathbf{x}_1, \mathbf{x}_{t+1} \in B(\mathbf{x}, r)} \mathbb{E}(Y_1^2 + Y_{t+1}^2 | \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_{t+1} = \mathbf{x}_{t+1}) < \infty.$$

Condition \mathcal{H}_δ (in which $\delta > 0$) guarantees a finite conditional moment of order $(2 + \delta)$ in a neighborhood of \mathbf{x} ; in the unconditional framework, Theorem 2 in Daouia et al. (2018) requires the analog $\mathbb{E}(Y_-^{2+\delta}) < \infty$. Finally, as in Section 2, the following stronger version of condition \mathcal{L}_m will be imposed to obtain better rates of convergence:

Condition \mathcal{D}_m The response Y has a finite second moment given $\mathbf{X} = \mathbf{x}$, and the conditional mean functions $\mathbb{E}(Y | \mathbf{X} = \cdot)$ and $\mathbb{E}(Y^2 | \mathbf{X} = \cdot)$ are continuously differentiable in a neighborhood of \mathbf{x} and have Lipschitz continuous gradients at \mathbf{x} .

3.2 Intermediate expectile estimation

We first derive the asymptotic distribution of an intermediate expectile estimator jointly with an intermediate quantile estimator. This will be key to the construction of an expectile-based estimator for the conditional tail index, and therefore to our extrapolation of conditional expectiles to extreme levels.

Theorem 4. *Assume that conditions \mathcal{M} , $\mathcal{A}(l_n, r_n)$, \mathcal{K} , \mathcal{H}_δ , \mathcal{L}_g , \mathcal{L}_m , \mathcal{L}_ω , \mathcal{B}_p , \mathcal{B}_m and \mathcal{B}_Ω hold with $\sum_{j=1}^{\infty} j^\eta [\alpha(j)]^{\delta/(2+\delta)} < \infty$ for some $\eta > \delta/(2 + \delta)$. Let $\tau_n \uparrow 1$ and $\kappa > 0$ be given, and let β_n be such that $1 - \beta_n = \kappa(1 - \tau_n)(1 + o(1))$ as $n \rightarrow \infty$. Assume further that $h_n \rightarrow 0$ is such that $nh_n^p(1 - \tau_n) \rightarrow \infty$, $nh_n^{p+2}(1 - \tau_n) \log^2(1 - \tau_n) \rightarrow 0$, $\sqrt{nh_n^p(1 - \tau_n)}A((1 - \tau_n)^{-1}|\mathbf{x}) = O(1)$ and $r_n(r_n/\sqrt{nh_n^p(1 - \tau_n)})^\delta \rightarrow 0$. Then*

$$\sqrt{nh_n^p(1 - \tau_n)} \left(\frac{\hat{e}_n(\tau_n|\mathbf{x})}{e(\tau_n|\mathbf{x})} - 1, \frac{\hat{q}_n(\beta_n|\mathbf{x})}{q(\beta_n|\mathbf{x})} - 1 \right) \xrightarrow{d} \mathcal{N} \left((0, 0), \frac{\int_{\mathbb{R}^p} K^2}{g(\mathbf{x})} \gamma^2(\mathbf{x}) \boldsymbol{\Sigma}(\mathbf{x}) \right),$$

where the 2×2 symmetric matrix $\boldsymbol{\Sigma}(\mathbf{x})$ has entries $\Sigma_{1,1}(\mathbf{x}) = 2\gamma(\mathbf{x})/(1 - 2\gamma(\mathbf{x}))$, $\Sigma_{2,2}(\mathbf{x}) = \kappa^{-1}$ and

$$\Sigma_{1,2}(\mathbf{x}) = \begin{cases} \kappa^{-1} & \text{if } \kappa \geq 1/\gamma(\mathbf{x}) - 1, \\ \left(\frac{1}{\gamma(\mathbf{x})} - 1 \right)^{\gamma(\mathbf{x})} \frac{\kappa^{-\gamma(\mathbf{x})}}{1 - \gamma(\mathbf{x})} - 1 & \text{if } \kappa < 1/\gamma(\mathbf{x}) - 1. \end{cases}$$

If conditions \mathcal{KS} , \mathcal{D}_g , \mathcal{D}_m and \mathcal{D}_ω hold, then condition $nh_n^{p+2}(1 - \tau_n) \log^2(1 - \tau_n) \rightarrow 0$ can be replaced by the weaker bias assumption $\sqrt{nh_n^p(1 - \tau_n)} \times h_n^2 \log^2(1 - \tau_n) \rightarrow \Delta(\mathbf{x}) \in [0, \infty)$, and if moreover $r_n h_n^p \rightarrow 0$, then the asymptotic mean $(0, 0)$ of the above Gaussian limit is replaced by $\Upsilon_K(\Delta(\mathbf{x}), \mathbf{x}) \times (1, 1)$.

If $((\mathbf{X}_t, Y_t))_{t \geq 1}$ is ρ -mixing with $\sum_{j=1}^{\infty} \rho(j) < \infty$ (instead of $\sum_{j=1}^{\infty} j^\eta [\alpha(j)]^{\delta/(2+\delta)} < \infty$ for some $\eta > \delta/(2 + \delta)$), condition \mathcal{B}_m can be dropped and condition \mathcal{H}_δ can be replaced by $\gamma(\mathbf{x}) < 1/(2 + \delta)$. If $((\mathbf{X}_t, Y_t))_{t \geq 1}$ is in fact also ψ -mixing with $\sum_{j=1}^{\infty} \psi(j) < \infty$ (instead of $\sum_{j=1}^{\infty} j^\eta [\alpha(j)]^{\delta/(2+\delta)} < \infty$ for some $\eta > \delta/(2 + \delta)$, or $\sum_{j=1}^{\infty} \rho(j) < \infty$), all conditions on (l_n) and (r_n) (including $\mathcal{A}(l_n, r_n)$) as well as conditions \mathcal{B}_p and \mathcal{B}_Ω can also be dropped, and condition \mathcal{H}_δ can be replaced by $0 < \gamma(\mathbf{x}) < 1/2$.

We obtain again the same asymptotic distribution as in the i.i.d. setting under weaker moment and regularity assumptions and with a faster optimal pointwise convergence rate, see Theorem 1 in Girard et al. (2022) and the discussion below our Theorem 1. By contrast, temporal mixing changes the asymptotic distribution of unconditional empirical intermediate expectiles, as shown in Davison et al. (2021).

3.3 Extreme expectile estimation

Extreme quantiles and expectiles being asymptotically proportional, in the sense that $e(\tau|\mathbf{x})/q(\tau|\mathbf{x}) \rightarrow (1/\gamma(\mathbf{x})-1)^{-\gamma(\mathbf{x})}$ as $\tau \uparrow 1$ (see Proposition 2.3 in Bellini and Di Bernardino, 2017), Weissman's approximation applies to tail expectiles as well, that is, $e(\tau'_n|\mathbf{x}) \approx ((1-\tau'_n)/(1-\tau_n))^{-\gamma(\mathbf{x})}e(\tau_n|\mathbf{x})$, for n large. Plugging in consistent estimators $\hat{\gamma}(\mathbf{x})$ of $\gamma(\mathbf{x})$ and $\bar{e}_n(\tau_n|\mathbf{x})$ of $e(\tau_n|\mathbf{x})$ (such as $\hat{e}_n(\tau_n|\mathbf{x})$) yields a Weissman-type estimator of $e(\tau'_n|\mathbf{x})$:

$$\hat{e}_{n,\tau_n}^W(\tau'_n|\mathbf{x}) = \left(\frac{1-\tau'_n}{1-\tau_n} \right)^{-\hat{\gamma}(\mathbf{x})} \bar{e}_n(\tau_n|\mathbf{x}).$$

The next result states that $\hat{e}_{n,\tau_n}^W(\tau'_n|\mathbf{x})$ also inherits the asymptotic distribution of $\hat{\gamma}(\mathbf{x})$.

Theorem 5. *Assume that condition \mathcal{M} holds, with $\gamma(\mathbf{x}) < 1$, $\rho(\mathbf{x}) < 0$ and $\mathbb{E}(Y_- | \mathbf{X} = \mathbf{x}) < \infty$. Let $\tau_n, \tau'_n \uparrow 1$ be such that $(1-\tau'_n)/(1-\tau_n) \rightarrow 0$ and assume that $v_n(\bar{e}_n(\tau_n|\mathbf{x})/e(\tau_n|\mathbf{x})-1) = O_{\mathbb{P}}(1)$ and $v_n(\hat{\gamma}(\mathbf{x}) - \gamma(\mathbf{x})) \xrightarrow{d} \Gamma$, where Γ is a nondegenerate distribution and $v_n \rightarrow \infty$. If $v_n A((1-\tau_n)^{-1}|\mathbf{x}) = O(1)$, $v_n/q(\tau_n|\mathbf{x}) = O(1)$ and $v_n/\log[(1-\tau_n)/(1-\tau'_n)] \rightarrow \infty$, then*

$$\frac{v_n}{\log[(1-\tau_n)/(1-\tau'_n)]} \left(\frac{\hat{e}_{n,\tau_n}^W(\tau'_n|\mathbf{x})}{e(\tau'_n|\mathbf{x})} - 1 \right) \xrightarrow{d} \Gamma.$$

Define now a purely expectile-based estimator of $\gamma(\mathbf{x})$ as

$$\hat{\gamma}_{\tau_n}^E(\mathbf{x}) = \left(1 + \frac{\hat{F}_n(\hat{e}_n(\tau_n|\mathbf{x})|\mathbf{x})}{1-\tau_n} \right)^{-1},$$

where $\hat{F}_n(\cdot|\mathbf{x}) = 1 - \hat{F}_n(\cdot|\mathbf{x})$ and τ_n is an intermediate sequence of expectile levels. We derive the asymptotic normality of $\hat{\gamma}_{\tau_n}^E(\mathbf{x})$ jointly with $\hat{e}_n(\tau_n|\mathbf{x})$ next.

Theorem 6. *Work under the conditions of Theorem 4. Assume further that $\sqrt{nh_n^p(1-\tau_n)}A((1-\tau_n)^{-1}|\mathbf{x}) \rightarrow \lambda_1(\mathbf{x}) \in \mathbb{R}$ and $\sqrt{nh_n^p(1-\tau_n)}/q(\tau_n|\mathbf{x}) \rightarrow \lambda_2(\mathbf{x}) \in \mathbb{R}$. Then*

$$\begin{aligned} \sqrt{nh_n^p(1-\tau_n)} \left(\hat{\gamma}_{\tau_n}^E(\mathbf{x}) - \gamma(\mathbf{x}), \frac{\hat{e}_n(\tau_n|\mathbf{x})}{e(\tau_n|\mathbf{x})} - 1 \right) \\ \xrightarrow{d} \mathcal{N} \left((b_E(\mathbf{x}), 0), \frac{\int_{\mathbb{R}^p} K^2}{g(\mathbf{x})} v_E(\mathbf{x}) \begin{pmatrix} 1 - \gamma(\mathbf{x}) & 1 \\ 1 & 2 \end{pmatrix} \right), \end{aligned}$$

where $v_E(\mathbf{x}) = \gamma^3(\mathbf{x})/(1-2\gamma(\mathbf{x}))$ and

$$b_E(\mathbf{x}) = \frac{\gamma(\mathbf{x})(1/\gamma(\mathbf{x})-1)^{1-\rho(\mathbf{x})}}{1-\gamma(\mathbf{x})-\rho(\mathbf{x})} \lambda_1(\mathbf{x}) + \gamma^2(\mathbf{x})(1/\gamma(\mathbf{x})-1)^{\gamma(\mathbf{x})+1} \mathbb{E}(Y|\mathbf{X}=\mathbf{x}) \lambda_2(\mathbf{x}).$$

If conditions \mathcal{KS} , \mathcal{D}_g , \mathcal{D}_m and \mathcal{D}_ω hold, under the weaker bias assumption $\sqrt{nh_n^p(1-\tau_n)} \times h_n^2 \log^2(1-\tau_n) \rightarrow \Delta(\mathbf{x}) \in [0, \infty)$ and if $r_n h_n^p \rightarrow 0$, the second component of the asymptotic mean of the above Gaussian limit is replaced by $\Upsilon_K(\Delta(\mathbf{x}), \mathbf{x})$.

