Inference for extremal regression with dependent heavy-tailed data

Abdelaati Daouia, Gilles Claude Stupfler, Antoine Usseglio-Carleve

WORKING PAPERS

N° 1324

August 2023

"Inference for extremal regression with dependent heavy-tailed data"

Abdelaati Daouia, Gilles Stupfler and Antoine Usseglio-Carleve

Toulouse School of Economics

INFERENCE FOR EXTREMAL REGRESSION WITH DEPENDENT HEAVY-TAILED DATA

BY ABDELAATI DAOUIA[1,a], GILLES STUPFLER[2,b] AND ANTOINE USSEGLIO-CARLEVE[3,c]

[1]*Toulouse School of Economics, University of Toulouse Capitole, France,* [a]*abdelaati.daouia@tse-fr.eu*

[2]*Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France,* [b]*gilles.stupfler@univ-angers.fr*

[3]*Avignon Université, LMA UPR 2151, 84000 Avignon, France,* [c]*antoine.usseglio-carleve@univ-avignon.fr*

Nonparametric inference on tail conditional quantiles and their least squares analogs, expectiles, remains limited to i.i.d. data. We develop a fully operational inferential theory for extreme conditional quantiles and expectiles in the challenging framework of $\alpha-$mixing, conditional heavy-tailed data whose tail index may vary with covariate values. This 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. Our assumptions hold in a wide range of models. 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.

1. Introduction.

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

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

where $\varrho_\tau(y) = |\tau - \mathbb{1}_{\{y\leq 0\}}|\,|y|$ denotes the quantile check function. Regression quantile estimators at the tails nonetheless 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, among others, insurance (large losses due to large claims), finance (large drops in stock indices) and natural sciences (large earthquake magnitudes, flood intensity, extreme rainfall). Existing approaches to extremal quantile regression in the heavy-tailed case fall into two main categories: linear quantile regression approaches, such as those of [6], [7] and [25], and at the opposite, nonparametric approaches such as those of [9] and [10]. Yet, tail quantile regression theory typically assumes that the data is independent and identically distributed (i.i.d.). Only [7] provides feasible inference tools for $\alpha-$mixing data, by assuming that conditional quantiles $q(\tau|\boldsymbol{x})$ are linear in $\boldsymbol{x}$, that the extreme value index is constant, and using self-normalized quantile regression statistics rather

MSC 2010 subject classifications: Primary 62G32; secondary 62G05, 62G08, 62G15, 62G20, 62G30.

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

than simpler asymptotic Gaussian confidence intervals. More broadly, the problem of inference on nonlinear extremal quantile regression remains untouched under serial dependence.

In actuarial and financial risk management, the robustness of quantiles may constitute a weakness as they are not sensitive to the severity of extreme losses. Their failure to satisfy the coherence property (introduced in [1]) is also a serious drawback. A better alternative in these respects is expectile regression [22], which focuses on the conditional expectiles

$$e(\tau|\boldsymbol{x}) = \underset{\theta\in\mathbb{R}}{\arg\min}\,\mathbb{E}\left([\eta_\tau(Y-\theta)-\eta_\tau(Y)]\,|\boldsymbol{X}=\boldsymbol{x}\right), \quad \tau\in(0,1), \tag{2}$$

where $\eta_\tau(y) = |\tau - \mathbb{1}_{\{y\leq 0\}}|\, y^2$ is an asymmetric quadratic loss function. Expectiles extend the mean as quantiles extend the median, and induce the only coherent and elicitable risk measure [29]. As such, they come endowed with a natural backtesting methodology. Besides, the τth expectile is in fact the τth quantile of a transformed distribution function, that is,

$$e(\tau|\boldsymbol{x}) = \inf\{y\in\mathbb{R}\,|\,E(y|\boldsymbol{x})\geq\tau\} \text{ with } E(y|\boldsymbol{x}) = \frac{\mathbb{E}[|Y-y|\mathbb{1}_{\{Y\leq y\}}|\boldsymbol{X}=\boldsymbol{x}]}{\mathbb{E}[|Y-y||\boldsymbol{X}=\boldsymbol{x}]}. \tag{3}$$

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 [3]. These properties and others have motivated the development of extremal expectile regression. The pioneering contribution of [24] is limited to elliptical heavy-tailed distributions, and the nonparametric approach of [17] hinges on the i.i.d. assumption. The approach of [16] can handle time series location-scale models, but is highly sensitive to model misspecification, makes the strong assumption of a constant tail index, and its 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 when the data comes from an $\alpha-$mixing dependent sequence $((\boldsymbol{X}_t, Y_t))_{t\geq 1}$. We allow heavy-tailed data for both regression modes, although moment restrictions are inevitably required for conditional extreme expectile inference, since expectiles extend the mean. The conditional tail index is allowed to depend on covariate values. Our unifying argument is that any quantile can be estimated by inverting an estimator of the associated distribution function. This is very beneficial theoretically, as it reduces the problem of investigating the asymptotic normality of extreme conditional quantile and expectile estimators to proving the asymptotic normality of row sums of identically distributed and dependent random variables, while a solution via empirical versions of (1) and (2) would be technically involved (see [13] and [15] in the unconditional setting and under a stronger $\beta-$mixing assumption). When the level τ is intermediate in the sense that $\tau = \tau_n \uparrow 1$ slowly enough as the sample size $n\to\infty$, 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 $nh_n^p(1-\tau_n)\to\infty$, where $h_n>0$ is the bandwidth sequence featuring in the kernel estimator. At properly extreme levels $\tau=\tau_n'$ such that $nh_n^p(1-\tau_n')=O(1)$, we consider Weissman-type estimators (after [27]) that rely on a conditional tail index estimator based on quantiles or expectiles.

Our asymptotic theory is obtained through a standard "big blocks separated by small blocks" argument, for algebraically fast mixing and under reasonable technical conditions on the distributional behavior of $\boldsymbol{X}$ and of Y given $\boldsymbol{X}$, marginally and through time; in particular, we make conditions on $(\boldsymbol{X}_1, \boldsymbol{X}_{t+1})$ and on the joint extreme value behavior of (Y_1, Y_{t+1}) given $(\boldsymbol{X}_1, \boldsymbol{X}_{t+1})$ that are weaker than typical conditions in the literature. Our framework encompasses, among others, location-scale models with possible temporal misspecification, a general class of nonlinear regression models containing Generalized Linear Models, and

autoregressive models. Surprisingly perhaps, the asymptotic distribution of the intermediate estimators remains the same as in the nonparametric i.i.d. setting [10, 17], unlike in the absence of covariates [13, 15]. We nonetheless improve upon results of the i.i.d. literature by deriving the asymptotic normality of our estimators at faster rates of convergence than had been found so far. Obtaining these rates, reminiscent of the optimal rates found in classical nonparametric statistics under twice differentiability of the function to be estimated, requires developing an innovative and careful approach to the quantification of bias in smoothed extremal regression estimators. In particular, we find that the variation in conditional extremes induces bias that bears a link to the *design bias* in the terminology of [26]; this had not been appreciated in the literature before. From the inferential point of view, the Weissman-type structure makes it possible to come up with very accurate and novel bias-corrected versions and precise approximations to the empirical variance of the estimators. We thus construct asymptotic Gaussian confidence intervals that substantially improve upon the naive solutions existing in the nonparametric extreme value regression literature, in which the problem of accurate Gaussian inference had been mostly ignored. The method is applied here to inference about extreme conditional quantiles and expectiles, but it can handle any kind of Weissman-type estimator: the bias correction methodology revolves upon identifying bias sources due to the extrapolation procedure, while the variance correction relies on a precise evaluation of the correlation between intermediate estimators and tail index estimators. We propose rules-of-thumb for the choices of the bandwidth h_n and of the sample fraction $1-\tau_n$ needed for tail extrapolation, resulting in confidence intervals achieving excellent coverage already for moderately large sample sizes.

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]. Further details about our technical conditions, the proofs of all theoretical results and extra finite-sample results are deferred to an online Supplementary Material document [12]. 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 $\mathbb{R}^p$, $\nabla f(\boldsymbol{x})$, $Jf(\boldsymbol{x})$ and $Hf(\boldsymbol{x})$ stand respectively for its gradient vector, Jacobian matrix, and Hessian matrix at the point $\boldsymbol{x}$. For a function $f = f(\boldsymbol{x},\boldsymbol{y})$ on $\mathbb{R}^p \times \mathbb{R}^q$, let $\nabla_{\boldsymbol{x}} f$ and $H_{\boldsymbol{x}} f$ be its partial gradient vector and Hessian matrix with respect to $\boldsymbol{x}$ (*i.e.* the first p components of its gradient vector and the submatrix made of the first p rows and columns of its Hessian matrix, respectively). 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.* Let $((\boldsymbol{X}_t, Y_t))_{t\geq 1}$ be a strictly stationary sequence of copies of $(\boldsymbol{X}, Y) \in \mathbb{R}^p \times \mathbb{R}$. Let $F(\cdot|\boldsymbol{x})$ denote the distribution function of Y given $\boldsymbol{X} = \boldsymbol{x}$, that is, $F(y|\boldsymbol{x}) = \mathbb{P}(Y \leq y | \boldsymbol{X} = \boldsymbol{x})$. Assume that $\boldsymbol{X}$ has a probability density function (p.d.f.) g on $\mathbb{R}^p$ and fix $\boldsymbol{x} \in \mathbb{R}^p$ with $g(\boldsymbol{x}) > 0$. Consider the following kernel estimator of $F(\cdot|\boldsymbol{x})$:

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

[1] Available on GitHub at `https://github.com/AntoineUC/Expectrem`

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

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

The standard roadmap for estimating extreme conditional quantiles at a level $\tau = \tau_n \to 1$ as $n \to \infty$ is to consider first intermediate ("extreme, but not too much") levels τ_n, for which $\widehat{q}_n(\tau_n|\boldsymbol{x})$ is a (relatively) consistent estimator of $q(\tau_n|\boldsymbol{x})$. Then, for the estimation of properly extreme quantiles $q(\tau'_n|\boldsymbol{x})$ without any restriction on τ'_n, one extrapolates such intermediate estimators using the shape of the tail of the underlying conditional distribution.

In order to assess the asymptotic behavior of this procedure, we introduce the following conditional, second-order regularly varying tails assumption about Y given $\boldsymbol{X} = \boldsymbol{x}$:

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

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

This standard condition, wherein $A(\cdot|\boldsymbol{x})$ is regularly varying with index $\rho(\boldsymbol{x})$ [see 14, 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 $\gamma(\boldsymbol{x})$. We therefore make the fundamental modeling assumption that

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

The $\alpha-$mixing (or strong mixing) assumption we shall make is expressed as follows: let, for any two positive integers $a \leq b$, $\mathcal{F}_a^b = \sigma(\{(\boldsymbol{X}_j, Y_j), a \leq j \leq b\})$ be the $\sigma-$algebra generated by $\{(\boldsymbol{X}_j, Y_j), a \leq j \leq b\}$, and say that $((\boldsymbol{X}_t, Y_t))_{t\geq 1}$ is $\alpha-$mixing if and only if

$$\alpha(n) = \sup_{k\geq 1} \sup_{A\in\mathcal{F}_1^k} \sup_{B\in\mathcal{F}_{k+n}^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \to 0 \text{ as } n \to \infty.$$

More restrictive assumptions that will be employed below to relax certain regularity conditions involve the $\beta-$, $\rho-$, $\phi-$ and $\psi-$mixing coefficients, see Section A.1 of the Supplementary Material document [12] for definitions and relationships between the different kinds of mixing. We introduce the following condition used to develop a "big blocks separated by small blocks" argument for the theory:

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

We also require reasonable regularity assumptions on the kernel function K and the probabilistic behavior of the covariate sequence $(\boldsymbol{X}_t)_{t\geq 1}$. Let $\|\cdot\|$ denote the Euclidean norm on $\mathbb{R}^p$ and $B(\boldsymbol{x}, r)$ be the open $\|\cdot\|-$ball with center $\boldsymbol{x}$ and radius $r > 0$. Full details for the rationale behind our conditions and their interpretation are given in Section A.2 of the Supplementary Material document [12].

