

Inference in Functional Linear Quantile Regression

Meng Li^{a,*}, Kehui Wang^b, Arnab Maity^c, Ana-Maria Staicu^c

^a*Department of Statistics, Rice University, Houston, TX*

^b*PPD Inc, Morrisville, NC*

^c*Department of Statistics, North Carolina State University, Raleigh, NC*

Abstract

In this paper, we study statistical inference in functional quantile regression for scalar response and a functional covariate. Specifically, we consider a functional linear quantile regression model where the effect of the covariate on the quantile of the response is modeled through the inner product between the functional covariate and an unknown smooth regression parameter function that varies with the level of quantile. The objective is to test that the regression parameter is constant across several quantile levels of interest. The parameter function is estimated by combining ideas from functional principal component analysis and quantile regression. An adjusted Wald testing procedure is proposed for this hypothesis of interest, and its chi-square asymptotic null distribution is derived. The testing procedure is investigated numerically in simulations involving sparse and noisy functional covariates and in a capital bike share data application. The proposed approach is easy to implement and the R code is published online at <https://github.com/xylimeng/fQR-testing>.

Keywords: Composite quantile regression, Functional principal component analysis, Functional quantile regression, Measurement error, Wald test.

2020 MSC: Primary 62G08, Secondary 62H15

1. Introduction

The advance in computation and technology generated an explosion of data that have functional characteristics. The need to analyze this type of data triggered a rapid growth of the functional data analysis (FDA) field; see [9, 35] for two comprehensive treatments. Most research in functional data analysis has primarily focused on mean regression (see, for example, [10, 19, 20, 41, 46]); only a few works accommodate higher-order moment effects [29, 38]. Quantile regression is appealing in many applications as it allows us to describe the entire conditional distribution of the response at various quantile levels. For example, in our capital bike share data application, it is of interest to study how the bike rental behavior of casual users in the previous day affects the upper quantiles of total bike rentals in the current day.

Quantile regression models for scalar responses and functional covariates have been introduced in [2]. Functional quantile regression (fQR) models essentially extend the standard quantile regression framework to account for functional covariates: the effect of the covariate on a particular quantile of the response is modeled through the inner product between the functional covariate and an unknown smooth regression parameter function that varies with the level of quantile. Cardot et al. [2] considered a smoothing splines-based approach to represent the functional covariates and derived its convergence rate; Kato [22] studied principal component analysis (PCA)-based estimation and established a sharp convergence rate. In [5] and [4] Crambes et al. discussed nonparametric quantile regression estimation and studied the theoretical properties of a support vector machine-based estimator, a method inspired from [30]. Yao et al. [47] considered regularized partially fQR model by additionally incorporating high-dimensional scalar covariates. Shi et al. [37] developed a procedure to test the adequacy of fQR based on functional PCA. Unlike these

*Corresponding author. Email address: meng@rice.edu

regularization and basis expansion-based methods, Ferraty et al. [8] and Chen and Müller [3] estimated the conditional quantile function by inverting the corresponding conditional distribution function; they too studied consistency properties of the regression estimator. Nevertheless, hitherto there is no available work on statistical inference of the quantile regression estimator under fQR. Additionally, existing functional quantile regression research often assumes that the functional covariate is observed either completely on the domain or at very dense grids of points and typically with little or no error contamination. In this work, we are interested in formally assessing whether the effect of the true smooth signal of the covariate, varies across several quantile levels of interest of the response, when the smooth signal is observed at finite grids and possibly perturbed with error and a functional linear quantile regression model is assumed. This problem is important in its own right, yielding a more comprehensive description of the relationship between the covariate and the conditional distribution of the response. Furthermore, formally assessing such a hypothesis is critical when one wishes to improve the estimation accuracy of the conditional quantile of the response at some specified level. Specifically, suppose for several quantile levels around the specified level of interest, there is no evidence that the effect of the latent covariate on these quantile levels of the response differs. In that case, one can improve the accuracy in estimating the covariate effect on the response at the specified level of interest by borrowing information across these quantile levels. For example, in the case of standard quantile regression with vector predictors, there has been a rich literature on the so-called composite quantile regression to aggregate information across quantile levels [21, 24, 49, 52].

In this paper, we assume a linear fQR model that relates the τ th quantile level of the response to the covariate through the inner product between a bivariate regression coefficient function and the true covariate signal. In the case when the true signal is measured at the same time points across the study, one naive way to test the null hypothesis that the effect of the true covariate signal is constant across several quantile levels of interest, is to treat the discretely observed functional covariates as high dimensional covariate and apply standard testing procedures (Wald test) in linear quantile regression for vector covariates [25]. As expected, such an approach results in inflated type I error rates due to the high correlation between the repeated measurements corresponding to the same subject; the situation gets progressively worse when the covariate includes noise. Another alternative is to consider a single number summary of the covariate, such as average or median, and carry out this hypothesis testing by employing standard testing methods in quantile regression. Our numerical investigation of this direction shows that while the Type I error rates are preserved well, the power is substantially affected.

We propose to represent the latent smooth covariate and the quantile regression parameter function using the same orthogonal basis system; this reduces the inner product part of the linear fQR model to an infinite sum of products of basis coefficients of the smooth covariate and parameter function. There are various options of orthogonal basis types: we consider the data-driven basis that is formed by the leading eigenbasis functions of the covariance of the true covariate signal and use the percentage of variance explained criterion to determine a finite truncation for this basis. While using a finite basis system reduces the dimensionality of the problem, an important challenge is handling the variability of the basis coefficients of the smooth latent signal, called functional principal component (fPC) scores. We develop the asymptotic distributions of the quantile estimators based on the estimated fPC scores, when the functional covariate is sampled at a fine grid of points (dense design). Finally, we introduce an adjusted Wald test statistic and develop its asymptotic null distribution. The introduced testing procedure shows excellent numerical results even in situations when the functional covariate is sampled at few and irregular time points across the study (sparse design) and the measurements are contaminated by error.

The theoretical study of the distribution of the quantile estimator based on the estimated fPC scores has important differences from the standard linear quantile regression with vector covariates. First, the predictors, fPC scores, are unknown and require estimation, which in turn introduces uncertainty; by comparison the vector covariates are known in the standard quantile setting counterpart. We show that asymptotically the quantile estimators are still unbiased, but their variances are inflated. This implies that, in this reduced framework, a direct application of the Wald testing procedure for null hypotheses involving regression parameters is not appropriate. Second, dealing with estimated fPC scores in this situation is different from the measurement errors in predictors setting. For the latter, it is typically assumed that the measurement error and the true predictors are mutually independent or that the errors are independent across subjects [42, 44, 45]. However, in the functional data setting the resulting errors, due to the difference between the estimated fPC scores and the true scores, are dependent on the true predictors and are also dependent across subjects. As a result, the theoretical investigation requires more careful quantification in terms of the estimated scores and the use of quantile loss.

This article makes three main contributions. First, we establish the asymptotic distribution of the coefficient estimator for both one single quantile level and multiple quantile levels for dense sampled functional covariates. To the best of our knowledge, this is the first work that studies inference of the quantile estimators; previous research in functional quantile regression focused on consistency and minimax rates (see [3, 22]), and most literature on inference in FDA is limited to the context of functional linear regression [1, 16, 26, 40, 48]. Second, we propose an adjusted Wald test statistic to formally assess that the quantile regression parameter is constant across specified quantile levels and derive its asymptotic null distribution. Third, we consider cases where the functional covariate is observed sparsely and contaminated with noise and illustrate through detailed numerical investigation that the testing procedure continues to have excellent performance. Furthermore, we demonstrate the usage of the composite quantile regression and the corresponding advantage in terms of estimation and prediction accuracy, using a capital bike rental data set. Composite quantile regression is well known to improve the efficiency of the quantile estimators at a single quantile level, which becomes especially useful for extreme quantiles [43]; nonetheless, more formal investigation of functional composite quantile regression is beyond the scope of this article.

The rest of the paper is organized as follows. Section 2 introduces the statistical framework, describes the null and alternative hypotheses, discusses a simpler approximation of the testing procedure, and presents the estimation approach. Section 3 develops the asymptotic normality of the proposed estimators, introduces the adjusted Wald test, and derives its null asymptotic distribution. Section 4 presents extensive simulation studies confirming the excellent performance of the proposed test procedure in various scenarios for both dense and sparse designs. Section 5 applies the proposed test to a bike rental data and illustrates the improvement of combined quantile regressions compared to a single level quantile regression after the proposed tests being used. Proofs of Theorem 1 and Theorem 2, as well as some additional useful results, are given in Section 6.

2. Methodology

2.1. Statistical framework

Suppose we observe data $\{Y_i, (t_{ij}, W_{ij})\}$ for $j \in \{1, \dots, m_i\}$ and $i \in \{1, \dots, n\}$, where Y_i is a scalar response variable, $\{W_{i1}, \dots, W_{im_i}\}$ is the evaluation of a latent and smooth process $X_i(\cdot)$ measured with noise at the finite grid of points $\{t_{i1}, \dots, t_{im_i}\}$ for $t_{ij} \in \mathcal{T}$, and \mathcal{T} is a bounded closed interval. It is assumed that the observed functional covariate is perturbed by white noise, i.e., $W_{ij} = X_i(t_{ij}) + e_{ij}$, where e_{ij} has mean 0 and variance σ^2 . Furthermore, we assume that the true functional signal $X_i(\cdot) \in L^2(\mathcal{T})$ with $\mathcal{T} = [0, 1]$, and $X_i(\cdot)$ are independent and identically distributed. Our objective is to formally assess whether the smooth covariate signal $X_i(\cdot)$ has constant effect at specified quantile levels of the response.

Let $Q_{Y_i|X_i}(\tau)$ be the conditional τ th quantile function of the response Y_i given the true covariate signal $X_i(\cdot)$ where $\tau \in (0, 1)$. We assume the following linear fQR model:

$$Q_{Y_i|X_i}(\tau) = \beta_0(\tau) + \int_0^1 \beta(t, \tau) X_i^c(t) dt, \quad (1)$$

where $\beta_0(\tau)$ is the quantile-level varying intercept function, and $\beta(t, \tau)$ is the bivariate regression coefficient function and the main object of interest. It is assumed that for a fixed quantile level τ , $\beta(t, \tau) \in L^2[0, 1]$ as a function of t . Here $X_i^c(t)$ is the de-meaned smooth covariate signal, defined as $X_i^c(t) = X_i(t) - EX_i(t)$. Model (1) is an extension of the standard linear quantile regression model [25] to functional covariates. It was first introduced by [2] and later considered by [3, 22]. For simplicity, in the following it is assumed that the smooth covariate signal has zero mean, i.e., $EX_i(t) = 0$ for all $t \in [0, 1]$.

Let $\mathcal{U} = \{\tau_1, \dots, \tau_L\}$ be a set with quantile levels of interest where $\tau_1 < \dots < \tau_L$. Motivated by the reasons mentioned in Section 1, our goal is to test the null hypothesis:

$$H_0 : \beta(\cdot, \tau_1) = \dots = \beta(\cdot, \tau_L), \quad (2)$$

against the alternative hypothesis $H_a : \beta(\cdot, \tau_\ell) \neq \beta(\cdot, \tau_{\ell'})$, for some $\ell \neq \ell' \in \{1, \dots, L\}$. This null hypothesis involves infinite dimensional objects, which is very different from the common null hypotheses considered in quantile regression.

One approach to simplify the null hypothesis is by using basis functions expansion. Specifically, let $\{\phi_k(\cdot)\}_{k \geq 1}$ be an orthogonal basis in $L^2[0, 1]$ such that $\int_0^1 \phi_k(t)\phi_{k'}(t)dt = 0$ if $k \neq k'$ and 1 if $k = k'$. We represent the unknown parameter function $\beta(\cdot, \tau)$ using this orthogonal basis $\beta(t, \tau) = \sum_{k \geq 1} \beta_k(\tau)\phi_k(t)$ where $\beta_k(\tau) = \int \beta(t, \tau)\phi_k(t)dt$ are unknown parameter loadings varying with the quantile level τ . It follows that the equality $\beta(\cdot, \tau_\ell) = \beta(\cdot, \tau_{\ell'})$ is equivalent to $\beta_k(\tau_\ell) = \beta_k(\tau_{\ell'})$, $k \geq 1$. Thus, the null hypothesis (2) can be written as $H_0 : \beta_k(\tau_1) = \beta_k(\tau_2) = \dots = \beta_k(\tau_L)$ for $k \geq 1$. Furthermore, we represent the smooth covariate using the same basis function as $X_i(t) = \sum_{k \geq 1} \xi_{ik}\phi_k(t)$ where $\xi_{ik} = \int X_i(t)\phi_k(t)dt$ are smooth covariate loadings. Then, the linear fQR model (1) can be equivalently represented as $Q_{Y_i|X_i}(\tau) = \beta_0(\tau) + \sum_{k=1}^{\infty} \beta_k(\tau)\xi_{ik}$. In practice the infinite summation is typically truncated to some finite truncation K . As a result the fQR model can be approximated by

$$Q_{Y_i|X_i}^K(\tau) = \beta_0(\tau) + \sum_{k=1}^K \beta_k(\tau)\xi_{ik}, \quad (3)$$

and the null hypothesis to be tested can be approximated by a reduced version

$$H_0^K : \begin{pmatrix} \beta_1(\tau_1) \\ \beta_2(\tau_1) \\ \vdots \\ \beta_K(\tau_1) \end{pmatrix} = \begin{pmatrix} \beta_1(\tau_2) \\ \beta_2(\tau_2) \\ \vdots \\ \beta_K(\tau_2) \end{pmatrix} = \dots = \begin{pmatrix} \beta_1(\tau_L) \\ \beta_2(\tau_L) \\ \vdots \\ \beta_K(\tau_L) \end{pmatrix}. \quad (4)$$

Let $\theta_\tau := (\beta_0(\tau), \beta_1(\tau), \dots, \beta_K(\tau))^T$ be the $(K+1)$ -dimensional parameter vector and $\zeta := (\theta_{\tau_1}^T, \dots, \theta_{\tau_L}^T)^T$ the full quantile regression parameter vector of dimension $L(K+1)$. Then the reduced null hypothesis (4) can be equivalently re-written as $H_0^K : R\zeta = 0$, where $R = R_1 \otimes R_2$ and

$$R_1 = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix}_{(L-1) \times L}, \quad R_2 = [\mathbf{0}_K, I_K]_{K \times (K+1)}.$$

Here $\mathbf{0}_K$ denotes the K -dimensional vector of zeros and I_K is the $K \times K$ dimensional identity matrix.