Again, mixing does not impact the bias or the variance of $\hat{\gamma}_{\tau_n}^E(\mathbf{x})$ obtained in the i.i.d. setting (see Theorem 4 in Girard et al., 2022). Strikingly, even though this estimator and $\hat{\gamma}_{\tau_n}^{(J)}(\mathbf{x})$ have very different structures and properties, the bias component $\Delta(\mathbf{x})$ does not appear in the limiting distribution of $\hat{\gamma}_{\tau_n}^E(\mathbf{x})$ either. The advantage of using the expectile-based estimator $\hat{\gamma}_{\tau_n}^E(\mathbf{x})$ is that its asymptotic variance is lower than the asymptotic variance $\gamma^2(\mathbf{x})$ of the conditional Hill estimator when $0 < \gamma(\mathbf{x}) < 0.4$, corresponding to many of the practical applications where a second (tail) moment exists but a third moment may not. The bias of $\hat{\gamma}_{\tau_n}^E(\mathbf{x})$ is, however, much stronger; we consequently discuss a bias-correction methodology in Section 5.2.

4 Regression models covered by our framework

We now draw a non-exhaustive list of examples satisfying our assumptions. The conditions involving the mixing coefficients or the sequences l_n and r_n hold automatically when the stochastic process $((\mathbf{X}_t, Y_t))_{t \geq 1}$ is geometrically α -mixing, namely, there exists $a \in (0, 1)$ such that $\alpha(n) = O(a^n)$. Besides, in the typical extreme value models where $A(t|\mathbf{x}) \propto t^{\rho(\mathbf{x})}$, our assumptions linking τ_n and h_n will be satisfied if $h_n = C_1 n^{-h}$ and $\tau_n = 1 - C_2 n^{-\tau}$, for any $C_1, C_2 > 0$ and suitably chosen $h, \tau > 0$, see the discussion below Theorem 1. We therefore focus in this section on assumptions \mathcal{M} , \mathcal{H}_δ , \mathcal{L}_g , \mathcal{L}_m , \mathcal{L}_ω , \mathcal{B}_p , \mathcal{B}_m , \mathcal{B}_Ω , \mathcal{D}_g , \mathcal{D}_m and \mathcal{D}_ω . We provide simple sufficient conditions and give an extended discussion with more insight, and sometimes weaker conditions in Section B. We also give a full treatment therein of the instructive case of m -dependent (including i.i.d.) observations.

Location-scale model with possible temporal misspecification Suppose that $Y_t = m(\mathbf{X}_t) + \sigma(\mathbf{X}_t)\varepsilon_t$ where m and $\sigma > 0$ are location and scale components, and (ε_t) is a stationary and centered sequence of unobserved heavy-tailed innovations independent from the sequence (\mathbf{X}_t) . The ε_t can be dependent, which yields a form of misspecified regression model.

Proposition 1 (Location-scale model). *Assume that $(\mathbf{X}_t)_{t \geq 1}$ is β -mixing (i.e. absolutely regular) and $(\varepsilon_t)_{t \geq 1}$ is strongly mixing. Suppose further that:*

- For any $t \geq 1$, the random pairs $(\mathbf{X}_1, \mathbf{X}_{t+1})$ have absolutely continuous distributions whose first marginal has a p.d.f. g such that $g(\mathbf{x}) > 0$, and the functions g , m and σ are continuously differentiable in a neighborhood of \mathbf{x} .
- The error ε has a p.d.f. f_ε which satisfies the asymptotic expansion

$$f_\varepsilon(z) = c_0 z^{-1/\gamma-1} (1 + d_0 z^{-a} + d'_0 z^{-a-b} (1 + o(1))) \text{ as } z \rightarrow \infty$$

where $\gamma > 0$, $c_0 > 0$, $d_0, d'_0 \neq 0$ and $a, b > 0$ are such that either $a \neq 1$, $a = 1 \neq b$, or $a = b = 1$ with then $2d'_0(1 + \gamma) \neq d_0^2(1 + 2\gamma)$.

Then conditions \mathcal{M} , \mathcal{L}_g , \mathcal{L}_ω , \mathcal{B}_p and \mathcal{B}_Ω hold. If moreover $\gamma < 1/(2 + \delta)$ and $\mathbb{E}(\varepsilon_-^{2+\delta}) < \infty$, then conditions \mathcal{H}_δ , \mathcal{L}_m and \mathcal{B}_m hold as well. If in addition g , m and σ are twice continuously differentiable in a neighborhood of \mathbf{x} , then conditions \mathcal{D}_g and \mathcal{D}_m also hold. If moreover f_ε is continuously differentiable in a neighborhood of infinity and satisfies the second-order von Mises condition $-z f'_\varepsilon(z)/f_\varepsilon(z) \rightarrow 1/\gamma + 1$, then condition \mathcal{D}_ω holds, with $\lim_{y \rightarrow \infty} \nabla_{\mathbf{x}} \log \bar{F}(y|\mathbf{x})/\log(y) = 0$.

The class of β -mixing processes covers many important cases such as Harris recurrent aperiodic Markov chains (Bradley, 2005, Corollary 3.6). Our conditions on r and σ cover all standard location-scale regression models, such as heteroscedastic linear regression with smooth conditional variance function, single-index models with smooth link functions, and additive models. The assumptions on f_ε hold in the vast majority of standard heavy-tailed models, such as the Fréchet, Burr, Student and Fisher distributions. We note that this setting cannot cover time series models such as autoregressive models, which require a specific treatment, because a key assumption is that the series (\mathbf{X}_t) must be independent from the series of errors (ε_t) .

Nonlinear regression model Let $F(\cdot, \boldsymbol{\theta})$ be a parametric family of heavy-tailed distribution functions on \mathbb{R} , where $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ is a finite-dimensional, convex and open set of parameters, and let $q(\cdot, \boldsymbol{\theta})$ be the associated quantile function (the left-continuous inverse of $y \mapsto F(y, \boldsymbol{\theta})$). Let (with a slight abuse of notation) $\boldsymbol{\theta} : \mathbf{x} \in \mathbb{R}^p \mapsto \boldsymbol{\theta}(\mathbf{x}) \in \Theta$ be a smooth mapping and consider the model $Y_t = q(U_t, \boldsymbol{\theta}(\mathbf{X}_t))$ where (U_t) is a stationary sequence of unobserved, uniformly distributed innovations independent from the series (\mathbf{X}_t) .

Proposition 2 (Nonlinear regression model). *Assume that $(\mathbf{X}_t)_{t \geq 1}$ is β -mixing and $(U_t)_{t \geq 1}$ is strongly mixing. Suppose further that:*

- For any $\boldsymbol{\theta}$, the survival function $\bar{F}(\cdot, \boldsymbol{\theta})$ is second-order regularly varying, with tail index $\gamma = \gamma(\boldsymbol{\theta})$, second-order parameter $\rho = \rho(\boldsymbol{\theta})$ and auxiliary function $A = A(\cdot | \boldsymbol{\theta})$. The function $\boldsymbol{\theta} \mapsto \bar{F}(y, \boldsymbol{\theta})$ is continuously differentiable for y large enough.
- For any $t \geq 1$, the random pairs $(\mathbf{X}_1, \mathbf{X}_{t+1})$ have absolutely continuous distributions whose first marginal has a p.d.f. g such that $g(\mathbf{x}) > 0$ and g is continuously differentiable in a neighborhood of \mathbf{x} .
- The parameter mapping $\boldsymbol{\theta} : \mathbb{R}^p \rightarrow \Theta$ is continuously differentiable in a neighborhood of \mathbf{x} . There are $y_0 > 0$ and a continuous function κ on Θ with

$$\forall \boldsymbol{\theta} \in \Theta, \sup_{y \geq y_0} \left\| \frac{\nabla_{\boldsymbol{\theta}} \log \bar{F}(y, \boldsymbol{\theta})}{\log(y)} \right\| = \sup_{y \geq y_0} \frac{1}{\log(y)} \frac{\|\nabla_{\boldsymbol{\theta}} \bar{F}(y, \boldsymbol{\theta})\|}{\bar{F}(y, \boldsymbol{\theta})} \leq \kappa(\boldsymbol{\theta}).$$

Then conditions \mathcal{M} , \mathcal{L}_g , \mathcal{L}_ω , \mathcal{B}_p and \mathcal{B}_Ω hold. Assume moreover that:

- The survival function $\bar{F}(\cdot, \boldsymbol{\theta})$ only puts mass on $[0, \infty)$.
- At the point \mathbf{x} , $\gamma(\boldsymbol{\theta}(\mathbf{x})) < 1/(2 + \delta)$.
- The first and second moments of the distribution $F(\cdot, \boldsymbol{\theta})$ define continuously differentiable functions m_1 and m_2 of $\boldsymbol{\theta}$ wherever they are defined.

Then conditions \mathcal{H}_δ , \mathcal{L}_m and \mathcal{B}_m hold. If the two moment functions m_1 and m_2 are twice continuously differentiable functions of $\boldsymbol{\theta}$ wherever they are defined and g and $\boldsymbol{\theta}(\cdot)$ are twice continuously differentiable in a neighborhood of \mathbf{x} , then conditions \mathcal{D}_g and \mathcal{D}_m hold. If in addition $\boldsymbol{\theta} \mapsto \gamma(\boldsymbol{\theta})$ is twice continuously differentiable on Θ and there exist $y_0 > 0$ and a continuous function κ on Θ with

$$\forall \boldsymbol{\theta} \in \Theta, \sup_{y \geq y_0} \left\{ \frac{1}{\log(y)} (\|\nabla_{\boldsymbol{\theta}} \log \bar{F}(y, \boldsymbol{\theta})\| + \|H_{\boldsymbol{\theta}} \log \bar{F}(y, \boldsymbol{\theta})\|) \right\} \leq \kappa(\boldsymbol{\theta})$$

then condition \mathcal{D}_ω holds, with $\lim_{y \rightarrow \infty} \nabla_{\mathbf{x}} \log \bar{F}(y | \mathbf{x}) / \log(y) = [\nabla \gamma(\boldsymbol{\theta}(\mathbf{x}))]^\top J_{\boldsymbol{\theta}}(\mathbf{x}) / \gamma^2(\boldsymbol{\theta}(\mathbf{x}))$.

Our conditions on the statistical model $F(\cdot, \boldsymbol{\theta})$ are mild and readily checked in typical heavy-tailed models such as the Fréchet, Burr, Generalized Pareto and half- t models (also called folded Student distribution, namely, the distribution of the absolute value of a Student random variable). If $\bar{F}(\cdot, \boldsymbol{\theta})$ puts mass on a neighborhood of $-\infty$, then extra assumptions on the left conditional tail (such as symmetry, or the left tail dominated by the right tail) are required to ensure that conditions \mathcal{H}_δ , \mathcal{L}_m , \mathcal{B}_m and \mathcal{D}_m hold.

Autoregressive model Consider the causal and invertible $\text{AR}(p)$ model $Y_t = \sum_{j=1}^p \phi_j Y_{t-j} + \varepsilon_t$, $t \in \mathbb{Z}$, where the polynomial $P(z) = 1 - \sum_{j=1}^p \phi_j z^j$ has no root inside the closed unit disk in \mathbb{C} , and (ε_t) is an i.i.d. sequence of innovations. Here Y_t should be understood as the stationary solution of the $\text{AR}(p)$ equations, and $\mathbf{X}_t = (Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})^\top$.

