

Functional partially linear quantile regression model

Ying Lu · Jiang Du · Zhimeng Sun

Received: 17 July 2012 / Published online: 9 April 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract This paper considers estimation of a functional partially quantile regression model whose parameters include the infinite dimensional function as well as the slope parameters. We show asymptotical normality of the estimator of the finite dimensional parameter, and derive the rate of convergence of the estimator of the infinite dimensional slope function. In addition, we show the rate of the mean squared prediction error for the proposed estimator. A simulation study is provided to illustrate the numerical performance of the resulting estimators.

Keywords Functional linear regression · Quantile regression · Asymptotic normality

Jiang Du's work is supported by the National Natural Science Foundation of China (No. 11271039, No. 11101015, No. 11261025), the specialised research fund for the doctoral program of higher education (No. 20091103120012) and the fund from the government of Beijing (No. 2011D005015000007). Sun's research was supported by grants from the Foundation of Academic Discipline Program at Central University of Finance and Economics and National Statistics Research Projects in 2012.

Y. Lu
College of Applied Sciences, Communication University of China,
Beijing 100024, People's Republic of China

J. Du (✉)
College of Applied Sciences, Beijing University of Technology,
Beijing 100124, People's Republic of China
e-mail: dujiang84@163.com

Z. Sun
School of statistics, Central University of Finance and Economics,
Beijing 100081, People's Republic of China

1 Introduction

Since first proposed by [Koenker and Bassett \(1978\)](#), quantile regression has emerged as an important statistical methodology. By estimating various conditional quantile functions, quantile regression complements the focus of classical least squares regression on the conditional mean and explores the effect of covariate on the location, scale and shape of the distribution of the response variable. It has been used in wide range of applications including economics, biology, and finance. A comprehensive review of the theory of quantile regression and some of the most recent development can be found in [Koenker \(2005\)](#).

In order to reduce severe modeling biases caused by the mis-specifying parametric models, there has been an upsurge of interests and efforts in nonparametric models. Although the nonparametric approach is useful in exploring hidden structures and reducing modeling biases, it can be too flexible to make the concise conclusions, and faces the curse of dimensionality due to a large number of covariates. In order to overcome these shortcomings, [Engle et al. \(1986\)](#) proposed the partially linear regression model (PLRM) which allowed some explanatory variables to act in a nonparametric manner, while the others to have a linear relation with the response variable. PLRM not only avoids the curse of dimensionality problem in nonparametric regression, but also retains the interpretation of the effect of the explanatory variables in linear regression. Therefore, PLRM has received considerable attention. For example, several papers have been published on this class of model for the independent and identically distributed case (see [Speckman 1988](#); [Shi and Li 1994](#); [Mammen and Geer 1997](#)) as well as for dependent data (see [Gao 1995](#); [Fan and Li 1999](#); [Liu 2011](#)), for more details see [Härdle et al. \(2000\)](#). Estimation methods in above mentioned papers are mainly based on the mean regression. As for partially linear quantile regression model, [He and Shi \(1996\)](#) used bivariate tensor-product B-splines to approximate the nonparametric function. [He and Liang \(2000\)](#) introduced partially linear quantile regression in errors-in-variables models.

There is a large collection of functional data literature on mean regression (see [Ramsay and Silverman 2005](#) for a review), but relatively few studies from a quantile regression perspective. Only [Cardot et al. \(2005\)](#) introduced penalized quantile regression when the covariates are functions. It is well known quantile regression approach is insensitive to outliers and more robust than the ordinary least squares method. Moreover, it may even work beautifully when the variance of random error is infinite, while the least squares method breaks down. In addition, fitting data at a set of quantiles provides a more comprehensive description of the response distribution than does the mean. In many applications, the functional impacts of the covariates on the response may vary at different percentiles of the distribution (see [Wang et al. 2009](#)). In many practical application, we are interested in the low or high quantile of the response variable when some explain covariates are functions and others may be classifiable variates. In the conditional growth charts, for example, one often uses the gender as classifiable variate because the adult male height is higher than woman's in general, and uses the entire growth phase as function. The ultimate outcome interest is the adult height. We are very much concerned about the low quantile of adult height, and want to decide which factors stunt height growth and which factors facilitate. All of these

motivate us to combine the quantile and semiparametric approaches with functional regression, which result in the functional partially linear quantile regression model (FPLQRM). It is obvious that the proposed FPLQRM is more flexible than functional linear quantile regression model proposed by Cardot et al. (2005), since it not only contains the nonparametric function, but also the linear variables which may be categorical variable. To the best of our knowledge, there is not yet any work that combines the semiparametric and quantile approaches in a functional semiparametric model in existing literature. We attempt to sort out this problem. However, there are many difficulties in this task. Firstly, unlike the square loss function, the quantile loss function is not differentiable at the origin, which makes it difficult to derive large sample properties of the resulting estimators. Secondly, one needs to estimate simultaneously the parametric and nonparametric component which have different convergence rates. In this paper, we develop new estimators for the parameters of functional partially linear quantile regression model from the perspective of the Karhunen-Loève expansion and show the consistence and asymptotical normality of the resulting estimators.

The remainder of the paper is organized as follows. Section 2 introduces functional partially linear quantile regression model. Section 3 then develops an approach. The large sample properties of the proposed estimators are given in Sect. 4. Simulation studies are presented in Sect. 5. The proofs of the main results are presented in the “Appendix”.

