Extremile Regression

Abdelaati Daouia, Irène Gijbels and Gilles Stupfler*†

Toulouse School of Economics, France

KU Leuven, Belgium

ENSAI & CREST, France

Abstract

Regression extremiles define a least squares analogue of regression quantiles. They are determined by weighted expectations rather than tail probabilities. Of special interest is their intuitive meaning in terms of expected minima and maxima. Their use appears naturally in risk management where, in contrast to quantiles, they fulfill the coherency axiom and take the severity of tail losses into account. In addition, they are comonotonically additive and belong to both the families of spectral risk measures and concave distortion risk measures. This paper provides the first detailed study exploring implications of the extremile terminology in a general setting of presence of covariates. We rely on local linear (least squares) check function minimization for estimating conditional extremiles and deriving the asymptotic normality of their estimators. We also extend extremile regression far into the tails of heavy-tailed distributions. Extrapolated estimators are constructed and their asymptotic theory is developed. Some applications to real data are provided.

Keywords: Asymmetric least squares; Extremes; Heavy tails; Regression extremiles; Regression quantiles; Tail index.

*Corresponding author.

†Abdelaati Daouia is Maître de Conférences, Université Toulouse 1 Capitole, TSE, 21 allée de Brienne, Manufacture des Tabacs, 31000 Toulouse, France (E-mail: abdelaati.daouia@tse-fr.eu). Irène Gijbels is Professor of Statistics, KU Leuven, Statistics and Risk Section, Celestijnenlaan, 200b - box 2400, 3001 Leuven, Belgium (E-mail: Irene.Gijbels@kuleuven.be). Gilles Stupfler is Lecturer in Statistics, ENSAI & CREST, Campus de Ker Lann, 51 Rue Blaise Pascal, 35172 Bruz Cedex (E-mail: gilles.stupfler@ensai.fr).

1 Introduction

A basic tool in different scientific fields for analyzing the impact of a set of regressors X on the distribution of a response Y is quantile regression. For $\tau \in (0,1)$, the conditional τth quantile of Y given $X = x$ is defined as the minimizer

$$q_\tau(x) \in \arg\min_{\theta \in \mathbb{R}} \mathbb{E}\left\{ [|\tau - \mathbb{1}(Y \leqslant \theta)| \cdot |Y - \theta| - |\tau - \mathbb{1}(Y \leqslant 0)| \cdot |Y|] \mid X = x \right\}, \tag{1}$$

with $\mathbb{1}(\cdot)$ being the indicator function. Subtracting $|\tau - \mathbb{1}(Y \leqslant 0)| \cdot |Y|$ in the expectation makes the integrand well-defined and finite without assuming $\mathbb{E}(|Y||X = x) < \infty$. A disadvantage of quantile regression is that quantiles only use the information on whether an observation is below or above some specific value. However, in a financial risk management context for example, not taking into account the effective magnitude of high values of losses, might not be wise. Conditional expectiles deal with this drawback, and lead to coherent and more realistic risk measures as compared to quantile-based risk measures, as evidenced by [1] and [4], among others. The conditional τth expectile is defined as

$$e_\tau(x) = \arg\min_{\theta \in \mathbb{R}} \mathbb{E}\left\{ \left[|\tau - \mathbb{1}(Y \leqslant \theta)| \cdot |Y - \theta|^2 - |\tau - \mathbb{1}(Y \leqslant 0)| \cdot |Y|^2\right] \mid X = x \right\}, \tag{2}$$

obtained in a similar way to $q_\tau(x)$ in (1) but replacing absolute deviations by squared deviations (Newey and Powell [8]). Expectiles depend on both the tail realizations and their probability, while quantiles only depend on the frequency of tail observations. An inconvenience of expectiles is their lack of transparent interpretation, due to the absence of a closed form expression of $e_\tau(x)$ as a solution to the asymmetric least squares problem (2), for all $\tau \neq \frac{1}{2}$. The absence of an explicit expression makes the treatment of expectiles a hard mathematical problem from the perspective of extreme value theory, for instance when it comes to estimating tail risk (Daouia *et al.* [4]).

Very recently, Daouia *et al.* [3] considered an alternative class to expectiles, called *extremiles*, which defines a new least squares analogue of quantiles. A starting point for the introduction of this class was that the unconditional τth quantile of Y, with continuous cumulative distribution function F, can alternatively be obtained from

$$q_\tau \in \arg\min_{\theta \in \mathbb{R}} \mathbb{E}\left\{ J_\tau(F(Y)) \cdot [|Y - \theta| - |Y|] \right\}, \tag{3}$$

where $J_\tau(\cdot) = K'_\tau(\cdot)$, with

$$K_\tau(t) = \begin{cases} 1 - (1-t)^{s(\tau)} & \text{if} \quad 0 < \tau \leqslant 1/2 \\ t^{r(\tau)} & \text{if} \quad 1/2 \leqslant \tau < 1 \end{cases} \tag{4}$$

being a distribution function with support $[0, 1]$, and $r(\tau) = s(1-\tau) = \log(1/2)/\log(\tau)$. See Section 2.1 in [3]. The *unconditional extremile of order* τ is then defined by substituting the absolute deviations with squared deviations, *i.e.*

$$\xi_\tau = \underset{\theta \in \mathbb{R}}{\arg\min}\, \mathbb{E}\left\{J_\tau(F(Y)) \cdot \left[|Y-\theta|^2 - |Y|^2\right]\right\}. \tag{5}$$

Unlike expectiles, extremiles can be motivated via several angles and enjoy various interpretations and closed form expressions. For an overview on this issue, and the specific merits related to these interpretations and explicit expressions, see Daouia *et al.* [3]. In the presence of covariates, one can define conditional extremiles by considering a conditional version of (5). It will be evidenced in Section 2 that conditional extremiles enjoy the same advantages as unconditional extremiles. Obviously statistical inference for conditional quantities, such as conditional quantiles, expectiles and extremiles, requires specific regression tools as compared to statistical inference for their unconditional counterparts.

The aim of this paper is to study conditional extremiles, *i.e.* to pursue extremile regression, in a general setting. The main contributions of this paper consist of (i) discussing probabilistic properties of regression extremiles; (ii) studying and establishing the asymptotic behaviour of their nonparametric estimators; (iii) investigating conditional extremile estimators when applied to the far tail (case $\tau = \tau_n \to 1$, as the sample size $n \to \infty$); and (iv) illustrating the practical use of noncentral conditional extremiles. We shall discuss below in Section 5 the various merits of extremile regression.

The paper is organized as follows. Section 2 presents the class of regression extremiles and their basic probabilistic properties. Section 3 deals with estimation of ordinary conditional extremiles for fixed orders τ. Extrapolated estimators of tail regression extremiles, for high orders $\tau = \tau_n \to 1$ as $n \to \infty$, are developed in Section 4 for heavy-tailed conditional distributions. Section 5 concludes. All the necessary mathematical proofs, practical implementation guidelines and simulation results are given in the supplementary file.

2 Class of regression extremiles

Let $X \in \mathbb{R}^d$ and $Y \in \mathbb{R}$ be two random variables. Denote by $F(\cdot|x)$ the cumulative distribution function of Y given $X = x$ and by $q_\tau(x) = F^{-1}(\tau|x) = \inf\{y \in \mathbb{R}|F(y|x) \geqslant \tau\}$ the related conditional quantile of order $\tau \in (0,1)$. For ease of presentation, we assume throughout the paper that $F(\cdot|x)$ is continuous. The order-τ extremile of this distribution function, as introduced in (5), defines the regression τth extremile of Y given $X = x$.

Definition 1 *Let Y given $X = x$ have a finite absolute first moment. Then, for any $\tau \in (0,1)$, the conditional order-τ extremile of Y given $X = x$ is*

$$\xi_\tau(x) = \underset{\theta \in \mathbb{R}}{\arg\min}\, \mathbb{E}\left\{ J_\tau\left(F(Y|X)\right) \cdot [|Y-\theta|^2 - |Y|^2] | X = x \right\}. \tag{6}$$

Particularly useful is to look at $\xi_\tau(x)$ as the following probability-weighted moment, expected maximum or expected minimum.

Proposition 1 *Let Y given $X = x$ have a finite absolute first moment. Then, for any $\tau \in (0,1)$, we have the following equivalent closed form expressions:*

$$\begin{aligned} \xi_\tau(x) &= \mathbb{E}\left[Y\, J_\tau\left(F(Y|X)\right)|X = x\right] = \int_0^1 J_\tau(t)\, q_t(x)\, dt = \int_0^1 q_t(x)\, dK_\tau(t), \\ \textit{and}\;\; \xi_\tau(x) &= \begin{cases} \mathbb{E}\left[\max\left(Y_x^1, \ldots, Y_x^r\right)\right] & \textit{when} \quad \tau = (1/2)^{1/r} \quad \textit{with} \;\; r \in \mathbb{N}\backslash\{0\}, \\ \mathbb{E}\left[\min\left(Y_x^1, \ldots, Y_x^s\right)\right] & \textit{when} \quad \tau = 1-(1/2)^{1/s} \quad \textit{with} \;\; s \in \mathbb{N}\backslash\{0\}, \end{cases} \end{aligned}$$

for independent observations Y_x^i drawn from the conditional distribution of Y given $X = x$.

In the central case $\tau = 1/2$, we have $r(\tau) = s(\tau) = 1$, and hence $\xi_\tau(x)$ reduces to the standard regression mean $\mathbb{E}(Y|X = x)$. The limit cases $\tau \uparrow 1$ (*i.e.* $r(\tau) \to \infty$) and $\tau \downarrow 0$ (*i.e.* $s(\tau) \to \infty$) lead to the upper and, respectively, lower endpoints of the support of $F(\cdot|x)$. Further important properties are established in the following.

Proposition 2 *(i) If Y given $X = x$ has a symmetric distribution with finite absolute first moment, then $\xi_{1-\tau}(x) = 2\,\mathbb{E}(Y|X = x) - \xi_\tau(x)$, for any $\tau \in (0,1)$.*

(ii) If $Y = m(X) + \sigma(X)\varepsilon$, where $\sigma(X) > 0$ and ε is independent of X and has a finite absolute first moment, then $\xi_\tau(x) = m(x) + \sigma(x)\xi_{\tau,\varepsilon}$, for any $\tau \in (0,1)$, where $\xi_{\tau,\varepsilon}$ denotes the τth extremile of ε.

An implication of Proposition 2 is that, for symmetric conditional distributions, the lower and upper extremile curves are symmetric about the regression mean. Also, the extremile curves are parallel to each other if the distribution of the response is homogeneous. These properties hold for conditional quantiles as well.

3 Estimation method

Our approach is a local linear estimation based on the definition (6) which is of particular relevance when considering flexible regression specifications such as local polynomial approximations. We restrict our analysis here to one-dimensional covariates X ($d = 1$).

3.1 Least squares kernel smoothing

For a generic estimator $\widehat{F}(\cdot|x)$ of $F(\cdot|x)$, the local linear check function minimization solves the weighted least squares problem

$$\underset{(\alpha,\beta)\in\mathbb{R}^2}{\arg\min} \sum_{i=1}^{n} J_\tau\left(\widehat{F}(Y_i|x)\right)\{Y_i - \alpha - \beta(x - X_i)\}^2 L\left(\frac{x - X_i}{h_n}\right) \tag{7}$$

to get the estimators $\breve{\alpha} = \breve{\xi}_{\mathrm{LL},\tau}(x)$ and $\breve{\beta} = \breve{\xi}'_{\mathrm{LL},\tau}(x)$ of $\xi_\tau(x)$ and $\xi'_\tau(x)$, respectively, where $L(\cdot)$ is a kernel function and $h_n > 0$ a bandwidth sequence. Standard weighted least squares theory leads to the following explicit solution

$$\begin{pmatrix} \breve{\alpha} \\ \breve{\beta} \end{pmatrix} = \left(\mathbf{X}_{\mathrm{LL}}^T \mathbf{W}_{\widehat{F},L} \mathbf{X}_{\mathrm{LL}}\right)^{-1} \mathbf{X}_{\mathrm{LL}}^T \mathbf{W}_{\widehat{F},L} \mathbf{Y},$$

where $\mathbf{Y}$ is the column vector of dimension n containing all Y_i, $i = 1, \ldots, n$, and $\mathbf{X}_{\mathrm{LL}}$ is the usual design matrix of the local linear fitting technique, *i.e.* the $n \times 2$ matrix with a vector of 1's as a first column, and where the second column consists of the values $x - X_i$, $i = 1, \ldots, n$. Furthermore, the weight matrix in the weighted least squares problem is

$$\mathbf{W}_{\widehat{F},L} = \mathrm{diag}\left(J_\tau\left(\widehat{F}(Y_i|x)\right) L\left(\frac{x - X_i}{h_n}\right)\right)_{i=1,\ldots,n}.$$

Clearly, the asymptotic behavior of $\widehat{F}(\cdot|x)$ will be crucial to the analysis of the asymptotic and finite-sample behavior of $\breve{\xi}_{\mathrm{LL},\tau}(x)$. Let us first discuss the properties of the latter estimator under some general high-level conditions, including the following assumptions:

(C1) The random vector (X, Y) has a joint density $f_{(X,Y)}$ which is twice continuously differentiable in its first argument and such that for each x_0, we can write

$$\sup_{x \in U} \left\{ f_{(X,Y)}(x, y) + |\partial_x f_{(X,Y)}(x, y)| + |\partial^2_{xx} f_{(X,Y)}(x, y)| \right\} \leqslant h(y)$$

for some neighborhood U of x_0, where h is a nonnegative measurable function satisfying $\int_{\mathbb{R}} (1 + |y|^{2+\delta}) h(y)\, dy < \infty$ for some $\delta > 0$;

(C2) The density f_X of X is continuous and positive on the interior of its support;

(C3) The kernel L is a symmetric and bounded density function with compact support.

