

ESTIMATION FOR EXTREME CONDITIONAL QUANTILES OF FUNCTIONAL QUANTILE REGRESSION

Hanbing Zhu¹, Riquan Zhang¹, Yehua Li² and Weixin Yao²

¹*East China Normal University and* ²*University of California, Riverside*

Abstract: Quantile regression as an alternative to modeling the conditional mean function provides a comprehensive picture of the relationship between a response and covariates. It is particularly attractive in applications focused on the upper or lower conditional quantiles of the response. However, conventional quantile regression estimators are often unstable at the extreme tails, owing to data sparsity, especially for heavy-tailed distributions. Assuming that the functional predictor has a linear effect on the upper quantiles of the response, we develop a novel estimator for extreme conditional quantiles using a functional composite quantile regression based on a functional principal component analysis and an extrapolation technique from extreme value theory. We establish the asymptotic normality of the proposed estimator under some regularity conditions, and compare it with other estimation methods using Monte Carlo simulations. Finally, we demonstrate the proposed method by empirically analyzing two real data sets.

Key words and phrases: Extrapolation, extreme quantile, extreme value theory, functional principal component analysis, functional quantile regression, heavy-tailed distribution.

1. Introduction

With modern technology related to data collection and storage, functional data have become increasingly available in many scientific fields, such as meteorology, chemistry, biomedicine, and neuroimaging (Zhu, Fan and Kong (2014); Yu, Kong and Mizera (2016); Miranda, Zhu and Ibrahim (2018)). The most striking feature of functional data is its inherent infinite dimensionality, which poses challenges both for theoretical analysis and statistical computation, and makes traditional multivariate statistical analysis methods no longer applicable. On the other hand, the infinite-dimensional structure of the data is also a rich source of potential useful information, which brings many opportunities for theoretical research and data application. For this very reason, functional data analysis (FDA) has attracted increasing interest in the statistical research community.

Corresponding author: Riquan Zhang, School of Statistics, East China Normal University, Shanghai 200062, China. E-mail: zhangriquan@163.com.

The monographs of Ramsay and Silverman (2005), Ferraty (2011), and Hsing and Eubank (2015) provide a comprehensive review of FDA statistical methods.

Quantile regression, first introduced by Koenker and Bassett (1978), is now widely used to analyze the effect of covariates on the conditional distribution of a response variable. By directly modeling the conditional quantile function, a quantile regression is more robust to outliers in response measurements, can better handle heterogeneity in the data, and provides a more comprehensive regression analysis than that of an ordinary least squares regression. As a result, quantile regression has attracted considerable applications in a variety of fields in recent decades (Koenker and Xiao (2002); Koenker (2005, 2017)).

In the FDA literature, works exist on estimating conditional quantiles using functional covariates. For example, Cardot, Crambes and Sarda (2005) considered a smoothing splines estimation for functional linear quantile regression models. Ferraty, Rabhi and Vieu (2005) studied the functional nonparametric estimation of conditional quantiles when the covariates are functional predictors. Chen and Müller (2012) proposed an “indirect” estimation approach for conditional quantiles by inverting an estimated conditional distribution function. Kato (2012) considered a functional principal component analysis (FPCA)-based estimation of a functional linear quantile regression, and established optimal convergence rates for his estimators in a minimax sense. Yao, Sue-Chee and Wang (2017) focused on a partially functional linear quantile regression with high-dimensional scalar covariates.

The aforementioned works on functional quantile regression have been restricted to a central quantile $\tau \in [\epsilon, 1 - \epsilon]$, where $0 < \epsilon < 1$ is a fixed positive number. This rules out studying the extreme tail quantiles of the response distribution. However, in many applications, such as the study of heavy rainfall (Friederichs and Hense (2007); Gardes and Girard (2010)), extreme warm temperatures (Dupuis, Sun and Wang (2015)), large portofolio losses (Odening and Hinrichs (2003); Schaumburg (2012)), and significant changes in the price of oil (Marimoutou, Raggad and Trabelsi (2009)), an important problem is to model and predict events that are rare, but that have significant social and economic impacts, which correspond to the upper or lower tails of the distribution. To address this problem, we need to study the extreme tail quantiles of the underlying distributions. However, data sparsity makes estimating extreme tail quantiles difficult, especially for heavy-tailed distributions. To the best of our knowledge, estimating the extreme conditional quantiles of a functional linear quantile regression with heavy-tailed distributions has not yet been considered in the literature. The aim of this study is to fill this gap.

In this paper, we develop a novel estimator for the extreme conditional quantiles of a functional linear quantile regression by combining ideas from FPCA and extreme value theory (EVT). The numerical studies and two real data applications demonstrate the superior performance of the new estimator compared with other competing estimators. EVT provides a useful mathematical tool for studying extreme events. Some work has been done on estimating extreme quantiles by integrating a quantile regression with EVT. For instance, Chernozhukov (2005) considered estimating the extreme quantiles of linear quantile regression models based on EVT. Li and Wang (2019) proposed a new estimator for extreme conditional quantiles of time series data. Zhang (2018) established a new asymptotic theory and inference approach for extremal quantile treatment-effect estimators. He, Cheng and Tong (2016) and Yoshida (2019) estimated the extremal quantiles of linear and nonparametric quantile regression models, respectively.

The rest of this paper is organized as follows. In Section 2, we present our proposed model and develop the estimation procedure. The theoretical properties of the estimators are studied in Section 3. In Section 4, we conduct Monte Carlo studies to demonstrate the finite-sample performance of our estimator, and compare it with that of other competing estimators. In Section 5, we demonstrate the proposed method by empirical analyses of two real data sets. Section 6 concludes the paper. All detailed proofs are deferred to the Supplementary Material.

2. Model and Estimation Method

2.1. Model

Let Y be a continuous scalar response variable of interest, and $X(\cdot)$ be a squared integrable and smooth random process supported on a closed interval \mathcal{I} . The conditional distribution of Y given $X = x$ is denoted by $F_Y(\cdot|x) = \mathbb{P}(Y \leq \cdot | X = x)$. Then, the τ th conditional quantile of Y given $X = x$ is defined as

$$Q_Y(\tau|x) = \inf\{y : F_Y(y|x) \geq \tau\}.$$

Suppose that we have a random sample Z_1, \dots, Z_n from some distribution $F(\cdot)$. If there exist sequences of constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\max_{1 \leq i \leq n} Z_i - b_n}{a_n} \leq z \right) = G_\gamma(z)$$

for each continuity point z of $G_\gamma(\cdot)$, where

$$G_\gamma(z) = \begin{cases} \exp\{-(1+\gamma z)^{-1/\gamma}\}, & \text{for } \gamma \neq 0, 1+\gamma z > 0, \\ \exp(-e^{-z}), & \text{for } \gamma = 0, \end{cases}$$

then $F(\cdot)$ is said to belong to the maximum domain of attraction (MDA) of an extreme value distribution G_γ , denoted as $F(\cdot) \in MDA(G_\gamma)$. The real-valued parameter γ is referred to as the extreme value index, and it dominates the tail behavior of the distribution G_γ .