2 Functional partially linear quantile regression model

Let Y be a real value random variable defined on a probability space (Ω, \mathcal{B}, P) . $\mathbf{Z} = (Z_1, Z_2, \dots, Z_p)$ is a p -dimensional random variables and let $\{X(t) : t \in \mathcal{F}\}$ be a zero mean, second-order stochastic process defined on (Ω, \mathcal{B}, P) with sample paths in $L^2(\mathcal{F})$, the Hilbert space containing square integrable functions with inner product $\langle x, y \rangle = \int_{\mathcal{F}} x(t)y(t)dt, \forall x, y \in L^2(\mathcal{F})$ and norm $\|x\| = \langle x, x \rangle^{1/2}$. Without loss of generality, we suppose throughout the paper that $\mathcal{F} = [0, 1]$. At a given quantile level $\tau \in (0, 1)$, the dependence between Y and (X, \mathbf{Z}) is expressed as

$$Y = \int_0^1 \beta(t, \tau)X(t)dt + \mathbf{Z}^T \boldsymbol{\theta}(\tau) + \varepsilon(\tau), \tag{1}$$

where $\varepsilon(\tau)$ is a random error whose τ th quantile equals zero, $\beta(t, \tau)$ is a square integrable function on $[0,1]$, $\boldsymbol{\theta}(\tau)$ is p -dimensional unknown real vector. In the rest of the article, we will suppress τ in $\boldsymbol{\theta}(\tau)$ and $\beta(t, \tau)$ for notational simplicity.

Remark 1 Model (1) generalizes both the linear quantile regression model and functional linear model which correspond to the cases $\beta = 0$ and $\boldsymbol{\theta} = 0$, respectively. If $\tau = 0.5$ and ε have symmetric distributions with a finite mean, then the median FPLQRM is equal to the conditional mean. Therefore, model (1) also includes the partially functional linear regression model proposed by Shin (2009) and includes the semi-functional partially linear regression model in Aneiros-Pérez and Vieu (2006) given by $Y = \beta^T \mathbf{z} + m(X) + \varepsilon$ when $m(X) = \int_0^1 \gamma(t)X(t)dt$.

3 Estimation methods

Let $\{(X_i, \mathbf{Z}_i, Y_i), i = 1, \dots, n\}$ be an independent and identically distributed sample which is generated from model (1). Define the covariance function and the empirical covariance function respectively as

$$K(s, t) = \text{Cov}(X(t), X(s))$$

and

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

The covariance function K defines a linear operator which maps a function f to Kf given by $(Kf)(u) = \int K(u, v)f(v)dv$. We shall assume that the linear operator with kernel K is positive definite. Let $\lambda_1 > \lambda_2 > \dots > 0$ and $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq 0$ be the ordered eigenvalue sequences of the linear operators with kernels K and \hat{K} , $\{\phi_j\}$ and $\{\hat{\phi}_j\}$ be the corresponding orthonormal eigenfunction sequences respectively. It is clear that the sequences $\{\phi_j\}$ and $\{\hat{\phi}_j\}$ each forms an orthonormal basis in $L^2([0, 1])$. Then, the spectral decompositions of the covariance functions K and \hat{K} can be written as

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

and

$$\hat{K}(s, t) = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{\phi}_j(s)\hat{\phi}_j(t),$$

respectively.

According to the Karhunen-Loève representation, we have

$$X(t) = \sum_{i=1}^{\infty} \xi_i \phi_i(t)$$

and

$$\beta(t) = \sum_{i=1}^{\infty} \gamma_i \phi_i(t) \quad (2)$$

where the ξ_i are uncorrelated random variables with mean 0 and variance $E[\xi_i^2] = \lambda_i$, and $\gamma_i = \langle \beta, \phi_i \rangle$, for more details see [Ramsay and Silverman \(2005\)](#). Substituting (2)

into model (1), we can get

$$Y = \sum_{j=1}^{\infty} \gamma_j \langle \phi_j, X \rangle + \mathbf{Z}^T \boldsymbol{\theta} + \varepsilon(\tau). \tag{3}$$

Therefore, the regression model in (3) can be well approximated by

$$Y \approx \sum_{j=1}^m \gamma_j \langle \phi_j, X \rangle + \mathbf{Z}^T \boldsymbol{\theta} + \varepsilon(\tau), \tag{4}$$

where $m \leq n$ is the truncation level that trades off approximation error against variability and typically diverges with n . We replace the ϕ_j by $\hat{\phi}_j$ for $j = 1, \dots, m$, model (4) can be rewritten as

$$Y \approx \mathbf{Z}^T \boldsymbol{\theta} + \mathbf{U}\boldsymbol{\gamma} + \varepsilon,$$

where $\mathbf{U} = \{\langle X, \hat{\phi}_j \rangle\}_{j=1, \dots, m}$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m)^T$. The quantile coefficient estimates of $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ can be obtained by minimizing

$$\sum_{i=1}^n \rho_{\tau} \left(Y_i - \mathbf{U}_i \boldsymbol{\gamma} - \mathbf{Z}_i^T \boldsymbol{\theta} \right), \tag{5}$$

where $\rho_{\tau}(u) = u \{ \tau - I(u < 0) \}$ is the quantile loss function. The solution to (5) satisfies the following gradient condition:

$$\sum_{i=1}^n \psi_{\tau} \left(Y_i - \mathbf{U}_i \boldsymbol{\gamma} - \mathbf{Z}_i^T \boldsymbol{\theta} \right) \left(\mathbf{U}_i, \mathbf{Z}_i^T \right)^T = \mathbf{0},$$

where ψ_{τ} is score function of ρ_{τ} .

4 Large sample properties

Before presenting the main asymptotic results, we first introduce some conditions required for our asymptotic properties. Throughout this paper, the constant C may change from line to line for convenience.

- C1: $E \|X\|^4 < C < \infty$.
- C2: For each j , $E[U_j^4] \leq C\lambda_j$. For the eigenvalues λ_j and Fourier coefficients γ_j , we require that $\lambda_j - \lambda_{j+1} \geq C^{-1} j^{-a-1}$ and $|\gamma_j| \leq Cj^{-b}$ for $j > 1$, $a > 1$ and $b > a/2 + 1$.
- C3: For the tuning parameter m , we assume that $m \sim n^{1/(a+2b)}$.
- C4: $E \|\mathbf{Z}\|^4 < \infty$.

Conditions C1–C3 are very common in functional linear regression model (see [Hall and Horowitz 2007](#); [Shin 2009](#)). Condition C4 is common in functional partially linear regression model (see [Shin 2009](#)).

