应用概率统计 第 33 卷
第 2 期 2017 年 4 月

Chinese Journal of Applied Probability and Statistics
Apr., 2017, Vol. 33, No. 2, pp. 170-190
doi: 10.3969/j.issn.1001-4268.2017.02.005

Estimation in Functional Partial Linear Composite Quantile Regression Model *

YU Ping[1,2] ZHANG ZhongZhan[1] DU Jiang[1]

([1] *College of Applied Sciences, Beijing University of Technology, Beijing, 100124, China*)

([2] *School of Mathematics and Computer Science, Shanxi Normal University, Linfen, 041000, China*)

Abstract: This paper studies estimation in functional partial linear composite quantile regression model in which the dependent variable is related to both a function-valued random variable in linear form and a real-valued random variable in nonparametric form. The functional principal component analysis and regression splines are employed to estimate the slope function and the nonparametric function respectively, and the convergence rates of the estimators are obtained under some regularity conditions. Simulation studies and a real data example are presented for illustration of the performance of the proposed estimators.

Keywords: functional data analysis; spline estimate; composite quantile regression; functional principal component analysis

2010 Mathematics Subject Classification: 62G08

Citation: Yu P, Zhang Z Z, Du J. Estimation in functional partial linear composite quantile regression model [J]. Chinese J. Appl. Probab. Statist., 2017, 33(2): 170–190.

§1. Introduction

Nowadays, the great progress of computational tools (both memory and capacity increasing) allows to create, store and work with large databases. In many cases, the dataset is made up of observations from a finite dimensional distribution, measured over a period of time or recorded at different spatial locations. When the temporal or spatial grid is fine enough, the sample can be considered as an observation of a random variable on a certain functional space. Analyzing this kind of data with standard multivariate methods and ignoring its functional feature may fail dramatically because of the curse of dimensionality, collinearity, or valuable information loss. Thus functional data analysis

*Yu and Zhang's research was partly supported by the National Natural Science Foundation of China (Grant No. 11271039), Education Ministry Funds for Doctor Supervisors and Fund from Collaborative Innovation Center on Capital Social Construction and Social Management (Grant No. 006000546615539). Yu and Du's research was supported by the National Natural Science Foundation of China (Grant No. 11501018).

Received November 17, 2015. Revised August 15, 2016.

(FDA) has been turned into one of the most active statistical fields in recent years. An overview of the basic methods of functional data analysis, computational aspects related with their practical application and important real data modeling can be seen in the pioneers books by [1, 2]. A detailed study on nonparametric FDA methodologies was developed in [3]. Statistical inference related with some FDA methods was recently studied in [4].

We also note that there are numerous works of functional data literature on functional linear regression models, see, e.g., [5–11] and references therein. However, relatively few studies form a quantile regression perspective. Quantile regression has been widely used since the seminal work of [12]. It is attractive not only by virtue of its robustness with respect to non-Gaussian errors but also because, by considering several quantiles simultaneously, it can provide a more complete picture of the conditional distribution of the response. Cardot et al.[13] studied smoothing splines estimators for functional nonparametric quantile regression models. More recently, Chen and Müller[14] developed a method for conditional quantile analysis. They first estimated the conditional distribution function under a generalized functional regression framework and then estimated the conditional quantile function by inverting the estimated conditional distribution function and derived the consistency of their estimator. Kato[15] studied estimation in functional linear quantile regression and established rates of convergence for proposed estimators and showed that these rates were optimal in a minimax sense under some smoothness assumptions on covariance kernel of the predictor and the slope function. Lu et al.[16] showed asymptotic normality of the estimator of the finite dimensional parameter vector and rate of convergence of estimator of the infinite dimensional slope function in functional partially linear quantile regression model. The references mentioned above are concerned mainly with quantile in functional linear models. We note that Zou and Yuan[17] proposed a composite quantile regression technique to combine information across different quantiles in a linear regression model. More references about composite quantile regression refers to [18–20] among others. Quantile regression can be treated as a special case of composite quantile regression. Therefore, we consider in this paper functional partial linear composite quantile regression models, which include not only nonparametric quantile regression model but also the models in [15] and [16] as special cases.

The rest of paper is organized as follows. In Sections 2, we introduce the functional linear partial regression model, and in Section 3, we present the functional principal component regression and spline-based method to estimate the function coefficient and nonparametric function, respectively. Section 4 gives the rates of convergence of the estimators under some regularity conditions. Section 5 discusses the tuning parameter selection problem and the computing algorithm for the proposed estimates. Some simulation results and an example are presented in Section 6. All proofs are given in the Appendix.

§2. Functional Partial Linear Regression Model

Let Y and Z be real-valued random variables defined on a probability space $(\Omega, \mathscr{B}, \mathsf{P})$, and let $\{X(t)\}$ be a second-order stochastic process valued in $H = L^2([0,1])$, the Hilbert space containing square integrable functions on $[0,1]$ with inner product $\langle x, y\rangle = \int_0^1 x(t) \cdot y(t)\mathrm{d}t$, $\forall\, x, y \in H$ and norm $\|x\| = \langle x, x\rangle^{1/2}$. The dependence between Y and $\{X, Z\}$ is expressed as

$$Y = \int_0^1 \beta_0(t)X(t)\mathrm{d}t + g_0(Z) + \epsilon, \tag{1}$$

where the functional coefficient $\beta_0(\cdot)$ belongs to H, $g_0(\cdot)$ is an unknown univariate smooth function of auxiliary variable Z, ϵ is a random error variable with median 0, and independent of $\{X, Z\}$. Model (1) generalizes both the nonparametric regression model and the functional linear model which correspond to the cases $\beta_0 = 0$ and $g_0 = 0$, respectively. Note that model (1) also includes the partial functional linear regression model proposed by [21] when $g_0(Z) = \boldsymbol{U}^\top\boldsymbol{\theta}$. Without loss of generality, we further assume that $Z \in [0,1]$, and X is centered, that is to say $\mathsf{E}(X(t)) = 0$, for all $t \in [0,1]$.

§3. Estimation Method

In this section, we will describe how to estimate the functional coefficient $\beta(\cdot)$ and nonparametric function $g(\cdot)$. As is discussed in [22], the functional principal component analysis (FPCA) is a benchmark basis for functional data. Furthermore, regression splines have some desirable properties in approximating a smooth function, and often provide good approximations with small number of knots. For these reasons, we approximate the functional coefficient by FPCA and nonparametric function by regression splines in the following.

Let (Z_i, X_i, Y_i), $i = 1, 2, \ldots, n$ be independent realizations of (Z, X, Y) generated by model (1). Define the covariance function and the empirical covariance function for $X(\cdot)$ respectively as

$$C(t,s) = \mathsf{Cov}\,(X(t), X(s)), \qquad \widehat{C}(t,s) = \frac{1}{n}\sum_{i=1}^{n} X_i(t)X_i(s).$$

The covariance function $C(t,s)$ defines a linear operator which maps a function f to Cf given by $Cf(s) = \int C(t,s)f(t)\mathrm{d}t$. We shall assume that the linear operator with kernel $C(t,s)$ is positive definite. Write the spectral decomposition of the covariance function

$C(t,s)$ and $\widehat{C}(t,s)$ as

$$C(t,s)=\sum_{i=1}^{\infty}\lambda_i v_i(t)v_i(s), \qquad \widehat{C}(t,s)=\sum_{i=1}^{\infty}\widehat{\lambda}_i\widehat{v}_i(t)\widehat{v}_i(s),$$

where

$$\lambda_1>\lambda_2>\cdots>0, \qquad \widehat{\lambda}_1\geqslant\widehat{\lambda}_2\geqslant\cdots\geqslant\widehat{\lambda}_{n+1}=\cdots=0$$

are respectively the ordered eigenvalue sequences of the linear operators with kernels $C(t,s)$ and $\widehat{C}(t,s)$, $\{v_i\}$ and $\{\widehat{v}_i\}$ are the corresponding orthonormal eigenfunction sequences. The sequence $\{v_i\}$ forms an orthonormal basis in H. Then, we have the following expansions in H:

$$X(t)=\sum_{i=1}^{\infty}\xi_i v_i(t), \qquad \beta_0(t)=\sum_{j=1}^{\infty}\beta_{0j}v_j(t),$$

where ξ_i and β_{0j} are defined by

$$\xi_i=\int_0^1 X(t)v_i(t)\mathrm{d}t, \qquad \beta_{0j}=\int_0^1 \beta_0(t)v_j(t)\mathrm{d}t,$$

ξ_i is referred to as the ith functional principal component score. It follows that $\xi_i s$ are uncorrelated random variables with mean zero and variance $\mathsf{Var}\,(\xi_i)=\lambda_i$, and

$$\langle\beta_0(\cdot),X(\cdot)\rangle=\sum_{j=1}^{\infty}\beta_{0j}\xi_j. \tag{2}$$

We shall use (2) to estimate $\beta_0(t)$ in the following. Now we turn to function $g_0(\cdot)$ in model (1). Following [23], we approximate $g(\cdot)$ by splines. Let $0=z_0<z_1<\cdots<z_{k_n}=1$ be a partition of the interval $[0,1]$. Let $N=k_n+l+1$ and $\pi_1(Z),\pi_2(Z),\cdots,\pi_N(Z)$ be the normalized B-splines of order $l+1$ based on the knot mesh $\{z_i\}$ that form a basis for the linear spline space. For the asymptotic theory in next section, we assume that the order $l+1$ is fixed but k_n depends on the sample size n. If the nonparametric function $g_0(\cdot)$ is sufficiently smooth, then it can be approximated as

$$g_0(\cdot)\approx\sum_{s=1}^{N}\alpha_{0s}\pi_s(\cdot). \tag{3}$$

Inserting (2) and (3) into (1), we get

$$Y\approx\sum_{j=1}^{m}\beta_{0j}\xi_j+\sum_{s=1}^{N}\alpha_{0s}\pi_s(Z)+\epsilon, \tag{4}$$

where the number of included components $m=m(n)$ and the number of spline basis $N=N(n)$ need to satisfy $1\leqslant m+N\leqslant n-1$, $m=m(n)\to\infty$ and $N=N(n)\to\infty$ as the sample size $n\to\infty$. Let $0<\tau_1<\tau_2<\cdots<\tau_K<1$, b_{0k} be the τ_k quantile of ϵ.