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

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

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

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

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

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 $\widehat{g}_n(\boldsymbol{x})$. Assumption $\mathcal{B}_p$ will be satisfied especially if, for all $t \geq 1$, the pair $(\boldsymbol{X}_1, \boldsymbol{X}_{t+1})$ has a p.d.f. g_t such that $\sup_{t\geq 1} g_t$ is bounded on $B(\boldsymbol{x},r) \times B(\boldsymbol{x},r)$ for some $r > 0$. Under condition $\mathcal{L}_g$, this local boundedness condition is automatically true if $(\boldsymbol{X}_t)$ is $\beta-$mixing. When $p \geq 2$, the causal and invertible AR(p) process does not satisfy this boundedness condition, but does satisfy assumption $\mathcal{B}_p$.

To control the variation in conditional extreme value behavior across the covariate space, we make a Lipschitz assumption on the log-conditional survival function at extreme levels.

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

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

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

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

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

This should be considered as a weak requirement compared with the existence of a (conditional) tail copula, as assumed in *e.g.* [13, 15] in the unconditional setting. It is an important ingredient in the quantification of the correlation between tail empirical moments.

These 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 better rates of convergence requires, similarly to classical nonparametric estimation, stronger regularity conditions: it is well-known that the optimal rate of convergence $n^{-2/(p+4)}$ of $\widehat{g}_n(\boldsymbol{x})$ to $g(\boldsymbol{x})$ is obtained by solving the bias-variance tradeoff if K is symmetric and g is twice differentiable at $\boldsymbol{x}$. This motivates the following additional assumptions.

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

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

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

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

The motivation for the assumption that $\nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})/\log(y)$ converges as $y \to \infty$ is the fact that, in the setup of conditional heavy tails, $\log \overline{F}(y|\boldsymbol{x})/\log(y) = -1/\gamma(\boldsymbol{x}) + \log L(y|\boldsymbol{x})/\log(y)$, where $L(\cdot|\boldsymbol{x})$ is slowly varying. In particular, $\log L(y|\boldsymbol{x})/\log(y) \to 0$

as $y \to \infty$ [see 14, Proposition B.1.9.1 p.36]. The assumption asks for this convergence to hold also when taking the gradient with respect to $\boldsymbol{x}$, *i.e.* the function $L(\cdot|\boldsymbol{x})$ should not vary too wildly in $\boldsymbol{x}$ when y is large. The finite limit of $\nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})/\log(y)$ as $y \to \infty$ will then be $\boldsymbol{\mu}(\boldsymbol{x}) = \nabla\gamma(\boldsymbol{x})/\gamma^2(\boldsymbol{x}) \in \mathbb{R}^p$. In summary, while condition $\mathcal{L}_\omega$ is a Lipschitz condition on the log-conditional survival function at extreme levels, condition $\mathcal{D}_\omega$ is essentially an appropriate analog about its gradient with respect to $\boldsymbol{x}$.

2.2. *Intermediate quantile estimation.* Let $\tau = \tau_n \uparrow 1$ as $n \to \infty$. We show that $\widehat{q}_n(\tau_n|\boldsymbol{x})$ is asymptotically normal if $nh_n^p(1-\tau_n) \to \infty$, *i.e.* $\tau_n \uparrow 1$ slowly enough. Hereafter, under condition $\mathcal{KS}$ and $\mathcal{D}_\omega$, we let $\Upsilon_K : (\Delta, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}^p \mapsto \frac{\gamma(\boldsymbol{x})\Delta}{2} \int_{\mathbb{R}^p} K(\boldsymbol{u})(\boldsymbol{u}^\top \boldsymbol{\mu}(\boldsymbol{x}))^2 \, d\boldsymbol{u}$.

THEOREM 2.1. *Assume that conditions $\mathcal{M}$, $\mathcal{A}(l_n, r_n)$, $\mathcal{K}$, $\mathcal{L}_g$, $\mathcal{L}_\omega$, $\mathcal{B}_p$ and $\mathcal{B}_\Omega$ 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 $n \to \infty$ (for $1 \leq j \leq J$). Assume further that $h_n \to 0$ is such that $nh_n^p(1-\tau_n) \to \infty$, $nh_n^{p+2}(1-\tau_n)\log^2(1-\tau_n) \to 0$ and $\sqrt{nh_n^p(1-\tau_n)}A((1-\tau_n)^{-1}|\boldsymbol{x}) = O(1)$, and that there is $\delta > 0$ such that $r_n(r_n/\sqrt{nh_n^p(1-\tau_n)})^\delta \to 0$. Then*

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

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

If $((\boldsymbol{X}_t, Y_t))_{t \geq 1}$ is $\rho-$mixing, condition $\sum_{j=1}^\infty j^\eta \alpha(j) < \infty$ can be replaced by summability of the $\rho-$mixing coefficient series. If $((\boldsymbol{X}_t, Y_t))_{t \geq 1}$ is also $\psi-$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 summability of $\rho-$mixing coefficients), all conditions on (l_n) and (r_n) (including $\mathcal{A}(l_n, r_n)$) as well as $\mathcal{B}_p$ and $\mathcal{B}_\Omega$ can also be dropped.

It is remarkable that the asymptotic distribution in Theorem 2.1 is exactly the one obtained in the i.i.d. setting by [10] under an unnecessary regularity assumption on conditional tails. This is not true in the unconditional setting, see [15]. The essential difference is that the kernel estimator only takes into account those pairs $(\boldsymbol{X}_t, Y_t)$ such that $\boldsymbol{X}_t$ are close enough to $\boldsymbol{x}$, 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 [21, p.784].

Theorem 2.1 actually holds under weaker bias assumptions than those of [10]. Indeed, when $((\boldsymbol{X}_t, Y_t))_{t \geq 1}$ is geometrically $\alpha-$mixing, and if $A(t|\boldsymbol{x}) \propto t^{\rho(\boldsymbol{x})}$, as is the case in a wide range of heavy-tailed models used in practice [see *e.g.* 2, Table 2.1 p.59], the optimal rate of convergence is $n^{\rho(\boldsymbol{x})/(1-(p+2)\rho(\boldsymbol{x}))}$ when $\mathcal{K}$, $\mathcal{L}_g$ and $\mathcal{L}_\omega$ hold, while it is equal to $n^{2\rho(\boldsymbol{x})/(2-(p+4)\rho(\boldsymbol{x}))}$ if $\mathcal{KS}$, $\mathcal{D}_g$ and $\mathcal{D}_\omega$ are satisfied, see Section A.3 of the Supplementary Material document [12] for details. In the latter setting, note that $p = 0$ yields the optimal convergence rate $n^{\rho(\boldsymbol{x})/(1-2\rho(\boldsymbol{x}))}$ of unconditional extreme value estimators in heavy-tailed models [see 14, p.77], while the case $\rho(\boldsymbol{x}) \to -\infty$, corresponding to the ideal but unrealistic scenario where all the Y_t such that $\boldsymbol{X}_t \in B(\boldsymbol{x}, h_n)$ 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 [5]. Of course, the nonparametric nature of the methodology coupled with the double sparsity of relevant data (in $\boldsymbol{x}$, due to kernel smoothing, and in y, due to taking only extreme values of the response into account) limits

the applicability of the method to low dimensions. The nonparametric approach is, however, indifferent to the structure of the conditional distribution, unlike techniques adapted to taking large dimensions into account that rely on strong model specifications such as linear quantile structures, see for instance [6, 7, 25].

The bias component $\Upsilon_K(\Delta, \boldsymbol{x})$ appearing under conditions $\mathcal{D}_g$ and $\mathcal{D}_\omega$ 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 $\boldsymbol{\mu}(\boldsymbol{x}) = \nabla\gamma(\boldsymbol{x})/\gamma^2(\boldsymbol{x})$. A key factor in extremal regression is therefore the *variation in conditional extremes*, while in standard regression it is also important to account for *changes of shape* in the regression function through its second derivatives [see *e.g.* 26, p.73]. We finally note that $\gamma(\boldsymbol{x})\boldsymbol{\mu}(\boldsymbol{x}) = \nabla\gamma(\boldsymbol{x})/\gamma(\boldsymbol{x}) = \nabla(\log\gamma)(\boldsymbol{x})$ is reminiscent of the *design bias* for classical regression in the terminology of [26]: the higher and less variable γ around $\boldsymbol{x}$, 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 τ_n' such that $nh_n^p(1-\tau_n') \to c < \infty$. Then, in a neighborhood of $\boldsymbol{x}$, very few or no top observations Y_t will be close to the extreme value $q(\tau_n'|\boldsymbol{x})$, so the empirical estimator $\widehat{q}_n(\tau_n'|\boldsymbol{x})$ will not be consistent. However, the conditional heavy tail assumption suggests the extrapolation formula $q(\tau_n'|\boldsymbol{x}) \approx ((1-\tau_n')/(1-\tau_n))^{-\gamma(\boldsymbol{x})} q(\tau_n|\boldsymbol{x})$ for n large. Plugging in consistent estimators $\widehat{\gamma}(\boldsymbol{x})$ of $\gamma(\boldsymbol{x})$ and $\overline{q}_n(\tau_n|\boldsymbol{x})$ of $q(\tau_n|\boldsymbol{x})$ yields a conditional Weissman-type estimator (see [27]) of $q(\tau_n'|\boldsymbol{x})$:

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

We prove below that the choice of $\widehat{\gamma}(\boldsymbol{x})$ is crucial since $\widehat{q}^{\,W}_{n,\tau_n}(\tau_n'|\boldsymbol{x})$ inherits its asymptotics.

THEOREM 2.2. *Assume that condition $\mathcal{M}$ holds with $\rho(\boldsymbol{x}) < 0$. Let $\tau_n, \tau_n' \uparrow 1$ be such that $(1-\tau_n')/(1-\tau_n) \to 0$ and assume that $v_n(\overline{q}_n(\tau_n|\boldsymbol{x})/q(\tau_n|\boldsymbol{x}) - 1) = O_{\mathbb{P}}(1)$ and $v_n(\widehat{\gamma}(\boldsymbol{x}) - \gamma(\boldsymbol{x})) \xrightarrow{d} \Gamma$, where Γ is a nondegenerate distribution and $v_n \to \infty$. If $v_n A((1-\tau_n)^{-1}|\boldsymbol{x}) = O(1)$ and $v_n/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$, then*

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

In our context, v_n is typically $\sqrt{nh_n^p(1-\tau_n)}$ and $\overline{q}_n(\tau_n|\boldsymbol{x}) = \widehat{q}_n(\tau_n|\boldsymbol{x})$. As a conditional tail index estimator, we use the so-called "kernel version of the Hill estimator" from [10]:

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

The asymptotic distribution of $\widehat{\gamma}^{(J)}_{\tau_n}(\boldsymbol{x})$ can be deduced from Theorem 2.1. Note that the number J of high quantiles is fixed; the case $J = J_n \to \infty$, which would correspond to the classical established theory of the Hill estimator [see *e.g.* 14, Section 3.2], cannot be handled using Theorem 2.1. Obtaining asymptotic results for a growing number of summands $J = J_n$ would involve the difficult study of conditional tail empirical quantile processes.