One complicating issue for FPLQRM comes from the dependence between \mathbf{Z} and X . Similar to [Shin \(2009\)](#), we let $\mathbf{Z} = \boldsymbol{\eta} + \langle \mathbf{g}, X \rangle$, where $\boldsymbol{\eta} = (\eta_1, \dots, \eta_p)$ is zero-mean random variable, $\mathbf{g} = (g_1, \dots, g_p)$ with $g_j \in L_2([0, 1])$, $j = 1, \dots, p$.

C5: $E[\boldsymbol{\eta}] = 0$ and $E[\boldsymbol{\eta}\boldsymbol{\eta}^T] = \Sigma$. Furthermore, we need that Σ is a positive definite matrix.

Condition C5 controls the limiting behaviour of the variance of $\hat{\boldsymbol{\theta}}$. [Speckman \(1988\)](#), [Moyeed and Diggle \(1994\)](#), [He et al. \(2002\)](#) and [Shin \(2009\)](#) used a similar device for modeling the dependence between parametric and nonparametric component.

The following Theorem 1 describes the rate of convergence of the functional slope parameter β and the asymptotic normality of constant slope parameters $\boldsymbol{\theta}$.

Theorem 1 *Under conditions C1–C5, we have*

$$\|\beta_0 - \hat{\beta}\|^2 = O_p\left(\delta_n^2\right),$$

and

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightarrow N\left(0, \varpi \Sigma^{-1}\right) \tag{6}$$

where $\delta_n^2 = n^{-(2b+1)/(a+2b)}$, $\varpi = \frac{\tau(1-\tau)}{f^2(0)}$ with $f(\varepsilon)$ is density function of the random error.

5 Simulation studies

In this section, we investigate the finite sample performance of the proposed estimation method with Monte Carlo simulation studies. We consider two sample sizes $n = 100$ and $n = 400$. We focus on $\tau = 0.25, 0.5$ and $\tau = 0.75$ in this study.

The data are generated from the following quantile regression model

$$Y = z_1\theta_1 + z_2\theta_2 + \int_0^1 X(t)\beta(t)dt + \varepsilon(\tau),$$

where z_1 follows the standard normal distribution and z_2 follows a Bernoulli distribution with 0.5 probability of being 1, $\theta_1 = 2$ and $\theta_2 = 1$. $\varepsilon(\tau) = \varepsilon - F^{-1}(\tau)$ with F being the CDF of ε . Here, $F^{-1}(\tau)$ is subtracted from ε to make the τ th quantile of $\varepsilon(\tau)$ zero for identifiability purpose. For the functional linear component, we take the same form as [Shin \(2009\)](#), that is, the functional coefficient $\beta(t) = \sqrt{2} \sin(\pi t/2) + 3\sqrt{2} \sin(3\pi t/2)$ and $X(t) = \sum_j \xi_j \phi_j(t)$, where the ξ_j are distributed as independent normal with mean 0 and variance $\lambda_j = ((j - 0.5)\pi)^{-2}$ and $\phi_j(t) = \sqrt{2} \sin((j - 0.5)\pi t)$.

We consider three cases for generating random error ε .

Table 1 Simulations results

Case	Sample size	Criterion	$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
			θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
Case 1	$n = 100$	Bias	-0.02	-0.01	-0.43	-0.45	0.48	-0.08
		MSE	14.07	19.37	13.43	18.10	14.13	19.79
	$n = 400$	Bias	-0.20	0.33	0.06	-0.20	-0.28	0.29
		MSE	6.84	9.88	6.33	8.94	7.00	9.92
Case 2	$n = 100$	Bias	-0.07	-0.68	-0.73	-0.70	0.30	0.36
		MSE	17.41	23.75	15.20	20.08	17.43	24.84
	$n = 400$	Bias	0.08	-0.67	-0.32	0.14	0.46	0.50
		MSE	8.33	12.26	7.06	10.10	8.26	12.21
Case 3	$n = 100$	Bias	-2.80	-18.63	1.39	-0.62	-1.37	16.48
		MSE	28.61	46.50	19.45	27.07	27.92	45.76
	$n = 400$	Bias	0.36	-5.49	0.61	-0.39	-0.52	5.53
		MSE	13.22	20.03	8.34	11.33	14.63	23.11

Bias and MSE of the estimated θ with different sample sizes and quantile levels for different cases ($\times 10^2$)

Case 1. ε follows a standard normal distribution.

Case 2. ε follows a $t(3)$ distribution. This yields a model with heavy-tailed.

Case 3. ε follows a standard Cauchy distribution. This yields a model in which the expectation of the response do not exist.

Throughout our numerical studies, we choose the number of eigenfunctions as the minimizer to the following Schwarz-type information criterion,

$$SIC(m) = \log \left\{ \sum_{i=1}^n \rho_{\tau} \left(Y_i - z_i^T \hat{\theta}_{(m)} - U_i \hat{\gamma}_{(m)} \right) \right\} + \frac{\log(n)}{2n} (m + p),$$

where $p = 2$, $\hat{\theta}_{(m)}$ and $\hat{\gamma}_{(m)}$ are the τ th quantile estimators obtained from minimizing (5) with m eigenfunctions; see He et al. (2002) and Wang et al. (2009) for a similar criterion for tuning parameters selection. Based on Condition C3, the optimal order of m should have the same order as $n^{1/(a+2b)}$ with $a > 1$ and $b > a/2 + 1$. Similar to Zhang and Liang (2011), we propose to choose the optimal knot number, m , from a neighborhood of $n^{1/5.5}$. In our simulation studies, we have used $[2/3Nr, 4/3Nr]$, where $Nr = \text{ceiling}(n^{1/5.5})$ and the function $\text{ceiling}(\cdot)$ returns the smallest integer not less than the corresponding element. Then the optimal knot number, m_{opt} , is the one which minimizes the SIC value. That is

$$m_{opt} = \arg \min_{m \in [2/3Nr, 4/3Nr]} SIC(m).$$

Base on 1,000 random perturbations, Table 1 summarizes the bias (Bias) and mean squared error (MSE) of the estimated θ with different sample sizes under different

