Statistica Sinica **34** (2024), 867-888
doi:https://doi.org/10.5705/ss.202021.0248

SIMULTANEOUS FUNCTIONAL QUANTILE REGRESSION

Boyi Hu[1], Xixi Hu[2], Hua Liu[3], Jinhong You[4] and Jiguo Cao*[1]

[1]*Simon Fraser University,* [2]*University of British Columbia,*
[3]*Xi'an Jiaotong University and* [4]*Shanghai University of Finance and Economics*

Abstract: The conventional method for functional quantile regression (FQR) is to fit the regression model for each quantile of interest separately. Therefore, the slope function of the regression, as a bivariate function of time and quantile, is estimated as a univariate function of time for each fixed quantile. However, there are several limitations to this conventional strategy. For example, it cannot guarantee the monotonicity of the conditional quantiles, nor can it control the smoothness of the slope estimator as a bivariate function. In this paper, we propose a new framework for FQR, in which we simultaneously fit the FQR model for multiple quantiles, with the help of a bivariate basis under some constraints, such that the estimated quantiles satisfy the monotonicity conditions and the smoothness of the slope estimator is controlled. The proposed estimator for the slope function is shown to be asymptotically consistent, and we establish its asymptotic normality. We use simulation to evaluate the finite-sample performance of the proposed method and compare it with that of the conventional method. We demonstrate the proposed method by analyzing the effects of daily temperature on bike rentals, and by investigating the relationship between children's growth history and their adult height.

Key words and phrases: Bivariate spline basis, functional data analysis, non-crossing quantiles.

1. Introduction

The uth quantile of a scalar response Y conditioning on a functional covariate $X(t)$, $Q_Y(u \mid X)$ can be modeled as

$$Q_Y(u \mid X) = c(u) + \int_{\mathscr{T}} X(t)\beta(t, u)dt, \tag{1.1}$$

where $X(t)$ is a stochastic process defined on a compact interval $\mathscr{T}$, and $\beta(t, u)$ is a bivariate slope function indexed by both time t and quantile u. Model (1.1) is called the functional quantile regression (FQR) model. The slope function $\beta(t, u)$ is of primary interest, because it describes how the quantile of the response variable is related to the functional covariate.

In the literature, a common strategy for estimating $\beta(t, u)$ is to treat it as a univariate function of t by first fixing the quantile u. However, this strategy has

*Corresponding author.

two major limitations. First, the slope function $\beta(t,u)$ is usually assumed to be smooth over both t and u. However, fitting the regression models for different quantiles separately cannot guarantee that the resulting estimator for $\beta(t,u)$ is smooth over u. Second, for some observations, the estimation of $Q_Y(u \mid X)$ may not be monotonically increasing in u, as it should be. These crossing quantiles can further lead to an invalid distribution estimation for the response variable.

In this paper, we address the above two limitations. In contrast to existing methods that estimate $\beta(t,u)$ as a univariate function of t for each fixed u, we propose using bivariate spline basis functions to approximate $\beta(t,u)$ directly, and then estimating the corresponding basis coefficients. Under our framework, the smoothness of the estimation is guaranteed by the smoothness of the bivariate spline approximation, which is ensured by adding some linear constraints on the spline coefficients. In addition, we impose extra linear constraints to mitigate the crossing-quantile problem. In this way, we ensure that the estimated quantiles for each subject are monotone. To some extent, the monotonicity problem can be addressed using monotonization techniques, as in Chernozhukov, Fernández-Val and Galichon (2009). However, this does not improve the estimation for $\beta(t,u)$, because the monotonicity of the quantiles is not considered in the estimation procedure for $\beta(t,u)$, and the monotonization is applied only to the estimated quantiles. For example, Kato (2012) proposed first estimating $\beta(t,u)$ for model (1.1), and then estimating conditional quantile functions based on the estimated $\beta(t,u)$. He adjusted any nonmonotone quantile functions to become monotone using the technique of Chernozhukov, Fernández-Val and Galichon (2009). However, the estimation for $\beta(t,u)$ was left unchanged.

The model we consider is an extension of the linear quantile regression (LQR) model, which describes the linear relationship between conditional quantiles of a scalar response and some predictor variables (Koenker and Bassett (1978)). By estimating multiple conditional quantiles, an LQR enables us to depict and then make inferences on the entire distribution of the response, conditioning on the predictors. LQR is well studied, and is used in many real-world applications (Koenker and Geling (2001); Wu, Ma and Yin (2015)).

Functional variables are becoming increasingly common in real-world applications. Functional data analysis is a comprehensive branch of statistics that provides a useful and convenient framework in which to analyze functional data with some high-dimensional structures, such as curves, images, and surfaces, which are so-called functional data. Numerous works estimate a quantile regression (QR) with a scalar response and some functional covariates, such as Cardot, Crambes and Sarda (2007), Chen and Müller (2012), Yu, Kong and Mizera (2016), Wang et al. (2019), and Zhang et al. (2021).

Model (1.1) was first formulated in Cardot, Crambes and Sarda (2005) as a natural extension of the classic LQR. The authors proposed a penalized spline estimator for $\beta(t,u)$ for a fixed u, without any dimension reduction on the functional covariate. Later, for the same model (1.1), Kato (2012) proposed first using a functional principal component analysis (FPCA) to truncate the functional covariate $X(t)$ for dimension reduction, and then estimating the slope function $\beta(t, u)$ for a fixed u using the conventional LQR framework. Kato (2012) also established an optimal convergence rate for the proposed estimator in the minimax sense.

The remainder of the paper is organized as follows. In Section 2, we introduce the model and the corresponding estimator for $\beta(t, u)$. In Section 3, we present the main theoretical results, in which we derive the asymptotic consistency and distribution of the proposed slope function estimator. In Section 4, we demonstrate the proposed estimation method for the slope function using two real-world applications. Section 5 concludes the paper.

2. Proposed Method

2.1. Estimation procedure

Let Y be a scalar random variable, and $X(t)$ be a random function with mean curve $\mu(t)$, where $t \in \mathscr{T}$, and $\mathscr{T} \subset \mathbb{R}$ is a compact set. Let $\mathbf{\Omega} = \mathscr{T} \times \mathscr{A}$, where $\mathscr{A} \subset (0, 1)$ is an interval. For any $u \in \mathscr{A}$, the uth quantile of Y given the functional covariate $X(t)$ is modeled by the following functional quantile model:

$$Q_Y(u \mid X) = c(u) + \int_{\mathscr{T}} X(t)\beta(t, u)dt. \tag{2.1}$$

To estimate the slope function $\beta(t, u)$ in (2.1), we propose first approximating $\beta(t, u)$ using bivariate splines, and then estimating the corresponding coefficients.

Multiple types of bivariate splines can be used for the approximation, such as tensor products of B-splines (Stone et al. (1997); Prautzsch, Boehm and Paluszny (2002); Zhang, Cao and Carroll (2017)) or bivariate Bernstein polynomials over triangulations (Lai and Schumaker (2007)), which is the approach we use to approximate the bivariate slope function in (2.1). Compared with the tensor products of B-splines, the triangulation technique of bivariate Bernstein polynomials enables local refinement; that is, we can flexibly adjust the number of bivariate basis functions with different resolutions in various local areas of the two-dimensional space $\mathscr{T} \times [0, 1]$, which is convenient in many applications. Of course, the Bernstein polynomials and triangulation technique are not required for the proposed method; other bivariate bases should also work.

Figure 1 shows an example of local refinement of a triangulation. The left panel of Figure 1 shows a triangulation over $[0, 1] \times [0, 1]$. The right panel of Figure 1 shows the triangulation after a local refinement by adding a new vertex D inside the triangle $\triangle \mathtt{ABC}$. The triangle $\triangle \mathtt{ABC}$ is further split into three triangles: $\triangle \mathtt{ABD}$, $\triangle \mathtt{BCD}$, and $\triangle \mathtt{ACD}$.