THEOREM 2.3. *Work under the conditions of Theorem 2.1, and assume in addition that $\sqrt{nh_n^p(1-\tau_n)}A\left((1-\tau_n)^{-1}|\boldsymbol{x}\right) \to \lambda(\boldsymbol{x}) \in \mathbb{R}$. Then,*

$$\sqrt{nh_n^p(1-\tau_n)}\left(\widehat{\gamma}^{(J)}_{\tau_n}(\boldsymbol{x}) - \gamma(\boldsymbol{x}), \frac{\widehat{q}_n(\tau_n|\boldsymbol{x})}{q(\tau_n|\boldsymbol{x})} - 1\right)$$

$$\xrightarrow{d} \mathcal{N}\left(\left(\frac{\lambda(\boldsymbol{x})}{\log(J!)}\sum_{j=2}^{J}\frac{j^{\rho(\boldsymbol{x})}-1}{\rho(\boldsymbol{x})},\,0\right),\,\frac{\int_{\mathbb{R}^p}K^2}{g(\boldsymbol{x})}\gamma^2(\boldsymbol{x})\begin{pmatrix}\frac{J(J-1)(2J-1)}{6\log^2(J!)} & 0\\ 0 & 1\end{pmatrix}\right).$$

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

The asymptotic variance of $\widehat{\gamma}_{\tau_n}^{(J)}(\boldsymbol{x})$ is minimal when $J=9$. We shall adopt this choice in our finite-sample experiments. Theorem 2.3 improves upon Corollary 2 in [10], by removing unnecessary assumptions about the right conditional tail, and by weakening a bias assumption, see the discussion below Theorem 2.1. The bias component Δ does not appear in the limiting distribution of $\widehat{\gamma}_{\tau_n}^{(J)}(\boldsymbol{x})$, which suggests that the local variation of conditional extremes is not as important in the estimation of the conditional shape parameter $\gamma(\boldsymbol{x})$ of the Pareto approximating distribution as it is for its scale $q(\tau_n|\boldsymbol{x})$.

3. Nonparametric extremal expectile regression.

3.1. *Framework.* Rewrite the conditional distribution function in (3) as $E(y|\boldsymbol{x})=1-\varphi^{(1)}(y|\boldsymbol{x})/(2\varphi^{(1)}(y|\boldsymbol{x})+y-m(\boldsymbol{x}))$ where $\varphi^{(a)}(y|\boldsymbol{x})=\mathbb{E}((Y-y)^a\mathbb{1}_{\{Y>y\}}|\boldsymbol{X}=\boldsymbol{x})$ and $m(\boldsymbol{x})=\mathbb{E}(Y|\boldsymbol{X}=\boldsymbol{x})$. Consider the following kernel smoother for $E(y|\boldsymbol{x})$:

$$\widehat{E}_n(y|\boldsymbol{x})=1-\frac{\widehat{\varphi}_n^{(1)}(y|\boldsymbol{x})}{2\widehat{\varphi}_n^{(1)}(y|\boldsymbol{x})+(y-\widehat{m}_n(\boldsymbol{x}))}\text{ with }\widehat{m}_n(\boldsymbol{x})=\frac{1}{nh_n^p\,\widehat{g}_n(\boldsymbol{x})}\sum_{t=1}^{n}Y_tK\left(\frac{\boldsymbol{x}-\boldsymbol{X}_t}{h_n}\right)$$

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

The estimator $\widehat{m}_n$ is the Nadaraya-Watson estimator of the regression function m. The characterization of conditional expectiles as $e(\tau|\boldsymbol{x})\equiv\inf\{y\in\mathbb{R}\,|\,E(y|\boldsymbol{x})\geq\tau\}$ (see [20]) implies that they can be estimated by their empirical counterparts

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

This is in fact exactly the *asymmetric least squares estimator* $\widehat{e}_n(\tau|\boldsymbol{x})=\arg\min_{\theta\in\mathbb{R}}\int_{\mathbb{R}}[\eta_\tau(y-\theta)-\eta_\tau(y)]\,d\widehat{F}_n(y|\boldsymbol{x})$ obtained by smoothing up the loss function defining $e(\tau|\boldsymbol{x})$ in (2).

The definition of conditional expectiles in (2) and (3) requires $\mathbb{E}(|Y|\,|\,\boldsymbol{X}=\boldsymbol{x})<\infty$. To obtain the asymptotic normality of $\widehat{e}_n(\tau|\boldsymbol{x})$ at intermediate levels, we make the following additional assumptions on conditional tail heaviness and regularity of conditional moments.

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

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

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

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

Condition $\mathcal{H}_\delta$ (in which $\delta>0$) guarantees a finite conditional moment of order $(2+\delta)$ in a neighborhood of $\boldsymbol{x}$; in the unconditional framework, Theorem 2 in [11] requires the analogous $\mathbb{E}(Y_-^{2+\delta})<\infty$. Conditions $\mathcal{L}_m$ and $\mathcal{B}_m$ ensure the convergence of the regression mean

at a standard rate, see Proposition A.1 in Section A.4 of the Supplementary Material document [12]. Finally, in the spirit of Section 2, a 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 $\boldsymbol{X} = \boldsymbol{x}$, and the conditional mean functions $\mathbb{E}(Y|\boldsymbol{X} = \cdot)$ and $\mathbb{E}(Y^2|\boldsymbol{X} = \cdot)$ are continuously differentiable in a neighborhood of $\boldsymbol{x}$ and have Lipschitz continuous gradients at $\boldsymbol{x}$.

An alternative approach uses the asymptotic proportionality between extreme quantiles and expectiles of heavy-tailed distributions, namely, $e(\tau|\boldsymbol{x})/q(\tau|\boldsymbol{x}) \to (1/\gamma(\boldsymbol{x}) - 1)^{-\gamma(\boldsymbol{x})}$ as $\tau \uparrow 1$ [see Proposition 2.3 in 3]. Plugging in the estimators $\widehat{\gamma}^{(J)}_{\tau_n}(\boldsymbol{x})$ and $\widehat{q}_n(\tau_n|\boldsymbol{x})$ results in the *quantile-based estimator*

$$\breve{e}_n(\tau_n|\boldsymbol{x}) \equiv \breve{e}^{(J)}_n(\tau_n|\boldsymbol{x}) = (1/\widehat{\gamma}^{(J)}_{\tau_n}(\boldsymbol{x}) - 1)^{-\widehat{\gamma}^{(J)}_{\tau_n}(\boldsymbol{x})}\widehat{q}_n(\tau_n|\boldsymbol{x}).$$

3.2. *Intermediate expectile estimation.* We first derive the asymptotic distribution of the intermediate expectile estimator $\widehat{e}_n(\tau_n|\boldsymbol{x})$ 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 3.1. *Assume that conditions $\mathcal{M}$, $\mathcal{A}(l_n, r_n)$, $\mathcal{K}$, $\mathcal{H}_\delta$, $\mathcal{L}_g$, $\mathcal{L}_m$, $\mathcal{L}_\omega$, $\mathcal{B}_p$, $\mathcal{B}_m$ and $\mathcal{B}_\Omega$ 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 $n \to \infty$. Assume further that $h_n \to 0$ is such that $nh_n^p(1-\tau_n) \to \infty$, $nh_n^{p+2}(1-\tau_n)\log^2(1-\tau_n) \to 0$, $\sqrt{nh_n^p(1-\tau_n)}A((1-\tau_n)^{-1}|\boldsymbol{x}) = O(1)$ and $r_n(r_n/\sqrt{nh_n^p(1-\tau_n)})^\delta \to 0$. Then*

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

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

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

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

If $((\boldsymbol{X}_t, Y_t))_{t\geq 1}$ is $\rho-$mixing with summability of the $\rho-$mixing coefficient series (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 $\mathcal{H}_\delta$ can be replaced by $\gamma(\boldsymbol{x}) < 1/(2+\delta)$. If $((\boldsymbol{X}_t, Y_t))_{t\geq 1}$ is in fact also $\psi-$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 summability of the $\rho-$mixing coefficients), all conditions on (l_n) and (r_n) (including $\mathcal{A}(l_n, r_n)$) as well as conditions $\mathcal{B}_p$ and $\mathcal{B}_\Omega$ can also be dropped, and condition $\mathcal{H}_\delta$ can be replaced by the weaker requirement that $0 < \gamma(\boldsymbol{x}) < 1/2$.

For $\widehat{e}_n(\tau_n|\boldsymbol{x})$, 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 convergence rate, see Theorem 1 in [17] and the discussion below our Theorem 2.1. By contrast, mixing changes the asymptotic distribution of unconditional empirical intermediate expectiles, see [13].

We turn to the asymptotic normality of the quantile-based expectile estimator $\breve{e}_n(\tau_n|\boldsymbol{x})$.

THEOREM 3.2. *Work under the conditions of Theorem 2.1 with $\gamma(\boldsymbol{x}) < 1$ and $\mathbb{E}(Y_- \mid \boldsymbol{X} = \boldsymbol{x}) < \infty$, and assume in addition that $\sqrt{nh_n^p(1-\tau_n)}A((1-\tau_n)^{-1}|\boldsymbol{x}) \to \lambda_1(\boldsymbol{x}) \in \mathbb{R}$ and $\sqrt{nh_n^p(1-\tau_n)}/q(\tau_n|\boldsymbol{x}) \to \lambda_2(\boldsymbol{x}) \in \mathbb{R}$. Then*

$$\sqrt{nh_n^p(1-\tau_n)}\left(\frac{\breve{e}_n(\tau_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})} - 1\right) \xrightarrow{d} \mathcal{N}\Bigg(b_1^{(J)}(\gamma(\boldsymbol{x}),\rho(\boldsymbol{x}))\lambda_1(\boldsymbol{x}) + b_2(\gamma(\boldsymbol{x}),m(\boldsymbol{x}))\lambda_2(\boldsymbol{x}),$$
$$\frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})}\gamma^2(\boldsymbol{x})\left(1 + [(1-\gamma(\boldsymbol{x}))^{-1} - \log(1/\gamma(\boldsymbol{x}) - 1)]^2 \frac{J(J-1)(2J-1)}{6\log^2(J!)}\right)\Bigg),$$

where

$$b_1^{(J)}(\gamma(\boldsymbol{x}),\rho(\boldsymbol{x})) = \frac{(1-\gamma(\boldsymbol{x}))^{-1} - \log(1/\gamma(\boldsymbol{x}) - 1)}{\log(J!)} \sum_{j=2}^{J} \frac{j^{\rho(\boldsymbol{x})} - 1}{\rho(\boldsymbol{x})}$$
$$- \left(\frac{(1/\gamma(\boldsymbol{x}) - 1)^{-\rho(\boldsymbol{x})}}{1 - \gamma(\boldsymbol{x}) - \rho(\boldsymbol{x})} + \frac{(1/\gamma(\boldsymbol{x}) - 1)^{-\rho(\boldsymbol{x})} - 1}{\rho(\boldsymbol{x})}\right)$$

and $b_2(\gamma(\boldsymbol{x}), m(\boldsymbol{x})) = -\gamma(\boldsymbol{x})(1/\gamma(\boldsymbol{x}) - 1)^{\gamma(\boldsymbol{x})} m(\boldsymbol{x})$.

If conditions $\mathcal{KS}$, $\mathcal{D}_g$ and $\mathcal{D}_\omega$ hold, under the weaker bias assumption $\sqrt{nh_n^p(1-\tau_n)} \times h_n^2 \log^2(1-\tau_n) \to \Delta \in [0,\infty)$ and if $r_n h_n^p \to 0$, the quantity $\Upsilon_K(\Delta, \boldsymbol{x})$ is added to the asymptotic mean of the above Gaussian limit.

As is the case for $\widehat{\gamma}_{\tau_n}^{(J)}(\boldsymbol{x})$, the asymptotic variance of $\breve{e}_n(\tau_n|\boldsymbol{x}) \equiv \breve{e}_n^{(J)}(\tau_n|\boldsymbol{x})$ is minimal when $J = 9$. An important benefit of this quantile-based estimator is that its asymptotic normality does not require conditions $\mathcal{H}_\delta$, $\mathcal{L}_m$ and $\mathcal{B}_m$. In particular, it applies as soon as a finite conditional first moment exists. The estimator $\breve{e}_n(\tau_n|\boldsymbol{x})$ is biased, however, and its variance is higher than the variance of $\widehat{e}_n(\tau_n|\boldsymbol{x})$ when $\gamma(\boldsymbol{x})$ is only moderately large.