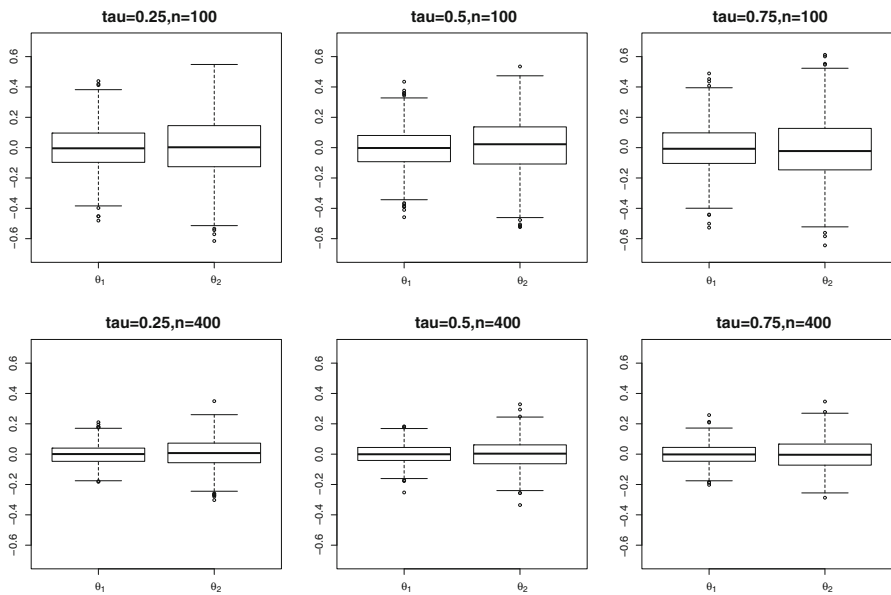


Fig. 1 The boxplots of resulting estimator of $\hat{\theta} - \theta_0$ at different quantile levels with different sample sizes for standard normal random error

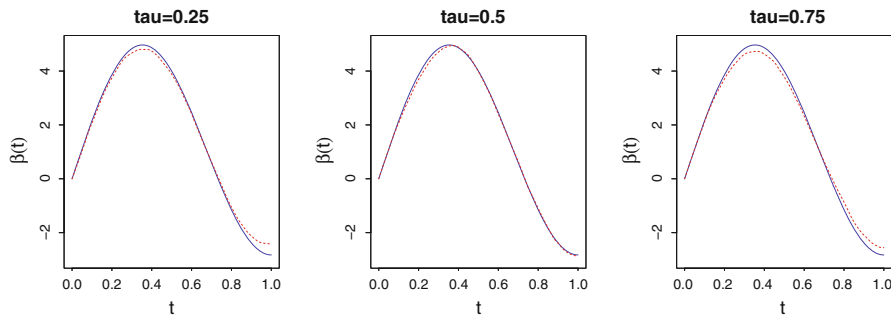


Fig. 2 The true $\beta(t)$ (blue) and $\hat{\beta}(t)$ (red) for standard normal random error with $n = 100$ (color figure online)

quantile level for different cases. Bias is reasonably small in general. We may conclude that this simulation study provides strong evidence in support of the asymptotic theory that we derive in Sect. 4.

Figure 1 shows the boxplots for standard normal random error with different sample sizes and quantile level. For the student t distribution with 3° and standard Cauchy random error distribution, the figures perform similarly with the standard normal distribution. To save space, we omit them. From the boxplots, we know that the proposed methods is consistent.

Figures 2, 3 and 4 demonstrate the performance of the curve estimation of the slope parameter $\beta(\cdot)$ for different cases under different levels with $n = 100$ and show that

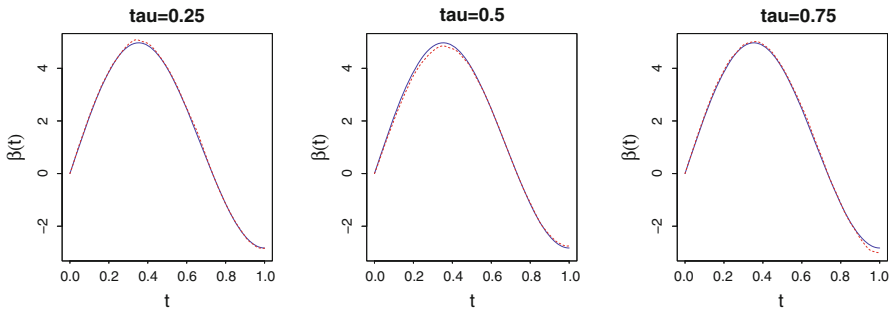


Fig. 3 The true $\beta(t)$ (blue) and $\hat{\beta}(t)$ (red) for standard Cauchy random error with $n = 100$ (color figure online)

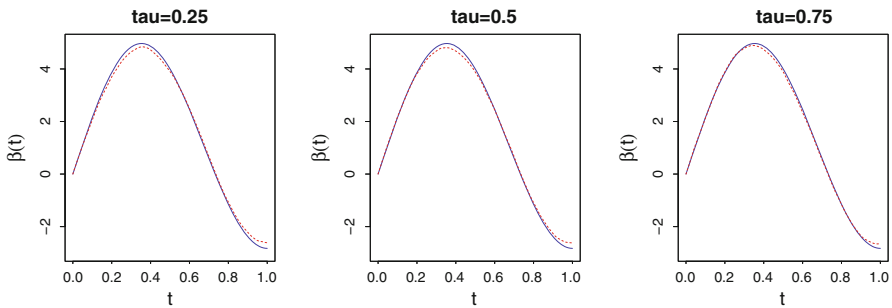


Fig. 4 The true $\beta(t)$ (blue) and $\hat{\beta}(t)$ (red) for $t(3)$ random error with $n = 100$ (color figure online)

the estimated curves are very close to the true curve $\beta(\cdot)$. We may conclude that the proposed estimation of the function $\beta(\cdot)$ performs reasonably well.

Acknowledgments The authors would like to thank the referees for their helpful comments that led an improvement of an early manuscript.