Suppose $\mathscr{A}$ is the interval containing multiple quantiles of interest. Our goal

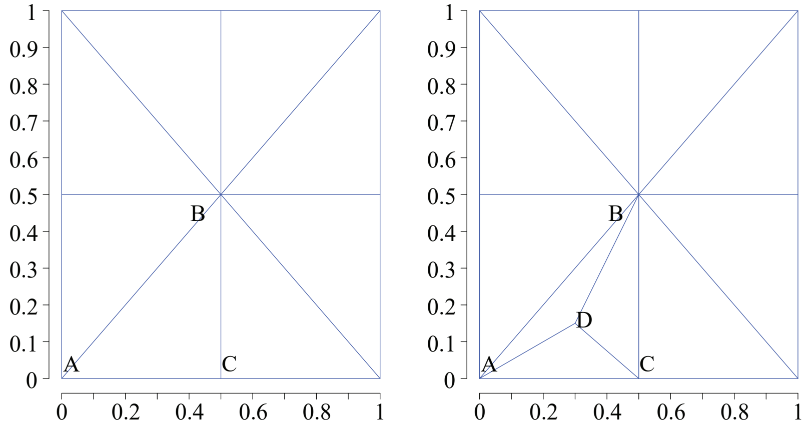


Figure 1. Example of local refinement of triangulation. The left panel shows a triangulation over $[0,1] \times [0,1]$. The right panel shows the triangulation after a local refinement by adding a new vertex D inside the triangle $\triangle \mathtt{ABC}$.

is to find a function $s(t,u) \in S_d^r(\Delta)$ that well approximates the slope function $\beta(t,u)$ on the domain $\mathscr{T} \times \mathscr{A}$. To make our writing and proofs in the subsequent sections clearer, we use $\{b_j(t,u)\}_{j=1}^J$ to denote the Bernstein polynomials defined over the triangulation $\Delta = \{\Lambda_1, \ldots, \Lambda_M\}$, where $j = 1, \ldots, J$ is the index for the polynomials. The relationship between J and M is $J = (d+2)(d+1)M/2$, because there are $(d+2)(d+1)/2$ Bernstein polynomials associated with each triangle of Δ. In addition, for each basis function $b_j(t,u)$, we denote its support by Δ_j, which is a specific triangle of Δ that is the support of $b_j(t,u)$. In other words, $b_j(t,u) \neq 0$ for $(t,u) \in \Delta_j$, and $b_j(t,u) = 0$ for $(t,u) \notin \Delta_j$. If two Bernstein polynomials $b_j(t,u)$ and $b_k(t,u)$ are associated with the same triangle, then Δ_j and Δ_k are identical.

The function $s(t,u) \in S_d^r(\mathscr{T} \times \mathscr{A})$ that approximates $\beta(t,u)$ can be written as a linear combination of Bernstein polynomials $\{b_j(t,u)\}_{j=1}^J$. Then, on the domain $\mathscr{T} \times \mathscr{A}$, we have the approximation

$$\beta(t,u) \approx s(t,u) = \sum_{j=1}^{J} \gamma_j b_j(t,u) \in S_d^r(\Delta), \tag{2.2}$$

where $\{\gamma_j\}_{j=1}^J$ are the corresponding coefficients.

Under some conventional assumptions on $X(t)$ (Yao, Müller and Wang (2005); Sang, Wang and Cao (2017); Nie et al. (2018); Nie and Cao (2020); Shi et al. (2021)) usually satisfied in real applications, by Mercer's theorem, $X(t)$ admits the decomposition

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t), \tag{2.3}$$

where $\phi_k(t)$, for $k = 1, \ldots$, are called functional principal components (FPCs) and ξ_k are called FPC scores. By the decomposition (2.3) and the approximation (2.2), model (2.1) can be approximately re-expressed as

$$\begin{aligned} Q_Y(u \mid X) &\approx c(u) + \int_{\mathscr{T}} \mu(t)\beta(t,u)dt + \int_{\mathscr{T}} \sum_{k=1}^{\infty} \xi_k \phi_k(t) s(t,u) dt, \\ &= c_0(u) + \int_{\mathscr{T}} \sum_{k=1}^{\infty} \xi_k \phi_k(t) s(t,u) dt, \end{aligned}$$

where $c_0(u) = c(u) + \int_{\mathscr{T}} \mu(t)\beta(t,u)dt$. Let $\{b_{0,j}(u)\}_{j=1}^{J_0}$ denote the univariate B-spline basis functions defined over the interval $\mathscr{A}$. Then, we further approximate $c_0(u)$ by $c_0(u) \approx \sum_{j=1}^{J_0} \gamma_{0,j} b_{0,j}(u) = \boldsymbol{b}_0^{\mathrm{T}}(u)\boldsymbol{\gamma}_0$, where $\boldsymbol{b}_0^{\mathrm{T}}(u) = (b_{0,1}(u), \ldots, b_{0,J_0}(u))$ and $\boldsymbol{\gamma}_0^{\mathrm{T}} = (\gamma_{0,1}(u), \ldots, \gamma_{0,J_0}(u))$.

In a functional data context, functional observations as infinite-dimensional subjects do not fit in the conventional LQR framework. In addition, the observed functional data are not always sufficiently smooth to use numerical integration to approximate the integral in (2.1). To address these problems and to extend the classic LQR to an FQR, we usually need to truncate the functional observations $\{x_i(t)\}_{i=1}^n$ to reduce the dimensionality and to smooth them. A plausible approach for the dimensionality reduction is to truncate $X(t)$ by using its first m FPCs obtained from the decomposition (2.3).

As mentioned in the introduction, in contrast to conventional methods, we estimate $\beta(t,u)$ as a bivariate function directly. Therefore, all quantiles of interest are considered simultaneously in the estimation procedure. Numerous papers have discussed the advantage of combining multiple QR models, such as the works of Zou and Yuan (2008), Kai, Li and Zou (2011), Zhao and Xiao (2014), and He et al. (2016). A common approach is to consider the sum of these models.

We know that for a real-valued random variable Y, the minimizer of $E\{\rho_u(Y-u)\}$ is the u-quantile of Y, where $\rho_u(x) = x\,(u - \mathbb{1}\{x<0\})$ is called the check function (Koenker and Bassett (1978)). Assume that we observe independent and identically distributed (i.i.d.) data pairs $\{y_i, x_i(t)\}_{i=1}^n$ as realizations of $\{Y, X(t)\}$. We use $A \in \mathscr{A}$ to denote a set of quantiles of interest, which are assumed to be uniformly distributed in $\mathscr{A}$, and use n_A to denote the cardinality of A. We first apply an FPCA on $\{y_i, x_i(t)\}_{i=1}^n$ to obtain the estimated FPCs $\{\hat{\phi}_k(t)\}$ and FPC scores $\{\hat{\xi}_{ik}\}_{k=1}^m$. Then, based on the approximation (2.2), a reasonable estimator for $\beta(t,u)$ should minimize the following loss function:

$$\frac{1}{nn_A} \sum_{r=1}^{n_A} \sum_{i=1}^{n} \rho_{u_r} \left(y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\boldsymbol{\gamma}_0 - \int_{\mathscr{T}} \sum_{k=1}^{m} \hat{\xi}_{ik} \hat{\phi}_k(t) s(t,u_r) dt \right), \tag{2.4}$$

with respect to $s(t,u) \in S_d^r(\Delta)$ and $\boldsymbol{\gamma}_0$.

The conventional LQR framework is designed for finite-dimensional subjects, and estimates a finite-dimensional slope parameter. Although functional observations can be truncated by using an FPCA into a finite dimension, the slope function $\beta(t,u)$ in the model (2.1) is still infinite-dimensional. As a result, a direct extension (2.4) of the conventional LQR framework to functional data can lead to an invalid estimation for $\beta(t,u)$, and the uniqueness of the minimizer of (2.4) cannot be guaranteed.

To clarify this, let $\{\hat{s}(t,u), \hat{\boldsymbol{\gamma}}_0\}$ be a minimizer of (2.4) and fix the truncation level at m. Assume there exists another function $s_1(t,u) \in S_d^r(\Delta)$, such that $s_1(t,u)$ is orthogonal to the first m estimated FPCs of $X(t)$. Then, $\{\hat{s}(t,u) + s_1(t,u), \hat{\boldsymbol{\gamma}}_0\}$ is another minimizer of (2.4). Specifically, using an FPCA, we obtain the $(m+1)$th FPC, denoted as $\hat{\phi}_{m+1}(t)$, which is orthogonal to the first m estimated FPCs, $\hat{\phi}_1(t), \ldots, \hat{\phi}_m(t)$. If there exists some measurable function $w(u)$ such that $\hat{s}(t,u) + w(u)\hat{\phi}_{m+1}(t)$ also belongs to the space $S_d^r(\Delta)$, then $\{\hat{s}(t,u) + w(u)\hat{\phi}_{m+1}(t), \hat{\boldsymbol{\gamma}}_0\}$ is also a minimizer of (2.4). This implies that $\hat{s}(t,u) + w(u)\hat{\phi}_{m+1}(t)$ is another estimator for $\beta(t,u)$. However, the information of $\hat{\phi}_k(t)$ for any $k \geq m+1$ is excluded from our estimation procedure when we choose the truncation level as m, and thus the estimator for $\beta(t,u)$ derived from the estimation procedure does not include any such information. Therefore, the objective function (2.4) derived directly from the conventional LQR is problematic in a functional data context.

