Statistica Sinica Preprint No: SS-2021-0248	
Title	Simultaneous Functional Quantile Regression
Manuscript ID	SS-2021-0248
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0248
Complete List of Authors	Boyi Hu, Xixi Hu, Hua Liu, Jinhong You and Jiguo Cao
Corresponding Authors	Jiguo Cao
E-mails	Jiguo_cao@sfu.ca
Notice: Accepted version subject to English editing.	

Simultaneous Functional Quantile Regression

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

[1] *Department of Statistics and Actuarial Science, Simon Fraser University*

[2] *Sauder School of Business, University of British Columbia*

[3] *School of Economics and Finance, Xi'an Jiaotong University*

[4] *School of Statistics and Management, 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. There are several limitations to this conventional strategy. For example, the monotonicity of conditional quantiles can not be guaranteed, and the smoothness of the slope estimator as a bivariate function can not be controlled. In this paper, we develop a new framework for functional quantile regression. We propose to simultaneously fit the functional quantile regression model for multiple quantiles with the help of a bivariate basis under some constraints so that the estimated quantiles satisfy the monotonicity conditions. Meanwhile, the smoothness of the slope estimator is controlled. The proposed estimator for the slope function is shown to be asymptotically consistent. In addition, we also establish its asymptotic normality. Simulation studies are implemented to evaluate the finite sample performance of the proposed method in comparison with the conventional method. The proposed method is further demonstrated by analyzing the impact of daily temperature on the bike rental and investigating the relationship between the children's growth history and their adult height.

Key words and phrases: Bivariate Spline Basis; Functional Data Analysis; Non-crossing Quantiles;

1. Introduction

The u-th 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. The model (1.1) is called the functional quantile regression 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 commonly used strategy for estimating $\beta(t,u)$ is to treat it as a univariate function of t by fixing the quantile u first. This conventional strategy has two major limitations. First, the slope function $\beta(t,u)$ is usually assumed to be smooth over both t and u, which is also favorable in real applications. 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 invalid distribution estimation for the response variable.

In this paper, we address the above two limitations. Different from the existing methods that $\beta(t,u)$ is estimated as the univariate function of t for each fixed u, we propose to use bivariate spline basis functions to approximate $\beta(t,u)$ directly and then estimate 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 further impose some extra linear constraints to mitigate the crossing-quantile problem. In this way, we can make sure that the estimated quantiles for each subject are monotone. The monotonicity issue, to some extent, can be fixed by using some monotonization techniques, such as Chernozhukov et al. (2009). But it can not improve the estimation for $\beta(t,u)$, because the monotonicity of the quantiles are not considered in the estimation procedure for $\beta(t,u)$ and the monotonization is only applied to the estimated quantiles. For example, Kato (2012) proposed first to estimate $\beta(t,u)$ for the model (1.1), and then to estimate conditional quantile functions based on the estimated $\beta(t,u)$. For the quantile functions that are not monotone, he adjusted them to become monotone by using the technique of Chernozhukov et al. (2009). However, the estimation for $\beta(t, u)$ was left unchanged.

The model we consider is an extension of 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, LQR allows us to depict and then make the inference on the entire distribution of the response conditioning on the predictors. Linear quantile regression has been well studied and makes many contributions to real-world applications (Koenker and Geling, 2001; Wu et al., 2015).

Nowadays, functional variables becomes more and more common in real-world applications. Functional data analysis has become a comprehensive branch of statistics that provides a useful and convenient framework to analyze functional data with some high dimensional structures, such as curves, images, and surfaces, which are so-called functional data. Estimation for a quantile regression with a scalar response and some functional covariates is a fruitful research topic as related questions arise in many recent applications, such as Cardot et al. (2007), Chen and Müller (2012), Yu et al. (2016), Wang et al. (2019) and Zhang et al. (2021).

The model (1.1) was first formulated in Cardot et al. (2005) as a natural extension of classical linear quantile regression. In the paper, a penalized spline estimator for $\beta(t,u)$ was proposed for a fixed u without any dimension reduction on the functional covariate. Later on, for the same model (1.1), Kato (2012) proposed to first use functional principal component analysis (FPCA) to truncate the functional covariate $X(t)$ for dimension reduction and then to estimate the slope function $\beta(t,u)$ for a fixed u based on the conventional linear quantile regression framework. Kato (2012) also established an optimal convergence rate for the proposed estimator under 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, the proposed estimation method for the slope function is illustrated using two real-world applications.

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 u-th quantile of Y given the functional covariate $X(t)$ is modelled 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 to first approximate $\beta(t, u)$ by bivariate splines, and then estimate the corresponding coefficients.

There are multiple types of bivariate splines that can be used for the approximation, such as tensor products of B-splines (Stone et al., 1997; Prautzsch et al., 2002; Zhang et al., 2017) and bivariate Bernstein polynomials over triangulations (Lai and Schumaker, 2007). In this paper, we choose the Bernstein polynomials over a triangulation to approximate the bivariate slope function in (2.1). In comparison with the tensor products of B-splines, the bivariate Bernstein polynomials enjoy the advantage that the triangulation technique allows the 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 the triangulation technique are not a must for the proposed method and other bivariate bases should also work.