Appendix

Let $\mathbf{Z} = (\mathbf{Z}_1^T, \dots, \mathbf{Z}_n^T)^T$ and $\mathbf{U} = (\mathbf{U}_1, \dots, \mathbf{U}_n)^T$ be the n by p design matrix and the n by m design matrix for the constant slope parametric and the functional slope parametric component, respectively. Also let $\mathbf{P} = \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T$, $\mathbf{Z}^* = (\mathbf{I} - \mathbf{P})\mathbf{Z}$, $\mathbf{Z}^* = (\mathbf{Z}_1^*, \dots, \mathbf{Z}_n^*)^T$, $\mathbf{S}_n = \mathbf{Z}^{*T}\mathbf{Z}^*$.

Lemma 1 Under conditions C1–C5, one has

$$\frac{1}{n}\mathbf{S}_n = \Sigma + o_p(1).$$

Proof Let $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$ and $\mathbf{Z} = (\mathbf{Z} - \boldsymbol{\Delta}) + \boldsymbol{\Delta} = \boldsymbol{\eta} + \boldsymbol{\Delta}$, where $\boldsymbol{\Delta} = (\langle \mathbf{g}, X_1 \rangle, \dots, \langle \mathbf{g}, X_n \rangle)^T$ and $\mathbf{g} = (g_1, \dots, g_p)^T$ is defined in Condition C5. We can

write

$$\begin{aligned}
 S_n &= ((I - P)Z)^T ((I - P)Z) \\
 &= (\eta + \Delta)^T (I - P)^T (I - P) (\eta + \Delta) \\
 &= \eta^T \eta + \Delta^T (I - P)^T (I - P) \Delta + \Delta^T (I - P)^T (I - P) \eta \\
 &\quad + \eta^T (I - P)^T (I - P) \Delta - \eta^T P^T P \eta \\
 &= \eta^T \eta + I_{n1} + I_{n2} + I_{n3} + I_{n4},
 \end{aligned}$$

where

$$\begin{aligned}
 I_{n1} &= \Delta^T (I - P)^T (I - P) \Delta, \\
 I_{n2} &= \eta^T (I - P)^T (I - P) \Delta, \\
 I_{n3} &= \Delta^T (I - P)^T (I - P) \eta, \\
 I_{n4} &= \eta^T P^T P \eta.
 \end{aligned}$$

Invoking the central limit theorem, one has

$$\frac{1}{n} \eta^T \eta \rightarrow \Sigma, \quad a.s. \tag{7}$$

Invoking (7), to prove Lemma 1, it is enough to show that

$$\|I_{nl}\| = o_p(n), \quad l = 1, 2, 3, 4.$$

By Condition C2, the fact $\|\phi_j - \hat{\phi}_j\|^2 = O_p(n^{-1}j^2)$ and the Karhunen-Loève representation, one has

$$\begin{aligned}
 \|\Delta - UM\|^2 &= \sum_{i=1}^n \sum_{l=1}^p \left\| \langle x_i, g_l \rangle - \sum_{j=1}^m \langle x_i, \hat{\phi}_j \rangle \lambda'_{lj} \right\|^2 \\
 &= \sum_{i=1}^n \sum_{l=1}^p \left\| \sum_{j=1}^{\infty} \phi_j \lambda_{lj} - \sum_{j=1}^m \hat{\phi}_j \lambda'_{lj} \right\|^2 \\
 &= \sum_{i=1}^n \sum_{l=1}^p \left\| \sum_{j=m+1}^{\infty} \phi_j \lambda_{lj} + \sum_{j=1}^m (\phi_j \lambda_{lj} - \hat{\phi}_j \lambda'_{lj}) \right\|^2 \\
 &\leq 2 \sum_{i=1}^n \sum_{l=1}^p \sum_{j=m+1}^{\infty} \left\| \langle x_i, \phi_j \rangle \lambda_{lj} \right\|^2 + 2 \sum_{i=1}^n \sum_{l=1}^p \left\| \sum_{j=1}^m (\phi_j \lambda_{lj} - \hat{\phi}_j \lambda'_{lj}) \right\|^2 \\
 &\leq 2 \sum_{i=1}^n \sum_{l=1}^p \sum_{j=m+1}^{\infty} j^{-2b} + 2 \sum_{i=1}^n \sum_{l=1}^p \sum_{j=1}^m \|\phi_j (\lambda_{lj} - \lambda'_{lj}) + (\phi_j - \hat{\phi}_j) \lambda'_{lj}\|^2
 \end{aligned}$$

$$\begin{aligned}
 &= O_p \left(n^{-\frac{2b-1}{a+2b}} \right) n + O_p \left(n \sum_{j=1}^m \|\phi_j - \hat{\phi}_j\|^2 \right) \\
 &= O_p \left(n^{-\frac{2b-1}{a+2b}} \right) n + O_p \left(n \sum_{j=1}^m n^{-1} j^2 \right) \\
 &= O_p \left(n n^{-\frac{2b-1}{a+2b}} \right) + O_p \left(\frac{m(2m+1)(m+1)}{6} \right) \\
 &= O_p \left(n^{\frac{a+1}{a+2b}} \right) + O_p(m^3),
 \end{aligned}$$

where $\lambda'_{lj} = \langle \hat{\phi}_j, g_l \rangle$ and $\lambda_{lj} = \langle \phi_j, g_l \rangle, l = 1, \dots, p$.

By Condition C3, there exists a matrix \mathbf{M} such that $\|\mathbf{\Delta} - \mathbf{UM}\|^2 = O_p(n^{a_1/(a+2b)})$, where $a_1 = \max(3, a + 1)$. In addition, as \mathbf{P} is a projection matrix, we have

$$\begin{aligned}
 \|(\mathbf{I} - \mathbf{P})\mathbf{\Delta}\|^2 &= \|\mathbf{\Delta} - \mathbf{UM}\|^2 + \|\mathbf{UM} - \mathbf{P}\mathbf{\Delta}\|^2 \\
 &\leq 2\|\mathbf{\Delta} - \mathbf{UM}\|^2 = O_p(n^{a_1/(a+2b)}).
 \end{aligned} \tag{8}$$