Proposition 3 (Autoregressive model). *Assume that ε is centered, square-integrable, and has a Lipschitz continuous, everywhere strictly positive p.d.f. f_ε with respect to the Lebesgue measure that satisfies the asymptotic expansion*

$$f_\varepsilon(z) = c_0 z^{-1/\gamma-1} (1 + d_0 z^{-a} + d'_0 z^{-a-b} (1 + o(1))) \text{ as } z \rightarrow \infty$$

where $\gamma > 0$, $c_0 > 0$, $d_0, d'_0 \neq 0$ and $a, b > 0$ are such that either $a \neq 1$, $a = 1 \neq b$, or $a = b = 1$ with then $2d'_0(1 + \gamma) \neq d_0^2(1 + 2\gamma)$.

Then conditions \mathcal{M} , \mathcal{L}_g , \mathcal{D}_m , \mathcal{L}_ω , \mathcal{B}_p and \mathcal{B}_Ω hold, and the process (Y_t) is geometrically β - and ρ -mixing. If ε also has a finite moment of order $(2 + \delta)$ and $\gamma < 1/(2 + \delta)$, then condition \mathcal{H}_δ holds. If moreover f_ε is continuously differentiable, with a uniformly bounded Lipschitz continuous derivative f'_ε , then condition \mathcal{D}_g holds. Finally, if f_ε also satisfies the second-order von Mises condition $-z f'_\varepsilon(z)/f_\varepsilon(z) \rightarrow 1/\gamma + 1$, then condition \mathcal{D}_ω holds, with $\lim_{y \rightarrow \infty} \nabla_{\mathbf{x}} \log \bar{F}(y|\mathbf{x})/\log(y) = 0$.

Condition \mathcal{B}_m is unnecessary because (Y_t) is ρ -mixing. Unlike in our other examples, checking condition \mathcal{B}_Ω is nontrivial, because the sequences $(\mathbf{X}_t) = ((Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})^\top)$ and (ε_t) are not independent. This is done by noting that (Y_t) is a Markov chain of order p and then by checking the conditions of Lemma 1(iii) (with $t_0 = p$).

5 Practical implementation

We discuss hyperparameter selection, bias and variance correction, and asymptotic confidence interval construction for both tail quantities of interest.

5.1 Selection of tuning parameters

In any extreme value technique, the intermediate level τ_n plays a crucial role. The value $k_n = n(1 - \tau_n)$, rounded to the next integer, can be viewed as the effective sample size for tail extrapolation. A larger k_n leads to larger bias, while smaller k_n results in more variance. In the regression case, in addition to τ_n (or k_n), one should determine the bandwidth h_n . Results in finite samples indicate that it is often enough to use the global bandwidth obtained by minimizing the mean integrated squared error $\mathbb{E}(\int_{\mathcal{X}} (\hat{g}_n(\mathbf{x}) - g(\mathbf{x}))^2 d\mathbf{x})$ of the density estimator over the support \mathcal{X} of X , that is,

$$h_{n,\star} = \left(\frac{p \int_{\mathbb{R}^p} K^2}{\int_{\mathcal{X}} \left(\int_{\mathbb{R}^p} (\mathbf{u}^\top H_g(\mathbf{x}) \mathbf{u}) K(\mathbf{u}) d\mathbf{u} \right)^2 d\mathbf{x}} \right)^{1/(p+4)} n^{-1/(p+4)}. \quad (5.1)$$

For example, when $p = 1$, the classical normal scale rule derived from (5.1), assuming a Gaussian p.d.f. g and a naive kernel $K(u) = 1/2$ on $[-1, 1]$ yields $\hat{h}_{n,\star} = (12\sqrt{\pi})^{1/5} \hat{\sigma}_n n^{-1/5}$, where $\hat{\sigma}_n$ is the empirical standard deviation of the X_t . We adopt this version in our examples with $p = 1$ and abuse notation by denoting it $h_{n,\star}$.

We turn to a choice of k_n optimizing the bias-variance tradeoff in extreme value estimation: if the chosen tail index estimator has an asymptotic variance $V(\mathbf{x})$ and an asymptotic bias component $\lambda(\mathbf{x})B(\mathbf{x})$, where $\lambda(\mathbf{x}) = \lim_{n \rightarrow \infty} \sqrt{k_n h_{n,\star}^p} A(n/k_n|\mathbf{x})$, we define

$$k_{n,\star}(\mathbf{x}) = \arg \min_{1 \leq k \leq n} \{ (kh_{n,\star}^p)^{-1} V(\mathbf{x}) + B^2(\mathbf{x}) A^2(n/k|\mathbf{x}) \}.$$

A common practice is to consider the very general case when $A(t|\mathbf{x}) = b(\mathbf{x})\gamma(\mathbf{x})t^{\rho(\mathbf{x})}$, for $b(\mathbf{x}) \in \mathbb{R}$ and $\rho(\mathbf{x}) < 0$, which is satisfied in the vast majority of heavy-tailed models, see among others Gomes and Pestana (2007). This yields the closed form expression

$$k_{n,\star}(\mathbf{x}) = \left(\frac{1}{-2\rho(\mathbf{x})b^2(\mathbf{x})\gamma^2(\mathbf{x})} \frac{V(\mathbf{x})}{B^2(\mathbf{x})} \right)^{1/(1-2\rho(\mathbf{x}))} h_{n,\star}^{-p/(1-2\rho(\mathbf{x}))} n^{-2\rho(\mathbf{x})/(1-2\rho(\mathbf{x}))}.$$

The quantities $b(\mathbf{x})$ and $\rho(\mathbf{x})$ are estimated using regression versions $\bar{b}(\mathbf{x})$ and $\bar{\rho}(\mathbf{x})$ of the estimators provided by the R function `mop` from the package `evt0`, see Section C.1 for explicit expressions. To estimate $B(\mathbf{x})$ and $V(\mathbf{x})$, note that when the tail index estimator is chosen as $\hat{\gamma}_{1-k_n/n}^{(J)}(\mathbf{x})$, we have $V(\mathbf{x}) = (\int_{\mathbb{R}^p} K^2/g(\mathbf{x}))v_q^{(J)}(\mathbf{x})$ and $B(\mathbf{x}) = B^{(J)}(\mathbf{x})$, with

$$v_q^{(J)}(\mathbf{x}) = \gamma^2(\mathbf{x}) \frac{J(J-1)(2J-1)}{6 \log^2(J!)} \quad \text{and} \quad B^{(J)}(\mathbf{x}) = \frac{1}{\log(J!)} \sum_{j=1}^J \frac{j^{\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})}.$$

When $\hat{\gamma}_{1-k_n/n}^E(\mathbf{x})$ is chosen, then $V(\mathbf{x}) = (\int_{\mathbb{R}^p} K^2/g(\mathbf{x}))v_E(\mathbf{x})$ and $B(\mathbf{x}) = B_E(\mathbf{x})$ where

$$v_E(\mathbf{x}) = \frac{\gamma^3(\mathbf{x})(1-\gamma(\mathbf{x}))}{1-2\gamma(\mathbf{x})} \quad \text{and} \quad B_E(\mathbf{x}) = \frac{\gamma(\mathbf{x})(1/\gamma(\mathbf{x})-1)^{1-\rho(\mathbf{x})}}{1-\gamma(\mathbf{x})-\rho(\mathbf{x})}.$$

These quantities are estimated by plugging in $\hat{g}_n(\mathbf{x})$ (with $h = h_{n,\star}$) and $\bar{\rho}(\mathbf{x})$ in place of $g(\mathbf{x})$ and $\rho(\mathbf{x})$, and a local regression version $\bar{\gamma}(\mathbf{x})$ of the Hill estimator in place of $\gamma(\mathbf{x})$, with $h = h_{n,\star}$ and effective sample size k corresponding to 25% of the local sample size $N_h(\mathbf{x}) = \sum_{t=1}^n \mathbb{1}_{\{\|\mathbf{x}_t - \mathbf{x}\| \leq h\}}$ (see Section C.1 for the expression of $\bar{\gamma}(\mathbf{x})$).

This results in data-driven parameter values $\hat{k}_{n,\star}^{(J)}(\mathbf{x})$ and $\hat{k}_{n,\star}^E(\mathbf{x})$, depending on the chosen tail index estimation technique. We omit the dependence of this selected value upon the tail index estimator chosen and we denote it again by $k_{n,\star}$ for the sake of brevity. The approximations made here might be far from the truth in certain difficult cases, but they afford effective data-based rules as we shall see on simulated and real data.

5.2 Bias correction guidelines

Our extrapolation procedures are based on the simplifying parametric assumption that the conditional tail is purely Pareto. The quality of this approximation deteriorates as $\rho(\mathbf{x})$ gets closer to 0, with the resulting bias being possibly very substantial then. Bias-reduced versions of $\hat{\gamma}_{1-k_n/n}^{(J)}(\mathbf{x})$ and $\hat{q}_{n,\tau_n}^W(\tau'_n|\mathbf{x})$ are

$$\begin{aligned} \hat{\gamma}_{1-k_n/n}^{(J,BR)}(\mathbf{x}) &= \hat{\gamma}_{1-k_n/n}^{(J)}(\mathbf{x}) \left(1 - \frac{1}{\log(J!)} \sum_{j=1}^J \frac{j^{\bar{\rho}(\mathbf{x})} - 1}{\bar{\rho}(\mathbf{x})} \bar{b}(\mathbf{x}) \left(\frac{n}{k_n} \right)^{\bar{\rho}(\mathbf{x})} \right) \\ \text{and } \hat{q}_{n,\tau_n}^{W,BR}(\tau'_n|\mathbf{x}) &= \hat{q}_{n,\tau_n}^W(\tau'_n|\mathbf{x}) \left(1 + \frac{\left(\frac{k_n}{n(1-\tau'_n)} \right)^{\bar{\rho}(\mathbf{x})} - 1}{\bar{\rho}(\mathbf{x})} \bar{b}(\mathbf{x}) \hat{\gamma}_{1-k_n/n}^{(J,BR)}(\mathbf{x}) \left(\frac{n}{k_n} \right)^{\bar{\rho}(\mathbf{x})} \right). \end{aligned}$$

These estimators are respectively inspired by Theorem 3 and by the approach of Gomes and Pestana (2007). We use $\hat{\gamma}_{1-k_n/n}^{(J,BR)}(\mathbf{x})$ within $\hat{q}_{n,\tau_n}^W(\tau'_n|\mathbf{x})$ in our subsequent implementation of $\hat{q}_{n,\tau_n}^{W,BR}(\tau'_n|\mathbf{x})$.

As for extreme conditional expectile-based estimation, we note that bias reduction is even more crucial, due to the fact that, in view of Theorems 5 and 6, the estimators contain

two sources of bias: one coming from the second-order regular variation framework, and the other stemming from the asymptotic proportionality between quantiles and expectiles. We suggest using the following bias-reduced version of $\widehat{\gamma}_{1-k_n/n}^E(\mathbf{x})$:

$$\begin{aligned}\widehat{\gamma}_{1-k_n/n}^{E,BR}(\mathbf{x}) &= \left(1 + \frac{\widehat{F}_n(\widehat{e}_n(1-k_n/n|\mathbf{x})|\mathbf{x})}{k_n/n} \frac{1}{1 + \widehat{r}(1-k_n/n|\mathbf{x})}\right)^{-1} \text{ where } \widehat{r}(1-k_n/n|\mathbf{x}) \\ &= \left(1 - \frac{\widehat{m}_n(\mathbf{x})}{\widehat{e}_n(1-k_n/n|\mathbf{x})}\right) \frac{1}{1-2k_n/n} \left(1 + \frac{\bar{b}(\mathbf{x})[\widehat{F}_n(\widehat{e}_n(1-k_n/n|\mathbf{x})|\mathbf{x})]^{-\bar{\rho}(\mathbf{x})}}{1 - \widehat{\gamma}_{1-k_n/n}^E(\mathbf{x}) - \bar{\rho}(\mathbf{x})}\right)^{-1} - 1.\end{aligned}$$

This is motivated by work in Girard et al. (2021b) in the unconditional setting. Here $\widehat{\gamma}_{1-k_n/n}^{E,BR}(\mathbf{x})$ is computed using the R function `tindexp` with argument `br=TRUE` (from the R package `Expectrem`) applied to the Y_t such that $\|\mathbf{X}_t - \mathbf{x}\| \leq h_{n,\star}$. We similarly compute a bias-reduced version of the extreme conditional expectile estimator \widehat{e}_{n,τ_n}^W using the R function `extExpect` with `br=TRUE` from that package; see Section C.2 for the full expression of this estimator $\widehat{e}_{n,\tau_n}^{W,BR}$.