In this study, we assume that the conditional distribution $F_Y(\cdot|x) \in MDA(G_\gamma)$. The MDA assumption is common in the literature on EVT, and is satisfied by most commonly used continuous distributions (Matthys and Beirlant (2003); Wang, Li and He (2012); Gomes and Guillou (2015); Gardes (2018)). For example, the uniform, Gaussian, t distribution, Pareto, and Cauchy belong to the MDA with various values of γ . For more details on the extreme value index and the MDA assumption, refer to de Haan and Ferreira (2006) and the references therein.

We focus on the case where $\gamma > 0$ (i.e., heavy-tailed distributions), for which the estimation of extreme quantiles is especially challenging because of data sparsity in the tail region. Without loss of generality, we focus on estimating the extremely high conditional quantiles of Y given $X = x$, that is, $Q_Y(\tau_n|x)$, where $\tau_n \rightarrow 1$ as $n \rightarrow \infty$. To this end, we assume that there exists a $\tau_0 \in (0, 1)$ such that

$$Q_Y(\tau|x) = \alpha(\tau) + \langle x, \beta_0 \rangle \text{ for } \tau \in [\tau_0, 1),$$

where $\langle x, \beta_0 \rangle = \int_{\mathcal{I}} x(t)\beta_0(t)dt$. Note that we only assume that the conditional quantile function has a functional linear structure when the quantile level exceeds τ_0 , which allows us to model the conditional quantile function without specifying a structure for $\tau \in (0, \tau_0)$. In addition, because τ_0 can be very close to one and the smooth slope function of the functional predictor varies little on the small interval $[\tau_0, 1)$, it is reasonable to assume that the quantile slope functions are common at the upper conditional quantiles of the response.

2.2. Estimation method

In what follows, the covariance function of X is defined as $K(s, t) = \text{Cov}\{X(s), X(t)\}$, for all $s, t \in \mathcal{I}$. We suppose that $K(s, t)$ is continuous and positive definite on $\mathcal{I} \times \mathcal{I}$. Then, $K(s, t)$ has the following eigen-decomposition (Hsing and Eubank (2015)):

$$K(s, t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s) \phi_j(t),$$

where $\lambda_1 > \lambda_2 > \dots > 0$ are eigenvalues, and $\{\phi_j\}_{j=1}^\infty$ is an orthonormal basis of $L_2(\mathcal{I})$ consisting of the eigenfunctions of the integral operator from $L_2(\mathcal{I})$ to itself with the kernel $K(s, t)$; that is,

$$\int_{\mathcal{I}} K(\cdot, t)\phi_j(t)dt = \lambda_j\phi_j(\cdot), \quad j \geq 1.$$

For notational convenience, we assume throughout that the functional covariate X has been centered. Because $\{\phi_j\}_{j=1}^\infty$ is an orthonormal basis of $L_2(\mathcal{I})$, we have the following expansion in $L_2(\mathcal{I})$:

$$X(t) = \sum_{j=1}^\infty \xi_j\phi_j(t), \quad \beta_0(t) = \sum_{j=1}^\infty \beta_j\phi_j(t),$$

where the principal component scores $\xi_j = \langle X, \phi_j \rangle$ are uncorrelated with $\mathbb{E}(\xi_j) = 0$ and $\text{Var}(\xi_j) = \lambda_j$, and $\beta_j = \langle \beta_0, \phi_j \rangle$. Suppose that $(X_1, Y_1), \dots, (X_n, Y_n)$ is a random sample from (X, Y) . Based on the sample X_1, \dots, X_n , the standard empirical estimation of $K(s, t)$ is

$$\widehat{K}(s, t) = \frac{1}{n} \sum_{i=1}^n X_i(s)X_i(t).$$

Let $\widehat{K}(s, t) = \sum_{j=1}^\infty \widehat{\lambda}_j \widehat{\phi}_j(s)\widehat{\phi}_j(t)$ be the eigen-decomposition of $\widehat{K}(s, t)$ (Yao, Müller and Wang (2005); Xiao et al. (2016)), where $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \dots \geq 0$ are eigenvalues, and $\{\widehat{\phi}_j\}_{j=1}^\infty$ is an orthonormal basis of $L_2(\mathcal{I})$ consisting of the eigenfunctions of the integral operator from $L_2(\mathcal{I})$ to itself with the kernel $\widehat{K}(s, t)$. Let ξ_{ij} be the principal component scores of X_i , namely, $\xi_{ij} = \langle X_i, \phi_j \rangle$, where each ξ_{ij} is estimated by $\widehat{\xi}_{ij} = \langle X_i, \widehat{\phi}_j \rangle$.

Our proposed estimator for β_0 is as follows:

$$\widehat{\beta}_0(t) = \sum_{j=1}^{m_n} \widehat{\beta}_j \widehat{\phi}_j(t),$$

where m_n is a cut-off level that satisfies $1 \leq m_n \leq n - 1$ and $m_n \rightarrow \infty$ as $n \rightarrow \infty$ (Yao, Müller and Wang (2005); Li, Wang and Carroll (2013); Zhu et al. (2019)), and $\widehat{\beta}_j$ is defined by

$$\left\{ \widehat{\alpha}(\tau_1), \dots, \widehat{\alpha}(\tau_l), \widehat{\beta}_1, \dots, \widehat{\beta}_{m_n} \right\} = \underset{a_1, \dots, a_l, b_1, \dots, b_{m_n}}{\operatorname{argmin}} \sum_{j=1}^l \sum_{i=1}^n \rho_{\tau_j} \left(Y_i - a_j - \sum_{q=1}^{m_n} \widehat{\xi}_{iq} b_q \right), \tag{2.1}$$

where $\rho_\tau(s) = \tau s - s\mathbf{1}(s < 0)$ is the check loss function and $\mathbf{1}(\cdot)$ is the indicator function, and $\tau_0 = \tau_1 < \tau_2 < \cdots < \tau_l = \tau^0$ is a sequence of quantile levels with $\tau_0 < \tau^0 < 1$. The estimator $(\widehat{\beta}_1, \dots, \widehat{\beta}_{m_n})^\top$ of $(\beta_1, \dots, \beta_{m_n})^\top$ is obtained by pooling information from multiple quantile levels (Koenker (2004); Zou and Yuan (2008)). The objective function in (2.1) can be solved efficiently using the iterative algorithm proposed by Hunter and Lange (2000).

Assume that $F_Y(\cdot|X_i)$ are continuous and strictly monotone. Let $\widehat{e}_i = Y_i - \langle X_i, \widehat{\beta}_0 \rangle$ and $v_i \in \{\tau : Q_Y(\tau|X_i) = Y_i\}$. Then, $Q_Y(v_i|X_i) = Y_i$ and $v_i \sim \text{Uniform}(0,1)$, for $i = 1, \dots, n$. It is easy to show that

$$\widehat{e}_i = \begin{cases} \alpha(v_i) + \langle X_i, \beta_0 - \widehat{\beta}_0 \rangle & \text{if } \tau_0 \leq v_i < 1, \\ Q_Y(v_i|X_i) - \langle X_i, \widehat{\beta}_0 \rangle & \text{otherwise.} \end{cases} \quad (2.2)$$

Denote the order statistics of $\{\widehat{e}_1, \dots, \widehat{e}_n\}$ by $\widehat{e}_{(1)} \leq \widehat{e}_{(2)} \leq \cdots \leq \widehat{e}_{(n)}$. Then, a Hill estimator (Hill (1975)) of the extreme value index γ is given by

$$\widehat{\gamma} = \frac{1}{k_n} \sum_{j=1}^{k_n} \log \frac{\widehat{e}_{(n-j+1)}}{\widehat{e}_{(n-k_n)}}, \quad (2.3)$$

where $k_n \rightarrow \infty$ and $k_n/n \rightarrow 0$ as $n \rightarrow \infty$. The intuitive argument behind this estimator is that the upper-order statistics $\{\widehat{e}_{(n-k_n)}, \dots, \widehat{e}_{(n)}\}$ of $\{\widehat{e}_1, \dots, \widehat{e}_n\}$ are asymptotically equivalent to those of $\{Q_Y(v_i|x=0), i = 1, \dots, n\}$; see the proof of Theorem 2 in the Supplementary Material.

