WORKING PAPERS

N° 1546

May 2024

"Extremile Regression"

Abdelaati Daouia and Gilles Stupfler

Toulouse School of Economics

Extremile Regression

Abdelaati Daouia[a,*], Gilles Stupfler[b†]

[a] Toulouse School of Economics, University of Toulouse Capitole, France
[b] Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France

Abstract

Extremiles are a least squares alternative to quantiles, determined by probability-weighted moments rather than tail probabilities. They benefit from several interpretations and closed form expressions that are equivalent for continuous distributions, and they characterize a distribution just as quantiles do. Their regression versions similarly define a least squares analog of regression quantiles. We give a comprehensive overview of the state of the art regarding probabilistic and statistical properties of unconditional extremiles and their regression counterparts and provide a comparison between extremiles and other important classes of indicators for the description of unconditional and conditional distributions on real data examples.

Extremiles and their equivalent formulations Let F be the cumulative distribution function of a real-valued random variable X. We assume throughout this article that F **is continuous**. The quantile function q of X is the left-continuous generalized inverse of X, that is, $q_\tau = F^{-1}(\tau) := \inf\{x \in \mathbb{R} \,|\, F(x) \geq \tau\}$ for $\tau \in (0,1)$. We have, for any $\tau \in (1/2, 1)$,

$$[F(q_\tau)]^{\log(1/2)/\log(\tau)} = \tau^{\log(1/2)/\log(\tau)} = \exp(\log(1/2)) = 1/2.$$

It is then immediate that q_τ is exactly the median of a random variable Z_τ having cumulative distribution function $K_\tau \circ F$, where, for any $t \in [0,1]$,

$$K_\tau(t) = \begin{cases} 1-(1-t)^{s(\tau)} & \text{if} \quad 0 < \tau \leq 1/2, \\ t^{r(\tau)} & \text{if} \quad 1/2 \leq \tau < 1, \end{cases} \quad \text{with } r(\tau) = s(1-\tau) = \log(1/2)/\log(\tau).$$

The function K_τ is itself a cumulative distribution function with support $[0,1]$. It follows that

$$q_\tau \in \underset{\theta\in\mathbb{R}}{\arg\min}\, \mathbb{E}(|Z_\tau - \theta| - |Z_\tau|) = \underset{\theta\in\mathbb{R}}{\arg\min}\, \mathbb{E}\{J_\tau(F(X))(|X-\theta| - |X|)\},$$

with the weight-generating function $J_\tau(\cdot) = K'_\tau(\cdot)$ on $(0,1)$. This means that q_τ can be viewed as the minimizer of an asymmetric L^1-loss function that is different in nature from the standard proposal of Koenker and Bassett (1978) and Koenker (2005). The extremile of order τ of X is then defined by substituting squared deviations in place of absolute deviations.

Definition 1 (Extremile)**.** Let X have a continuous distribution function F. If X has a finite first moment, the *extremile* of order $\tau \in (0,1)$ is equivalently defined as $\xi_\tau = \mathbb{E}(Z_\tau)$, where Z_τ has cumulative distribution function $K_\tau \circ F$, and as

$$\xi_\tau = \underset{\theta\in\mathbb{R}}{\arg\min}\, \mathbb{E}\{J_\tau(F(X))(|X-\theta|^2 - |X|^2)\}.$$

In other words

$$\xi_\tau = \frac{\mathbb{E}\{XJ_\tau(F(X))\}}{\mathbb{E}\{J_\tau(F(X))\}}.$$

*E-mail: abdelaati.daouia@tse-fr.eu. Orcid: 0000-0003-2621-8860
†Corresponding author. E-mail: gilles.stupfler@univ-angers.fr. Orcid: 0000-0003-2497-9412

The presence of the quantity $-|X|^2$ in the loss function ξ_τ is a minimizer of makes it well-defined and finite as soon as $|X|$ has a finite first moment, as can be seen from the triangle inequality; obviously $\xi_{1/2} = \mathbb{E}(X)$ and $\tau \mapsto \xi_\tau$ is a continuous increasing function mapping $(0, 1)$ onto the support of X.

We now discuss four additional equivalent definitions of the τth extremile of a continuous distribution, in terms of probability-weighted moments, expected extreme values, Choquet integrals, and spectral/distortion risk measures. From the definition of extremiles, it can be seen that

$$\xi_\tau = \mathbb{E}\{X J_\tau(F(X))\},$$

since $\mathbb{E}\{J_\tau(F(X))\} = 1$ by continuity of F (of which a consequence is that $U = F(X)$ has a standard uniform distribution). Consequently

$$\xi_\tau = \begin{cases} s(\tau) M_{1,0,s(\tau)-1} & \text{for} \quad 0 < \tau \leq 1/2, \\ r(\tau) M_{1,r(\tau)-1,0} & \text{for} \quad 1/2 \leq \tau < 1, \end{cases}$$

with

$$M_{p,r,s} = \mathbb{E}\{X^p (F(X))^r (1 - F(X))^s\}.$$

The probability-weighted moments $M_{p,r,s}$, where p, r, s are nonnegative real numbers, were introduced by Greenwood et al. (1979). They form the backbone of several extreme value analysis procedures; see *e.g.* Beirlant et al. (2004) and de Haan and Ferreira (2006).