Theorem 1 *Assume that conditions (C1)–(C3) hold, and that $\widehat{F}(\cdot|x)$ is a uniformly consistent estimator of $F(\cdot|x)$ satisfying*

$$\begin{aligned} &\frac{1}{\sqrt{nh_n}} \sum_{i=1}^{n} J_\tau\left(\widehat{F}(Y_i|x)\right) \{Y_i - \alpha - \beta(x - X_i)\} \left(\frac{x - X_i}{h_n}\right)^j L\left(\frac{x - X_i}{h_n}\right) \\ &= \frac{1}{\sqrt{nh_n}} \sum_{i=1}^{n} J_\tau\left(F(Y_i|x)\right) \{Y_i - \alpha - \beta(x - X_i)\} \left(\frac{x - X_i}{h_n}\right)^j L\left(\frac{x - X_i}{h_n}\right) + o_{\mathbb{P}}(1) \quad (8) \end{aligned}$$

locally uniformly in $(\alpha, \beta) \in \mathbb{R}^2$, for $j = 0, 1$. Let $h_n \to 0$ be such that $nh_n^5 \to 0$, as $n \to \infty$. Then, for any x interior to the support of X,

$$\sqrt{nh_n} \left\{ \check{\xi}_{LL,\tau}(x) - \xi_\tau(x) \right\} \xrightarrow{d} \mathcal{N}\left(0, \frac{\|L\|_2^2}{f_X(x)} V_\tau(x)\right), \quad \text{as} \quad n \to \infty, \tag{9}$$

where $V_\tau(x) = \mathbb{E}\left[J_\tau^2\left(F(Y|x)\right) \{Y - \xi_\tau(x)\}^2 \mid X = x\right]$.

Assumption (8) is evidently the central condition to be checked as part of Theorem 1. It ensures that the asymptotic analysis of $\check{\xi}_{\mathrm{LL},\tau}(x)$ can be performed by replacing $\widehat{F}(\cdot|x)$ with $F(\cdot|x)$. It should also be noted that in the higher-dimensional setting with $d-$dimensional covariates X, nonparametric estimators converge in general at the rate $\sqrt{nh^d}$, for a given bandwidth sequence $h = h_n \to 0$. The associated bias condition, which makes it possible to find the optimal rate of convergence of the estimator, is typically $nh^{d+4} \to c \in (0, \infty)$. This is realized for h having order $n^{-1/(d+4)}$, resulting in the optimal convergence rate $n^{2/(d+4)}$. This gets slower as d grows, and is an example of the well-known curse of dimensionality phenomenon.

The next corollary gives a simpler result for estimators $\widehat{F}(\cdot|x)$ having the typical rate of convergence $n^{2/5}$. Examples of such estimators include the traditional Nadaraya-Watson estimator [7], the nearest-neighbor estimator [9] and the (improved) local linear estimator [5]. For those estimators $\widehat{F}(\cdot|x)$, we derive the asymptotic normality of $\breve{\xi}_{\mathrm{LL},\tau}(x)$ when estimating noncentral regression extremiles $\xi_\tau(x)$.

Corollary 1 *Assume that conditions (C1)–(C3) hold. Let $\widehat{F}(\cdot|x)$ be an estimator of $F(\cdot|x)$ satisfying $n^{2/5}\sup_{y\in\mathbb{R}}\left|\widehat{F}(y|x)-F(y|x)\right| = \mathrm{O}_{\mathbb{P}}(1)$. Let finally $h_n \to 0$ be such that $nh_n^5 \to 0$, as $n\to\infty$. Then, with the notation of Theorem 1, for any x interior to the support of X and any $\tau \in (0, 1-1/\sqrt{2}] \cup [1/\sqrt{2}, 1)$, we have the convergence (9).*

The condition $\tau \in (0, 1-1/\sqrt{2}] \cup [1/\sqrt{2}, 1)$ should not be viewed as a restriction in practice. Indeed, by Proposition 1, regression extremiles in the right tail ($\tau \geqslant 1/2$) are most easily interpreted when the power $r(\tau) = \log(1/2)/\log(\tau)$ in (4) is an integer, since then $\xi_\tau(x) = \mathbb{E}\left[\max\left(Y_x^1,\ldots,Y_x^{r(\tau)}\right)\right]$, for independent observations Y_x^i drawn from the conditional distribution of Y given $X = x$. In this case, the condition $\tau \in [1/\sqrt{2}, 1)$ is equivalent to $r(\tau) \geqslant 2$, and hence all expected maxima and corresponding extremiles are covered by this condition, except for the conditional expectation $\xi_{1/2}(x) = \mathbb{E}(Y|X=x)$ whose estimation obviously does not require extremile regression. Likewise, regression extremiles in the left tail ($\tau \leqslant 1/2$) are interpreted as $\xi_\tau(x) = \mathbb{E}\left[\min\left(Y_x^1,\ldots,Y_x^{r(1-\tau)}\right)\right]$ when $r(1-\tau) \in \mathbb{N}\backslash\{0\}$. In this case, the condition $\tau \in (0, 1-1/\sqrt{2}]$ is equivalent to $r(1-\tau) \geqslant 2$, and so apart from $\xi_{1/2}(x) = \mathbb{E}(Y|X=x)$, all expected minima and corresponding extremiles are covered by this condition.

3.2 Empirical data examples

We now illustrate the usefulness of extremile regression on two real datasets about triceps skinfold variation and motorcycle insurance payouts. The first dataset *'dataTriceps'*, kindly sent by Keming Yu, comprises triceps skinfold measurements of 892 girls and women up to age 50, recorded in three Gambian villages during the dry season of 1989. To understand the evolution of triceps skinfold with age, Yu and Jones [12] proposed to look at a collection of estimated quantiles as a function of age. The obtained fits via local linear check function minimization are graphed in Figure 1 (left panel). To calculate these conditional quantiles, we used the locally polynomial quantile regression function `lprq` of the R package `quantreg`, in conjunction with the optimal bandwidth h_{q_τ} chosen by the Yu and Jones [12] selection method. The competing conditional extremiles $\breve{\xi}_{\mathrm{LL},\tau}$ in (7), obtained with the bandwidth h_{ξ_τ} from our automatic selection strategy developed in Section B of the supplementary file, are given in the same figure (left panel) in solid lines, along with some 95% pointwise asymptotic confidence intervals in dashed lines. To calculate the probability weights $J_\tau\left(\widehat{F}(Y_i|x)\right)$ in $\breve{\xi}_{\mathrm{LL},\tau}(x)$, we used in all our examples the local linear estimator $\widehat{F}(\cdot|x)$. In the absence of a rule-of-thumb bandwidth selection for estimated expectiles via the check function method of Yao and Tong [11], we superimpose in the same figure (left panel) the expectile curves corresponding to h_{q_τ} (dashed lines) and those corresponding to h_{ξ_τ} (solid lines); the difference between the resulting expectile curves is negligible although h_{q_τ} and h_{ξ_τ} are appreciably different for each τ.

The messages yielded by the three regression methods are broadly similar, indicating particularly that adulthood corresponds to a much greater variability in triceps skinfold compared to childhood. Still, expectiles beyond the regression mean exhibit less evidence of the obvious variation and over-dispersion of the triceps skinfold as age increases: the widening of extreme expectiles seems to be rather "narrow". By contrast, there is a distinct tendency for the noncentral extremiles and quantiles to be more spread, suggesting better capability of fitting both location and sparseness in data points. That said, extremile regression seems to be beneficial at least in producing smoother and more pleasing fits of conditional location and spread beyond the regression mean. Of course, the quantile curves can be smoothed by resorting to local linear double-kernel smoothing, but this is unnecessary for extremiles. Moreover, we are not aware of any ready-made procedure for constructing asymptotic confidence intervals of conditional quantiles and expectiles based directly on the limit distributions of their local linear estimators.

The advantages of extremile regression at the tails become even more pronounced when considering heavy-tailed scenarios as is the case in most actuarial and financial applications. The second dataset *'dataOhlsson'*, available from the R package `insuranceData`, contains 670 motorcycle-related claims recorded from 1994 to 1998 by the Swedish insurer Wasa. The scatterplot and local linear fits are given in Figure 1 (right panel). Here, tail regression extremiles show more alertness and reactivity to unexpected high losses than their expectile counterparts. They also exhibit better smoothness and stability than their quantile competitors and do not show any crossing, unlike the unpleasant quantile crossings that are incompatible with what occurs at the population level.

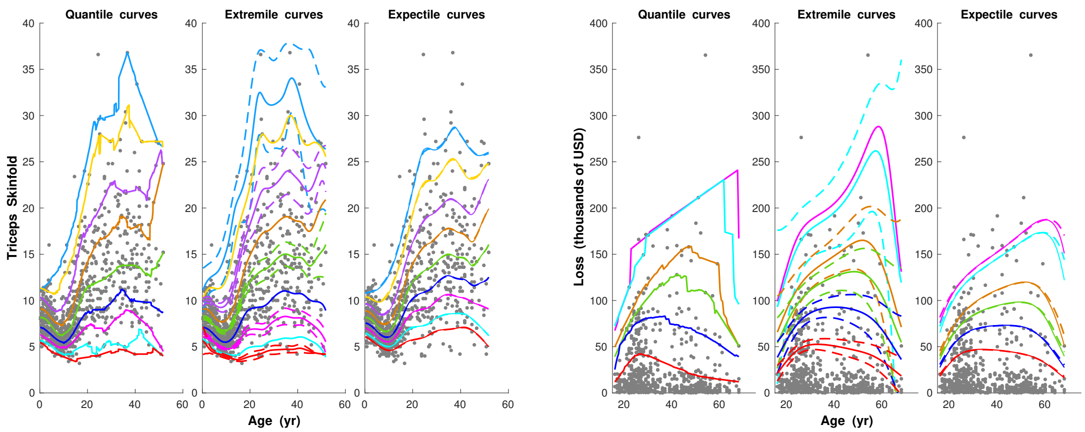

Figure 1: *Left panel: dataTriceps, with smoothed* 1%, 3%, 10%, 25%, 50%, 75%, 90%, 97% *and* 99% *quantile (left), extremile (middle) and expectile curves (right) in solid lines, and* 95% *pointwise asymptotic confidence intervals for* $\xi_{.01}$, $\xi_{.1}$, $\xi_{.5}$, $\xi_{.9}$, $\xi_{.99}$ *in dashed lines. Right panel: dataOhlsson, with smoothed* 75%, 90%, 95%, 97%, 99% *and* 99.2% *regression curves (solid), and confidence intervals for* $\xi_{.75}$, $\xi_{.9}$, $\xi_{.95}$, $\xi_{.97}$ *and* $\xi_{.99}$ *(dashed).*

Note that Proposition 1 provides a straightforward interpretation of the regression extremile $\xi_\tau(x)$ by making use of the asymmetry level forms $\tau = (1/2)^{1/r(\tau)}$, for $\tau \geqslant 1/2$, and $\tau = 1-(1/2)^{1/r(1-\tau)}$ for $\tau \leqslant 1/2$. Intuitively, for example in the case of motorcycle insurance claims, in the right tail, $\xi_\tau(x) \equiv \mathbb{E}[\max(Y_x^1, \ldots, Y_x^{r(\tau)})]$ gives the expected maximum claim amount among $r(\tau)$ potential claimants aged x years, with $r(.97) \approx 22.75$, $r(.99) \approx 68.96$, $r(.992) \approx 86.29$, $r(.993) \approx 98.67$, and $r(.994) \approx 115.17$. Interestingly, the regression quantile of the same order τ has the "dual" intuitive meaning as $q_\tau(x) \equiv \text{Median}[\max(Y_x^1, \ldots, Y_x^{r(\tau)})]$. For a general level τ, we still keep the intuitive meaning of $\xi_\tau(x)$ as an expected maximum on the right tail ($\frac{1}{2} \leqslant \tau < 1$) in the sense that

$$\mathbb{E}\left[\max\left(Y_x^1, \ldots, Y_x^{\lfloor r(\tau) \rfloor}\right)\right] \leqslant \xi_\tau(x) \leqslant \mathbb{E}\left[\max\left(Y_x^1, \ldots, Y_x^{\lfloor r(\tau) \rfloor + 1}\right)\right],$$

where $\lfloor \cdot \rfloor$ denotes the floor function.

4 Extremal regression

In this section, we focus on *extremal regression* of a response variable $Y \in \mathbb{R}$ given a vector of covariates $X \in \mathbb{R}^d$. This translates into considering the order $\tau = \tau_n' \to 1$ or $\tau_n' \to 0$ as the sample size n goes to infinity. To ease the presentation, we restrict our extreme-value analysis to the case $\tau \to 1$. Similar considerations evidently apply to the left tail $\tau \to 0$.

4.1 Model assumption

We assume for the sake of simplicity that the response Y given $X = x$ is positive and $\mathbb{E}(Y|X = x) < \infty$. We focus on the challenging domain of attraction of heavy-tailed conditional distributions that better describe the tail structure and sparseness of the data in most applications in financial and natural sciences [2, 6, 10]. More precisely, we assume that the conditional tail quantile function $t \mapsto q_{1-t^{-1}}(x)$ is second-order regularly varying:

$$\text{(E)} \quad \forall y > 0, \ \lim_{t\to\infty} \frac{1}{A(t|x)} \left(\frac{q_{1-(ty)^{-1}}(x)}{q_{1-t^{-1}}(x)} - y^{\gamma(x)} \right) = y^{\gamma(x)} \frac{y^{\rho(x)} - 1}{\rho(x)}$$

for some parameters $0 < \gamma(x) < 1$, $\rho(x) \leqslant 0$ and an auxiliary function $A(\cdot|x)$ having constant sign, with $A(t|x) \to 0$ as $t \to \infty$. We use throughout the convention that $(y^b - 1)/b = \log y$ for $b = 0$, so that the right-hand side reads $y^{\gamma(x)} \log y$ if the second-order parameter $\rho(x)$ is zero. The index $\gamma(x) > 0$ tunes the tail heaviness of the conditional distribution of Y given $X = x$, with higher positive values indicating heavier conditional tails. The assumption $\gamma(x) < 1$ is tailored to our requirement that $\mathbb{E}(Y|X = x) < \infty$.