Let $H_Y(t|x=0) = \inf\{y : F_Y(y|x=0) \geq 1 - 1/t\} = F_Y^{-1}(1 - 1/t|x=0)$ for $t > 1$; that is, $H_Y(t|x=0)$ is the $(1 - 1/t)$ th quantile of $F_Y(\cdot|x=0)$. By Theorem 1.1.6 and Lemma 1.2.9 in de Haan and Ferreira (2006), for a heavy-tailed distribution $F_Y(\cdot|x=0)$, when $t \rightarrow \infty$, we have

$$\frac{H_Y(tz|x=0)}{H_Y(t|x=0)} \rightarrow z^\gamma, \quad (2.4)$$

for any $z > 0$. It is easy to obtain that $\alpha(\tau) = H_Y\{1/(1 - \tau)|x=0\}$, for $\tau \in [\tau_0, 1)$. Therefore,

$$\begin{aligned} & \frac{H_Y(1/(1 - p_n) \cdot (1 - p_n)/(1 - \tau_n)|x=0)}{H_Y(1/(1 - p_n)|x=0)} \cdot \left(\frac{1 - \tau_n}{1 - p_n}\right)^\gamma \\ &= \frac{H_Y(1/(1 - \tau_n)|x=0)}{H_Y(1/(1 - p_n)|x=0)} \cdot \left(\frac{1 - \tau_n}{1 - p_n}\right)^\gamma = \frac{\alpha(\tau_n)}{\alpha(p_n)} \cdot \left(\frac{1 - \tau_n}{1 - p_n}\right)^\gamma \rightarrow 1, \text{ as } n \rightarrow \infty, \end{aligned} \quad (2.5)$$

where $\tau_n \rightarrow 1$ as $n \rightarrow \infty$ and $p_n = 1 - k_n/n$. Inspired by (2.5), we estimate $\alpha(\tau_n)$ by

$$\widehat{\alpha}(\tau_n) = \left(\frac{1 - p_n}{1 - \tau_n} \right)^{\widehat{\gamma}} \widehat{\alpha}(p_n), \tag{2.6}$$

where $\widehat{\alpha}(p_n) = \widehat{e}_{(n-k_n)}$ is an estimator of $\alpha(p_n)$, and the p_n th quantile of $F_Y(\cdot|x = 0)$. Here, (2.6) is a Weissman estimator (Weissman (1978)) of $\alpha(\tau_n)$. Note that τ_n is assumed to be much closer to one than p_n (please also refer to the condition in Theorem 3). In (2.6), we use the EVT to extrapolate the tail estimation from p_n to τ_n in order to provide a stable estimate of $\alpha(\tau_n)$, which is otherwise difficult to estimate, owing to data sparsity in the extreme tails. More justifications on using (2.6) to estimate extreme quantiles instead of a sample quantile of \widehat{e}_i can be found in Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011). Hence, we estimate the τ_n th conditional quantile of Y given $X = x$ by

$$\widehat{Q}_Y(\tau_n|x) = \widehat{\alpha}(\tau_n) + \langle x, \widehat{\beta}_0 \rangle.$$

Our simulations suggest that the extreme conditional quantiles obtained using our proposed estimation approach are much more stable than those obtained using the conventional functional quantile regression estimation method (Kato (2012)) or functional kernel estimation approach (Gardes and Girard (2012)).

3. Theoretical Results

For notational simplicity, let $F_\tau(\cdot|X_i)$ be the conditional distribution of $\varepsilon_{i\tau}$ given X_i , and let $f_\tau(\cdot|X_i)$ be the corresponding density function, where $\varepsilon_{i\tau} = Y_i - \alpha(\tau) - \langle X_i, \beta_0 \rangle$ and $\tau \in [\tau_0, 1)$. To study the asymptotic properties of the estimators, some regularity conditions are needed. The following assumptions are imposed. Throughout the paper, C denotes a generic positive constant, the value of which varies in different places.

(C1) $\mathbb{E}(\|X\|^4) < \infty$, where $\|X\|^2 = \int_{\mathcal{I}} X^2(t)dt$.

(C2) $\mathbb{E}(\xi_j^4) \leq C\lambda_j^2$, for $j \in \{1, 2, \dots\}$.

(C3) For some $\nu > 1$, $C^{-1}j^{-\nu} \leq \lambda_j \leq Cj^{-\nu}$ and $\lambda_j - \lambda_{j+1} \geq Cj^{-\nu-1}$, for $j \in \{1, 2, \dots\}$.

(C4) For some $\zeta > \nu/2 + 1$, $|\beta_j| \leq Cj^{-\zeta}$, for $j \in \{1, 2, \dots\}$.

(C5) $m_n \asymp n^{1/(\nu+2\zeta)}$. For positive r_n and s_n , $r_n \asymp s_n$ means that r_n/s_n is bounded away from zero and infinity.

(C6) $\varepsilon_{i\tau}$ are independent and identically distributed (i.i.d.), and $f_\tau(\cdot|X_i)$ and their derivatives are continuous and bounded away from zero and infinity uniformly over i in an interval that contains zero for all $\tau \in [\tau_0, \tau^0]$.

(C7) The first derivative $H'_Y(\cdot|x=0)$ of $H_Y(\cdot|x=0)$ exists, and satisfies $\lim_{t \rightarrow \infty} tH'_Y(t|x=0)/H_Y(t|x=0) = \gamma$.

Remark 1. If the FPCA approach is used to analyze a functional linear regression, (C1)–(C5) are standard conditions (Cai and Hall (2006); Hall and Horowitz (2007)). Condition (C2) is automatically satisfied if X is a Gaussian random process, because in this case, ξ_j are Gaussian. In assumption (C3), ν measures the smoothness of the covariance function K . The second part of assumption (C3) requires that the spaces between λ_j are not too small to ensure that each individual ϕ_j is identifiable. Assumption (C4) determines the smoothness of the slope function β_0 . Assumption (C5) is a technical condition for proving Theorem 1. For more discussion on these assumptions, see Hall and Horowitz (2007). Assumption (C6) is a regularity condition on error quantiles (Kato (2012; Yao, Sue-Chee and Wang (2017))), and is common in the quantile regression literature when X is a multivariate covariate (Kai, Li and Zou (2011); Ma and He (2016)). Condition (C7) is a technical condition for proving Theorem 2, and is called the von Mises condition (see Corollary 1.1.12 in de Haan and Ferreira (2006)). Indeed, $\lim_{t \rightarrow \infty} tH'_Y(t|x=0)/H_Y(t|x=0) = C$ is enough to complete our proof of Theorem 2, where C is an arbitrary fixed positive constant.