In the particular case where $\tau \geq 1/2$ and $r(\tau)$ is an integer, it is easily seen that Z_τ has the distribution of the maximum of $r(\tau)$ independent copies of X, and similarly, when $\tau \leq 1/2$ and $s(\tau)$ is an integer, Z_τ has the distribution of $s(\tau)$ independent copies of X. In general,

$$\mathbb{E}\{\max(X_1, \ldots, X_{\lfloor r(\tau) \rfloor})\} \leq \xi_\tau \leq \mathbb{E}\{\max(X_1, \ldots, X_{\lfloor r(\tau) \rfloor + 1})\} \text{ if } \frac{1}{2} \leq \tau < 1$$

and

$$\mathbb{E}\{\min(X_1, \ldots, X_{\lfloor s(\tau) \rfloor + 1})\} \leq \xi_\tau \leq \mathbb{E}\{\min(X_1, \ldots, X_{\lfloor s(\tau) \rfloor})\} \text{ if } 0 < \tau \leq \frac{1}{2},$$

where $\lfloor \cdot \rfloor$ denotes the floor function and $X_1, X_2, \ldots$ are i.i.d. observations from X. We also have $Z_\tau \stackrel{d}{=} \varphi_\tau(X)$, with

$$\varphi_\tau(x) = q_{K_\tau^{-1}(F(x))}.$$

This means that extremiles can be viewed as expected maxima or minima when $\tau \in \{1/\sqrt[k]{2}, 1 - 1/\sqrt[k]{2}, k = 1, 2, \ldots\}$, and are otherwise bracketed by such expected maxima and minima, with the bracketing getting narrower as $\tau \downarrow 0$ or $\tau \uparrow 1$. It follows from this expression of extremiles as expected extreme values that the extremile function $\tau \mapsto \xi_\tau$ characterizes probability distributions having a finite first moment, see Proposition 2(ii) in Daouia et al. (2019).

We finally briefly touch upon the interpretation of extremiles from the point of view of risk measure theory; an expanded discussion can be found in Daouia et al. (2019). Denote by $x_\star = \inf\{x \in \mathbb{R} \,|\, F(x) > 0\}$ and $x^\star = \sup\{x \in \mathbb{R} \,|\, F(x) < 1\}$ the lower and upper endpoints of the support of X, respectively. Then

$$\xi_\tau = \begin{cases} -\displaystyle\int_{x_\star}^{0} \left\{1 - (1 - F(x))^{s(\tau)}\right\} \mathrm{d}x + \int_0^{x^\star} (1 - F(x))^{s(\tau)} \,\mathrm{d}x & \text{for} \quad 0 \leq \tau \leq 1/2, \\ -\displaystyle\int_{x_\star}^{0} (F(x))^{r(\tau)} \,\mathrm{d}x + \int_0^{x^\star} \left\{1 - (F(x))^{r(\tau)}\right\} \mathrm{d}x & \text{for} \quad 1/2 \leq \tau \leq 1. \end{cases}$$

This makes an extremile a Choquet integral. Extremiles can also be viewed as weighted integrals of the quantile function, in the sense that

$$\xi_\tau = \int_0^1 q_t \,\mathrm{d}K_\tau(t) = \int_0^1 J_\tau(t) q_t \,\mathrm{d}t.$$

This, in turn, allows one to view extremiles as spectral/distortion risk measures in the sense of Wang (1996) and Acerbi (2002), and the properties of the function J_τ ensure that extremiles above the mean induce a coherent (namely, location and positive scale equivariant, monotonic, and subadditive) and comonotonically additive risk measure in the sense of Artzner et al. (1999) and Bassett et al. (2004). Since the weight-generating function $J_\tau(\cdot)$ is monotonically increasing for $\tau \geq 1/2$ and decreasing for $\tau \leq 1/2$, the extremile ξ_τ depends by construction on all feasible values of X, putting more weight on its right tail for $\tau \geq 1/2$ and more weight on its left tail for $\tau \leq 1/2$. It follows that ξ_τ is sensitive to the magnitude of extreme values of X for any order $\tau \in (0,1)$, whereas the τth quantile q_τ is determined solely by the probability level τ and not affected by the behavior of the distribution of X beyond q_τ.

Estimation Given a random sample $X_1, X_2, \ldots, X_n$ from the distribution of X, an estimator of the extremile ξ_τ is easily obtained from its formulation as a probability-weighted moment $\xi_\tau = \mathbb{E}\{X J_\tau(F(X))\}$, *viz.*

$$\widehat{\xi}_\tau^{\mathrm{LM}} = \frac{1}{n}\sum_{i=1}^{n} J_\tau\left(\frac{i}{n}\right) X_{i:n},$$

where $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ denotes the ordered version of $(X_1, \ldots, X_n)$. Alternatively, one may estimate ξ_τ by replacing the unknown distribution function F with its empirical version $\widehat{F}_n : x \mapsto n^{-1}\sum_{i=1}^{n} \mathbb{1}\{X_i \leq x\}$ in the weighted integral expression

$$\xi_\tau = \int_0^1 q_t \, \mathrm{d}K_\tau(t) = \int_0^1 F^{-1}(t) \, \mathrm{d}K_\tau(t).$$

This yields the following L-statistic generated by the measure $\mathrm{d}K_\tau$:

$$\widehat{\xi}_\tau^{\mathrm{L}} = \int_0^1 \widehat{q}_t \, \mathrm{d}K_\tau(t) = \sum_{i=1}^{n} \left\{ K_\tau\left(\frac{i}{n}\right) - K_\tau\left(\frac{i-1}{n}\right) \right\} X_{i:n}.$$

This estimator is another L-statistic, whose asymptotic properties are closely linked to those of $\widehat{\xi}_\tau^{\mathrm{LM}}$ since the finite differences built on the function K_τ in the estimator $\widehat{\xi}_\tau^{\mathrm{L}}$ can be approximated using the derivative J_τ of K_τ when n is large.