3.3. *Extreme expectile estimation.* The asymptotic proportionality of extreme quantiles and expectiles entails $e(\tau_n'|\boldsymbol{x}) \approx ((1-\tau_n')/(1-\tau_n))^{-\gamma(\boldsymbol{x})} e(\tau_n|\boldsymbol{x})$, for n large. Plugging in consistent estimators $\widehat{\gamma}(\boldsymbol{x})$ of $\gamma(\boldsymbol{x})$ and $\overline{e}_n(\tau_n|\boldsymbol{x})$ of $e(\tau_n|\boldsymbol{x})$ (such as $\widehat{e}_n(\tau_n|\boldsymbol{x})$ or $\breve{e}_n(\tau_n|\boldsymbol{x})$) yields a Weissman-type estimator of $e(\tau_n'|\boldsymbol{x})$:

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

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

THEOREM 3.3. *Assume that condition $\mathcal{M}$ holds, with $\gamma(\boldsymbol{x}) < 1$, $\rho(\boldsymbol{x}) < 0$ and $\mathbb{E}(Y_- \mid \boldsymbol{X} = \boldsymbol{x}) < \infty$. Let $\tau_n, \tau_n' \uparrow 1$ be such that $(1-\tau_n')/(1-\tau_n) \to 0$ and assume that $v_n(\overline{e}_n(\tau_n|\boldsymbol{x})/e(\tau_n|\boldsymbol{x}) - 1) = O_{\mathbb{P}}(1)$ and $v_n(\widehat{\gamma}(\boldsymbol{x}) - \gamma(\boldsymbol{x})) \xrightarrow{d} \Gamma$, where Γ is a nondegenerate distribution and $v_n \to \infty$. If $v_n A((1-\tau_n)^{-1}|\boldsymbol{x}) = O(1)$, $v_n/q(\tau_n|\boldsymbol{x}) = O(1)$ and $v_n/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$, then*

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

Let now (τ_n) be an intermediate sequence. Define an expectile-based estimator of $\gamma(\boldsymbol{x})$ as

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

where $\widehat{\overline{F}}_n(\cdot|\boldsymbol{x}) = 1 - \widehat{F}_n(\cdot|\boldsymbol{x})$. We derive the joint asymptotic normality of $(\widehat{\gamma}^E_{\tau_n}(\boldsymbol{x}), \widehat{e}_n(\tau_n|\boldsymbol{x}))$.

THEOREM 3.4. *Work under the conditions of Theorem 3.1. Assume further that* $\sqrt{nh_n^p(1-\tau_n)}A((1-\tau_n)^{-1}|\boldsymbol{x}) \to \lambda_1(\boldsymbol{x}) \in \mathbb{R}$ *and* $\sqrt{nh_n^p(1-\tau_n)}/q(\tau_n|\boldsymbol{x}) \to \lambda_2(\boldsymbol{x}) \in \mathbb{R}$. *Then*

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

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

$$b_E(\boldsymbol{x}) = \frac{\gamma(\boldsymbol{x})\left(1/\gamma(\boldsymbol{x}) - 1\right)^{1-\rho(\boldsymbol{x})}}{1-\gamma(\boldsymbol{x})-\rho(\boldsymbol{x})}\lambda_1(\boldsymbol{x}) + \gamma^2(\boldsymbol{x})(1/\gamma(\boldsymbol{x})-1)^{\gamma(\boldsymbol{x})+1} m(\boldsymbol{x})\lambda_2(\boldsymbol{x}).$$

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

Again, mixing does not impact the asymptotic distribution of $\widehat{\gamma}^E_{\tau_n}(\boldsymbol{x})$ obtained in the i.i.d. setting [see Theorem 4 in 17] and the bias component Δ does not appear in the limit. The advantage of using $\widehat{\gamma}^E_{\tau_n}(\boldsymbol{x})$ is its low asymptotic variance when $\gamma(\boldsymbol{x})$ is moderately high, at the expense of a strong bias. We 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 $((\boldsymbol{X}_t, Y_t))_{t\geq 1}$ is geometrically $\alpha-$mixing, namely, there exists $a \in (0,1)$ such that $\alpha(n) = O(a^n)$. Besides, in the typical extreme value models where $A(t|\boldsymbol{x}) \propto t^{\rho(\boldsymbol{x})}$, 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 2.1. We therefore focus in this section on the validity of assumptions $\mathcal{M}$, $\mathcal{H}_\delta$ (for a given $\delta > 0$), $\mathcal{L}_g$, $\mathcal{L}_m$, $\mathcal{L}_\omega$, $\mathcal{B}_p$, $\mathcal{B}_m$, $\mathcal{B}_\Omega$, $\mathcal{D}_g$, $\mathcal{D}_m$ and $\mathcal{D}_\omega$. We provide an extended discussion with more insight, and sometimes weaker conditions, in Section B of the Supplementary Material document [12]. 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 $Y_t = m(\boldsymbol{X}_t) + \sigma(\boldsymbol{X}_t)\varepsilon_t$ 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 $(\boldsymbol{X}_t)$. The ε_t can be dependent, yielding a possibly misspecified regression model in the sense that relevant, serially correlated covariates can be missing in $\boldsymbol{X}_t$ but left in the error ε_t.

PROPOSITION 4.1 (Location-scale model). *Assume that* $(\boldsymbol{X}_t)_{t\geq 1}$ *is* $\beta-$*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* $(\boldsymbol{X}_1, \boldsymbol{X}_{t+1})$ *have absolutely continuous distributions whose first marginal has a p.d.f.* g *such that* $g(\boldsymbol{x}) > 0$, *and the functions* g, m *and* σ *are continuously differentiable in a neighborhood of* $\boldsymbol{x}$.

• *The errors ε_t have a common p.d.f. f_ε with respect to the Lebesgue measure such that $f_\varepsilon(z) = c_0 z^{-1/\gamma-1}(1 + d_0 z^{-a} + d_0' z^{-a-b}(1 + o(1)))$ as $z \to \infty$, 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 $\mathcal{M}$, $\mathcal{L}_g$, $\mathcal{L}_\omega$, $\mathcal{B}_p$ and $\mathcal{B}_\Omega$ hold. If moreover $\gamma < 1/(2+\delta)$ and $\mathbb{E}(\varepsilon_-^{2+\delta}) < \infty$, then conditions $\mathcal{H}_\delta$, $\mathcal{L}_m$ and $\mathcal{B}_m$ hold as well. If in addition g, m and σ are twice continuously differentiable in a neighborhood of $\boldsymbol{x}$, 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 $-z f_\varepsilon'(z)/f_\varepsilon(z) \to 1/\gamma + 1$, then condition $\mathcal{D}_\omega$ holds, with $\lim_{y\to\infty} \nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})/\log(y) = 0$.

The class of $\beta-$mixing processes covers many important cases such as Harris recurrent aperiodic Markov chains [4, Corollary 3.6]. Our conditions on m 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 $(\boldsymbol{X}_t)$ 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 $y \mapsto F(y, \boldsymbol{\theta})$). We abuse notation and let $\boldsymbol{\theta} = \boldsymbol{\theta}(\cdot) : \boldsymbol{x} \in \mathbb{R}^p \mapsto \boldsymbol{\theta}(\boldsymbol{x}) \in \Theta$ be a smooth mapping, and we consider the model $Y_t = q(U_t, \boldsymbol{\theta}(\boldsymbol{X}_t))$ where (U_t) is a stationary sequence of unobserved, uniformly distributed innovations independent from the series $(\boldsymbol{X}_t)$. Then Y given $\boldsymbol{X} = \boldsymbol{x}$ has distribution function $F(\cdot, \boldsymbol{\theta}(\boldsymbol{x}))$: when $\boldsymbol{\theta}(\cdot)$ is linear, this setting covers the (V)GLM model of [28] for a univariate response variable.

PROPOSITION 4.2 (Nonlinear regression model). *Assume that $(\boldsymbol{X}_t)_{t\geq 1}$ is $\beta-$mixing and $(U_t)_{t\geq 1}$ is strongly mixing. Suppose further that:*

• *For any $\boldsymbol{\theta}$, the survival function $\overline{F}(\cdot, \boldsymbol{\theta})$ 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 $\boldsymbol{\theta} \mapsto \overline{F}(y, \boldsymbol{\theta})$ is continuously differentiable for y large enough.*
• *For any $t \geq 1$, the random pairs $(\boldsymbol{X}_1, \boldsymbol{X}_{t+1})$ have absolutely continuous distributions whose first marginal has a p.d.f. g such that $g(\boldsymbol{x}) > 0$ and g is continuously differentiable in a neighborhood of $\boldsymbol{x}$.*
• *The parameter mapping $\boldsymbol{\theta} : \mathbb{R}^p \to \Theta$ is continuously differentiable in a neighborhood of $\boldsymbol{x}$. There are $y_0 > 0$ and a continuous function κ on Θ with*

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

Then conditions $\mathcal{M}$, $\mathcal{L}_g$, $\mathcal{L}_\omega$, $\mathcal{B}_p$ and $\mathcal{B}_\Omega$ hold. Assume moreover that:

• *The survival function $\overline{F}(\cdot, \boldsymbol{\theta})$ only puts mass on $[0, \infty)$.*
• *At the point $\boldsymbol{x}$, $\gamma(\boldsymbol{\theta}(\boldsymbol{x})) < 1/(2+\delta)$.*
• *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 $\mathcal{H}_\delta$, $\mathcal{L}_m$ and $\mathcal{B}_m$ hold. If 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 $\boldsymbol{x}$, then conditions $\mathcal{D}_g$ and $\mathcal{D}_m$ hold. If $\boldsymbol{\theta} \mapsto \gamma(\boldsymbol{\theta})$ is also twice continuously differentiable on Θ and there exist $y_0 > 0$ and a continuous function κ on Θ with

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

then condition $\mathcal{D}_\omega$ holds, with $\lim_{y\to\infty} \nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x}) / \log(y) = [\nabla\gamma(\boldsymbol{\theta}(\boldsymbol{x}))]^\top J\boldsymbol{\theta}(\boldsymbol{x})/\gamma^2(\boldsymbol{\theta}(\boldsymbol{x}))$.

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. If $\overline{F}(\cdot, \boldsymbol{\theta})$ puts mass on a neighborhood of $-\infty$, then extra assumptions on the left conditional tail (such as symmetry) are required to ensure that conditions $\mathcal{H}_\delta$, $\mathcal{L}_m$, $\mathcal{B}_m$ and $\mathcal{D}_m$ hold.

Autoregressive model. Consider the causal and invertible 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 AR(p) equations, and $\boldsymbol{X}_t = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})^\top$.

PROPOSITION 4.3 (Autoregressive model). *Assume that the common distribution of the ε_t 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 $f_\varepsilon(z) = c_0 z^{-1/\gamma-1}(1 + d_0 z^{-a} + d_0' z^{-a-b}(1 + o(1)))$ as $z \to \infty$, 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 $\mathcal{M}$, $\mathcal{L}_g$, $\mathcal{D}_m$, $\mathcal{L}_\omega$, $\mathcal{B}_p$ and $\mathcal{B}_\Omega$ hold, and the process (Y_t) is geometrically $\beta-$ and $\rho-$mixing. If ε also has a finite moment of order $(2+\delta)$ and $\gamma < 1/(2+\delta)$, then condition $\mathcal{H}_\delta$ 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 $-z f_\varepsilon'(z)/f_\varepsilon(z) \to 1/\gamma + 1$, then condition $\mathcal{D}_\omega$ holds, with $\lim_{y\to\infty} \nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x}) / \log(y) = 0$.

Condition $\mathcal{B}_m$ is unnecessary because (Y_t) is $\rho-$mixing. Unlike in our other examples, checking condition $\mathcal{B}_\Omega$ is nontrivial, because the sequences $(\boldsymbol{X}_t) = ((Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})^\top)$ 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 A.1(iii) of the Supplementary Material document [12] (with $t_0 = p$). Handling models whose natural covariate is infinite-dimensional, such as the ARMA and GARCH models containing lagged values of the innovation, or moving maxima processes such as the m4 process of [23], in which joint distributions of pairs of responses at different time points are never absolutely continuous, should be viewed as substantially harder and worthy of future research.