Figure 1 shows the 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 ver-

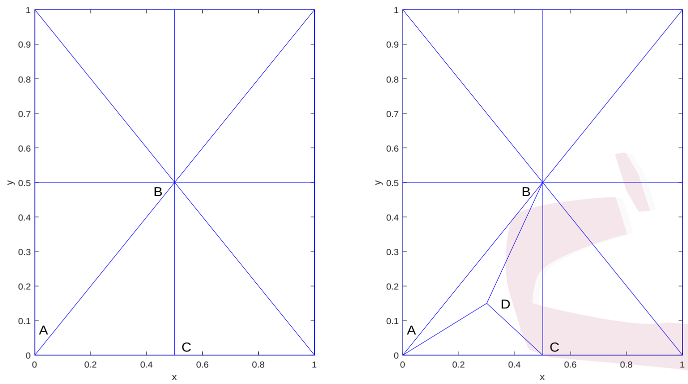


Figure 1: Example of local refinement of triangulation. The left panel shows a triangulation over $[0,1] \times [0,1]$. The right panel shows the triangulation after a local refinement by adding a new vertex D inside the triangle $\triangle\texttt{ABC}$.

tex D inside the triangle $\triangle\texttt{ABC}$. The triangle $\triangle\texttt{ABC}$ is further split into three triangles: $\triangle\texttt{ABD}$, $\triangle\texttt{BCD}$ and $\triangle\texttt{ACD}$.

Suppose that $\mathscr{A}$ is the interval containing multiple quantiles of interest. Our goal is to find a function $s(t,u) \in S_d^r(\Delta)$ that can well approximate 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 with $\Delta_j =$ the 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 $X(t)$, which are commonly assumed in the literature (Yao et al., 2005; Sang et al., 2017; Nie et al., 2018; Nie and Cao, 2020; Shi et al., 2021) and are usually satisfied in real applications, by the 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 $\phi_k(t)$, $k = 1, \ldots$, are called functional principal components (FPC) and ξ_k are called functional principal component 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 $\boldsymbol{\gamma}_0^{\mathrm{T}} = (\gamma_{0,1}(u), \ldots, \gamma_{0,J_0}(u))$.

Under the functional data context, functional observations as the infinite dimension subjects can not fit in the conventional framework of linear quantile regression. In addition, the observed functional data is not always smooth enough for using numerical integration to approximate the integral in (2.1). To tackle these problems and to extend the classical linear quantile regression to functional quantile regression, 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 functional principal components obtained from the decomposition (2.3).

As we have mentioned in the introduction, different from the conventional methods, we want to estimate $\beta(t,u)$ as a bivariate function directly. Therefore, all the quantiles of interest should be considered simultaneously in the estimation procedure. There are a lot of papers that have discussed the advantage of combining multiple quantile regression models, such as Zou and Yuan (2008); Kai et al. (2011); Zhao and Xiao (2014); He et al. (2016). One commonly used approach is to consider the sum of those 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 data pairs $\{y_i, x_i(t)\}_{i=1}^n$ as the 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 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 framework of linear quantile regression is designed for finite dimensional subjects, and the slope parameter to be estimated in classical linear quantile regression is also finite-dimensional. Although the functional observations can be truncated by FPCA into a finite dimension, the slope function $\beta(t,u)$ in the model (2.1) is still infinite-dimensional. As a consequence, the direct extension (2.4) of the conventional linear quantile regression framework to functional data can lead to invalid estimation for $\beta(t,u)$ and the uniqueness of the minimizer of (2.4) cannot be guaranteed.

To clarify this, let $\{\hat{s}(t,u), \hat{\gamma}_0\}$ be a minimizer of (2.4) and fix the truncation level at m. Assume that 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 $\{\hat{s}(t,u) + s_1(t,u), \hat{\gamma}_0\}$ is another minimizer of (2.4). More specifically, by FPCA, we can obtain the $(m+1)$-th FPC denoted as $\hat{\phi}_{m+1}(t)$, which is orthogonal to the first m estimated functional principal components, $\hat{\phi}_1(t), \ldots, \hat{\phi}_m(t)$. 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 $\left\{\hat{s}(t,u) + w(u)\hat{\phi}_{m+1}(t), \hat{\gamma}_0\right\}$ 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 the estimator for $\beta(t,u)$ derived from the estimation procedure should not include any information about $\hat{\phi}_k(t)$ for any $k \geq m+1$. Therefore, the objective function (2.4) derived directly from the conventional linear quantile regression is problematic under the functional data context.