Yet another estimator of ξ_τ is provided by solving the empirical least squares problem

$$\arg\min_{\theta \in \mathbb{R}} \sum_{i=1}^{n} J_\tau\left(\frac{i}{n}\right) |X_{i:n} - \theta|^2,$$

which yields the closed form expression

$$\widehat{\xi}_\tau^{\mathrm{M}} = \frac{\sum_{i=1}^{n} J_\tau(i/n) X_{i:n}}{\sum_{i=1}^{n} J_\tau(i/n)} = \frac{\widehat{\xi}_\tau^{\mathrm{LM}}}{\frac{1}{n}\sum_{i=1}^{n} J_\tau(i/n)}.$$

Since the denominator in the last equality converges to 1 as $n \to \infty$, the L-statistic $\widehat{\xi}_\tau^{\mathrm{LM}}$ in the numerator is nothing but a linearized variant of the M-estimator $\widehat{\xi}_\tau^{\mathrm{M}}$.

Finally, when $s(\tau)$ or $r(\tau)$ is a positive integer, the extremile ξ_τ can be estimated using unbiased probability-weighted estimators provided by Landwehr et al. (1979): the moments $M_{1,0,s}$ and $M_{1,r,0}$ are respectively estimated by

$$\widehat{M}_{1,0,s} = \frac{1}{n}\sum_{i=1}^{n-s} \left(\prod_{j=1}^{s} \frac{n-i+1-j}{n-j} \right) X_{i:n} \quad \text{and} \quad \widehat{M}_{1,r,0} = \frac{1}{n}\sum_{i=r+1}^{n} \left(\prod_{j=1}^{r} \frac{i-j}{n-j} \right) X_{i:n}.$$

Then the statistic

$$\widehat{\xi}_\tau^{\mathrm{PWM}} = \begin{cases} s(\tau)\widehat{M}_{1,0,s(\tau)-1} & \text{for} \quad 0 < \tau \le 1/2, \\ r(\tau)\widehat{M}_{1,r(\tau)-1,0} & \text{for} \quad 1/2 \le \tau < 1, \end{cases}$$

is an unbiased estimator of the extremile ξ_τ when $\tau \in \{1/\sqrt[k]{2}, 1-1/\sqrt[k]{2}, k = 1, 2, \ldots\}$. It can be seen that the estimators $\widehat{\xi}_\tau^{\mathrm{L}}$, $\widehat{\xi}_\tau^{\mathrm{LM}}$, $\widehat{\xi}_\tau^{\mathrm{M}}$ and $\widehat{\xi}_\tau^{\mathrm{PWM}}$ have the same asymptotic distribution, see Daouia et al. (2019) for a complete discussion. We state below a theoretical result on the consistency, asymptotic normality and Berry-Esséen rate of uniform convergence relative to the estimator $\widehat{\xi}_\tau^{\mathrm{L}}$.

Theorem 1 (Asymptotic theory for the L-statistic $\widehat{\xi}_\tau^{\mathrm{L}}$)**.** *If the X_i are i.i.d. copies of X, then, for any $\tau \in (0,1)$:*

- *If $\mathbb{E}|X|^p < \infty$ for some $p > 1$, then $\lim_{n\to\infty} \widehat{\xi}_\tau^{\mathrm{L}} = \xi_\tau$ with probability 1 as $n \to \infty$.*
- *If $\mathbb{E}|X|^p < \infty$ for some $p > 2$, then $\sqrt{n}(\widehat{\xi}_\tau^{\mathrm{L}} - \xi_\tau) \to \mathcal{N}(0, v_\tau)$ weakly as $n \to \infty$, where*

$$v_\tau = \iint_{[0,1]^2} (\min(s,t) - st) J_\tau(s) J_\tau(t) \, \mathrm{d}F^{-1}(s) \, \mathrm{d}F^{-1}(t).$$

- *If $\mathbb{E}|X|^3 < \infty$, then*

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\frac{\sqrt{n}}{\sqrt{v_\tau}} \left(\widehat{\xi}_\tau^{\mathrm{L}} - \xi_\tau \right) \le t \right) - \Phi(t) \right| = \mathrm{O}(1/\sqrt{n})$$

for any $\tau \in [1-1/\sqrt[3]{2}, 1/\sqrt[3]{2}]$, where Φ stands for the standard normal distribution function.

Real data illustration: Hurricane losses To contrast the estimation of extremiles with the estimation of other popular quantities for the description of a distribution, we consider a dataset on inflation-adjusted (to 1981 using the U.S. Residential Construction Index) hurricane losses that occurred between 1949 and 1980. This sample $(x_1, \ldots, x_n)$, of size $n = 35$, is also considered in Jones and Zitikis (2003). We compare the following estimates:

- The empirical quantile $\widehat{q}_\tau = x_{\lceil n\tau \rceil:n}$, where $\lceil \cdot \rceil$ is the ceiling function.
- The empirical Expected Shortfall

$$\widehat{\mathrm{ES}}_\tau = \frac{1}{\lfloor n(1-\tau) \rfloor} \sum_{i=1}^{\lfloor n(1-\tau) \rfloor} x_{n-i+1:n}.$$