The estimates $\widehat{\beta}(t)=\sum\limits_{j=1}^{m}\widehat{\beta}_j\widehat{v}_j(t)$ and $\widehat{g}(Z)=\sum\limits_{s=1}^{N}\widehat{\alpha}_s\pi_s(Z)$ are obtained by solving following composite quantile regression (CQR) [17]:

$$\min\sum_{k=1}^{K}\sum_{i=1}^{n}\rho_{\tau_k}\Big(Y_i-b_k-\sum_{j=1}^{m}\beta_j\widehat{\xi}_{ij}-\sum_{s=1}^{N}\alpha_s\pi_s(Z_i)\Big),\tag{5}$$

with respect to β_j, $j=1,2,\cdots,m$, α_s, $s=1,2,\ldots,N$ and b_k, $k=1,2,\ldots,K$, where $\rho_\tau(u)=u(\tau-I(u<0))$ is the quantile loss function [24] and $\widehat{\xi}_{ij}=\langle X_i,\widehat{v}_j\rangle$. The objective function in (5) is a mixture of the objective functions from different quantile regression models. Typically, we use the equally spaced quantiles: $\tau_k=k/(K+1)$ for $k=1,2,\cdots,K$.

§4. Asymptotic Results

In this section, we derive consistency for the proposed estimators. Let F_ϵ be the cumulative distribution function and f_ϵ be the density function of ϵ_i conditional on (X_i,Z_i). For a vector $\boldsymbol{v}$ in R^d we let $|\boldsymbol{v}|$ be its Euclidian norm. In the remaining part of the paper, we denote by C a generic positive constant, which may take different values at different places, and use $a_n\sim b_n$ to express that a_n/b_n is bounded away from zero and infinity as $n\to\infty$. We will make the following assumptions:

C1. $\mathsf{E}\|X\|^4\leqslant C$ and $\mathsf{E}[\xi_j^4]\leqslant C\lambda_j^2$ for all $j\geqslant 1$.

C2. For some $a>1$, $C^{-1}j^{-a}\leqslant\lambda_j\leqslant Cj^{-a}$ and $\lambda_j-\lambda_{j+1}\geqslant C^{-1}j^{-a-1}$ for all $j\geqslant 1$.

C3. For some $b>a/2+1$, $|\beta_j|\leqslant Cj^{-b}$ for all $j\geqslant 1$.

C4. There exist some $r=p+v\geqslant 1$ and a constant $C\in(0,\infty)$ such that $|g^{(p)}(s)-g^{(p)}(t)|\leqslant C|s-t|^v$, for any $0\leqslant s,t\leqslant 1$.

C5. The density function $f_Z(z)$ of random variable Z has a compact support $[0,1]$ and $f_Z(z)$ is bounded away from zero and infinity on $[0,1]$.

C6. f_ϵ is bounded from infinity in whole support, and bounded away from zero at all b_{0k}s, and has bounded first derivatives in the neighborhoods of $b_{0k}s$.

C7. The eigenvalues of $\boldsymbol{\Lambda}_n/n$ are bounded away from infinity and zero for sufficiently large n, where $\boldsymbol{\Lambda}_n$ is defined in the Appendix.

Remark 1 C1 – C3 are very common in the functional linear regression model [7,8]. Specially, C1 ensures the consistency of $\widehat{C}(t,s)$, C2 prevents the spacings among eigenvalues being too small for identification of the slope function $\beta(t)$, and C3 makes the slope function sufficiently smooth relative to the covariance function $C(t,s)$. C4 states the smoothness condition on $g(\cdot)$, which describes a requirement on the best convergence rate that $g(\cdot)$

can be estimated. C5 is very common in nonparametric regression [25]. C6 is a standard assumption in quantile regression [26]. C7 is parallel to the condition A6 in [36], which is used to obtain the convergence rate of the estimators.

Theorem 2 Suppose that conditions C1 – C7 hold, $m \sim n^{1/(a+2b)}$ and $N \sim n^{1/(2r+1)}$, we have

$$\|\widehat{\beta}(\cdot) - \beta_0(\cdot)\|^2 = O_p(n^{-(2b-1)/(a+2b)}) + O_p(n^{-2b/(a+2b)+1/(2r+1)}),$$
$$\|\widehat{g}(\cdot) - g_0(\cdot)\|^2 = O_p(n^{-2r/(2r+1)}) + O_p(n^{-(a+2b-1)/(a+2b)}).$$

The proofs of following three corollaries are similar to Theorem 2, and thus omitted.

Corollary 3 Suppose that conditions C1 – C7 hold, if $m \sim N \sim n^{1/(2r+1)}$ and $1/2 + a < r \leqslant (a + 2b - 1)/2$, we have

$$\|\widehat{\beta}(\cdot) - \beta_0(\cdot)\|^2 = O_p(n^{-(2r-a)/(2r+1)}), \qquad \|\widehat{g}(\cdot) - g_0(\cdot)\|^2 = O_p(n^{-2r/(2r+1)}).$$

Corollary 4 Suppose that conditions C1 – C7 hold, if $m \sim N \sim n^{1/(a+2b)}$ and $r > (a + 2b - 1)/2$, we have

$$\|\widehat{\beta}(\cdot) - \beta_0(\cdot)\|^2 = O_p(n^{-(2b-1)/(a+2b)}), \qquad \|\widehat{g}(\cdot) - g_0(\cdot)\|^2 = O_p(n^{-(a+2b-1)/(a+2b)}).$$

Corollary 5 Suppose that conditions C1 – C7 hold, if $m \sim N \sim n^{1/(a+2b)} \sim n^{1/(2r+1)}$, we have

$$\|\widehat{\beta}(\cdot) - \beta_0(\cdot)\|^2 = O_p(n^{-(2b-1)/(a+2b)}), \qquad \|\widehat{g}(\cdot) - g_0(\cdot)\|^2 = O_p(n^{-2r/(2r+1)}).$$

The result of Corollary 5 indicates that the estimator $\widehat{\beta}$ has the same rate of convergence as the estimators of [8] and [15], which are optimal in minimax sense. The rate of convergence for $\widehat{g}$ is the same as the optimal global convergence rate established by [27].

§5. Implementation

In this section, we shall discuss the selection of tuning parameters, and introduce computing algorithm to minimize (5) in Section 3.

5.1 Tuning Parameter Selection

Throughout our numerical studies, we use B-splines of order 4 to approximate the nonparametric component, and following the idea of [28], let N be the integer part of

$n^{1/5}$. In addition, we need to know how to choose tuning parameter m. Similar to [16], we choose m as the minimizer to the following Schwarz-type information criterion

$$\mathrm{BIC}(m) = \ln\Big(\sum_{k=1}^{K}\sum_{i=1}^{n}\rho_{\tau_k}\Big(Y_i - \widehat{b}_k - \sum_{j=1}^{m}\widehat{\beta}_j\widehat{\xi}_{ij} - \sum_{s=1}^{N}\widehat{\alpha}_s\pi_s(Z_i)\Big)\Big) + \frac{\ln n}{2n}(m + l + k_n + 1),$$

where $l = 3$, $\widehat{\beta}_j$ and $\widehat{\alpha}_s$ are the CQR estimators obtained from minimizing (5) with m eigenfunctions. We also can refer to [26] for a similar criterion for tuning parameters selection.