5.3 Pointwise asymptotic confidence intervals

5.3.1 Extremal conditional quantiles

We use the equivalent version of Theorem 2 on the log-scale, which tends to be more accurate in practice. Applied to the bias-reduced version $\widehat{q}_{n,1-k_{n,\star}/n}^{W,BR}(\tau'_n|\mathbf{x})$, this suggests that $\log(\widehat{q}_{n,1-k_{n,\star}/n}^{W,BR}(\tau'_n|\mathbf{x})/q(\tau'_n|\mathbf{x}))$ is approximately Gaussian centered with variance $(\int_{\mathbb{R}^p} K^2/g(\mathbf{x}))v_q^{(J)}(\mathbf{x})$, with $v_q^{(J)}(\mathbf{x}) = \gamma^2(\mathbf{x}) \times J(J-1)(2J-1)/(6\log^2(J!))$. A first 95% asymptotic Gaussian confidence interval for $q(\tau'_n|\mathbf{x})$ is then

$$\begin{aligned}\widehat{I}_{q,1}(\tau'_n|\mathbf{x}) &= \left[\widehat{q}_{n,1-k_{n,\star}/n}^{W,BR}(\tau'_n|\mathbf{x}) \exp\left(\pm \frac{\sqrt{\int_{\mathbb{R}^p} K^2/g_n(\mathbf{x})} \widehat{v}_q^{(J)}(\mathbf{x})}{\sqrt{k_{n,\star} h_{n,\star}^p}} \log\left(\frac{k_{n,\star}}{n(1-\tau'_n)}\right) z_{0.975}\right) \right] \\ &\quad \text{with } \widehat{v}_q^{(J)}(\mathbf{x}) = \frac{J(J-1)(2J-1)}{6\log^2(J!)} (\widehat{\gamma}_{1-k_{n,\star}/n}^{(J,BR)}(\mathbf{x}))^2,\end{aligned}$$

and where z_τ is the τ th quantile of the standard normal distribution. This relies exclusively on the asymptotic distribution of the tail index estimator used in the extrapolation. For small sample sizes, the variability of the intermediate quantile estimator also has an impact on the variance of the extrapolated estimator. A direct application of Theorem 3 provides a refined version $\widetilde{v}_q^{(J)}(\mathbf{x})$ of the asymptotic variance estimate of $\log(\widehat{q}_{n,1-k_{n,\star}/n}^{W,BR}(\tau'_n|\mathbf{x})/q(\tau'_n|\mathbf{x}))$ and thus a corrected 95% asymptotic Gaussian confidence interval, as

$$\begin{aligned}\widehat{I}_{q,2}(\tau'_n|\mathbf{x}) &= \left[\widehat{q}_{n,1-k_{n,\star}/n}^{W,BR}(\tau'_n|\mathbf{x}) \exp\left(\pm \frac{\sqrt{\int_{\mathbb{R}^p} K^2/g_n(\mathbf{x})} \widetilde{v}_q^{(J)}(\mathbf{x})}{\sqrt{k_{n,\star} h_{n,\star}^p}} \log\left(\frac{k_{n,\star}}{n(1-\tau'_n)}\right) z_{0.975}\right) \right], \\ &\quad \text{with } \widetilde{v}_q^{(J)}(\mathbf{x}) = \left(\frac{J(J-1)(2J-1)}{6\log^2(J!)} + \frac{1}{\log^2(k_{n,\star}/[n(1-\tau'_n)])}\right) (\widehat{\gamma}_{1-k_{n,\star}/n}^{(J,BR)}(\mathbf{x}))^2.\end{aligned}$$

Simulation evidence shows that this correction improves coverage for low sample sizes.

5.3.2 Extremal conditional expectiles

Our extreme conditional expectile estimators are constructed using the same Weissman extrapolation argument: we build upon Theorems 5 and 6 to deduce a first 95% asymptotic Gaussian confidence interval for $e(\tau'_n|\mathbf{x})$ as

$$\widehat{I}_{E,1}(\tau'_n|\mathbf{x}) = \left[\widehat{e}_{n,1-k_{n,\star}/n}^{W,BR}(\tau'_n|\mathbf{x}) \exp \left(\pm \frac{\sqrt{\int_{\mathbb{R}^p} K^2 \widehat{v}_E(\mathbf{x})}}{\widehat{g}_n(\mathbf{x})} \frac{\widehat{v}_E(\mathbf{x})}{\sqrt{k_{n,\star} h_{n,\star}^p}} \log \left(\frac{k_{n,\star}}{n(1-\tau'_n)} \right) z_{0.975} \right) \right]$$

$$\text{with } \widehat{v}_E(\mathbf{x}) = \frac{(\widehat{\gamma}_{1-k_{n,\star}/n}^{E,BR}(\mathbf{x}))^3 (1 - \widehat{\gamma}_{1-k_{n,\star}/n}^{E,BR}(\mathbf{x}))}{1 - 2\widehat{\gamma}_{1-k_{n,\star}/n}^{E,BR}(\mathbf{x})}.$$

For a low-to-moderate sample size n , the empirical variance of the estimates tends to be very far from the asymptotic variance. This is not only due to neglecting the correlation between the conditional tail index estimator and the empirical expectiles, but also to the use of the asymptotic proportionality between extreme quantiles and expectiles in the derivation of the asymptotic results. An inspection of the proofs of Theorems 4 and 6 (see Section C.3) suggests that a corrected asymptotic variance matrix for the random vector

$$\sqrt{k_{n,\star} h_{n,\star}^p} \left(\frac{n\widehat{F}_n(\widehat{e}_n(1 - k_{n,\star}/n|\mathbf{x})|\mathbf{x})}{k_{n,\star}} - \left(\frac{1}{\gamma(\mathbf{x})} - 1 \right), \frac{\widehat{e}_n(1 - k_{n,\star}/n|\mathbf{x})}{e(1 - k_{n,\star}/n|\mathbf{x})} - 1 \right)$$

is $(\int_{\mathbb{R}^p} K^2/g(\mathbf{x}))\mathbf{T}_n(\mathbf{x})$ where the 2×2 symmetric matrix $\mathbf{T}_n(\mathbf{x})$ has components

$$T_{n,11}(\mathbf{x}) = \frac{2(1-\gamma(\mathbf{x}))^2}{\gamma(\mathbf{x})(1-2\gamma(\mathbf{x}))} \frac{\kappa_{1,n}(\mathbf{x})}{\kappa_{2,n}^2(\mathbf{x})} - 2 \frac{1-\gamma(\mathbf{x})}{\gamma(\mathbf{x})} \frac{\sqrt{\kappa_{1,n}(\mathbf{x})}}{\kappa_{2,n}(\mathbf{x})} + \frac{1-\gamma(\mathbf{x})}{\gamma(\mathbf{x})},$$

$$T_{n,12}(\mathbf{x}) = -\frac{2\gamma(\mathbf{x})(1-\gamma(\mathbf{x}))}{1-2\gamma(\mathbf{x})} \frac{\kappa_{1,n}(\mathbf{x})}{\kappa_{2,n}(\mathbf{x})} + \gamma(\mathbf{x})\sqrt{\kappa_{1,n}(\mathbf{x})}, \quad T_{n,22}(\mathbf{x}) = \frac{2\gamma^3(\mathbf{x})}{1-2\gamma(\mathbf{x})} \kappa_{1,n}(\mathbf{x}),$$

$$\text{with } \kappa_{1,n}(\mathbf{x}) = \frac{1-2k_{n,\star}/n}{1-m(\mathbf{x})/e(1-k_{n,\star}/n|\mathbf{x})} \text{ and } \kappa_{2,n}(\mathbf{x}) = 1 - \frac{\gamma(\mathbf{x})m(\mathbf{x})}{e(1-k_{n,\star}/n|\mathbf{x})}.$$

Note that $\kappa_{1,n}(\mathbf{x})$ and $\kappa_{2,n}(\mathbf{x})$ indeed converge to 1 as $n \rightarrow \infty$. They are estimated by replacing $\gamma(\mathbf{x})$, $m(\mathbf{x})$ and $e(1 - k_{n,\star}/n|\mathbf{x})$ with $\widehat{\gamma}_{1-k_{n,\star}/n}^{E,BR}(\mathbf{x})$, $\widehat{m}_n(\mathbf{x})$ and $\widehat{e}_n(1 - k_{n,\star}/n|\mathbf{x})$, respectively, yielding estimators $\widehat{\kappa}_{1,n}(\mathbf{x})$ and $\widehat{\kappa}_{2,n}(\mathbf{x})$ and therefore $\widehat{T}_{n,11}(\mathbf{x})$, $\widehat{T}_{n,12}(\mathbf{x})$ and $\widehat{T}_{n,22}(\mathbf{x})$. Then, observe that up order $\sqrt{k_{n,\star} h_{n,\star}^p}$, we have $\widehat{\gamma}_{1-k_{n,\star}/n}^{E,BR}(\mathbf{x}) \approx \mathcal{G}(n\widehat{F}_n(\widehat{e}_n(1 - k_{n,\star}/n|\mathbf{x})|\mathbf{x})/k_{n,\star})$, where $\mathcal{G}(u) = 1/(1+u)$. A power series expansion suggests

$$\sqrt{k_{n,\star} h_{n,\star}^p} \left(\mathcal{G}(n\widehat{F}_n(\widehat{e}_n(1 - k_{n,\star}/n|\mathbf{x})|\mathbf{x})/k_{n,\star}) - \mathcal{G}(1/\gamma(\mathbf{x}) - 1) \right)$$

$$= \sum_{k=1}^{\infty} \frac{(-1)^k \gamma^{k+1}(\mathbf{x})}{(k_{n,\star} h_{n,\star}^p)^{(k-1)/2}} \left(\sqrt{k_{n,\star} h_{n,\star}^p} \left(\frac{n\widehat{F}_n(\widehat{e}_n(1 - k_{n,\star}/n|\mathbf{x})|\mathbf{x})}{k_{n,\star}} - \left(\frac{1}{\gamma(\mathbf{x})} - 1 \right) \right) \right)^k$$

because $\mathcal{G}^{(k)}(u) = (-1)^k k!/(1+u)^{k+1}$ for any $k \geq 1$. [This is conceptually simpler than an Edgeworth expansion, which would approximate the p.d.f. of $\widehat{\gamma}_{1-k_{n,\star}/n}^{E,BR}(\mathbf{x})$.] Taking only the first term above leads to the delta-method for $\widehat{\gamma}_{1-k_{n,\star}/n}^{E,BR}(\mathbf{x})$; to obtain a finer approximation, we use all terms up to order 4. Based on our calculation of $\mathbf{T}_n(\mathbf{x})$, an asymptotic approximation of the covariance matrix of $\sqrt{k_{n,\star} h_{n,\star}^p}(\widehat{\gamma}_{1-k_{n,\star}/n}^{E,BR}(\mathbf{x}) - \gamma(\mathbf{x}), \widehat{e}_n(1 - k_{n,\star}/n|\mathbf{x}) - e(1 - k_{n,\star}/n|\mathbf{x}))$