To overcome this problem, we propose to penalize 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 under functional data context. Therefore, a roughness penalty for $s(t,u)$ is also used during the estimation procedure. Roughness penalty is a very useful tool to control the smoothness of functions through the estimation procedure. Detailed discussions on roughness penalty for functional data can be found in Ramsay and Silverman (2002), Cardot et al. (2003), Ramsay and Silverman (2005), Ramsay et al. (2009) and Cao and Ramsay (2010). In this paper 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 three tuning parameters representing the weights corresponding to second derivatives in different directions. More specifically, ω_0 is the weight corresponding to $\frac{\partial^2 s}{\partial t^2}$. ω_1 is the weight corresponding to $\frac{\partial^2 s}{\partial t \partial u}$. ω_2 is the weight corresponding to $\frac{\partial^2 s}{\partial u^2}$. Since in our estimation procedure, we include a tuning parameter $\lambda_{2,n}$ for the whole roughness penalty $R(s;\omega_0,\omega_1,\omega_2)$, then ω_0 can be fixed as a constant $\omega_0 = 1$. In addition, if the smoothness of the target slope function along quantile index and functional index can be assumed to be identical or without much difference, 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 to $R(s)$,

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

which is the most commonly used roughness penalty discussed in the literature listed above.

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

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

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 $\boldsymbol{B}(t,u) = (b_1(t,u),\cdots,b_J(t,u))^{\mathrm{T}}$ 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, the degree d, and the 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 the continuous first partial derivatives over both t and u. An useful technique to remove the constraint (2.7) is QR decomposition (Wang et al., 2020). For a given $\boldsymbol{H}$, by 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 a 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 we use n_T to denote the cardinality of T. By (2.6), the penalty $\|s\|^2_{L^2(\boldsymbol{\Omega})}$ can be approximated by $\|s\|^2_{L^2(\boldsymbol{\Omega})} \approx \frac{1}{n_A n_T}\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}^{\mathrm{T}}_{A,T}\boldsymbol{\gamma}$, where $\boldsymbol{B}_{A,T}$ is a J by $n_A n_T$ matrix with the jth row of $\boldsymbol{B}_{A,T}$ being the evaluations of 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 as the matrix form $\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{\gamma}$, 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 $L_0(\boldsymbol{\theta}, \boldsymbol{\gamma}_0) = (nn_A)^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\rho_{u_r}\left(y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\boldsymbol{\gamma}_0 - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\boldsymbol{\theta}\right)$, the whole quantile loss based on FPCA. Then by (2.6) and (2.9), the minimization problem (2.5) can be converted into the following,

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

where $\hat{\boldsymbol{\xi}}_i = (\hat{\xi}_{i1}, \ldots, \xi_{im})^{\mathrm{T}}$, $\hat{\boldsymbol{P}}(u)$ 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 $(\hat{\gamma}_0, \hat{\boldsymbol{\theta}})$, 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 the estimated conditional quantile functions of all the subjects to be monotone, some extra linear constraints on $\boldsymbol{\theta}$ can be further imposed. Specifically, given $(\hat{\gamma}_0, \hat{\boldsymbol{\theta}})$, the estimated u-quantile of the ith subject is $\hat{Q}_Y(u \mid X = x_i) = \boldsymbol{b}_0^{\mathrm{T}}(u)\hat{\gamma}_0 + \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u)\boldsymbol{Q}\hat{\boldsymbol{\theta}}$.

The monotonicity of $\hat{Q}_Y(u \mid X = x_i)$ can be approximately expressed as $\hat{Q}_Y(u_r \mid X = x_i) \leq \hat{Q}_Y(u_r' \mid X = x_i)$ 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 into the optimization,

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

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 each other (Bondell et al., 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 is concerned with the computational aspect of the minimization problem (2.10). As introduced in Koenker and Bassett (1978), for a specific quantile u, the minimization of loss function derived from the classical linear quantile regression model is equivalent to a constrained linear programming problem.

For the proposed method, we need to solve the minimization problem (2.10). Following the same idea in Koenker and Bassett (1978), (2.10) can also be formulated into 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 want to further impose the monotonicity constraints (2.12) on (2.13), then the constrained optimization can be similarly formulated 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 $i = 1, \ldots, n$, $r = 1, \ldots, n_A$ and any $u_r < u_r'$, $u_r, u_r' \in A$.

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

- Derive the coefficients in (2.13) with or without the monotonicity constraints (2.12), such as $\{\hat{\boldsymbol{\xi}}_i\}_{i=1}^n$, $\{\hat{\boldsymbol{P}}(u)\}_{u\in A}$, $\boldsymbol{Q}$, etc. 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 by using numerical integration based on Simpson's rule.

- With all the preparations in the previous step, we can code the quadratic programming problem (2.13) with or without the constraints (2.12) in

MATLAB and solve it in MATLAB as well.