  In other words, $\widehat{\mathrm{ES}}_\tau$ is the sample mean of the observations above the empirical quantile $\widehat{q}_\tau$.
- The empirical expectile

$$\widehat{e}_\tau = \arg\min_{\theta\in\mathbb{R}} \sum_{i=1}^n |\tau - \mathbb{1}\{x_i \le \theta\}| (x_i - \theta)^2$$

  which is the sample counterpart of the population expectile

$$e_\tau = \arg\min_{\theta\in\mathbb{R}} \mathbb{E}\{|\tau - \mathbb{1}\{X \le \theta\}|(X-\theta)^2 - |\tau - \mathbb{1}\{X \le 0\}| X^2\}$$

  as defined in Newey and Powell (1987). Expectiles form another class of asymmetric least squares quantities extending the mean, since $e_{1/2} = \mathbb{E}(X)$.
- The empirical extremile $\widehat{\xi}_\tau^{\mathrm{L}}$.

These four estimates are represented in Figure 1 against the level $\tau \in [1/2, 1]$. It is apparent that while $\widehat{q}_\tau$ and $\widehat{\mathrm{ES}}_\tau$ are non-smooth step functions, the sample expectile and extremile are smooth functions of τ due to their least squares nature; in particular, while in the upper tail a small change in τ can trigger a large jump in the values of the estimated quantile and Expected Shortfall, this is not the case for the extremile and expectile. Moreover, when comparing the four estimated quantities at the same level τ, it can be seen that the Expected Shortfall is substantially larger (and hence, in risk assessment terms, more conservative) than the sample extremile $\widehat{\xi}_\tau$, itself larger than the estimated expectile. One may in fact show that the inequalities $e_\tau < \xi_\tau < \mathrm{ES}_\tau$ are always true for τ sufficiently close to 1 when the underlying distribution is heavy-tailed and has a finite variance, see Daouia et al. (2019).

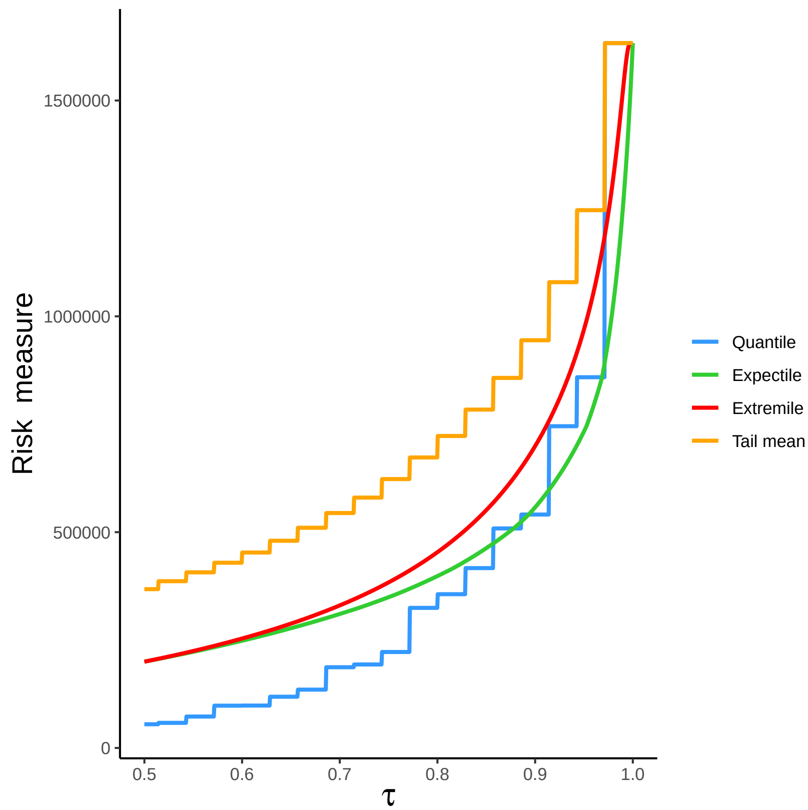

Figure 1: Hurricane losses data ($n = 35$) – Empirical Expected Shortfall $\widehat{\mathrm{ES}}_\tau$ (orange), expectile $\widehat{e}_\tau$ (green), extremile $\widehat{\xi}_\tau^{\mathrm{L}}$ (red) and quantile $\widehat{q}_\tau$ (blue) at level $\tau \in [0.5, 1]$.

Regression extremiles With appropriate care, extremiles can be adapted to the conditional setting where a univariate response variable $Y \in \mathbb{R}$ is recorded alongside a finite-dimensional covariate $\boldsymbol{X} \in \mathbb{R}^p$, so as to understand the influence the variable $\boldsymbol{X}$ has on Y. Recall that an unconditional extremile ξ_τ of Y is the mean of the random variable having distribution function $K_\tau \circ F$, where F is the unconditional distribution function of Y. A conditional version of the extremile can then be defined by replacing F by the conditional distribution function $y \mapsto F(y|\boldsymbol{x}) = \mathbb{P}(Y \leq y|\boldsymbol{X} = \boldsymbol{x})$ (this function is always well-defined by the disintegration theorem for probability measures on $\mathbb{R}^{p+1}$). This suggests the following definition.

Definition 2 (Regression extremile)**.** Let $Y \in \mathbb{R}$ have a conditional continuous distribution function $y \mapsto F(y|\boldsymbol{x})$ given $\boldsymbol{X} = \boldsymbol{x}$, where $\boldsymbol{X}$ is a finite-dimensional covariate and $\boldsymbol{x}$ lies in the support of $\boldsymbol{X}$.