To overcome this problem, we propose penalizing the L^2-norm of the approximation $s(t,u)$ during the estimation procedure. In addition, the roughness of the slope function estimator $s(t,u)$ is also a concern in a functional data context, and so we a roughness penalty for $s(t,u)$ during the estimation procedure. Roughness penalties are useful for controlling the smoothness of functions in an estimation procedure; see Ramsay and Silverman (2002), Cardot, Ferraty and Sarda (2003), Ramsay and Silverman (2005), Ramsay, Hooker and Graves (2009), and Cao and Ramsay (2010). We consider the following roughness penalty $R(s;\omega_0,\omega_1,\omega_2)$:

$$R(s;\omega_0,\omega_1,\omega_2) = \sum_{\Lambda\in\Delta} \int_{\Lambda} \sum_{d_1+d_2=2} \omega_{d_1} \binom{2}{d_1} \left[\nabla_t^{d_1} \nabla_u^{d_2} s(t,u)\right]^2 dtdu,$$

where ω_0, ω_1, and ω_2 are tuning parameters representing the weights corresponding to second derivatives in different directions. More specifically, ω_0 is the weight corresponding to $\partial^2 s/\partial t^2$, ω_1 is the weight corresponding to $\partial^2 s/\partial t \partial u$, and ω_2 is the weight corresponding to $\partial^2 s/\partial u^2$. Because we include a tuning parameter $\lambda_{2,n}$ in our estimation procedure for the whole roughness penalty $R(s;\omega_0,\omega_1,\omega_2)$, ω_0 can be fixed as a constant $\omega_0 = 1$. In addition, if the smoothness of the target slope functions along the quantile index and the functional index can be assumed to be similar, then we can simply set $\omega_0 = \omega_1 = \omega_2 = 1$ to reduce the computational cost. Then, $R(s; \omega_0, \omega_1, \omega_2)$ becomes

$$R(s) = \sum_{\Lambda \in \Delta} \int_{\Lambda} \sum_{d_1+d_2=2} \binom{2}{d_1} \left(\nabla_t^{d_1} \nabla_u^{d_2} \boldsymbol{B}^{\mathrm{T}}(t,u)\boldsymbol{\gamma}\right)^2 dtdu,$$

which is the most common roughness penalty discussed in the literature.

Let $\lambda_{1,n}$ and $\lambda_{2,n}$ be nonnegative tuning parameters. Then, we estimate the slope function $\beta(t,u)$ in (2.1) by minimizing

$$\begin{aligned}&\frac{1}{nn_A}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\rho_{u_r}\left(y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\boldsymbol{\gamma}_0 - \int_{\mathscr{T}}\sum_{k=1}^{m}\hat{\xi}_{ik}\hat{\phi}_k(t)s(t,u)dt\right)\\ &+ \lambda_{1,n}\|s\|^2_{L^2(\boldsymbol{\Omega})} + \lambda_{2,n}R(s;\omega_0,\omega_1,\omega_2),\end{aligned} \tag{2.5}$$

with respect to $s(t,u) \in S_d^r(\Delta)$ and $\boldsymbol{\gamma}_0$, where the norm $\|s\|^2_{L^2(\boldsymbol{\Omega})}$ is defined as $\|s\|^2_{L^2(\boldsymbol{\Omega})} = \int_{\mathscr{T}\times\mathscr{A}} s^2(t,u)dtdu$.

For any $s(t,u) \in S_d^r(\Delta)$, we have the expression

$$s(t,u) = \sum_{j=1}^{J} \gamma_j b_j(t,u) = \boldsymbol{B}^{\mathrm{T}}(t,u)\boldsymbol{\gamma}, \tag{2.6}$$

where $\boldsymbol{B}(t,u) = (b_1(t,u), \ldots, b_J(t,u))^{\mathrm{T}}$ and $\boldsymbol{\gamma}$ is the vector of coefficients satisfying some linear constraint,

$$\boldsymbol{H}\boldsymbol{\gamma} = \boldsymbol{0}. \tag{2.7}$$

The constraint (2.7) ensures that $s(t,u) = \boldsymbol{B}^{\mathrm{T}}(t,u)\boldsymbol{\gamma} \in C^r(\mathscr{T}\times\mathscr{A})$. The matrix $\boldsymbol{H}$ depends on the triangulation Δ, degree d, and smoothness parameter r of the spline space $S_d^r(\Delta)$ (Lai and Schumaker (2007)). For example, when $r = 1$, $s(t,u)$ is assumed to have continuous first partial derivatives over both t and u. A useful way of removing the constraint (2.7) is to use a QR decomposition (Wang et al. (2020)). For a given $\boldsymbol{H}$, by the QR decomposition, we have

$$\boldsymbol{H}^{\mathrm{T}} = (\boldsymbol{Q}^*, \boldsymbol{Q})\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{pmatrix}, \tag{2.8}$$

where $(\boldsymbol{Q}^*, \boldsymbol{Q})$ is a matrix with orthogonal columns, and $\boldsymbol{R}$ is an upper triangle matrix with nonzero diagonal elements. With the decomposition (2.8), the constraint $\boldsymbol{H}\boldsymbol{\gamma} = \boldsymbol{0}$ can be removed by rewriting $\boldsymbol{\gamma}$ as

$$\boldsymbol{\gamma} = \boldsymbol{Q}\boldsymbol{\theta}. \tag{2.9}$$

Suppose we observe $X_i(t)$ for $t \in T$, and use n_T to denote the cardinality of T. By (2.6), the penalty $\|s\|^2_{L^2(\boldsymbol{\Omega})}$ can be approximated by $\|s\|^2_{L^2(\boldsymbol{\Omega})} \approx (1/n_A n_T)\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{\gamma}$, where $\boldsymbol{B}_{A,T}$ is a J-by-$n_A n_T$ matrix, with its jth row being the evaluations of the Bernstein polynomials $b_j(t,u)$, for all $t \in T$ and $u \in A$. The roughness penalty $R(s;\omega_0,\omega_1,\omega_2)$ or $R(s)$ can also be written in matrix form as $\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{\gamma}$, where the matrix $\boldsymbol{D}$ is a J-by-J positive-definite and block-diagonal matrix, with each block corresponding to one triangle of the triangulation Δ, and the size of each block depends on the degree d.

Define $L_0(\boldsymbol{\theta},\boldsymbol{\gamma}_0) = (nn_A)^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\rho_{u_r}(y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\boldsymbol{\gamma}_0 - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\boldsymbol{\theta})$ as the whole quantile loss based on an FPCA. Then, by (2.6) and (2.9), the minimization problem (2.5) can be converted to

$$\min_{\boldsymbol{\theta},\boldsymbol{\gamma}_0} \; L_0(\boldsymbol{\theta},\boldsymbol{\gamma}_0) + \lambda_{1,n}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\theta} + \lambda_{2,n}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\boldsymbol{\theta}, \tag{2.10}$$

where $\hat{\boldsymbol{\xi}}_i = (\hat{\xi}_{i1},\ldots,\xi_{im})^{\mathrm{T}}$, and $\hat{\boldsymbol{P}}(u)$ is an $m \times J$ matrix, with the (k,j)-entry being $\hat{p}_{k,j}(u) = \int_{(t,u)\in\Delta_j}\hat{\phi}_k(t)b_j(t,u)dt$. Note that, for the matrix $\hat{\boldsymbol{P}}(u)$ and a specific u, say $u = u_r \in A$, many entries of $\hat{\boldsymbol{P}}(u_r)$ are zeros, because the integral $\int_{(t,u)\in\Delta_j}\hat{\phi}_k(t)b_j(t,u_r)dt$ is equal to zero if the triangle Δ_j, which is the support of $b_j(t,u_r)$, does not intersect with the horizontal line $u = u_r$.

If we denote the minimizer of (2.10) by $(\hat{\boldsymbol{\gamma}}_0,\hat{\boldsymbol{\theta}})$, then our proposed estimator for $\beta(t,u)$ in (2.1) is

$$\hat{\beta}(t,u) = \boldsymbol{B}^{\mathrm{T}}(t,u)\boldsymbol{Q}\hat{\boldsymbol{\theta}}. \tag{2.11}$$

In practice, to guarantee that the estimated conditional quantile functions of all the subjects are monotone, we impose some extra linear constraints on $\boldsymbol{\theta}$. Specifically, given $(\hat{\boldsymbol{\gamma}}_0,\hat{\boldsymbol{\theta}})$, the estimated u-quantile of the ith subject is $\hat{Q}_Y(u \mid X = x_i) = \boldsymbol{b}_0^{\mathrm{T}}(u)\hat{\boldsymbol{\gamma}}_0 + \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u)\boldsymbol{Q}\hat{\boldsymbol{\theta}}$.

The monotonicity of $\hat{Q}_Y(u \mid X = x_i)$ can be approximately expressed as $\hat{Q}_Y(u_r \mid X = x_i) \le \hat{Q}_Y(u'_r \mid X = x_i)$, for any $u_r < u'_r$, $u_r, u'_r \in A$. Then, a reasonable way to mimic the monotonicity of these quantile functions is to impose the following constraints on the optimization:

$$\{\boldsymbol{b}_0^{\mathrm{T}}(u_r) - \boldsymbol{b}_0^{\mathrm{T}}(u'_r)\}\boldsymbol{\gamma}_0 + \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\{\hat{\boldsymbol{P}}(u_r) - \hat{\boldsymbol{P}}(u'_r)\}\boldsymbol{Q}\boldsymbol{\theta} \le \mathbf{0}$$

for any quantile $u_r < u'_r$ and any $i = 1,\ldots,n$, which guarantee that the estimated conditional quantiles of $Y \mid X_i(t)$ do not cross (Bondell, Reich and Wang (2010); Liu and Wu (2011)). Then, we can solve (2.10) under the constraints

$$\{\boldsymbol{b}_0^{\mathrm{T}}(u_r) - \boldsymbol{b}_0^{\mathrm{T}}(u'_r)\}\boldsymbol{\gamma}_0 + \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\{\hat{\boldsymbol{P}}(u_r) - \hat{\boldsymbol{P}}(u'_r)\}\boldsymbol{Q}\boldsymbol{\theta} \le \mathbf{0} \tag{2.12}$$

for all $i = 1,\ldots,n$ and any $u_r < u'_r, u_r, u'_r \in A$.