In addition to the above assumptions, we also need to introduce a second-order condition on the distribution $F \in MDA(G_\gamma)$, where F is continuous and strictly monotone. Define $V(t) = \inf\{y : F(y) \geq 1 - 1/t\} = F^{-1}(1 - 1/t)$, for $t > 1$. Then, $F \in MDA(G_\gamma)$ implies that there exists a positive function $a_1(\cdot)$ such that for $z > 0$, as $t \rightarrow \infty$,

$$\frac{V(tz) - V(t)}{a_1(t)} \rightarrow \frac{z^\gamma - 1}{\gamma}, \quad (3.1)$$

where for $\gamma = 0$ the right-hand side is interpreted as $\log z$. For more details on (3.1), see Theorem 1.1.6 in de Haan and Ferreira (2006).

To get the asymptotic results, it is usually necessary to assume that the following second-order condition holds (see Corollary 2.3.4 and Theorem 2.3.12 in de Haan and Ferreira (2006)):

$$\frac{(V(tz) - V(t))/a_1(t) - (z^\gamma - 1)/\gamma}{a_2(t)} \rightarrow \frac{1}{\delta} \left(\frac{z^{\gamma+\delta} - 1}{\gamma + \delta} - \frac{z^\gamma - 1}{\gamma} \right), \quad (3.2)$$

for any $z > 0$ as $t \rightarrow \infty$, where $a_2(\cdot)$ is a positive or negative function and $a_2(t) \rightarrow 0$ as $t \rightarrow \infty$, and $a_2(\cdot) \in RV(\delta)$ with $\delta \leq 0$. Here, $a_2(\cdot) \in RV(\delta)$ means that $a_2(\cdot)$ is a regularly varying (RV) function with index δ ; that is, namely, $\lim_{t \rightarrow \infty} a_2(tz)/a_2(t) = z^\delta$, for all $z > 0$. When at least one of δ and γ is equal to zero, the right-hand side of (3.2) is equal to

$$\begin{cases} \frac{1}{\gamma} \left(z^\gamma \log z - \frac{z^\gamma - 1}{\gamma} \right) & \text{if } \delta = 0 \neq \gamma, \\ \frac{1}{\delta} \left(\frac{z^\delta - 1}{\delta} - \log z \right) & \text{if } \delta \neq 0 = \gamma, \\ \frac{1}{2} (\log z)^2 & \text{if } \delta = 0 = \gamma. \end{cases}$$

Most commonly used families of continuous distributions satisfy condition (3.2). For instance, a t distribution with degrees of freedom v satisfies (3.2) with $\gamma = 1/v$ and $\delta = -2/v$, and a normal distribution satisfies (3.2) with $\gamma = \delta = 0$.

For $\gamma > 0$, the second-order condition (3.2) is equivalent to there existing a positive or negative function $A(\cdot)$ with $\lim_{t \rightarrow \infty} A(t) = 0$ and $A(t) \in RV(\delta)$, with $\delta \leq 0$, such that for all $z > 0$,

$$A(t)^{-1} \left\{ \frac{V(tz)}{V(t)} - z^\gamma \right\} \rightarrow z^\gamma \frac{z^\delta - 1}{\delta}, \text{ as } t \rightarrow \infty. \tag{3.3}$$

We say that $V(\cdot)$ satisfies the second-order condition indexed by (γ, δ, A) when (3.3) holds.

The following Theorems 1 and 2 present, respectively, the convergence rate of the estimator for the slope function and the asymptotic normality of the estimator for the extreme value index.

Theorem 1. *Suppose that conditions (C1)–(C6) hold. Then, we have*

$$\left\| \widehat{\beta}_0 - \beta_0 \right\|^2 = O_p \left(n^{-2\zeta - 1/(\nu + 2\zeta)} \right).$$

Theorem 2. *Let $k_n \rightarrow \infty$, $k_n/n \rightarrow 0$, and $k_n^{\gamma+2} n^{-\gamma - (2\zeta - 2 - \nu)/(4(\nu + 2\zeta))} \rightarrow 0$ as $n \rightarrow \infty$. Suppose that $H_Y(\cdot|x = 0)$ satisfies the second-order condition indexed by (γ, δ, A) , and $\gamma > 0$, $\delta < 0$, and $\sqrt{k_n} A(n/k_n) \rightarrow \eta \in \mathbb{R}$. Then, under conditions (C1)–(C7), we have*

$$\sqrt{k_n} (\widehat{\gamma} - \gamma) \xrightarrow{d} N \left(\frac{\eta}{1 - \delta}, \gamma^2 \right).$$

Remark 2. Deriving the asymptotic properties of the estimators requires some

general conditions on k_n , the number of tail observations used in the estimation procedure. The condition $\lim_{n \rightarrow \infty} k_n^{\gamma+2} n^{-\gamma-(2\zeta-2-\nu)/(4(\nu+2\zeta))} = 0$ is used to make sure that the upper-order statistics of $\{\widehat{e}_1, \dots, \widehat{e}_n\}$ behave similarly to those of $\{Q_Y(v_i|x=0), i=1, \dots, n\}$. It is easy to see that $k_n = n^\alpha$ satisfies all the conditions on k_n , where $0 < \alpha < \min(2\delta/(2\delta-1), (4\gamma(\nu+2\zeta)+2\zeta-2-\nu)/(4(\gamma+2)(\nu+2\zeta)))$. Hence, in theory, there exists a wide range of choices for a proper k_n . For instance, for the $t(1)$ distribution with $\gamma = 1$ and $\delta = -2$, $\min(2\delta/(2\delta-1), (4\gamma(\nu+2\zeta)+2\zeta-2-\nu)/(4(\gamma+2)(\nu+2\zeta))) > 1/3$. In Section 4, we recommend a practical choice of k_n that performs well in our simulation studies.

The following Theorems 3 and 4 present the asymptotic normality of $\widehat{\alpha}(\tau_n)$ and $\widehat{Q}_Y(\tau_n|x)$, respectively.

Theorem 3. Assume that $nq_n = o(k_n)$ and $|\log(nq_n)| = o(\sqrt{k_n})$, where $q_n = 1 - \tau_n$. Then, under the same assumptions as in Theorem 2, we have

$$\frac{\sqrt{k_n}}{\log\{k_n/(nq_n)\}} \left\{ \frac{\widehat{\alpha}(\tau_n)}{\alpha(\tau_n)} - 1 \right\} \xrightarrow{d} N\left(\frac{\eta}{1-\delta}, \gamma^2\right).$$

Theorem 4. Assume that $\sqrt{k_n}[\log\{k_n/(nq_n)\}]^{-1} q_n^\gamma n^{-(2\zeta-2-\nu)/(4(\nu+2\zeta))} \rightarrow 0$. Under the conditions of Theorem 3, we have

$$\frac{\sqrt{k_n}}{\log\{k_n/(nq_n)\}} \left\{ \frac{\widehat{Q}_Y(\tau_n|x)}{Q_Y(\tau_n|x)} - 1 \right\} \xrightarrow{d} N\left(\frac{\eta}{1-\delta}, \gamma^2\right).$$

Remark 3. The condition $nq_n = o(k_n)$ guarantees that τ_n th quantile is an extreme quantile level and gives the upper bound on q_n . The condition $|\log(nq_n)| = o(\sqrt{k_n})$ gives the lower bound on q_n , which restricts the range of extrapolation. Obviously, $q_n \asymp n^{-1}$ satisfies both of these conditions. The asymptotic normality of $\widehat{Q}_Y(\tau_n|x)$ even holds for some $\tau_n > 1 - 1/n$, which means that it is beyond the range of the available data.