If Y has a finite first conditional moment, the *regression extremile* (or conditional extremile) of order $\tau \in (0,1)$ of Y given $\boldsymbol{X} = \boldsymbol{x}$ is defined as

$$\xi_\tau(\boldsymbol{x}) = \underset{\theta \in \mathbb{R}}{\arg\min}\, \mathbb{E}\{J_\tau(F(Y|\boldsymbol{x}))(|Y-\theta|^2 - |Y|^2) \,|\, \boldsymbol{X} = \boldsymbol{x}\}.$$

Equivalently, $\xi_\tau(\boldsymbol{x})$ is the mean of $K_\tau \circ F(\cdot|\boldsymbol{x})$, and equals the expectation $\mathbb{E}\{Y J_\tau(F(Y|\boldsymbol{x})) \,|\, \boldsymbol{X} = \boldsymbol{x}\}$.

Importantly, the location and positive scale equivariance of extremiles entails that, in a location-scale model of the form $Y = m(\boldsymbol{X}) + \sigma(\boldsymbol{X})\varepsilon$, where $\sigma(\boldsymbol{X}) > 0$ and ε is independent of $\boldsymbol{X}$ and has a finite first moment, we have $\xi_\tau(\boldsymbol{X}) = m(\boldsymbol{X}) + \sigma(\boldsymbol{X})\xi_{\tau,\varepsilon}$, for any $\tau \in (0,1)$, where $\xi_{\tau,\varepsilon}$ denotes the τth unconditional extremile of the error term ε. This means that in homoskedastic regression models, regression extremile curves are parallel to one another, just like regression quantile curves and regression expectile curves are.

Regression extremiles can be estimated using, for instance, local linear estimation. For a generic estimator $\widehat{F}(\cdot|\boldsymbol{x})$ of $F(\cdot|\boldsymbol{x})$, such as a kernel, spline, or wavelet smoothing estimator, and given a kernel function (namely, a probability density function) L on $\mathbb{R}^p$ and a positive bandwidth sequence (h_n), the local linear check function minimization solves the weighted least squares problem

$$\underset{(\alpha,\boldsymbol{\beta}) \in \mathbb{R}\times\mathbb{R}^p}{\arg\min} \sum_{i=1}^n J_\tau(\widehat{F}(Y_i|\boldsymbol{x}))\left(Y_i - \alpha - (\boldsymbol{x}-\boldsymbol{X}_i)^\top \boldsymbol{\beta}\right)^2 L\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_n}\right)$$

to get the estimators $\widehat{\alpha} = \widehat{\xi}_\tau^{\mathrm{LL}}(\boldsymbol{x})$ and $\widehat{\boldsymbol{\beta}}$ of $\xi_\tau(\boldsymbol{x})$ and its gradient at the point $\boldsymbol{x}$, respectively. Weighted least squares theory leads to the explicit solution

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

Here $\boldsymbol{Y}$ is the column vector containing the Y_i, $1 \le i \le n$, the matrix $\mathbf{X}_{\mathrm{LL}}$ is the $n \times (p+1)$ design matrix of the local linear fitting technique, *i.e.* whose first column has all its entries equal to 1, and whose $(j+1)$th column is made of the values $x_j - X_{i,j}$, $1 \le i \le n$, where x_j (resp. $X_{i,j}$) is the jth entry of $\boldsymbol{x}$ (resp. $\boldsymbol{X}_i$), and the diagonal weight matrix $\mathbf{W}_{\widehat{F},L}$ is

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

The asymptotic behavior of $\widehat{F}(\cdot|\boldsymbol{x})$ is naturally crucial in the analysis of the asymptotic and finite-sample behavior of $\widehat{\alpha} = \widehat{\xi}_\tau^{\mathrm{LL}}(\boldsymbol{x})$. Corollary 1 in Daouia et al. (2022) shows that, in the case $p = 1$ of regression upon a one-dimensional covariate, and if certain reasonable regularity conditions on the distributions of X and $Y|X$ are satisfied, then pointwise in x belonging to the interior of the support of X,

$$\sqrt{nh_n}\left(\widehat{\xi}_\tau^{\mathrm{LL}}(x) - \xi_\tau(x)\right) \to \mathcal{N}\left(0, \frac{\int_{\mathbb{R}} L^2}{f_X(x)}V_\tau(x)\right)$$

weakly as $n \to \infty$, where f_X is the probability density function of X and $V_\tau(x) = \mathbb{E}(J_\tau^2(F(Y|x))\{Y - \xi_\tau(x)\}^2 \,|\, X = x)$, as long as the preliminary estimator $\widehat{F}(y|x)$ is $n^{2/5}-$consistent uniformly in $y \in \mathbb{R}$, $h_n \to 0$ is such that $nh_n^5 \to 0$, and $\tau \in (0, 1-1/\sqrt{2}] \cup [1/\sqrt{2}, 1)$. It is interesting to note that the asymptotic variance $V_\tau(x)$ is **not** merely an adaptation to the conditional setup of the asymptotic variance of the L-estimator in Theorem 1, which would be

$$v_\tau(x) = \iint_{[0,1]^2} (\min(s,t) - st)J_\tau(s)J_\tau(t)\,\mathrm{d}F^{-1}(s|x)\,\mathrm{d}F^{-1}(t|x).$$