2.3 Tuning Parameter Selection

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

For the truncation level m, we suggest to use following BIC criterion to choose m,

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

For the selection of penalty parameters $\lambda_{1,n}$ and $\lambda_{2,n}$, ideally, leave-one-out cross-validation should be the best way to do it. However, the computational cost for each fitting is expensive, therefore, in practice, we usually use five-fold or ten-fold cross-validation to select the parameter values. As we will show in the next section, the value of $\lambda_{2,n}$ depends on the value of $\lambda_{1,n}$. For this reason, we propose a sequential procedure to choose values for $\lambda_{1,n}$ and $\lambda_{2,n}$. and the specific cross-validation procedure is described as follows. We use ten-fold cross-validation as an example.

We first use the complete sample $\{x_i(t)\}_{i=1}^{n}$ to estimate FPCs $\{\hat{\phi}_k(t)\}_{k=1}^{m}$, and corresponding scores $\{\hat{\boldsymbol{\xi}}_i\}_{i=1,\ldots,n}$. Then for a fixed m, we use ten-fold cross validation to find the optimal value for 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, denoted 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, denoted 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 will repeat this sequential selection procedure for multiple values of m, and then choose the optimal value for m based on the 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 independent and identically distributed.

(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 $k \geq 1$, $C^{-1}k^{-\alpha} \leq \kappa_k \leq Ck^{-\alpha}$, $\kappa_k - \kappa_{k+1} \geq C^{-1}k^{-\alpha-1}$.

(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 $\zeta > \alpha/2 + 1$, $\sup_{u\in\mathscr{A}} |\beta_k(u)| \leq Ck^{-\zeta}$, $k = 1, \ldots$, 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 the scenario of dependent data is not considered in this paper. A2 are commonly assumed restrictions on the moments of $X(t)$ and ξ_k. There is no condition on the moment of Y needed. A3 is adapted from (A3) of Kato (2012), which ensures the identifiability of $\phi_k(t)$ as well as the estimation accuracy of $\hat{\phi}_k(t)$. A4 are common conditions on the conditional distribution and density functions of Y under quantile regression 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 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 as previously defined. The following theorem gives the rate of convergence of the slope function estimator $\hat{\beta}(t, u)$ for a given truncation level m when the 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

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

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

Theorem 1. *Under the conditions A1-A5, and assume further that* $|\Delta| = o\left(m^{-(1+2\alpha)/(2d+2)} n^{-3/(2d+2)}\right)$, *and* $n_A^{-1}|\Delta|^{-1} m^{(\alpha-1)/3} = o(1)$. *Suppose the tun-*

ing 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*

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

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 $\left\|\hat{\beta}(t,u) - \beta(t,u)\right\|_{L^2(\mathbf{\Omega})}$ in Theorem 1 is decreasing as the sample size n becomes larger, and is increasing with a larger the truncation level m (i.e, adding more FPCs in 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 can 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 in the supplementary material, there exist $\boldsymbol{\gamma}_0^*$ and $\boldsymbol{\theta}^*$ such that

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

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

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

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 its (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 $\boldsymbol{\Sigma} = (2n\mathbf{\Sigma}_2)^{-1}\boldsymbol{U}_2/n\,(2n\mathbf{\Sigma}_2)^{-1}$.

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. Due to the fact that the number of quantile levels, n_A, is used to ensure a good estimate of the bivariate function in the quantile interval, n_A shouldn't be too small. Meanwhile, larger n_A will result in larger number of triangle basis functions, which will increase the variance of the estimator. So 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 presents how to construct a simultaneous confidence region (SCR) for $\beta(t, u)$. Let $\mathit{\Gamma}_{min}(\cdot)$ and $\mathit{\Gamma}_{max}(\cdot)$ represent the minimum and maximum eigenvalues 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 further assume that* $\Gamma_{min}(\boldsymbol{\Sigma})$ *and* $\Gamma_{max}(\boldsymbol{\Sigma})$ *are bounded away from 0 and* ∞ *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 0 defined on* $\boldsymbol{\Omega}$ *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 $a \in (0,1)$,

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

where $\boldsymbol{\Omega}_s$ *as a subset of* $\boldsymbol{\Omega}$ *becomes denser as* $n \to \infty$, *and* $Q_{\beta}(a) = (2\log|\boldsymbol{\Omega}_s|)^{1/2} - (2\log|\boldsymbol{\Omega}_s|)^{-1/2}\{\log(-0.5\log(1-a)) + 0.5[\log(\log|\boldsymbol{\Omega}_s|) + \log 4\pi]\}$. *Then an asymptotic* $100(1-a)\%$ *simultaneous confidence region (SCR) for* $\beta(t,u)$ *over* $\boldsymbol{\Omega}_s$ *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, which is widely applied in the series approximating estimations (Yu et al., 2020, 2021). By consistently estimating the asymptotic variance $\sigma_\beta(t,u)$, the result in Theorem 2 can be used 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 simultaneous confidence region 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

The Urban population is growing rapidly in recent years. Meanwhile, air pollution, greenhouse gas emissions, and other environmental problems are getting worse and worse as an increasing number of people need to drive to work. As an alternative to driving to work, especially in big cities, which are facing traffic, environmental and health issues, biking is a healthy and eco-friendly way.