If the loadings ξ_{ik} 's were known, then model (3) is exactly the conventional quantile regression model. In such case, a standard Wald testing procedure for H_0^K is typically formulated as $T_W = (\widehat{R\zeta})^T (\widehat{R\Gamma_\zeta} R^T)^{-1} \widehat{R\zeta}$, where $\widehat{\zeta}$ is the quantile regression estimator of ζ and $\widehat{\Gamma_\zeta}$ is a consistent estimator of the covariance of $\widehat{\zeta}$ conditional on the true loadings ξ_{ik} 's; see [25, Chapter 3] for a review of existing methods. However, in practice the loadings of the smooth covariate signal ξ_{ik} are unknown, and a valid approach has to account for such uncertainty.

Depending on the choice of the orthogonal basis, the approaches used to select the finite truncation K and to develop the theoretical properties for the quantile regression estimators differ. Several choices have been commonly used in functional data analysis literature: Fourier basis functions [39], Wavelet basis [33] or orthogonal B-splines [36, 50]. One important aspect to keep in mind when selecting the basis functions is how to handle the finite truncation K . In this paper we consider the orthogonal basis given by the eigenfunctions of the covariance of the smooth covariate signal $X_i(\cdot)$. Let $G(s, t) := \text{Cov}\{X_i(s), X_i(t)\}$ be the covariance of $X_i(\cdot)$; Mercer's theorem gives the following spectral decomposition of the covariance $G(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t)$, where $\{\phi_k(\cdot), \lambda_k\}_k$ are the pairs of eigenfunctions and corresponding eigenvalues. The eigenvalues λ_k 's are nondecreasing and nonnegative and the eigenfunctions $\phi_k(\cdot)$'s are mutually orthogonal functions in $L^2[0, 1]$. Using the Karhunen-Loève expansion, the zero-mean smooth covariate $X_i(\cdot)$ can be represented as $X_i(t) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t)$, where $\xi_{ik} = \int_0^1 X_i(t)\phi_k(t)dt$ are commonly known as functional principal component (fPC) scores of $X_i(\cdot)$, satisfying that $E(\xi_{ik}) = 0$, $\text{Var}(\xi_{ik}) = \lambda_k$ and uncorrelated over k . A popular way to select the finite truncation, or equivalently the number of leading eigenfunctions, is the percentage of variance explained; alternative options for selecting the finite truncation K are considered in [32] and [28].

2.2. Estimation procedure

We discuss estimation for the case when the functional covariate is observed on a fine grid of points, a setting known in the literature by the name of dense sampling design. Nevertheless, our procedure can be successfully applied

to the case when the covariate is observed on an irregular sampling design with few points (sparse sampling design) and contaminated with noise, as illustrated later in the numerical investigation. When the sampling design is dense, and thus m_i is very large for each i , a common approach in functional data analysis is “smoothing first, then estimation” [48]. Specifically, we first reconstruct each trajectory $\widehat{X}_i(\cdot)$ from the data $(t_{ij}, W_{ij})_{j=1}^{m_i}$ using penalized regression splines, while one can also use any other appropriate smoothing method such as the local polynomial kernel smoothing technique [6]. Let $\bar{X}(\cdot)$ be the sample mean of these reconstructed trajectories and denote by $\widehat{X}_i^c(t) := \widehat{X}_i(t) - \bar{X}(t)$ the centered covariate. Furthermore, let $S(\cdot, \cdot)$ be the sample covariance of $\widehat{X}_i(t)$; the spectral decomposition of $S(\cdot, \cdot)$ yields the pairs of estimated eigenfunctions and eigenvalues $\{\widehat{\phi}_k(\cdot), \widehat{\lambda}_k\}_k$. The theoretical properties of the estimated eigenfunctions $\widehat{\phi}_k(\cdot)$ have been well studied in the literature; see [11, 13, 48] among others. As eigenfunctions $\phi_k(\cdot)$ and $\widehat{\phi}_k(\cdot)$ are both defined up to a change in sign, we assume that the sign of $\widehat{\phi}_k(\cdot)$ is chosen such that $\int \phi_k(t)\widehat{\phi}_k(t)dt \geq 0$ throughout the paper. Finally the fPC scores ξ_{ik} are estimated as $\widehat{\xi}_{ik} = \int \widehat{X}_i^c(t)\widehat{\phi}_k(t)dt$; in practice numerical integration is used to approximate the integral; see also [31].

Using the estimated fPC scores $\widehat{\xi}_{ik}$ ’s, the quantile regression parameter of the approximated linear fQR model, θ_τ , is estimated by

$$\widehat{\theta}_\tau = \arg \min_{(b_0, b_1, \dots, b_K)^T \in \mathbb{R}^{K+1}} \sum_{i=1}^n \rho_\tau \left(y_i - b_0 - \sum_{k=1}^K \widehat{\xi}_{ik} b_k \right), \quad (5)$$

where $\rho_\tau(x) = x\{\tau - I(x < 0)\}$ is the quantile loss function and $I(x < 0)$ is the indicator function that equals 1 if $x < 0$ and 0 otherwise. Although throughout this article we focus on a homogeneous truncation level K to ease presentation, the proposed method easily generalizes to the case in which K varies with τ . We next move on to studying the theoretical properties of the quantile regression estimator in (5).

3. Theoretical properties

3.1. Assumptions

Let $F_i(y) = P(Y_i < y | X_i(\cdot))$, and $f_i(\cdot)$ be the corresponding density function. We make the following assumptions:

- A1. $\{Y_i, X_i(\cdot), e_i(\cdot)\}_{i=1}^n$ are independent and identically distributed (i.i.d.) as $\{Y, X(\cdot), e(\cdot)\}$, and $X(\cdot)$ and $e(\cdot)$ are independent where $E\{e(t)\} = 0$ and $\text{Cov}\{e(t), e(t')\} = \sigma^2 I(t = t')$ for any t, t' ;
- A2. The conditional distribution $F_i(\cdot)$ is twice continuously differentiable and the corresponding density function $f_i(\cdot)$ is uniformly bounded away from 0 and ∞ at points $Q_{Y_i|X_i}(\tau)$;
- A3. The functional covariates $X(\cdot)$ satisfy that $E\{X(t_1)X(t_2)X(t_3)X(t_4)\} < \infty$ uniformly for $(t_1, t_2, t_3, t_4) \in [0, 1]^4$;
- A4. There exists a finite number p_0 such that $\lambda_1 > \lambda_2 > \dots > \lambda_{p_0} > 0$ and $\lambda_k = 0$ if $k > p_0$.

A2 and the i.i.d. assumption in A1 are standard in quantile regression with vector covariates; see [25, Ch. 4]. A1 assumes that the functional covariates $X_i(\cdot)$ ’s are observed with independent white noise $e_i(\cdot)$, making the model more realistic compared to error free assumptions made by [22]. The assumption A3 holds for Gaussian processes and is common in the FDA literature; for example, see [13] and the discussion therein.

Finally A4 requires that the functional covariate has a finite number of non-zero eigenvalues, making the approximate model (3) exact, with $K = p_0$. This strong assumption has been employed previously in the literature [31, 32]. In numerical studies, we found that A4 is not needed in order for the testing procedure to show excellent performance in terms of size and power; see, for example, the simulation study in Section 4.4 under the more general model (1) when p_0 is divergent. This seems to indicate that A4 is for theoretical convenience. One possible way to relax this assumption is to replace it with a condition on the number of principal components that are relevant in describing the dependence between the functional covariate and the response. Another possibility is to remove it entirely and show that the functional quantile regression $Q_{Y_i|X_i}^K(\tau)$ approximates the original model with negligible error. Nonetheless, our attempts to prove the main results by relaxing A4 in these directions have not been productive, partly due on one hand to the complication in the interweave of the quantile loss function and infinitely dimensional functional data and on the other hand to the focus on hypothesis testing, as opposed to estimation. Specifically, A4 is critical to ensure a root- n rate for the estimated coefficient functions formulated in Theorem 1 and subsequently to derive the test’s null distribution. As noted in the preceding section, even under A4, inference on fQR based on the estimated fPC scores differs from the standard multivariate quantile regression with vector covariates in the key aspect that estimation of

the fPC scores induces a specific type of measurement error. Unlike the existing measurement error in covariates literature relying on certain independence assumptions [42, 44, 45], measurement errors in the estimated fPC scores are dependent on the true predictors and are also dependent across subjects. This consideration requires a more careful quantification in terms of the estimated scores and the use of quantile loss. In this article, we focus on addressing the challenge posed by measurement error in quantile regression with intricate dependence induced by functional covariates, and leave developments to relax A4 to future research.

The following assumptions are commonly used when describing a dense sampling design [31, 48]. For convenient mathematical derivations, we assume that there are the same number of observations per subject, i.e., $m_i = m$ for all i .

B1. The time points $t_{ij} \stackrel{i.i.d.}{\sim} g(\cdot)$ for $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$, where the density $g(\cdot)$ has bounded support $[0, 1]$ and is continuously differentiable.

B2. $m \geq Cn^{c_m}$ where $c_m > 5/4$ and C is some constant.

For our theoretical development, we require the following condition for the kernel bandwidth h_X that is used in smoothing the functional covariates.

C1. $h_X = O(n^{-c_m/5})$.

3.2. Asymptotic distribution

The following theorem gives the asymptotic distribution of the quantile estimator. Kato [22] gave the minimax rate of the coefficient function estimation when there is no measurement error on the discrete functional covariates. The author assumed that the number of eigenvalues is infinite instead of finite as in our assumption A4. Our established root- n rate crucially depends on A4, which facilitates downstream inference. One would need to properly scale the estimator using a slower rate and derive the asymptotic distribution, for both $\widehat{\theta}_\tau$ and test statistics constructed via $\widehat{\theta}_\tau$, should A4 be relaxed. We denote D_0 as the diagonal matrix whose diagonal entries are $(1, \lambda_1, \dots, \lambda_{p_0})$ and $D_1(\tau) = E[f_i\{Q_{Y_i|X_i}(\tau)\}\xi_i\xi_i^T]$ which is positive definite, where $\xi_i = (1, \xi_{i1}, \dots, \xi_{ip_0})^T$. Similarly, we denote $\widehat{\xi}_i = (1, \widehat{\xi}_{i1}, \dots, \widehat{\xi}_{ip_0})^T$. When $K = p_0$, the hypothesis H_0^K in (4) is equivalent to H_0 in (2), and the truncation model in (3) does not incur approximation error as the residual $\sum_{k>K} \beta_k(\tau)\xi_{ik}$ degenerates to zero owing to its zero variance.

Theorem 1. Denote by $\widehat{\theta}_\tau$ the quantile regression estimator defined by (5) for $K = p_0$, where $\tau \in (0, 1)$. Under Conditions A1–A4, B1–B2, C1, we have

$$\sqrt{n}(\widehat{\theta}_\tau - \theta_\tau) \xrightarrow{d} N\{0, \tau(1 - \tau)D_1^{-1}(\tau)D_0D_1^{-1}(\tau) + \Theta_\tau\Sigma_0\Theta_\tau\}, \quad (6)$$

where $\Theta_\tau = 1_{(p_0+1) \times (p_0+1)} \otimes \theta_\tau^T$ and the matrix Σ_0 is defined in Section 6 which does not depend on τ . Moreover, $\widehat{\zeta} = (\widehat{\theta}_{\tau_1}^T, \dots, \widehat{\theta}_{\tau_L}^T)^T$ is asymptotically multivariate normal centered at $\zeta = (\theta_{\tau_1}^T, \dots, \theta_{\tau_L}^T)^T$, and for $1 \leq \ell \neq \ell' \leq L$ the asymptotic covariance matrix for $\widehat{\theta}_{\tau_\ell}$ and $\widehat{\theta}_{\tau_{\ell'}}$ is given by

$$Acov\{\sqrt{n}(\widehat{\theta}_{\tau_\ell} - \theta_{\tau_\ell}), \sqrt{n}(\widehat{\theta}_{\tau_{\ell'}} - \theta_{\tau_{\ell'}})\} = \{\min(\tau_\ell, \tau_{\ell'}) - \tau_\ell\tau_{\ell'}\}D_1^{-1}(\tau_\ell)D_0D_1^{-1}(\tau_{\ell'}) + \Theta_{\tau_\ell}\Sigma_0\Theta_{\tau_{\ell'}}. \quad (7)$$

Remark that the asymptotic covariances in both (6) and (7) contain two components: a Huber [18] sandwich term that is typical in quantile regression theory and a “variance inflation” term. Specifically, if the true scores ξ_i ’s were observed, then the asymptotic variance of $\widehat{\theta}_\tau$ would be $\tau(1 - \tau)D_1^{-1}(\tau)D_0D_1^{-1}(\tau)$, and the asymptotic covariance matrix for $\widehat{\theta}_{\tau_\ell}, \widehat{\theta}_{\tau_{\ell'}}$ would be $\{\min(\tau_\ell, \tau_{\ell'}) - \tau_\ell\tau_{\ell'}\}D_1^{-1}(\tau_\ell)D_0D_1^{-1}(\tau_{\ell'})$; see [25, 34]. The variance inflation terms, $\Theta_\tau\Sigma_0\Theta_\tau$ in (6) and $\Theta_{\tau_\ell}\Sigma_0\Theta_{\tau_{\ell'}}$ in (7), quantify the effect of uncertainty in estimating the fPC scores on the quantile regression estimators. Thus, when the covariates are functional data, the asymptotic distribution of $\widehat{\theta}_\tau$ is unbiased but the variance is inflated where the variance inflation terms depend on the true parameter value θ_τ .

The proof of Theorem 1 is detailed in Section 6. The reasoning follows two main steps: 1) approximate the estimated fPC scores $\widehat{\xi}_i$ ’s by linear combinations of random vectors of the true fPC scores ξ_i ; and 2) show that the approximation error in the predictors is negligible to the quantile loss function. Step 1 crucially relies on the dense design assumption B2. This allows to employ various bounds on both the estimated eigenfunctions and the difference $\widehat{X}_i(\cdot) - X_i(\cdot)$, which in turn enables us to derive a fine-grained characterization of the estimated scores (Lemma 1); see the supplementary materials for more detail.

3.3. Adjusted Wald test

Using the asymptotic properties of the quantile regression estimators, we are now ready to develop a Wald type testing procedure for assessing the general null hypothesis (2) or its finite reduced version (4) represented in vector form by $H_0^K : R\zeta = 0$. Recall that $\zeta = (\theta_{\tau_1}^T, \dots, \theta_{\tau_L}^T)^T$ denotes the full quantile regression parameter, and $\hat{\zeta} = (\hat{\theta}_{\tau_1}^T, \dots, \hat{\theta}_{\tau_L}^T)^T$ is its estimator.