To see why, suppose that Y given X is uniformly distributed on $[0,\phi(x)]$ for some positive function ϕ, take $\tau > 1/2$ and set $r = r(\tau) > 1$ for the sake of clarity. Then it is a straightforward calculus exercise to get

$$v_\tau(x) = \phi^2(x) \iint_{[0,1]^2} (\min(s,t) - st) J_\tau(s) J_\tau(t)\,\mathrm{d}s\,\mathrm{d}t = \phi^2(x) \frac{r^2}{(r+1)^2(2r+1)}$$

since $J_\tau(s) = rs^{r-1}$. To calculate $V_\tau(x)$, note that

$$\xi_\tau(x) = \int_0^1 J_\tau(t)\, q_t(x)\,\mathrm{d}t = \phi(x) \int_0^1 t\, J_\tau(t)\,\mathrm{d}t = \phi(x)\frac{r}{r+1}.$$

Letting U be a uniform random variable on $[0,1]$, a straightforward calculation then entails

$$\begin{aligned} V_\tau(x) = \mathbb{E}(J_\tau^2(U)\,\{\phi(x)U - \xi_\tau(x)\}^2) &= \phi^2(x) \int_0^1 J_\tau^2(t) \left(t - \frac{r}{r+1}\right)^2 \mathrm{d}t \\ &= \phi^2(x) \frac{r^3}{(r+1)^2(2r-1)(2r+1)}. \end{aligned}$$

Then $V_\tau(x)/v_\tau(x) = r/(2r-1) < 1$. This means that the estimator $\widehat{\xi}_\tau^{\mathrm{LL}}(x)$ is more efficient than the regression version of $\widehat{\xi}_\tau^{\mathrm{L}}$, namely,

$$\int_0^1 \widehat{F}^{-1}(t|x)\,\mathrm{d}K_\tau(t).$$

Guidelines for the automatic choice of bandwidth In the univariate regression setting $p = 1$, an optimal choice of the bandwidth parameter $h = h_{\xi_\tau}$ can be constructed in terms of the optimal bandwidth h_{q_τ} for regression quantile estimation, whose automatic selection is well-established. Further details can be found in the online supplement to Daouia et al. (2022). Since the conditional extremile $\xi_\tau(x)$ and quantile $q_\tau(x) = F^{-1}(\tau|x)$ are, respectively, the mean and the median of the distribution function $K_\tau(F(\cdot|x))$, the bandwidths h_{ξ_τ} and h_{q_τ} correspond to optimal choices of the bandwidth parameter for kernel regression mean and median estimation, respectively. Following Yu and Jones (1998),

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

where $q_\tau''(x)$ and $\xi_\tau''(x)$ are the second derivatives of $x \mapsto q_\tau(x)$ and $x \mapsto \xi_\tau(x)$, respectively, and $\sigma_{Z_\tau^X}^2(x)$ and $f_{Z_\tau^X}(\cdot|x)$ stand for the conditional variance and probability density function relative to the distribution function $K_\tau(F(\cdot|x))$, respectively. A useful rule-of-thumb calculation can then be carried out as in Yu and Jones (1998) by assuming that the conditional distribution of Y given $X = x$ is Gaussian with mean μ_x and variance σ_x^2. This yields

$$\begin{aligned} & f_{Z_\tau^X}(q_\tau(x)|x) = \sigma_x^{-1} J_\tau(\tau) \phi(\Phi^{-1}(\tau)) \\ & \text{and} \;\; \sigma_{Z_\tau^X}^2(x) = \sigma_x^2 V_{K_\tau \circ \Phi} := \sigma_x^2 \int_0^1 \left(\Phi^{-1}(t) - \mu_{K_\tau\circ\Phi}\right)^2 J_\tau(t)\,\mathrm{d}t, \end{aligned}$$

where ϕ and Φ are the standard normal density and distribution functions, $V_{K_\tau\circ\Phi}$ is the variance corresponding to the distribution function $K_\tau \circ \Phi$, and $\mu_{K_\tau\circ\Phi}$ is the mean of this same distribution, or equivalently the τth extremile of Φ, which is independent of x and readily calculated numerically. If one further makes the simplifying assumption that $q_t''(x)$ is constant with respect to t, then $\xi_\tau''(x) = q_\tau''(x)$ and

$$\left(\frac{h_{\xi_\tau}}{h_{q_\tau}}\right)^5 = 4\sigma_{Z_\tau^X}^2(x)(f_{Z_\tau^X}(q_\tau(x)|x))^2 = 4V_{K_\tau\circ\Phi}\{J_\tau(\tau)\phi(\Phi^{-1}(\tau))\}^2.$$

This crucially does not depend on σ_x^2, and suggests the following bandwidth selection algorithm:

- Use ready-made methods to select $h_{\xi_{1/2}} = h_{\mathbb{E}(Y|x)}$, 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` developed by Hayfield and Racine (2008);

- Using the automatic method of Yu and Jones (1998), find the optimal bandwidth for smoothing the τth conditional quantile, that is, $h_{q_\tau} = h_{\xi_{1/2}}\{\tau(1-\tau)/(\phi(\Phi^{-1}(\tau)))^2\}^{1/5}$;

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

Real data illustration: Triceps skinfold length Consider first data, previously analyzed by Yu and Jones (1998), on triceps skinfold measurements of 892 girls and women up to age 50, recorded in three Gambian villages during the dry season of 1989. We compare, for the conditional distribution of triceps skinfold length:

- Regression quantile estimates calculated using the `lprq` routine of the R package `quantreg` maintained by Koenker (2023), with the optimal bandwidth $h = h_{q_\tau}$ chosen by the Yu and Jones (1998) selection method;

- Regression extremile estimates obtained with the bandwidth $h = h_{\xi_\tau}$ calculated using the automatic selection strategy described above, along with 95% pointwise asymptotic confidence intervals;

- Regression expectile estimates of Yao and Tong (1996), first with $h = h_{q_\tau}$ and then with $h = h_{\xi_\tau}$, in the absence of a selection strategy for h in the problem of local linear conditional expectile estimation.