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.* 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, one should also 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 $\mathbb{E}(\int_{\mathcal{X}}(\widehat{g}_n(\boldsymbol{x})-g(\boldsymbol{x}))^2 d\boldsymbol{x})$ of the density estimator over the support $\mathcal{X}$ of $\boldsymbol{X}$, that is,

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

For example, when $p=1$, the classical normal scale rule derived from (4), assuming a Gaussian p.d.f. g and a naive kernel $K(u)=1/2$ on $[-1,1]$ yields $\widehat{h}_{n,\star}=(12\sqrt{\pi})^{1/5}\widehat{\sigma}_n n^{-1/5}$, where $\widehat{\sigma}_n$ 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 $V(\boldsymbol{x})$ and an asymptotic bias component $\lambda(\boldsymbol{x})B(\boldsymbol{x})$, where $\lambda(\boldsymbol{x})=\lim_{n\to\infty}\sqrt{k_n h_{n,\star}^p}A(n/k_n|\boldsymbol{x})$, we define

$$k_{n,\star}(\boldsymbol{x})=\underset{1\le k\le n}{\arg\min}\{(kh_{n,\star}^p)^{-1}V(\boldsymbol{x})+B^2(\boldsymbol{x})A^2(n/k|\boldsymbol{x})\}.$$

A common practice is to consider the very general case when $A(t|\boldsymbol{x})=b(\boldsymbol{x})\gamma(\boldsymbol{x})t^{\rho(\boldsymbol{x})}$, for $b(\boldsymbol{x})\in\mathbb{R}$ and $\rho(\boldsymbol{x})<0$, see among others [19]. This yields the closed form expression

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

The quantities $b(\boldsymbol{x})$ and $\rho(\boldsymbol{x})$ are estimated using naive kernel regression versions $\overline{b}(\boldsymbol{x})$ and $\overline{\rho}(\boldsymbol{x})$ of the moment estimators provided by the R function `mop` from the `Expectrem` package, see Section C.1 of the Supplementary Material document [12]. To estimate $B(\boldsymbol{x})$ and $V(\boldsymbol{x})$, note that when the tail index estimator is chosen as $\widehat{\gamma}^{(J)}_{1-k_n/n}(\boldsymbol{x})$, we have $V(\boldsymbol{x})=(\int_{\mathbb{R}^p}K^2/g(\boldsymbol{x}))v_q^{(J)}(\boldsymbol{x})$ and $B(\boldsymbol{x})=B^{(J)}(\boldsymbol{x})$, with

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

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

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

These quantities are estimated by plugging in $\widehat{g}_n(\boldsymbol{x})$ (with $h=h_{n,\star}$) and $\overline{\rho}(\boldsymbol{x})$ in place of $g(\boldsymbol{x})$ and $\rho(\boldsymbol{x})$, and a naive kernel regression version $\overline{\gamma}(\boldsymbol{x})$ of the Hill estimator in place of $\gamma(\boldsymbol{x})$, with $h=h_{n,\star}$ and effective sample size k corresponding to 25% of the local sample size $N_h(\boldsymbol{x})=\sum_{t=1}^n \mathbb{1}_{\{\|\boldsymbol{X}_t-\boldsymbol{x}\|\le h\}}$ (see Section C.1 of [12] for the expression of $\overline{\gamma}(\boldsymbol{x})$).

This results in data-driven choices $\widehat{k}^{(J)}_{n,\star}(\boldsymbol{x})$ and $\widehat{k}^{E}_{n,\star}(\boldsymbol{x})$, depending on the tail index estimation technique. We omit the dependence of this selected value upon the tail index estimator and we denote it again by $k_{n,\star}$ for the sake of brevity. Our choices may not be optimal in certain difficult cases, but they afford effective data-based rules on our simulated and real data.

5.2. *Bias correction guidelines.* The quality of the pure Pareto approximation to conditional heavy tails deteriorates as $\rho(\boldsymbol{x})$ gets closer to 0, with the resulting bias being possibly very substantial then. Bias-reduced versions of $\widehat{\gamma}^{(J)}_{1-k_n/n}(\boldsymbol{x})$ and $\widehat{q}^{W}_{n,\tau_n}(\tau_n'|\boldsymbol{x})$ are

$$\widehat{\gamma}^{(J,BR)}_{1-k_n/n}(\boldsymbol{x})=\widehat{\gamma}^{(J)}_{1-k_n/n}(\boldsymbol{x})\left(1-\frac{1}{\log(J!)}\sum_{j=1}^{J}\frac{j^{\overline{\rho}(\boldsymbol{x})}-1}{\overline{\rho}(\boldsymbol{x})}\overline{b}(\boldsymbol{x})\left(\frac{n}{k_n}\right)^{\overline{\rho}(\boldsymbol{x})}\right)$$

$$\text{and } \widehat{q}^{W,BR}_{n,\tau_n}(\tau_n'|\boldsymbol{x}) = \widehat{q}^{W}_{n,\tau_n}(\tau_n'|\boldsymbol{x}) \left(1 + \frac{\left(\frac{k_n}{n(1-\tau_n')}\right)^{\overline{\rho}(\boldsymbol{x})} - 1}{\overline{\rho}(\boldsymbol{x})} \overline{b}(\boldsymbol{x}) \widehat{\gamma}^{(J,BR)}_{1-k_n/n}(\boldsymbol{x}) \left(\frac{n}{k_n}\right)^{\overline{\rho}(\boldsymbol{x})} \right).$$

These estimators are respectively inspired by Theorem 2.3 and by the approach of [19]. We use $\widehat{\gamma}^{(J,BR)}_{1-k_n/n}(\boldsymbol{x})$ within $\widehat{q}^{W}_{n,\tau_n}(\tau_n'|\boldsymbol{x})$ in our subsequent implementation of $\widehat{q}^{W}_{n,\tau_n}(\tau_n'|\boldsymbol{x})$.

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 3.3 and 3.4, 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 $\widehat{\gamma}^{E}_{1-k_n/n}(\boldsymbol{x})$:

$$\widehat{\gamma}^{E,BR}_{1-k_n/n}(\boldsymbol{x}) = \left(1 + \frac{\widehat{\overline{F}}_n(\widehat{e}_n(1-k_n/n|\boldsymbol{x})|\boldsymbol{x})}{k_n/n} \frac{1}{1+\widehat{r}(1-k_n/n|\boldsymbol{x})} \right)^{-1} \text{ where } \widehat{r}(1-k_n/n|\boldsymbol{x})$$

$$= \left(1 - \frac{\widehat{m}_n(\boldsymbol{x})}{\widehat{e}_n(1-k_n/n|\boldsymbol{x})} \right) \frac{1}{1-2k_n/n} \left(1 + \frac{\overline{b}(\boldsymbol{x})[\widehat{\overline{F}}_n(\widehat{e}_n(1-k_n/n|\boldsymbol{x})|\boldsymbol{x})]^{-\overline{\rho}(\boldsymbol{x})}}{1-\widehat{\gamma}^{E}_{1-k_n/n}(\boldsymbol{x}) - \overline{\rho}(\boldsymbol{x})} \right)^{-1} - 1.$$

This is a kernel regression version of the work in [18] in the unconditional setting. Here $\widehat{\gamma}^{E,BR}_{1-k_n/n}(\boldsymbol{x})$ is computed using the R function `tindexp` with argument `br=TRUE` (from the R package `Expectrem`) applied to the Y_t such that $\|\boldsymbol{X}_t - \boldsymbol{x}\| \leq h_{n,\star}$. Bias-reduced versions $\widehat{e}^{W,BR}_{n,\tau_n}(\tau_n'|\boldsymbol{x})$ and $\breve{e}^{W,BR}_{n,\tau_n}(\tau_n'|\boldsymbol{x})$ of the extreme conditional expectile estimators $\widehat{e}^{W}_{n,\tau_n}(\tau_n'|\boldsymbol{x})$ (extrapolating $\widehat{e}_n(\tau_n|\boldsymbol{x})$) and $\breve{e}^{W}_{n,\tau_n}(\tau_n'|\boldsymbol{x})$ (extrapolating $\breve{e}_n(\tau_n|\boldsymbol{x})$), computed using the R function `extExpect` with `br=TRUE` from that package, are described in Section C.2 of the Supplementary Material document [12].

5.3. *Pointwise asymptotic confidence intervals.*

5.3.1. *Extremal conditional quantiles.* We use the equivalent version of Theorem 2.2 on the log-scale, which tends to be more accurate in practice. Combined with Theorem 2.3, this suggests that $\log(\widehat{q}^{W,BR}_{n,1-k_{n,\star}/n}(\tau_n'|\boldsymbol{x})/q(\tau_n'|\boldsymbol{x}))$ is approximately Gaussian centered with variance $(\int_{\mathbb{R}^p} K^2/g(\boldsymbol{x}))v_q^{(J)}(\boldsymbol{x})$, with $v_q^{(J)}(\boldsymbol{x}) = \gamma^2(\boldsymbol{x}) \times J(J-1)(2J-1)/(6\log^2(J!))$. A first 95% asymptotic Gaussian confidence interval for $q(\tau_n'|\boldsymbol{x})$ is then

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

$$\text{with } \widehat{v}_q^{(J)}(\boldsymbol{x}) = \frac{J(J-1)(2J-1)}{6\log^2(J!)} (\widehat{\gamma}^{(J,BR)}_{1-k_{n,\star}/n}(\boldsymbol{x}))^2,$$

and where z_τ is the τth 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. By Theorem 2.3, $\widehat{\gamma}^{(J)}_{\tau_n}(\boldsymbol{x})$ is asymptotically independent of $\widehat{q}_n(\tau_n|\boldsymbol{x})$, so a straightforward calculation provides a refined version $\widetilde{v}_q^{(J)}(\boldsymbol{x})$ of the asymptotic variance estimate of $\log(\widehat{q}^{W,BR}_{n,1-k_{n,\star}/n}(\tau_n'|\boldsymbol{x})/q(\tau_n'|\boldsymbol{x}))$, and thus a corrected 95%

asymptotic Gaussian confidence interval as

$$\widehat{I}_{q,2}(\tau_n'|\boldsymbol{x}) = \left[\widehat{q}^{W,BR}_{n,1-k_{n,\star}/n}(\tau_n'|\boldsymbol{x}) \exp\left(\pm \frac{\sqrt{\frac{\int_{\mathbb{R}^p} K^2}{\widehat{g}_n(\boldsymbol{x})}\widetilde{v}_q^{(J)}(\boldsymbol{x})}}{\sqrt{k_{n,\star}h^p_{n,\star}}} \log\left(\frac{k_{n,\star}}{n(1-\tau_n')}\right) z_{0.975}\right)\right],$$

$$\text{with } \widetilde{v}_q^{(J)}(\boldsymbol{x}) = \left(\frac{J(J-1)(2J-1)}{6\log^2(J!)} + \frac{1}{\log^2(k_{n,\star}/[n(1-\tau_n')])}\right)(\widehat{\gamma}^{(J,BR)}_{1-k_{n,\star}/n}(\boldsymbol{x}))^2.$$

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 3.3 and 3.4 to deduce a first 95% asymptotic Gaussian confidence interval for $e(\tau_n'|\boldsymbol{x})$ as