2.2. Computation

This subsection examines the computational aspect of the minimization problem (2.10). As introduced in Koenker and Bassett (1978), for a specific quantile u, the minimization of the loss function derived from the classic LQR model is equivalent to a constrained linear programming problem.

For the proposed method, we need to solve the minimization problem (2.10). Following Koenker and Bassett (1978), (2.10) can be formulated as the minimization of the following quadratic programming problem with respect to $\boldsymbol{\theta}$, $\boldsymbol{\gamma}_0$, and $\{w_{i,r}, v_{i,r}\}_{i=1,\ldots,n,r=1,\ldots,n_A}$:

$$\frac{1}{nn_A}\sum_{r=1}^{n_A}\left\{u_r\sum_{i=1}^{n} w_{i,r} + (1-u_r)\sum_{i=1}^{n} v_{i,r}\right\} + \lambda_{1,n}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\theta} + \lambda_{2,n}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\boldsymbol{\theta}, \tag{2.13}$$

subject to $y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\boldsymbol{\gamma}_0 - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\boldsymbol{\theta} = w_{i,r} - v_{i,r}$, $w_{i,r} \geq 0$, and $v_{i,r} \geq 0$, for all $i = 1, \ldots, n$ and $r = 1, \ldots, n_A$.

If we further impose the monotonicity constraints (2.12) on (2.13), then the constrained optimization can be formulated similarly as the following problem with respect to $\boldsymbol{\theta}$, $\boldsymbol{\gamma}_0$, and $\{w_{i,r}, v_{i,r}\}_{i=1,\ldots,n,r=1,\ldots,n_A}$:

$$\frac{1}{nn_A}\sum_{r=1}^{n_A}\left\{u_r\sum_{i=1}^{n} w_{i,r} + (1-u_r)\sum_{i=1}^{n} v_{i,r}\right\} + \lambda_{1,n}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\theta} + \lambda_{2,n}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\boldsymbol{\theta},$$

subject to $y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\boldsymbol{\gamma}_0 - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\boldsymbol{\theta} = w_{i,r} - v_{i,r}$, $\{\boldsymbol{b}_0^{\mathrm{T}}(u_r) - \boldsymbol{b}_0^{\mathrm{T}}(u_r')\}\boldsymbol{\gamma}_0 + \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\{\hat{\boldsymbol{P}}(u_r) - \hat{\boldsymbol{P}}(u_r')\}\boldsymbol{Q}\boldsymbol{\theta} \leq \boldsymbol{0}$, $w_{i,r} \geq 0$, and $v_{i,r} \geq 0$, for all $i = 1, \ldots, n$ and $r = 1, \ldots, n_A$, and any $u_r < u_r'$, $u_r, u_r' \in A$.

In summary, the complete algorithm can be split into two parts:

- Derive the coefficients in (2.13), with or without the monotonicity constraints (2.12), such as $\{\hat{\boldsymbol{\xi}}_i\}_{i=1}^n$, $\{\hat{\boldsymbol{P}}(u)\}_{u\in A}$, $\boldsymbol{Q}$, and so on. Specifically, we first derive the estimated FPCs $\{\hat{\phi}_k(t)\}_{k=1}^m$ and corresponding scores $\{\hat{\boldsymbol{\xi}}_i\}_{i=1}^n$. Next, we compute the matrices related to the bivariate spline basis, $\boldsymbol{B}^{\mathrm{T}}(t,u)$, $\boldsymbol{Q}$, $\boldsymbol{B}_{A,T}$, and $\boldsymbol{D}$. Given $\{\hat{\phi}_k(t)\}_{k=1}^m$ and $\boldsymbol{B}^{\mathrm{T}}(t,u)$, $\{\hat{\boldsymbol{P}}(u)\}_{u\in A}$ are approximated using numerical integration based on Simpson's rule.

- We can code and solve the quadratic programming problem (2.13), with or without the constraints (2.12), using Matlab.

2.3. Tuning parameter selection

In our proposed method, to estimate $c(u)$ and $\beta(t,u)$ in model (2.1), we need to first decide the truncation level m and the values of the tuning parameters $\lambda_{1,n}$ and $\lambda_{2,n}$.

To choose the truncation level m, we suggest using the following BIC:

$$BIC(m) = \log\left(n^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\rho_{u_r}\left\{y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\hat{\boldsymbol{\gamma}}_0 - \int_{\mathscr{T}}\sum_{k=1}^{m}\hat{\xi}_{ik}\hat{\phi}_k(t)\hat{\beta}(t,u)dt\right\}\right) + \frac{(m+1)\log n}{n}. \tag{2.14}$$

We recommend using leave-one-out cross-validation to select the penalty parameters $\lambda_{1,n}$ and $\lambda_{2,n}$. However, the computational cost for each fitting is expensive. Therefore, in practice, we usually use five-fold or 10-fold cross-validation to select the parameter values. As we show in the next section, the value of $\lambda_{2,n}$ depends on the value of $\lambda_{1,n}$. Therefore, we propose a sequential procedure for choosing values for $\lambda_{1,n}$ and $\lambda_{2,n}$. The cross-validation procedure is described as follows. We use 10-fold cross-validation as an example.

We first use the complete sample $\{x_i(t)\}_{i=1}^{n}$ to estimate the FPCs $\{\hat{\phi}_k(t)\}_{k=1}^{m}$ and corresponding scores $\{\hat{\boldsymbol{\xi}}_i\}_{i=1,\dots,n}$. Then, for a fixed m, we use 10-fold cross-validation to find the optimal values for the tuning parameters $\lambda_{1,n}$ and $\lambda_{2,n}$. More specifically, we first apply the cross-validation on the following objective function with only one penalty $\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\theta}$:

$$L_{n,1}(\boldsymbol{\theta},\boldsymbol{\gamma}_0) = L_0(\boldsymbol{\theta},\boldsymbol{\gamma}_0) + \lambda_{1,n}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\theta},$$

to decide the optimal value for $\lambda_{1,n}$ among all candidates, and denote it as $\hat{\lambda}_{1,n}$.

Next, based on $\hat{\lambda}_{1,n}$, we apply the cross-validation again on the full objective function with two penalties,

$$L_{n,2}(\boldsymbol{\theta},\boldsymbol{\gamma}_0) = L_0(\boldsymbol{\theta},\boldsymbol{\gamma}_0) + \hat{\lambda}_{1,n}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\theta} + \lambda_{2,n}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\boldsymbol{\theta},$$

to find the optimal value for $\lambda_{2,n}$ among all candidates, and denote it as $\hat{\lambda}_{2,n}$. Then, $(\hat{\lambda}_{1,n},\hat{\lambda}_{2,n})$ are the optimal values for $(\lambda_{1,n},\lambda_{2,n})$ for the current truncation level m. We repeat this sequential selection procedure for multiple values of m, and then choose the optimal value for m based on criterion (2.14).

3. Theoretical Results

To investigate the asymptotic properties of the proposed slope function estimator $\hat{\beta}(t,u)$ defined in (2.11), we assume the following conditions on the distribution of the random function $X(t)$, the conditional distribution of $Y \mid X(t)$, and the slope function $\beta(t,u)$:

(A1) $\{Y_i, X_i(t)\}_{i=1}^n$ are i.i.d.

(A2) $\int_{\mathscr{T}} E\left(X^4(t)\right) dt < \infty$, and $E\left(\xi_k^4\right) < C\kappa_k^2$, for all $k \geq 1$.

(A3) For some $\alpha > 1$ and for any $k \geq 1$, $C^{-1}k^{-\alpha} \leq \kappa_k \leq Ck^{-\alpha}$ and $\kappa_k - \kappa_{k+1} \geq C^{-1}k^{-\alpha-1}$.

(A4) $\partial F_{Y|X}(y \mid X)/\partial y \vee |\partial^2 F_{Y|X}(y \mid X)/\partial y^2| \leq C$, and $\inf_{u\in\mathscr{A}} f_{Y|X}(Q_{Y|X}(u \mid X) \mid X) \geq C^{-1}$.

(A5) $\beta(t,u) \in W_q^{d+1}(\mathscr{T}\times\mathscr{A})$, and for some $\zeta > \alpha/2+1$, $\sup_{u\in\mathscr{A}} |\beta_k(u)| \leq Ck^{-\zeta}$, for $k = 1, \ldots$, where $W_q^{d+1}(\mathscr{T}\times\mathscr{A})$ is a Sobolev space defined over $\mathscr{T}\times\mathscr{A}$, and $\beta_k(u) = \int_{\mathscr{T}} \beta(t,u)\phi_k(t)dt$.