4. Simulation Studies

4.1. Tuning parameter selection

The proposed estimation method depends on the tuning parameter m_n and the number of upper-order statistics k_n used in the extreme value index estimation. We use the Bayesian information criterion (BIC) to select m_n . Specifically, for a fixed m_n , the BIC is defined as

$$\text{BIC}(m_n) = \log \left\{ \sum_{j=1}^l \sum_{i=1}^n \rho_{\tau_j} \left(Y_i - \hat{\alpha}(\tau_j) - \sum_{q=1}^{m_n} \hat{\xi}_{iq} \hat{\beta}_q \right) \right\} + \frac{m_n \log n}{n}.$$

The m_n that minimizes $\text{BIC}(m_n)$ is the cut-off level we choose.

In the EVT, the optimal k_n is often chosen to minimize the mean squared error (MSE) of the proposed estimator (de Haan and Ferreira (2006)). However, the optimal k_n depends on the unknown extreme value index γ and the unknown regularly varying function $A(\cdot)$ in (3.3), which are difficult to estimate in practice. Drees, de Haan and Resnick (2000) and de Sousa and Michailidis (2004) proposed plotting the value of the extreme value index γ estimator as a function of k_n and selecting a “stable” point, which is called the Hill plot approach. Hall (1990) and Beirlant, Vynckier and Teugels (1996) suggested choosing k_n based on a bootstrap method and regression diagnostics on a Pareto quantile plot, respectively. More discussion on the choice of k_n can be found in Guillou and Hall (2001), Gomes, Figueiredo and Neves (2012), Gardes, Guillou and Schorgen (2012), and Bader, Yan and Zhang (2018). Even though many approaches have been proposed to select k_n , there is no widely accepted method, and its choice requires more research. In our study, combining our numerical investigation and the Hill plot approach, we set $k_n = \lceil 4n^{1/4} \rceil$ for all of our numerical studies, where $\lceil \cdot \rceil$ denotes the integer part. Our simulations suggest that both the extreme value index and the extreme conditional quantiles obtained by our proposed estimation approach perform well under such a choice.

4.2. Numerical results

We now conduct simulation studies to verify the performance of our proposed method (EFQR). We also illustrate that the proposed method improves on the functional kernel estimate (FKE) proposed by Gardes and Girard (2012) and the conventional functional linear quantile regression estimate (CQR) at high tails. The data that we use are generated from the following model:

$$Y_i = \int_0^1 X_i(t) \beta_0(t) dt + \varepsilon_i, \quad i = 1, \dots, n,$$

where $\beta_0(t) = \sum_{j=1}^{50} \beta_j \phi_j(t)$, $X_i(t) = \sum_{j=1}^{50} \xi_{ij} b_j \phi_j(t)$, $\beta_1 = 0.5$, $\beta_j = 4j^{-2}$, for $j > 1$, $b_j = 1/j$, $\phi_1(t) = 1$, $\phi_{j+1}(t) = 2^{1/2} \cos(j\pi t)$, $\xi_{ij} \sim \text{Uniform}(-\sqrt{3}, \sqrt{3})$, for $j \geq 1$, and ε_i are i.i.d. random variables. Hence, the τ th conditional quantile of Y given $X = X_i$ is

$$Q_Y(\tau|X_i) = \alpha(\tau) + \int_0^1 X_i(t)\beta_0(t)dt,$$

where $\alpha(\tau)$ is the τ th quantile of ε_i .

We consider three sample sizes $n = 200, 500$, and 1000 . For each sample size, we replicate the simulation 500 times. We sample ε_i from either a $t(1)$ distribution ($\gamma = 1$) or the Pareto distribution with $\gamma = 0.25$ or 0.5 . For each simulated data set, we apply the proposed method to estimate $Q_Y(\tau|x)$ at $\tau = 0.99$ and 0.995 , and compare it with CQR and FKE. Here, the semi-metric we use is the L_2 distance between functions, and the bandwidth, kernel functions, and other tuning parameters are selected as described in the simulation studies of Gardes and Girard (2012). For our method, we set $\tau_0 = 0.85$, $\tau^0 = 0.95$, and get the estimator $(\hat{\beta}_1, \dots, \hat{\beta}_{m_n})^\top$ of $(\beta_1, \dots, \beta_{m_n})^\top$ by minimizing the combined quantile objective function (2.1) across quantiles $\tau_1, \tau_2, \dots, \tau_{10}$ that are equally spaced between $[\tau_0, \tau^0]$. We evaluate the performance of an estimator by its MSE. For an estimator $\hat{Q}_Y(\tau|x)$ of $Q_Y(\tau|x)$, its MSE is defined as

$$\frac{1}{500} \sum_{j=1}^{500} \left(\hat{Q}_Y^{(j)}(\tau|x) - Q_Y(\tau|x) \right)^2,$$

where $\hat{Q}_Y^{(j)}(\tau|x)$ is the estimator of $Q_Y(\tau|x)$ obtained from the j th data set. Tables 1–3 report the MSEs of different estimators of $Q_Y(\tau|x)$ at $x(t) = 0$ or $x(t) = \sum_{j=1}^{50} 1.5b_j\phi_j(t)$.

5. Real Data Analysis

5.1. Diffusion tensor imaging data

We apply the proposed method to a data set from a study on cognitive disorders using diffusion tensor images. The study was conducted on 100 multiple sclerosis patients at the Johns Hopkins Hospital with multiple clinical visits (Goldsmith et al. (2011, 2012)). This cerebral data set is available in the R package “refund.” To quantify the cognitive disorder, each patient received a paced auditory serial addition test (PASAT) at every visit, which is the most commonly used examination of cognitive function affected by multiple sclerosis, with scores ranging from 0 to 60. The scalar response of interest is the PASAT score, and the functional covariates are the mean diffusivity profile of the corpus callosum tract (CCA) and the parallel diffusivity profile of the right corticospinal tract (RCST). There are 93 and 55 locations along the CCA and RCST, respectively. We refer to Greven et al. (2010) and Kong, Staicu and Maity (2016) for more detailed

Table 1. MSEs of different estimators of $Q_Y(\tau|x)$, with errors from the Pareto distribution with $\gamma = 0.25$.

n		$x(t) = 0$		$x(t) = \sum_{j=1}^{50} 1.5b_j\phi_j(t)$	
		$\tau = 0.99$	$\tau = 0.995$	$\tau = 0.99$	$\tau = 0.995$
200	CQR	14.45	43.55	84.52	139.58
	FKE	10.71	36.16	28.33	61.27
	EFQR	9.15	31.64	10.59	31.73
500	CQR	4.71	14.21	39.82	71.78
	FKE	3.41	13.52	13.37	24.86
	EFQR	3.04	9.17	3.43	10.35
1000	CQR	2.39	6.30	23.40	55.54
	FKE	2.03	5.24	6.79	13.93
	EFQR	1.58	4.99	1.68	5.32