Instead of owing bikes, renting bikes is considered a more economical and environmental-friendly alternative. Nowadays, bike-sharing systems have become an essential part of urban mobility in many major cities, and the number of cities that are becoming bike-friendly is increasing day by day.

As an outdoor activity, customers' rental behaviors are affected by weather conditions. A successful bike business needs to have a good strategy to adjust the supply of available bikes to meet the demands based on the weather conditions. Therefore, it is of great interest to quantify the weather condition's effect on the bike rental. Weather conditions can be measured using a wide range of factors. In this article, we investigate the relationship between the total daily number of bike rentals and the hourly temperature.

The data set we use in this study is from the Capital Bike Share study (Fanaee-T and Gama, 2014), which contains rentals to cyclists without membership in the Capital Bike Share program in Washington D.C. from January 1st, 2011 to December 31th 2012. The hourly counts of casual bike rentals every day, the weather conditions, and the hourly temperature measurements are all recorded in the data set. The demands of bike rentals are quite different between weekdays and weekends. In the study, we restrict our analysis to the data observed on the weekends. Specifically, we only consider the temperature measurements and the counts of bike rentals obtained between 7:00 and 17:00 on Saturdays and Sundays without raining or snowing. The goal of our analysis is to investigate how the hourly temperature affects the lower, middle, and upper quantiles of the daily total bike rentals during the weekends.

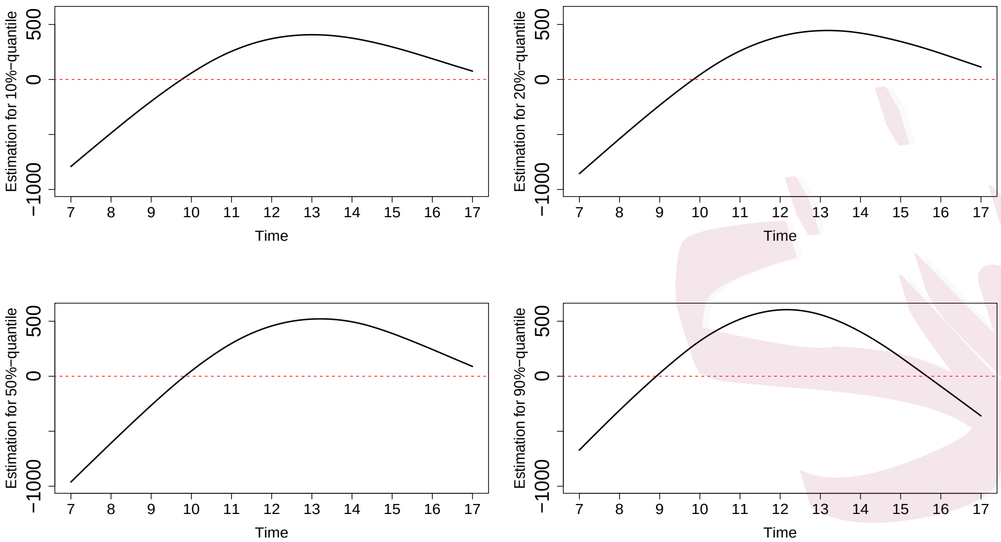


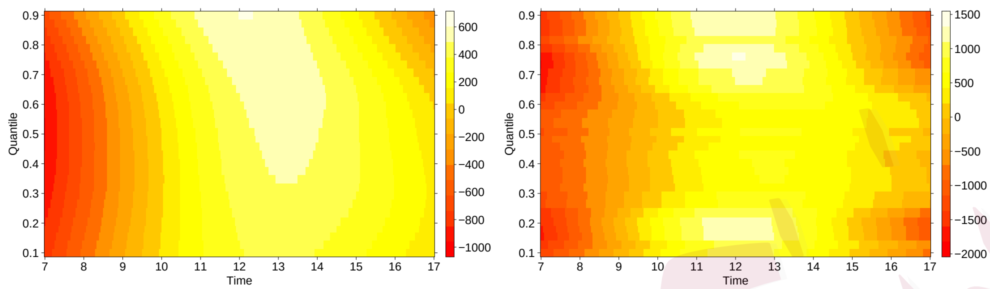
Figure 2: The estimated slope function $\hat{\beta}(t, u)$ for the regression model (2.1) at quantiles $u = 10\%, 20\%, 50\%, 90\%$ based on the data collected from Capital Bike Share program in Washington D.C. during 7:00 to 17:00 every weekend.

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 in the noon and afternoon, which are the peak demand periods for bike rentals. Since the temperature in the early morning is usually much cooler than the temperature in the noon and afternoon, the cumulative effect of temperature on the bike rental is positive. It indicates that as long as the overall temperature of that day is not too low, then the lower bounds of the bike rental will usually not be bad because the 10% and 20%

quantiles measure the worst situations for the bike rental demand. The result on the 50% quantile displayed in the bottom left panel in Figure 2 represents the normal situation. It shares a similar pattern with the lower quantiles that the cool morning and the warm noon are preferred for a normal bike demand.