We define a modified version of Wald test, called the *adjusted Wald test*, by ignoring the variance inflation terms in the above asymptotic covariances. Let $\Sigma(\tau_\ell, \tau_{\ell'}) = \sigma(\tau_\ell, \tau_{\ell'})D_1^{-1}(\tau_\ell)D_0D_1^{-1}(\tau_{\ell'})$ with $\sigma(\tau_\ell, \tau_{\ell'})$ set to $\tau_\ell(1 - \tau_\ell)$ if $\ell = \ell'$, and $\{\min(\tau_\ell, \tau_{\ell'}) - \tau_\ell\tau_{\ell'}\}$ otherwise. Then the asymptotic covariance matrix of $\hat{\zeta}$ without the inflation terms, denoted as $\Gamma_{\hat{\zeta}}^a$ in which the superscript a indicates the adjustment by ignoring the inflation terms, i.e.,

$$\Gamma_{\hat{\zeta}}^a = \begin{matrix} L(K+1) \times L(K+1) \end{matrix} = \begin{bmatrix} \Sigma(\tau_1, \tau_1) & \Sigma(\tau_1, \tau_2) & \cdots & \Sigma(\tau_1, \tau_L) \\ \Sigma(\tau_2, \tau_1) & \Sigma(\tau_2, \tau_2) & \cdots & \Sigma(\tau_2, \tau_L) \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma(\tau_L, \tau_1) & \Sigma(\tau_L, \tau_2) & \cdots & \Sigma(\tau_L, \tau_L) \end{bmatrix}. \quad (8)$$

Let $\hat{\Gamma}_{\hat{\zeta}}^a$ be a consistent estimator of $\Gamma_{\hat{\zeta}}^a$ constructed similarly to (8) but with a consistent estimator $\hat{\Sigma}(\tau_\ell, \tau_{\ell'})$ of $\Sigma(\tau_\ell, \tau_{\ell'})$. The adjusted Wald test is given by

$$T_n = n(\hat{R}\hat{\zeta})^T (\hat{R}\hat{\Gamma}_{\hat{\zeta}}^a \hat{R}^T)^{-1} \hat{R}\hat{\zeta}. \quad (9)$$

This test is not a proper Wald test as the covariance matrix used is not the valid covariance of $\hat{\zeta}$. The following result studies the asymptotic null distribution of T_n assuming $K = p_0$.

Theorem 2. Assume the regularity conditions A1–A4, B1–B2 and C1 hold. If the null hypothesis is true, $R\zeta = 0$, then the asymptotic distribution of T_n is χ_K^2 .

The proof of this result relies on the observation that if $\Gamma_{\hat{\zeta}}$ is the proper covariance of $\hat{\zeta}$ as described by Theorem 1, then $R(\Gamma_{\hat{\zeta}}^a - \Gamma_{\hat{\zeta}})R^T = 0$. Intuitively, this is because the inflation terms in (6) and (7) possess a sandwich structure with a constant matrix enclosed by Θ_τ , which is zeroed out if left multiplied by R under the null hypothesis that $R\zeta = 0$. Thus, although the estimation of the fPC scores yields inflated covariance of the regression estimator, its effect on testing the null hypothesis (2) is negligible. Nevertheless, if one is interested in testing a different type of null hypothesis for ζ , such as nonlinear functionals, then this variance inflated term has to be taken into account for a proper testing procedure.

We construct $\hat{\Sigma}(\tau_\ell, \tau_{\ell'})$ for $1 \leq \ell, \ell' \leq L$ by a plug-in estimator that uses $\hat{D}_0 = \sum_{i=1}^n \hat{\xi}_i \hat{\xi}_i^T / n$ and $\hat{D}_1(\tau) = \sum_{i=1}^n \hat{f}_i(\xi_i^T \theta_\tau) \hat{\xi}_i \hat{\xi}_i^T / n$ to estimate D_0 and $D_1(\tau)$, respectively. The consistency of these estimators can be proved by law of large numbers-based arguments together with Lemma 1 that discusses the closeness between $\hat{\xi}_i$ and ξ_i . For the estimation of $\hat{f}_i(\xi_i^T \theta_\tau)$ in $D_1(\tau)$, we use the difference quotient method proposed by [15] and substitute the estimates $\hat{\xi}_i$ and $\hat{\theta}_\tau$. Theorem 2 implies that, for testing the null hypothesis of equal functional covariate effect across various quantile levels, the common Wald test based on the estimated fPC scores provides a valid testing procedure. The adjusted Wald test, that disregards the variance component due to the estimation uncertainty of the fPC scores, has a chi-square asymptotic null distribution.

If the number of principal components p_0 is replaced by its consistent estimator K_n , then the null distribution of the test statistic T_n is approximately $\chi_{K_n}^2$ for large n . In other words, the difference in the respective cumulative distribution functions, $P(T_n \leq t) - P(\chi_{K_n}^2 \leq t)$, goes to zero for any $t \in \mathbb{R}$; it implies that the critical value of T_n is asymptotically the same as that of $\chi_{K_n}^2$. Functional data analysis literature provides a rich menu of possibilities for selecting p_0 , such as the percentage of variance explained (PVE) criterion and a Bayesian information criterion (BIC) proposed by [32]. The BIC method is proved to be consistent for both sparse and dense functional data. The PVE criterion is defined as

$$K_n = \min \left\{ p : \sum_{i=1}^p \hat{\lambda}_i / \sum_{i=1}^q \hat{\lambda}_i \geq \text{PVE} \right\},$$

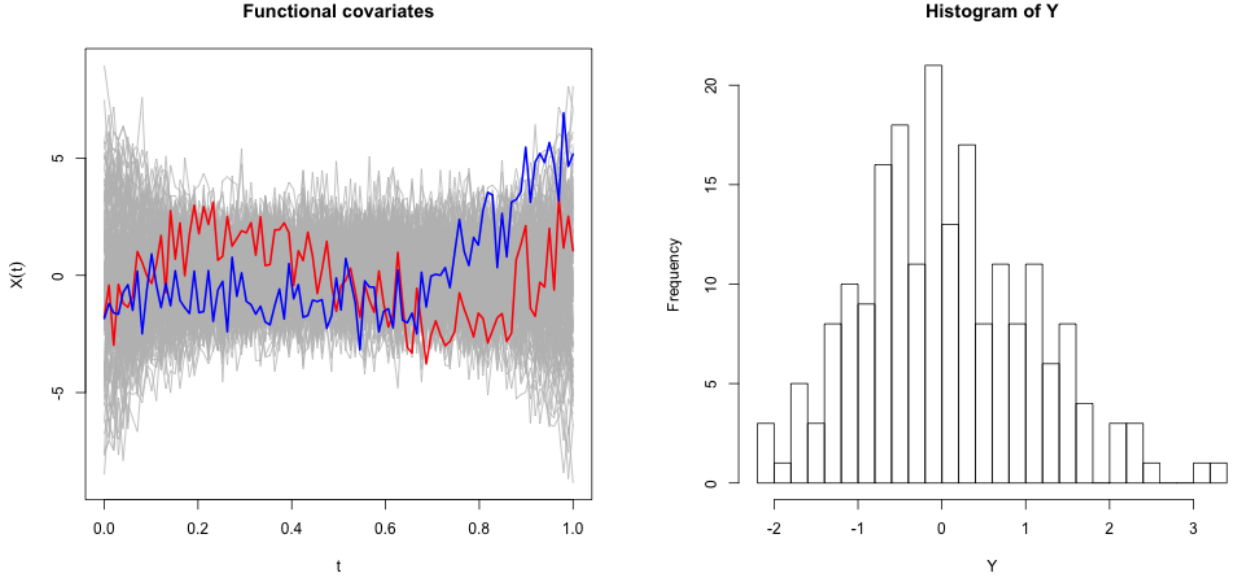


Fig. 1: Simulated data when $n = 200$ and $\gamma = 1$. The left panel plots the functional covariates and two randomly selected curves are highlighted in blue and red; the right panel is the histogram of the response.

where q is the number of estimated eigenvalues and PVE some user-defined threshold that approaches one. The widely used PVE approach also leads to consistent estimators of p_0 given that the number of estimated eigenvalues is greater than p_0 and eigenvalues are estimated consistently, which may well be true in many applications and particularly as suggested by our extensive simulation studies. We shall use the method of PVE in the remaining sections, and we have found that it leads to accurate estimate of p_0 in finite sample performance.

We would like to point out that the asymptotic power of the adjusted Wald test is obtainable using a non-central chi-square distribution. However, the expression is complicated without involving stronger assumptions on $X_i(\cdot)$'s, since the equation $R(\Gamma_{\hat{\zeta}}^a - \Gamma_{\hat{\zeta}})R^T = 0$ does not generally hold under the alternative thus a Wald-type test requires an estimate of the matrix Σ_0 .

4. Simulation

4.1. Settings

The simulated data is of the form $\{Y_i, (t_{ij}, W_{ij})_{j=1}^{m_i}\}$ for $i \in \{1, \dots, n\}$, where Y_i is the scalar response and $W_{ij} = X_i(t_{ij}) + e_{ij}$ is the functional covariate contaminated with measurement error e_{ij} , $t_{ij} \in [0, 1]$, and $X_i(\cdot)$ is the true functional covariate. We generate the data from the following heteroscedastic model: $Y_i = \int X_i(t)tdt + \{1 + \gamma \int X_i(t)t^2dt\}\epsilon$ with $\epsilon \sim N(0, 1)$. This leads to a quantile regression model of the form (1) with $\beta_0(\tau) = \Phi^{-1}(\tau)$, and $\beta(t, \tau) = t + \gamma t^2 \Phi^{-1}(\tau)$. Note that the functional coefficient $\beta(t, \tau)$ is nonlinear in t when $\gamma \neq 0$. Here the scalar γ controls the heteroscedasticity and determines how the coefficient function $\beta(\cdot, \tau)$ ($\tau \in \mathcal{U}$) varies across τ . Specifically, if $\gamma = 0$ then the effect of $X_i(\cdot)$ is constant across different quantile levels of $Y_i|X_i(\cdot)$, while if $\gamma \neq 0$ then the effect of $X_i(\cdot)$ varies across different quantile levels of $Y_i|X_i(\cdot)$.

The true functional covariate $X_i(\cdot)$ is generated from a Gaussian process with zero mean and covariance function $\text{cov}\{X_i(s), X_i(t)\} = \sum_{k \geq 1} \lambda_k \phi_k(s)\phi_k(t)$, where $\lambda_k = (1/2)^{k-1}$ for $k = 1, 2, 3$ and $\lambda_k = 0$ for $k \geq 4$, and $\{\phi_k(\cdot)\}_k$ are the orthonormal Legendre polynomials on $[0, 1]$: $\phi_1(t) = \sqrt{3}(2t - 1)$, $\phi_2(t) = \sqrt{5}(6t^2 - 6t + 1)$, $\phi_3(t) = \sqrt{7}(20t^3 - 30t^2 + 12t - 1)$. It is assumed that the measurement error $e_{ij} \sim N(0, \sigma^2)$. Fig. 1 plots simulated data when $n = 200$, $\gamma = 1$, and $\sigma = 1$.

Table 1: Type I error of the adjusted Wald-type test at significant level $\alpha \in \{0.01, 0.05, 0.10\}$ under dense design. We test H_0 at two sets of quantile levels: $\mathcal{U}_1 = \{0.1, 0.2, 0.3, 0.4\}$ and $\mathcal{U}_2 = \{0.1, 0.2, 0.6, 0.7\}$. Results are based on 5000 simulations.

Scenario	n	0.01	0.05	0.10	Scenario	n	0.01	0.05	0.10
$\sigma = 1$ $\tau \in \mathcal{U}_1$	100	0.021	0.060	0.104	$\sigma = 1$ $\tau \in \mathcal{U}_2$	100	0.030	0.076	0.123
	500	0.014	0.057	0.107		500	0.015	0.062	0.116
	1000	0.017	0.052	0.106		1000	0.015	0.059	0.112
	2000	0.011	0.051	0.101		2000	0.010	0.053	0.103
	5000	0.010	0.054	0.105		5000	0.012	0.056	0.103

The objective is to test the null hypothesis $H_0 : \beta(\cdot, \tau_\ell) = \beta(\cdot, \tau_{\ell'})$ for $\tau_\ell, \tau_{\ell'} \in \mathcal{U}$, that the effect of the true functional covariate on the conditional distribution of the response is the same for all the quantile levels in a given set \mathcal{U} . When $\gamma = 0$, the coefficient function $\beta(\cdot, \tau)$ is independent of τ , which means that null hypothesis is true; when $\gamma \neq 0$ then $\beta(\cdot, \tau)$ varies with τ and thus the null hypothesis is false. We consider two sets of quantile levels: $\mathcal{U}_1 = \{0.1, 0.2, 0.3, 0.4\}$ for one-sided quantile levels, and $\mathcal{U}_2 = \{0.1, 0.2, 0.6, 0.7\}$ for two-sided quantile levels.

We implement the proposed adjusted Wald test using a number of fPC selected via the PVE criterion with PVE=95%. We use the R package *refund* [17] to estimate the fPC scores, where the individual trajectory is reconstructed using penalized regression splines via the function *gam* and the smoothing parameter is selected using the restricted maximum likelihood approach. We investigate the performance of the proposed test for low and high level of measurement error in the functional covariate ($\sigma = 0.05$ and $\sigma = 1$ respectively), for varying sample sizes n from 100 to 5000. For the functional covariates, we consider a dense design in Section 4.2, a sparse design in Section 4.3, and a setting when p_0 diverges in Section 4.4.

4.2. Dense design

We first consider a dense design for the functional covariates: the grid of points for each i is an equispaced grid of $m_i = 100$ timepoints in $[0, 1]$. We are not aware of any testing procedures for testing the null hypothesis of constant effect at various quantile levels, when the covariate is functional; however we can exploit this particular setting and pretend the covariates are vectors and thus use or directly extend existing testing procedures from quantile regression. In particular, we consider three alternative approaches: (1) treat the observed functional covariate as vector and use the common Wald test for vector covariates in quantile regression (NaiveQR); (2) summarize observed functional covariates via a single number summary of the functional covariate in conjunction with the Wald test (SSQR); and (3) treat the observed functional covariate as a vector, reduce the dimensionality using principal component analysis and then apply the Wald test using the vector of principal component scores (pcaQR). For the pcaQR approach, the number of principal components are selected via PVE and using a level PVE=95%. The Wald test for vector covariates in these three approaches is described in [25, Chapter 3.2.3].