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

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

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. Calculating the errors incurred in using these two approximations requires, first of all, an accurate quantification of the variance matrix of the random vector

$$\sqrt{k_{n,\star}h^p_{n,\star}}\left(\frac{n\widehat{\overline{F}}_n(\widehat{e}_n(1-k_{n,\star}/n|\boldsymbol{x})|\boldsymbol{x})}{k_{n,\star}} - \left(\frac{1}{\gamma(\boldsymbol{x})}-1\right), \frac{\widehat{e}_n(1-k_{n,\star}/n|\boldsymbol{x})}{e(1-k_{n,\star}/n|\boldsymbol{x})}-1\right)$$

since, up to order $1/\sqrt{k_{n,\star}h^p_{n,\star}}$, we have $\widehat{\gamma}^{E,BR}_{1-k_{n,\star}/n}(\boldsymbol{x}) \approx \mathcal{G}(n\widehat{\overline{F}}_n(\widehat{e}_n(1-k_{n,\star}/n|\boldsymbol{x})|\boldsymbol{x})/k_{n,\star})$, where $\mathcal{G}(u) = 1/(1+u)$. An inspection of the proofs of Theorems 3.1 and 3.4 (see Section C.3 of the Supplementary Material document [12]) suggests that a corrected asymptotic variance matrix for this random vector is $(\int_{\mathbb{R}^p} K^2/g(\boldsymbol{x}))\boldsymbol{T}_n(\boldsymbol{x})$ where the 2×2 symmetric matrix $\boldsymbol{T}_n(\boldsymbol{x})$ has components

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

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

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

Replacing $\gamma(\boldsymbol{x})$, $m(\boldsymbol{x})$ and $e(1-k_{n,\star}/n|\boldsymbol{x})$ with $\widehat{\gamma}^{E,BR}_{1-k_{n,\star}/n}(\boldsymbol{x})$, $\widehat{m}_n(\boldsymbol{x})$ and $\widehat{e}_n(1-k_{n,\star}/n|\boldsymbol{x})$ yields estimators $\widehat{\kappa}_{1,n}(\boldsymbol{x})$ and $\widehat{\kappa}_{2,n}(\boldsymbol{x})$ and therefore $\widehat{T}_{n,11}(\boldsymbol{x})$, $\widehat{T}_{n,12}(\boldsymbol{x})$ and $\widehat{T}_{n,22}(\boldsymbol{x})$. Then

$$\sqrt{k_{n,\star}h^p_{n,\star}}(\widehat{\gamma}^{E,BR}_{1-k_{n,\star}/n}(\boldsymbol{x}) - \gamma(\boldsymbol{x}))$$

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

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

through a power series expansion. [This is conceptually simpler than an Edgeworth expansion, which would approximate the p.d.f. of $\widehat{\gamma}^{E,BR}_{1-k_{n,\star}/n}(\boldsymbol{x})$.] Taking only the first term above leads to the delta-method for $\widehat{\gamma}^{E,BR}_{1-k_{n,\star}/n}(\boldsymbol{x})$; to obtain a finer approximation, we use all terms up to order 4. Based on our calculation of $\boldsymbol{T}_n(\boldsymbol{x})$, an asymptotic approximation of the covariance matrix of $\sqrt{k_{n,\star}h_{n,\star}^p}(\widehat{\gamma}^{E,BR}_{1-k_{n,\star}/n}(\boldsymbol{x})-\gamma(\boldsymbol{x}), \widehat{e}_n(1-k_{n,\star}/n|\boldsymbol{x})/e(1-k_{n,\star}/n|\boldsymbol{x})-1)$ is $(\int_{\mathbb{R}^p}K^2/g(\boldsymbol{x}))\boldsymbol{S}_n(\boldsymbol{x})$, where the symmetric matrix $\boldsymbol{S}_n(\boldsymbol{x})$ has components

$$S_{n,11}(\boldsymbol{x}) = \gamma^4(\boldsymbol{x})T_{n,11}(\boldsymbol{x})\left(1+8\frac{\gamma^2(\boldsymbol{x})T_{n,11}(\boldsymbol{x})}{k_{n,\star}h_{n,\star}}\times\frac{\int_{\mathbb{R}^p}K^2}{g(\boldsymbol{x})}\right),$$

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

We denote by $\widehat{\boldsymbol{S}}_n(\boldsymbol{x})$ the associated estimator, which hinges upon the estimators $\widehat{\boldsymbol{T}}_n(\boldsymbol{x})$ and $\widehat{\gamma}^{E,BR}_{1-k_{n,\star}/n}(\boldsymbol{x})$ previously introduced. The final step is to recall that (see Section C.2 of [12])

$$\log\frac{\widehat{e}^{W,BR}_{n,1-k_{n,\star}/n}(\tau_n'|\boldsymbol{x})}{e(\tau_n'|\boldsymbol{x})} \approx \log\frac{\widehat{e}^{W}_{n,1-k_{n,\star}/n}(\tau_n'|\boldsymbol{x})}{e(\tau_n'|\boldsymbol{x})} - (\widehat{\gamma}^{E,BR}_{1-k_{n,\star}/n}(\boldsymbol{x})-\gamma(\boldsymbol{x}))\log(\kappa_{1,n}(\boldsymbol{x}))$$

$$\approx \left(\log\left(\frac{k_{n,\star}}{n(1-\tau_n')}\right) - \log(\kappa_{1,n}(\boldsymbol{x}))\right)(\widehat{\gamma}^{E,BR}_{1-k_{n,\star}/n}(\boldsymbol{x})-\gamma(\boldsymbol{x})) + \log\frac{\widehat{e}_n(1-k_{n,\star}/n|\boldsymbol{x})}{e(1-k_{n,\star}/n|\boldsymbol{x})}.$$

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

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

This results in the corrected 95% asymptotic Gaussian confidence interval

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

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

6. Simulation study.

6.1. *Models and setup.* For the sake of brevity, we only report here results in the nonlinear (Burr) regression model, with covariate dimension $p=1$. Other cases spanning our list of worked-out examples, in dimensions $p=1$ and 2, are presented in Section C.4 of the Supplementary Material document [12].

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

- $X_t = \Phi(Z_t)$, where Φ is the standard Gaussian distribution function 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 $\Sigma_{t+1}=(\omega+\alpha Z_t^2+\beta\Sigma_t^2)^{1/2}$.

• (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 $\{0, 1/r, \ldots, (r-1)/r\}$, and $r = 5$. The U_t are standard uniform and $\alpha-$mixing, see Section C.4 of [12] for further details.

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) The polynomial model $\gamma(x) = 0.15 + 0.5x(1-x)$;
(NL-S) The 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 of [12]). 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, we select $h_n = h_{n,\star}$ and $\tau_n = \tau_{n,\star} = 1 - k_{n,\star}/n$ as in Section 5.1, and take $J = 9$ in the conditional tail index estimator $\widehat{\gamma}^{(J,BR)}_{1-k_{n,\star}/n}(x)$. Our quantile (resp. expectile) estimators are compared with the non-extrapolated quantile estimator $\widehat{q}_n(\tau'_n|x)$ (resp. the non-extrapolated expectile estimator $\widehat{e}_n(\tau'_n|x)$) and the simple extrapolated version $\widehat{q}^{W}_{n,1-k_{n,\star}/n}(\tau'_n|x)$ based on $\widehat{\gamma}^{(J)}_{1-k_{n,\star}/n}(x)$ (resp. the simple extrapolated version $\widehat{e}^{W}_{n,1-k_{n,\star}/n}(\tau'_n|x)$ based on $\widehat{\gamma}^{E}_{1-k_{n,\star}/n}(x)$). Our proposed 95% asymptotic Gaussian confidence intervals for $q(\tau'_n|x)$ are compared with the following competing intervals:

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

with $\overline{v}^{(J)}_q(\boldsymbol{x}) = \frac{J(J-1)(2J-1)}{6\log^2(J!)}(\widehat{\gamma}^{(J)}_{1-k_{n,\star}/n}(\boldsymbol{x}))^2$, similar to $\widehat{I}_{q,1}(\tau'_n|\boldsymbol{x})$ (without correction), and

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

based on an application of Theorem 2.1 written on the log-scale, which does not feature any extrapolation at all. Likewise, for $e(\tau'_n|\boldsymbol{x})$, we compute

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

with $\overline{v}_E(\boldsymbol{x}) = \frac{(\widehat{\gamma}^{E}_{1-k_{n,\star}/n}(\boldsymbol{x}))^3(1-\widehat{\gamma}^{E}_{1-k_{n,\star}/n}(\boldsymbol{x}))}{1-2\widehat{\gamma}^{E}_{1-k_{n,\star}/n}(\boldsymbol{x})}$, similar to $\widehat{I}_{E,1}(\tau'_n|\boldsymbol{x})$ (without correction), and

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

suggested by an equivalent version of Theorem 3.1 using the log-scale. This will make it possible to assess the benefits of the extrapolation procedure and bias correction scheme.

6.2. *Results.* We represent in Figure 3 boxplots of the extreme conditional quantile and asymmetric least squares expectile estimates, as well as the coverage probabilities of the intervals $\widehat{I}_{q,j}(\tau_n'|x)$ and $\widehat{I}_{E,j}(\tau_n'|x)$, for $1 \leq j \leq 4$, in models (NL-P), (NL-S) and (NL-C). Quantitative results at selected values of x are provided in Tables C.1, C.2 and C.3 in Section C.4 of the Supplementary Material document [12]. We discuss the conclusions from a full set of results for sample sizes $n \in \{1{,}000, 5{,}000, 10{,}000\}$ in an expanded list of models found in Section C.4 of [12], see Figures C.1–C.8 therein. Comparing quantile-based expectile estimates with their bias-reduced asymmetric least squares counterparts (see Figure C.1) reveals that $\widehat{e}^{W,BR}_{n,\tau_n}(\tau_n'|\boldsymbol{x})$ appears to have the smallest bias overall, although it has a slightly higher variance than the extrapolated, bias-reduced quantile-based estimator $\check{e}^{W,BR}_{n,\tau_n}(\tau_n'|\boldsymbol{x})$. This makes it difficult to decide which one of the asymmetric least squares or quantile-based expectile estimators performs best. We note that the main advantage of the quantile-based approach is to not rely upon the finite second moment assumption. Since our simulation setting considers models where the conditional second moment is finite, we focus throughout this section on the proposed extreme conditional quantile estimators as well as the asymmetric least squares conditional expectile estimators and their associated confidence intervals.

Even though 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. This is obvious from the coverage probabilities of $\widehat{I}_{E,4}(\tau_n'|x)$, often very close to 1. While bias reduction and variance correction for extrapolated estimates are arguably not crucial for quantile estimation at large values of n, and in fact bias correction can slightly deteriorate coverage probabilities of the Gaussian confidence interval, bias reduction is however valuable when the second-order parameter $\rho(x)$ gets close to 0 (see the top panels of Figure C.6), 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 third row of Figures C.2–C.7. 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 second row of Figures C.2–C.7. 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 A.10 in Section A.4 of [12]. 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 third row of Figures C.2–C.7. A coverage close to the nominal level is also achieved in the more challenging dimension $p = 2$, see the third column of Figure C.8, although misspecified values of the second-order parameters ρ and b were used in this difficult situation in order to better handle the finite-sample variability of the resulting quantile and expectile estimates. It should 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. As evidenced in the bottom panels of Figures C.2–C.7, the median length of the interval $\widehat{I}_{q,2}(\tau_n'|x)$ (resp. $\widehat{I}_{E,2}(\tau_n'|x)$) was found to be close to that of $\widehat{I}_{q,1}(\tau_n'|x)$ and $\widehat{I}_{q,3}(\tau_n'|x)$ (resp. $\widehat{I}_{E,1}(\tau_n'|x)$ and $\widehat{I}_{E,3}(\tau_n'|x)$) and always substantially shorter than that of $\widehat{I}_{q,4}(\tau_n'|x)$ (resp. $\widehat{I}_{E,4}(\tau_n'|x)$),

thus reflecting the added value of the extrapolation procedure combined with the bias and variance correction schemes.