4.2 Estimation procedure and main results

Here we consider the estimation of $\xi_\tau(x)$ when $\tau = \tau_n' \uparrow 1$ at an arbitrary rate as $n \to \infty$. Under assumption (E), we have by Proposition 3 of [3], applied to the conditional distribution of Y given $X = x$, that $\xi_{\tau_n'}(x) \sim q_{\tau_n'}(x)\, \mathcal{G}\left(\gamma(x)\right)$ as $n \to \infty$, where $\mathcal{G}(s) := \Gamma(1-s)\{\log 2\}^s$ and Γ is the Gamma function. This motivates the estimator

$$\widehat{\xi}^{\star}_{\tau_n'}(x) := \widehat{q}^{\star}_{\tau_n'}(x)\, \mathcal{G}\left(\widehat{\gamma}(x)\right) \tag{10}$$

obtained by substituting in suitable estimators $\widehat{q}^{\star}_{\tau_n'}(x)$ of $q_{\tau_n'}(x)$ and $\widehat{\gamma}(x)$ of $\gamma(x)$. Nonparametric local estimates of the tail quantities $q_{\tau_n'}(x)$ and $\gamma(x)$ have been proposed in the last decade by [2, 6, 10], among others. Prominent among these contributions is the Weissman quantile-type estimator

$$\widehat{q}^{\star}_{\tau_n'}(x) \equiv \widehat{q}^{\star}_{\tau_n',\tau_n}(x) := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}(x)} \widehat{q}_{\tau_n}(x), \tag{11}$$

where $\widehat{\gamma}(x)$ and $\widehat{q}_{\tau_n}(x)$ are consistent estimators of $\gamma(x)$ and $q_{\tau_n}(x)$, with $\tau_n < \tau_n'$ being a tuning sequence to be selected jointly with h_n. Combining (10) and (11), we arrive at

$$\widehat{\xi}^{\star}_{\tau_n'}(x) \equiv \widehat{\xi}^{\star}_{\tau_n',\tau_n}(x) = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}(x)} \widehat{q}_{\tau_n}(x)\, \mathcal{G}\left(\widehat{\gamma}(x)\right). \tag{12}$$

In Theorem A.1 in the Supplementary Material document, we establish the asymptotic distribution of $\widehat{\xi}^{\star}_{\tau_n'}(x)$ in its general form (12), for generic estimators $\widehat{\gamma}(x)$ and $\widehat{q}_{\tau_n}(x)$, under standard assumptions in the literature of conditional extremes. Here, we specialize the discussion to well-specified estimators $\widehat{q}_{\tau_n}(x)$ and $\widehat{\gamma}(x)$ in the generic form (12) of $\widehat{\xi}^{\star}_{\tau_n'}(x)$. We consider the Nadaraya-Watson type estimator $\widehat{q}_{\tau_n}(x) \equiv \widehat{F}^{-1}_{\mathrm{NW}}(\tau_n|x)$, where

$$\widehat{F}_{\mathrm{NW}}(y|x) := \sum_{i=1}^{n} \mathbb{1}(Y_i \leqslant y) L\left(\frac{x-X_i}{h_n}\right) \Bigg/ \sum_{i=1}^{n} L\left(\frac{x-X_i}{h_n}\right). \tag{13}$$

As for the choice of the conditional tail index estimator $\widehat{\gamma}(x)$, we will use in the sequel the notation $\alpha_n := 1-\tau_n$ and $p_n := 1-\tau_n'$, and consider the kernel estimator of [2]:

$$\widehat{\gamma}(x) = \sum_{j=1}^{J} \left[\log \widehat{q}_{1-t_j\alpha_n}(x) - \log \widehat{q}_{1-\alpha_n}(x)\right] \Bigg/ \sum_{j=1}^{J} \log(1/t_j), \tag{14}$$

where $(1 = t_1 > t_2 > \cdots > t_J > 0)$ is a decreasing list of J weights. Note that, unlike [2], we do not assume differentiability of the conditional distribution function, and therefore the distribution of Y given X is allowed to have atoms. The asymptotic normality of the corresponding regression extremile estimator

$$\widehat{\xi}^{\star}_{1-p_n}(x) := \left(\frac{\alpha_n}{p_n}\right)^{\widehat{\gamma}(x)} \widehat{q}_{1-\alpha_n}(x)\, \mathcal{G}\left(\widehat{\gamma}(x)\right)$$

follows under the following additional regularity conditions:

(K1) The functions $1/\gamma$ and f_X are Lipschitz continuous and $x \mapsto \log \ell(y|x)/\log y$, where $\ell(y|x) := y^{1/\gamma(x)}[1 - F(y|x)]$, satisfies, for a norm $\|\cdot\|$ on $\mathbb{R}^d$,

$$\exists c > 0,\ \exists y_0 > 1,\ \sup_{y \geqslant y_0} \left| \frac{\log \ell(y|x)}{\log y} - \frac{\log \ell(y|x')}{\log y} \right| = \sup_{y \geqslant y_0} \frac{1}{\log y} \left| \log \frac{\ell(y|x)}{\ell(y|x')} \right| \leqslant c\|x - x'\|.$$

(K2) The kernel L is a bounded density with support included in the unit ball of $\mathbb{R}^d$.

Theorem 2 *Suppose (E) and (K1)–(K2) hold. Let $x \in \mathbb{R}^d$ be such that $f_X(x) > 0$. Assume further that $\rho(x) < 0$ and, as $n \to \infty$,*

1. *$\alpha_n \to 0$, $nh_n^d\alpha_n \to \infty$ and $nh_n^d p_n \to c < \infty$;*
2. *$\log(\alpha_n/p_n)/\sqrt{nh_n^d\alpha_n} \to 0$, $nh_n^{d+2}\alpha_n \log^2 \alpha_n \to 0$ and $\sqrt{nh_n^d\alpha_n}\, A((1-\alpha_n)^{-1}|x) \to 0$.*

Then we have

$$\frac{\sqrt{nh_n^d\alpha_n}}{\log(\alpha_n/p_n)} \left(\frac{\widehat{\xi}^{\star}_{1-p_n}(x)}{\xi_{1-p_n}(x)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\|L\|_2^2}{f_X(x)} V_J\, \gamma^2(x)\right) \quad \textit{as} \quad n \to \infty,$$

where

$$V_J = \left(\sum_{j=1}^{J} \frac{2(J-j)+1}{t_j} - J^2 \right) \Bigg/ \left(\sum_{j=1}^{J} \log(1/t_j) \right)^2.$$

When choosing, for instance, the harmonic sequence $t_j = 1/j$, the variance of the limiting distribution is minimal for $J = 9$ with $V_9 \approx 1.25$ (see [2]). We shall discuss below concrete applications where $(t_j = 1/j)_{1\leqslant j\leqslant 9}$ are employed with the discretized tuning parameter $\alpha_n = k/n$ for a sequence of integers $k \in [1, n)$. A data-driven method for selecting both k and the bandwidth h_n in practice is described in Supplement C.

4.3 Insurance payouts

This section returns to our motivating data set *'dataOhlsson'* and explores estimation and inference for extreme risk associated with motorcycle insurance claims. It can be seen in Figure 1 that this data ($n = 670$) features heavy tails and data sparsity in the tail areas. Figure 2 (top left) plots the tail index estimates $\widehat{\gamma}(x)$ versus k, for the empirical quartiles x of X. The plot shows stability over the region $k \in [50, 90]$, which suggests to pick out the pointwise estimates $\widehat{\gamma}(x)$ over this interval. The top right panel plots the final estimates $\widehat{\gamma}(x)$ versus x obtained via our data-driven device (red curve), along with the estimates using $k \in \{50, 70, 90\}$. It is remarkable that the automatic selection points towards similar results as these k values from the stable region. The four estimated curves indicate tail indices $\widehat{\gamma}(x) \in [0.25, 0.65]$ which, as expected, reflect a strong conditional tail heaviness. Hereafter in this extremal regression study, we focus mainly on the estimates obtained for $x \leqslant 55$ to mitigate data scarcity and boundary effects beyond this range.

To estimate conditional extremiles $\xi_{\tau_n'}(x)$ at extreme levels, Supplement D gives Monte Carlo evidence that the extrapolated estimates $\widehat{\xi}^{\star}_{\tau_n'}(x)$ in (12) are more efficient relative to the direct estimates $\check{\xi}_{\mathrm{LL},\tau_n'}(x)$ from ordinary local linear regression in (7). For $\tau_n' = .99$, the middle panel of Figure 2 plots $\widehat{\xi}^{\star}_{\tau_n'}(x)$ versus k (left), for the empirical quartiles x of X, and plots the final estimates $x \mapsto \widehat{\xi}^{\star}_{\tau_n'}(x)$ (right), obtained by using the automatic selection and three values of k (selected in the stable region $[5, 25]$ of the plots shown on the left-hand side). The data-driven method affords a smoother and more stable estimated curve (in red). This extrapolated curve $x \mapsto \widehat{\xi}^{\star}_{\tau_n'}(x)$ is superimposed in the bottom panel (left) with the curve $x \mapsto \check{\xi}_{\mathrm{LL},\tau_n'}(x)$ of the local linear estimator (in solid blue), along with their corresponding asymptotic pointwise 95% confidence intervals (in dashed blue for $\check{\xi}_{\mathrm{LL},\tau_n'}(x)$ and dashed red for $\widehat{\xi}^{\star}_{\tau_n'}(x)$). There is a substantial difference between $\widehat{\xi}^{\star}_{\tau_n'}(x)$ and $\check{\xi}_{\mathrm{LL},\tau_n'}(x)$, which may exceed 80,000 USD for claimants' ages x outside the interval $[30, 40]$. The extrapolated quantile (Value at Risk) estimator $\widehat{q}^{\star}_{\tau_n'}(x)$, described in (11), is also graphed in the same figure (in dashed green). It lies below the extremile competitor (in solid red), close and sometimes beyond its lower confidence bound (in dashed red), with a gap that may exceed 54,000 USD. The extrapolated estimate $\widehat{\xi}^{\star}_{\tau_n'}(x)$ shows much greater reactivity to the shape of the conditional tail, as can be observed on the right end of the plot where $\widehat{\xi}^{\star}_{\tau_n'}(x)$ visibly takes into account the few very large claims incurred by older customers to produce a more prudent measure of extreme risk relative to these older claimants.

The bottom right panel plots the extrapolated $\widehat{\xi}^{\star}_{\tau_n'}(x)$ and ordinary $\check{\xi}_{\mathrm{LL},\tau_n'}(x)$ estimators of the tail extremile $\xi_{\tau_n'}(x) \equiv \mathbb{E}[\max(Y_x^1, \ldots, Y_x^{r(\tau_n')})]$, in solid and dashed lines respectively, for $\tau_n' = .992$ [*i.e.* $r(\tau_n') \approx 86$], $\tau_n' = .995$ [*i.e.* $r(\tau_n') \approx 138$] and $\tau_n' = .998$ [*i.e.* $r(\tau_n') \approx 346$]. With the increase of the security level τ_n', in contrast to the non-extrapolated estimator $\breve{\xi}_{\mathrm{LL},\tau_n'}(x)$, the refined version $\widehat{\xi}^{\star}_{\tau_n'}(x)$ becomes clearly more alert to the claims' severity and better captures the magnitude of the two most extreme losses.

5 Concluding remarks

The use of regression extremiles appears naturally in the context of risk handling, where their interpretability is straightforward and they are fully operational in practice. Beyond their remarkable merits from the point of view of the axiomatic theory of risk measures, why should statisticians and practitioners care about extremile regression? A first distinctive property of extremile regression is that, in contrast to its quantile and expectile competitors, the local linear estimators have an explicit form that is straightforward to compute, without recourse to any approximation algorithm. A second unexpected result is that the asymptotic variance of these estimators is not merely an adaptation to the conditional setup of the asymptotic variance arising in the unconditional case from [3]. In particular, this makes inference on regression extremiles much easier than inference on regression quantiles and expectiles. We are not aware of any ready-made procedure for constructing asymptotic confidence intervals of both conditional quantiles and expectiles based on the limit distributions of their local linear estimators. A further distinctive advantage of using local linear extremiles is that they suggest better capability of fitting both location and spread of data points beyond the regression mean, especially for heavy-tailed distributions. They provide much smoother and more stable fits than their quantile counterparts and do not suffer from crossings, as illustrated through both the concrete applications on triceps skinfold variation and motorcycle insurance payouts. In the class of light-tailed conditional distributions, population extremiles and quantiles are equivalent in the tail. In this class, the merits of local linear extremile estimators lead us then to favor their use over quantile estimators. It should finally be pointed out that we restrict our local linear kernel smoothing analysis to one-dimensional covariates. Extensions of our Theorem 1 and Corollary 1 to multiple covariates are obviously of interest as well.

Acknowledgments

The authors are grateful to the Editor, an Associate Editor, and three reviewers for their very valuable comments, which led to a much improved presentation of the results. The research of A. Daouia and G. Stupfler is supported by the French National Research Agency under the grant ANR-19-CE40-0013/ExtremReg project. A. Daouia also acknowledges funding from the ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program). I. Gijbels gratefully acknowledges support from the Research Fund KU Leuven (projects GOA/12/014 and C16/20/002), and from Research Grant FWO G0D6619N (Flemish Science Foundation). G. Stupfler also acknowledges support from an AXA Research Fund Award on "Mitigating risk in the wake of the COVID-19 pandemic".

References

[1] Bellini, F., Klar, B., Müller, A. and Gianin, E.R. (2014). Generalized quantiles as risk measures. *Insurance Math. Econom.*, **54**, 41–48.