Table 1 summarizes the empirical Type I error rates of the adjusted Wald test when testing H_0 at one-sided quantile levels (\mathcal{U}_1) as well as two-sided quantile levels (\mathcal{U}_2), when the functional covariate is observed with large ($\sigma = 1$) measurement error. The results are presented for three significance levels $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.10$; they indicate irrespective of quantile levels set or magnitude of the measurement error the Type I error rates are slightly inflated for moderate sample sizes. Nevertheless the empirical Type I error rates converge to the nominal level. The empirical Type I error rates for the alternative approaches are presented in Table 2. As expected the NaiveQR approach has very poor performance. The NaiveQR approach does hypothesis testing when the covariates are highly correlated; this leads to numerical instability due to singularity of the design matrix. Therefore NaiveQR produces many missing values (reported as “—”) in the table, and yields inflated empirical Type I error rates for any significance level. Results for $\sigma = 0.05$ are similar and omitted here.

The pcaQR approach gives relatively good performance when the magnitude of the error is small ($\sigma = 0.05$): the empirical Type I error is close to the nominal level in results not reported here. However, Table 1 shows that as the error variance increases ($\sigma = 1$), the empirical rejection probabilities are either excessively inflated when $n \in \{1000, 2000, 5000\}$, or there are too many missing values when $n \in \{100, 500\}$. The results are not surprising, because in the case of large error variance, a direct application of principal component analysis yields a large number of principal components. As a consequence, the application of the classical Wald test for vector covariate leads

Table 2: Type I error of alternative approaches at significant level $\alpha \in \{0.01, 0.05, 0.10\}$ under dense design. We test H_0 at two sets of quantile levels: $\mathcal{U}_1 = \{0.1, 0.2, 0.3, 0.4\}$ and $\mathcal{U}_2 = \{0.1, 0.2, 0.6, 0.7\}$. Results are based on 5000 simulations. When one method returns error (due to singularity of the design matrix) in more than 20% replications, we report it as “–”.

Scenario	n	NaiveQR			SSQR			pcaQR		
		0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
$\sigma = 1$ $\tau \in \mathcal{U}_1$	100	–	–	–	0.008	0.033	0.071	–	–	–
	500	–	–	–	0.008	0.036	0.080	–	–	–
	1000	–	–	–	0.010	0.049	0.092	0.996	0.999	1.000
	2000	1.000	1.000	1.000	0.009	0.048	0.097	1.000	1.000	1.000
	5000	1.000	1.000	1.000	0.008	0.053	0.099	0.999	1.000	1.000
$\sigma = 1$ $\tau \in \mathcal{U}_2$	100	–	–	–	0.009	0.040	0.077	–	–	–
	500	–	–	–	0.009	0.050	0.096	–	–	–
	1000	1.000	1.000	1.000	0.009	0.046	0.095	1.000	1.000	1.000
	2000	1.000	1.000	1.000	0.010	0.048	0.099	1.000	1.000	1.000
	5000	1.000	1.000	1.000	0.011	0.051	0.100	1.000	1.000	1.000

to numerical instability due to singularity of the design matrix, in a similar way to the NaiveQR approach. The performance of SSQR approach is very good for all the scenarios considered and across various sample sizes: the empirical Type I error rates are close to the nominal levels. This is expected, as in the case when H_0 holds, the functional covariate effect is through its mean, and this effect is invariant over quantile levels.

Next we evaluate the performance in terms of empirical rejection probabilities when the null hypothesis is not true. We only focus on the proposed adjusted Wald testing and SSQR procedures, as they have the correct size. Fig. 2 illustrates the power curves based on 2000 simulations for large noise with $\sigma = 1$; the results are similar in the case of low noise ($\sigma = 0.05$) and for brevity are not included. The adjusted Wald procedure is much more powerful than SSQR irrespective of the departure from the null hypothesis as reflected by the coefficient γ . For example, when $\gamma = 1$ the probability to correctly reject H_0 using the adjusted Wald is about 100% when the sample size is 500 or more, whereas the counterpart obtained with SSQR is less than 70% even when the sample size increases to 5000. These results are not surprising, as SSQR summarizes the entire functional covariate through a single scalar, while the proposed adjusted Wald test employs the full functional covariate.

4.3. Sparse design

Next, we study the performance of the adjusted Wald testing procedure when the functional covariate is observed sparsely and with measurement error. We set an overall grid of 101 equispaced points in $[0, 1]$ and consider two settings: a ‘moderately sparse’ sampling design with $m_i = 50$ randomly generated time points to form t_{i1}, \dots, t_{im_i} for each i , and a ‘highly sparse’ design with $m_i = 10$. Other aspects of the data generating process follow the dense design described in the previous section. We use the adjusted Wald test which relies on sparse fPCA techniques, that estimate the fPC scores ξ_{ik} ’s using conditional expectation proposed by [46]. When the sampling design of the functional covariate is sparse, there are no obvious reasonable alternative approaches to compare. Thus in this section we only discuss the performance of the proposed Wald-type procedure.

Table 3 shows the empirical Type I error when the noise level $\sigma = 1$. They show excellent performance of the adjusted Wald test in maintaining the nominal levels for moderately large sample size ($n = 1000$ or larger) under both moderately sparse and sparse sampling design of the functional covariate. Fig. 3 shows the power of the adjusted Wald test for moderately sparse and highly sparse designs for $\sigma = 1$. It indicates that the sparsity of the functional covariates slightly affects the proposed functional Wald-type procedure, as expected. Nevertheless the adjusted Wald test continues to display excellent performance. The results are similar for low level of measurement error and for brevity are omitted here.

4.4. Divergent p_0

In this section, we study the performance of the proposed adjusted Wald test when Assumption A4 is violated. We follow the same settings in Section 4.1 but the eigen values and eigenfunctions to generate the functional covariate

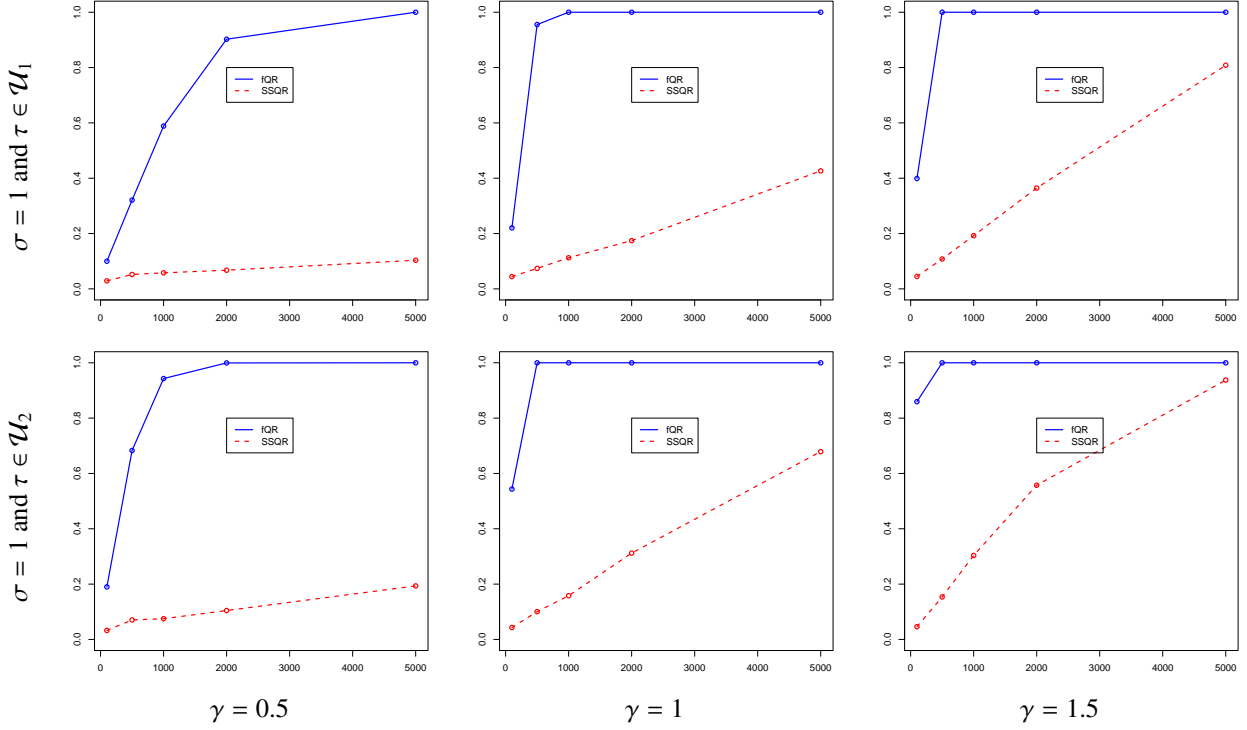


Fig. 2: Power curves of the adjusted Wald test and SSQR under dense design. We test H_0 at two sets of quantile levels: $\mathcal{U}_1 = \{0.1, 0.2, 0.3, 0.4\}$ and $\mathcal{U}_2 = \{0.1, 0.2, 0.6, 0.7\}$. The x -axis is the sample size $n \in \{100, 500, 1000, 2000, 5000\}$. Results are based on 2000 simulations.

Table 3: Type I error of the adjusted Wald test at significance level $\alpha \in \{0.01, 0.05, 0.10\}$ under sparse design. We test H_0 at two sets of quantile levels: $\mathcal{U}_1 = \{0.1, 0.2, 0.3, 0.4\}$ and $\mathcal{U}_2 = \{0.1, 0.2, 0.6, 0.7\}$. The missing rate is 50% for moderate sparsity and 90% for high sparsity. Results are based on 5000 simulations.

Scenario	n	missing rate = 50%			missing rate = 90%		
		0.01	0.05	0.10	0.01	0.05	0.10
$\sigma = 1$ $\tau \in \mathcal{U}_1$	100	0.021	0.063	0.104	0.024	0.075	0.119
	500	0.014	0.055	0.104	0.011	0.058	0.110
	1000	0.014	0.055	0.106	0.013	0.052	0.101
	2000	0.011	0.055	0.106	0.013	0.053	0.103
	5000	0.011	0.052	0.100	0.010	0.048	0.100
$\sigma = 1$ $\tau \in \mathcal{U}_2$	100	0.026	0.075	0.120	0.034	0.092	0.143
	500	0.016	0.058	0.110	0.021	0.069	0.119
	1000	0.013	0.057	0.106	0.014	0.063	0.114
	2000	0.011	0.053	0.100	0.011	0.056	0.108
	5000	0.010	0.048	0.103	0.011	0.049	0.100

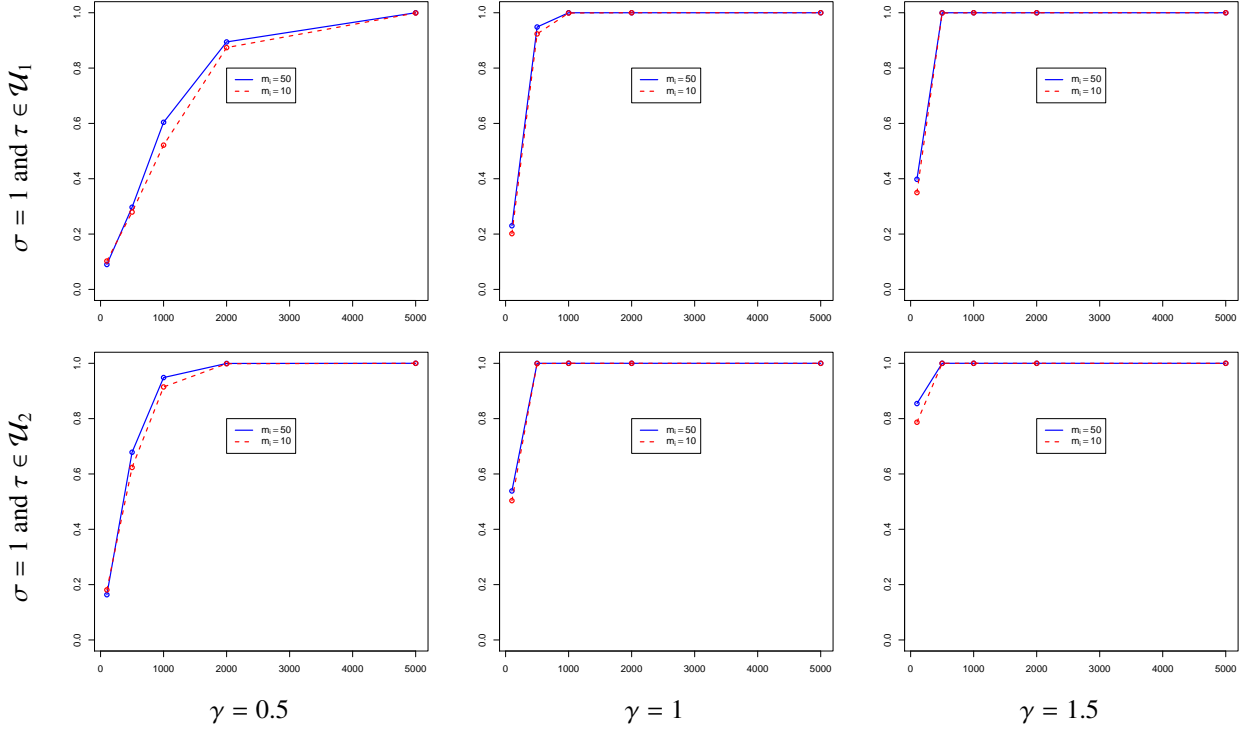


Fig. 3: Power curves of the adjusted Wald test for moderately sparse design with $m_i = 50$ (blue) and highly sparse design with $m_i = 10$ (red). We test H_0 at two sets of quantile levels: $\mathcal{U}_1 = \{0.1, 0.2, 0.3, 0.4\}$ and $\mathcal{U}_2 = \{0.1, 0.2, 0.6, 0.7\}$. The x -axis is the sample size $n \in \{100, 500, 1000, 2000, 5000\}$. Results are based on 2000 simulations.

are given by $\lambda_k = (1/2)^{k-1}$ for $k \in \{1, \dots, \lfloor \sqrt{n} \rfloor\}$ and $\lambda_k = 0$ for $k > \lfloor \sqrt{n} \rfloor$ where $\lfloor \cdot \rfloor$ is the floor function; the eigen function ϕ_k is the k th function in the Fourier basis $\{\sqrt{2} \cos(2\pi t), \sqrt{2} \sin(2\pi t), \sqrt{2} \cos(4\pi t), \sqrt{2} \sin(4\pi t), \dots\}$. We set $\sigma = 1$ for the measurement error in the functional covariate.

Table 4 presents the Type I error rates of the adjusted Wald test under various designs. We can see that even Assumption A4 is violated, the proposed test matches the nominal level when the sample size is large, for both dense and sparse designs. Fig. 4 plots the power curves when $\gamma \in \{0.5, 1, 1.5\}$, which indicates similar performance to the case where p_0 is a small constant. Therefore, it seems that the proposed Wald test continues to show desirable performance when Assumption A4 does not hold, at least under the simulation settings. A theoretical justification may be an interesting research topic.