By Condition C4 and the strong law of large numbers, one has $\frac{1}{n}\boldsymbol{\eta}^T \boldsymbol{\eta}$ converges almost surely to Σ . For $k \neq l$, one has

$$\begin{aligned}
 E \left\{ (\boldsymbol{\eta}^T \mathbf{P}\boldsymbol{\eta})_{kl} \right\}^2 &= (\Sigma_{kl})^2 (\text{trace}(\mathbf{P}))^2 + \left\{ \Sigma_{kk} \Sigma_{ll} + (\Sigma_{kl})^2 \right\} \text{trace}(\mathbf{P}\mathbf{P}^T) \\
 &\quad + \left\{ E[\boldsymbol{\eta}_{1k}^2 \boldsymbol{\eta}_{1l}^2] - 2(\Sigma_{kl})^2 - \Sigma_{kk} \Sigma_{ll} \right\} \sum_s \mathbf{P}_{ss}.
 \end{aligned} \tag{9}$$

In addition, as \mathbf{P} is a projection matrix, this expression is $O(m)$. Since \mathbf{P} is a positive semidefinite matrix, when $k = l$,

$$E \left\{ (\boldsymbol{\eta}^T \mathbf{P}\boldsymbol{\eta})_{kk} \right\} = \Sigma_{kk} \text{trace}(\mathbf{P}) = O(m). \tag{10}$$

Invoking (9) and (10), we have

$$\|\mathbf{P}\boldsymbol{\eta}\|^2 = O_p(m). \tag{11}$$

Similarly, we have

$$\|(\mathbf{I} - \mathbf{P})\boldsymbol{\eta}\|^2 = O_p(m). \tag{12}$$

Invoking (11) and (12) and Condition C2, we have

$$\begin{aligned}
 \|I_{n1}\| &= \|(\mathbf{I} - \mathbf{P})\mathbf{\Delta}\|^2 = O_p(n^{a_1/(a+2b)}) = o_p(n), \\
 \|I_{n2}\| &= \|I_{n3}\| = \|\boldsymbol{\eta}^T (\mathbf{I} - \mathbf{P})^T (\mathbf{I} - \mathbf{P}) \mathbf{\Delta}\|^2
 \end{aligned}$$

$$\begin{aligned} &\leq \|(\mathbf{I} - \mathbf{P})\boldsymbol{\eta}\|^2 \|(\mathbf{I} - \mathbf{P})\boldsymbol{\Delta}\|^2 \\ &= O_p(m) O_p\left(n^{\frac{a_1}{a+2b}}\right) \\ &= o_p(n), \end{aligned}$$

and

$$\|I_{n4}\| = \|\boldsymbol{\eta}^T \mathbf{P}^T \mathbf{P} \boldsymbol{\eta}\|^2 = O_p(m) = o_p(n).$$

The proof is hence complete. □

Lemma 2 *Under conditions C1–C5, we have*

$$\mathbf{S}_n^{-\frac{1}{2}} \mathbf{Z}^{*T} \boldsymbol{\psi}(\boldsymbol{\varepsilon}) \rightarrow N(0, \tau(1 - \tau)I_p),$$

where $\boldsymbol{\psi}(\boldsymbol{\varepsilon}) = (\psi(\varepsilon_1), \dots, \psi(\varepsilon_n))^T$.

Proof Invoking $\mathbf{Z} = \boldsymbol{\eta} + \boldsymbol{\Delta}$, we have

$$\begin{aligned} \mathbf{S}_n^{-\frac{1}{2}} \mathbf{Z}^{*T} \boldsymbol{\psi}(\boldsymbol{\varepsilon}) &= \mathbf{S}_n^{-\frac{1}{2}} ((\mathbf{I} - \mathbf{P})\mathbf{Z})^T \boldsymbol{\psi}(\boldsymbol{\varepsilon}) \\ &= \mathbf{S}_n^{-\frac{1}{2}} ((\mathbf{I} - \mathbf{P})(\boldsymbol{\eta} + \boldsymbol{\Delta}))^T \boldsymbol{\psi}(\boldsymbol{\varepsilon}) \\ &= \mathbf{S}_n^{-\frac{1}{2}} \boldsymbol{\eta}^T \boldsymbol{\psi}(\boldsymbol{\varepsilon}) - \mathbf{S}_n^{-\frac{1}{2}} (\mathbf{P}\boldsymbol{\eta})^T \boldsymbol{\psi}(\boldsymbol{\varepsilon}) + \mathbf{S}_n^{-\frac{1}{2}} ((\mathbf{I} - \mathbf{P})\boldsymbol{\Delta})^T \boldsymbol{\psi}(\boldsymbol{\varepsilon}). \end{aligned}$$

By Lemma 1, (8) and (11), we have

$$\|\mathbf{S}_n^{-\frac{1}{2}} (\mathbf{P}\boldsymbol{\eta})^T\|^2 = o_p(1) \quad \text{and} \quad \|\mathbf{S}_n^{-\frac{1}{2}} ((\mathbf{I} - \mathbf{P})\boldsymbol{\Delta})^T \boldsymbol{\psi}(\boldsymbol{\varepsilon})\|^2 = o_p(1).$$

Thus,

$$\mathbf{S}_n^{-\frac{1}{2}} \mathbf{Z}^{*T} \boldsymbol{\psi}(\boldsymbol{\varepsilon}) = \mathbf{S}_n^{-\frac{1}{2}} \boldsymbol{\eta}^T \boldsymbol{\psi}(\boldsymbol{\varepsilon}) + o_p(1).$$

By condition C5 and the central limit theorem, one has

$$\mathbf{S}_n^{-\frac{1}{2}} \mathbf{Z}^{*T} \boldsymbol{\psi}(\boldsymbol{\varepsilon}) \rightarrow N(0, \tau(1 - \tau)I_p).$$

□

Proof of Theorem 1 Let

$$\boldsymbol{\xi} \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\gamma} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix} = \begin{pmatrix} f(0)S_n^{\frac{1}{2}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ \mathbf{H}_m^{\frac{1}{2}}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_m) + \mathbf{H}_m^{-\frac{1}{2}} \mathbf{U}\mathbf{Z}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \end{pmatrix},$$

where $\mathbf{H}_m = m\mathbf{U}^T\mathbf{U}$. Let $\hat{\boldsymbol{\xi}} = \boldsymbol{\xi}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\gamma}}) = (\hat{\boldsymbol{\xi}}_1^T, \hat{\boldsymbol{\xi}}_2^T)^T$. Now, we show that $\|\hat{\boldsymbol{\xi}}\| = O_p(\delta_n)$. To do so, let $\tilde{\mathbf{Z}}_i = \frac{1}{f(0)}\mathbf{S}_n^{-\frac{1}{2}}\mathbf{Z}_i$, $\tilde{\mathbf{U}}_i = \mathbf{H}_m^{-1}\mathbf{U}_i$, $R_i = \sum_{j=1}^m \langle x_i, \hat{\phi}_j \rangle \gamma_{j0} - \int_0^1 \beta(t)x(t)dt$.