5.2 Computing Algorithm

Similar to [29], we propose the following computing algorithm to estimate the parameters in (5). Specifically, the optimization problem of the CQR method can be formulated as a linear programming (LP) problem. To derive the LP formulation, we introduce $(2n) \times K$ slack variables $\{(u_{ik}^{+}, u_{ik}^{-}),\ i = 1, 2, \ldots, n,\ k = 1, 2, \ldots, K)\}$ that satisfy the equality constraints $Y_i - b_k - \sum\limits_{j=1}^{m}\beta_j\widehat{\xi}_{ij} - \sum\limits_{s=1}^{N}\alpha_s\pi_s(Z_i) = u_{ik}^{+} - u_{ik}^{-}$, where $u_{ik}^{+} \geqslant 0$ and $u_{ik}^{-} \geqslant 0$ represent the positive and negative parts of $Y_i - b_k - \sum\limits_{j=1}^{m}\beta_j\widehat{\xi}_{ij} - \sum\limits_{s=1}^{N}\alpha_s\pi_s(Z_i)$. Similarly, write $b_k = b_k^{+} - b_k^{-}$, $(b_k^{+} \geqslant 0,\ b_k^{-} \geqslant 0)$, $\beta_j = \beta_j^{+} - \beta_j^{-}$, $(\beta_j^{+} \geqslant 0,\ \beta_j^{-} \geqslant 0)$ and $\alpha_s = \alpha_s^{+} - \alpha_s^{-}$, $(\alpha_s^{+} \geqslant 0,\ \alpha_s^{-} \geqslant 0)$. Then the CQR estimators in (5) can be obtained by minimizing

$$\sum_{k=1}^{K}\sum_{i=1}^{n}\tau_k u_{ik}^{+} + (1 - \tau_k)u_{ik}^{-} \tag{6}$$

subject to

$$\begin{aligned}
&b_k^{+} - b_k^{-} + u_{ik}^{+} - u_{ik}^{-} + \sum_{j=1}^{m}(\beta_j^{+} - \beta_j^{-})\widehat{\xi}_{ij} + \sum_{s=1}^{N}(\alpha_s^{+} - \alpha_s^{-})\pi_s(Z_i) = Y_i,\\
&b_k^{+} \geqslant 0, \quad b_k^{-} \geqslant 0, \quad \beta_j^{+} \geqslant 0, \quad \beta_j^{-} \geqslant 0, \quad \alpha_s^{+} \geqslant 0, \quad \alpha_s^{-} \geqslant 0,\\
&u_{ik}^{+} \geqslant 0, \quad u_{ik}^{-} \geqslant 0, \quad 1 \leqslant i \leqslant n, \quad 1 \leqslant k \leqslant K, \quad 1 \leqslant j \leqslant m, \quad 1 \leqslant s \leqslant m.
\end{aligned}$$

The R lpSolve package can be used to implement the above LP problem.

§6. Numerical Examples

6.1 Simulation Studies

In this subsection we investigate the finite sample performance of the proposed estimation methods with Monte Carlo simulation studies. The data sets are generated from the following model:

$$Y = \int_0^1 \beta(t)X(t)\mathrm{d}t + g(Z) + \epsilon. \tag{7}$$

For the functional linear component, the design is similar to [21], that is, the functional coefficient $\beta(t) = \sqrt{2}\sin(\pi t/2) + 3\sqrt{2}\sin(3\pi t/2)$ and $X(t) = \sum_{j=1}^{50} \xi_j v_j(t)$, where the ξ_js are distributed as the independent normal with mean 0 and variance $\lambda_j = ((j-0.5)\pi)^{-2}$, $v_j(t) = \sqrt{2}\sin((j-0.5)\pi t)$. The nonparametric component is taken as $g(Z) = \sin(2\pi Z)$, and the random variable Z is distributed uniformly on $[0,1]$. The following three distributions of random error are considered to compare the CQR estimators and the least squares (LS) estimators: (1) ϵ follows $N(0,1)$ normal distribution; (2) ϵ follows $t(3)$ distribution; (3) ϵ follows standard Cauchy distribution. The latter yields a model in which the expectation of the response does not exist. We take $\tau_k = k/8$, $k = 1, 2, \ldots, 7$, k_n as the integer part of $n^{1/5}$, and $l = 3$ in the simulation.

In addition to the comparison of CQR with the LS regression, we also compare the B-spline approximation used in previous section with local linear approximation [30] to the nonparametric function $g(Z)$ in the composite quantile regression model (LCQR). In the simulation, the bandwidth $h = 0.5n^{1/5}$ and biweight kernel function are used for the local linear approximation.

To assess the performance of estimates for slope function $\beta(\cdot)$ and nonparametric function $g(\cdot)$, we employ the following error criteria [10]:

$$\mathrm{MSE}_1 = \sum_{k=1}^{n_1} (\widehat{\beta}(t_k) - \beta(t_k))^2/n_1, \qquad \mathrm{MSE}_2 = \sum_{i=1}^{n_2} (\widehat{g}(z_i) - g(z_i))^2/n_2,$$

where $\{t_k, k = 1, 2, \ldots, n_1\}$ and $\{z_i, i = 1, 2, \ldots, n_2\}$ are grid points chosen to be equally spaced in the domains of function $\beta(\cdot)$ and $g(\cdot)$, respectively. $n_1 = n_2 = 200$ are used.

Table 1 MSEs of $\beta(\cdot)$ and $g(\cdot)$

	n	N(0,1)			t(3)			Cauchy(0,1)		
		LS	CQR	LCQR	LS	CQR	LCQR	LS	CQR	LCQR
$\beta(\cdot)$	100	0.419	0.442	0.438	0.937	0.601	0.593	–	1.358	1.188
	200	0.204	0.215	0.197	0.425	0.280	0.274	–	0.519	0.491
$g(\cdot)$	100	0.066	0.070	0.065	0.193	0.108	0.102	–	0.256	2.170
	200	0.032	0.034	0.035	0.087	0.047	0.055	–	0.108	0.232

Based on 1 000 repetitions, Table 1 summarizes the MSEs of the estimated $\beta(\cdot)$ and $g(\cdot)$ with different sample sizes under three kinds of random error. In the table, "–" means the value is much larger than others and we do not report them. We can see that the MSEs decrease as the sample size increases. As is anticipated, the CQR is slightly less efficient than the LS regression for normal distributed errors, but stays robust when the distribution tails of error become heavier, while the LS estimator breaks down for Cauchy distributed errors. This verifies that the CQR estimation is necessary for the error with heavy tails.

The comparison between CQR and LCQR is more or less complicated. Unexpectedly, we find that the estimator of the slope obtained from LCQR performs better than that from CQR. However, the local linear estimator of the nonparametric component may be more affected by heavier tails of the errors, although it is similar to the B-spline estimator for both normal distributed errors and $t(3)$ distributed errors. We also note that the bandwidth for local linear approximation is fixed to $h = 0.5n^{1/5}$ and not optimized considering the computation time of the simulation. In fact, the computing times of the LCQR simulation with R on a PC (processor: Intel(R)Core(TM)2 i5-3470 CPU @ 3.20 GHz, 3.47 GB) takes 8.7 hours for one case, which is about 8 times of CQR.

Figures 1−3 demonstrate the average curve estimates of slope parameter $\beta(\cdot)$ and nonparametric function $g(\cdot)$ over the 1 000 repetitions for three error cases with $n = 200$ (Figure 3 does not show the failed LS estimator). The plots for $n = 100$ are very similar, thus are omitted to save space.

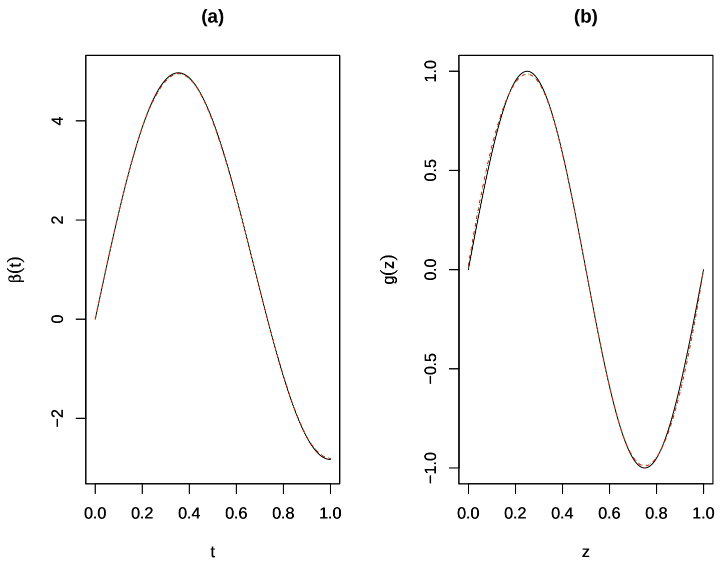

Figure 1 The true β and g and their estimators for $\mathrm{N}(0, 1)$ random error with $n = 200$: true-solid line, LS-dashed line, CQR-dotted line

6.2 Application to Spectra Data

In this subsection, we apply our proposed CQR estimation to the tecator data, which is available at http://lib.stat.cmu.edu/datasets/tecator, and has been widely used in the

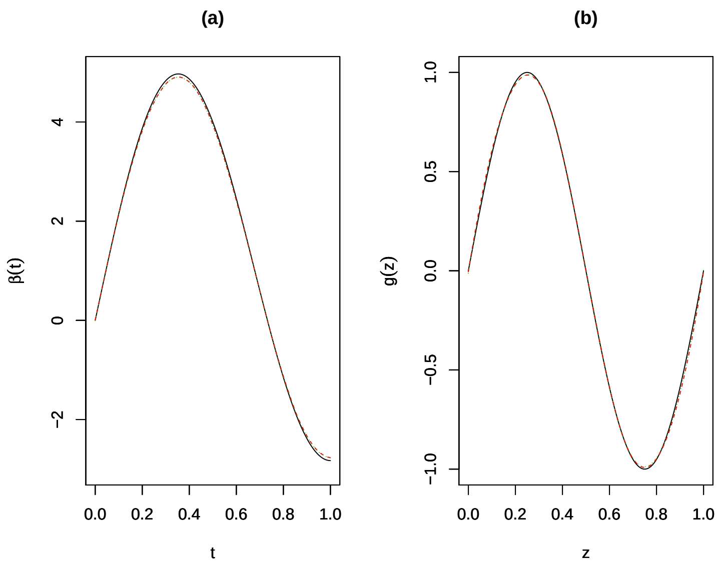

Figure 2 The true β and g and their estimators for $\mathrm{t}(3)$ random error with $n=200$: true-solid line, LS-dashed line, CQR-dotted line