(A6) There exists a finite number p_0 such that $\kappa_k = 0$, for all $k \geq p_0$.

The i.i.d. assumption is conventional, and we do not consider the scenario of dependent data in this paper. A2 provides commonly assumed restrictions on the moments of $X(t)$ and ξ_k. There is no condition on the moment of Y. A3 is adapted from (A3) of Kato (2012), and ensures the identifiability of $\phi_k(t)$ and the estimation accuracy of $\hat{\phi}_k(t)$. A4 provides common conditions on the conditional distribution and density functions of Y in a QR context. A5 determines the estimation accuracy of $\hat{\beta}(t,u)$ by using the truncated functional covariate, and the Sobolev space assumption ensures that bivariate splines can be used to approximate $\beta(t,u)$. A6 implies that the functional covariate $X_i(t)$ can be represented by a finite number of pairs of FPCs and corresponding FPC scores.

For a triangle Λ, let $|\Lambda|$ be the length of its longest edge, and then, for a triangulation Δ, we define $|\Delta| := \max\{|\Lambda| : \Lambda \in \Delta\}$ (i.e., the length of the longest edge of all triangles in the triangulation Δ). Recall that n_A and n_T represent the cardinalities of A and T, respectively, as defined previously. The following theorem gives the rate of convergence of the slope function estimator $\hat{\beta}(t,u)$ for a given truncation level m when an FPCA is used to reduce the dimension of the functional covariate.

For any fixed $u \in (0,1)$, we use $\beta_u(t)$ to denote $\beta(t,u)$ and use $\hat{\beta}_u(t)$ to denote $\hat{\beta}(t,u)$. Define

$$A_1 = \left\{ r \in (1,\ldots,n_A) : \left\| \hat{\beta}_{u_r}(t) - \beta_{u_r}(t) \right\|_{L^2} \geq M\kappa_m^{-1/2} m^{1/2} n^{-1/2}, \right.$$
$$\left. \text{for some constant } M > 0 \right\},$$

where $\|\hat{\beta}_{u_r}(t) - \beta_{u_r}(t)\|_{L^2} = \{\int_{\mathscr{T}} (\hat{\beta}_{u_r}(t) - \beta_{u_r}(t))^2 dt\}^{1/2}$. The set A_1 can be regarded as an index set of quantiles for which the estimations are not good enough.

Theorem 1. *Under conditions* A1–A5*, assume that* $|\Delta| = o(m^{-(1+2\alpha)/(2d+2)} n^{-3/(2d+2)})$ *and* $n_A^{-1}|\Delta|^{-1}m^{(\alpha-1)/3} = o(1)$*. Suppose the tuning parameters* $\lambda_{1,n}$ *and* $\lambda_{2,n}$ *satisfy* $\lambda_{1,n} \asymp n_A^{-1}n_T^{-1}m^{-1/2}n|\Delta|^{d+1}$ *and* $\lambda_{2,n} = o(\lambda_{1,n}n_A^{-1}n_T^{-1}|\Delta|^4)$*. Then,*

$$\left\|\hat{\beta}(t,u) - \beta(t,u)\right\|_{L^2(\boldsymbol{\Omega})} \approx O_p\left(\kappa_m^{-1/2}m^{1/2}n^{-1/2} \vee m^{-(2\zeta+1)/2}\right).$$

In addition, for A_1*, we have* $|A_1| = o_p(m^{-1-\alpha}n^{-1/2}n_A)$*.*

Remark 1. The first term of the stochastic order of $\|\hat{\beta}(t,u) - \beta(t,u)\|_{L^2(\boldsymbol{\Omega})}$ in Theorem 1 decreases as the sample size n increases, and is increasing with the truncation level m (i.e., adding FPCs to the estimation). The second term represents the information loss if we include too few FPCs in the estimation procedure. Then, based on condition A5, we obtain a theoretically optimal truncation level $m \asymp n^{1/(\alpha+2\zeta)}$.

The following theorem presents the asymptotic distribution of the slope estimator $\hat{\beta}(t,u)$. We now assume that p_0 is known and finite, as in Li et al. (2022). Under A6 and by Lemma 1 and Lemma 3 presented in the Supplementary Material, there exist $\boldsymbol{\gamma}_0^*$ and $\boldsymbol{\theta}^*$ such that

$$\sup_{(t,u)\in\mathscr{T}\times\mathscr{A}} |\beta(t,u) - \boldsymbol{B}^{\mathrm{T}}(t,u)\boldsymbol{Q}\boldsymbol{\theta}^*| \le C_1|\Delta|^{d+1} \text{ and } \sup_{u\in\mathscr{A}} |c(u) - \boldsymbol{b}_0^{\mathrm{T}}(u)\boldsymbol{\gamma}_0^*| \le C_2|\Delta|^{d+1},$$

for some constants C_1 and C_2. Let $\boldsymbol{\Gamma}^* = (\boldsymbol{\gamma}_0^*, \boldsymbol{\theta}^*)^{\mathrm{T}}$, $\boldsymbol{Z_i}(u) = [\boldsymbol{b}_0^{\mathrm{T}}(u), \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u)\boldsymbol{Q}]$, $\tilde{\boldsymbol{B}}(t,u) = (\boldsymbol{0}_{1\times n_B}, \boldsymbol{B}^{\mathrm{T}}(t,u)\boldsymbol{Q})^{\mathrm{T}}$, and $\tilde{\boldsymbol{Z}}_i = (\boldsymbol{Z}_i^{\mathrm{T}}(u_1), \ldots, \boldsymbol{Z}_i^{\mathrm{T}}(u_{n_A}))$. Then, define $\boldsymbol{\Sigma}_1 = n_A^{-1}\sum_{r=1}^{n_A} E\left[f_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*)\boldsymbol{Z}_i^{\mathrm{T}}(u_r)\boldsymbol{Z}_i(u_r)\right]$ and

$$\boldsymbol{\Sigma}_2 = \frac{1}{2n}\boldsymbol{\Sigma}_1 + \lambda_{1,n}\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q} \end{bmatrix} + \lambda_{2,n}\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q} \end{bmatrix},$$

where f_i is the conditional density of $Y_i \mid X_i(t)$. Let $\boldsymbol{U}_1$ be an n_A-by-n_A matrix with the (r,r')-entry being $u_r \wedge u_{r'} - u_r u_{r'}$, for any $r, r' = 1, \ldots, n_A$. Define $\boldsymbol{U}_2 = n_A^{-2}E\left[\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\boldsymbol{U}_1\tilde{\boldsymbol{Z}}_i\right]$ and $\boldsymbol{\Sigma} = (2n\boldsymbol{\Sigma}_2)^{-1}\boldsymbol{U}_2/n\,(2n\boldsymbol{\Sigma}_2)^{-1}$.

Theorem 2. *Under the conditions of Theorem* 1, A6*, and* $n_A n|\Delta|^{d+2} = o(1)$*, as* $n \to \infty$ *and* $n_A \to \infty$*, for fixed* (t,u)*, we have*

$$\sigma_\beta^{-1/2}(t,u)\left\{\hat{\beta}(t,u) - \beta(t,u)\right\} \to N(0,1)$$

in distribution, where $\sigma_\beta(t,u) = \tilde{\boldsymbol{B}}^{\mathrm{T}}(t,u)\boldsymbol{\Sigma}\tilde{\boldsymbol{B}}(t,u)$*.*

Remark 2. Because we use the number of quantile levels, n_A, to ensure a good estimate of the bivariate function in the quantile interval, n_A should not be too small. However, larger n_A will result in a larger number of triangle basis functions, which will increase the variance of the estimator. Thus, in our theorems, n_A needs to satisfy $n_A^{-1}|\Delta|^{-1}m^{(\alpha-1)/3} = o(1)$ and $n_A n|\Delta^{d+2}| = o(1)$.

The next theorem describes how to construct a simultaneous confidence region (SCR) for $\beta(t,u)$. Let $\Gamma_{min}(\cdot)$ and $\Gamma_{max}(\cdot)$ represent the minimum and maximum eigenvalues, respectively, of a square matrix. Let $\mathbf{\Omega}_s$ denote the set of vertices of the triangulation Δ, and $|\mathbf{\Omega}_s|$ denote the cardinality of the set $\mathbf{\Omega}_s$.

Theorem 3. *Under the conditions of Theorem 2, and assuming that $\Gamma_{min}(\boldsymbol{\Sigma})$ and $\Gamma_{max}(\boldsymbol{\Sigma})$ are bounded away from zero and ∞, respectively, with probability tending to one as $n \to \infty$,*

(1) *As $n, n_A \to \infty$, we have*

$$\sigma_\beta^{-1/2}(t,u)\left\{\hat{\beta}(t,u) - \beta(t,u)\right\} \to \vartheta(t,u) \tag{3.1}$$

in distribution, where $\vartheta(t,u)$ is a Gaussian random field with mean zero defined on $\mathbf{\Omega}$ with the covariance function

$$\begin{aligned} C(t,u,t',u') &:= Cov\left(\vartheta(t,u), \vartheta(t',u')\right) \\ &= \sigma_\beta^{-1/2}(t,u)\sigma_\beta^{-1/2}(t',u')\tilde{\boldsymbol{B}}^{\mathrm{T}}(t,u)\boldsymbol{\Sigma}\tilde{\boldsymbol{B}}(t',u'). \end{aligned}$$

Specifically, $C(t,u,t,u) = Var\left(\vartheta(t,u)\right) = 1$.