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

[3] Daouia, A., Gijbels, I. and Stupfler, G. (2019). Extremiles: A new perspective on asymmetric least squares. *J. Amer. Statist. Assoc.*, **114**, 1366–1381.

[4] Daouia, A., Girard, S. and Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles. *J. R. Stat. Soc. Ser. B*, **80**, 263–292.

[5] Daouia, A. and Park, B.U. (2013). On Projection-Type Estimators of Multivariate Isotonic Functions. *Scand. J. Stat.*, **40**, 363–386.

[6] Goegebeur, Y., Guillou, A. and Osmann, M. (2017). A local moment type estimator for an extreme quantile in regression with random covariates. *Comm. Statist. Theory Methods*, **46**, 319–343.

[7] Horváth, L. and Yandell, B.S. (1988). Asymptotics of conditional empirical processes. *J. Multivariate Anal.*, **26**, 184–206.

[8] Newey, W.K. and Powell, J.L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, **55**, 819–847.

[9] Stute, W. (1986). Conditional empirical processes. *Ann. Statist.*, **14**, 638–647.

[10] Stupfler, G. (2013). A moment estimator for the conditional extreme-value index. *Electron. J. Stat.*, **7**, 2298–2343.

[11] Yao, Q. and Tong, H. (1996). Asymmetric least squares regression and estimation: A nonparametric approach. *J. Nonparametr. Stat.*, **6**, 273–292.

[12] Yu, K. and Jones, M.C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.*, **93**, 228–237.

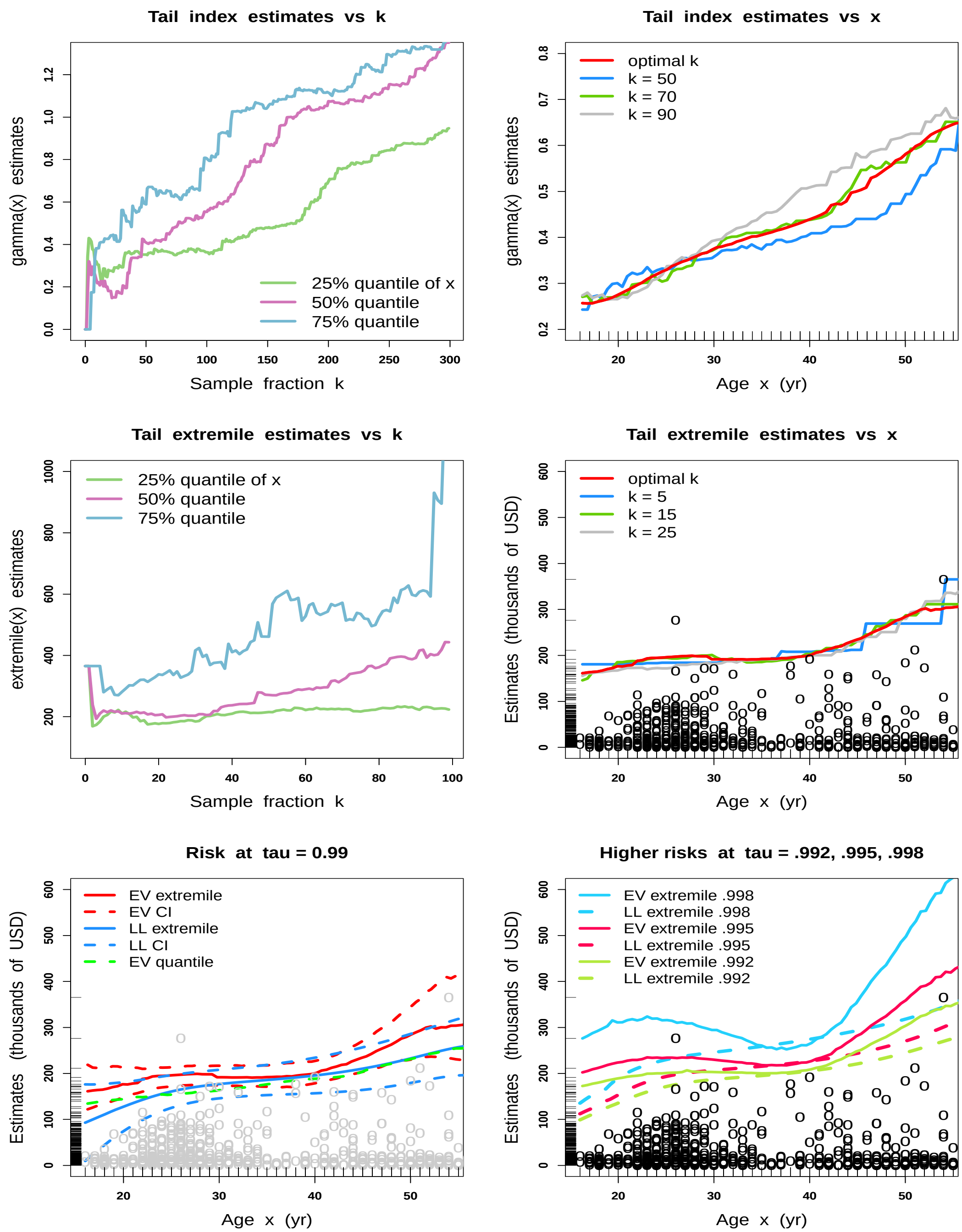

Figure 2: *Top left: plots of* $\widehat{\gamma}(x)$ *versus* k*, for the quartiles of* X*; Top right: estimates* $x \mapsto \widehat{\gamma}(x)$ *obtained via the automatic selection (red) and three selected* k *values; Middle left: plots of* $\widehat{\xi}^{\star}_{.99}(x)$ *versus* k*; Middle right: final estimates* $x \mapsto \widehat{\xi}^{\star}_{.99}(x)$*; Bottom left: Estimates* $\widehat{\xi}^{\star}_{.99}(x)$ *and* $\breve{\xi}_{LL,.99}(x)$ *(solid red and blue), corresponding* 95% *confidence intervals (dashed red and blue), and* $\widehat{q}^{\star}_{.99}(x)$ *(dashed green); Bottom right: The estimators* $\widehat{\xi}^{\star}_{\tau'_n}(x)$ *(solid lines) and* $\breve{\xi}_{LL,\tau'_n}(x)$ *(dashed lines), for* $\tau'_n = .992$ *(green),* $\tau'_n = .995$ *(red), and* $\tau'_n = .998$ *(blue).*

Supplementary Material for
"**Extremile regression**"

This supplement contains the proofs of all theoretical results in the main paper, along with auxiliary results in Section A. It also discusses in Section B the rationale behind our automatic bandwidth selector that was employed in Section 3 of the main article. The data-driven selection method of the two tuning parameters h_n and α_n that was used in Section 4 of the main article is described below in Section C. Finally, we provide some simulation experiments in Section D.

A Proofs

Proof of Proposition 1. Definition 1 states that

$$\xi_\tau(x) = \arg\min_{\theta\in\mathbb{R}} \mathbb{E}\left\{ J_\tau\left(F(Y|X)\right) \cdot [|Y-\theta|^2 - |Y|^2]|X=x \right\}.$$

The right-hand side defines a convex polynomial of degree 2 whose derivative is 0 at

$$\xi_\tau(x) = \frac{\mathbb{E}\left[Y\, J_\tau\left(F(Y|X)\right)|X=x\right]}{\mathbb{E}\left[J_\tau\left(F(Y|X)\right)|X=x\right]}.$$

Now

$$\mathbb{E}\left[J_\tau\left(F(Y|X)\right)|X=x\right] = \int_{y\in\mathbb{R}} J_\tau\left(F(y|x)\right)\, dF(y|x) = \int_0^1 J_\tau(u)\, du = 1$$

by continuity of $F(\cdot|x)$, so that

$$\xi_\tau(x) = \mathbb{E}\left[Y\, J_\tau\left(F(Y|X)\right)|X=x\right] = \int_{y\in\mathbb{R}} y\, J_\tau\left(F(y|x)\right)\, dF(y|x) = \int_0^1 J_\tau(t)\, q_t(x)\, dt$$
$$= \int_0^1 q_t(x)\, dK_\tau(t)$$

using the identity $K'_\tau = J_\tau$. Besides, we have by Definition 1 that $\xi_\tau(x) = \mathbb{E}(Z_\tau^x)$ where Z_τ^x has cumulative distribution function $F_{Z_\tau^x} = K_\tau(F(\cdot|x))$. When $\tau = (1/2)^{1/r}$, $r \in \mathbb{N}\backslash\{0\}$, this cumulative distribution function becomes

$$\forall z \in \mathbb{R},\ F_{Z_\tau^x}(z) = K_\tau(F(z|x)) = [F(z|x)]^r = \mathbb{P}(\max\left(Y_x^1, \ldots, Y_x^r\right) \leqslant z)$$

so that $Z_\tau^x \stackrel{d}{=} \max\left(Y_x^1, \ldots, Y_x^r\right)$ and therefore $\xi_\tau(x) = \mathbb{E}\left[\max\left(Y_x^1, \ldots, Y_x^r\right)\right]$. Similarly, when $\tau = 1-(1/2)^{1/s}$, $s \in \mathbb{N}\backslash\{0\}$, one finds the complementary distribution function of Z_τ^x as

$$\forall z \in \mathbb{R},\ 1 - F_{Z_\tau^x}(z) = 1 - K_\tau(F(z|x)) = [1-F(z|x)]^s = \mathbb{P}(\min\left(Y_x^1, \ldots, Y_x^s\right) > z)$$

so that $Z_\tau^x \stackrel{d}{=} \min\left(Y_x^1, \ldots, Y_x^s\right)$ and thus $\xi_\tau(x) = \mathbb{E}\left[\min\left(Y_x^1, \ldots, Y_x^s\right)\right]$, as required. ∎

Proof of Proposition 2. Statement (i) is a direct corollary of Proposition 2(iv) in Daouia *et al.* (2019) applied to the conditional distribution of Y given $X = x$. To show (ii), we note that, by independence of X and ε and positivity of $\sigma(X)$,

$$q_t(x) = m(x) + \sigma(x)q_{t,\varepsilon}$$

where $t \mapsto q_{t,\varepsilon}$ denotes the quantile function of ε. By our Proposition 1,

$$\xi_\tau(x) = \int_0^1 J_\tau(t)\, q_t(x)\, dt = m(x) + \sigma(x) \int_0^1 J_\tau(t)\, q_{t,\varepsilon}\, dt = m(x) + \sigma(x)\xi_{t,\varepsilon}$$

since J_τ has integral 1. ■

To prove Theorem 1 we need the following preliminary result on certain population and empirical smoothed moments.

Lemma A.1. *Assume that conditions (C1)–(C3) hold. Assume also that $h_n \to 0$ and $nh_n \to \infty$, as $n \to \infty$. Let x be interior to the support of X.*

(i) We have, for any nonnegative integer j, any $k \in \{1,2\}$ and any $l \in \{0,1,2\}$, that

$$\mathfrak{m}_{j,k,l,n}(x) := \mathbb{E}\left\{ J_\tau^k\left(F(Y|x)\right) Y^l \left(\frac{x-X}{h_n}\right)^{jk} \frac{1}{h_n^k} L^k\left(\frac{x-X}{h_n}\right)\right\}$$

$$\sim \frac{1}{h_n^{k-1}} \mathbb{E}(Y^l J_\tau^k\left(F(Y|x)\right)|X = x) f_X(x) \int_{\mathbb{R}} z^{jk} L^k(z)\, dz \quad \text{as} \quad n \to \infty.$$

In addition, there is $\delta > 0$ such that for any nonnegative integer j,

$$\mathbb{E}\left\{ J_\tau^{2+\delta}\left(F(Y|x)\right) |Y|^{2+\delta} \left|\frac{x-X}{h_n}\right|^{j(2+\delta)} \frac{1}{h_n} L^{2+\delta}\left(\frac{x-X}{h_n}\right)\right\} = \mathrm{O}(1) \quad \text{as} \quad n \to \infty.$$

(ii) With the above notation,

$$\frac{1}{h_n}\left(\mathfrak{m}_{0,1,0,n}(x)\mathfrak{m}_{1,1,1,n}(x) - \mathfrak{m}_{1,1,0,n}(x)\mathfrak{m}_{0,1,1,n}(x)\right)$$

$$\to -f_X(x) \int_{\mathbb{R}} [y - \xi_\tau(x)] J_\tau\left(F(y|x)\right) \partial_x f_{(X,Y)}(x,y)\, dy \int_{\mathbb{R}} z^2 L(z)\, dz \quad \text{as} \quad n \to \infty$$

(note that the limit is indeed well-defined and finite).