Table 2. MSEs of different estimators of $Q_Y(\tau|x)$, with errors from the Pareto distribution with $\gamma = 0.5$.

n		$x(t) = 0$		$x(t) = \sum_{j=1}^{50} 1.5b_j\phi_j(t)$	
		$\tau = 0.99$	$\tau = 0.995$	$\tau = 0.99$	$\tau = 0.995$
200	CQR	19.58	98.61	161.64	368.80
	FKE	15.22	50.43	50.53	120.27
	EFQR	10.14	50.24	12.06	47.99
500	CQR	6.16	22.31	62.03	150.49
	FKE	6.11	20.29	15.79	43.98
	EFQR	3.42	14.03	6.71	33.79
1000	CQR	2.95	10.29	34.90	96.67
	FKE	2.26	8.18	9.84	25.11
	EFQR	1.91	7.55	3.49	13.71

descriptions of this data set. For illustration, Figure 1 shows 100 trajectories for the two functional predictors CCA and RCST. After deleting observations with missing data, we use the remaining 229 observations for our statistical analysis.

In our analysis, we take the response variable Y as the centered $-\log(\text{PASAT score}+1)$, and let the two functional covariates $X_1(\cdot)$ and $X_2(\cdot)$ be centered CCA and RCST, respectively. Then, we consider the following functional linear model

$$Y = \int_{\mathcal{I}_1} X_1(t)\beta_1(t)dt + \int_{\mathcal{I}_2} X_2(t)\beta_2(t)dt + \varepsilon, \tag{5.1}$$

where $\mathcal{I}_1 = [0, 93]$ and $\mathcal{I}_2 = [0, 55]$. We first estimate model (5.1) using the classic functional least squares method, which assumes the error ε follows a normal

Table 3. MSEs of different estimators of $Q_Y(\tau|x)$, with errors from $t(1)$ distribution.

n		$x(t) = 0$		$x(t) = \sum_{j=1}^{50} 1.5b_j\phi_j(t)$	
		$\tau = 0.99$	$\tau = 0.995$	$\tau = 0.99$	$\tau = 0.995$
200	CQR	3,703.41	49,629.05	5,693.75	45,504.57
	FKE	433.95	5,787.58	878.25	4,939.58
	EFQR	569.96	4,670.20	482.37	4,325.60
500	CQR	412.92	2,099.68	1,563.31	4,148.55
	FKE	193.37	1,339.83	332.02	1,631.768
	EFQR	175.55	1,321.12	183.28	1,584.86
1000	CQR	173.33	1,990.58	669.89	2,824.52
	FKE	112.82	741.48	152.94	896.48
	EFQR	94.15	686.07	70.97	719.05

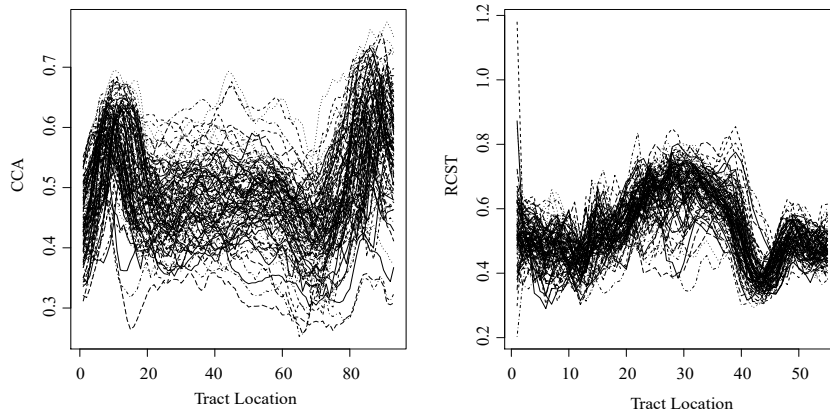


Figure 1. 100 trajectories for the two functional predictors CCA and RCST.

distribution. Figure 2 displays a histogram and a QQ plot of the resulting standardized residuals, which are heavily right-skewed. The QQ plot also indicates that the assumption of a normal distribution for this data set is seriously violated. Furthermore, we perform a Shapiro–Wilk test of normality for the resulting standardized residuals and find that the p -value is less than 2.2×10^{-16} .

These observations motivate us to apply our functional quantile approach as a robust alternative. The τ th conditional quantile of Y given X_1 and X_2 is

$$Q_Y(\tau|X_1, X_2) = \alpha(\tau) + \int_{\mathcal{I}_1} X_1(t)\beta_1(t)dt + \int_{\mathcal{I}_2} X_2(t)\beta_2(t)dt,$$

where $\alpha(\tau)$ is the τ th quantile of ε .

By setting the quantile $\tau = 0.5$ and using the BIC proposed in Subsection

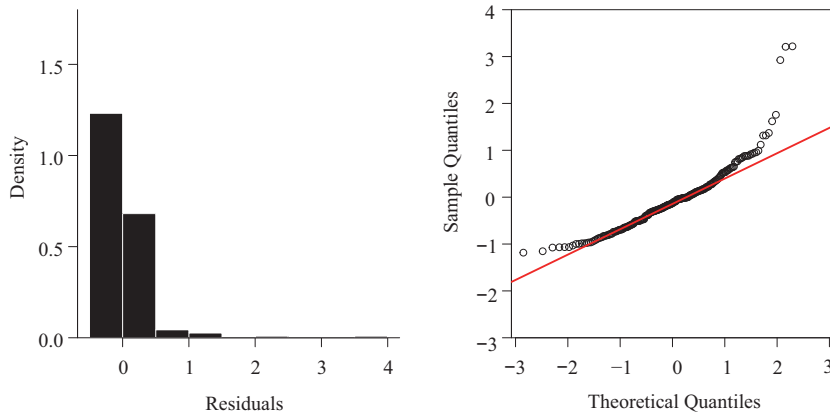


Figure 2. Histogram (left) and normal QQ plot (right) of the standardized residuals from the diffusion tensor image data using model (5.1).

4.1 to select the cut-off levels, we obtain the estimators of $\beta_1(\cdot)$ and $\beta_2(\cdot)$, with $m_{n1} = m_{n2} = 6$. We set $k_n = \lceil 4n^{1/4} \rceil$ in estimation of γ , the same as in the simulation studies. To assess the predictive performance of our proposed estimator, and to compare it with the conventional functional linear quantile regression estimator, we randomly select 129 observations as a training data set I_1 , and the remaining 100 observations as a testing data set I_2 . If $Q_Y(\tau|X_1, X_2)$ is the true τ th conditional quantile of the response Y , we have $\mathbb{E}[\mathbf{1}\{Y > Q_Y(\tau|X_1, X_2)\}] = \mathbb{P}\{Y > Q_Y(\tau|X_1, X_2)\} = 1 - \tau$. Therefore, the standardized exceedance proportion (SEP) as an assessment criterion for the predictive performance,

$$\text{SEP} \left(\widehat{Q}_Y(\tau|X_1, X_2) \right) = \frac{\sum_{i \in I_2} \mathbf{1}\{Y_i - \widehat{Q}_{Y_i}(\tau|X_{i1}, X_{i2}) > 0\} / |I_2| - (1 - \tau)}{\sqrt{\tau(1 - \tau) / |I_2|}},$$

should be small, if the estimation method gives reasonable estimates, where $|I_2|$ is the cardinality of the set I_2 . We show the average of the absolute value of SEP across 500 repetitions for our proposed estimation method and the conventional functional linear quantile regression estimation method at different quantile levels in Table 4.