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 late afternoon after 15:30. This may be due to the fact that a high temperature in the morning deters unnecessary bike rental at noon and afternoon. If the temperature is high in the morning, then the temperature of the whole day is usually very high as well. In addition, the late afternoon is usually the hottest time of the day, and on some days, the late afternoon temperature can be too high for biking. On the other hand, a cool morning may indicate a comfortable biking temperature for the peak demand periods: the noon and the afternoon. Since 90% quantile almost measures the most ideal situation for the bike rental demand, this plot indicates that for a high bike rental demand, 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%, while



(a) (b)

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

it gradually becomes negative in the late afternoon for quantiles u from 60% to 90%.

In comparison with the proposed method, Figure 3(b) shows the heat map of the estimation for $\beta(t, u)$ derived from the conventional method Kato (2012). We can observe that this estimation is not smooth. In addition to that, the proposed method can overcome the issue of the monotonicity of the quantile estimates. Figure 4 shows the comparison of 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

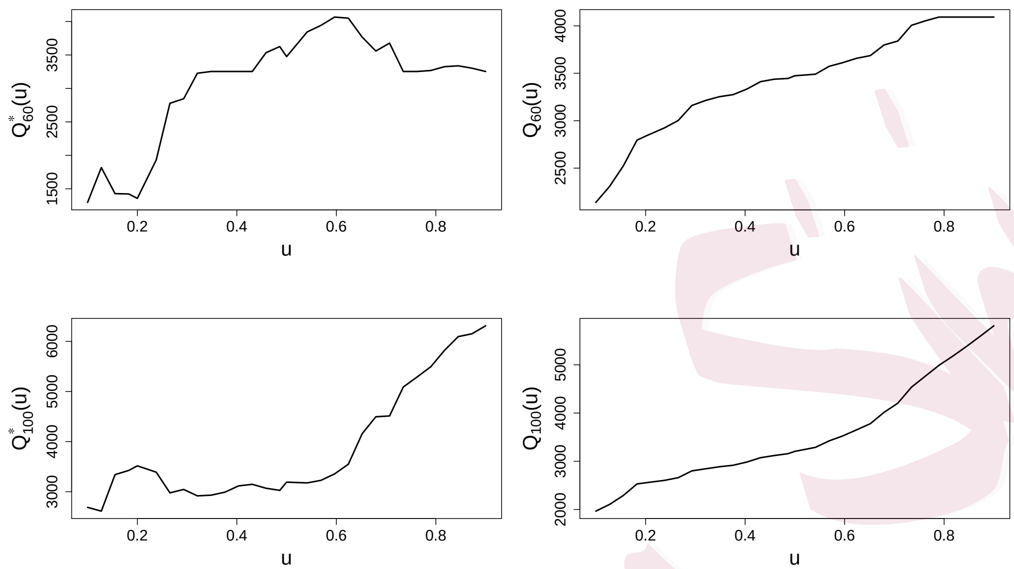


Figure 4: Estimated quantile functions of the 60th and the 100th subjects derived from the conventional method (Kato, 2012) (shown in the left two panels) and the proposed method (shown in the right two panels) based on the data collected from the Capital Bike Share program in Washington D.C. during 7:00 to 17:00 every weekend.

of the 60$^{\text{th}}$ and the 100$^{\text{th}}$ subjects 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 directly 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, while $Q_{60}(u)$ and $Q_{100}(u)$ are both monotonically increasing in $u \in [0.1, 0.9]$.

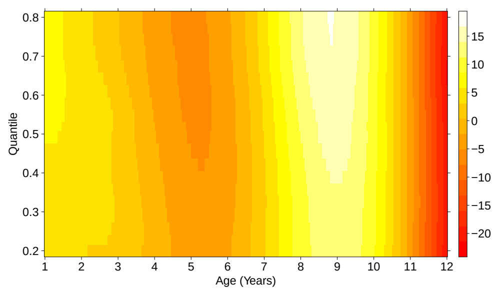
4.2 Berkeley Growth Data

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 to evaluate the health and growth progress of children. If the predicted adult height of a child has an abnormally small lower quantile, then interventions should be considered during their teenage year to treat any potential health problem that affects height growth.

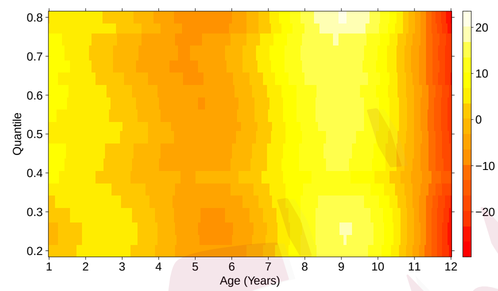
To investigate this relationship, we use the children's growth history between one-year-old and twelve-year-old as a functional covariate (Chen and Müller, 2012), and the conditional quantile of their eighteen-year-old heights as the response variable. We apply the proposed method to the Berkeley growth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