(iii) With the notation of (i), as $n \to \infty$,

$$\frac{\mathfrak{m}_{0,1,1,n}(x)\mathfrak{m}_{2,1,0,n}(x) - \mathfrak{m}_{1,1,1,n}(x)\mathfrak{m}_{1,1,0,n}(x)}{\mathfrak{m}_{0,1,0,n}(x)\mathfrak{m}_{2,1,0,n}(x) - [\mathfrak{m}_{1,1,0,n}(x)]^2}$$

$$= \xi_\tau(x) + \frac{h_n^2}{f_X(x)} \int_{\mathbb{R}} z^2 L(z)\, dz \left(\frac{1}{2} \int_{\mathbb{R}} [y - \xi_\tau(x)] J_\tau\left(F(y|x)\right) \partial^2_{xx} f_{(X,Y)}(x,y)\, dy\right.$$

$$\left. - \int_{\mathbb{R}} [y - \xi_\tau(x)] J_\tau\left(F(y|x)\right) \partial_x f_{(X,Y)}(x,y)\, dy \int_{\mathbb{R}} J_\tau\left(F(y|x)\right) \frac{\partial_x f_{(X,Y)}(x,y)}{f_X(x)}\, dy\right) + \mathrm{o}(h_n^2).$$

(iv) We have, for any nonnegative integer j:

$$\frac{1}{nh_n}\sum_{i=1}^{n} J_\tau\left(F(Y_i|x)\right)\left(\frac{x-X_i}{h_n}\right)^j L\left(\frac{x-X_i}{h_n}\right) \xrightarrow{\mathbb{P}} f_X(x)\int_{\mathbb{R}} z^j L(z)\,dz \quad \text{as} \quad n\to\infty.$$

Proof of Lemma A.1. Note that

$$\mathfrak{m}_{j,k,l,n}(x) = \frac{1}{h_n^{k-1}}\int\int_{\mathbb{R}^2} J_\tau^k\left(F(y|x)\right) y^l z^{jk} L^k(z) f_{(X,Y)}(x-h_n z, y)\,dz\,dy. \tag{A.1}$$

By the dominated convergence theorem (note that J_τ is bounded), this entails

$$h_n^{k-1}\mathfrak{m}_{j,k,l,n}(x) \to \int_{\mathbb{R}} y^l J_\tau^k\left(F(y|x)\right) f_{(X,Y)}(x,y)\,dy \int_{\mathbb{R}} z^{jk} L^k(z)\,dz \quad \text{as} \quad n\to\infty.$$

Noting that $f_{(X,Y)}(x,y) = f_{Y|X}(y|x) f_X(x)$ gives the convergence stated in (i). The O(1) statement follows in exactly the same way.

To show (ii), we note that, combining (A.1), a Taylor expansion of $z\mapsto f_{(X,Y)}(x-h_n z, y)$ in a neighborhood of 0, and the equality $\int_{\mathbb{R}} zL(z)\,dz = 0$, we find

$$\mathfrak{m}_{1,1,l,n}(x) = -h_n\int_{\mathbb{R}} y^l J_\tau\left(F(y|x)\right)\partial_x f_{(X,Y)}(x,y)\,dy\int_{\mathbb{R}} z^2 L(z)\,dz + \mathrm{o}(h_n) \tag{A.2}$$

for $l = 0, 1$. Using this asymptotic equivalence in conjunction with $\mathfrak{m}_{0,1,0,n}(x) \to f_X(x)$ and $\mathfrak{m}_{0,1,1,n}(x) \to f_X(x)\xi_\tau(x)$ (as a consequence of (i)) shows (ii).

Expansion (iii) rests on the following identity:

$$\begin{aligned}&\frac{\mathfrak{m}_{0,1,1,n}(x)\mathfrak{m}_{2,1,0,n}(x) - \mathfrak{m}_{1,1,1,n}(x)\mathfrak{m}_{1,1,0,n}(x)}{\mathfrak{m}_{0,1,0,n}(x)\mathfrak{m}_{2,1,0,n}(x) - [\mathfrak{m}_{1,1,0,n}(x)]^2}\\ &= \left(\frac{\mathfrak{m}_{0,1,1,n}(x)}{\mathfrak{m}_{0,1,0,n}(x)} - \frac{\mathfrak{m}_{1,1,1,n}(x)\mathfrak{m}_{1,1,0,n}(x)}{\mathfrak{m}_{0,1,0,n}(x)\mathfrak{m}_{2,1,0,n}(x)}\right)\left(1 - \frac{[\mathfrak{m}_{1,1,0,n}(x)]^2}{\mathfrak{m}_{0,1,0,n}(x)\mathfrak{m}_{2,1,0,n}(x)}\right)^{-1}.\end{aligned}$$

Note further that, up to order h_n^2,

$$\mathfrak{m}_{0,1,l,n}(x) - f_X(x)\left[\mathbb{1}(l=0) + \xi_\tau(x)\mathbb{1}(l=1)\right] \approx \frac{h_n^2}{2}\int_{\mathbb{R}} y^l J_\tau\left(F(y|x)\right)\partial^2_{xx} f_{(X,Y)}(x,y)\,dy\int_{\mathbb{R}} z^2 L(z)\,dz$$

for $l = 0, 1$. Combining this expansion with the convergence $\mathfrak{m}_{2,1,0,n}(x) \to f_X(x)\int_{\mathbb{R}} z^2 L(z)\,dz$ and (A.2) provides the desired result after some straightforward calculations.

Convergence (iv) is obtained by remarking that

$$\mathbb{E}\left[\frac{1}{nh_n}\sum_{i=1}^{n} J_\tau\left(F(Y_i|x)\right)\left(\frac{x-X_i}{h_n}\right)^j L\left(\frac{x-X_i}{h_n}\right)\right] = \mathfrak{m}_{j,1,0,n}(x)$$

$$\text{and} \quad \mathrm{Var}\left[\frac{1}{nh_n}\sum_{i=1}^{n} J_\tau\left(F(Y_i|x)\right)\left(\frac{x-X_i}{h_n}\right)^j L\left(\frac{x-X_i}{h_n}\right)\right] = \frac{1}{n}\left(\mathfrak{m}_{j,2,0,n}(x) - \{\mathfrak{m}_{j,1,0,n}(x)\}^2\right).$$

The conclusion then follows from a combination of (i) and Chebyshev's inequality. □

Proof of Theorem 1. Define

$$(\alpha_n(x), \beta_n(x)) = \underset{(\alpha,\beta)\in\mathbb{R}^2}{\arg\min}\, \mathbb{E}\left\{ J_\tau\left(F(Y|x)\right)(Y - \alpha - \beta(x - X))^2 L\left(\frac{x - X}{h_n}\right)\right\}.$$

This minimizer is indeed well-defined and unique, and we have

$$\alpha_n(x) = \frac{\mathfrak{m}_{0,1,1,n}(x)\mathfrak{m}_{2,1,0,n}(x) - \mathfrak{m}_{1,1,1,n}(x)\mathfrak{m}_{1,1,0,n}(x)}{\mathfrak{m}_{0,1,0,n}(x)\mathfrak{m}_{2,1,0,n}(x) - [\mathfrak{m}_{1,1,0,n}(x)]^2}$$

$$\text{and} \quad \beta_n(x) = \frac{1}{h_n}\,\frac{\mathfrak{m}_{0,1,0,n}(x)\mathfrak{m}_{1,1,1,n}(x) - \mathfrak{m}_{1,1,0,n}(x)\mathfrak{m}_{0,1,1,n}(x)}{\mathfrak{m}_{0,1,0,n}(x)\mathfrak{m}_{2,1,0,n}(x) - [\mathfrak{m}_{1,1,0,n}(x)]^2}$$

with the notation of Lemma A.1. Note now that

$$\sqrt{nh_n}\left(\breve{\xi}_{\mathrm{LL},\tau}(x) - \alpha_n(x), h_n(\breve{\beta} - \beta_n(x))\right) = \underset{(u,v)\in\mathbb{R}^2}{\arg\min}\, \psi_n(u, v)$$

with

$$\psi_n(u, v) = \sum_{i=1}^n J_\tau\left(\widehat{F}(Y_i|x)\right)\left\{Y_i - \left[\alpha_n(x) + \frac{u}{\sqrt{nh_n}}\right] - \left[\beta_n(x) + \frac{v/h_n}{\sqrt{nh_n}}\right](x - X_i)\right\}^2 L\left(\frac{x - X_i}{h_n}\right)$$
$$- \sum_{i=1}^n J_\tau\left(\widehat{F}(Y_i|x)\right)\{Y_i - \alpha_n(x) - \beta_n(x)(x - X_i)\}^2 L\left(\frac{x - X_i}{h_n}\right).$$

The objective function $\psi_n(u, v)$ is clearly continuous and convex; by Theorem 5 in Knight (1999), it is enough to analyze the asymptotic properties of $\psi_n(u, v)$, rather than those of the minimizer. Expanding and simplifying, we find

$$\psi_n(u, v) = -\frac{2}{\sqrt{nh_n}}\sum_{i=1}^n J_\tau\left(\widehat{F}(Y_i|x)\right)\{Y_i - \alpha_n(x) - \beta_n(x)(x - X_i)\}\left\{u + v\frac{x - X_i}{h_n}\right\} L\left(\frac{x - X_i}{h_n}\right)$$
$$+ \frac{1}{nh_n}\sum_{i=1}^n J_\tau\left(\widehat{F}(Y_i|x)\right)\left\{u + v\frac{x - X_i}{h_n}\right\}^2 L\left(\frac{x - X_i}{h_n}\right).$$

Since $\alpha_n(x)$ and $\beta_n(x)$ have finite limits by Lemma A.1(ii) and (iii), we can rewrite $\psi_n(u, v)$ as

$$\psi_n(u, v) = -\frac{2}{\sqrt{nh_n}}\sum_{i=1}^n J_\tau\left(F(Y_i|x)\right)\{Y_i - \alpha_n(x) - \beta_n(x)(x - X_i)\}\left\{u + v\frac{x - X_i}{h_n}\right\} L\left(\frac{x - X_i}{h_n}\right)$$
$$+ \frac{1}{nh_n}\sum_{i=1}^n J_\tau\left(F(Y_i|x)\right)\left\{u + v\frac{x - X_i}{h_n}\right\}^2 L\left(\frac{x - X_i}{h_n}\right) + \mathrm{o}_{\mathbb{P}}(1).$$

[Here assumption (8) in the statement of Theorem 1 was used to replace $\widehat{F}(Y_i|x)$ by $F(Y_i|x)$ in the first term, and the uniform consistency of $\widehat{F}(\cdot|x)$ was used together with Lemma A.1(iv) for $\tau = 1/2$ to replace $\widehat{F}(Y_i|x)$ by $F(Y_i|x)$ in the second term.] By Lemma A.1(iv) and the equality $\int_{\mathbb{R}} zL(z)\,dz = 0$,

$$\frac{1}{nh_n}\sum_{i=1}^n J_\tau\left(F(Y_i|x)\right)\left\{u + v\frac{x - X_i}{h_n}\right\}^2 L\left(\frac{x - X_i}{h_n}\right) \xrightarrow{\mathbb{P}} f_X(x)\left(u^2 + v^2\int_{\mathbb{R}} z^2 L(z)\,dz\right).$$

We use this convergence to rewrite $\psi_n(u,v)$ in yet another form:

$$\psi_n(u,v) = -2u\mathcal{S}_{n,1}(x) - 2v\mathcal{S}_{n,2}(x) + u^2 f_X(x) + v^2 f_X(x)\int_{\mathbb{R}} z^2 L(z)\,dz + o_{\mathbb{P}}(1)$$

$$\text{with}\quad \mathcal{S}_{n,1}(x) = \frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n} J_\tau\left(F(Y_i|x)\right)\{Y_i - \alpha_n(x) - \beta_n(x)(x - X_i)\}\, L\left(\frac{x - X_i}{h_n}\right)$$

$$\text{and}\quad \mathcal{S}_{n,2}(x) = \frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n} J_\tau\left(F(Y_i|x)\right)\{Y_i - \alpha_n(x) - \beta_n(x)(x - X_i)\}\, \frac{x - X_i}{h_n} L\left(\frac{x - X_i}{h_n}\right).$$

Note that $\mathcal{S}_{n,1}(x)$ and $\mathcal{S}_{n,2}(x)$ are sums of independent, identically distributed, and centered random variables (the latter by definition of $\alpha_n(x)$ and $\beta_n(x)$). By Lemma A.1(i), (ii) and (iii), tedious but straightforward calculations, and the Lyapunov central limit theorem (see *e.g.* Theorem 27.3 in p.362 of Billingsley, 1999), $(\mathcal{S}_{n,1}(x), \mathcal{S}_{n,2}(x))$ converges weakly to a random pair $(\mathcal{S}_1, \mathcal{S}_2)$ having a bivariate centered normal distribution. In particular

$$\begin{aligned}\mathrm{Var}(\mathcal{S}_1) &= \lim_{n\to\infty}\mathbb{E}\left[J_\tau^2\left(F(Y|x)\right)\{Y - \alpha_n(x) - \beta_n(x)(x - X)\}^2 \frac{1}{h_n} L^2\left(\frac{x - X}{h_n}\right)\right] \\ &= f_X(x)\mathbb{E}\left[J_\tau^2\left(F(Y|x)\right)\{Y - \xi_\tau(x)\}^2 \mid X = x\right]\|L\|_2^2, \qquad \text{(A.3)}\end{aligned}$$

where Lemma A.1(ii) and (iii) were used to get $\alpha_n(x) \to \xi_\tau(x)$ and $\beta_n(x) = \mathrm{O}(1)$. Consequently

$$\psi_n(u,v) \xrightarrow{d} -2u\mathcal{S}_1 - 2v\mathcal{S}_2 + u^2 f_X(x) + v^2 f_X(x)\int_{\mathbb{R}} z^2 L(z)\,dz$$

in the sense of finite-dimensional convergence. By Theorem 5 in Knight (1999),

$$\sqrt{nh_n}\left(\breve{\xi}_{\mathrm{LL},\tau}(x) - \alpha_n(x), h_n(\breve{\beta} - \beta_n(x))\right) \xrightarrow{d} \left(\frac{\mathcal{S}_1}{f_X(x)}, \frac{\mathcal{S}_2}{f_X(x)\int_{\mathbb{R}} z^2 L(z)\,dz}\right).$$

Combining the convergence of the first marginal with (A.3) yields

$$\sqrt{nh_n}\left(\breve{\xi}_{\mathrm{LL},\tau}(x) - \alpha_n(x)\right) \xrightarrow{d} \mathcal{N}\left(0, V_\tau(x)\frac{\|L\|_2^2}{f_X(x)}\right).$$

Combining finally Lemma A.1(iii) with the assumption $nh_n^5 \to 0$ completes the proof. ∎

Proof of Corollary 1. Note that since $\tau \in (0, 1 - 1/\sqrt{2}] \cup [1/\sqrt{2}, 1)$, the function J_τ is Lipschitz continuous. Consequently

$$n^{2/5}\sup_{y\in\mathbb{R}}\left|J_\tau(\widehat{F}(y|x)) - J_\tau(F(y|x))\right| = \mathrm{O}_{\mathbb{P}}(1).$$