5. Application

In this section we consider the capital bike sharing study and discuss the application of the proposed testing procedure to formally assess whether the effect of the previous day casual bike rentals on the current day total bike rentals varies across several quantile levels. The bike data [7] is recorded by the Capital Bikeshare System (CBS), Washington D.C., USA, which is available at <http://capitalbikeshare.com/system-data>. As the new generation of bike rentals, bike sharing systems possess membership, rental and return automatically. With currently over 500 bike-share programs around the world [27] and the fast growing trend, data analysis on these systems regarding the effects to public traffic and the environment has become popular. The bike data includes hourly rented bikes for casual users that are collected during January 1st 2011 to December 31st 2012, for a total of 731 days.

Our objective is to formally assess how the previous day casual bike rentals, $X_i(\cdot)$, affects the distribution of the current day total bike rentals counts, Y_i , where $i \in \{1, \dots, 730\}$ denote the i th day starting from January 2nd 2011. A subsequent interest is to predict the 90% quantile of the total casual bike rentals. Fig. 5 plots the hourly profiles of casual bike rentals (left) and the histogram of the total casual bike rentals (right).

Table 4: Type I error of the adjusted Wald test at significance level $\alpha \in \{0.01, 0.05, 0.10\}$ when p_0 is divergent under dense design (no missing) and sparse design. We test H_0 at two sets of quantile levels: $\mathcal{U}_1 = \{0.1, 0.2, 0.3, 0.4\}$ and $\mathcal{U}_2 = \{0.1, 0.2, 0.6, 0.7\}$. The missing rate is 50% for moderate sparsity and 90% for high sparsity. Results are based on 5000 simulations.

\mathcal{U}	n	no missing			missing rate = 50%			missing rate = 90%		
		0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
\mathcal{U}_1	100	0.045	0.107	0.161	0.049	0.107	0.162	0.040	0.096	0.147
	500	0.025	0.085	0.146	0.022	0.081	0.141	0.026	0.084	0.137
	1000	0.015	0.067	0.123	0.019	0.072	0.132	0.013	0.062	0.120
	2000	0.015	0.063	0.120	0.016	0.062	0.115	0.015	0.063	0.119
	5000	0.010	0.052	0.107	0.010	0.059	0.117	0.012	0.055	0.101
\mathcal{U}_2	100	0.070	0.149	0.222	0.061	0.142	0.212	0.051	0.125	0.193
	500	0.035	0.105	0.169	0.037	0.109	0.174	0.025	0.087	0.148
	1000	0.025	0.084	0.148	0.022	0.079	0.140	0.024	0.077	0.132
	2000	0.020	0.072	0.127	0.016	0.067	0.123	0.015	0.063	0.123
	5000	0.013	0.056	0.111	0.014	0.058	0.104	0.012	0.057	0.110

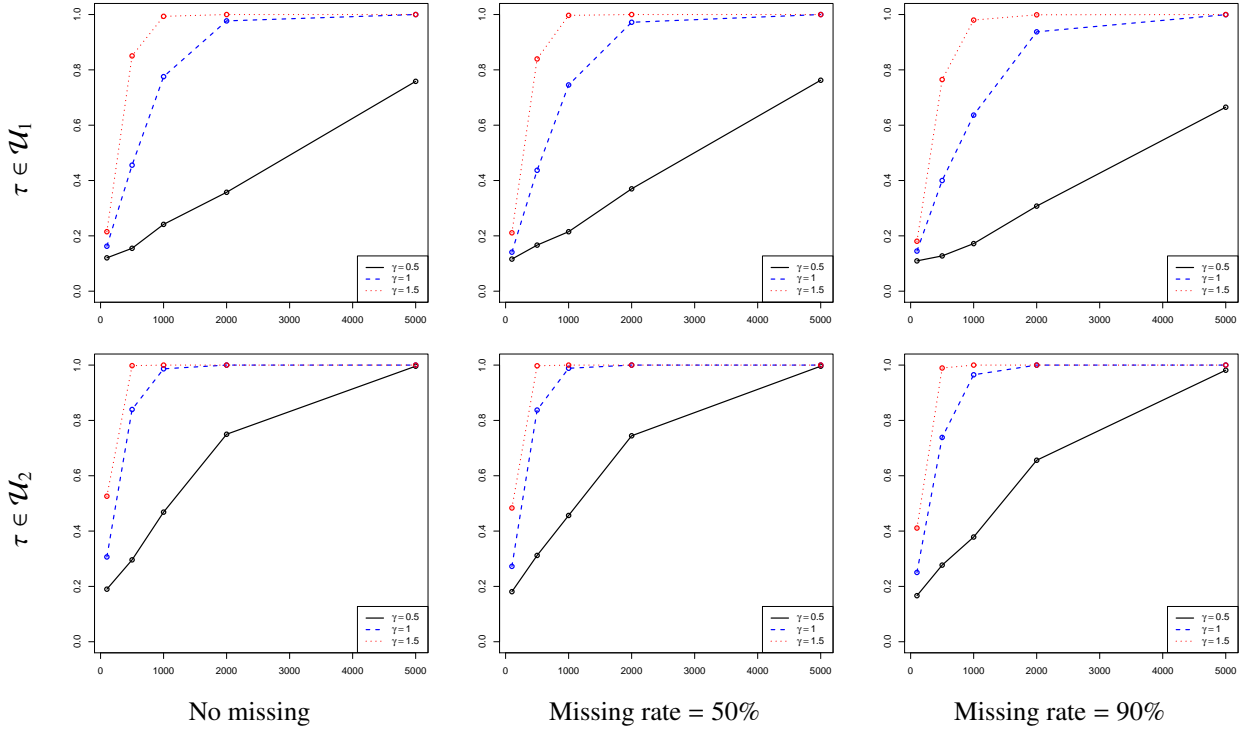


Fig. 4: Power curves of the adjusted Wald test when p_0 is divergent. We test H_0 at two sets of quantile levels: $\mathcal{U}_1 = \{0.1, 0.2, 0.3, 0.4\}$ and $\mathcal{U}_2 = \{0.1, 0.2, 0.6, 0.7\}$. In each plot, the x -axis is the sample size $n \in \{100, 500, 1000, 2000, 5000\}$, and a larger γ corresponds to more deviation from the null hypothesis. Results are based on 2000 simulations.

We assume the functional quantile regression model (1), $Q_{Y_i|X_i}(\tau) = \beta_0(\tau) + \int \beta(t, \tau) X_i^c(t) dt$, where Y_i is the total bike casual bike rentals for the current day and $X_i(\cdot)$ is the true profile of the casual bike rentals recorded in the previous day. As described earlier $\beta_0(\cdot)$ is the quantile varying intercept function and $\beta(\cdot, \tau)$ is the slope parameter and quantifies the effect of the functional covariate at the τ th quantile level of the distribution of the response.

To address the first objective we consider a set of quantile levels and use the proposed testing procedure to test the

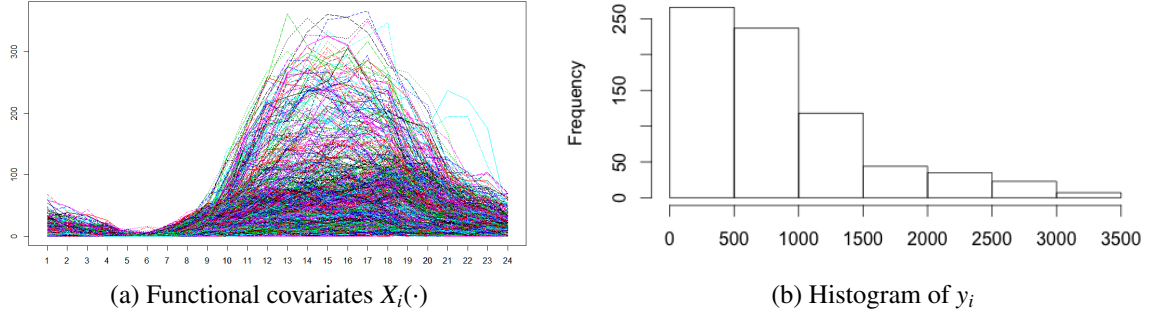


Fig. 5: Bike rental data (casual users). The left panel plots hourly bike rentals for casual users on the previous day $X_i(t)$ for t ranges from 0 to 24 hours, and the right panel plots the histogram of the total casual bike rentals on the current day y_i , where $i \in \{1, \dots, 730\}$.

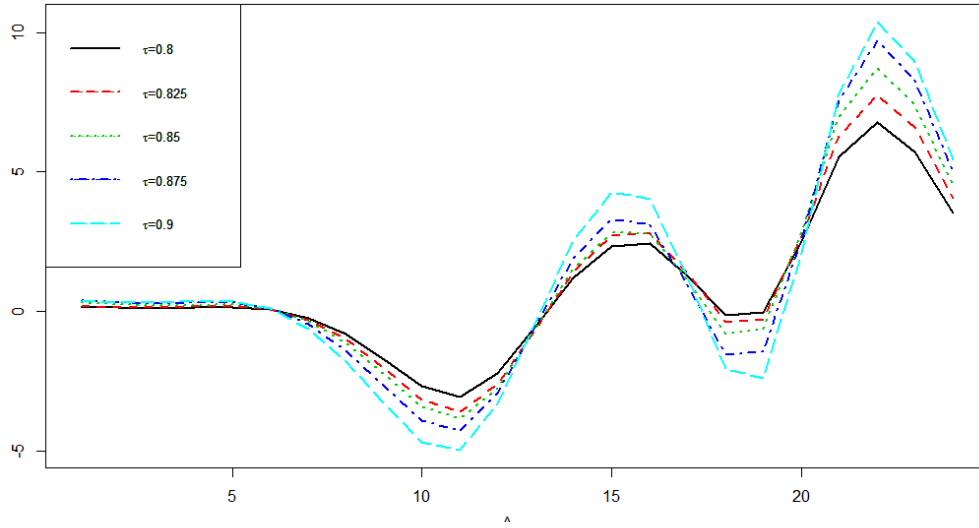


Fig. 6: Estimated $\beta(\cdot, \tau)$ by the proposed method at various quantile levels for the capital bike sharing study. The x -axis ranges from 0 to 24 hours.

null hypothesis

$$H_0 : \beta(\cdot, 0.20) = \beta(\cdot, 0.40) = \beta(\cdot, 0.60) = \beta(\cdot, 0.80).$$

The number of fPC is selected using $\text{PVE} = 99\%$; this choice selects three fPC. We use the adjusted Wald test T_n and its asymptotic null distribution; the resulting p -value is close to zero indicating overwhelming evidence that low and large number of bike rentals are affected differently by the hourly rentals on the previous day.

Next we turn to the problem of predicting the 90% quantile of the total bike rentals for the current day. When some quantile coefficients in a region of quantile levels are constant, we may improve the estimator's efficiency by borrowing information from neighboring quantiles to estimate the common coefficients, especially when the quantile level of interest is high. Here consider the quantile level set $\mathcal{U} = \{0.8, 0.825, 0.85, 0.875, 0.9\}$ around the 90%th quantile. We apply the proposed method to estimate the coefficient functions at various quantile levels \mathcal{U} as shown in Fig. 6. The corresponding adjusted Wald test leads to a p -value = 0.466, which suggests that the quantile coefficients are not significantly different across the quantile levels. We consider combined quantile regression at \mathcal{U} by using the methods of quantile average estimator (QAE) and composite regression of quantiles (CRQ) with equal weights; see [24, 43] for more technical details. We denote the single quantile regression estimation at the 90th quantile by RQ.

We use 1000 bootstrap samples to study the efficiency of the three estimators. Fig. 7 plots the bootstrap means and standard errors of the estimates of $\beta(\cdot, 0.9)$ by QAE, CRQ and RQ. The QAE and CRQ estimators have smaller

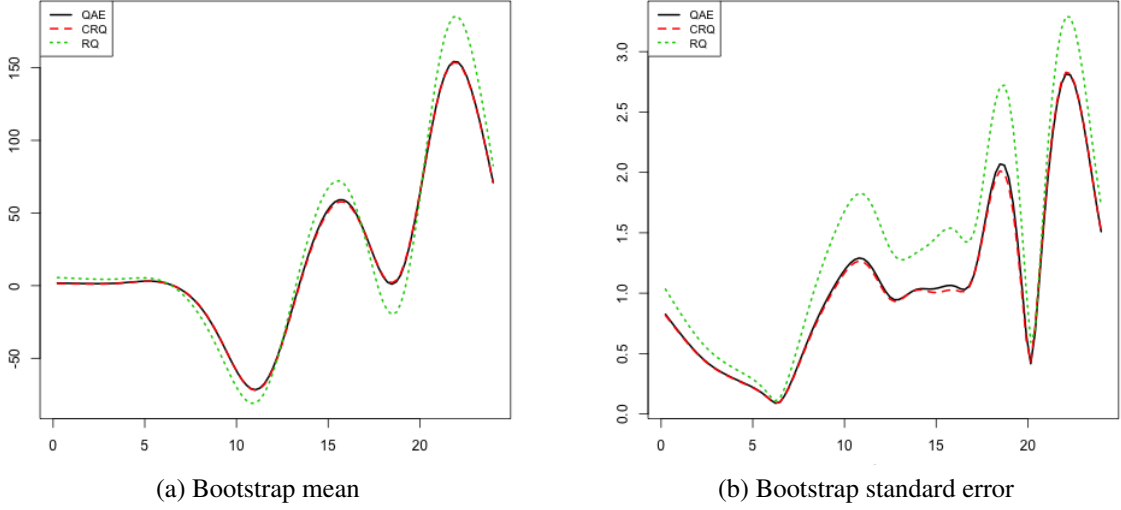


Fig. 7: Bootstrap means (left) and standard errors (right) when estimating $\beta(\cdot, 0.9)$ by QAE, CRQ, and RQ. QAE and CRQ reduce the standard error of RQ. The x -axis ranges from 0 to 24 hours.

Table 5: Prediction errors from different methods averaged over 1000 cross-validations. The maximum standard error of each row is reported in the last column. QAE and CRQ that combine information at various quantile levels tend to yield smaller prediction errors than RQ at more extreme quantile levels.

τ	QAE	CRQ	RQ	SE
0.8	154.163	153.073	152.396	0.277
0.825	146.163	145.598	145.504	0.268
0.85	137.028	136.758	137.071	0.259
0.875	126.138	125.949	126.819	0.252
0.9	112.774	112.842	113.823	0.238

standard errors uniformly for all t , indicating efficiency gain by combining information across quantile levels. We also observe that the number of fPC is either 3 or 4 in all bootstrap samples, suggesting that the assumption A4 is reasonable in this data application.