$k_{n,\star}/n|\mathbf{x})/e(1 - k_{n,\star}/n|\mathbf{x}) - 1)$ is $(\int_{\mathbb{R}^p} K^2/g(\mathbf{x}))\mathbf{S}_n(\mathbf{x})$, where the symmetric matrix $\mathbf{S}_n(\mathbf{x})$ has components

$$\begin{aligned} S_{n,11}(\mathbf{x}) &= \gamma^4(\mathbf{x})T_{n,11}(\mathbf{x}) \left(1 + 8 \frac{\gamma^2(\mathbf{x})T_{n,11}(\mathbf{x})}{k_{n,\star}h_{n,\star}} \times \frac{\int_{\mathbb{R}^p} K^2}{g(\mathbf{x})}\right), \\ S_{n,12}(\mathbf{x}) &= -\gamma^2(\mathbf{x})T_{n,12}(\mathbf{x}) \left(1 + 3 \frac{\gamma^2(\mathbf{x})T_{n,11}(\mathbf{x})}{k_{n,\star}h_{n,\star}} \times \frac{\int_{\mathbb{R}^p} K^2}{g(\mathbf{x})}\right) \text{ and } S_{n,22}(\mathbf{x}) = T_{n,22}(\mathbf{x}). \end{aligned}$$

We denote $\widehat{\mathbf{S}}_n(\mathbf{x})$ the associated estimator, which hinges upon the estimators $\widehat{\mathbf{T}}_n(\mathbf{x})$ and $\widehat{\gamma}_{1-k_{n,\star}/n}^{E,BR}(\mathbf{x})$ previously introduced. The final step is to recall that (see Section C.2)

$$\begin{aligned} \log \frac{\widehat{e}_{n,1-k_{n,\star}/n}^{W,BR}(\tau'_n|\mathbf{x})}{e(\tau'_n|\mathbf{x})} &\approx \log \frac{\widehat{e}_{n,1-k_{n,\star}/n}^W(\tau'_n|\mathbf{x})}{e(\tau'_n|\mathbf{x})} - (\widehat{\gamma}_{1-k_{n,\star}/n}^{E,BR}(\mathbf{x}) - \gamma(\mathbf{x})) \log(\kappa_{1,n}(\mathbf{x})) \\ &\approx \left(\log \left(\frac{k_{n,\star}}{n(1 - \tau'_n)} \right) - \log(\kappa_{1,n}(\mathbf{x})) \right) (\widehat{\gamma}_{1-k_{n,\star}/n}^{E,BR}(\mathbf{x}) - \gamma(\mathbf{x})) + \log \frac{\widehat{e}_n(1 - k_{n,\star}/n|\mathbf{x})}{e(1 - k_{n,\star}/n|\mathbf{x})}. \end{aligned}$$

The variance of $(\sqrt{k_{n,\star}h_{n,\star}^p}/\log(k_{n,\star}/(n(1 - \tau'_n)))) \times \log(\widehat{e}_{n,1-k_{n,\star}/n}^{W,BR}(\tau'_n|\mathbf{x})/e(\tau'_n|\mathbf{x}))$ is then estimated by $(\int_{\mathbb{R}^p} K^2/\widehat{g}_n(\mathbf{x}))\widetilde{v}_E(\mathbf{x})$, where

$$\widetilde{v}_E(\mathbf{x}) = \frac{\widehat{S}_{n,11}(\mathbf{x})\widehat{\mathcal{L}}_n^2 + 2\widehat{S}_{n,12}(\mathbf{x})\widehat{\mathcal{L}}_n + \widehat{S}_{n,22}(\mathbf{x})}{\log^2(k_{n,\star}/(n(1 - \tau'_n)))} \text{ and } \widehat{\mathcal{L}}_n = \log(k_{n,\star}/(n(1 - \tau'_n)\widehat{\kappa}_{1,n}(\mathbf{x}))).$$

This results in the corrected 95% asymptotic Gaussian confidence interval

$$\widehat{I}_{E,2}(\tau'_n|\mathbf{x}) = \left[\widehat{e}_{n,1-k_{n,\star}/n}^{W,BR}(\tau'_n|\mathbf{x}) \exp \left(\pm \frac{\sqrt{\int_{\mathbb{R}^p} K^2 \widetilde{v}_E(\mathbf{x})}}{\sqrt{k_{n,\star}h_{n,\star}^p}} \log \left(\frac{k_{n,\star}}{n(1 - \tau'_n)} \right) z_{0.975} \right) \right].$$

It is shown below to have a greatly improved coverage probability compared to $\widehat{I}_{E,1}(\tau'_n|\mathbf{x})$.

6 Simulation study

6.1 Models and setup

For the sake of brevity, we only report here results in the following nonlinear (Burr) regression model, with covariate dimension $p = 1$. Other cases spanning our list of worked-out examples are presented in Section C.4.

We consider a nonlinear Burr process $Y_t = ((1 - U_t)^{\rho(X_t)} - 1)^{-\gamma(X_t)/\rho(X_t)}$ where:

- $X_t = \Phi(Z_t)$, with Φ being the standard Gaussian c.d.f. and (Z_t) (simulated using the `garch.sim` routine from the R package TSA) is a GARCH(1,1) process with $\omega = 0.25$, $\alpha = 0.75$, $\beta = 0.2$, *i.e.* $Z_{t+1} = \Sigma_{t+1}\delta_{t+1}$, where the δ_t are i.i.d. standard Gaussian and Σ_{t+1} is defined recursively as $\Sigma_{t+1} = (\omega + \alpha Z_t^2 + \beta \Sigma_t^2)^{1/2}$.
- $\varepsilon_t = \Phi_\nu^{-1}(U_t)$, where Φ_ν is the Student c.d.f. and $\nu = 10/3$, and (U_t) is defined recursively as $U_0 \sim \text{Uniform}[0, 1]$ and, for $t \geq 1$, $U_t = \frac{1}{r}U_{t-1} + \eta_t$, where the η_t are i.i.d. uniformly drawn over $\{0, 1/r, \dots, (r-1)/r\}$, and $r = 5$.

We fix $\rho(x) = -1$ for all $x \in [0, 1]$ and consider three different models for $\gamma(x)$, $x \in [0, 1]$:

(NL-P) A polynomial trend $\gamma(x) = 0.5x(1 - x) + 0.15$;

(NL-S) A sinusoidal model $\gamma(x) = 0.2 + 0.05 \sin(2\pi x)$;

(NL-C) The constant model $\gamma(x) = 0.2$.

In these three cases, $\gamma(x) \in (0, 1/2)$ for any $x \in [0, 1]$. The true value of the conditional quantile is $q(\tau|x) = ((1 - \tau)^{-1} - 1)^{\gamma(x)}$. The theoretical conditional expectile $e(\tau|x)$ is computed numerically using the function `eburr` in the R package `Expectrem`.

We simulate $N = 1,000$ replications of size $n = 10,000$ of each of these models (results for $n \in \{1,000, 5,000\}$ are found in Section C.4). We estimate extreme conditional quantiles and expectiles at level $\tau'_n = 1 - 10/n = 0.999$. We let $K(u) = 0.5 \mathbb{1}_{\{|u| \leq 1\}}$ be the uniform kernel, $h_n = h_{n,\star}$ and $\tau_n = \tau_{n,\star} = 1 - k_{n,\star}/n$ selected as in Section 5.1 and, as far as extreme quantile estimation is concerned, $J = 9$ in the conditional tail index estimator $\hat{\gamma}_{1-k_{n,\star}/n}^{(J, BR)}(x)$. Our quantile (resp. expectile) estimators are compared with the non-extrapolated quantile estimator $\hat{q}_n(\tau'_n|x)$ (resp. the non-extrapolated expectile estimator $\hat{e}_n(\tau'_n|x)$) and the simple extrapolated version $\hat{q}_{n,1-k_{n,\star}/n}^W(\tau'_n|x)$ based on $\hat{\gamma}_{1-k_{n,\star}/n}^{(J)}$ (resp. the simple extrapolated version $\hat{e}_{n,1-k_{n,\star}/n}^W(\tau'_n|x)$ based on $\hat{\gamma}_{1-k_{n,\star}/n}^E(x)$). Our proposed 95% asymptotic Gaussian confidence intervals for $q(\tau'_n|x)$ are compared with the following competing intervals:

$$\hat{I}_{q,3}(\tau'_n|x) = \left[\hat{q}_{n,1-k_{n,\star}/n}^W(\tau'_n|x) \exp \left(\pm \frac{\sqrt{\frac{\int_{\mathbb{R}} K^2}{\hat{g}_n(x)} \bar{v}_q^{(J)}(x)}}{\sqrt{k_{n,\star} h_{n,\star}}} \log \left(\frac{k_{n,\star}}{n(1 - \tau'_n)} \right) z_{0.975} \right) \right],$$

with $\bar{v}_q^{(J)}(x) = \frac{J(J-1)(2J-1)}{6 \log^2(J!)} (\hat{\gamma}_{1-k_{n,\star}/n}^{(J)}(x))^2$, based on naive Weissman extrapolation and a direct application of Theorem 2 on the log-scale, and

$$\hat{I}_{q,4}(\tau'_n|x) = \left[\hat{q}_n(\tau'_n|x) \exp \left(\pm \hat{\gamma}_{1-k_{n,\star}/n}^{(J)}(x) \frac{\sqrt{\frac{\int_{\mathbb{R}} K^2}{\hat{g}_n(x)}}}{\sqrt{n h_{n,\star} (1 - \tau'_n)}} z_{0.975} \right) \right],$$

based on an equivalent version of Theorem 1 on the log-scale, which does not feature any extrapolation at all. Likewise, for $e(\tau'_n|x)$, we compute

$$\hat{I}_{e,3}(\tau'_n|x) = \left[\hat{e}_{n,1-k_{n,\star}/n}^W(\tau'_n|x) \exp \left(\pm \frac{\sqrt{\frac{\int_{\mathbb{R}} K^2}{\hat{g}_n(x)} \bar{v}_E(x)}}{\sqrt{k_{n,\star} h_{n,\star}}} \log \left(\frac{k_{n,\star}}{n(1 - \tau'_n)} \right) z_{0.975} \right) \right],$$

with $\bar{v}_E(x) = \frac{(\hat{\gamma}_{1-k_{n,\star}/n}^E(x))^3 (1 - \hat{\gamma}_{1-k_{n,\star}/n}^E(x))}{1 - 2\hat{\gamma}_{1-k_{n,\star}/n}^E(x)}$, inspired by Theorem 5, and

$$\hat{I}_{e,4}(\tau'_n|x) = \left[\hat{e}_n(\tau'_n|x) \exp \left(\pm \sqrt{\frac{2(\hat{\gamma}_{1-k_{n,\star}/n}^E(x))^3}{1 - 2\hat{\gamma}_{1-k_{n,\star}/n}^E(x)}} \times \frac{\sqrt{\frac{\int_{\mathbb{R}} K^2}{\hat{g}_n(x)}}}{\sqrt{n h_{n,\star} (1 - \tau'_n)}} z_{0.975} \right) \right],$$

suggested by an equivalent version of Theorem 4 on the log-scale. This will make it possible to assess the benefits of the extrapolation procedure and of the bias correction scheme.

6.2 Results

We represent in Figure 3 boxplots of the extreme conditional quantile and expectile estimates and the coverage probabilities of the intervals $\hat{I}_{q,j}(\tau'_n|x)$ and $\hat{I}_{E,j}(\tau'_n|x)$, for $1 \leq j \leq 4$,

in models (NL-P), (NL-S) and (NL-C). We discuss the conclusions from a full set of results, in an expanded list of models and for sample sizes $n \in \{1,000, 5,000, 10,000\}$, that are found in Section C.4 of the Supplementary Material, see Figures C.1–C.6 therein.