(2) *For any $a \in (0,1)$,*

$$\lim_{n\to\infty} P\left\{\sup_{(t,u)\in\mathbf{\Omega}_s}\left|\sigma_\beta^{-1/2}(t,u)\left\{\hat{\beta}(t,u) - \beta(t,u)\right\}\right| \le Q_\beta(a)\right\} = 1 - a, \tag{3.2}$$

where $\mathbf{\Omega}_s$ as a subset of $\mathbf{\Omega}$ becomes denser as $n \to \infty$, and $Q_\beta(a) = (2\log|\mathbf{\Omega}_s|)^{1/2} - (2\log|\mathbf{\Omega}_s|)^{-1/2}\{\log(-0.5\log(1-a)) + 0.5[\log(\log|\mathbf{\Omega}_s|) + \log 4\pi]\}$. Then, an asymptotic $100(1-a)\%$ SCR for $\beta(t,u)$ over $\mathbf{\Omega}_s$ is given by $\hat{\beta}(t,u) \pm \sigma_\beta^{1/2}(t,u)Q_\beta(a)$.

Remark 3. In Theorem 2, the condition $n_A n|\Delta|^{d+2} = o(1)$ is used for undersmoothing of the slope estimator, and is widely applied in series approximating estimations (Yu et al. (2020, 2021)). By consistently estimating the asymptotic variance $\sigma_\beta(t,u)$, we can use the result in Theorem 2 to establish the pointwise confidence interval of the slope function. Compared with the asymptotic $100(1-a)\%$ point confidence interval in Theorem 2, $\hat{\beta}(t,u) \pm \sigma_\beta^{1/2}(t,u)z_a$, the width of the SCR in Theorem 3 for any $(t,u) \in \mathbf{\Omega}_s$ is inflated by the rate $Q_\beta(a)/z_a$, where z_a is the a-quantile of the standard normal distribution.

4. Applications

4.1. The capital bike share program

Urban population growth and increasing air pollution, greenhouse gas emissions, and other environmental problems have led to some people using bicycles as a healthy and eco-friendly alternative to driving to work, especially in big cities.

Rather than owning a bicycle, many people rent one as an economical and environmentally friendly alternative. As a result, bike-sharing systems have become an essential part of urban mobility in many major cities.

Because cycling is an outdoor activity, customers' rental behaviors are affected by weather conditions. Thus, a successful business needs to have a good strategy to adjust the supply of available bicycles to meet demand based on weather conditions. Here, we seek to quantify the effect of weather conditions on bicycle rentals, focusing on the relationship between the total daily number of rentals and hourly temperature.

The data set is taken from a study (Fanaee-T and Gama (2014)) on rentals to cyclists without membership in the Capital Bike Share program in Washington D.C. from January 1, 2011, to December 31, 2012. The data set includes hourly counts of casual bike rentals every day, the weather conditions, and hourly temperature measurements. The demand for bicycles differs between weekdays and weekends. We restrict our analysis to data observed on weekends. Specifically, we consider the temperature measurements and the counts of rentals between 7:00 and 17:00 on Saturdays and Sundays without rain or snow. The goal of our analysis is to investigate how the hourly temperature affects the lower, middle, and upper quantiles of the daily total rentals on weekends.

Figure 2 shows the estimated slope function $\hat{\beta}(t, u)$ for $u = 10\%, 20\%, 50\%$, and 90%. In the top two panels of Figure 2, the slope function is negative in the early morning and becomes positive at noon and in the afternoon, which are the peak demand periods for bike rentals. Because the temperature in the early morning is usually cooler than later in the day, the cumulative effect of temperature on rentals is positive. Here, the lower bounds of rentals are given by the 10% and 20% quantiles, indicating low rental demand. The result on the 50% quantile, displayed in the bottom-left panel in Figure 2, represents the normal situation, and shows a similar pattern to that of the lower quantiles.

The bottom-right panel in Figure 2 shows that when $u = 90\%$, the slope function is negative in the early morning before 9:00 and in the late afternoon after 15:30. This may be because higher temperatures in the morning deter rentals at noon and during the afternoon. If the temperature is high in the morning, then it is likely to remain high during the day. In addition, the late afternoon is usually the hottest time of the day, and may be too hot for biking. On the other hand, a cool morning may indicate a comfortable biking temperature

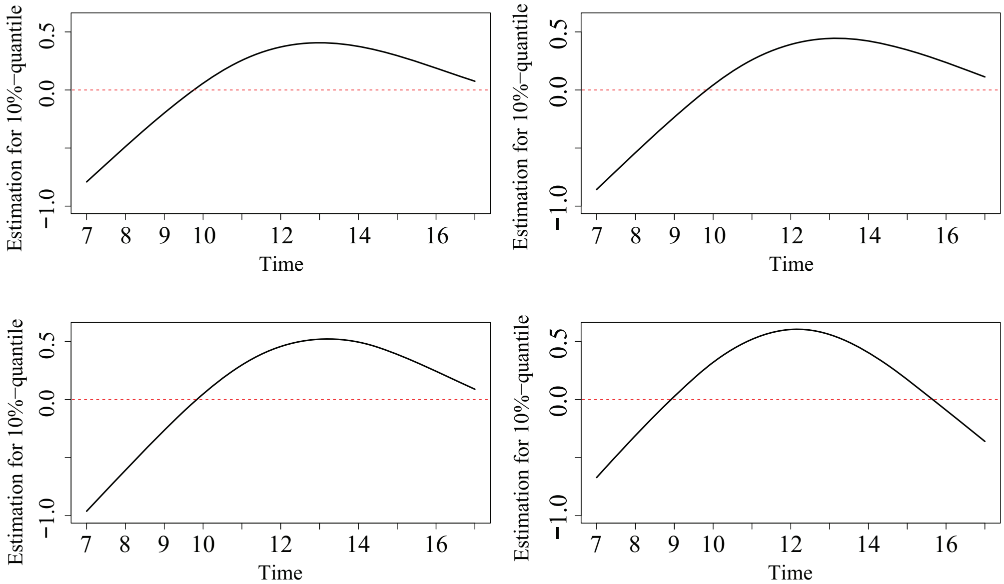


Figure 2. The estimated slope function $\hat{\beta}(t, u)$ for the regression model (2.1) at quantiles $u = 10\%, 20\%, 50\%, 90\%$ based on data collected from the Capital Bike Share program in Washington D.C. from 7:00 to 17:00 on weekends. The unit of y-axis is per 1,000 bicycles.

for the peak demand periods at noon and in the afternoon. The 90% quantile indicates high bike rental demand, showing that the weather needs to be cool in the morning, and comfortable or moderate in the afternoon.

To give an overall visualization of the estimated $\hat{\beta}(t, u)$, Figure 3(a) displays the heat map of $\hat{\beta}(t, u)$ estimated from the proposed method for the time t from 7:00 to 17:00 and the quantile u from 10% to 90%. The estimated slope function $\hat{\beta}(t, u)$ is positive after 9:00 for the quantiles u from 10% to 60%, and gradually becomes negative in the late afternoon for quantiles u from 60% to 90%.

Figure 3(b) shows the heat map of the estimation for $\beta(t, u)$ derived from the conventional method Kato (2012), which is not smooth. In addition, the proposed method can overcome the problem of the monotonicity of the quantile estimates. Figure 4 compares the estimated quantile functions of the 60th and the 100th subjects derived from the conventional method (Kato (2012)) and the proposed method. Let $Q^*_{60}(u)$ and $Q^*_{100}(u)$ be the estimated quantile functions of the 60th and the 100th subjects, respectively, derived from the conventional method (Kato (2012)), and $Q_{60}(u)$ and $Q_{100}(u)$ be the corresponding estimated quantile functions derived from the proposed method. We can observe that $Q^*_{60}(u)$ and $Q^*_{100}(u)$ are not monotone over the interval $u \in [0.1, 0.9]$, as they should be, whereas $Q_{60}(u)$ and $Q_{100}(u)$ are both monotonically increasing in $u \in [0.1, 0.9]$.