Furthermore, we conduct a cross-validation by randomly selecting 50% of the data as the training data set and using the other half as the test data set. We use 1000 replications and calculate the prediction error for each replication and each $\tau \in \mathcal{U}$ as follows:

$$\text{PE} = \sum_{i \in \text{test sample}} \rho_{\tau}(y_i - \widehat{\xi}_i^T \widehat{\theta}_{\tau}),$$

where the estimated coefficients $\widehat{\theta}_{\tau}$ are based on the training data and the summation is over the test data. The RQ estimates are obtained separately at each $\tau \in \mathcal{U}$, while the QAE and CRQ estimates are shared across \mathcal{U} . The averaged prediction errors are reported in Table 5. We can see that the application of QAE and CRQ improves the prediction significantly for the 87.5%th and 90%th quantiles; differences among the three methods are not significant at the lower quantiles. This makes sense since data sparsity becomes more severe for more extreme quantile levels. Hence, incorporating lower quantile levels improves efficiency at higher levels, while it may not benefit the prediction performance at lower quantile levels by considering more extreme levels.

6. Proofs

In this section, we prove Theorem 1 and Theorem 2, as well as auxiliary results needed in the proofs, including Lemma 1, Lemma 2, and Lemma 3. We use $\|\cdot\|_{L^2}$ as the L^2 -norm for a function and $\|\cdot\|$ as the Euclidean norm for a vector.

6.1. Proofs of Theorem 1 and Theorem 2

Proof of Theorem 1. The proof proceeds in three steps. In step 1, we approximate the estimated scores $\widehat{\xi}_i$'s by linear combinations of ξ_i 's. In step 2, we obtain the asymptotic distribution of $\widehat{\theta}_\tau$ at a single quantile level. In step 3, we extend the results in step 2 to multiple quantile levels.

Step 1 (Approximation of the estimated scores). Most of the existing literature has been focused on establishing error bounds for estimated eigenvalues and eigenfunctions; see for example [11, 12] and the discussion therein. The following lemma instead characterizes the accuracy in predicting the fPC scores.

Lemma 1. *Under Assumptions A4, B1, B2 and C1, we have*

$$E\|\widehat{\xi}_i - \xi_i\|^2 = o(n^{-1/2}). \quad (10)$$

In addition,

$$\max_{1 \leq i \leq n} |\widehat{\xi}_i - \xi_i - n^{-1/2} B \xi_i| = O_p(n^{-1}), \quad (11)$$

where B is a $(p_0 + 1) \times (p_0 + 1)$ dimensional matrix with the bottom right $p_0 \times p_0$ block matrix equal to B^+ described next and the rest of the elements equal to zero. Here $B^+ = (b_{kk'})$ is a $p_0 \times p_0$ random matrix such that $b_{kk} = 0$ for $k \in \{1, \dots, p_0\}$ and $b_{kk'} = n^{-1/2}(\lambda_k - \lambda_{k'})^{-1} \left(\sum_{i=1}^n \xi_{ik} \xi_{ik'} \right)$ if $k \neq k'$.

The result in (11) indicates that the leading term of $\widehat{\xi}_i - \xi_i$ is $n^{-1/2} B \xi_i$, which is a linear combination of ξ_i with a random weight matrix B that does not depend on i .

Step 2 (Quantile regression on estimated scores). We focus on a single quantile level τ in this step. For any $\delta \in \mathbb{R}^{p_0+1}$, let

$$Z_n(\delta) = \sum_{i=1}^n \{\rho_\tau(\widehat{u}_i - \widehat{\xi}_i^T \delta / \sqrt{n}) - \rho_\tau(\widehat{u}_i)\},$$

where $\widehat{u}_i = y_i - \widehat{\xi}_i^T \theta_\tau$. Then $Z_n(\delta)$ is a convex function which is minimized at $\widehat{\delta}_n = \sqrt{n}(\widehat{\theta}_\tau - \theta_\tau)$. Therefore, the asymptotic distribution of $\widehat{\delta}_n$ is determined by the limiting behavior of $Z_n(\delta)$. Let $\psi_\tau(t) = \tau - I(t < 0)$. According to the Knight's identity [23], we can decompose $Z_n(\delta)$ into two parts: $Z_n(\delta) = Z_{1n}(\delta) + Z_{2n}(\delta)$, where

$$Z_{1n}(\delta) = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \widehat{\xi}_i^T \delta \psi_\tau(\widehat{u}_i), \quad Z_{2n}(\delta) = \sum_{i=1}^n \int_0^{\widehat{\xi}_i^T \delta / \sqrt{n}} \{I(\widehat{u}_i \leq s) - I(\widehat{u}_i \leq 0)\} ds = \sum_{i=1}^n Z_{2ni}(\delta). \quad (12)$$

In order to show (6), it is sufficient to prove that

$$Z_n(\delta) \xrightarrow{d} -\delta^T W(\tau) + \frac{1}{2} \delta^T D_1(\tau) \delta, \quad (13)$$

where $W(\tau) \sim N\{0, \tau(1-\tau)D_0 + D_1(\tau)\Sigma_0(\tau)D_1(\tau)\}$, since one can apply the convexity lemma [34] to the quadratic form of δ in (13).

We next derive the limiting distributions of $Z_{1n}(\delta)$ and $Z_{2n}(\delta)$. For $Z_{1n}(\delta)$, similarly to its definition in (12), we define $Z_{1n}^*(\delta)$ based on the true scores ξ_i :

$$Z_{1n}^*(\delta) = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i^T \delta \psi_\tau(u_i),$$

where $u_i = y_i - Q_{Y_i|X_i}(\tau) = y_i - \xi_i^T \theta_\tau = y_i - \sum_{k=0}^{p_0} \xi_{ik} \beta_k(\tau)$. By a direct application of the central limit theorem (CLT), we obtain that the asymptotic distribution of $Z_{1n}^*(\delta)$ is $N(0, \tau(1-\tau)\delta^T D_0 \delta)$. However, when the predictors are estimated with errors, the difference $Z_{1n}(\delta) - Z_{1n}^*(\delta)$ is non-negligible. Lemma 2 provides a representation of $Z_{1n}(\delta)$ by explicitly formulating this difference.

Lemma 2. *Under Assumptions A4, B1, B2 and C1,*

$$Z_{1n}(\delta) = \delta^T \left[-\frac{1}{\sqrt{n}} \sum_{i=1}^n \{\xi_i \psi_\tau(u_i) - D_1(\tau) d_i\} \right] + o_p(1),$$

where $d_i = (0, d_{i1}, \dots, d_{ip_0})^T$ and $d_{ik} = \sum_{r=1, r \neq k}^{p_0} (\lambda_k - \lambda_r)^{-1} \xi_{ik} \xi_{ir} \beta_r(\tau)$, $k \geq 1$.

Since $\xi_i \psi_\tau(u_i) - D_1(\tau) d_i$ are i.i.d., Lemma 2 allows us to directly apply Linderberg's CLT to obtain the asymptotic distribution of $Z_{1n}(\delta)$. Note that $E\{\xi_i \psi_\tau(u_i)\} = 0$ and $\text{Var}\{\xi_i \psi_\tau(u_i)\} = \tau(1 - \tau)D_0$. In addition, $E d_i = 0$ because ξ_{ik} and ξ_{ir} are uncorrelated and have mean 0 (when $r \neq k$). Let the matrix $\Sigma(\tau)$ be the covariance matrix of d_i whose first row and first column is all 0 and the $(k + 1, k' + 1)$ th element ($k, k' = 1, \dots, p_0$) is given by $\text{Cov}(d_{ik}, d_{ik'}) = \theta_\tau^T A^{k,k'} \theta_\tau$ for some $(p_0 + 1)$ by $(p_0 + 1)$ matrix $A^{k,k'}$. The first row and first column of $A^{k,k'}$ are all 0, and simple calculation yields its bottom right block $A^{k,k',+} = (\sigma_{j,j'})$:

$$\sigma_{j,j'} = \begin{cases} 0, & \text{if } j = k \text{ or } j' = k', \\ (\lambda_k - \lambda_j)^{-1}(\lambda_{k'} - \lambda_{j'})^{-1} E(\xi_{1k} \xi_{1j} \xi_{1k'} \xi_{1j'}), & \text{otherwise.} \end{cases} \quad (14)$$

Let $\Theta_\tau = 1_{(p_0+1) \times (p_0+1)} \otimes \theta_\tau^T$, and Σ_0 be a $(p_0 + 1)^2$ by $(p_0 + 1)^2$ matrix whose $(k + 1, k' + 1)$ th block is $A^{k,k'}$ ($k, k' = 1, \dots, p_0$) and $(k + 1, k' + 1)$ th block is $0_{(p_0+1) \times (p_0+1)}$ for $k = 0$ or $k' = 0$. Then $\Sigma(\tau)$ can be rewritten as $\Sigma(\tau) = \Theta_\tau \Sigma_0 \Theta_\tau^T$. Furthermore, we have

$$\text{Cov}\{\xi_i \psi_\tau(u_i), d_i\} = E\{\psi_\tau(u_i) \xi_i^T d_i\} = E\{\xi_i^T d_i E\psi_\tau(u_i) | \xi_i\} = 0,$$

which leads to

$$-\frac{1}{\sqrt{n}} \sum_{i=1}^n \{\xi_i \psi_\tau(u_i) - D_1(\tau) d_i\} \xrightarrow{d} N(0, \tau(1 - \tau)D_0 + D_1(\tau)\Sigma(\tau)D_1(\tau)).$$

Hence, we have $Z_{1n}(\delta) \xrightarrow{d} -\delta^T W(\tau)$ where $W(\tau) \sim N(0, \tau(1 - \tau)D_0 + D_1(\tau)\Sigma(\tau)D_1(\tau))$. Consequently, the following result for $Z_{2n}(\delta)$ concludes the asymptotic distribution in (13).

Lemma 3. *Under Assumptions A4, B1, B2 and C1, we have*

$$Z_{2n}(\delta) = \frac{1}{2} \delta^T D_1(\tau) \delta + o_p(1).$$

Step 3 (Asymptotic distributions across quantile levels). When considering various quantile levels, the same arguments can be made via a convex optimization and the limiting distribution of the objective function. The asymptotic covariance in (7) is obtained by the covariance between $\xi_i \psi_{\tau_\ell}(u_i) + D_1(\tau_\ell) d_i(\tau_\ell)$ and $\xi_i \psi_{\tau_{\ell'}}(u_i) + D_1(\tau_{\ell'}) d_i(\tau_{\ell'})$, following similar calculation as in (14). \square

Proof of Theorem 2. We just need to show that $R(\Gamma_{\hat{\zeta}}^a - \Gamma_{\hat{\zeta}})R^T = 0$. The (ℓ, ℓ') th block of the matrix $\Gamma_{\hat{\zeta}}^a - \Gamma_{\hat{\zeta}}$ is $\Theta_{\tau_\ell} \Sigma_0 \Theta_{\tau_{\ell'}}^T$, where $1 \leq \ell, \ell' \leq L$. Therefore, we have $\Gamma_{\hat{\zeta}}^a - \Gamma_{\hat{\zeta}} = A \Sigma_0 A^T$ where $A = (\Theta_{\tau_1}, \dots, \Theta_{\tau_L})^T$ is a $(p_0 + 1)L \times (p_0 + 1)$ matrix. Noting that $\Theta_{\tau_\ell} = 1_{(p_0+1) \times (p_0+1)} \otimes \theta_\ell^T$ for $\ell \in \{1, \dots, L\}$, we have $A^T = 1_{(p_0+1) \times (p_0+1)} \otimes \zeta^T$ and thus $A = 1_{(p_0+1) \times (p_0+1)} \otimes \zeta$. Therefore, when $R\zeta = 0$, it follows that $RA = 1_{(p_0+1) \times (p_0+1)} \otimes (R\zeta) = 0$. This completes the proof. \square

6.2. Proofs of lemmas

Proof of Lemma 1. The bound in (10) follows from standard bounds for the estimated eigenfunctions and covariance kernel in the FDA literature. According to Theorem 1 in [11], we have

$$\|\widehat{\phi}_k - \phi_k\|_{L^2} \leq 8^{1/2} s_k^{-1} \|\widehat{G} - G\|,$$

where $s_k = \min_{r \leq k} (\lambda_r - \lambda_{r+1})$ and $\|\widehat{G} - G\| = [\int_0^1 \int_0^1 \{\widehat{G}(u, v) - G(u, v)\}^2 du dv]^{1/2}$. Therefore,

$$|\widehat{z}_{ik} - \xi_{ik}| = \left| \int_0^1 X_i(t) \{\widehat{\phi}_k(t) - \phi_k(t)\} dt \right| \leq \|X_i\|_{L^2} \cdot \|\widehat{\phi}_k - \phi_k\|_{L^2} \leq \text{constant} \cdot \|X_i\|_{L^2} s_k^{-1} \cdot \|\widehat{G} - G\|,$$

which leads to $\|\widehat{\xi}_i - \xi_i\| \leq \text{constant} \cdot \|X_i\|_{L^2} s_{p_0}^{-1} \|\widehat{G} - G\|$. For any $c > 0$, invoking the bound $E\|\widehat{G} - G\|^c \leq \text{constant} \cdot n^{-c/2}$ [12, Lemma 3.3] leads to

$$E\|\widehat{\xi}_i - \xi_i\|^c \leq \text{constant} \cdot s_{p_0}^{-c} (E\|\widehat{G} - G\|^{2c})^{1/2} \leq \text{constant} \cdot s_{p_0}^{-c} n^{-c/2}.$$

Thus, for finite p_0 , we have $E\|\widehat{\xi}_i - \xi_i\|^c = o(n^{-c/4})$; in particular, there holds $\sqrt{n} E\|\widehat{\xi}_i^T - \xi_i^T\|^2 = o(1)$.

Next we prove the representation in (11). Let \widetilde{G} be the estimator of the kernel G based on the fully observed covariate $X_i(\cdot)$, and recall that \widehat{G} is the estimate based on the discretized W_{ij} with measurement error. Denote $\widetilde{\mathcal{Z}} = \sqrt{n}(\widetilde{G} - G)$ and $\widehat{\mathcal{Z}} = \sqrt{n}(\widehat{G} - G)$. We use the notation $\int \widehat{\mathcal{Z}}\phi_k\phi_{k'}$ to denote $\int_0^1 \int_0^1 \widehat{\mathcal{Z}}(u, v)\phi_k(u)\phi_{k'}(v)dudv$.