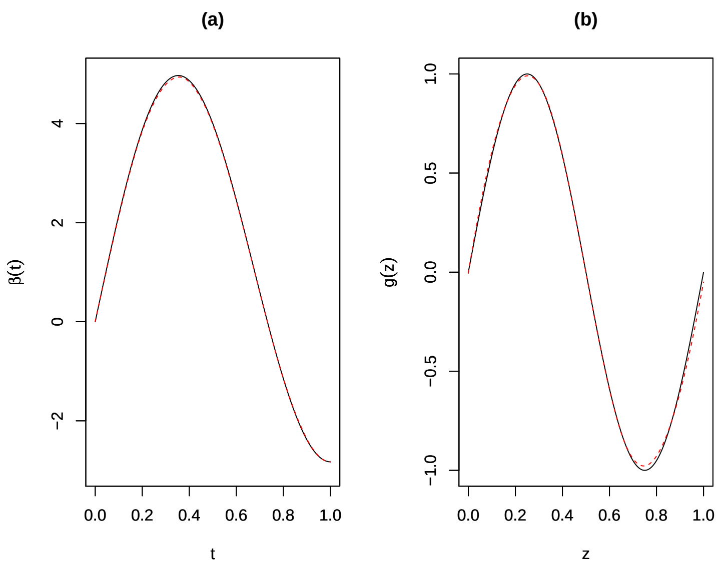

Figure 3 The true β and g and their estimators for standard Cauchy random error with $n=200$: true-solid line, CQR-dotted line

context of functional data[3,31]. For each food sample, the spectrum of the absorbance recorded on a Tecator Infratec Food and Feed Analyzer working in the wavelength range $850-1\,050$ nm by the near-infrared transmission (NIT) principle is provided also with the fat, protein, and moisture contents, measured in percent and determined by analytic chemistry.

We denote the fat content as Y_i, the protein content as Z_i, the moisture content as U_i, and the absorbance as $X_i(t)$. $X_i(t) = \log_{10}(p_0/p)$, where t is the wavelegenth of the light, p_0 is the intensity of the light before passing through the meat sample, and p is the intensity of the light after it passes through the meat sample. These absorbances can be took as a discrete approximation to the continuous record, $X_i(t)$. We use B-splines to convert discrete grid form of $X_i(t)$ to the functional form and fix the number of knots to be 100. The degree of spline functions has been chosen to be 4. To evaluate the performance of the models, similar to [32], the sample is divided into two groups: the training sample, $I_1 = \{(X_i, Y_i, Z_i, U_i), i = 1, 2, \ldots, 165\}$, and the test sample $I_2 = \{(X_i, Y_i, Z_i, U_i), i = 166, 167, \ldots, 215\}$. The training sample is used to estimate the parameters and the test sample is employed to verify the quality of predictions. For this, we computed the mean square error of prediction (MSEP), which is defined by

$$\text{MSEP} = \frac{1}{50} \sum_{i \in I_2} (Y_i - \widehat{Y}_i)^2 / \mathsf{Var}_{I_2}(Y_i),$$

where $\widehat{Y}_i$ is the predicted value based on the training sample and Var_{I_2} is the sample variance of the response variable from the test sample.

According to Figure 4 and examination of the fat, water and protein data, we can obtain that there are no obvious outliers and the LS estimator can be considered. For comparison, we compute the MSEPs corresponding to the following 4 models:

$$\begin{aligned}
Y &= \int_{850}^{1\,050} X(t)\beta(t)\mathrm{d}t + g_1(U) + \epsilon \quad &&\text{(model 1)},\\
Y &= \int_{850}^{1\,050} X(t)\beta(t)\mathrm{d}t + aU + \epsilon \quad &&\text{(model 2)},\\
Y &= \int_{850}^{1\,050} X(t)\beta(t)\mathrm{d}t + g_2(Z) + \epsilon \quad &&\text{(model 3)},\\
Y &= \int_{850}^{1\,050} X(t)\beta(t)\mathrm{d}t + bZ + \epsilon \quad &&\text{(model 4)}.
\end{aligned}$$

Model 2 and model 4 are belong to partial functional linear regression model [21]. For composite quantile regression, we use $\tau_k = k/10$, $k = 1, 2, \ldots, 9$ in computation. According to Table 2, we can obtain that the results from CQR and LS are very similar, and the relationship between fat content and moisture content or protein content with nonlinearity are more accurate than linearity. Considering prediction performance, MSEP of model 1 is the smallest. This may imply that a nonparametric component of U has a strong effect additive to the linear term of X. Moreover, we can also consider model

$$Y = \int_{850}^{1\,050} X(t)\beta(t)\mathrm{d}t + g_1(U) + g_2(Z) + \epsilon,$$

which intuitively has smaller MSEP than model 1. But this goes beyond the topic of this paper and is a subject of works in progress. In addition, comparing the decrease in MSEP from model 2 to model 1, we may anticipate that the gain of adding g_2 to model 1 would be limited.

Table 2 The MSEPs of different models

	model 1	model 2	model 3	model 4
CQR	0.035	0.516	0.234	0.484
LS	0.036	0.489	0.244	0.464

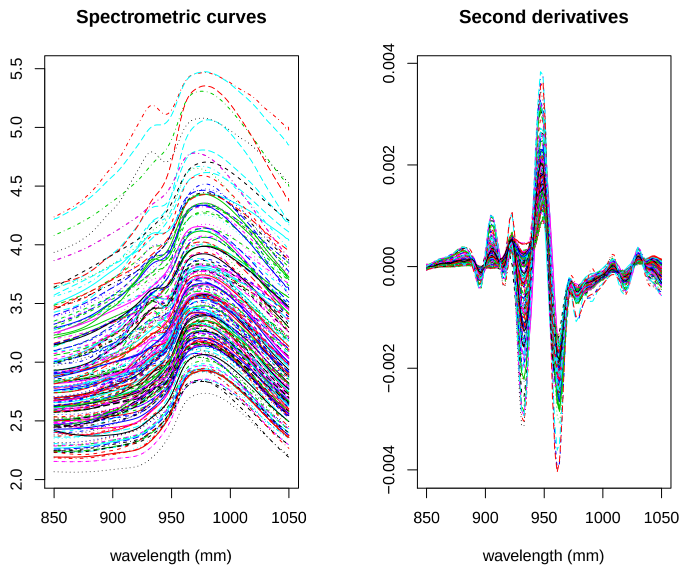

Figure 4 The spectrometric curves and their second derivative curves

Appendix: Proof of Theorems

The following lemma, which follows easily from Theorem 12.7 of [33], is stated for easy reference.

Lemma 6 Under condition C4, there exists a spline function $g^*(\cdot) = \sum\limits_{s=1}^{N} \alpha_{0s}\pi_s(\cdot)$, called spline approximation of $g_0(\cdot)$, such that

$$\sup_{z\in[0,1]} |g^*(z) - g_0(z)| \leqslant CN^{-r}.$$

For convenience, we define $\boldsymbol{U}_i = \{\widehat{\xi}_{i1}, \widehat{\xi}_{i2}, \ldots, \widehat{\xi}_{im}\}^{\mathsf{T}}$, $\boldsymbol{\pi}_i = \{\pi_1(Z_i), \pi_2(Z_i), \ldots, \pi_N(Z_i)\}^{\mathsf{T}}$. $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_n)^{\mathsf{T}}$ and $\boldsymbol{U} = (\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots, \boldsymbol{U}_n)^{\mathsf{T}}$ be the $n \times N$ and $n \times m$ quasi-design matrix for nonparametric function and functional slope parametric component, respectively. We denote by T_n the set of random vectors $((X_1,Z_1),(X_2,Z_2),\ldots,(X_n,Z_n))$, $\boldsymbol{\alpha}_0=\{\alpha_{01},\alpha_{02},\ldots,\alpha_{0N}\}^\mathsf{T}$, $\boldsymbol{\beta}_0=\{\beta_{01},\beta_{02},\ldots,\beta_{0m}\}^\mathsf{T}$, $\boldsymbol{b}_0=\{b_{01},b_{02},\ldots,b_{0K}\}^\mathsf{T}$, $\boldsymbol{\alpha}=\{\alpha_1,\alpha_2,\ldots,\alpha_N\}^\mathsf{T}$, $\boldsymbol{\beta}=\{\beta_1,\beta_2,\ldots,\beta_m\}^\mathsf{T}$, $\boldsymbol{b}=\{b_1,b_2,\ldots,b_K\}^\mathsf{T}$, $\boldsymbol{D}=\mathrm{diag}\{\sqrt{n}f_\epsilon(b_{01}),\sqrt{n}f_\epsilon(b_{02}),\ldots,\sqrt{n}f_\epsilon(b_{0K})\}$, $\boldsymbol{P}=\boldsymbol{\pi}(\boldsymbol{\pi}^\mathsf{T}\boldsymbol{\pi})^{-1}\boldsymbol{\pi}^\mathsf{T}$, $\boldsymbol{H}_n^2=N\boldsymbol{\pi}^\mathsf{T}\boldsymbol{\pi}$, $\boldsymbol{U}^*=(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{U}=(\boldsymbol{U}_1^*,\boldsymbol{U}_2^*,\ldots,\boldsymbol{U}_n^*)^\mathsf{T}$ and $\boldsymbol{\Lambda}_n=\boldsymbol{U}^{*\mathsf{T}}\boldsymbol{U}^*$.