Note that $\|\phi_j - \hat{\phi}_j\|^2 = O_p(n^{-1}j^2)$, one has

$$\begin{aligned} \|R_i\|^2 &= \left\| \sum_{j=1}^m \langle x_i, \hat{\phi}_j \rangle \gamma_{j0} - \sum_{j=1}^{\infty} \langle x_i, \phi_j \rangle \gamma_{j0} \right\|^2 \\ &\leq \left\| \sum_{j=1}^m \langle x_i, \hat{\phi}_j \rangle \gamma_{j0} - \sum_{j=1}^m \langle x_i, \phi_j \rangle \gamma_{j0} \right\|^2 + \left\| \sum_{j=m+1}^{\infty} \langle x_i, \phi_j \rangle \gamma_{j0} \right\|^2 \\ &\leq \sum_{j=1}^m \|\hat{\phi}_j - \phi_j\|^2 |\gamma_{j0}|^2 + \sum_{j=m+1}^{\infty} |\gamma_{j0}|^2 \\ &= \sum_{j=1}^m O_p(n^{-1}j^{2-2b}) + \sum_{j=m+1}^{\infty} j^{-2b} \\ &= O_p\left(n^{-\frac{2b-1}{a+2b}}\right) + O\left(n^{-(2b-1)/(a+2b)}\right) \\ &= O_p\left(n^{-\frac{2b-1}{a+2b}}\right). \end{aligned}$$

Thus, one has

$$\sum_{i=1}^n \rho_{\tau}(Y_i - \mathbf{Z}_i^T \boldsymbol{\theta} - \mathbf{U}_i^T \boldsymbol{\gamma}) = \sum_{i=1}^n \rho_{\tau}\left(\varepsilon_i - \tilde{\mathbf{Z}}_i^T \boldsymbol{\xi}_1 - \tilde{\mathbf{U}}_i^T \boldsymbol{\xi}_2 - R_i\right),$$

which is minimized at $\hat{\boldsymbol{\xi}}$.

By similar arguments to these of Lemma 1 of Cardot et al. (2005) for any $\kappa > 0$, there exists L_{κ} such that

$$P \left\{ \inf_{\|\boldsymbol{\xi}\| > L_{\kappa} \delta_n} \sum_{i=1}^n \rho_{\tau}\left(\varepsilon_i - \tilde{\mathbf{Z}}_i^T \boldsymbol{\xi}_1 - \tilde{\mathbf{U}}_i^T \boldsymbol{\xi}_2 - R_i\right) > \sum_{i=1}^n \rho_{\tau}(\varepsilon_i - R_i) \right\} > 1 - \kappa.$$

On the other hand, we have

$$\sum_{i=1}^n \rho_{\tau}\left(\varepsilon_i - \tilde{\mathbf{Z}}_i^T \hat{\boldsymbol{\xi}}_1 - \tilde{\mathbf{U}}_i^T \hat{\boldsymbol{\xi}}_2 - R_i\right) = \inf_{\boldsymbol{\xi} \in R^{m+p}} \rho_{\tau}\left(\varepsilon_i - \tilde{\mathbf{Z}}_i^T \boldsymbol{\xi}_1 - \tilde{\mathbf{U}}_i^T \boldsymbol{\xi}_2 - R_i\right). \tag{13}$$

Thus, we have

$$\sum_{i=1}^n \rho_{\tau}\left(\varepsilon_i - \tilde{\mathbf{Z}}_i^T \hat{\boldsymbol{\xi}}_1 - \tilde{\mathbf{U}}_i^T \hat{\boldsymbol{\xi}}_2 - R_i\right) < \sum_{i=1}^n \rho_{\tau}(\varepsilon_i - R_i).$$

Then connecting this with Eq. (13), we obtain

$$P \left\{ \inf_{\|\xi\| > L\kappa\delta_n} \sum_{i=1}^n \rho_\tau \left(\varepsilon_i - \tilde{\mathbf{Z}}_i^T \xi_1 - \tilde{\mathbf{U}}_i^T \xi_2 - R_i \right) > \sum_{i=1}^n \rho_\tau \left(\varepsilon_i - \tilde{\mathbf{Z}}_i^T \hat{\xi}_1 - \tilde{\mathbf{U}}_i^T \hat{\xi}_2 - R_i \right) \right\} > 1 - \kappa.$$

Thus, $\|\hat{\xi}\| = O_p(\delta_n)$. This together with Lemma 1, and the definition of $\hat{\xi}$, one has

$$\|\hat{\theta} - \theta_0\| = \left\| \frac{1}{f(0)} \mathbf{S}_n^{-\frac{1}{2}} \hat{\xi}_1 \right\| = O_p \left(n^{-\frac{1}{2}} \|\hat{\xi}_1\| \right) = O_p(n^{-\frac{1}{2}} \delta_n).$$

Note that

$$\begin{aligned} \|\hat{\beta}(t) - \beta_0(t)\|^2 &= \left\| \sum_{j=1}^m \hat{\gamma}_j \hat{\phi}_j - \sum_{j=1}^\infty \gamma_j \phi_j \right\|^2 \\ &\leq 2 \left\| \sum_{j=1}^m \hat{\gamma}_j \hat{\phi}_j - \sum_{j=1}^m \gamma_j \phi_j \right\|^2 + 2 \left\| \sum_{j=m+1}^\infty \gamma_j \phi_j \right\|^2 \\ &\leq 4 \left\| \sum_{j=1}^m (\hat{\gamma}_j - \gamma_j) \hat{\phi}_j \right\|^2 + 4 \left\| \sum_{j=1}^m \gamma_j (\hat{\phi}_j - \phi_j) \right\|^2 + 2 \sum_{j=m+1}^\infty \gamma_j^2 \\ &= K_{n1} + K_{n2} + K_{n3}. \end{aligned}$$