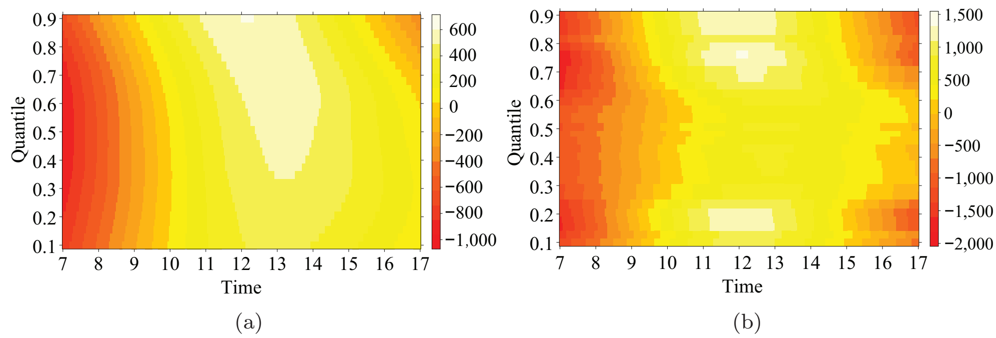


Figure 3. Heat maps of the estimated slope function $\hat{\beta}(t,u)$ for the regression model (2.1) derived using the proposed method (Panel (a)) and the conventional method (Kato (2012)) (Panel (b)) based on the data collected from the Capital Bike Share program in Washington D.C. from 7:00 to 17:00 on weekends.

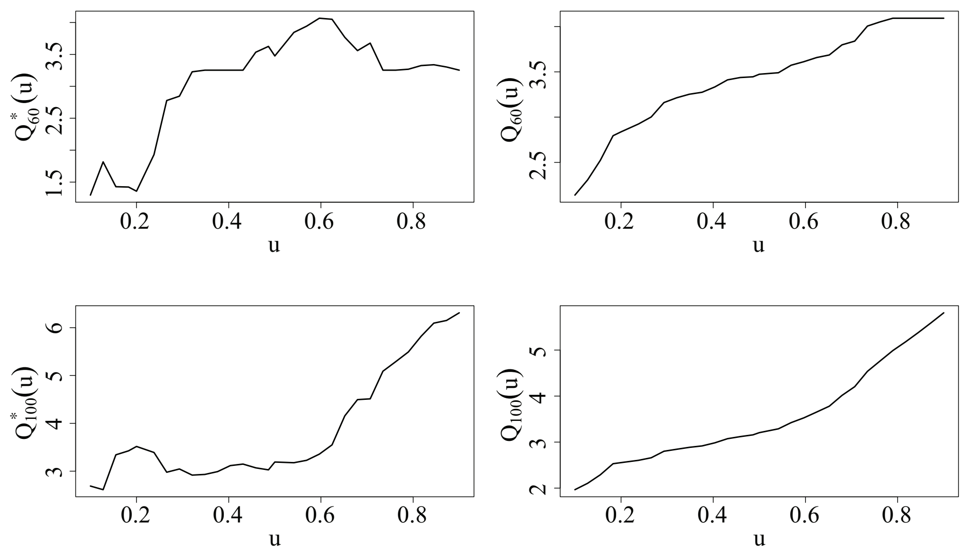
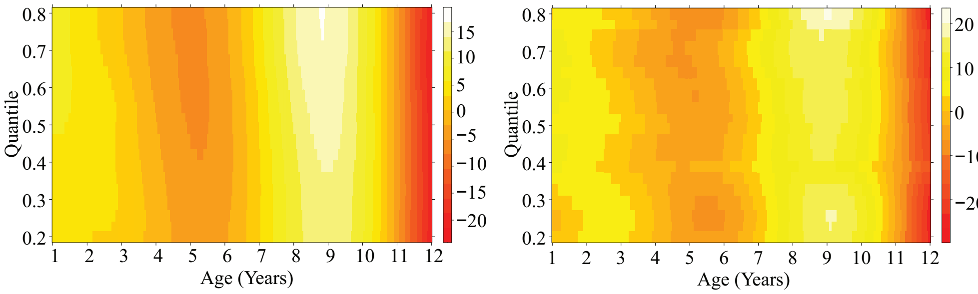


Figure 4. Estimated quantile functions of the 60th and 100th subjects derived from the conventional method (Kato (2012)) (shown in the left two panels) and the proposed method (shown in the right two panels) based on the data collected from the Capital Bike Share program in Washington D.C. from 7:00 to 17:00 on weekends. The unit of y-axis is per 1,000 bicycles.

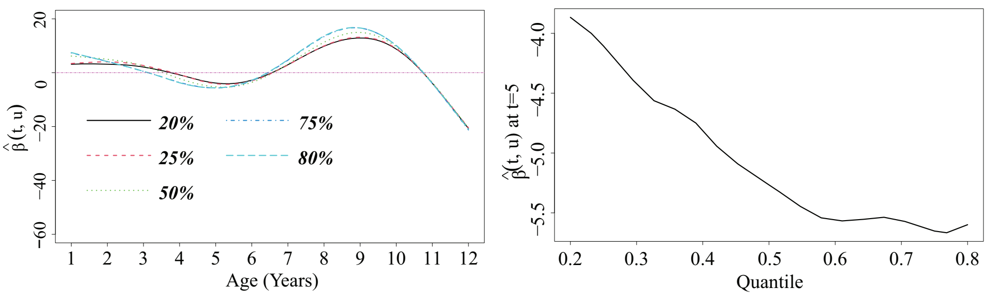
4.2. Berkeley growth data

Child's height growth is an important health indicator, and abnormal growth usually implies an underlying health problem or growth disorder. It is thus helpful to understand the relationship between children's growth history and their adult height in order to evaluate their health and growth progress. If the predicted



(a) (b)

Figure 5. Heat maps of the estimated slope function $\hat{\beta}(t,u)$ for the regression model (2.1) derived using the proposed method (Panel (a)) and using the conventional method Kato (2012) (Panel (b)) based on Berkeley growth data for children between the ages of one and 12.



(a) (b)

Figure 6. The estimated slope function $\hat{\beta}(t,u)$ for the regression model (2.1) at $u = 20\%, 25\%, 50\%, 75\%$, and 80% for age t from one to 12, and the estimated slope function $\hat{\beta}(t,u)$ at age $t = 5$ for u from 20% to 80% based on the Berkeley growth data.

adult height of a child has an abnormally small lower quantile, then interventions should be considered during their teenage years to treat any potential health problems that might affect height growth.

To investigate this relationship, we use children's growth history between the ages of one and 12 as a functional covariate (Chen and Müller (2012)), and the conditional quantile of their eighteen-year-old height as the response variable. We apply the proposed method to the Berkeley growth data (Tuddenham and Snyder (1954)) to estimate the slope function $\beta(t,u)$ from model (2.1).

Figure 5(a) displays $\hat{\beta}(t,u)$ for $u \in [0.2, 0.8]$ and $t \in [1, 12]$. The major variation of $\hat{\beta}(t,u)$ along the direction of u (y-axis variable) occurs between the ages of one and six. For any fixed age $t \geq 7$, $\hat{\beta}(t,u)$ does not change significantly as a function of u.

Figure 6(a) displays $\beta(t,u)$ as a function of t for $u = 20\%, 25\%, 50\%, 75\%$,

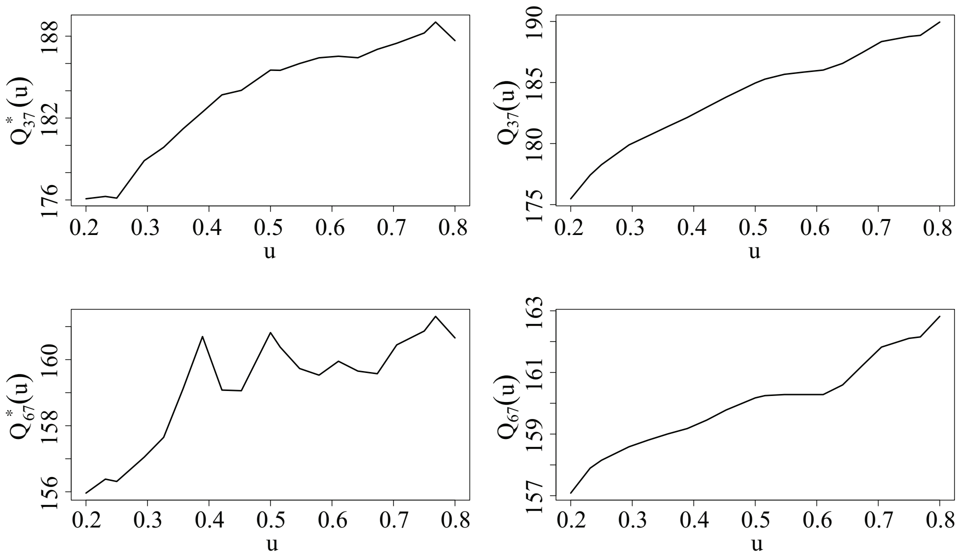


Figure 7. Estimated quantile functions of the 37th and 67th subjects derived from the conventional method (Kato (2012)) (shown in the left two panels) and the proposed method (shown in the right two panels) based on the Berkeley growth data for ages one to 12.

and 80%. It shows that children's growth history between the ages of seven and 11 is always positively correlated with the quantiles of their adult height. This interval may be regarded as a growth spurt. If a child has a significantly lower height compared with the normal level during the growth spurt period, then an intervention should be considered.