Lemma 7 Under the conditions C1−C7, there exist two constants $C_1,C_2>0$ with probability tending to 1 such that

$$C_1\leqslant\frac{\lambda_{\min}(\boldsymbol{\Lambda}_n/n)}{\lambda_m}\leqslant C_2,$$

where $\lambda_{\min}(\boldsymbol{\Lambda}_n/n)$ denotes the minimum eigenvalue of $\boldsymbol{\Lambda}_n/n$.

Proof of Lemma 7 Observe that

$$\frac{1}{n}\boldsymbol{\Lambda}_n=\frac{1}{n}\boldsymbol{U}^{*\mathsf{T}}\boldsymbol{U}^*=\frac{1}{n}\boldsymbol{U}^\mathsf{T}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{U}=\frac{1}{n}\boldsymbol{U}^\mathsf{T}\boldsymbol{U}-\frac{1}{n}\boldsymbol{U}^\mathsf{T}\boldsymbol{P}\boldsymbol{U}.$$

Let $\boldsymbol{a}\in R^m$, satisfying $\boldsymbol{a}^\mathsf{T}\boldsymbol{a}=1$, then

$$\begin{aligned}\frac{1}{n}\boldsymbol{a}^\mathsf{T}\boldsymbol{U}^\mathsf{T}\boldsymbol{U}\boldsymbol{a}&=\frac{1}{n}\sum_{l=1}^m\sum_{k=1}^m a_l\sum_{i=1}^n\langle X_i,\widehat{v}_l\rangle\langle X_i,\widehat{v}_k\rangle a_k\\&=\frac{1}{n}\sum_{l=1}^m\sum_{k=1}^m\sum_{i=1}^n a_l\langle X_i,\widehat{v}_l\rangle\langle X_i,\widehat{v}_k\rangle a_k\\&=\frac{1}{n}\sum_{i=1}^n\Big\langle X_i,\sum_{l=1}^m a_l\widehat{v}_l\Big\rangle\Big\langle X_i,\sum_{k=1}^m a_k\widehat{v}_k\Big\rangle\\&=\frac{1}{n}\sum_{i=1}^n\int_0^1\int_0^1X_i(t)X_i(s)\sum_{l=1}^m a_l\widehat{v}_l(t)\sum_{k=1}^m a_k\widehat{v}_k(s)\mathrm{d}t\mathrm{d}s\\&=\int_0^1\int_0^1\widehat{C}(t,s)\sum_{l=1}^m a_l\widehat{v}_l(t)\sum_{k=1}^m a_k\widehat{v}_k(s)\mathrm{d}t\mathrm{d}s\\&=\sum_{l=1}^m\sum_{k=1}^m a_la_k\int_0^1\int_0^1\widehat{C}(t,s)\widehat{v}_l(t)\widehat{v}_k(s)\mathrm{d}t\mathrm{d}s\\&=\sum_{l=1}^m\widehat{\lambda}_la_l^2.\end{aligned}$$

Therefore, $\widehat{\lambda}_m\leqslant\sum\limits_{l=1}^m\widehat{\lambda}_la_l^2\leqslant\widehat{\lambda}_1$. According to $\|\lambda_j-\widehat{\lambda}_j\|^2=O_p(n^{-1})$ (see, e.g., Theorem 2.7 of [4]), we have

$$C_1\leqslant\frac{\lambda_{\min}(\boldsymbol{U}^\mathsf{T}\boldsymbol{U}/n)}{\lambda_m}\leqslant C_2.$$

Invoking Lemma 5.1 of [34] and Theorem 4 of [25], we see that the eigenvalues of $n^{-1}N\boldsymbol{\pi}^\mathsf{T}\boldsymbol{\pi}$ are bounded away from zero and infinity. Since $\boldsymbol{I}-\boldsymbol{P}$ is the orthogonal projection matrix, by Theorem 5.9 of [35], we have

$$\lambda_j\Big(\frac{1}{n}\boldsymbol{\Lambda}_n\Big)\leqslant\lambda_j\Big(\frac{1}{n}\boldsymbol{U}^\mathsf{T}\boldsymbol{U}\Big),\qquad j=1,2,\ldots,m,$$

where $\lambda_j(\boldsymbol{A})$ is the jth largest nonzero eigenvalue of $\boldsymbol{A}$. Then, Lemma 7 follows from condition C7. □

Lemma 8 Under the conditions of Theorem 2, one has

$$|\widehat{\boldsymbol{\theta}}| = O_p(\delta_n), \tag{8}$$

where $\delta_n = \sqrt{N+m}$ and $\widehat{\boldsymbol{\theta}}$ is defined as (9).

Proof of Lemma 8 Let

$$\boldsymbol{\theta}\begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \boldsymbol{\theta}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}(\boldsymbol{b}-\boldsymbol{b}_0) \\ \boldsymbol{\Lambda}_n{}^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) \\ \sqrt{1/N}\boldsymbol{H}_n(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0) + \sqrt{N}\boldsymbol{H}_n{}^{-1}\boldsymbol{\pi}^{\mathsf{T}}\boldsymbol{U}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) \end{pmatrix}$$

and

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}}, \widehat{\gamma}) = (\widehat{\boldsymbol{\theta}}_1{}^{\mathsf{T}}, \widehat{\boldsymbol{\theta}}_2{}^{\mathsf{T}}, \widehat{\boldsymbol{\theta}}_3{}^{\mathsf{T}})^{\mathsf{T}}. \tag{9}$$

Now, we show that $|\widehat{\boldsymbol{\theta}}| = O_p(\delta_n)$. To do so, let $\widetilde{\boldsymbol{V}}_k = (0, \ldots, 1/[\sqrt{n}f_\epsilon(b_{0k})], \ldots, 0)^{\mathsf{T}}$, $\widetilde{\boldsymbol{U}}_i = \boldsymbol{\Lambda}_n^{-1/2}\boldsymbol{U}_i^*$, $\widetilde{\boldsymbol{\pi}}_i = N^{1/2}\boldsymbol{H}_n^{-1}\boldsymbol{\pi}_i$, $R_i = \boldsymbol{U}_i^{\mathsf{T}}\boldsymbol{\beta}_0 - \int_0^1 \beta_0(t)X_i(t)\mathrm{d}t + \boldsymbol{\pi}_i^{\mathsf{T}}\boldsymbol{\alpha}_0 - g(Z_i)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{\mathsf{T}}, \boldsymbol{\theta}_2^{\mathsf{T}}, \boldsymbol{\theta}_3^{\mathsf{T}})^{\mathsf{T}}$ and

$$f_{ik}(\boldsymbol{\theta}) = \rho_{\tau_k}(\epsilon_i - b_{0k} - \widetilde{\boldsymbol{V}}_k^{\mathsf{T}}\boldsymbol{\theta}_1 - \widetilde{\boldsymbol{U}}_i^{\mathsf{T}}\boldsymbol{\theta}_2 - \widetilde{\boldsymbol{\pi}}_i^{\mathsf{T}}\boldsymbol{\theta}_3 - R_i). \tag{10}$$

Then, one has

$$\begin{aligned} L(\boldsymbol{\theta}) &\equiv \sum_{k=1}^{K}\sum_{i=1}^{n}\rho_{\tau_k}(Y_i - b_k - \boldsymbol{U}_i^{\mathsf{T}}\boldsymbol{\beta} - \boldsymbol{\pi}_i^{\mathsf{T}}\boldsymbol{\alpha}) \\ &= \sum_{k=1}^{K}\sum_{i=1}^{n}\rho_{\tau_k}(\epsilon_i - b_{0k} - \widetilde{\boldsymbol{V}}_k^{\mathsf{T}}\boldsymbol{\theta}_1 - \widetilde{\boldsymbol{U}}_i^{\mathsf{T}}\boldsymbol{\theta}_2 - \widetilde{\boldsymbol{\pi}}_i^{\mathsf{T}}\boldsymbol{\theta}_3 - R_i) \\ &= \sum_{k=1}^{K}\sum_{i=1}^{n} f_{ik}(\boldsymbol{\theta}). \end{aligned}$$

We can obtain (11) with similar arguments to these of Lemma 1 of [13]. Using Lemma 9 and Lemma 10, which will be addressed later, for any $\kappa > 0$, we can find L_κ sufficiently large such that

$$\mathsf{P}\Big\{\inf_{|\boldsymbol{\theta}|\geqslant L_\kappa\delta_n} L(\boldsymbol{\theta}) > \sum_{k=1}^{K}\sum_{i=1}^{n}\rho_{\tau_k}(\epsilon_i - R_i - b_{0k})\Big\} > 1-\kappa, \tag{11}$$

when n is large enough. On the other hand, we have

$$L(\widehat{\boldsymbol{\theta}}) = \inf_{\boldsymbol{\theta}\in R^{K+m+N}} L(\boldsymbol{\theta}). \tag{12}$$

Thus, we have

$$L(\widehat{\boldsymbol{\theta}}) \leqslant \sum_{k=1}^{K}\sum_{i=1}^{n}\rho_{\tau_k}(\epsilon_i - R_i - b_{0k}).$$

Then connecting this with (11), we obtain that when n is large enough,

$$\mathsf{P}\Big\{\inf_{|\boldsymbol{\theta}|\geqslant L_\kappa\delta_n} L(\boldsymbol{\theta}) > L(\widehat{\boldsymbol{\theta}})\Big\} > 1-\kappa.$$

Thus, $|\widehat{\boldsymbol{\theta}}| = O_p(\delta_n)$. This achieves the proof of Lemma 8. □

Similar to identity [36]

$$|r-s|-|r| = -s(I(r>0)-I(r<0)) + 2\int_0^s [I(r\leqslant t)-I(r\leqslant 0)]\mathrm{d}t,$$

we have

$$\rho_\tau(r-s)-\rho_\tau(r) = s(I(r<0)-\tau) + \int_0^s [I(r\leqslant t)-I(r\leqslant 0)]\mathrm{d}t. \tag{13}$$

Before giving the results of Lemma 9 and Lemma 10, we first point out that it is equivalent to prove Lemma 9 and Lemma 10 with $\rho_\tau(u)$ replaced by $|u|$ [13, 37].