Since L has compact support and $nh_n^5 \to 0$, it follows that, for any nonnegative integer j,

$$\begin{aligned}&\left|\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n} J_\tau\left(\widehat{F}(Y_i|x)\right)\{Y_i - \alpha - \beta(x - X_i)\}\left(\frac{x - X_i}{h_n}\right)^j L\left(\frac{x - X_i}{h_n}\right)\right. \\ &\left. - \frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n} J_\tau\left(F(Y_i|x)\right)\{Y_i - \alpha - \beta(x - X_i)\}\left(\frac{x - X_i}{h_n}\right)^j L\left(\frac{x - X_i}{h_n}\right)\right| \\ &= o_{\mathbb{P}}\left(\frac{1}{nh_n}\sum_{i=1}^{n}(1 + |Y_i|)L\left(\frac{x - X_i}{h_n}\right)\right) \qquad \text{(A.4)}\end{aligned}$$

locally uniformly in $(\alpha, \beta) \in \mathbb{R}^2$. Conclude the proof by noting that

$$\mathbb{E}\left(\frac{1}{nh_n}\sum_{i=1}^n (1+|Y_i|)L\left(\frac{x-X_i}{h_n}\right)\right) = \int\int_{\mathbb{R}^2} (1+|y|)L(z)f_{(X,Y)}(x-h_n z, y)\,dz\,dy$$
$$\to f_X(x)\mathbb{E}[(1+|Y|)\,|\,X=x]$$

by the dominated convergence theorem and assumption (C1). Similarly

$$\mathrm{Var}\left(\frac{1}{nh_n}\sum_{i=1}^n (1+|Y_i|)L\left(\frac{x-X_i}{h_n}\right)\right) = \mathrm{O}\left(\frac{1}{nh_n}\right) \to 0.$$

Using the Chebyshev inequality entails

$$\frac{1}{nh_n}\sum_{i=1}^n (1+|Y_i|)L\left(\frac{x-X_i}{h_n}\right) \xrightarrow{\mathbb{P}} f_X(x)\mathbb{E}[(1+|Y|)\,|\,X=x]$$

which, combined with (A.4), shows that

$$\left|\frac{1}{\sqrt{nh_n}}\sum_{i=1}^n J_\tau\left(\widehat{F}(Y_i|x)\right)\{Y_i-\alpha-\beta(x-X_i)\}\left(\frac{x-X_i}{h_n}\right)^j L\left(\frac{x-X_i}{h_n}\right)\right.$$
$$\left.-\frac{1}{\sqrt{nh_n}}\sum_{i=1}^n J_\tau\left(F(Y_i|x)\right)\{Y_i-\alpha-\beta(x-X_i)\}\left(\frac{x-X_i}{h_n}\right)^j L\left(\frac{x-X_i}{h_n}\right)\right| = \mathrm{o}_{\mathbb{P}}(1)$$

locally uniformly in $(\alpha, \beta) \in \mathbb{R}^2$. Applying Theorem 1 completes the proof. ■

Before moving to the proof of Theorem 2, we first establish the asymptotic distribution of $\widehat{\xi}^{\star}_{\tau'_n}(x)$ under the general form (11), for generic estimators $\widehat{\gamma}(x)$ and $\widehat{q}_{\tau_n}(x)$.

Theorem A.1. *Suppose (E) holds with $\rho(x) < 0$ and, as $n \to \infty$,*

1. *$\tau_n \to 1$, $(1-\tau'_n)/(1-\tau_n) \to 0$ and $v_n \to \infty$ such that $v_n/\log[(1-\tau_n)/(1-\tau'_n)] \to \infty$;*
2. *$v_n\left(\widehat{q}_{\tau_n}(x)/q_{\tau_n}(x) - 1\right) = \mathrm{O}_{\mathbb{P}}(1)$;*
3. *$v_n\left(\widehat{\gamma}(x) - \gamma(x)\right) \xrightarrow{d} Z_x$, where Z_x is a nondegenerate limit;*
4. *$\dfrac{v_n}{\log[(1-\tau_n)/(1-\tau'_n)]}A((1-\tau_n)^{-1}|x) \to 0$ and $\dfrac{v_n}{\log[(1-\tau_n)/(1-\tau'_n)]}(1-\tau'_n) \to 0$.*

Then $\dfrac{v_n}{\log[(1-\tau_n)/(1-\tau'_n)]}\left(\dfrac{\widehat{\xi}^{\star}_{\tau'_n}(x)}{\xi_{\tau'_n}(x)} - 1\right) \xrightarrow{d} Z_x$ *as $n \to \infty$.*

Proof of Theorem A.1. Set $\alpha_n = 1-\tau_n$, $p_n = 1-\tau'_n$ and $d_n = \alpha_n/p_n$. We have

$$\frac{v_n}{\log d_n}\left(\frac{\widehat{\xi}^{\star}_{1-p_n}(x)}{\xi_{1-p_n}(x)} - 1\right) = \frac{v_n}{\log d_n}\left(\frac{\widehat{q}^{\star}_{1-p_n}(x)}{q_{1-p_n}(x)} - 1\right)\frac{q_{1-p_n}(x)}{\xi_{1-p_n}(x)}\mathcal{G}\left(\widehat{\gamma}(x)\right)$$
$$+\frac{v_n}{\log d_n}\left(\mathcal{G}\left(\widehat{\gamma}(x)\right) - \mathcal{G}\left(\gamma(x)\right)\right)\frac{q_{1-p_n}(x)}{\xi_{1-p_n}(x)} + \frac{v_n}{\log d_n}\left[\mathcal{G}\left(\gamma(x)\right) - \frac{\xi_{1-p_n}(x)}{q_{1-p_n}(x)}\right]\frac{q_{1-p_n}(x)}{\xi_{1-p_n}(x)},$$

where $q_{1-p_n}(x)/\xi_{1-p_n}(x) \to 1/\mathcal{G}(\gamma(x))$, $v_n\{\mathcal{G}(\widehat{\gamma}(x)) - \mathcal{G}(\gamma(x))\} = \mathrm{O}_{\mathbb{P}}(1)$, and $\log d_n \to \infty$ as $n \to \infty$. Following the lines of the proof of Theorem 4.3.8 in de Haan and Ferreira (2006), the first term above therefore converges to Z_x. The second term, meanwhile, converges to 0 by a straightforward application of the Delta-method. Finally, the third term satisfies

$$\frac{v_n}{\log d_n}\left|\mathcal{G}(\gamma(x)) - \frac{\xi_{1-p_n}(x)}{q_{1-p_n}(x)}\right| \frac{q_{1-p_n}(x)}{\xi_{1-p_n}(x)} = \mathrm{O}\left(\frac{v_n}{\log d_n}\left[|A(1/p_n|x)| + p_n\right]\right)$$
$$= \mathrm{O}\left(\frac{v_n}{\log d_n}\left[|A(1/\alpha_n|x)| + p_n\right]\right)$$

by Proposition 4 in Daouia *et al.* (2019), since $|A(\cdot|x)|$ is regularly varying with index $\rho(x) < 0$. The left-hand side thus converges to 0, completing the proof. ■

Let us now turn to the asymptotic normality of

$$\widehat{\xi}^{\star}_{1-p_n}(x) := \left(\frac{\alpha_n}{p_n}\right)^{\widehat{\gamma}(x)} \widehat{q}_{1-\alpha_n}(x)\,\mathcal{G}(\widehat{\gamma}(x))$$

in Theorem 2, where $\widehat{q}_\tau(x) \equiv \widehat{F}^{-1}_{\mathrm{NW}}(\tau|x)$ and

$$\widehat{\gamma}(x) = \sum_{j=1}^{J}\left[\log \widehat{q}_{1-t_j\alpha_n}(x) - \log \widehat{q}_{1-\alpha_n}(x)\right] \Big/ \sum_{j=1}^{J} \log(1/t_j).$$

Proof of Theorem 2. We first show the joint convergence of the estimators $\widehat{q}_{1-t_j\alpha_n}(x)$, for $j \in \{1, \ldots, J\}$. Let $\alpha_{n,j} = t_j\alpha_n$, $v_n = \sqrt{nh_n^d\alpha_n}$, choose $z_1, \ldots, z_J \in \mathbb{R}$ and focus on the probability

$$\pi_n(z_1, \ldots, z_J) = \mathbb{P}\left(\bigcap_{j=1}^{J}\left\{v_n\left(\frac{\widehat{q}_{1-\alpha_{n,j}}(x)}{q_{1-\alpha_{n,j}}(x)} - 1\right) \leqslant z_j\right\}\right)$$
$$= \mathbb{P}\left(\bigcap_{j=1}^{J}\left\{\widehat{q}_{1-\alpha_{n,j}}(x) \leqslant q_{1-\alpha_{n,j}}(x)\,(1 + z_j/v_n)\right\}\right).$$

Putting $\overline{F}(\cdot|x) := 1 - F(\cdot|x)$ and $\widehat{\overline{F}}_{\mathrm{NW}}(\cdot|x) := 1 - \widehat{F}_{\mathrm{NW}}(\cdot|x)$, and using that for all y and α, $\widehat{q}_\alpha(x) \leqslant y \Leftrightarrow \widehat{\overline{F}}_{\mathrm{NW}}(y|x) \leqslant 1 - \alpha$, we find that

$$\pi_n(z_1, \ldots, z_J) = \mathbb{P}\left(\bigcap_{j=1}^{J}\left\{\widehat{\overline{F}}_{\mathrm{NW}}(q_{1-\alpha_{n,j}}(x)\,(1 + z_j/v_n)\,|x) \leqslant \alpha_{n,j}\right\}\right).$$

Letting $y_n = q_{1-\alpha_n}(x)$ and $y_{n,j} = q_{1-\alpha_{n,j}}(x)(1 + z_j/v_n)$, we rewrite $\pi_n(z_1, \ldots, z_J)$ as

$$\mathbb{P}\left(\bigcap_{j=1}^{J}\left\{\sqrt{nh_n^d\overline{F}(y_n|x)}\left(\frac{\widehat{\overline{F}}_{\mathrm{NW}}(y_{n,j}|x)}{\overline{F}(y_{n,j}|x)} - 1\right) \leqslant \sqrt{nh_n^d\overline{F}(y_n|x)}\left(\frac{\alpha_{n,j}}{\overline{F}(y_{n,j}|x)} - 1\right)\right\}\right).$$

By Lemma 1(ii) in Daouia *et al.* (2020) applied to the conditional distribution of Y given $X = x$, the second-order condition (E) yields

$$\lim_{t\to\infty}\frac{1}{A(t|x)}\left(\frac{1/\overline{F}(q_{1-t^{-1}}(x)|x)}{t} - 1\right) = 0.$$

In our case, this implies, for any $j \in \{1, \ldots, J\}$,

$$\frac{\alpha_{n,j}}{\overline{F}(q_{1-\alpha_{n,j}}(x)|x)} - 1 = \mathrm{o}\left(|A(1/\alpha_{n,j}|x)|\right) = \mathrm{o}\left(|A(1/\alpha_n|x)|\right) = \mathrm{o}\left(1/\sqrt{nh_n^d\overline{F}(y_n|x)}\right),$$

since $nh_n^d\overline{F}(y_n|x) = nh_n^d\overline{F}(q_{1-\alpha_n}(x)|x) = v_n^2(1+\mathrm{o}(1))$. Moreover, the second-order condition is equivalent to the following convergence:

$$\forall y > 0,\ \lim_{t\to\infty} \frac{1}{A(1/\overline{F}(t|x)|x)} \left(\frac{\overline{F}(ty|x)}{\overline{F}(t|x)} - y^{-1/\gamma(x)}\right) = y^{-1/\gamma(x)} \frac{y^{\rho(x)/\gamma(x)} - 1}{\gamma(x)\rho(x)}.$$

This second-order convergence is known to be locally uniform in $y \in (0, \infty)$ (see for instance Lemma 2 in Stupfler, 2019) and therefore, by a Taylor expansion,

$$\frac{\overline{F}(q_{1-\alpha_{n,j}}(x)|x)}{\overline{F}(y_{n,j}|x)} - 1 = \frac{z_j}{\gamma(x)} \frac{1}{v_n} + \mathrm{o}\left(|A(1/\alpha_{n,j}|x)|\right) = \frac{z_j}{\gamma(x)} \frac{1}{\sqrt{nh_n^d\overline{F}(y_n|x)}} (1 + \mathrm{o}(1)).$$

It follows that

$$\pi_n(z_1, \ldots, z_J) = \mathbb{P}\left(\bigcap_{j=1}^{J} \left\{\sqrt{nh_n^d\overline{F}(y_n|x)} \left(\frac{\widehat{\overline{F}}_{\mathrm{NW}}(y_{n,j}|x)}{\overline{F}(y_{n,j}|x)} - 1\right) \leqslant \frac{z_j}{\gamma(x)} + \mathrm{o}(1)\right\}\right).$$

Besides, again by the regular variation assumption, $y_{n,j}/y_n \to t_j^{-\gamma(x)}$ as $n \to \infty$. An inspection of the proof of Theorem 1 in Daouia *et al.* (2011) shows that, even though this result is formulated under the assumption "$y_{n,j} = a_j y_n$ for $j = 1, \ldots, J$", it is in fact valid under the weaker assumption "$y_{n,j} = a_j y_n(1 + \mathrm{o}(1))$ for $j = 1, \ldots, J$". Applying this result proves that the random vector

$$\left\{\sqrt{nh_n^d\overline{F}(y_n|x)} \left(\frac{\widehat{\overline{F}}_{\mathrm{NW}}(y_{n,j}|x)}{\overline{F}(y_{n,j}|x)} - 1\right)\right\}_{1\leqslant j\leqslant J}$$

converges weakly to a centered Gaussian random vector with covariance matrix $(\|L\|_2^2/f_X(x))C(x)$, where $C_{i,j}(x) = 1/t_{\min(i,j)}$. Conclude that

$$\left\{v_n\left(\frac{\widehat{q}_{1-\alpha_{n,j}}(x)}{q_{1-\alpha_{n,j}}(x)} - 1\right)\right\}_{1\leqslant j\leqslant J} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_J, \frac{\|L\|_2^2}{f_X(x)}\gamma^2(x)C(x)\right).$$