Figure 6(b) shows the estimated slope function $\hat{\beta}(t, u)$ as a function of u from 0.2 to 0.8 when $t = 5$, which is a negative function for any $u \in [0.2, 0.8]$. It indicates that the early growth spurt is not always a good indicator of a child's adult height. The early spurt may decrease children's potential to have a higher adult height because of sex hormone levels in their bodies (Soliman et al. (2014)). These children grow taller than other kids when they are young. However, their skeletons mature more rapidly. Consequently, they may stop growing at an early age, and have an average or a below average height as adults.

Similarly to the previous application, Figures 5 and 7 compare the performance of the proposed method and the conventional method. In Figure 7, $Q^*_{37}(u)$ and $Q^*_{67}(u)$ are defined in the same way as the previous application. Clearly, the quantile estimations obtained from the conventional method (Kato (2012)) are not monotone over the interval $u \in [0.2, 0.8]$, whereas the proposed method guarantees the desired monotonicity.

5. Conclusion

We have proposed a novel framework enabling a simultaneous FQR to overcome the two major limitations of conventional methods. When the true slope function is not a univariate function of time, our framework provides a better estimation for the slope function than that of the conventional estimation strategy, which estimates the slope function as a univariate function by first fixing the quantile index. This advantage of the proposed method is examined using simulation studies in a comparison with the method of (Kato (2012)). In addition, the proposed framework addresses the two major limitations of conventional methods. Within the proposed framework, the estimated conditional quantile functions are guaranteed to be monotone and their smoothness can be controlled.

In the current model (1.1), we consider only a single functional covariate. This may not be flexible enough to capture all information in the data. Chen and Müller (2012) proposed a generalized version of model (1.1) by using the composition of some link function and the linear functional of the functional covariate. In practice, it is common for several accompanying scalar covariates to be observed along with the functional covariate. Thus, Tang and Kong (2017) include a linear combination of the scalar covariates in the model. Moreover, we often observe multiple functional covariates simultaneously. To consider multiple functional covariates, Ma et al. (2019) extended the model to incorporate a linear combination of multiple functional covariates with different slope functions.

Although we present our method based on model (1.1), it can be extended to different settings of the FQR model, such as sparse functional observations (Yao, Müller and Wang (2005); Che et al. (2017)). Therefore, in future work, we will extend our framework to include multivariate functional covariates and finite-dimensional covariates. We will also investigate the properties and performance of our method in the scenario of sparse functional observations.

Supplementary Material

The online Supplementary Material presents detailed simulation studies and technical proofs of the asymptotic results. We provide the code and data required to reproduce the numerical results in the simulation studies and applications at `https://github.com/caojiguo/FunQR/`.

Acknowledgments

The authors gratefully acknowledge the useful comments and constructive suggestions provided by the editor, Professor Rong Chen, the associate editor, and two anonymous referees. Dr. Liu's research was supported by the National Natural Science Foundation of China (NSFC, No. 12201487), the Project funded by China Postdoctoral Science Foundation (No. 2022M722544), and the

Fundamental Research Funds for the Central Universities (SK2022044). Dr. You's research was supported by the National Natural Science Foundation of China (NSFC, No. 11971291) and Innovative Research Team of Shanghai University of Finance and Economics (2022110918). Dr. Cao's research was supported by a Discovery grant (RGPIN-2023-04057) from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Bondell, H. D., Reich, B. J. and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika* **97**, 825–838.

Cao, J. and Ramsay, J. O. (2010). Linear mixed effects modeling by parameter cascading. *Journal of the American Statistical Association* **105**, 365–374.

Cardot, H., Crambes, C. and Sarda, P. (2005). Quantile regression when the covariates are functions. *Journal of Nonparametric Statistics* **17**, 841–856.

Cardot, H., Crambes, C. and Sarda, P. (2007). Ozone pollution forecasting using conditional mean and conditional quantiles with functional covariates. In *Statistical Methods for Biostatistics and Related Fields*, 221–243. Springer, New York.

Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica* **13**, 571–591.

Che, M., Kong, L., Bell, R. C. and Yuan, Y. (2017). Trajectory modeling of gestational weight: A functional principal component analysis approach. *PloS One* **12**, e0186761.

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., Fernández-Val, I. and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* **96**, 559–575.

Fanaee-T, H. and Gama, J. (2014). Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence* **2**, 113–127.

He, Q., Kong, L., Wang, Y., Wang, S., Chan, T. A. and Holland, E. (2016). Regularized quantile regression under heterogeneous sparsity with application to quantitative genetic traits. *Computational Statistics & Data Analysis* **95**, 222–239.

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

Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *Journal of the American Statistical Association* **96**, 458–468.

Lai, M.-J. and Schumaker, L. L. (2007). *Spline Functions on Triangulations*. Cambridge University Press, Cambridge.

Li, M., Wang, K., Maity, A. and Staicu, A.-M. (2022). Inference in functional linear quantile regression. *Journal of Multivariate Analysis* **190**, 104985.

Liu, Y. and Wu, Y. (2011). Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. *Journal of Nonparametric Statistics* **23**, 415–437.

Ma, H., Li, T., Zhu, H. and Zhu, Z. (2019). Quantile regression for functional partially linear model in ultra-high dimensions. *Computational Statistics & Data Analysis* **129**, 135–147.

Nie, Y. and Cao, J. (2020). Sparse functional principal component analysis in a new regression framework. *Computational Statistics & Data Analysis* **152**, 107016.

Nie, Y., Wang, L., Liu, B. and Cao, J. (2018). Supervised functional principal component analysis. *Statistics and Computing* **28**, 713–723.

Prautzsch, H., Boehm, W. and Paluszny, M. (2002). *Bézier and B-spline Techniques*. Springer Science & Business Media.

Ramsay, J. O., Hooker, G. and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis*. Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. 2nd Edition. Springer, New York.

Sang, P., Wang, L. and Cao, J. (2017). Parametric functional principal component analysis. *Biometrics* **73**, 802–810.

Shi, H., Dong, J., Wang, L. and Cao, J. (2021). Functional principal component analysis for longitudinal data with informative dropout. *Statistics in Medicine* **40**, 712–724.

Soliman, A., De Sanctis, V., Elalaily, R. and Bedair, S. (2014). Advances in pubertal growth and factors influencing it: Can we increase pubertal growth? *Indian Journal of Endocrinology and Metabolism* **18**, S53–S62.

Stone, C. J., Hansen, M., Kooperberg, C. and Truong, Y. K. (1997). Polynomial splines and their tensor products in extended linear modeling (with discussion). *The Annals of Statistics* **25**, 1371–1470.

Tang, Q. and Kong, L. (2017). Quantile regression in functional linear semiparametric model. *Statistics* **51**, 1342–1358.

Tuddenham, R. and Snyder, M. (1954). Physical growth of California boys and girls from birth to eighteen years. *University of California Publications in Child Development* **1**, 183–364.

Wang, Y., Kong, L., Jiang, B., Zhou, X., Yu, S., Zhang, L. et al. (2019). Wavelet-based Lasso in functional linear quantile regression. *Journal of Statistical Computation and Simulation* **89**, 1111–1130.

Wang, Y., Wang, G., Wang, L. and Ogden, R. T. (2020). Simultaneous confidence corridors for mean functions in functional data analysis of imaging data. *Biometrics* **76**, 427–437.

Wu, Y., Ma, Y. and Yin, G. (2015). Smoothed and corrected score approach to censored quantile regression with measurement errors. *Journal of the American Statistical Association* **110**, 1670–1683.

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.

Yu, D., Kong, L. and Mizera, I. (2016). Partial functional linear quantile regression for neuroimaging data analysis. *Neurocomputing* **195**, 74–87.

Yu, S., Wang, G., Wang, L., Liu, C. and Yang, L. (2020). Estimation and inference for generalized geoadditive models. *Journal of the American Statistical Association* **115**, 761–774.

Yu, S., Wang, G., Wang, L. and Yang, L. (2021). Multivariate spline estimation and inference for image-on-scalar regression. *Statistica Sinica* **31**, 1–39.

Zhang, X., Cao, J. and Carroll, R. J. (2017). Estimating varying coefficients for partial differential equation models. *Biometrics* **73**, 949–959.

Zhang, Z., Wang, X., Kong, L. and Zhu, H. (2021). High-dimensional spatial quantile function-on-scalar regression. *Journal of the American Statistical Association*.

Zhao, Z. and Xiao, Z. (2014). Efficient regressions via optimally combining quantile information. *Econometric Theory* **30**, 1272–1314.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics* **36**, 1108–1126.

Boyi Hu

Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

E-mail: boyi_hu@sfu.ca

Xixi Hu

Sauder School of Business, University of British Columbia, Vancouver, BC V6T 1Z2, Canada.

E-mail: xixi.hu@sauder.ubc.ca

Hua Liu

School of Economics and Finance, Xi'an Jiaotong University, Xi'an, Shaanxi 710061, China.

E-mail: liuhua_sufe@163.com

Jinhong You

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China.

E-mail: johnyou07@163.com

Jiguo Cao

Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

E-mail: jjiguo_cao@sfu.ca

(Received July 2021; accepted September 2022)