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), whose large values represent large losses. In a full-fledged data analysis one would of course be interested in understanding the extreme value behavior of Y_t given the full history of the S&P 500 index; in this illustrative example we focus on predicting the conditional extreme quantiles and expectiles of Y_t given $X_t = x$, a fixed value. 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 of [12]. 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. 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 $\widehat{q}^{W,BR}_{n,1-k_{n,\star}/n}(\tau_n'|x)$ and $\widehat{e}^{W,BR}_{n,1-k_{n,\star}/n}(\tau_n'|x)$, along with their respective asymptotic 95% confidence intervals $\widehat{I}_{q,2}(\tau_n'|x)$ and $\widehat{I}_{E,2}(\tau_n'|x)$. The purely empirical estimates $\widehat{q}_n(\tau_n'|x)$ and $\widehat{e}_n(\tau_n'|x)$ are graphed in the left panel along with their respective asymptotic 95% confidence intervals $\widehat{I}_{q,4}(\tau_n'|x)$ and $\widehat{I}_{E,4}(\tau_n'|x)$. In agreement to what was observed in Section 6.2, the intervals $\widehat{I}_{q,4}(\tau_n'|x)$ and $\widehat{I}_{E,4}(\tau_n'|x)$ are almost overall wider than $\widehat{I}_{q,2}(\tau_n'|x)$ and $\widehat{I}_{E,2}(\tau_n'|x)$, respectively.

The estimated curves, confidence intervals, and regression mean all point towards a linear trend. This motivated us to perform a residual-based extreme value estimation from the ordinary linear regression model $Y_t = -0.00030 + 1.24X_t + \varepsilon_t$. We construct the residuals $\widehat{\varepsilon}_t$ and we calculate corresponding residual-based, bias-corrected extreme quantile estimates $\widehat{q}_{n,\tau_n,\varepsilon}(\tau_n')$, following [19], and bias-corrected extreme expectile estimates $\widehat{e}_{n,\tau_n,\varepsilon}(\tau_n')$, following [18], in which we chose $\tau_n = 1 - 200/n = 0.915$ after a graphical inspection of the Hill plot of residuals, with associated asymptotic 95% Gaussian confidence intervals on the log-scale. This yields conditional extreme point quantile and expectile estimates of the Y_t as $\overline{q}_{n,\tau_n}(\tau_n'|x) = -0.00030 + 1.24x + \widehat{q}_{n,\tau_n,\varepsilon}(\tau_n')$ and $\overline{e}_{n,\tau_n}(\tau_n'|x) = -0.00030 + 1.24x + \widehat{e}_{n,\tau_n,\varepsilon}(\tau_n')$, and their corresponding confidence intervals through the same linear translation, all of them graphed in the right panel of Figure 1. In this panel, we also represented

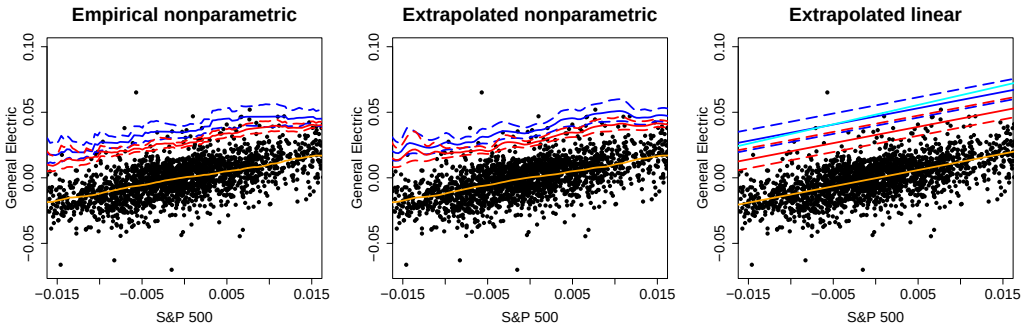

FIG 1. *Stock market data. Blue line: Estimated conditional quantile at level* $\tau_n' = 0.995$ *(with* 95% *confidence interval in dashed line), red line: Estimated conditional expectile at level* $\tau_n' = 0.995$ *(with* 95% *confidence interval in dashed line). Left panel: Nonparametric empirical estimates, middle panel: Extrapolated nonparametric estimates, right panel: Extrapolated estimates based on the linear regression model. In the left and middle panels, the orange line is the Nadaraya-Watson estimate. In the right panel, the orange line is the ordinary least squares line, and the cyan line represents the extremal quantile regression estimate of [8].*

the corresponding (extrapolated) extremal linear quantile regression estimate of [8], produced using their `extrap.rq` routine in which the preliminary estimate of the tail index is calculated using their `summary.rq.hill` routine at the intermediate level 0.95, chosen according to their recommendations. 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 dataset[2] records, for each tornado that has occurred in the United States between 1 January 2010 and 31 December 2019, the associated monetary loss (`loss`), its starting and ending latitude and longitude (`slat`, `slon`, `elat` and `elon`), and the length and width of the area 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 $\boldsymbol{X} = (X_1, X_2)$, that is, $Y = \texttt{loss}/(\texttt{len} \times \texttt{wid})$, $X_1 = (\texttt{slon} + \texttt{elon})/2$, and $X_2 = (\texttt{slat} + \texttt{elat})/2$. To keep the analysis simple and illustrative of our results, we do not attempt to conduct a conditional extreme value analysis of losses given the random path of the tornado, and we do not incorporate other covariates (such as population density) that can contribute to the monetary loss. This results in a sample $(\boldsymbol{X}_t, Y_t)$ 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. We focus on the part of the US to the 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).

[2]Available at `https://www.spc.noaa.gov/wcm/\#data`

The heavy tail model assumption was checked using local Generalized Pareto QQ-plots [see pp. 90-91 in 14] omitted for the sake of brevity. Tail index estimates were found to exceed 1 in many locations $\boldsymbol{x}$, which prevents the use of the expectile risk measure.

Our target is the extremal conditional risk measure $q(\tau_n'|\boldsymbol{x})$ at each location $\boldsymbol{x}$, 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 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 that 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 $\int_{\mathbb{R}^2}(\boldsymbol{u}^\top H_g(\boldsymbol{x})\boldsymbol{u})\,K(\boldsymbol{u})d\boldsymbol{u}$, involved in (4), 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 $X_2 \in [18, 49]$ of the data points 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 $H_g(\boldsymbol{x})$. Letting $\kappa(u) = (15/16)(1-u^2)^2$, for $u \in [-1, 1]$, be the one-dimensional quartic kernel and $K(\boldsymbol{u}) = (16/(5\pi))\kappa(\|\boldsymbol{u}\|)$ 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 $\boldsymbol{x}$, representing one of the 21,935 cities located east of the 100th meridian picked in the United States Cities Database[3], we chose the corresponding local optimal hyperparameter $k_{n,\star} = \widehat{k}_{n,\star}(\boldsymbol{x})$ as described above in Section 5.1, with the only difference that we set $\overline{\rho}(\boldsymbol{x}) \equiv -1$ and $\overline{b}(\boldsymbol{x}) \equiv 1$, as recommended in Section C.4 of [12] when the covariate has dimension 2. Each city $\boldsymbol{x}$ 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 Nadaraya-Watson estimates of the conditional mean of losses per squared yard and the 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 $\boldsymbol{x}$, 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. We also note 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 $\widehat{I}_{q,4}(0.995|\boldsymbol{x})$ based on purely empirical estimates are much wider than those provided by our bias- and variance-corrected proposal $\widehat{I}_{q,2}(0.995|\boldsymbol{x})$ based

[3] Available at `https://simplemaps.com/data/us-cities`

Location $\boldsymbol{x}$ (State)	$N_{h_{n,\star}}(\boldsymbol{x})$	$\widehat{\gamma}^{(J,BR)}_{1-k_{n,\star}/n}(\boldsymbol{x})$	$\widehat{q}_n(0.995\|\boldsymbol{x})$ $\widehat{I}_{q,4}(0.995\|\boldsymbol{x})$	$\widehat{q}^{W,BR}_{n,1-k_{n,\star}/n}(0.995\|\boldsymbol{x})$ $\widehat{I}_{q,2}(0.995\|\boldsymbol{x})$
New York (NY)	413	1.11	16.71 [2.46, 113.35]	63.53 [18.13, 222.70]
Charleston (SC)	839	0.96	48.08 [12.54, 184.28]	56.15 [22.58, 139.64]
Nashville (TN)	2,317	0.95	16.57 [8.31, 33.04]	24.51 [14.62, 41.08]
Captiva (FL)	205	0.93	236.74 [36.08, 1553.36]	144.67 [43.50, 481.13]
New Orleans (LA)	1,427	0.98	27.46 [11.44, 65.88]	31.01 [16.42, 58.58]
Woodson (TX)	958	1.53	118.37 [16.53, 847.72]	390.59 [100.95, 1511.26]
Kansas City (MO)	1,326	1.30	45.99 [11.29, 187.43]	69.00 [25.57, 186.21]
Minneapolis (MN)	620	0.93	34.09 [9.54, 121.79]	40.49 [17.01, 96.40]
Harrisville (MI)	472	0.79	24.86 [4.76, 129.71]	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.*

on extrapolated quantile estimates, and overrepresent the uncertainty about high regression quantiles, in agreement with our conclusions based on simulated data.

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 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.

[4]Available at `https://hazards.fema.gov/nri/tornado`

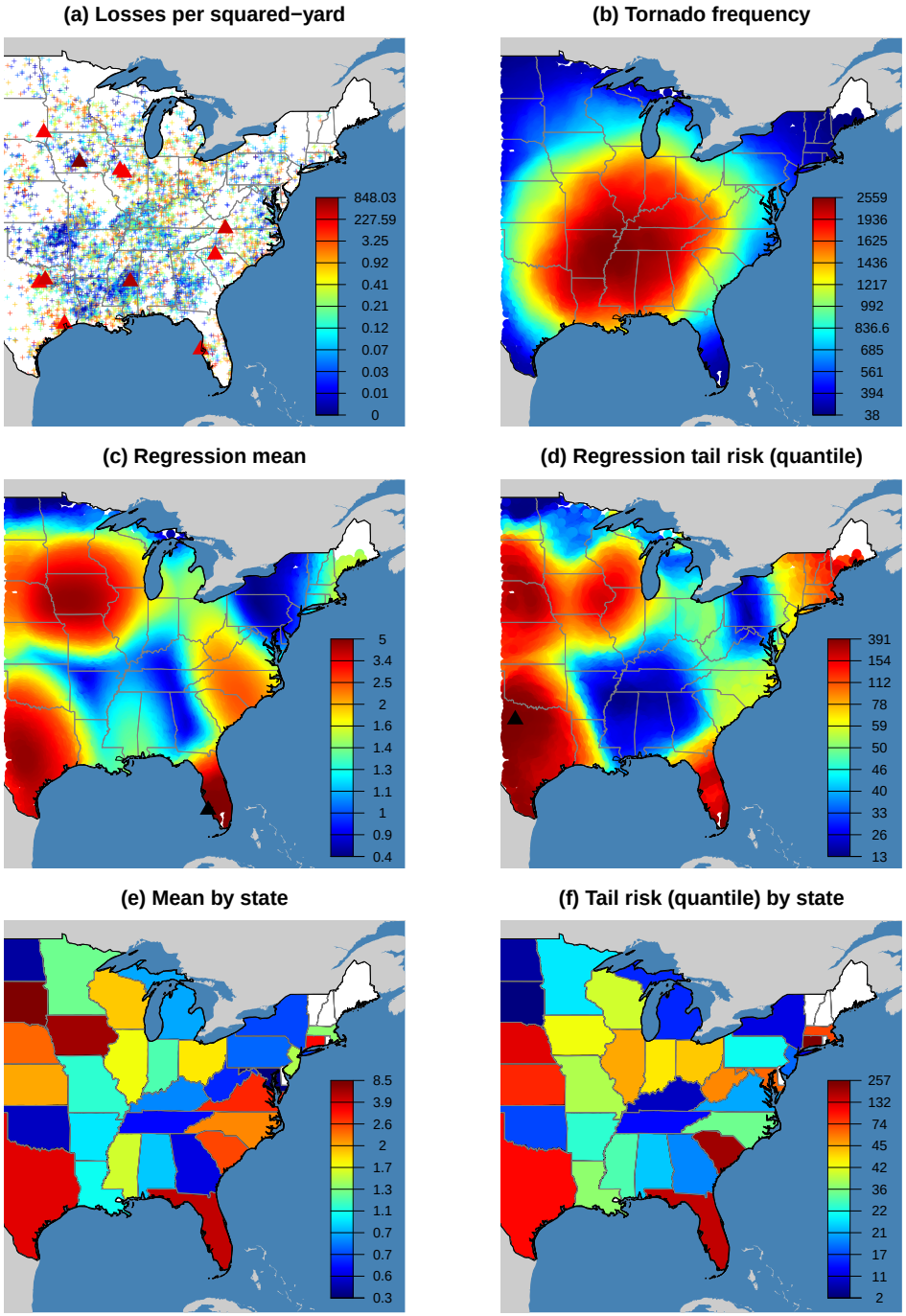

FIG 2. *Tornado losses data. Top row, left: Data across the eastern half of the US, right: Local number of observations in the $h_{n,\star}-$ball. Middle row, left: Estimated conditional mean of losses per squared yard, right: Extrapolated conditional quantile estimate of those losses at level $\tau_n' = 0.995$. 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 [19]. 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).*