Using once again the second-order condition yields the expansion

$$\sum_{j=1}^{J}\left[\log q_{1-\alpha_{n,j}}(x) - \log q_{1-\alpha_n}(x)\right] \Big/ \sum_{j=1}^{J} \log(1/t_j) = \gamma(x) + \mathrm{o}(1/v_n).$$

A simple application of the delta-method now yields

$$v_n\left(\widehat{\gamma}(x) - \gamma(x)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\|L\|_2^2}{f_X(x)} V_J\, \gamma^2(x)\right).$$

Applying Theorem A.1 concludes the proof. ∎

B An automatic bandwidth selector

As with any smoothing techniques, tuning the degree of smoothing, reflected in our setup through bandwidth selection, is a major issue in practice. Our main goal here is to select an automatic bandwidth h_{ξ_τ} for smoothing the τth regression extremile curve via kernel local linear fitting. As a matter of fact, we can express h_{ξ_τ} in terms of the optimal bandwidth h_{q_τ} for regression quantile estimation, whose automatic selection is nowadays well-established in the literature. Since the conditional extremile $\xi_\tau(\cdot)$ and quantile $q_\tau(\cdot) = F^{-1}(\tau|\cdot)$ are, respectively, the mean and the median of the same conditional distribution of a random variable Z_τ^X given X, whose explicit distribution function is $F_{Z_\tau^X|X}(\cdot|x) = K_\tau(F(\cdot|x))$, the bandwidths h_{ξ_τ} and h_{q_τ} actually correspond to the optimal choices of h_n for regression mean and median estimation, respectively. Yu and Jones (1998, p.231) have already established a neat and practical connection between h_{mean} (*i.e.* the optimal choice h_{ξ_τ} of h_n for regression mean estimation) and h_{median} (*i.e.* the optimal choice h_{q_τ} of h_n for regression median estimation). More precisely, they have found that

$$\left(\frac{h_{\text{mean}}}{h_{\text{median}}}\right)^5 \equiv \left(\frac{h_{\xi_\tau}}{h_{q_\tau}}\right)^5 = \frac{4\,[q_\tau''(x)]^2 \cdot \sigma^2_{Z_\tau^X|X}(x) \cdot [f_{Z_\tau^X|X}\,(q_\tau(x)|x)]^2}{[\xi_\tau''(x)]^2},$$

where $q_\tau''(x)$ and $\xi_\tau''(x)$ are the second derivatives (with respect to x) of the conditional median $q_\tau(x)$ and mean $\xi_\tau(x)$, respectively, and $\sigma^2_{Z_\tau^X|X}(x)$ and $f_{Z_\tau^X|X}\,(\cdot|x)$ stand respectively for the conditional variance and density of Z_τ^X given $X = x$. By the same arguments as in Yu and Jones (1998, p.231) we can turn to the usual type of rule-of-thumb calculations of the conditional variance and density, based on assuming that the conditional distribution of Y given $X = x$ is Gaussian with mean μ_x and variance σ_x^2. Then it is easily seen that

$$\begin{aligned} f_{Z_\tau^X|X}\,(q_\tau(x)|x) &= J_\tau(\tau) \cdot \sigma_x^{-1} \cdot \phi\left(\Phi^{-1}(\tau)\right), \\ \sigma^2_{Z_\tau^X|X}(x) &= \sigma_x^2 \cdot V_{K_\tau\circ\Phi} := \sigma_x^2 \int_0^1 \left(\Phi^{-1}(t) - \mu_{K_\tau\circ\Phi}\right)^2 J_\tau(t)dt, \end{aligned}$$

where ϕ and Φ are the standard normal density and distribution functions, and $V_{K_\tau\circ\Phi}$ is the variance corresponding to the distribution function $K_\tau \circ \Phi$, with $\mu_{K_\tau\circ\Phi}$ being the mean of the distribution $K_\tau \circ \Phi$, or equivalently, the τth extremile of Φ, which is independent of x and readily calculated. Also, following Yu and Jones (1998, Discussion after Equation (7)), it seems reasonable as a first-order approximation to assume that $q_t''(x)$ is constant with respect to t. Hence

$$\frac{\xi_\tau''(x)}{q_\tau''(x)} = \int_0^1 \frac{q_t''(x)}{q_\tau''(x)} J_\tau(t)dt = \int_0^1 J_\tau(t)dt = 1.$$

In summary, we get a ready-to-use bandwidth selector for local linear kernel estimation as follows:

a. Use ready-made methods to select $h_{\xi_{1/2}}$, the optimal choice of bandwidth for regression mean estimation, *e.g.*, the cross-validation method implemented in the function `npregbw` of the R package `np`;

b. Find the optimal bandwidth $h_{q_\tau} = h_{\xi_{1/2}}\,\{\tau(1-\tau)/[\phi\,(\Phi^{-1}(\tau))]^2\}^{1/5}$, for smoothing the τth conditional quantile by the mean of the automatic method of Yu and Jones (1998);

c. Use the selected extremile bandwidth $h_{\xi_\tau} = h_{q_\tau}\,\{4\,V_{K_\tau\circ\Phi} \cdot [J_\tau(\tau) \cdot \phi\,(\Phi^{-1}(\tau))]^2\}^{1/5}$,

where (as specified above) ϕ and Φ are the standard normal density and distribution functions, and $V_{K_\tau\circ\Phi}$ is the readily calculated variance corresponding to the distribution function $K_\tau \circ \Phi$.

C Practical guidelines for selecting (h_n, α_n)

The quantile, extremile and tail index estimators $\widehat{q}^{\star}_{\tau'_n}(x)$, $\widehat{\xi}^{\star}_{\tau'_n}(x)$ and $\widehat{\gamma}(x)$ described in (11), (12) and (14), respectively, all depend on the intermediate regression quantiles $\widehat{q}_{1-t_j\alpha_n}(x)$ and on the choice of the tuning parameters h_n and $\alpha_n \equiv 1-\tau_n$. First, we apply the method of Li and Racine (2008) to compute the least-squares cross-validated bandwidths via the function `npcdistbw` of the R package `np`. These bandwidths are optimal for $\widehat{F}_{\mathrm{NW}}(\cdot|x)$ in (13). Then we compute the conditional quantiles $\widehat{q}_{1-t_j\alpha_n}(x) \equiv \widehat{F}^{-1}_{\mathrm{NW}}(1-t_j\alpha_n|x)$ by inverting $\widehat{F}_{\mathrm{NW}}(\cdot|x)$ via a direct adaptation of the function `kernesti.quantile` of the package `regpro` allowing it to use the Epanechnikov kernel. In our concrete application to motorcycle insurance data ($n = 670$) in Section 4.3 of the main paper, the global least-squares cross-validated bandwidth obtained with this Epanechnikov kernel is $h_n \approx 7.16$ (years).

As for selecting the parameter α_n, a usual practice is to set $\alpha_n = k/n$ for a sequence of integers $k = k_n \in \{1, \ldots, n-1\}$, then to plot the graph of the extreme value estimator of interest, say $\widehat{\gamma}(x)$ versus k for each x fixed, and finally to choose $k = k(x)$ from the first stable region of the plot. To this end, similarly to El Methni and Stupfler (2017) and Daouia *et al.* (2020), we employ a simple data-driven method based on balancing the potential estimation bias and variance:

- The first step consists in plotting the estimates $\widehat{\gamma}(x)$ versus k, for various values of x that correspond, for instance, to the $10\%, 20\%, \ldots, 90\%$ empirical quantiles of X, as can be seen below in Figure 1 (top) for motorcycle insurance claims. Note that, in Figure 2 (top left) of the main paper, we restrict our attention to the three empirical quartiles of X. The plots show global stability over the region $k \in [50, 90]$, which suggests to pick out the desired pointwise estimates $\widehat{\gamma}(x)$ over this interval;

- The second step, after identification of the first possible stable region globally shared by the plots for the different covariates x, is fully automatic. It consists first in computing the standard deviations of the estimator over a moving window large enough to cover around 60% up to 80% of k values in the selected potential stable region. Then, we determine the first window over which the standard deviation has a local minimum, and is less than the average standard deviation across all windows. Finally, we take the average of the estimates within this stable window as the final estimate. More specifically, for motorcycle insurance data, we computed the standard deviations of the tail index kernel smoothing estimator $\widehat{\gamma}(x)$ over a moving window of 32 successive values of k in the stable part $[50, 90]$ of its plots.

When for instance $\tau'_n = .99$, the plots of the extrapolated extremile estimator $\widehat{\xi}^{\star}_{\tau'_n}(x)$ versus k, for the same 9 empirical quantiles x of X as above, are graphed below in Figure 1 (middle) below. These plots show a potential global stable region over $k \in [5, 25]$. We computed the standard deviations of $\widehat{\xi}^{\star}_{\tau'_n}(x)$ over a moving window of around 13 successive values of k in this selected range. The resulting plots for the extrapolated quantile estimator $\widehat{q}^{\star}_{\tau'_n}(x)$, displayed below in Figure 1 (bottom), show a global stability over the range $k \in [20, 40]$. We computed the standard deviations of this estimator over a moving window of length 13 in this selected range.

D Some Monte Carlo evidence

As can be seen from Figure 1 in the main paper, due to data sparsity in the tail areas of *'dataAutoBi'* and *'dataOhlsson'*, direct estimates $\breve{\xi}_{\mathrm{LL},\tau'_n}(x) \equiv \breve{\alpha}$ from ordinary extremile regression in (7)

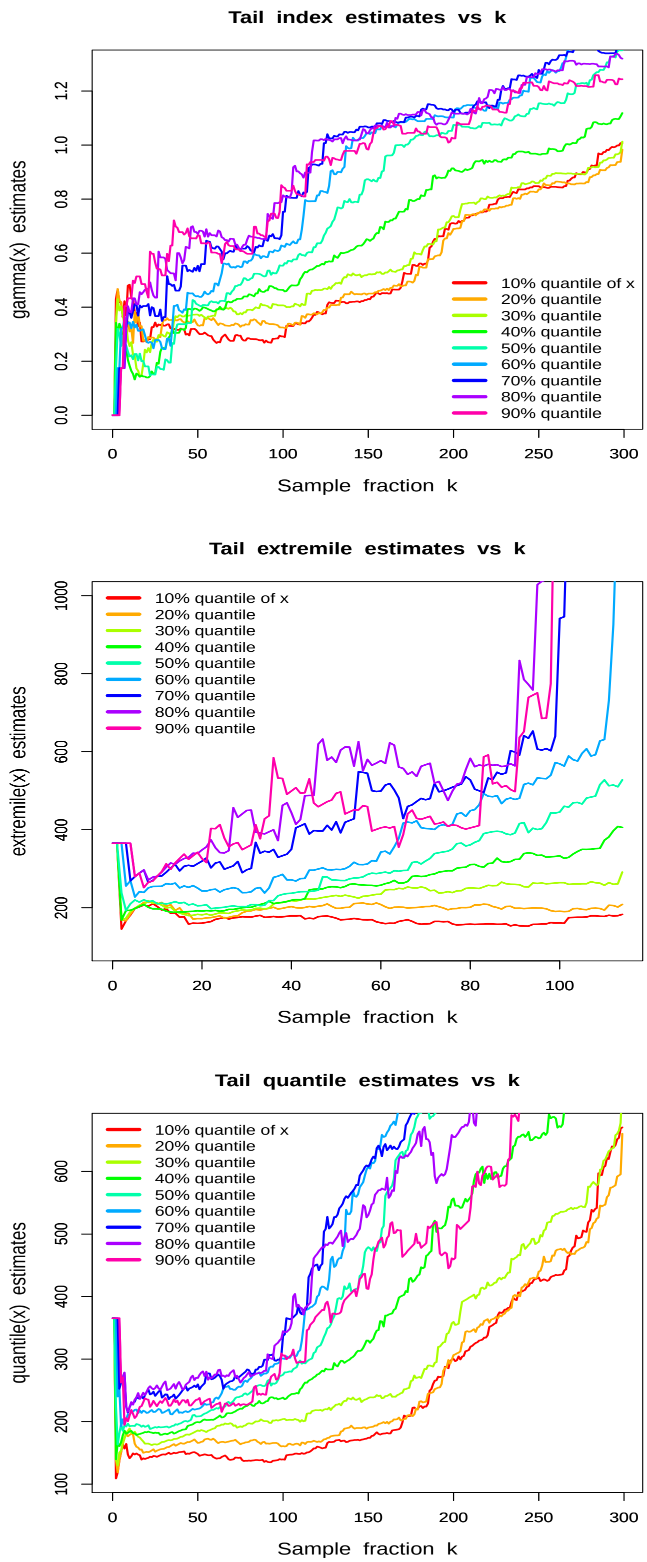

Figure 1: *(Top) plots of* $\widehat{\gamma}(x)$ *versus* k, *(Middle) plots of* $\widehat{\xi}^{\star}_{.99}(x)$ *versus* k, *(Bottom) plots of* $\widehat{q}^{\star}_{.99}(x)$ *versus* k, *for the* $10\%, 20\%, \ldots, 90\%$ *empirical quantiles* x *of* X.

may suffer from high variability at tails, especially for heavy-tailed distributions. To estimate conditional extremiles $\xi_{\tau_n'}(x)$ in the very far tail where very few or no observations are available, it is most efficient to use the extrapolated estimates $\widehat{\xi}^{\star}_{\tau_n'}(x)$ in (12). To illustrate this, we consider a uniformly distributed covariate variable X on the unit interval in conjunction with two conditional distributions for the response Y given $X = x \in [0,1]$, namely

- Fréchet distribution with tail index $\gamma(x)$:

$$F(y|x) = e^{-y^{-1/\gamma(x)}}, \; y > 0.$$

- Pareto distribution with tail index $\gamma(x)$:

$$F(y|x) = 1 - y^{-1/\gamma(x)}, \; y > 1.$$

All the experiments were performed for the sample size $n = 670$, as in our concrete application to insurance payouts. Inspired by the shape of the estimated conditional tail index obtained in this application, we used in all our simulations the function

$$\gamma(x) = 0.5 + 0.15\sqrt{x}, \quad \text{for all} \quad x \in [0,1].$$

For selecting the bandwidth h_n in each competing estimator, we used the automatic data-driven methods described above in Section C.