The results in Table 4 demonstrate that our proposed estimator is superior to the conventional functional quantile estimator in terms of the SEP criterion. Moreover, we can see that CQR tends to underestimate the high quantiles, especially at the extreme tails. In contrast, the proportions of exceeding EFQR are closer to $(1 - \tau)$ at the high quantiles that we considered. This suggests that our

Table 4. The average of the absolute value of SEP of different estimation methods, where CQR is the conventional functional linear quantile regression estimator and EFQR is our proposed estimator.

τ	0.95	0.97	0.99	0.995	0.999
CQR	2.29	3.46	7.67	10.96	25.11
EFQR	0.41	0.14	0.55	0.45	0.16

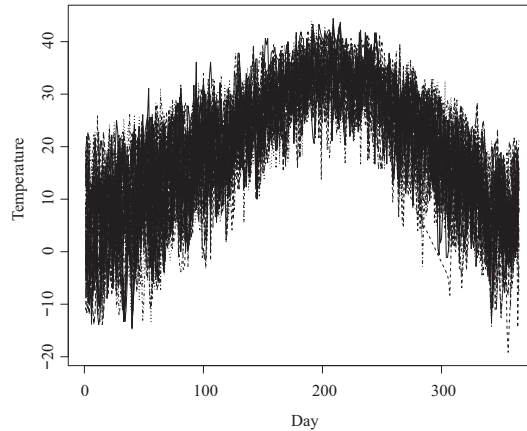


Figure 3. 50 randomly selected trajectories for daily maximum temperature.

proposed estimation method is more suitable for estimating extreme quantiles.

5.2. The Kansas precipitation data

Next, we apply the proposed method to a Kansas precipitation data set to study the relationship between temperature trajectories and extreme precipitation. As climate continues to change around the world, there is growing concern that increasing temperatures will cause extreme precipitation events, including flood or drought. Extreme precipitation has a significant impact on agriculture, wild fire prevention, the economy, public safety, and so on.

Our data are obtained from the National Climatic Data Center (<https://www.ncdc.noaa.gov/data-access>), where we have annual precipitation and daily maximum temperatures for each of the 104 counties in Kansas from 1990 to 2011. Let the response Y be the centered annual precipitation for a specific year and county; the functional predictor $X(t)$ is the centered daily maximum temperature trajectory with the time domain $\mathcal{I} = [0, 365]$. To illustrate the functional predictor, we show in Figure 3 50 randomly selected trajectories for daily maximum temperature.

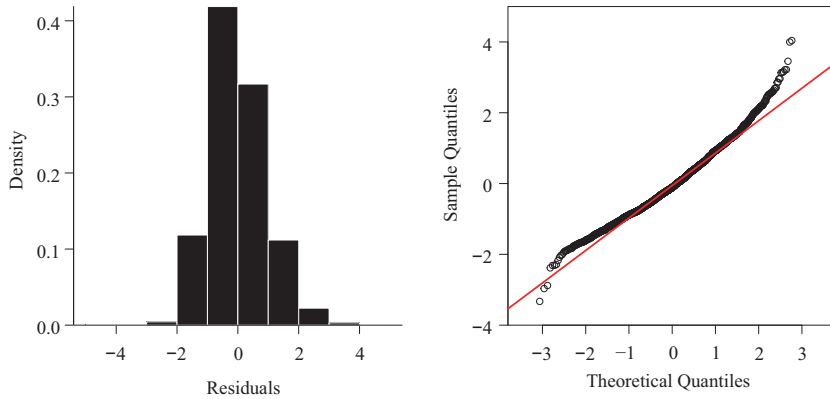


Figure 4. Histogram (left) and normal QQ plot (right) of the standardized residuals from the Kansas precipitation data.

Table 5. The average of the absolute value of the SEP of different estimation methods.

τ	0.95	0.97	0.99	0.995	0.999
CQR	0.35	0.29	0.77	1.14	2.94
EFQR	0.33	0.05	0.38	0.26	1.10

We first use this precipitation data set to fit a functional linear model using a functional least squares regression. Figure 4 displays the histogram and the QQ plot of the resulting standardized residuals. As one can see, the resulting standardized residuals are right-skewed and do not follow a normal distribution.

Next, we apply our functional quantile approach to this data set and assume the τ th conditional quantile of Y given X is

$$Q_Y(\tau|X) = \alpha(\tau) + \int_{\mathcal{I}} X(t)\beta(t)dt.$$

Similarly to the presentation in Subsection 5.1, we set $k_n = \lceil 4n^{1/4} \rceil$ and use the BIC proposed in Subsection 4.1 to select the cut-off level and obtain $m_n = 4$. We conduct a cross-validation study to evaluate the predictive performance of our proposed estimator and the conventional functional quantile estimator. We randomly select 1,288 samples as the training data set I_1 , and the remaining 1,000 samples as the validation data set I_2 . The cross-validation is repeated 200 times, and the average of the absolute value of the SEP across the 200 repetitions is reported in Table 5.

Similarly to the findings in Table 4, the results in Table 5 show that our

proposed method offers a more accurate estimation of high conditional quantiles than that of the conventional functional linear quantile regression estimation method.

6. Conclusion

We propose a new estimation method for extreme conditional quantiles of functional quantile regression with heavy-tailed distributions by first estimating the intermediate conditional quantiles, and then extrapolating the intermediate conditional quantile estimate to extreme tails based on some regularity conditions on tail behaviors. The asymptotic properties of the proposed estimators are established using FPCA and EVT. Our simulation studies suggest that the proposed estimator of high conditional quantiles is much more efficient than the conventional functional linear quantile regression estimator and the functional kernel estimator.

We have supposed that the quantile slope functions are common at the upper conditional quantiles of the response. It would be interesting to consider the case where these functions differ across these quantiles. This is left to future research.

Supplementary Material

The online Supplementary Material includes proofs of Theorems 1–4 and some lemmas required to prove these theorems.

Acknowledgments

Zhang's research was partially supported by the National Natural Science Foundation of China (11971171), 111 Project (B14019) and Project of National Social Science Fund of China (15BTJ027), and Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, ECNU. Li's research was partially supported by the US National Institutes of Health, grant 5R21AG058198.

References

- Bader, B., Yan, J. and Zhang, X. (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *The Annals of Applied Statistics* **12**, 310–329.
- Beirlant, J., Vynckier, P. and Teugels, J. L. (1996). Tail index estimation, pareto quantile plots regression diagnostics. *Journal of the American statistical Association* **91**, 1659–1667.
- Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics* **34**, 2159–2179.