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

More specifically, Figure 6(a) displays $\beta(t, u)$ as a function of t for $u = 20\%, 25\%, 50\%, 75\%$, and 80%. It shows that the children's growth history be-



(a)

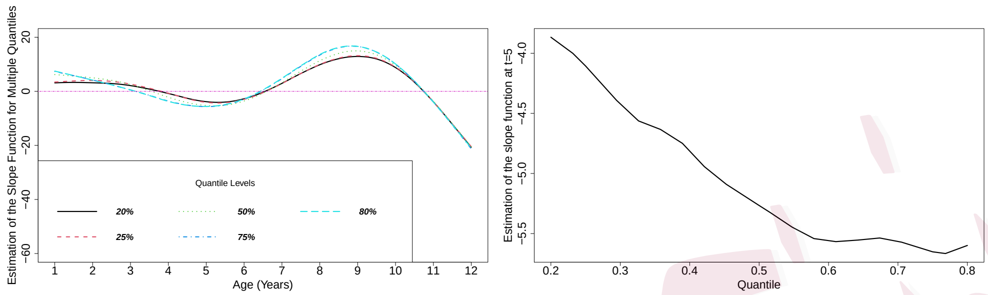


(b)

Figure 5: The heat maps of the estimated slope function $\hat{\beta}(t, u)$ for the regression model (2.1) derived from the proposed method (Panel (a)) and the conventional method Kato (2012) (Panel (b)) based on the Berkeley growth data for one-year-old to twelve-year-old.

tween age seven years and eleven years is always positively correlated with the quantiles of their adult height. This interval may be regarded as a growth spurt. If one child has a significantly lower height compared to the normal level during the growth spurt period, then some 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 sign for children's adult height. The early spurt may decrease children's potential to have a higher adult height due to the sex hormone levels in their bodies (Soliman et al., 2014).



(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% over the age t from one-year-old to twelve-year-old 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.

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 eventually, end up having average or below average height as adults.

Similar to the previous application, Figure 5 and Figure 7 also show the comparison of 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 $u \in [0.2, 0.8]$, while the proposed method can guarantee the desired monotonicity.

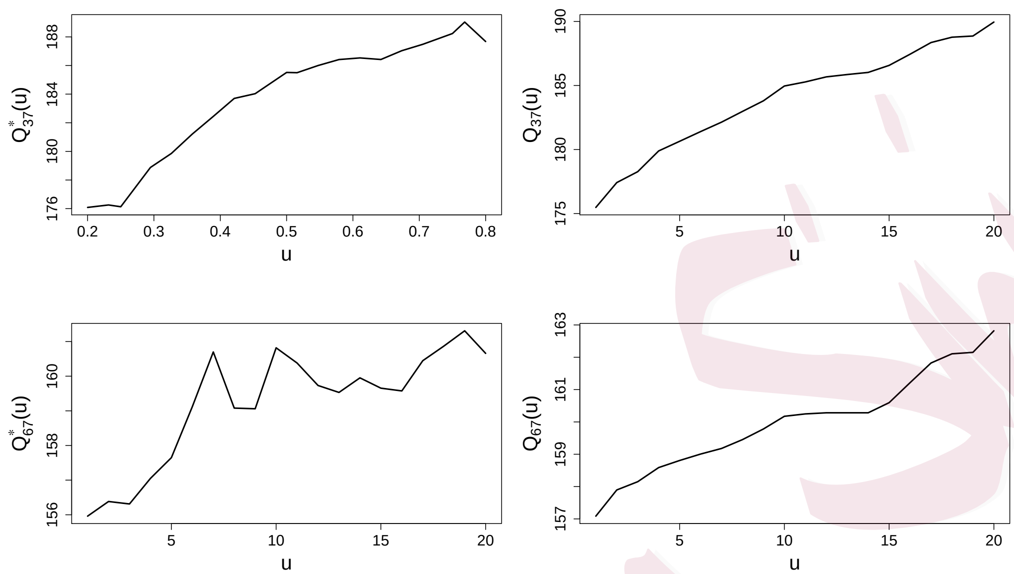


Figure 7: Estimated quantile functions of the 37th and the 67th subjects derived from the conventional method (Kato, 2012) (shown in the left two panels) and the proposed method (shown in the right two panels) based on the Berkeley growth data for one-year-old to twelve-year-old.

5. Conclusions and Discussion

We propose a novel framework for the simultaneous functional quantile regression to overcome the two major limitations of the conventional methods. When the true slope function is not a univariate function of time index, our framework can provide a better estimation for the slope function compared with the conventional estimation strategy that estimates the slope function as the univariate function by first fixing the quantile index. This advantage of the proposed method is examined by simulation studies in comparison with the method of (Kato, 2012). In addition, the proposed framework can solve the two major limitations of the 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 only consider a single functional covariate. It may not be flexible enough to capture all of the information from data. In Chen and Müller (2012), they proposed a generalized version of the model (1.1) by using the composition of some link function and the linear functional of the functional covariate. In practice, it is common that several accompanying scalar covariates are observed along with the functional covariate. For this reason, in Tang and Kong (2017), a linear combination of scalar covariates is included in the model. Moreover, we often observe multiple functional covariates simultaneously in applications. To take multiple functional covariates into account, Ma et al. (2019) further extended the model to incorporate a linear combination of multiple functional covariates with different slope functions.