Since $\{\phi_k : k \geq 1\}$ forms a basis of the L^2 space on $[0, 1]$, we have $\widehat{\phi}_k = \sum_{k'=1}^{\infty} a_{kk'}\phi_{k'}$, where $k \in \{1, \dots, p_0\}$ and the generalized Fourier coefficients $a_{kk'} = \int_0^1 \widehat{\phi}_k(t)\phi_{k'}(t)dt$. Furthermore, we have the following expansion for $a_{kk'}$'s:

$$a_{kk} = 1 + O_p(n^{-1}); \quad a_{kk'} = n^{-1/2}(\lambda_k - \lambda_{k'})^{-1} \int \widehat{\mathcal{Z}}\phi_k\phi_{k'} + O_p(n^{-1}) \text{ if } k \neq k',$$

according to (2.6) and (2.7) in [11]. Therefore, for $k \in \{1, \dots, p_0\}$, we have

$$\int_0^1 X_i(t)\{\widehat{\phi}_k(t) - \phi_k(t)\}dt = \sum_{k'=1}^{p_0} \{a_{kk'} - I(k' = k)\}\xi_{ik'} = \sum_{k'=1, k' \neq k}^{p_0} n^{-1/2}(\lambda_k - \lambda_{k'})^{-1} \int \widehat{\mathcal{Z}}\phi_k\phi_{k'}\xi_{ik'} + O_p(n^{-1}).$$

A direct calculation gives that

$$\int \widehat{\mathcal{Z}}\phi_k\phi_{k'} = n^{-1/2} \sum_{i=1}^n \xi_{ik}\xi_{ik'} - n^{1/2}\bar{\xi}_k\bar{\xi}_{k'}$$

for $k, k' = 1, \dots, p_0$ and $k \neq k'$, where $\bar{\xi}_k = n^{-1} \sum_{i=1}^n \xi_{ik}$. Since $n^{1/2}\bar{\xi}_k\bar{\xi}_{k'} = n^{-1/2} \cdot (n^{1/2}\bar{\xi}_k) \cdot (n^{1/2}\bar{\xi}_{k'}) = n^{-1/2} \cdot O_p(1) \cdot O_p(1) = O_p(n^{-1/2})$, we have $\int \widehat{\mathcal{Z}}\phi_k\phi_{k'} = n^{-1/2} \sum_{i=1}^n \xi_{ik}\xi_{ik'} + O_p(n^{-1/2})$. The same approximation holds when using $\widehat{\mathcal{Z}}$ since $\widehat{\mathcal{Z}} - \widetilde{\mathcal{Z}}$ is uniformly $o_p(n^{-1/2})$ as shown by [48]. Consequently,

$$\int_0^1 X_i(t)\{\widehat{\phi}_k(t) - \phi_k(t)\}dt = \sum_{k'=1, k' \neq k}^{p_0} n^{-1}(\lambda_k - \lambda_{k'})^{-1} \left(\sum_{i=1}^n \xi_{ik}\xi_{ik'} \right) \xi_{ik'} + O_p(n^{-1}). \quad (15)$$

This approximation will not be affected if we use $\widehat{X}_i(\cdot)$ instead of the true curve $X_i(\cdot)$ because the difference $\widehat{X}_i(\cdot) - X_i(\cdot)$ is negligible uniformly for all i (e.g., see Theorem 2 in [48] or Lemma 1 in [51]). Let a p_0 -dimension random matrix $B^+ = (b_{kk'})$ where $b_{kk'} = 0$ if $k = k'$ and $b_{kk'} = n^{-1/2}(\lambda_k - \lambda_{k'})^{-1} \left(\sum_{i=1}^n \xi_{ik}\xi_{ik'} \right)$ if $k \neq k'$. Let B be a $(p_0 + 1) \times (p_0 + 1)$ zero matrix but the bottom right block is replaced by B^+ , then the right hand side in (15) becomes $n^{-1/2}B\xi_i + O_p(n^{-1})$. Consequently, we have

$$\widehat{\xi}_i - \xi_i = n^{-1/2}B\xi_i + O_p(n^{-1}).$$

This completes the proof by noting that all the stochastic bounds starting from $O_p(n^{-1})$ in a_{kk} do not depend on i . \square

Proof of Lemma 2. We first decompose the difference between $Z_{1n}(\delta)$ and $Z_{1n}^*(\delta)$ into three parts S_1, S_2 and S_3 as follows:

$$\begin{aligned} Z_{1n}(\delta) - Z_{1n}^*(\delta) &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \widehat{\xi}_i^T \delta \psi_\tau(\widehat{u}_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^n z_i^T \delta \psi_\tau(u_i) \\ &= \left(-\frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{\xi}_i^T - \xi_i^T) \delta \{\psi_\tau(\widehat{u}_i) - \psi_\tau(u_i)\} \right) + \left(-\frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{\xi}_i^T - \xi_i^T) \delta \psi_\tau(u_i) \right) + \left(-\frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i^T \delta \{\psi_\tau(\widehat{u}_i) - \psi_\tau(u_i)\} \right) \\ &=: S_1 + S_2 + S_3. \end{aligned}$$

The proof proceeds in three steps: $S_2 = o_p(1)$ (Step i), $S_1 = o_p(1)$ (Step ii), and $S_3 = n^{-1/2}\delta^T D_1(\tau) \sum_{i=1}^n d_i + o_p(1)$ (Step iii). Step i and Step ii indicate that the first two terms S_1 and S_2 are negligible, and it is sufficient to show that $E(S_2^2) = o(1)$ and $E|S_1| = o(1)$ according to Chebyshev's inequality. The third term S_3 is challenging to analyze since the function of $\psi_\tau(\cdot)$ is not differentiable. In Step iii, we approximate the term S_3 mainly using the uniform approximation on $\psi_\tau(\cdot)$.

Step i. First notice that $E\{\psi_\tau(u_i)|\xi_i, \widehat{\xi}_i\} = 0$ and $E\{\psi_\tau(u_i)^2|\xi_i, \widehat{\xi}_i\} = \tau - \tau^2$. Therefore, we have $E(S_2) = 0$, and further

$$E(S_2^2) = E \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{\xi}_i^T \delta - \xi_i^T \delta) \psi_\tau(u_i) \right\}^2 = \frac{1}{n} \sum_{i=1}^n \sum_{i'=1}^n E \left\{ (\widehat{\xi}_i^T - \xi_i^T) \delta \psi_\tau(u_i) \cdot (\widehat{\xi}_{i'}^T - \xi_{i'}^T) \delta \psi_\tau(u_{i'}) \right\}.$$

For $i = i'$,

$$\begin{aligned} \mathbb{E}\left\{\left(\widehat{\xi}_i^T - \xi_i^T\right)\delta\psi_\tau(u_i) \cdot \left(\widehat{\xi}_{i'}^T - \xi_{i'}^T\right)\delta\psi_\tau(u_{i'})\right\} &= \mathbb{E}\left\{\left(\widehat{\xi}_i^T \delta - \xi_i^T \delta\right)\psi_\tau(u_i)\right\}^2 \\ &= \mathbb{E}\left\{\left(\widehat{\xi}_i^T \delta - \xi_i^T \delta\right)^2 \mathbb{E}\{\psi_\tau^2(u_i)|\xi_i, \widehat{\xi}_i\}\right\} = \tau(1 - \tau)\mathbb{E}\left\{\left(\widehat{\xi}_i^T \delta - \xi_i^T \delta\right)^2\right\}. \end{aligned}$$

Since $\widehat{\xi}_i$ are identically distributed for all i , we have $\mathbb{E}\left\{\left(\widehat{\xi}_i^T \delta - \xi_i^T \delta\right)^2\right\} = \mathbb{E}\left\{\left(\widehat{\xi}_1^T \delta - \xi_1^T \delta\right)^2\right\}$. For $i \neq i'$, we have $\mathbb{E}\left\{\left(\widehat{\xi}_i^T - \xi_i^T\right)\delta\psi_\tau(u_i) \cdot \left(\widehat{\xi}_{i'}^T - \xi_{i'}^T\right)\delta\psi_\tau(u_{i'})\right\} = 0$ by noting that

$$\mathbb{E}\{\psi_\tau(u_i)\psi_\tau(u_{i'})|\xi_i, \widehat{\xi}_i, \xi_{i'}, \widehat{\xi}_{i'}\} = \mathbb{E}\{\psi_\tau(u_i)|\xi_i, \widehat{\xi}_i, \xi_{i'}, \widehat{\xi}_{i'}\} \cdot \mathbb{E}\{\psi_\tau(u_{i'})|\xi_i, \widehat{\xi}_i, \xi_{i'}, \widehat{\xi}_{i'}\} = 0.$$

Therefore, $\mathbb{E}(S_2^2) = \tau(1 - \tau)\mathbb{E}\left\{\left(\widehat{\xi}_1^T \delta - \xi_1^T \delta\right)^2\right\} = O(\mathbb{E}\|\widehat{\xi}_i - \xi_i\|^2) = o(1)$.

Step ii. For S_1 , we first introduce the notation

$$\Delta_i = \mathbb{E}(\psi_\tau(\widehat{u}_i)|\xi_i, \widehat{\xi}_i) = \tau - F_i(\widehat{\xi}_i^T \theta_\tau) = F_i(\xi_i^T \theta_\tau) - F_i(\widehat{\xi}_i^T \theta_\tau).$$

For each i , this random variable Δ_i satisfies that

$$\Delta_i = \mathbb{E}(\psi_\tau(\widehat{u}_i) - \psi_\tau(u_i)|\xi_i, \widehat{\xi}_i), \quad (16)$$

$$|\Delta_i| = \mathbb{E}(|\psi_\tau(\widehat{u}_i) - \psi_\tau(u_i)||\xi_i, \widehat{\xi}_i). \quad (17)$$

The result given in (16) is obtained by noting that $\psi_\tau(u_i)$ has mean 0 conditional on ξ_i , while (17) holds because $|\psi_\tau(\widehat{u}_i) - \psi_\tau(u_i)| = I\{\min(\widehat{\xi}_i^T \theta_\tau, \xi_i^T \theta_\tau) < y_i < \max(\widehat{\xi}_i^T \theta_\tau, \xi_i^T \theta_\tau)\}$.

By Taylor's theorem, for any $a, b \in \mathbb{R}$, we have

$$F(a + b) - F(a) = f(a)b + b^2 \int_0^1 f'(a + tb)(1 - t)dt =: f(a)b + \frac{b^2}{2}R(a, b),$$

where $|R(a, b)| \leq C_0$. Therefore,

$$\Delta_i = -(\widehat{\xi}_i^T \theta_\tau - \xi_i^T \theta_\tau)f_i(\xi_i^T \theta_\tau) + (\widehat{\xi}_i^T \theta_\tau - \xi_i^T \theta_\tau)^2 R(\widehat{\xi}_i^T, \xi_i^T),$$

where $|R(\widehat{\xi}_i^T, \xi_i^T)| \leq 2C_0$. We also have the bound

$$\mathbb{E}\Delta_i^2 \leq \text{constant} \cdot \mathbb{E}\|\widehat{\xi}_i - \xi_i\|^2 = o(n^{-1/2}).$$

Therefore, $|S_1| \leq \frac{1}{\sqrt{n}} \sum_{i=1}^n |(\widehat{\xi}_i^T \delta - \xi_i^T \delta)| \cdot |\psi_\tau(\widehat{u}_i) - \psi_\tau(u_i)|$ and consequently

$$\begin{aligned} \mathbb{E}|S_1| &\leq \frac{1}{\sqrt{n}} \mathbb{E}\left\{\sum_{i=1}^n |(\widehat{\xi}_i^T \delta - \xi_i^T \delta)| |\Delta_i|\right\} \\ &= \sqrt{n} \mathbb{E}|(\widehat{\xi}_1^T \delta - \xi_1^T \delta) \Delta_1| \leq \sqrt{n \mathbb{E}\|\widehat{\xi}_i - \xi_i\|^2 \mathbb{E}\Delta_1^2} = o(1). \end{aligned}$$

Step iii. Define

$$R_n(t) = \sum_{i=1}^n \xi_i \{\psi_\tau(u_i - \xi_i^T t) - \psi_\tau(u_i)\},$$

for any vector such that $\|t\| \leq C$ for some constant C . Then the uniform approximation [14] indicates that

$$\sup \|R_n(t) - \mathbb{E}\{R_n(t)\}\| = O_p(n^{1/2}(\log n)\|t\|^{1/2}).$$

On the other hand,

$$\begin{aligned} \mathbb{E}\{R_n(t)\} &= \sum_{i=1}^n \mathbb{E}[\xi_i \{F_i(\xi_i^T \theta_\tau) - F_i(\xi_i^T \theta_\tau - \xi_i^T t)\}] = n \mathbb{E}[\xi_1 \{F_1(\xi_1^T \theta_\tau) - F_1(\xi_1^T \theta_\tau - \xi_1^T t)\}] \\ &= -n \mathbb{E}\xi_1 \xi_1^T f_1(\xi_1^T \theta_\tau) t + O(n \mathbb{E}\|\xi_1\|^3 \|t\|^2) = -n D_1(\tau) t + O(n \|t\|^2). \end{aligned}$$

Therefore,

$$R_n(t) = -nD_1(\tau)t + O(n\|t\|^2) + O_p\left(n^{1/2}(\log n)\|t\|^{1/2}\right). \quad (18)$$

Note that $\widehat{u}_i = u_i + \xi_i^T \theta_\tau - \widehat{\xi}_i^T \theta_\tau$ and $\widehat{\xi}_i - \xi_i = B\xi_i$ up to a negligible term $O_p(n^{-1})$. Then

$$-\frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i^T \delta \{\psi_\tau(\widehat{u}_i) - \psi_\tau(u_i)\} = -n^{-1/2} R_n(n^{-1/2} B\theta_\tau) + o_p(1),$$

where the term $o_p(1)$ is obtained by the same arguments used in Step ii via conditional expectation and Taylor theorem. Substituting $t = n^{-1/2} B\theta_\tau$ into (18) and noting that $\|n^{-1/2} B\theta_\tau\| = O_p(n^{-1/2})$, we obtain that $R_n(n^{-1/2} B\theta_\tau) = -n^{1/2} D_1(\tau) B\theta_\tau + O(1) + O_p(n^{1/4} \log n)$, leading to

$$S_3 = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i^T \delta \{\psi_\tau(\widehat{u}_i) - \psi_\tau(u_i)\} = \delta^T D_1(\tau) B\theta_\tau + o_p(1).$$

According to the definition of B in (11), it is easy to verify that $B\theta_\tau = n^{-1/2} \sum_{i=1}^n d_i$, where $d_i = (0, d_{i1}, \dots, d_{ip_0})$ and $d_{ik} = \sum_{r=1, r \neq k}^{p_0} (\lambda_k - \lambda_r)^{-1} \xi_{ik} \xi_{ir} \beta_r(\tau)$ for $k \geq 1$. Therefore, it follows that $S_3 = n^{-1/2} \delta^T D_1(\tau) \sum_{i=1}^n d_i + o_p(1)$, which concludes Step iii. This completes the proof. \square