All these estimates are represented, for $\tau \in \{0.01, 0.03, 0.1, 0.25, 0.5, 0.75, 0.9, 0.97, 0.99\}$, in the top panel of Figure 2. The conclusions that can be drawn from the three regression methods broadly concur in the sense that adulthood appears to be correlated with a much greater variability in triceps skinfold length compared to childhood. However, expectiles beyond the regression mean exhibit less evidence of the obvious variation and over-dispersion of triceps skinfold length as age increases, while noncentral extremiles and quantiles are more spread out, suggesting a better ability of the latter two notions to model location and sparseness. Finally, extremile regression estimates and their confidence intervals are smooth in τ and x, which is not the case for quantile regression; moreover, as far as we are aware, there is no ready-made procedure for the construction of pointwise asymptotic confidence intervals for conditional quantiles and expectiles based on the limiting distributions of their local linear estimators. Available devices seem to be mainly based on semiparametric statistics, bootstrap or more sophisticated techniques such as, for instance, in Sobotka et al. (2013).

Real data illustration: Automobile insurance claims The second dataset consists of $n = 1{,}037$ automobile bodily injury claims collected in 2002 by the American Institute for Chartered Property Casualty Underwriters and the Insurance Institute of America. The scatterplot of the claimants' losses *versus* their ages is shown in Figure 2 (bottom panel), along with the local linear quantile, extremile, and expectile smoothers (at the 0.75, 0.9, 0.95, 0.97, 0.99, 0.992, 0.993 and 0.994 levels). In this insurance example, tail regression extremiles appear to be more alert to unexpected high losses than their expectile counterparts. As in the previous data example, they are smoother as a function of the covariate value than regression quantiles; moreover, regression extremile curves do not appear to cross, unlike quantile regression curves, even though crossings are obviously incompatible with the definition of population regression quantiles.

Discussion The state of the art in extremile regression has so far not considered the case where (conditional) distributions are discrete. In this setting, the successive formulations of extremiles that we have highlighted stop being equivalent: for example, when the distribution of X has an atom,

$$\frac{\mathbb{E}\{XJ_\tau(F(X))\}}{\mathbb{E}\{J_\tau(F(X))\}} \neq \mathbb{E}\{XJ_\tau(F(X))\} = \int_0^1 J_\tau(t) q_t \,\mathrm{d}t$$

in general, because $F(X)$ is no longer standard uniform. This raises the natural question of what would be the "canonical" definition of an extremile.

Extremile estimation theory is currently restricted to independent and identically distributed data points. The independent data assumption makes it possible to utilize existing powerful results on L-statistics in order to derive the asymptotic normality of the estimators we have discussed. A natural question, to be considered in particular from the risk analysis viewpoint in finance and insurance, is to extend extremile estimation and inferential theory to the setup where the data-generating process features serial dependence, starting with the mixing conditions of Bradley (2005).

As is the case with quantiles and expectiles, there are several equivalent definitions of extremiles for continuous univariate distributions, but reasonable extensions to the multivariate setting are much harder to come by. In this setting and at least for quantile constructions, one needs to balance axiomatic considerations (Serfling, 2002), interpretability, and computational difficulties. The construction of a well-behaved notion of multivariate extremile keeping one or several of its intuitive interpretations and scalable to high dimensions is a difficult problem that constitutes, in our view, an important avenue of further research.

Acknowledgments This research was supported by the French National Research Agency under the grants ANR-19-CE40-0013 (ExtremReg project), ANR-17-EURE-0010 (EUR CHESS) and ANR-11-LABX-0020-01 (Centre Henri Lebesgue), the TSE-HEC ACPR Chair "Regulation and systemic risks", and the Chair Stress Test, RISK Management and Financial Steering of the Foundation Ecole Polytechnique.

References

Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518.

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

Bassett, G. W., Koenker, R., and Kordas, G. (2004). Pessimistic portfolio allocation and Choquet expected utility. *Journal of Financial Econometrics*, 2(4):477–492.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, Chichester, England.

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

Daouia, A., Gijbels, I., and Stupfler, G. (2019). Extremiles: A new perspective on asymmetric least squares. *Journal of the American Statistical Association*, 114(527):1366–1381.

Daouia, A., Gijbels, I., and Stupfler, G. (2022). Extremile regression. *Journal of the American Statistical Association*, 117(539):1579–1586.

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

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, 15(5):1049–1054.

Hayfield, T. and Racine, J. S. (2008). Nonparametric Econometrics: The np package. *Journal of Statistical Software*, 27(5):1–32.

Jones, B. L. and Zitikis, R. (2003). Empirical estimation of risk measures and related quantities. *North American Actuarial Journal*, 7(4):44–54.

Koenker, R. (2005). *Quantile Regression. Econometric Society Monographs* **38**, Cambridge University Press.