Lemma 9 Under the conditions of Theorem 2, for all $\kappa>0$, there exists $L=L_\kappa$ such that

$$\lim_{n\to\infty}\mathsf{P}\Big[\inf_{|\boldsymbol{\theta}|=1}\sum_{i=1}^n (f_{ik}(L\delta_n\boldsymbol{\theta})-f_{ik}(0)-\mathsf{E}[f_{ik}(L\delta_n\boldsymbol{\theta})-f_{ik}(0)]\,|\,T_n) > \kappa(m+N)^2\Big] = 0.$$

Proof of Lemma 9 Firstly, noting that $\|v_j-\widehat{v}_j\|^2 = O_p(n^{-1}j^2)$ [21], one has

$$\begin{aligned}
|R_i|^2 &= \Big|\boldsymbol{U}_i^{\mathsf{T}}\boldsymbol{\beta}_0 - \int_0^1 \beta(t)X(t)\mathrm{d}t + \boldsymbol{\pi}_i^{\mathsf{T}}\boldsymbol{\alpha}_0 - g(Z_i)\Big|^2\\
&\leqslant 4\Big|\sum_{j=1}^m \langle X_i, \widehat{v}_j - v_j\rangle\beta_{0j}\Big|^2 + 4\Big|\sum_{j=m+1}^\infty \langle X_i, v_j\rangle\beta_{0j}\Big|^2 + 2|\boldsymbol{\pi}_i^{\mathsf{T}}\boldsymbol{\alpha}_0 - g(Z_i)|^2\\
&= 4A_1 + 4A_2 + 2A_3.
\end{aligned}$$

For A_1, by condition C1 and the Hölder inequality, it is obtained

$$\begin{aligned}
A_1 = \Big|\sum_{j=1}^m \langle X_i, v_j-\widehat{v}_j\rangle\beta_{0j}\Big|^2 &\leqslant Cm\sum_{j=1}^m \|v_j-\widehat{v}_j\|^2|\beta_{0j}|^2\\
&\leqslant Cm\sum_{j=1}^m O_p(n^{-1}j^{2-2b}) = O_p(n^{-(a+4b-4)/(a+2b)}).
\end{aligned}$$

As for A_2, due to

$$\mathsf{E}\Big\{\sum_{j=m+1}^\infty \langle X_i, v_j\rangle\beta_{0j}\Big\} = 0,$$

and

$$\begin{aligned}
\mathsf{Var}\Big\{\sum_{j=m+1}^\infty \langle X_i, v_j\rangle\beta_{0j}\Big\} &= \sum_{j=m+1}^\infty \lambda_j\beta_{0j}^2 \leqslant C\sum_{j=m+1}^\infty j^{-(a+2b)}\\
&= O(n^{-(a+2b-1)/(a+2b)}),
\end{aligned}$$

one has

$$A_2 = O_p(n^{-(a+2b-1)/(a+2b)}).$$

By Lemma 6, one has

$$A_3 = O_p(N^{-2r}).$$

Taking these together, we have

$$|R_i|^2 = O_p(n^{-(a+2b-1)/(a+2b)} + N^{-2r}). \tag{14}$$

Using the definition of functions f_{ik}, we have

$$\begin{aligned}
&\sup_{|\boldsymbol{\theta}|\leqslant 1} \sum_{i=1}^{n} (f_{ik}(L\delta_n\boldsymbol{\theta}) - f_{ik}(0) - \mathsf{E}[f_{ik}(L\delta_n\boldsymbol{\theta}) - f_{ik}(0)] \,|\, T_n) \\
= &\sup_{|\boldsymbol{\theta}|\leqslant 1} \sum_{i=1}^{n} (|\epsilon_i - b_{0k} - L\delta_n(\widetilde{\boldsymbol{V}}_k^{\mathsf{T}}\boldsymbol{\theta}_1 + \widetilde{\boldsymbol{U}}_i^{\mathsf{T}}\boldsymbol{\theta}_2 + \widetilde{\boldsymbol{\pi}}_i^{\mathsf{T}}\boldsymbol{\theta}_3) - R_i| - |\epsilon_i - b_{0k} - R_i| \\
&- \mathsf{E}[|\epsilon_i - b_{0k} - L\delta_n(\widetilde{\boldsymbol{V}}_k^{\mathsf{T}}\boldsymbol{\theta}_1 + \widetilde{\boldsymbol{U}}_i^{\mathsf{T}}\boldsymbol{\theta}_2 + \widetilde{\boldsymbol{\pi}}_i^{\mathsf{T}}\boldsymbol{\theta}_3) - R_i| - |\epsilon_i - b_{0k} - R_i|]).
\end{aligned}$$

Denoting $\Delta_i(\boldsymbol{\theta}) = |\epsilon_i - b_{0k} - L\delta_n(\widetilde{\boldsymbol{V}}_k^{\mathsf{T}}\boldsymbol{\theta}_1 + \widetilde{\boldsymbol{U}}_i^{\mathsf{T}}\boldsymbol{\theta}_2 + \widetilde{\boldsymbol{\pi}}_i^{\mathsf{T}}\boldsymbol{\theta}_3) - R_i| - |\epsilon_i - b_{0k} - R_i|$. To prove Lemma 9, it suffices to show that for any $\kappa > 0$, there exists $L = L_\kappa$ such that

$$\lim_{n\to\infty} \mathsf{P}\Big(\sup_{|\boldsymbol{\theta}|\leqslant 1} \sum_{i=1}^{n} [\Delta_i(\boldsymbol{\theta}) - \mathsf{E}(\Delta_i(\boldsymbol{\theta}) \,|\, T_n)] > \kappa(N+m)^2\Big) = 0.$$

Let $\mathscr{C} \equiv \{\boldsymbol{\theta} : \boldsymbol{\theta} \in R^{K+m+N}, |\boldsymbol{\theta}| \leqslant 1\}$. As $\mathscr{C}$ is a compact set, we can cover it with open balls, that is $\mathscr{C} = \bigcup_{j=1}^{K_n} \mathscr{C}_j$, with chosen for all j from 1 to K_n, such that

$$\operatorname{diam}(\mathscr{C}_j) \leqslant \frac{\kappa\sqrt{N+m}}{L\sqrt{n}}, \qquad j = 1, 2, \ldots, K_n$$

and

$$K_n \leqslant \Big(\frac{L\sqrt{n}}{\kappa\sqrt{N+m}}\Big)^{K+m+N}. \tag{15}$$

According to condition C6, Lemma 5.1 of [34] and Lemma 7, respectively, we can easily conclude that

$$|\widetilde{\boldsymbol{V}}_k|^2 = O_p\Big(\frac{1}{n}\Big), \qquad |\widetilde{\boldsymbol{\pi}}_i|^2 = O_p\Big(\frac{N}{n}\Big), \qquad |\widetilde{\boldsymbol{U}}_i|^2 = O_p\Big(\frac{m}{n}\Big), \tag{16}$$

where $1 \leqslant k \leqslant K$, $1 \leqslant i \leqslant n$. Now, for $j = 1, 2, \ldots, K_n$, let $\boldsymbol{\theta}_j^* \in \mathscr{C}_j$. Using the definition of $\Delta_i(\boldsymbol{\theta})$ and (16), we have

$$\begin{aligned}
&\min_{j=1,2,\ldots,K_n} \sum_{i=1}^{n} [\Delta_i(\boldsymbol{\theta}) - \mathsf{E}(\Delta_i(\boldsymbol{\theta}) \,|\, T_n)] - [\Delta_i(\boldsymbol{\theta}_j^*) - \mathsf{E}(\Delta_i(\boldsymbol{\theta}_j^*) \,|\, T_n)] \\
&\leqslant C(m+N)^{3/2}\kappa \leqslant \frac{\kappa}{4}(m+N)^2,
\end{aligned} \tag{17}$$

and

$$\sup_{\boldsymbol{\theta}\in\mathscr{C}} |\Delta_i(\boldsymbol{\theta})| \leqslant \frac{CL\delta_n^2}{\sqrt{n}}. \tag{18}$$

Besides, for $\boldsymbol{\theta}$ fixed in $\mathscr{C}$, with the same arguments as before, one has

$$\sum_{i=1}^{n} \mathsf{Var}\,(\Delta_i(\boldsymbol{\theta})\,|\,T_n) \leqslant CL^2(N+m)^2. \tag{19}$$