Proof of Lemma 3. Recall that $Z_{2n} = \sum_{i=1}^n Z_{2ni}$, where

$$Z_{2ni}(\delta) = \int_0^{\widehat{\xi}_i^T \delta / \sqrt{n}} \{I(\widehat{u}_i \leq s) - I(\widehat{u}_i \leq 0)\} ds.$$

First, we have

$$\mathbb{E}[Z_{2ni}(\delta) | \xi_i, \widehat{\xi}_i] = \int_0^{\widehat{\xi}_i^T \delta / \sqrt{n}} F_i(\widehat{\xi}_i^T \theta_\tau + s) - F_i(\widehat{\xi}_i^T \theta_\tau) ds = \frac{1}{\sqrt{n}} \int_0^{\widehat{\xi}_i^T \delta} F_i(\widehat{\xi}_i^T \theta_\tau + \frac{t}{\sqrt{n}}) - F_i(\widehat{\xi}_i^T \theta_\tau) dt.$$

Therefore, by Taylor's theorem, we have

$$\mathbb{E}[Z_{2ni}(\delta) | \xi_i, \widehat{\xi}_i] = \frac{1}{\sqrt{n}} \int_0^{\widehat{\xi}_i^T \delta} f_i(\widehat{\xi}_i^T \theta_\tau) \frac{t}{\sqrt{n}} + \frac{t^2}{2n} R(\widehat{\xi}_i^T \delta, \frac{t}{\sqrt{n}}) dt = \frac{1}{2n} \delta^T \widehat{\xi}_i f_i(\widehat{\xi}_i^T \theta_\tau) \widehat{\xi}_i^T \delta + R_{ni},$$

where R_{ni} is the remainder satisfying that $|R_{ni}| \leq cn^{-3/2} |\widehat{\xi}_i^T \delta|^3$. Consequently,

$$\mathbb{E}[Z_{2ni}(\delta) | \xi_i, \widehat{\xi}_i] = \frac{1}{2} \cdot \delta^T \frac{1}{n} \widehat{\xi}_i f_i(\widehat{\xi}_i^T \theta_\tau) \widehat{\xi}_i^T \delta + R_{ni}.$$

Therefore, the unconditional expectation of $Z_{2ni}(\delta)$ is

$$\mathbb{E}Z_{2ni}(\delta) = \mathbb{E}\{\mathbb{E}[Z_{2ni}(\delta) | \xi_i, \widehat{\xi}_i]\} = \frac{1}{2} \cdot \delta^T \mathbb{E}\left(\frac{1}{n} \widehat{\xi}_i f_i(\widehat{\xi}_i^T \theta_\tau) \widehat{\xi}_i^T\right) \delta + \mathbb{E}R_{ni} = \frac{1}{2} \cdot \frac{1}{n} \delta^T D_1(\tau) \delta + \mathbb{E}(R_{ni}),$$

leading to

$$\mathbb{E}Z_{2n} = \frac{1}{2} \delta^T D_1(\tau) \delta + \sum_{i=1}^n \mathbb{E}(R_{ni}).$$

The second term $\sum_{i=1}^n \mathbb{E}(R_{ni})$ is negligible because

$$\left| \sum_{i=1}^n \mathbb{E}(R_{ni}) \right| \leq \sum_{i=1}^n \mathbb{E}|R_{ni}| \leq cn^{-3/2} \sum_{i=1}^n \mathbb{E}|\widehat{\xi}_i^T \delta|^3 = cn^{-1/2} \mathbb{E}|\widehat{\xi}_1^T \delta|^3 \quad (19)$$

$$\leq O(n^{-1/2}) \cdot (\mathbb{E}\|\widehat{\xi}_1\|_2^3) \|\delta\|_2^3 = O(n^{-1/2}) \cdot O(1) = o(1), \quad (20)$$

where the last step is due to the fact that $\|\widehat{\xi}_1\|_2 = \|\widehat{X}_1\|_2 \leq \|\widehat{X}_1 - X_1\|_2 + \|X_1\|_2$.

We next will show that $\max_{i=1,\dots,n} \|\xi_i\|/\sqrt{n} \xrightarrow{P} 0$. Note that $\|\xi_i\|^2 = 1 + \xi_{i1}^2 + \dots + \xi_{ip_0}^2$ for $i \in \{1, \dots, n\}$, and $\|\xi_i\|^2$'s are i.i.d. with a finite second moment $E\|\xi_i\|^2 = 1 + \lambda_1 + \dots + \lambda_{p_0} < \infty$. For any $\epsilon > 0$, we have

$$\begin{aligned} P\left(\max_{i=1,\dots,n} \frac{\|\xi_i\|}{\sqrt{n}} > \epsilon\right) &\leq \sum_{i=1}^n P(\|\xi_i\| > \sqrt{n}\epsilon) \leq \frac{1}{n\epsilon^2} \sum_{i=1}^n E\{\|\xi_i\|^2 I(\|\xi_i\| > \sqrt{n}\epsilon)\} \\ &= \frac{1}{\epsilon^2} E\{\|\xi_1\|^2 I(\|\xi_1\| > \sqrt{n}\epsilon)\} \rightarrow 0, \end{aligned}$$

according to the dominated convergence theorem. It implies that $\max_{i=1,\dots,n} \|\widehat{\xi}_i\|/\sqrt{n} = o_p(1)$ since $\|\widehat{\xi}_i - \xi_i\| = o_p(1)$ uniformly for all i 's. Consequently, $\text{Var}(Z_{2n}|\xi_i$'s, $\widehat{\xi}_i$'s) $\leq \max_{i=1,\dots,n} \|\widehat{\xi}_i\| \delta/\sqrt{n} \cdot E(Z_{2ni}|\xi_i$'s, $\widehat{\xi}_i$'s) $= o_p(1)$, i.e., the conditional variance converges to 0 in probability. Therefore, following the martingale argument in the proof of Theorem 2 in [34], we have $Z_{2n} - E(Z_{2n}) = o_p(1)$. This completes the proof. \square

Acknowledgments

We thank the Editor, the Associate Editor and two anonymous referees for constructive comments that helped to improve the paper. The authors would like to acknowledge the support of NSF DMS 2015569 (Li) and DMS 1454942 (Staicu) and NIH 5P01 CA142538-09 (Staicu).

References

- [1] G. Cao, S. Wang, L. Wang, Estimation and inference for functional linear regression models with partially varying regression coefficients, *Stat* 9 (2020) e286.
- [2] H. Cardot, C. Crambes, P. Sarda, Quantile regression when the covariates are functions, *Nonparametric Statistics* 17 (2005) 841–856.
- [3] K. Chen, H.-G. Müller, Conditional quantile analysis when covariates are functions, with application to growth data, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74 (2012) 67–89.
- [4] C. Crambes, A. Gannoun, Y. Henchiri, Weak consistency of the Support Vector Machine Quantile Regression approach when covariates are functions, *Statistics and Probability Letters* 81 (2011) 1847–1858.
- [5] C. Crambes, A. Gannoun, Y. Henchiri, Support vector machine quantile regression approach for functional data: Simulation and application studies, *Journal of Multivariate Analysis* 121 (2013) 50–68.
- [6] J. Fan, I. Gijbels, *Local Polynomial Modelling and Its Applications*, Monographs on Statistics and Applied Probability, Chapman & Hall, 1996.
- [7] H. Fanaee-T, J. Gama, Event labeling combining ensemble detectors and background knowledge, *Progress in Artificial Intelligence* 2 (2014) 113–127.
- [8] F. Ferraty, A. Rabhi, P. Vieu, Conditional quantiles for dependent functional data with application to the climatic *El Niño* phenomenon, *Sankhyā: The Indian Journal of Statistics* 67 (2005) 378–398.
- [9] F. Ferraty, P. Vieu, *Nonparametric Functional Data Analysis*, Springer, New York, 2006.
- [10] J. Gertheiss, A. Maity, A.-M. Staicu, Variable selection in generalized functional linear models, *Stat* 2 (2013) 86–101.
- [11] P. Hall, M. Hosseini-Nasab, On properties of functional principal components analysis, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68 (2006) 109–126.
- [12] P. Hall, M. Hosseini-Nasab, Theory for high-order bounds in functional principal components analysis, *Mathematical Proceedings of the Cambridge Philosophical Society* 146 (2009) 225–256.
- [13] P. Hall, H.-G. Müller, J.-L. Wang, Properties of principal component methods for functional and longitudinal data analysis, *The Annals of Statistics* 34 (2006) 1493–1517.
- [14] X. He, Q.-M. Shao, A general Bahadur representation of M -estimators and its application to linear regression with nonstochastic designs, *The Annals of Statistics* 24 (1996) 2608–2630.
- [15] W. Hendricks, R. Koenker, Hierarchical spline models for conditional quantiles and the demand for electricity, *Journal of the American statistical Association* 87 (1992) 58–68.
- [16] L. Horvath, P. Kokoszka, M. Reimherr, Two sample inference in functional linear models, *The Canadian Journal of Statistics* 37 (2009) 571–591.
- [17] L. Huang, F. Scheipl, J. Goldsmith, J. Gellar, J. Harezlak, M. W. McLean, B. Swihart, L. Xiao, C. Crainiceanu, P. Reiss, refund: Regression with Functional Data, 2015. R package version 0.1-13.
- [18] P. J. Huber, The behavior of maximum likelihood estimates under nonstandard conditions, in: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I: Statistics, University of California Press: Berkeley, CA, USA, 1967, pp. 221–233.
- [19] A. E. Ivanescu, A.-M. Staicu, F. Scheipl, S. Greven, Penalized function-on-function regression, *Computational Statistics* 30 (2015) 539–568.

- [20] C.-R. Jiang, J.-L. Wang, Covariate adjusted functional principal components analysis for longitudinal data., *The Annals of Statistics* 38 (2010) 1194–1226.
- [21] L. Jiang, H. D. Bondell, H. Wang, Interquantile shrinkage and variable selection in quantile regression, *Computational Statistics & Data Analysis* 69 (2014) 208–219.
- [22] K. Kato, Estimation in functional linear quantile regression, *The Annals of Statistics* 40 (2012) 3108–3136.
- [23] K. Knight, Limiting distributions for L_1 regression estimators under general conditions, *The Annals of Statistics* 26 (1998) 755–770.
- [24] R. Koenker, A note on L-estimates for linear models, *Statistics & Probability Letters* 2 (1984) 323–325.
- [25] R. Koenker, *Quantile Regression*, volume 38, Cambridge university press, 2005.
- [26] D. Kong, A.-M. Staicu, A. Maity, Classical testing in functional linear models, *Journal of Nonparametric Statistics* 28 (2016) 813–838.
- [27] J. Larsen, Policy institute: Bike-sharing programs hit the streets in over 500 cities worldwide, http://www.earth-policy.org/plan_b_updates/2013/update112, 2013.
- [28] E. R. Lee, H. Noh, B. U. Park, Model selection via Bayesian information criterion for quantile regression models, *Journal of the American Statistical Association* 109 (2014) 216–229.
- [29] M. Li, A.-M. Staicu, H. D. Bondell, Incorporating covariates in skewed functional data models, *Biostatistics* 16 (2015) 413–426.
- [30] Y. Li, Y. Liu, J. Zhu, Quantile regression in reproducing kernel Hilbert spaces, *Journal of the American Statistical Association* 102 (2007) 255–268.
- [31] Y. Li, N. Wang, R. J. Carroll, Generalized functional linear models with semiparametric single-index interactions, *Journal of the American Statistical Association* 105 (2010) 621–633.
- [32] Y. Li, N. Wang, R. J. Carroll, Selecting the number of principal components in functional data, *Journal of the American Statistical Association* 108 (2013) 1284–1294.
- [33] J. S. Morris, R. J. Carroll, Wavelet-based functional mixed models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68 (2006) 179–199.
- [34] D. Pollard, Asymptotics for least absolute deviation regression estimators, *Econometric Theory* 7 (1991) 186–199.
- [35] J. Ramsay, B. Silverman, *Functional Data Analysis*, Springer Series in Statistics, Springer, 2005.
- [36] A. Redd, A comment on the orthogonalization of B-spline basis functions and their derivatives, *Statistics and Computing* 22 (2012) 251–257.
- [37] G. Shi, J. Du, Z. Sun, Z. Zhang, Checking the adequacy of functional linear quantile regression model, *Journal of Statistical Planning and Inference* 210 (2021) 64–75.
- [38] A.-M. Staicu, C. M. Crainiceanu, D. S. Reich, D. Ruppert, Modeling functional data with spatially heterogeneous shape characteristics, *Biometrics* 68 (2012) 331–343.
- [39] A.-M. Staicu, S. N. Lahiri, R. J. Carroll, Significance tests for functional data with complex dependence structure, *Journal of Statistical Planning and Inference* 156 (2015) 1–13.
- [40] Y.-R. Su, C.-Z. Di, L. Hsu, Hypothesis testing in functional linear models, *Biometrics* 73 (2017) 551–561.
- [41] J. Usset, A.-M. Staicu, A. Maity, Interaction models for functional regression, *Computational Statistics & Data Analysis* 94 (2016) 317–330.
- [42] H. J. Wang, L. A. Stefanski, Z. Zhu, Corrected-loss estimation for quantile regression with covariate measurement errors, *Biometrika* 99 (2012) 405–421.
- [43] K. Wang, H. J. Wang, Optimally combined estimation for tail quantile regression, *Statistica Sinica* 26 (2016) 295–311.
- [44] Y. Wei, R. J. Carroll, Quantile regression with measurement error, *Journal of the American Statistical Association* 104 (2009) 1129–1143.
- [45] Y. Wu, Y. Ma, G. Yin, Smoothed and corrected score approach to censored quantile regression with measurement errors, *Journal of the American Statistical Association* in press (2015).
- [46] F. Yao, H.-G. Müller, J.-L. Wang, Functional data analysis for sparse longitudinal data, *Journal of the American Statistical Association* 100 (2005) 577–590.
- [47] F. Yao, S. Sue-Chee, F. Wang, Regularized partially functional quantile regression, *Journal of Multivariate Analysis* 156 (2017) 39–56.
- [48] J.-T. Zhang, J. Chen, Statistical inferences for functional data, *The Annals of Statistics* 35 (2007) 1052–1079.
- [49] Z. Zhao, Z. Xiao, Efficient regressions via optimally combining quantile information, *Econometric Theory* 30 (2014) 1272–1314.
- [50] L. Zhou, J. Z. Huang, R. J. Carroll, Joint modelling of paired sparse functional data using principal components, *Biometrika* 95 (2008) 601–619.
- [51] H. Zhu, F. Yao, H. H. Zhang, Structured functional additive regression in reproducing kernel hilbert spaces, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76 (2014) 581–603.
- [52] H. Zou, M. Yuan, Composite quantile regression and the oracle model selection theory, *The Annals of Statistics* 36 (2008) 1108–1126.