- Cardot, H., Crambes, C. and Sarda, P. (2005). Quantile regression when the covariates are functions. *Journal of Nonparametric Statistics* **17**, 841–856.
- Chen, K. and Müller, H.-G. (2012). Conditional quantile analysis when covariates are functions, with application to growth data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74**, 67–89.
- Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics* **33**, 806–839.
- Chernozhukov, V. and Fernández-Val, I. (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *The Review of Economic Studies* **78**, 559–589.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer Science & Business Media, New York.
- de Sousa, B. and Michailidis, G. (2004). A diagnostic plot for estimating the tail index of a distribution. *Journal of Computational and Graphical Statistics* **13**, 974–995.
- Drees, H., de Haan, L. and Resnick, S. (2000). How to make a hill plot. *The Annals of Statistics* **28**, 254–274.
- Dupuis, D., Sun, Y. and Wang, H. J. (2015). Detecting change-points in extremes. *Statistics and Its Interface* **8**, 19–31.
- Ferraty, F. (2011). *Recent advances in functional data analysis and related topics*. Springer Science & Business Media, New York.
- Ferraty, F., Rabhi, A. and Vieu, P. (2005). Conditional quantiles for dependent functional data with application to the climatic “el niño” phenomenon. *Sankhyā: The Indian Journal of Statistics* **67**, 378–398.
- Friederichs, P. and Hense, A. (2007). Statistical downscaling of extreme precipitation events using censored quantile regression. *Monthly weather review* **135**, 2365–2378.
- Gardes, L. (2018). Tail dimension reduction for extreme quantile estimation. *Extremes* **21**, 57–95.
- Gardes, L. and Girard, S. (2010). Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. *Extremes* **13**, 177–204.
- Gardes, L. and Girard, S. (2012). Functional kernel estimators of large conditional quantiles. *Electronic Journal of Statistics* **6**, 1715–1744.
- Gardes, L., Guillou, A. and Schorgen, A. (2012). Estimating the conditional tail index by integrating a kernel conditional quantile estimator. *Journal of Statistical Planning and Inference* **142**, 1586–1598.
- Goldsmith, J., Crainiceanu, C. M., Caffo, B. and Reich, D. (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **61**, 453–469.
- Goldsmith, J., Crainiceanu, C. M., Caffo, B. S. and Reich, D. S. (2011). Penalized functional regression analysis of white-matter tract profiles in multiple sclerosis. *Neuroimage* **57**, 431–439.
- Gomes, M. I., Figueiredo, F. and Neves, M. M. (2012). Adaptive estimation of heavy right tails: Resampling-based methods in action. *Extremes* **15**, 463–489.
- Gomes, M. I. and Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: A review. *International Statistical Review* **83**, 263–292.
- Greven, S., Crainiceanu, C., Caffo, B. and Reich, D. (2010). Longitudinal functional principal component analysis. *Electronic Journal of Statistics* **4**, 1022–1054.

- Guillou, A. and Hall, P. (2001). A diagnostic for selecting the threshold in extreme value analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**, 293–305.
- Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of multivariate analysis* **32**, 177–203.
- Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics* **35**, 70–91.
- He, F., Cheng, Y. and Tong, T. (2016). Estimation of extreme conditional quantiles through an extrapolation of intermediate regression quantiles. *Statistics & Probability Letters* **113**, 30–37.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**, 1163–1174.
- Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons, Chichester.
- Hunter, D. R. and Lange, K. (2000). Quantile regression via an MM algorithm. *Journal of Computational and Graphical Statistics* **9**, 60–77.
- Kai, B., Li, R. and Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *The Annals of Statistics* **39**, 305–332.
- Kato, K. (2012). Estimation in functional linear quantile regression. *The Annals of Statistics* **40**, 3108–3136.
- Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis* **91**, 74–89.
- Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, Cambridge.
- Koenker, R. (2017). Quantile regression: 40 years on. *Annual Review of Economics* **9**, 155–176.
- Koenker, R. and Bassett, J. (1978). Regression quantiles. *Econometrica* **46**, 33–50.
- Koenker, R. and Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica* **70**, 1583–1612.
- Kong, D., Staicu, A.-M. and Maity, A. (2016). Classical testing in functional linear models. *Journal of Nonparametric Statistics* **28**, 813–838.
- Li, D. and Wang, H. J. (2019). Extreme quantile estimation for autoregressive models. *Journal of Business & Economic Statistics* **37**, 661–670.
- Li, Y., Wang, N. and Carroll, R. J. (2013). Selecting the number of principal components in functional data. *Journal of the American Statistical Association* **108**, 1284–1294.
- Ma, S. and He, X. (2016). Inference for single-index quantile regression models with profile optimization. *The Annals of Statistics* **44**, 1234–1268.
- Marimoutou, V., Raggad, B. and Trabelsi, A. (2009). Extreme value theory and value at risk: Application to oil market. *Energy Economics* **31**, 519–530.
- Matthys, G. and Beirlant, J. (2003). Estimating the extreme value index and high quantiles with exponential regression models. *Statistica Sinica* **13**, 853–880.
- Miranda, M. F., Zhu, H. and Ibrahim, J. G. (2018). Tprpm: Tensor partition regression models with applications in imaging biomarker detection. *The Annals of Applied Statistics* **12**, 1422–1450.
- Odening, M. and Hinrichs, J. (2003). Using extreme value theory to estimate value-at-risk. *Agricultural Finance Review* **63**, 55–73.

- Ramsay and Silverman (2005). *Functional Data Analysis*. Springer-Verlag, New York.
- Schaumburg, J. (2012). Predicting extreme value at risk: Nonparametric quantile regression with refinements from extreme value theory. *Computational Statistics & Data Analysis* **56**, 4081–4096.
- Wang, H. J., Li, D. and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association* **107**, 1453–1464.
- Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association* **73**, 812–815.
- Xiao, L., Zipunnikov, V., Ruppert, D. and Crainiceanu, C. (2016). Fast covariance estimation for high-dimensional functional data. *Statistics and Computing* **26**, 409–421.
- Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.
- Yao, F., Sue-Chee, S. and Wang, F. (2017). Regularized partially functional quantile regression. *Journal of Multivariate Analysis* **156**, 39–56.
- Yoshida, T. (2019). Nonparametric smoothing for extremal quantile regression with heavy tailed distributions. *arXiv preprint arXiv:1903.03242*.
- Yu, D., Kong, L. and Mizera, I. (2016). Partial functional linear quantile regression for neuroimaging data analysis. *Neurocomputing* **195**, 74–87.
- Zhang, Y. (2018). Extremal quantile treatment effects. *The Annals of Statistics* **46**, 3707–3740.
- Zhu, H., Fan, J. and Kong, L. (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association* **109**, 1084–1098.
- Zhu, H., Zhang, R., Yu, Z., Lian, H. and Liu, Y. (2019). Estimation and testing for partially functional linear errors-in-variables models. *Journal of Multivariate Analysis* **170**, 296–314.
- Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics* **36**, 1108–1126.

Hanbing Zhu

School of Statistics, Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, East China Normal University, Shanghai 200062, China.

E-mail: zhuhbecnu@163.com

Riquan Zhang

School of Statistics, Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, East China Normal University, Shanghai 200062, China.

E-mail: zhangriquan@163.com

Yehua Li

Department of Statistics, University of California, Riverside, CA 92521, USA.

E-mail: yehuali@ucr.edu

Weixin Yao

Department of Statistics, University of California, Riverside, CA 92521, USA.

E-mail: weixin.yao@ucr.edu

(Received May 2020; accepted February 2021)