We are now able to prove Lemma 9. By inequalities (15), (17)–(19) and Bernstein inequality, we obtain, as $n \to +\infty$,

$$\begin{aligned}
&\mathsf{P}\Big(\sup_{|\boldsymbol{\theta}|\leqslant 1}\sum_{i=1}^{n}[\Delta_i(\boldsymbol{\theta}) - \mathsf{E}(\Delta_i(\boldsymbol{\theta})\,|\,T_n)] > \kappa(m+N)^2\,\Big|\,T_n\Big)\\
\leqslant{}& \mathsf{P}\Big(\max_{j=1,2,\ldots,K_n}\sum_{i=1}^{n}[\Delta_i(\boldsymbol{\theta}_j^*) - \mathsf{E}(\Delta_i(\boldsymbol{\theta}_j^*)\,|\,T_n)] > \frac{\kappa}{2}(m+N)^2\,\Big|\,T_n\Big)\\
\leqslant{}& \sum_{j=1}^{K_n}\mathsf{P}\Big(\sum_{i=1}^{n}[\Delta_i(\boldsymbol{\theta}_j^*) - \mathsf{E}(\Delta_i(\boldsymbol{\theta}_j^*)\,|\,T_n)] > \frac{\kappa}{2}(m+N)^2\,\Big|\,T_n\Big)\\
\leqslant{}& 2\exp\Big\{\ln\Big(\frac{L\sqrt{n}}{\sqrt{(N+m)}\,\kappa}\Big)^{K+m+N} - \frac{\kappa^2(N+m)^4/4}{CL^2(N+m)^2 + \kappa CL(m+N)^3/\sqrt{n}}\Big\} \to 0.
\end{aligned}$$

The bound on the right hand side does not depend on the sample T_n. Hence, if we take the expectation on both sides, the direction of inequality remain unchanged. This achieves the proof of Lemma 9. $\square$

Lemma 10 Under the conditions of Theorem 2, for all $\kappa > 0$, there exists $L = L_\kappa$ (sufficiently large) such that

$$\lim_{n\to\infty}\mathsf{P}\Big[\inf_{|\boldsymbol{\theta}|=1}\sum_{i=1}^{n}\mathsf{E}[f_{ik}(L\delta_n\boldsymbol{\theta}) - f_{ik}(0)\,|\,T_n] > (m+N)^2\Big] > 1-\kappa.$$

Proof of Lemma 10 According to (13), we have

$$\mathsf{E}(\rho_{\tau_k}(\epsilon_i - b_{0k} - a - b) - \rho_{\tau_k}(\epsilon_i - b_{0k} - b)\,|\,T_n) = \frac{1}{2}f_\epsilon(b_{0k})a^2 + f_\epsilon(b_{0k})ab + o((a+b)^2).$$

Considering $|\boldsymbol{\theta}| = 1$, if we set $R_i' = 2R_i$, then

$$\begin{aligned}
&\sum_{i=1}^{n}\mathsf{E}[\rho_{\tau_k}(\epsilon_i - b_{0k} - L\delta_n(\widetilde{\boldsymbol{V}}_k^\mathsf{T}\boldsymbol{\theta}_1 + \widetilde{\boldsymbol{U}}_i^\mathsf{T}\boldsymbol{\theta}_2 + \widetilde{\boldsymbol{\pi}}_i^\mathsf{T}\boldsymbol{\theta}_3) - R_i) - \rho_{\tau_k}(\epsilon_i - b_{0k} - R_i)\,|\,T_n]\\
={}& \frac{1}{2}\sum_{i=1}^{n} f_\epsilon(b_{0k})[L^2\delta_n^2(\widetilde{\boldsymbol{V}}_k^\mathsf{T}\boldsymbol{\theta}_1 + \widetilde{\boldsymbol{U}}_i^\mathsf{T}\boldsymbol{\theta}_2 + \widetilde{\boldsymbol{\pi}}_i^\mathsf{T}\boldsymbol{\theta}_3)^2 + L\delta_n(\widetilde{\boldsymbol{V}}_k^\mathsf{T}\boldsymbol{\theta}_1 + \widetilde{\boldsymbol{U}}_i^\mathsf{T}\boldsymbol{\theta}_2 + \widetilde{\boldsymbol{\pi}}_i^\mathsf{T}\boldsymbol{\theta}_3)R_i']\\
&+ o((L\delta_n(\widetilde{\boldsymbol{V}}_k^\mathsf{T}\boldsymbol{\theta}_1 + \widetilde{\boldsymbol{U}}_i^\mathsf{T}\boldsymbol{\theta}_2 + \widetilde{\boldsymbol{\pi}}_i^\mathsf{T}\boldsymbol{\theta}_3) + R_i')^2)\\
\geqslant{}& \frac{1}{4}\sum_{i=1}^{n} f_\epsilon(b_{0k})[L^2\delta_n^2(\widetilde{\boldsymbol{V}}_k^\mathsf{T}\boldsymbol{\theta}_1 + \widetilde{\boldsymbol{U}}_i^\mathsf{T}\boldsymbol{\theta}_2 + \widetilde{\boldsymbol{\pi}}_i^\mathsf{T}\boldsymbol{\theta}_3)^2 - {R_i'}^2] + o_p((m+N)^2).
\end{aligned}$$

According to the definition of f_{ik} and (16), we have

$$\begin{aligned}
&\frac{1}{(m+N)^2}\inf_{|\boldsymbol{\theta}|=1}\sum_{i=1}^{n}\mathsf{E}[f_{ik}(L\delta_n\boldsymbol{\theta})-f_{ik}(0)\,|\,T_n]\\
&\geqslant \frac{1}{4(m+N)^2}\inf_{|\boldsymbol{\theta}|=1}\sum_{i=1}^{n}f_\epsilon(b_{0k})[L^2\delta_n^2(\widetilde{\boldsymbol{V}}_k^{\mathsf{T}}\boldsymbol{\theta}_1+\widetilde{\boldsymbol{U}}_i^{\mathsf{T}}\boldsymbol{\theta}_2+\widetilde{\boldsymbol{\pi}}_i^{\mathsf{T}}\boldsymbol{\theta}_3)^2-R_i'^2]+o_p(1)\\
&\geqslant CL^2-O_p\Big(\frac{1}{m+N}\Big)+o_p(1),
\end{aligned}$$

which can be made arbitrarily large (as $n\to\infty$) by choosing L.

This leads to

$$\lim_{n\to\infty}\mathsf{P}\Big[\frac{1}{(m+N)^2}\inf_{|\boldsymbol{\theta}|=1}\sum_{i=1}^{n}\mathsf{E}[f_{ik}(L\delta_n\boldsymbol{\theta})-f_{ik}(0)\,|\,T_n]>1\Big]=1.$$

The proof of Lemma 10 is hence completed. □

Proof of Theorem 2 By the definition of $\widehat{\boldsymbol{\theta}}$, Lemma 7 and Lemma 8, one has

$$|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0|^2=|\boldsymbol{\Lambda}_n^{-1/2}\widehat{\boldsymbol{\theta}}_2|^2=O_p(n^{-(2b-1)/(a+2b)})+O_p(n^{-2b/(a+2b)+1/(2r+1)}).$$

Observe that

$$\begin{aligned}
\|\widehat{\beta}(t)-\beta_0(t)\|^2&=\Big\|\sum_{j=1}^{m}\widehat{\beta}_{0j}\widehat{v}_j-\sum_{j=1}^{\infty}\beta_{0j}v_j\Big\|^2\\
&\leqslant 2\Big\|\sum_{j=1}^{m}\widehat{\beta}_j\widehat{v}_j-\sum_{j=1}^{m}\beta_{0j}v_j\Big\|^2+2\Big\|\sum_{j=m+1}^{\infty}\beta_{0j}v_j\Big\|^2\\
&\leqslant 4\Big\|\sum_{j=1}^{m}(\widehat{\beta}_j-\beta_{0j})\widehat{v}_j\Big\|^2+4\Big\|\sum_{j=1}^{m}\beta_{0j}(\widehat{v}_j-v_j)\Big\|^2+2\sum_{j=m+1}^{\infty}\beta_{0j}^2\\
&=4B_1+4B_2+2B_3.
\end{aligned}$$

By orthogonality of $\{\widehat{v}_j\}$ and $\|v_j-\widehat{v}_j\|^2=O_p(n^{-1}j^2)$, one has

$$\begin{aligned}
B_1&=\Big\|\sum_{j=1}^{m}(\widehat{\beta}_j-\beta_{0j})\widehat{v}_j\Big\|^2\leqslant\sum_{j=1}^{m}|\widehat{\beta}_j-\beta_{0j}|^2=|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0|^2\\
&=O_p(n^{-(2b-1)/(a+2b)})+O_p(n^{-2b/(a+2b)+1/(2r+1)}),\\
B_2&=\Big\|\sum_{j=1}^{m}\beta_{0j}(\widehat{v}_j-v_j)\Big\|^2\leqslant m\sum_{j=1}^{m}\|\widehat{v}_j-v_j\|^2\beta_{0j}^2\leqslant\frac{m}{n}O_p\Big(\sum_{j=1}^{m}j^2\beta_{0j}^2\Big)\\
&=O_p\Big(n^{-1}m\sum_{j=1}^{m}j^{2-2b}\Big)=O_p(n^{-1}m)=o_p(n^{-(2b-1)/(a+2b)}),\\
B_3&=\sum_{j=m+1}^{\infty}\beta_{0j}^2\leqslant C\sum_{j=m+1}^{\infty}j^{-2b}=O(n^{-(2b-1)/(a+2b)}).
\end{aligned}$$