While empirical point estimates without extrapolation seem at first glance to perform respectably, their variance is substantially larger than that of extrapolated estimates, and most importantly the lack of extrapolation makes accurate inference impossible, as can be seen from the very low coverage probabilities of $\widehat{I}_{q,4}(\tau'_n|x)$ and $\widehat{I}_{E,4}(\tau'_n|x)$. Extrapolation is therefore necessary in order to correctly infer extreme quantiles and expectiles. While bias reduction and variance correction for extrapolated estimates are arguably not crucial for quantile estimation at large values of n , bias reduction is indeed valuable when the second-order parameter $\rho(x)$ gets close to 0 (see the top panels of Figure C.5), and variance correction brings noticeable improvements when n is moderately large ($n = 1,000$), with coverage probabilities typically improving by about 5% when using our proposed interval $\widehat{I}_{q,2}(\tau'_n|x)$ instead of $\widehat{I}_{q,1}(\tau'_n|x)$, see the bottom panels of Figures C.1–C.6. Bias and variance correction are, however, of prime importance when estimating and inferring extreme conditional expectiles: standard extrapolated estimates are heavily biased, with the asymptotic consistency property not visibly evidenced even for large n , see the middle panels of Figures C.1–C.6. This is due to the presence of a typically very large bias term proportional to the reciprocal of the extreme conditional quantile in the expectile extrapolation formula that is key for the asymptotic normality of extrapolated expectile estimators, see Lemma 10 in Section A.4. In addition, the uncorrected expression of the asymptotic variance of the extrapolated expectile estimates yields values very far from their sample variance, while our corrected proposal gets very close to this observed variance, see the rightmost panels in Figure 3. These are the main reasons why the intervals $\widehat{I}_{E,1}(\tau'_n|x)$ and $\widehat{I}_{E,3}(\tau'_n|x)$ have very poor coverage, while $\widehat{I}_{E,2}(\tau'_n|x)$ achieves a coverage remarkably close to the nominal rate when $n \geq 5,000$ and performs much better than its competitors when $n = 1,000$, see the bottom panels of Figures C.1–C.6. It should also be highlighted that in a substantial number of cases, the non-corrected variance estimate $\widehat{v}_E(x)$ was not positive and therefore the associated confidence interval $\widehat{I}_{E,3}(\tau'_n|x)$ could not be calculated; by contrast, the interval $\widehat{I}_{E,2}(\tau'_n|x)$ was always well-defined.

We conclude that the Gaussian asymptotic theory, when properly corrected in order to account for the presence of conditional heavy tails, provides reasonably accurate inference. It should not come as a surprise that good inferential results require a fairly large sample size, of the order of several thousands, due to the four main difficulties of the problem: (i) Temporal mixing, which impacts the trustworthiness of the asymptotic Gaussian limits, (ii) Nonparametric smoothing, inducing an approximation bias due to local variation of the underlying distribution, (iii) Sparsity in x due to the nonparametric regression framework, and (iv) Sparsity in y due to the extreme value context. Sparsity in x is a particular concern in the difficult model (AR), where the covariate, being a lagged response, takes unbounded values and where a correct inference can only be expected where observations Y_t concentrate.

7 Real data analysis

7.1 Stock returns data

The first dataset, available from the R package HRW, contains the $n = 2,363$ values of the excess daily stock returns (daily log-returns minus risk-free interest rate) on General Electric and the S&P 500 index, from 1 November 1993 to 31 March 2003. Let X_t (resp. Y_t)

be the negative of the excess daily log-return of the S&P 500 index (resp. General Electric), so that large values represent large losses. About half of the data is negative, which is detrimental to the calculation of our bias-reduced estimators that require only positive values, see Section C.1. We thus shift most of the data above 0 by subtracting to the Y_t their empirical unconditional 10% quantile, approximately equal to -0.023 , *i.e.* we apply our procedures to $Y'_t = Y_t + 0.023$ before shifting back to the original position.

As in our simulation study, we infer extremal regression quantiles and expectiles at level $\tau'_n = 1 - 10/n \approx 0.995$ by following the methodology described in Section 5.1. This results in the global bandwidth $h_{n,\star} \approx 0.0045$. We represent in the middle panel of Figure 1 (restricted to the interval $X_t \in [-0.015, 0.015]$ which contains more than 80% of the data) our bias-reduced extrapolated estimates $\hat{q}_{n,1-k_{n,\star}/n}^{W,BR}(\tau'_n|x)$ and $\hat{e}_{n,1-k_{n,\star}/n}^{W,BR}(\tau'_n|x)$, along with their respective asymptotic 95% confidence intervals $\hat{I}_{q,2}(\tau'_n|x)$ and $\hat{I}_{E,2}(\tau'_n|x)$. The purely empirical competitors $\hat{q}_n(\tau'_n|x)$ and $\hat{e}_n(\tau'_n|x)$ are graphed in the left panel along with their respective asymptotic 95% confidence intervals $\hat{I}_{q,4}(\tau'_n|x)$ and $\hat{I}_{E,4}(\tau'_n|x)$. In agreement to what was observed in Section 6.2, the intervals $\hat{I}_{q,4}(\tau'_n|x)$ and $\hat{I}_{E,4}(\tau'_n|x)$ are unrealistically narrower than $\hat{I}_{q,2}(\tau'_n|x)$ and $\hat{I}_{E,2}(\tau'_n|x)$, respectively.

The estimated curves, confidence intervals, and regression in mean in the left and middle panels all point towards a linear trend. This motivated us to perform a residual-based extreme value estimation from the ordinary linear regression model for comparison purposes. We thus construct the residuals $\hat{\varepsilon}_t = Y'_t - (-0.00030 + 1.24X_t)$, and we calculate corresponding residual-based, bias-corrected extreme quantile estimates $\hat{q}_{n,\tau_n,\varepsilon}(\tau'_n)$, following Gomes and Pestana (2007), and bias-corrected extreme expectile estimates $\hat{e}_{n,\tau_n,\varepsilon}(\tau'_n)$, following Girard et al. (2021b), in which we chose $\tau_n = 1 - 200/n = 0.915$ after a graphical inspection of a standard Hill plot. Their associated asymptotic 95% Gaussian confidence intervals on the log-scale are also constructed. This yields conditional extreme point quantile and expectile estimates of the Y'_t as

$$\bar{q}_{n,\tau_n}(\tau'_n|x) = -0.00030 + 1.24x + \hat{q}_{n,\tau_n,\varepsilon}(\tau'_n), \quad \bar{e}_{n,\tau_n}(\tau'_n|x) = -0.00030 + 1.24x + \hat{e}_{n,\tau_n,\varepsilon}(\tau'_n)$$

and their corresponding confidence intervals through the same shift. These are then all shifted back to the original location of the Y_t and graphed in the right panel of Figure 1. Although the middle and right panels yield a broadly similar message, the confidence intervals from the linear model are wider than our bias- and variance-corrected nonparametric confidence intervals. This is probably due to the unavoidable assumption of constant tail index in the linear model, whose validity is unclear here. The extremal linear regression estimates are thus inevitably driven by the few largest observations in the data cloud. By contrast, our nonparametric method is able to finely differentiate conditional extreme value behavior when x varies. Accurate inference on both extremal regression modes, without recourse to the *a priori* assumptions of linearity and common tail, is crucial in order to produce correct tail risk appraisal. This is especially important for conditional expectiles that typically result in more liberal assessments of risk than conditional quantiles because they satisfy the diversification principle, and that here appear indeed to induce less conservative risk measurements.

7.2 Tornado losses data

This dataset² records, for each tornado that has occurred in the United States between 1 January 2010 and 31 December 2019, the associated monetary loss (**loss**), its starting

²Available at <https://www.spc.noaa.gov/wcm/#data>

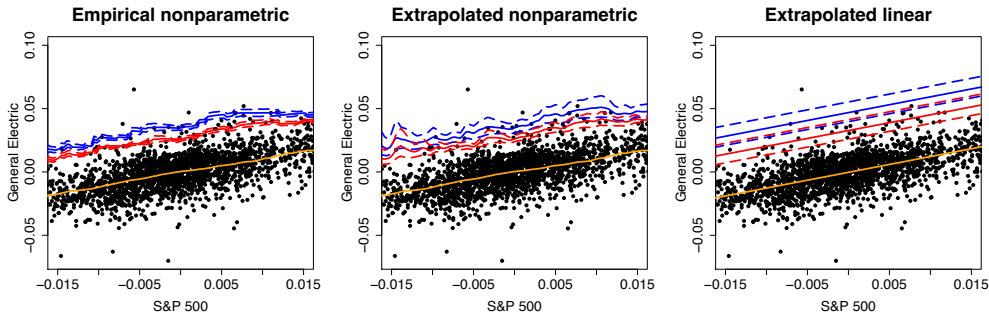


Figure 1: Stock market data. Blue line: Estimated conditional quantile at level $\tau'_n = 0.995$ (with 95% confidence interval in dashed line), red line: Estimated conditional expectile at level $\tau'_n = 0.995$ (with 95% confidence interval in dashed line). Left panel: Nonparametric empirical estimates, middle panel: Extrapolated nonparametric estimates, right panel: Extrapolated estimates based on the linear regression model. The orange line is the Nadaraya-Watson estimate in the left and middle panels, and the ordinary least squares line in the right panel.

and ending latitude and longitude (`slat`, `slon`, `elat` and `elon`), and the length and width of the area traveled over by the tornado (`len` and `wid`). We focus on the loss per surface unit Y (in USD) in terms of the tornado's (average) geographical location $\mathbf{X} = (X_1, X_2)$, that is, $Y = \text{loss}/(\text{len} \times \text{wid})$, $X_1 = (\text{slon} + \text{elon})/2$, and $X_2 = (\text{slat} + \text{elat})/2$. This results in a sample (\mathbf{X}_t, Y_t) of size $n = 6,360$ across the whole of the US (excluding Alaska and Hawaii), including the major Joplin, Missouri, tornado which caused a total loss of 2.8 billion USD on 22 May 2011. A fine analysis of the geographic variability of damages due to infrequent, high-impact tornadoes is of utmost interest to the insurance industry and government agencies. We conduct such an analysis in American cities located east of the 100th meridian west, due to sparsity of recorded tornadoes to the west of this geographical limit. The data, over the studied area, is represented in Figure 2 (a). The heavy tail model assumption was checked using local Generalized Pareto QQ-plots (see pp. 90-91 in de Haan and Ferreira, 2006) omitted for the sake of brevity. Tail index estimates were found to exceed 1 in many locations \mathbf{x} , which prevents the use of the expectile risk measure.

Our target is the extremal conditional risk measure $q(\tau'_n|\mathbf{x})$ at each location \mathbf{x} , where $\tau'_n = 0.995$ corresponds to a catastrophic loss exceeded (on average) only once every 200 cases at the location of interest. This is a reasonable choice in this setup given that, for instance, the state of Florida has recorded around 200 tornado events over the 10-year period we examine. A major hurdle to address this goal in this two-dimensional setting is the practical calculation of the optimal bandwidth $h_{n,\star}$. The usual rule-of-thumb calculations for the evaluation of $\int_{\mathbb{R}^2} (\mathbf{u}^\top H_g(\mathbf{x})\mathbf{u}) K(\mathbf{u}) d\mathbf{u}$, involved in (5.1), based on a bivariate Gaussian assumption lead to tedious calculations and unappealing results, and a uniform distribution over the covariate space cannot be chosen as it would have an identically zero Hessian matrix. Instead, the crude observation that the latitude $X_2 \in [18, 49]$ in the data concentrates around its median while the longitude $X_1 \in [-100, -66.5]$ appears more uniformly scattered suggests, for the specific purpose of calculating $h_{n,\star}$ only, to make the simplifying assumption that

$$g(x_1, x_2) \propto \mathbb{1}_{[-100, -66.5]}(x_1) \times (49 - x_2)(x_2 - 18) \mathbb{1}_{[18, 49]}(x_2).$$

This leads to a diagonal nonzero Hessian matrix $H_g(\mathbf{x})$. Letting $\kappa(u) = (15/16)(1 - u^2)^2$, for $u \in [-1, 1]$, be the one-dimensional quartic kernel and $K(\mathbf{u}) = (16/(5\pi))\kappa(\|\mathbf{u}\|)$ be

its isotropic version on \mathbb{R}^2 , leads to the global spatial bandwidth $h_{n,\star} \approx 5.47$. Then, for each geographic location \mathbf{x} , representing one of the 21,935 cities located east of the 100th meridian picked in the United States Cities Database³, we chose the corresponding local optimal hyperparameter $k_{n,\star} = \widehat{k}_{n,\star}(\mathbf{x})$ as described above in Section 5.1, with the only difference that we set $\bar{\rho}(\mathbf{x}) \equiv -1$ and $\bar{b}(\mathbf{x}) \equiv 1$. This is motivated by an inspection of preliminary local estimates, which seem to vary around these values; setting constant estimates $\bar{\rho}(\mathbf{x})$ and $\bar{b}(\mathbf{x})$ also produces smoother fits. Each city \mathbf{x} has in its $h_{n,\star}$ -vicinity 1,100 observations on average, with 90% of locations reporting at least 400 observations. Figure 2 (b) displays the historical frequency of tornadoes, showing that the area most often hit mainly comprises the states of Alabama, Mississippi, Louisiana, Arkansas, Missouri, Kentucky, and Tennessee.