Now we consider K_{n1} , by the fact that the sequences $\{\hat{\phi}_j\}$ forms an orthonormal basis in $L^2([0, 1])$, one has

$$\begin{aligned} K_{n1} &= \left\| \sum_{j=1}^m (\hat{\gamma}_j - \gamma_j) \hat{\phi}_j \right\|^2 \\ &\leq \sum_{j=1}^m (\hat{\gamma}_j - \gamma_j)^2 \\ &\leq \|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_m\|^2. \end{aligned}$$

By Lemma 1 of Stone (1985), it is easy to show that \mathbf{H}_m is positive definite for sufficiently large n . Therefore, one has

$$\begin{aligned} \|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_m\|^2 &\leq C(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_m)^T \mathbf{H}_m (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_m) \\ &\leq O_p(n^{-1} \|\hat{\xi}_2\|^2) + O_p(\|\hat{\theta} - \theta_0\|) = O_p(\delta_n^2). \end{aligned}$$

As a result, we have $K_{n1} = O_p(\delta_n^2)$.

$$\begin{aligned} K_{n2} &\leq m \sum_{j=1}^m \|\hat{\phi}_j - \phi_j\|^2 \gamma_j^2 \leq n^{-1} m \sum_{j=1}^m j^2 \gamma_j^2 \\ &= O_p\left(n^{-1} m \sum_{j=1}^m j^{2-2b}\right) = O_p\left(n^{-\frac{a+3b-1}{a+2b}}\right) = o_p\left(\delta_n^2\right), \end{aligned}$$

$$K_{n3} = \sum_{j=m+1}^{\infty} \gamma_j^2 \leq C \sum_{j=m+1}^{\infty} j^{-2b} = O(n^{-\frac{2b-1}{a+2b}}) = o\left(\delta_n^2\right).$$

Therefore, one has

$$\|\hat{\beta} - \beta\|^2 = O_p\left(\delta_n^2\right).$$

Next we will show the asymptotic normality of $\hat{\theta}$, let $\xi_1^* = \frac{1}{f(0)} S_n^{-\frac{1}{2}} \sum_{i=0}^n Z_i^* \psi_{\tau}(\varepsilon_i)$, according to Lemmas 1 and 2, ξ_1^* is asymptotically normal with variance-covariance $\frac{\tau(1-\tau)}{f^2(0)} I_p$.

On the other hand, similar to He and Shi (1996), we can proof that $\|\hat{\xi}_1^* - \hat{\xi}_1\| = o_p(1)$. Thus,

$$\hat{\xi}_1 = \hat{\xi}_1^* + o_p(1) = \frac{1}{f(0)} S_n^{-\frac{1}{2}} \sum_{i=0}^n Z_i^* \psi_{\tau}(\varepsilon_i) + o_p(1).$$

Obviously,

$$\sqrt{n}(\hat{\theta}_0 - \theta_0) \rightarrow N\left(0, \frac{\tau(1-\tau)}{f^2(0)} \Sigma\right).$$

This completes the proof of Theorem 1.

References

Aneiros-Pérez G, Vieu P (2006) Semi-functional partial linear regression. *Stat Prob Lett* 76:1102–1110
 Cardot H, Crambes C, Sarda P (2005) Quantile regression when the covariates are functions. *J Nonparametr Stat* 17:841–856
 Engle R, Granger C, Rice J, Weiss A (1986) Semiparametric estimates of the relation between weather and electricity sales. *J Am Stat Assoc* 81:310–320
 Fan YQ, Li Q (1999) Root-n-consistent estimation of partially linear time series models. *J Nonparametr Stat* 11:251–269
 Gao JT (1995) Asymptotic theory for partly linear models. *Commun Stat Theory Methods* 24:1985–2009
 Hall P, Horowitz JL (2007) Methodology and convergence rates for functional linear regression. *Ann Stat* 35:70–91

- Härdle W, Liang H, Gao JT (2000) Partially linear models. Physica Verlag, Heidelberg
- He X, Liang H (2000) Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statistica Sinica* 1:129–140
- He X, Shi PD (1996) Bivariate tensor-product B-spline in a partly linear model. *J Multivar Anal* 58:162–181
- He X, Zhu ZY, Fung WK (2002) Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* 89:579–590
- Koenker R (2005) Quantile regression. Cambridge University Press, Cambridge
- Koenker R, Bassett G (1978) Regression quantiles. *Econometrica* 46:33–51
- Liu Q (2011) Asymptotic normality for the partially linear EV models with longitudinal data. *Commun Stat Theory Methods* 40:1149–1158
- Mammen E, Geer S (1997) Penalized quasi-likelihood estimation in partial linear models. *Ann Stat* 25:1014–1035
- Moyeed RA, Diggle PJ (1994) Rate of convergence in semiparametric modeling of longitudinal data. *Aust J Stat* 36:75–93
- Ramsay J, Silverman B (2005) Functional data analysis, 2nd edn. Springer, New York
- Shi P, Li G (1994) On the rate of convergence of minimum L_1 -norm estimates in a partly linear model. *Commun Stat Theory Methods* 23:175–196
- Shin H (2009) Partial functional linear regression. *J Stat Plan Inference* 139:3405–3418
- Speckman P (1988) Kernel smoothing in partial linear models. *J R Stat Soc Ser B* 50:413–436
- Stone CJ (1985) Additive regression and other nonparametric models. *Ann Stat* 13:689–705
- Wang H, Zhu ZY, Zhou JH (2009) Quantile regression in partially linear varying coefficient models. *Ann Stat* 37:3841–3866
- Zhang X, Liang H (2011) Focused information criterion and model averaging for generalized additive partial linear models. *Ann Stat* 39:174–200