Thus, we have

$$\|\widehat{\beta}(t)-\beta_0(t)\|^2=O_p(n^{-(2b-1)/(a+2b)})+O_p(n^{-2b/(a+2b)+1/(2r+1)}).$$

On the other hand, by the definition of $\widehat{\boldsymbol{\theta}}$, one has

$$|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0|^2=|\sqrt{N}\boldsymbol{H}_n^{-1}\widehat{\boldsymbol{\theta}}_3-N\boldsymbol{H}_n{}^{-2}\boldsymbol{\pi}^{\mathsf{T}}\boldsymbol{U}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)|^2.$$

We note that

$$|\sqrt{N}\boldsymbol{H}_n^{-1}\widehat{\boldsymbol{\theta}}_3|^2=N\widehat{\boldsymbol{\theta}}_3^{\mathsf{T}}\boldsymbol{H}_n^{-2}\widehat{\boldsymbol{\theta}}_3=O_p\Big(\frac{N(N+m)}{n}\Big), \tag{20}$$

similarly

$$\begin{aligned}\Big|\frac{N}{n}\boldsymbol{H}_n^{-2}\boldsymbol{\pi}^{\mathsf{T}}\boldsymbol{U}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\Big|^2&=\frac{N^2}{n^2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)^{\mathsf{T}}\boldsymbol{U}^{\mathsf{T}}\boldsymbol{\pi}\boldsymbol{H}_n^{-4}\boldsymbol{\pi}^{\mathsf{T}}\boldsymbol{U}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\&=O_p\Big(\frac{N(N+m)}{n}\Big).\end{aligned} \tag{21}$$

As a result, we obtain

$$|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0|^2=O_p\Big(\frac{N(N+m)}{n}\Big). \tag{22}$$

Using Lemma 6, Hölder inequality and (22), one has

$$\begin{aligned}\|\widehat{g}-g_0\|^2&=\Big\|\sum_{s=1}^{N}\widehat{\alpha}_s\pi_s-g_0\Big\|^2\\&\leqslant 2\|\widehat{g}-g^*\|^2+2\|g^*-g_0\|^2\\&=2\Big\|\sum_{s=1}^{N}(\widehat{\alpha}_s-\alpha_{0s})\pi_s\Big\|^2+2\Big\|\sum_{s=1}^{N}\alpha_{0s}\pi_s-g_0\Big\|^2\\&=O\Big(\frac{1}{N}\Big)|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0|^2+2\Big\|\sum_{s=1}^{N}\alpha_{0s}\pi_s-g_0\Big\|^2\\&=O_p\Big(\frac{N+m}{n}\Big)+O(N^{-2r})\\&=O_p(n^{-2r/(2r+1)})+O_p(n^{-(a+2b-1)/(a+2b)}).\end{aligned}$$

This achieves the proof of Theorem 2. □

References

[1] Ramsay J O, Silverman B W. *Applied Functional Data Analysis: Methods and Case Studies* [M]. New York: Springer, 2002.

[2] Ramsay J O, Silverman B W. *Functional Data Analysis* [M]. 2nd ed. New York: Springer, 2005.

[3] Ferraty F, Vieu P. *Nonparametric Functional Data Analysis: Theory and Practice* [M]. New York: Springer, 2006.

[4] Horváth L, Kokoszka P. *Inference for Functional Data with Applications* [M]. New York: Springer, 2012.

[5] Cardot H, Ferraty F, Sarda P. Functional linear model [J]. *Statist. Probab. Lett.*, 1999, **45(1)**: 11–22.

[6] Yao F, Müller H G, Wang J L. Functional linear regression analysis for longitudinal data [J]. *Ann. Statist.*, 2005, **33(6)**: 2873–2903.

[7] Cai T T, Hall P. Prediction in functional linear regression [J]. *Ann. Statist.*, 2006, **34(5)**: 2159–2179.

[8] Hall P, Horowitz J L. Methodology and convergence rates for functional linear regression [J]. *Ann. Statist.*, 2007, **35(1)**: 70–91.

[9] Crambes C, Kneip A, Sarda P. Smoothing splines estimators for functional linear regression [J]. *Ann. Statist.*, 2009, **37(1)**: 35–72.

[10] Lian H. Functional partial linear model [J]. *J. Nonparametr. Stat.*, 2011, **23(1)**: 115–128.

[11] Cai T T, Yuan M. Minimax and adaptive prediction for functional linear regression [J]. *J. Amer. Statist. Assoc.*, 2012, **107(499)**: 1201–1216.

[12] Koenker R, Bassett G. Regression Quantiles [J]. *Econometrica*, 1978, **46(1)**: 33–50.

[13] Cardot H, Crambes C, Sarda P. Quantile regression when the covariates are functions [J]. *J. Nonparametr. Stat.*, 2005, **17(7)**: 841–856.

[14] Chen K H, Müller H G. Conditional quantile analysis when covariates are functions, with application to growth data [J]. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 2012, **74(1)**: 67–89.

[15] Kato K. Estimation in functional linear quantile regression [J]. *Ann. Statist.*, 2012, **40(6)**: 3108–3136.

[16] Lu Y, Du J, Sun Z M. Functional partially linear quantile regression model [J]. *Metrika*, 2014, **77(2)**: 317–332.

[17] Zou H, Yuan M. Composite quantile regression and the oracle model selection theory [J]. *Ann. Statist.*, 2008, **36(3)**: 1108–1126.

[18] Kai B, Li R Z, Zou H. Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression [J]. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 2010, **72(1)**: 49–69.

[19] Jiang R, Zhou Z G, Qian W M, et al. Single-index composite quantile regression [J]. *J. Korean Statist. Soc.*, 2012, **41(3)**: 323–332.

[20] Kong E F, Xia Y C. An adaptive composite quantile approach to dimension reduction [J]. *Ann. Statist.*, 2014, **42(4)**: 1657–1688.

[21] Shin H. Partial functional linear regression [J]. *J. Statist. Plann. Inference*, 2009, **139(10)**: 3405–3418.

[22] Delaigle A, Hall P. Methodology and theory for partial least squares applied to functional data [J]. *Ann. Statist.*, 2012, **40(1)**: 322–352.

[23] He X M, Zhu Z Y, Fung W K. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure [J]. *Biometrika*, 2002, **89(3)**: 579–590.

[24] Koenker R. *Quantile Regression* [M]. Cambridge, UK: Cambridge University Press, 2005.

[25] Chen H. Polynomial splines and nonparametric regression [J]. *J. Nonparametr. Statist.*, 1991, **1(1-2)**: 143–156.

[26] Wang H X J, Zhu Z Y, Zhou J H. Quantile regression in partially linear varying coefficient models [J]. *Ann. Statist.*, 2009, **37(6B)**: 3841–3866.

[27] Stone C J. Optimal rates of convergence for nonparametric estimators [J]. *Ann. Statist.*, 1980, **8(6)**: 1348–1360.

[28] He X M, Fung W K, Zhu Z Y. Robust estimation in generalized partial linear models for clustered data [J]. *J. Amer. Statist. Assoc.*, 2005, **100(472)**: 1176–1184.

[29] Bang S W, Jhun M. Simultaneous estimation and factor selection in quantile regression via adaptive sup-norm regularization [J]. *Comput. Statist. Data Anal.*, 2012, **56(4)**: 813–826.

[30] Fan J Q, Gijbels I. *Local Polynomial Modelling and Its Applications* [M]. London: Chapman and Hall, 1996.

[31] Aneiros-Pérez G, Vieu P. Partial linear modelling with multi-functional covariates [J]. *Comput. Statist.*, 2015, **30(3)**: 647–671.

[32] Aneiros-Pérez G, Vieu P. Semi-functional partial linear regression [J]. *Statist. Probab. Lett.*, 2006, **76 (11)**: 1102–1110.

[33] Schumaker L L. *Spline Functions: Basic Theory* [M]. New York: Wiley, 1981.

[34] Shi P D, Li G Y. Global convergence rates of B-Spline M-Estimators in nonparametric regression [J]. *Statist. Sinica*, 1995, **5(1)**: 303–318.

[35] Yanai H, Takeuchi K, Takane Y. *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition* [M]. New York: Springer, 2011.

[36] Knight K. Limiting distributions for L_1 regression estimators under general conditions [J]. *Ann. Statist.*, 1998, **26(2)**: 755–770.

[37] He X M, Shi P D. Convergence rate of b-spline estimators of nonparametric conditional quantile functions [J]. *J. Nonparametr. Statist.*, 1994, **3(3-4)**: 299–308.

函数型部分线性复合分位数回归模型的估计

余 平[1,2] 张忠占[1] 杜 江[1]

([1]北京工业大学应用数理学院, 北京, 100124; [2]山西师范大学数学与计算机学院, 临汾, 041000)

摘 要： 本文研究函数型部分线性复合分位数回归模型的估计问题. 我们采用函数型主成分分析方法分析斜率函数, 回归样条逼近非参数函数. 在相当宽松的条件下给出斜率函数和非参数函数的收敛速度. 最后通过理论模拟和实例分析来评价我们提出的方法.

关键词： 函数型数据分析; 样条估计; 复合分位数回归; 函数型主成分分析

中图分类号： O212.7