The estimated conditional mean of losses per squared yard (Nadaraya-Watson estimates) and extrapolated bias-reduced conditional quantile estimates are shown in Figures 2 (c) and (d), respectively. The first conclusion is that the area most exposed to tornadoes is actually not the riskiest in terms of average and/or extremal conditional losses per tornado. By contrast, tornadoes in the states of Florida, Texas, Oklahoma, Nebraska, South Dakota, Iowa and their surroundings are found to carry the most extreme risk, with a 99.5%-regression Value-at-Risk exceeding 80 USD per squared yard, even though the frequency of tornadoes there is substantially lower. The large difference in order of magnitude between the regression mean and extremal quantile reflects the great variability and tail heaviness of the conditional loss distribution; an important benefit of the nonparametric approach is its ability to accurately identify conditional extreme value behavior, without recourse to any strong *a priori* spatio-temporal model specification. According to the results obtained at 9 selected cities, reported in Table 1, the conditional tail index varies with \mathbf{x} , with a minimum of 0.79 achieved at Harrisville, MI, as opposed to 1.53 in Woodson, TX, which is the location with maximal estimated extreme quantile risk. It should also be noted that, among others, Charleston, SC and New Orleans, LA have very similar tail index estimates but completely different tail quantile estimates, owing to a strong geographical heterogeneity in the scale parameter of the loss distribution. Most importantly, the confidence intervals $\widehat{I}_{q,4}(0.995|\mathbf{x})$ based on purely empirical estimates are much narrower than our corrected proposal $\widehat{I}_{q,2}(0.995|\mathbf{x})$ based on extrapolated quantile estimates, which underestimate the uncertainty of high regression quantiles.

Our results are compared with the straightforward unconditional approach estimating, in each state, the mean and Value-at-Risk at level 99.5% from the univariate sample of losses in this state only, see Figures 2 (e) and (f). This approach yields quantile point estimates in the riskiest areas that are up to 67% lower than those of the regression method, and produces unrealistic discontinuities in the estimates, see the examples of Texas-Oklahoma-Kansas, Nebraska-South Dakota and Georgia-South Carolina, while the local nature of our proposed approach eliminates these unpleasant discontinuities between neighboring states by combining regional tail information. A similar countywide analysis in the spirit of risk assessment exercises such as those reported by the US Federal Emergency Management Agency (FEMA)⁴ cannot be carried out here, because certain counties did not report any observation.

³Available at <https://simplemaps.com/data/us-cities>

⁴Available at <https://hazards.fema.gov/nri/tornado>

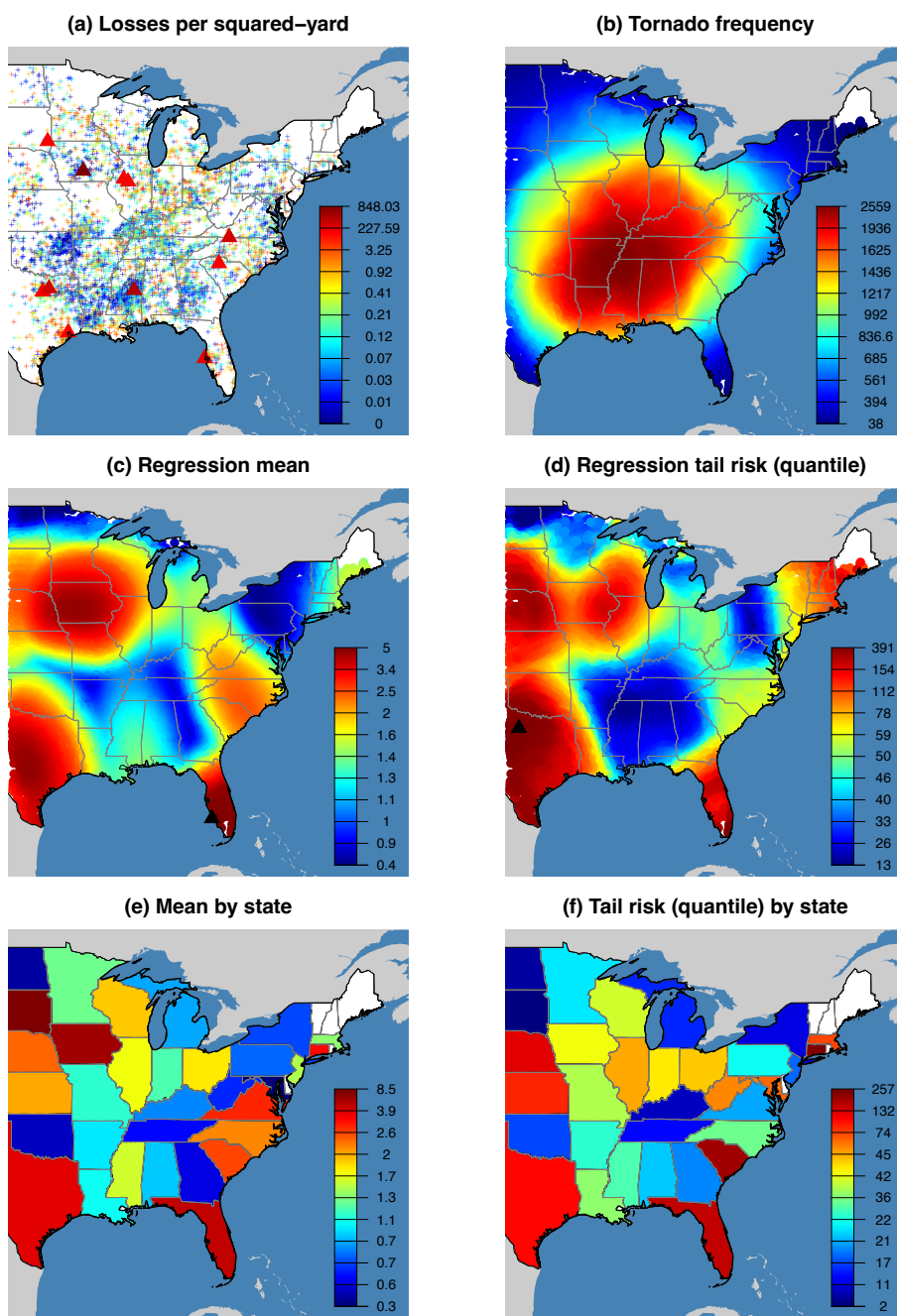


Figure 2: Tornado losses data. Top row, left: Data across the eastern half of the US (the data across the western half is not represented for consistency with the other plots), right: Local number of observations in the $h_{n,\star}$ -ball. Middle row, left: Estimated conditional mean of losses per squared yard, right: Extrapolated conditional quantile estimate of those losses at level $\tau'_n = 0.995$. Cities with the highest estimated conditional average loss and extreme loss are marked with a black triangle in the left and right panels, respectively. Bottom row, left: Unconditional statewide estimation, using the sample average, right: Using the bias-reduced extreme quantile estimator of Gomes and Pestana (2007). Losses (all panels except (b), in USD) and tornado frequency (panel (b)) are indicated by a color scheme, ranging from dark blue (lowest) to dark red (highest).

Location \mathbf{x} (State)	$N_{h_{n,*}}(\mathbf{x})$	$\hat{\gamma}_{1-k_{n,*}/n}^{(J, BR)}(\mathbf{x})$	$\hat{q}_n(0.995 \mathbf{x})$ $\hat{I}_{q,4}(0.995 \mathbf{x})$	$\hat{q}_{n,1-k_{n,*}/n}^{W, BR}(0.995 \mathbf{x})$ $\hat{I}_{q,2}(0.995 \mathbf{x})$
New York (NY)	413	1.11	16.71 [11.99, 23.29]	63.53 [18.13, 222.70]
Charleston (SC)	1,019	1.01	340.91 [271.17, 428.59]	259.92 [80.65, 837.72]
Nashville (TN)	2,317	0.95	16.57 [14.10, 19.47]	24.51 [14.62, 41.08]
Captiva (FL)	205	0.93	236.74 [173.66, 322.73]	144.67 [43.50, 481.13]
New Orleans (LA)	1,427	0.98	27.46 [22.69, 33.22]	31.01 [16.42, 58.58]
Woodson (TX)	958	1.53	118.37 [81.12, 172.72]	390.59 [100.95, 1511.26]
Kansas City (MO)	1,326	1.30	45.99 [34.53, 61.27]	69.00 [25.57, 186.21]
Minneapolis (MN)	620	0.93	34.09 [26.82, 43.33]	40.49 [17.01, 96.40]
Harrisville (MI)	472	0.79	24.86 [19.00, 32.52]	29.32 [10.28, 83.67]

Table 1: Tornado losses data. Results at selected cities (first column), with the number of neighboring observations (second column), conditional tail index estimate (third column), empirical conditional quantile estimate at level 0.995 (fourth column) and extrapolated bias-reduced conditional quantile estimate at the same level (fifth column), along with the 95% asymptotic confidence interval corresponding to each quantile estimate in brackets. Captiva, FL is the city with maximal estimated average loss; Woodson, TX is the city with maximal extrapolated conditional quantile estimate; Harrisville, MI is the city with minimal estimated conditional tail index.