We compared the performance of the normalized estimators $\breve{\xi}_{\mathrm{LL},\tau_n'}(x)/\xi_{\tau_n'}(x)$ and $\widehat{\xi}^{\star}_{\tau_n'}(x)/\xi_{\tau_n'}(x)$ by computing their averaged mean-squared error (MSE in log-scale) and bias over 200 simulations, for the three covariate values $x \in \{0.25, 0.5, 0.75\}$ and the four extreme levels $\tau_n' \in \{0.99, 0.992, 0.995, 0.998\}$. Figure 2 plots the resulting Monte Carlo estimates at $x = 0.25$, versus the sample fraction $k = 1, \ldots, \lfloor n/\log(n^{0.9})\rfloor$. By construction, the estimates related to the ordinary local linear estimator $\breve{\xi}_{\mathrm{LL},\tau_n'}(x)$ (horizontal red lines) do not depend on the choice of the intermediate sequence k. Only those related to the extrapolated extreme-value estimator $\widehat{\xi}^{\star}_{\tau_n'}(x)$ (blue lines) require such a choice. To do so, we identified in each plot the first region of k values (a window large enough to contain at least 15 successive values of k) on which the MSE estimates look stable and reasonably small. Then, we took the average of the Monte Carlo estimates within this stable region as the final pointwise estimate, indicated by a horizontal green line in each plot.

The final pointwise MSE and bias estimates obtained for the three covariate values $x \in \{0.25, 0.5, 0.75\}$ are displayed in Table 1. Apart from a handful of cases indicated in blue, where the difference in performance marginally favors the non-extrapolated estimator, $\widehat{\xi}^{\star}_{\tau_n'}(x)$ seems to be substantially more accurate than $\breve{\xi}_{\mathrm{LL},\tau_n'}(x)$. That the non-extrapolated estimator $\breve{\xi}_{\mathrm{LL},\tau_n'}(x)$ may in very specific situations perform better than an extrapolated estimator such as $\widehat{\xi}^{\star}_{\tau_n'}(x)$ is neither unprecedented nor unexpected; it was shown recently in Girard *et al.* (2020), in the context of the nonparametric estimation of conditional expectiles, that at extreme levels such as the levels 0.99, 0.992, 0.995 and 0.998 we are considering here, non-extrapolated expectile estimators may have lower biases (but always much higher variance) than their extrapolated counterparts. This happens because non-extrapolated estimators such as $\breve{\xi}_{\mathrm{LL},\tau_n'}(x)$ tend to put a very high weight on the few (locally) largest observations relevant to extreme value estimation, whereas an extrapolated estimator such as $\widehat{\xi}^{\star}_{\tau_n'}(x)$ typically uses many more data points to infer the local shape parameter and anchor intermediate conditional extremile. The latter will thus be always much more stable (*i.e.* will have a much lower variance) but might be more biased due to its reliance on possibly non-extreme data points.

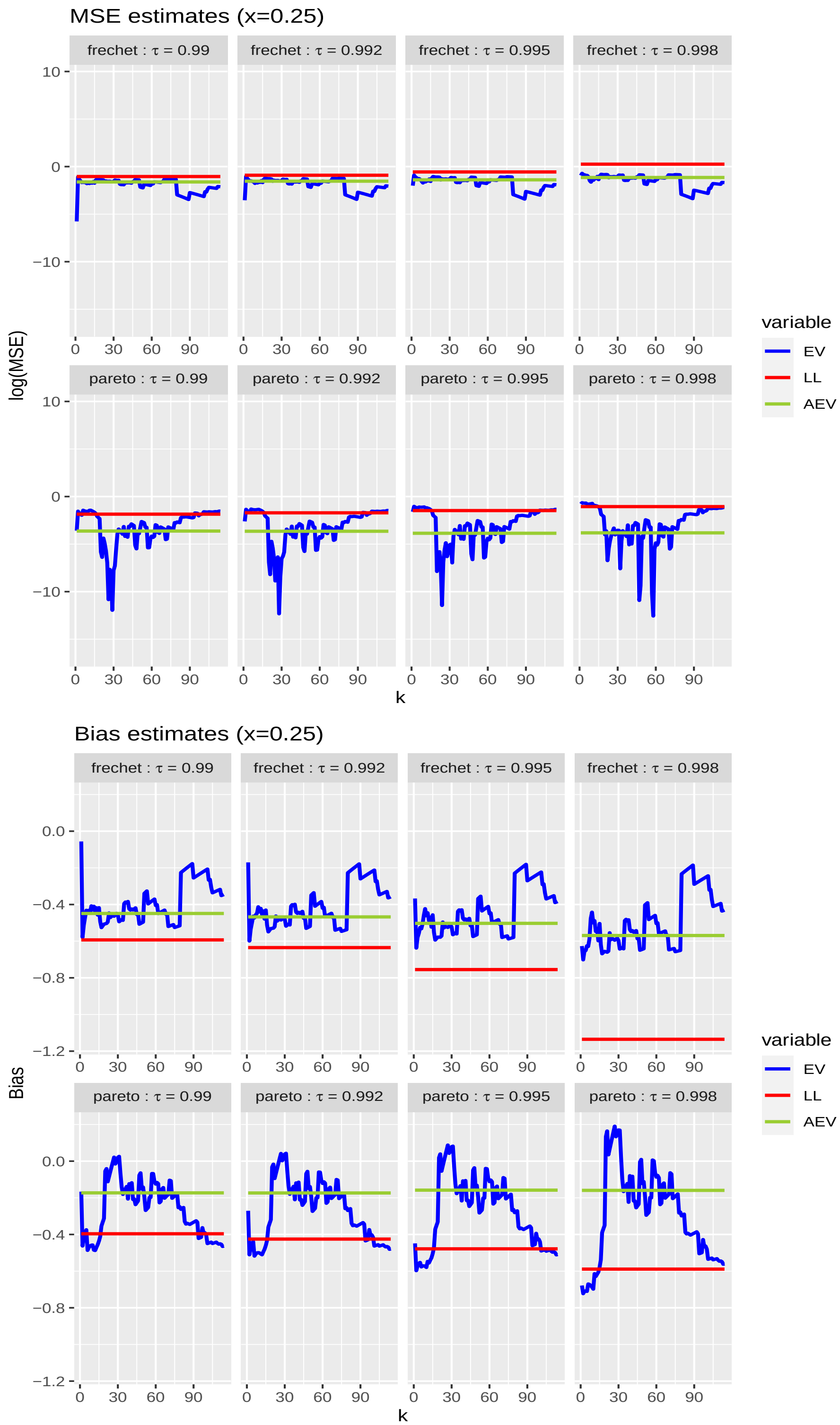

Figure 2: *Averaged MSE in log-scale (top) and bias (bottom) of* $\breve{\xi}_{LL,\tau_n'}(x)/\xi_{\tau_n'}(x)$ *in red (LL) and* $\widehat{\xi}^{\star}_{\tau_n'}(x)/\xi_{\tau_n'}(x)$ *in blue (EV) versus* k*, computed for* $x = 0.25$ *over 200 Monte Carlo simulations. In green (AEV), MSE and bias of the average of the Monte Carlo estimates (EV) within the selected stable region.*

log **MSE** ($x = 0.25$)

$F(\cdot\|x)$	$\tau_n' = 0.99$		$\tau_n' = 0.992$		$\tau_n' = 0.995$		$\tau_n' = 0.998$	
	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$
Fréchet	-1.044	-1.615 $k \in [2, 75]$	-0.908	-1.534 $k \in [2, 75]$	-0.562	-1.391 $k \in [1, 75]$	0.253	-1.145 $k \in [1, 75]$
Pareto	-1.851	-3.628 $k \in [33, 80]$	-1.710	-3.649 $k \in [35, 80]$	-1.474	-3.859 $k \in [30, 80]$	-1.058	-3.817 $k \in [15, 90]$

Bias ($x = 0.25$)

$F(\cdot\|x)$	$\tau_n' = 0.99$		$\tau_n' = 0.992$		$\tau_n' = 0.995$		$\tau_n' = 0.998$	
	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$
Fréchet	-0.593	-0.448 $k \in [2, 75]$	-0.635	-0.467 $k \in [2, 75]$	-0.754	-0.502 $k \in [1, 75]$	-1.135	-0.569 $k \in [1, 75]$
Pareto	-0.396	-0.173 $k \in [33, 80]$	-0.425	-0.173 $k \in [35, 80]$	-0.478	-0.158 $k \in [30, 80]$	-0.588	-0.159 $k \in [15, 90]$

log **MSE** ($x = 0.5$)

$F(\cdot\|x)$	$\tau_n' = 0.99$		$\tau_n' = 0.992$		$\tau_n' = 0.995$		$\tau_n' = 0.998$	
	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$
Fréchet	-7.001	-6.719 $k \in [20, 40]$	-5.304	-5.532 $k \in [15, 45]$	-3.276	-3.603 $k \in [50, 75]$	-1.200	-3.507 $k \in [15, 114]$
Pareto	-2.777	-2.820 $k \in [56, 75]$	-2.498	-2.701 $k \in [50, 75]$	-2.018	-2.041 $k \in [1, 75]$	-1.341	-1.680 $k \in [1, 114]$

Bias ($x = 0.5$)

$F(\cdot\|x)$	$\tau_n' = 0.99$		$\tau_n' = 0.992$		$\tau_n' = 0.995$		$\tau_n' = 0.998$	
	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$
Fréchet	-0.030	-0.003 $k \in [20, 40]$	-0.070	0.038 $k \in [15, 45]$	-0.194	-0.193 $k \in [50, 75]$	-0.548	-0.090 $k \in [15, 114]$
Pareto	-0.249	0.257 $k \in [56, 75]$	-0.286	0.282 $k \in [50, 75]$	-0.364	0.491 $k \in [1, 75]$	-0.511	0.620 $k \in [1, 114]$

log **MSE** ($x = 0.75$)

$F(\cdot\|x)$	$\tau_n' = 0.99$		$\tau_n' = 0.992$		$\tau_n' = 0.995$		$\tau_n' = 0.998$	
	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$
Fréchet	1.156	0.789 $k \in [30, 60]$	1.184	0.906 $k \in [30, 60]$	1.147	0.196 $k \in [30, 90]$	0.621	-0.012 $k \in [30, 114]$
Pareto	-4.965	-5.070 $k \in [22, 43]$	-3.969	-4.330 $k \in [15, 45]$	-2.716	-3.189 $k \in [1, 60]$	-1.418	-2.009 $k \in [45, 60]$

Bias ($x = 0.75$)

$F(\cdot\|x)$	$\tau_n' = 0.99$		$\tau_n' = 0.992$		$\tau_n' = 0.995$		$\tau_n' = 0.998$	
	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$	$\check{\xi}_{\mathrm{LL},\tau_n'}(x)$	$\widehat{\xi}^{\star}_{\tau_n'}(x)$
Fréchet	1.782	1.531 $k \in [30, 60]$	1.807	1.625 $k \in [30, 60]$	1.775	1.310 $k \in [30, 90]$	1.364	1.367 $k \in [30, 114]$
Pareto	-0.083	-0.005 $k \in [22, 43]$	-0.137	-0.061 $k \in [15, 45]$	-0.257	-0.079 $k \in [1, 60]$	-0.491	0.404 $k \in [45, 60]$

Table 1: *Final MSE (in log-scale) and bias estimates of* $\check{\xi}_{LL,\tau_n'}(x)/\xi_{\tau_n'}(x)$ *and* $\widehat{\xi}^{\star}_{\tau_n'}(x)/\xi_{\tau_n'}(x)$*, for* $x \in \{0.25, 0.5, 0.75\}$ *and* $\tau_n' \in \{0.99, 0.992, 0.995, 0.998\}$*. Those of* $\widehat{\xi}^{\star}_{\tau_n'}(x)/\xi_{\tau_n'}(x)$ *are obtained as the average of the Monte Carlo estimates within the selected stable region of* k*-values. The cases where* $\widehat{\xi}^{\star}_{\tau_n'}(x)$ *does not outperform* $\check{\xi}_{LL,\tau_n'}(x)$ *are indicated in blue.*

References

Billingsley, P. (1999). *Convergence of Probability Measures* (second edition), Wiley Series in Probability and Statistics.

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

Daouia, A., Gijbels, I. and Stupfler, G. (2019). Extremiles: A new perspective on asymmetric least squares. *Journal of the American Statistical Association*, **114**, 1366–1381.

Daouia, A., Girard, S. and Stupfler, G. (2020). Tail expectile process and risk assessment. *Bernoulli*, **26**, 531–556.

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

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

Girard, S., Stupfler, G. and Usseglio-Carleve, A. (2020). Nonparametric extreme conditional expectile estimation, available at `https://hal.archives-ouvertes.fr/hal-02114255v3`.

Knight, K. (1999). Epi-convergence in distribution and stochastic equi-semicontinuity. Technical report, University of Toronto.

Li, Q. and Racine, J.S. (2008). Nonparametric Estimation of Conditional CDF and Quantile Functions with Mixed Categorical and Continuous Data. *Journal of Business and Economic Statistics*, **26**, 423–434.

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

Yu, K. and Jones, M.C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, **93**, 228–237.