Although we present our method based on the model (1.1), our proposed method can be further extended to different settings of the functional quantile regression model, such as sparse functional observations (Yao et al., 2005; Che et al., 2017). Therefore, for the future work, we will extend our framework to the model that contains 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 Materials

The supplementary document presents the detailed simulation studies and technical proofs of the asymptotic results. We provide the computing codes and data to reproduce the numerical results in the simulation studies and applications at the website `https://github.com/caojiguo/FunQR/`.

Acknowledgements

The authors gratefully acknowledge the Editor, Professor Rong Chen, the Associate Editor and two anonymous referees for their useful comments and constructive suggestions. Their feedback greatly helped to improve the quality of the paper. Dr. Liu's research was supported by the Fundamental Research Funds for the Central Universities (SK2022044). Dr. You's research was supported in part by the National Natural Science Foundation of China (11971291). Dr. Cao's research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant (RGPIN-2018-06008).

References

Bondell, H. D., B. J. Reich, and H. Wang (2010). Noncrossing quantile regression curve estimation. *Biometrika 97*(4), 825–838.

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

Cardot, H., C. Crambes, and P. Sarda (2005). Quantile regression when the covariates are functions. *Journal of Nonparametric Statistics 17*(7), 841–856.

Cardot, H., C. Crambes, and P. Sarda (2007). Ozone pollution forecasting using conditional mean and conditional quantiles with functional covariates. In *Statistical methods for biostatistics and related fields*, pp. 221–243. New York: Springer.

Cardot, H., F. Ferraty, and P. Sarda (2003). Spline estimators for the functional linear model. *Statistica Sinica 13*(3), 571–591.

Che, M., L. Kong, R. C. Bell, and Y. Yuan (2017). Trajectory modeling of gestational weight: a functional principal component analysis approach. *PloS one 12*(10), e0186761.

Chen, K. and H.-G. Müller (2012). Conditional quantile analysis when covariates are functions, with application to growth data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74*(1), 67–89.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika 96*(3), 559–575.

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

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

Kai, B., R. Li, and H. Zou (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Annals of statistics 39*(1), 305–332.

Kato, K. (2012). Estimation in functional linear quantile regression. *The Annals of Statistics 40*(6), 3108–3136.

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

Koenker, R. and O. Geling (2001). Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association 96*(454), 458–468.

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

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

Liu, Y. and Y. Wu (2011). Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. *Journal of Nonparametric Statistics 23*(2), 415–437.

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

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

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

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

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

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

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (Second ed.). New York: Springer.

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

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

Soliman, A., V. De Sanctis, R. Elalaily, and S. Bedair (2014). Advances in pubertal growth and factors influencing it: Can we increase pubertal growth? *Indian Journal of Endocrinology and Metabolism 18*(Suppl 1), S53–S62.

Stone, C. J., M. Hansen, C. Kooperberg, and Y. K. Truong (1997). Polynomial splines and their tensor products in extended linear modeling (with discussion). *The Annals of Statistics 25*(4), 1371–1470.

Tang, Q. and L. Kong (2017). Quantile regression in functional linear semiparametric model. *Statistics 51*(6), 1342–1358.

Tuddenham, R. and M. Snyder (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., L. Kong, B. Jiang, X. Zhou, S. Yu, L. Zhang, and G. Heo (2019). Wavelet-based lasso in functional linear quantile regression. *Journal of Statistical Computation and Simulation 89*(6), 1111–1130.

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

Wu, Y., Y. Ma, and G. Yin (2015). Smoothed and corrected score approach to censored quantile regression with measurement errors. *Journal of the American Statistical Association 110*(512), 1670–1683.

Yao, F., H.-G. Müller, and J.-L. Wang (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association 100*(470), 577–590.

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

Yu, S., G. Wang, L. Wang, C. Liu, and L. Yang (2020). Estimation and inference for generalized geoadditive models. *Journal of the American Statistical Association 115*(530), 761–774.

Yu, S., G. Wang, L. Wang, and L. Yang (2021). Multivariate spline estimation and inference for image-on-scalar regression. *Statistica Sinica 31*(3), 1–39.

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

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

Zhao, Z. and Z. Xiao (2014). Efficient regressions via optimally combining quantile information. *Econometric Theory 30*(6), 1272–1314.

Zou, H. and M. Yuan (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics 36*(3), 1108–1126.

Boyi Hu and Jiguo Cao

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

E-mails: boyih@sfu.ca and jiguo_cao@sfu.ca

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

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

Hua Liu, School of Economics and Finance, Xi'an Jiaotong University, Xi'an 710061, Shaanxi, 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