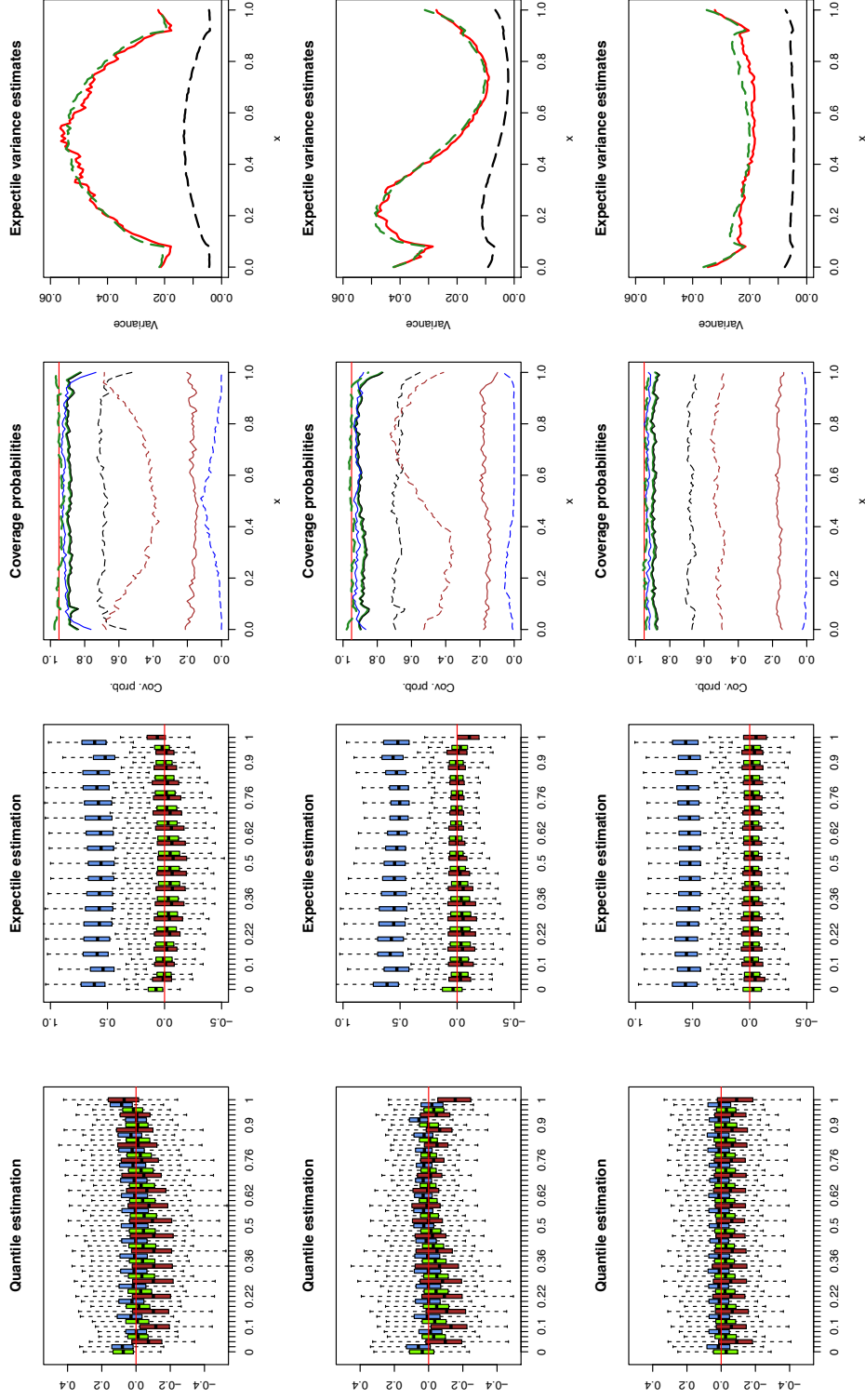


Figure 3: Simulation results. From left to right, boxplots of $\log(\hat{q}_{1-k_{n,*}/n}^{W,BR}(\tau'_n|x)/q(\tau'_n|x))$ (green), $\log(\hat{q}_{1-k_{n,*}/n}^W(\tau'_n|x)/q(\tau'_n|x))$ (blue) and $\log(\hat{q}_n(\tau'_n|x)/q(\tau'_n|x))$ (brown); boxplots of $\log(\hat{e}_{1-k_{n,*}/n}^{W,BR}(\tau'_n|x)/e(\tau'_n|x))$ (green), $\log(\hat{e}_{1-k_{n,*}/n}^W(\tau'_n|x)/e(\tau'_n|x))$ (blue) and $\log(\hat{e}_n(\tau'_n|x)/e(\tau'_n|x))$ (brown); empirical pointwise coverage probabilities of the asymptotic 95% confidence intervals $\hat{I}_{q,1}(\tau'_n|x)$ (full black line), $\hat{I}_{q,2}(\tau'_n|x)$ (full green line), $\hat{I}_{q,3}(\tau'_n|x)$ (full blue line), $\hat{I}_{q,4}(\tau'_n|x)$ (full brown line), $\hat{I}_{E,1}(\tau'_n|x)$ (dashed black line), $\hat{I}_{E,2}(\tau'_n|x)$ (dashed green line), $\hat{I}_{E,3}(\tau'_n|x)$ (dashed blue line) and $\hat{I}_{E,4}(\tau'_n|x)$ (dashed brown line), with the target 95% nominal level in full red line; comparison of the pointwise empirical variances of the $(\sqrt{k_{n,*}}h_{n,*}/\log(k_{n,*}/(n(1-\tau'_n)))) \log(\hat{e}_{1-k_{n,*}/n}^{W,BR}(\tau'_n|x)/e(\tau'_n|x))$ (full red line), the pointwise median of the corrected variance estimates $(\int_{\mathbb{R}^p} K^2/\hat{g}_n(x)) \tilde{v}_E(x)$ (dashed green line) and the pointwise median of the naive variance estimates $(\int_{\mathbb{R}^p} K^2/\hat{g}_n(x)) \hat{v}_E(x)$ (dashed black line). Top row: model (NL-P), middle row: model (NL-S), bottom row: model (NL-C), all with $n = 10,000$ and estimates at level $\tau'_n = 0.999$.

Acknowledgments

This research was supported by the French National Research Agency under the grants ANR-19-CE40-0013 and ANR-17-EURE-0010. G. Stupfler also acknowledges support from an AXA Research Fund Award on “Mitigating risk in the wake of the COVID-19 pandemic”.

References

- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.
- Bellini, F. and Di Bernardino, E. (2017). Risk management with expectiles. *The European Journal of Finance*, 23(6):487–506.
- Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular Variation*. Cambridge University Press.
- Bosq, D. (1998). *Nonparametric statistics for stochastic processes: estimation and prediction*, volume 110. Springer Science & Business Media.
- Bradley, R. C. (1981). Central limit theorems under weak dependence. *Journal of Multivariate Analysis*, 11(1):1–16.
- Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2:107–144.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods (Second Edition)*. Springer.
- Bücher, A. and Ruppert, M. (2013). Consistent testing for a constant copula under strong mixing based on the tapered block multiplier technique. *Journal of Multivariate Analysis*, 116:208–229.
- Chaudhuri, P. (1991). Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis*, 39(2):246–269.
- Chen, J., Li, D., Linton, O., and Lu, Z. (2018). Semiparametric ultra-high dimensional model averaging of nonlinear dynamic time series. *Journal of the American Statistical Association*, 113(522):919–932.
- Chen, X., Linton, O., and Robinson, P. M. (2000). The estimation of conditional densities. In Puri, M. L., editor, *Asymptotics in Statistics and Probability: Papers in Honor of George Gregory Roussas*, pages 71–84. VSP International Science Publishers, Netherlands.
- Chernick, M. R. (1981). A limit theorem for the maximum of autoregressive processes with uniform marginal distributions. *Annals of Probability*, 9(1):145–149.
- Chernozhukov, V. and Fernández-Val, I. (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *Review of Economic Studies*, 78(2):559–589.

- Daouia, A., Gardes, L., and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19(5B):2557–2589.
- Daouia, A., Gardes, L., Girard, S., and Lekina, A. (2011). Kernel estimators of extreme level curves. *TEST*, 20(2):311–333.
- Daouia, A., Girard, S., and Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society: Series B*, 80(2):263–292.
- Daouia, A., Girard, S., and Stupfler, G. (2020). Tail expectile process and risk assessment. *Bernoulli*, 26(1):531–556.
- Davison, A., Padoan, S., and Stupfler, G. (2021). Extreme expectile estimation for heavy-tailed time series. Available at <https://arxiv.org/abs/2004.04078>.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer-Verlag, New York.
- Doukhan, P. (1994). *Mixing: Properties and Examples*. Springer, New York.
- Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli*, 9(4):617–657.
- Ehm, W., Gneiting, T., Jordan, A., and Krüger, F. (2016). Of quantiles and expectiles: consistent scoring functions, choquet representations and forecast rankings. *Journal of the Royal Statistical Society: Series B*, 78:505–562.
- El Methni, J., Gardes, L., and Girard, S. (2014). Non-parametric estimation of extreme risk measures from conditional heavy-tailed distributions. *Scandinavian Journal of Statistics*, 41(4):988–1012.
- Fraga Alves, M., Gomes, M. I., and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, 60(2):193–214.
- Francq, C. and Zakoïan, J.-M. (2019). *GARCH Models (Second Edition)*. John Wiley & Sons, Hoboken, NJ.
- Gardes, L. and Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes*, 17(1):45–75.
- Gardes, L. and Stupfler, G. (2019). An integrated functional Weissman estimator for conditional extreme quantiles. *REVSTAT: Statistical Journal*, 17(1):109–144.
- Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2021a). Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. *Annals of Statistics*, 49(6):3358–3382.
- Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2021b). On automatic bias reduction for extreme expectile estimation. Available at <https://hal.archives-ouvertes.fr/hal-03086048>.
- Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2022). Nonparametric extreme conditional expectile estimation. *Scandinavian Journal of Statistics*, 49(1):78–115.

- Gomes, M. I., Brillhante, M. F., and Pestana, D. (2016). New reduced-bias estimators of a positive extreme value index. *Communications in Statistics-Simulation and Computation*, 45(3):833–862.
- Gomes, M. I. and Martins, M. J. (2002). “Asymptotically unbiased” estimators of the tail index based on external estimation of the second order parameter. *Extremes*, 5(1):5–31.
- Gomes, M. I. and Pestana, D. (2007). A sturdy reduced-bias extreme quantile (var) estimator. *Journal of the American Statistical Association*, 102(477):280–292.
- Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B*, 44(1):37–42.
- Ibragimov, I. (1962). Some limit theorems for stationary processes. *Theory of Probability & Its Applications*, 7(4):349–382.
- Jones, M. (1994). Expectiles and M-quantiles are quantiles. *Statistics & Probability Letters*, 20:149–153.
- Linton, O. and Xiao, Z. (2013). Estimation of and inference about the Expected Shortfall for time series with infinite variance. *Econometric Theory*, 29(4):771–807.
- 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.
- Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics*, 24(2):165–179.
- Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of non-linear ARCH time series: Strong convergence and asymptotic normality. *Econometric Theory*, 11(2):258–289.
- McMurry, T. L. and Politis, D. N. (2008). Minimally biased nonparametric regression and autoregression. *REVSTAT - Statistical Journal*, 6(2):123–150.
- Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847.
- Padoan, S. A. and Stupfler, G. (2022). Joint inference on extreme expectiles for multivariate heavy-tailed distributions. *Bernoulli*, 28:1021–1048.
- Philipp, W. and Webb, G. R. (1973). An invariance principle for mixing sequences of random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 25(3):223–237.
- Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer.
- Rio, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes*. In *Probability Theory and Stochastic Modelling*, volume 80, Springer-Verlag, Berlin-Heidelberg.
- Rootzén, H. (2009). Weak convergence of the tail empirical process for dependent sequences. *Stochastic Processes and their Applications*, 119(2):468–490.

- Roussas, G. G. (1990). Nonparametric regression estimation under mixing conditions. *Stochastic Processes and Their Applications*, 36(1):107–116.
- Stupfler, G. (2013). A moment estimator for the conditional extreme-value index. *Electronic Journal of Statistics*, 7:2298–2343.
- Stupfler, G. (2016). Estimating the conditional extreme-value index under random right-censoring. *Journal of Multivariate Analysis*, 144:1–24.
- Stupfler, G. and Usseglio-Carleve, A. (2021). Composite bias-reduced L^p -quantile-based estimators of extreme quantiles and expectiles. *Canadian Journal of Statistics*, to appear, available at <https://hal.archives-ouvertes.fr/hal-03197015>.
- Usseglio-Carleve, A. (2018). Estimation of conditional extreme risk measures from heavy-tailed elliptical random vectors. *Electronic Journal of Statistics*, 12(2):4057–4093.
- Utev, S. A. (1990). On the central limit theorem for φ -mixing arrays of random variables. *Theory of Probability and its Applications*, 35(1):131–139.
- Vilar-Fernández, J. A. and Vilar-Fernández, J. M. (1998). Recursive estimation of regression functions by local polynomial fitting. *Annals of the Institute of Statistical Mathematics*, 50(4):729–754.
- Wang, H. J., Li, D., and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, 107(500):1453–1464.
- Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer.
- 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.
- Ziegel, J. F. (2016). Coherence and elicibility. *Mathematical Finance*, 26(4):901–918.