Koenker, R. (2023). *quantreg: Quantile Regression*. R package version 5.97.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). Probability weighted moments compared with some traditional techniques in estimating gumbel parameters and quantiles. *Water Resources Research*, 15(5):1055–1064.

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847.

Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica*, 56(2):214—232.

Sobotka, F., Kauermann, G., Schulze Waltrup, L., and Kneib, T. (2013). On confidence intervals for semiparametric expectile regression. *Statistics and Computing*, 23(2):135–148.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 26(1):71–92.

Yao, Q. and Tong, H. (1996). Asymmetric least squares regression and estimation: A nonparametric approach. *Journal of Nonparametric Statistics*, 6(2–3):273–292.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, 93(441):228–237.

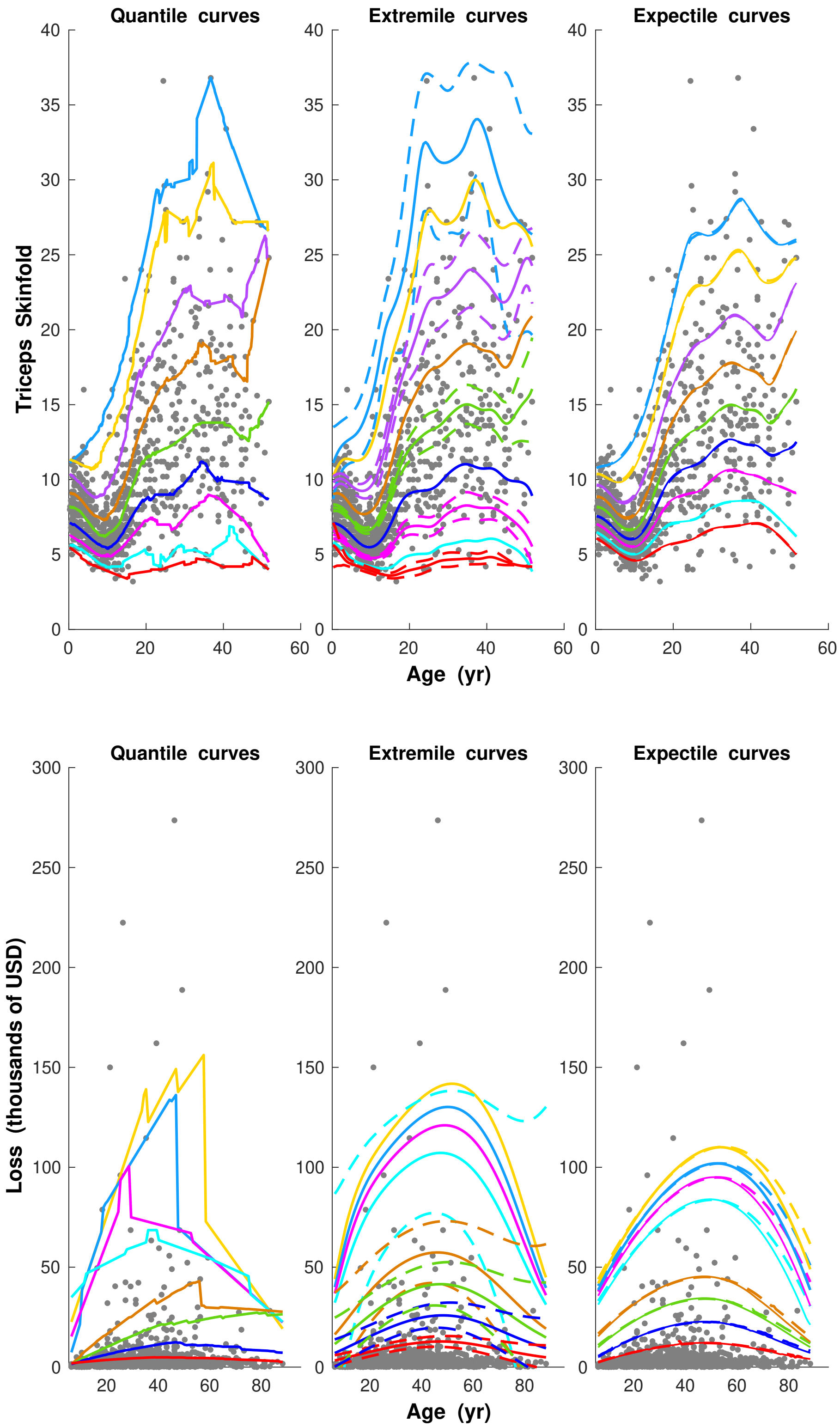

Figure 2: (Top) Triceps skinfold data ($n = 892$) – Smoothed 1%, 3%, 10%, 25%, 50%, 75%, 90%, 97% and 99% quantile (left), extremile (middle) and expectile curves (right) in solid lines, and 95% pointwise asymptotic confidence intervals for $\xi_{0.01}(x)$, $\xi_{0.1}(x)$, $\xi_{0.5}(x)$, $\xi_{0.9}(x)$, $\xi_{0.99}(x)$ in dashed lines. (Bottom) Automobile insurance data ($n = 1{,}037$) – Smoothed 75%, 90%, 95%, 97%, 99%, 99.2%, 99.3% and 99.4% quantile (left), extremile (middle) and expectile curves (right) in solid lines, and 95% pointwise asymptotic confidence intervals for $\xi_{0.75}(x)$, $\xi_{0.9}(x)$, $\xi_{0.95}(x)$, $\xi_{0.97}(x)$ and $\xi_{0.99}(x)$ in dashed lines.