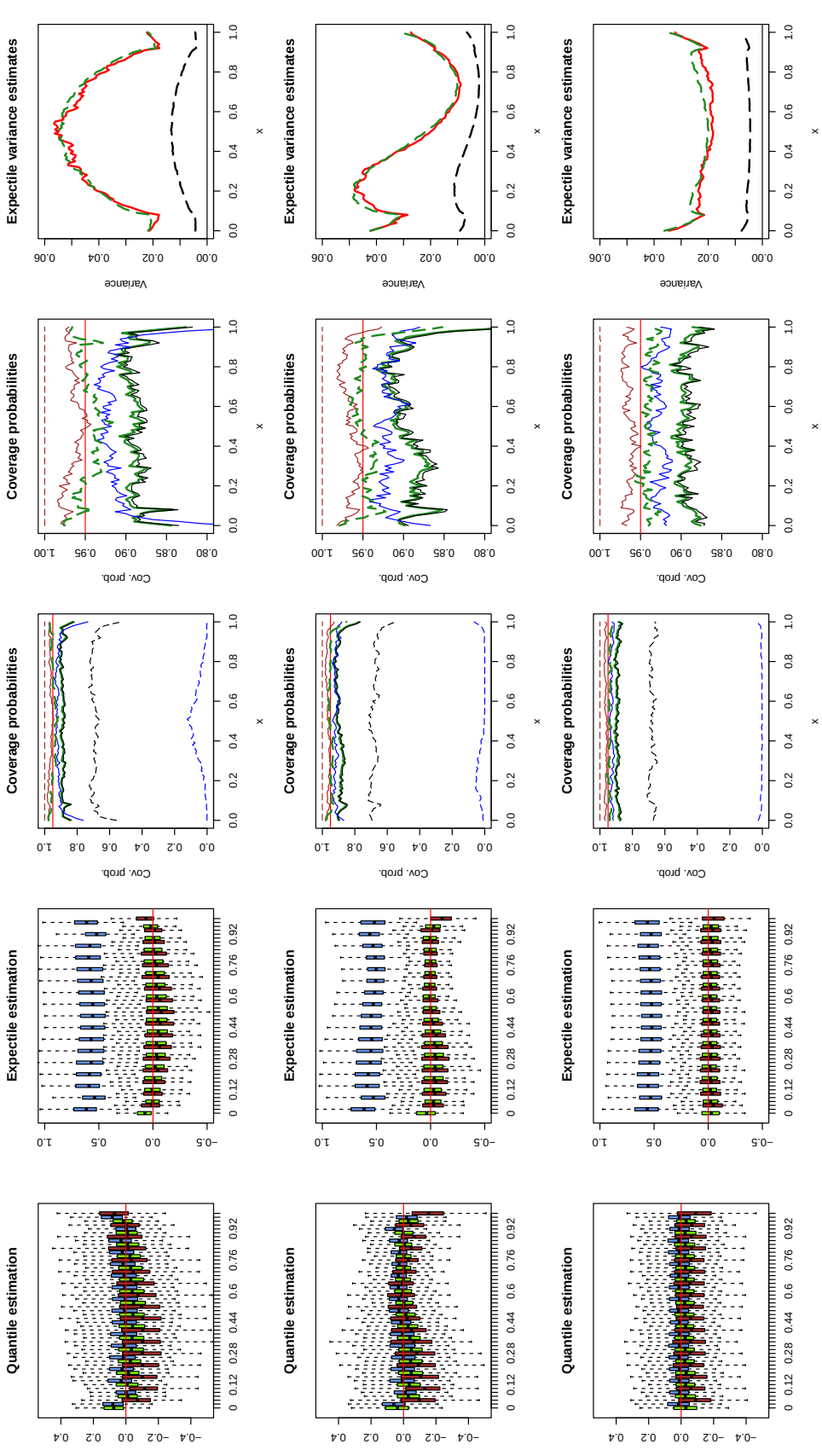

FIG 3. *Simulation results in dimension $p = 1$. From left to right, boxplots of $\log(\widehat{q}^{W,BR}_{1-k_{n,\star}/n}(\tau'_n|x)/q(\tau'_n|x))$ (green), $\log(\widehat{q}^{W}_{1-k_{n,\star}/n}(\tau'_n|x)/q(\tau'_n|x))$ (blue) and $\log(\widehat{q}_n(\tau'_n|x)/q(\tau'_n|x))$ (brown); boxplots of $\log(\widehat{e}^{W,BR}_{1-k_{n,\star}/n}(\tau'_n|x)/e(\tau'_n|x))$ (green), $\log(\widehat{e}^{W}_{1-k_{n,\star}/n}(\tau'_n|x)/e(\tau'_n|x))$ (blue) and $\log(\widehat{e}_n(\tau'_n|x)/e(\tau'_n|x))$ (brown); empirical pointwise coverage probabilities of the asymptotic 95% confidence intervals $\widehat{I}_{q,1}(\tau'_n|x)$ (full black line), $\widehat{I}_{q,2}(\tau'_n|x)$ (full green line), $\widehat{I}_{q,3}(\tau'_n|x)$ (full blue line), $\widehat{I}_{q,4}(\tau'_n|x)$ (full brown line), $\widehat{I}_{E,1}(\tau'_n|x)$ (dashed black line), $\widehat{I}_{E,2}(\tau'_n|x)$ (dashed green line), $\widehat{I}_{E,3}(\tau'_n|x)$ (dashed blue line) and $\widehat{I}_{E,4}(\tau'_n|x)$ (dashed brown line), with the target 95% nominal level in full red line; same figure, zooming around nominal coverage; comparison of the pointwise empirical variances of the $(\sqrt{k_{n,\star}}h_{n,\star}/\log(k_{n,\star}/(n(1-\tau'_n))))\log(\widehat{e}^{W,BR}_{1-k_{n,\star}/n}(\tau'_n|x)/e(\tau'_n|x))$ (full red line), the pointwise median of the corrected variance estimates $(\int_{\mathbb{R}} K^2/\widehat{g}_n(x))\widehat{v}_E(x)$ (dashed green line) and the pointwise median of the naïve variance estimates $(\int_{\mathbb{R}} K^2/\widehat{g}_n(x))\widehat{v}_E(x)$ (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. The authors acknowledge an anonymous Associate Editor and three anonymous reviewers for their helpful comments that led to a much improved article.

Funding. Support from the ANR (grants ANR-19-CE40-0013 and ANR-17-EURE-0010) and the Centre Henri Lebesgue (ANR-11-LABX-0020-01) is gratefully acknowledged. A. Daouia and G. Stupfler acknowledge support from the TSE-HEC ACPR Chair and an AXA Research Fund Award.

SUPPLEMENTARY MATERIAL

Supplementary material for "Inference for extremal regression with dependent heavy-tailed data" The supplementary material document [12] contains further details about our technical conditions and an expanded discussion of the rates of pointwise convergence of our estimators. We then provide the proofs of all theoretical results in the main paper and a full analysis of our worked-out regression examples, preceded by auxiliary results and their proofs. We finally provide further details about our bias and variance correction procedures, and extra finite-sample results.

REFERENCES

[1] ARTZNER, P., DELBAEN, F., EBER, J. M. and HEATH, D. (1999). Coherent measures of risk. *Mathematical Finance* **9** 203–228.

[2] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.

[3] BELLINI, F. and DI BERNARDINO, E. (2017). Risk management with expectiles. *The European Journal of Finance* **23** 487–506.

[4] BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* **2** 107–144.

[5] CHAUDHURI, P. (1991). Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis* **39** 246–269.

[6] CHERNOZHUKOV, V. (2005). Extremal quantile regression. *Annals of Statistics* **33** 806–839.

[7] 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** 559–589.

[8] CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and KAJI, T. (2017). Extremal Quantile Regression. In *Handbook of Quantile Regression* (R. Koenker, V. Chernozhukov, X. He and L. Peng, eds.) Chapman and Hall/CRC.

[9] DAOUIA, A., GARDES, L. and GIRARD, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli* **19** 2557–2589.

[10] DAOUIA, A., GARDES, L., GIRARD, S. and LEKINA, A. (2011). Kernel estimators of extreme level curves. *TEST* **20** 311–333.

[11] 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** 263–292.

[12] DAOUIA, A., STUPFLER, G. and USSEGLIO-CARLEVE, A. (2023). Supplement to "Inference for extremal regression with dependent heavy-tailed data". DOI: 10.1214/[provided by typesetter].

[13] DAVISON, A. C., PADOAN, S. A. and STUPFLER, G. (2023). Tail risk inference via expectiles in heavy-tailed time series. *Journal of Business and Economic Statistics* **41** 876–889.

[14] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.

[15] DREES, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* **9** 617–657.

[16] GIRARD, S., STUPFLER, G. and USSEGLIO-CARLEVE, A. (2021). Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. *Annals of Statistics* **49** 3358–3382.

[17] GIRARD, S., STUPFLER, G. and USSEGLIO-CARLEVE, A. (2022). Nonparametric extreme conditional expectile estimation. *Scandinavian Journal of Statistics* **49** 78–115.

[18] GIRARD, S., STUPFLER, G. and USSEGLIO-CARLEVE, A. (2022). On automatic bias reduction for extreme expectile estimation. *Statistics & Computing* **32** 64.

[19] GOMES, M. I. and PESTANA, D. (2007). A sturdy reduced-bias extreme quantile (VaR) estimator. *Journal of the American Statistical Association* **102** 280–292.
[20] JONES, M. C. (1994). Expectiles and M-quantiles are quantiles. *Statistics & Probability Letters* **20** 149–153.
[21] LINTON, O. and XIAO, Z. (2013). Estimation of and inference about the Expected Shortfall for time series with infinite variance. *Econometric Theory* **29** 771–807.
[22] NEWEY, W. K. and POWELL, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55** 819–847.
[23] SMITH, R. L. and WEISSMAN, I. (1996). Characterization and Estimation of the Multivariate Extremal Index. Technical Report, University of North Carolina.
[24] USSEGLIO-CARLEVE, A. (2018). Estimation of conditional extreme risk measures from heavy-tailed elliptical random vectors. *Electronic Journal of Statistics* **12** 4057–4093.
[25] 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** 1453–1464.
[26] WASSERMAN, L. (2006). *All of Nonparametric Statistics*. Springer.
[27] WEISSMAN, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association* **73** 812–815.
[28] YEE, T. W. and WILD, C. J. (1996). Vector generalized additive models. *Journal of the Royal Statistical Society: Series B* **58** 481–493.
[29] ZIEGEL, J. F. (2016). Coherence and elicitability. *Mathematical Finance* **26** 901–918.