Research Article

Peng Cao* and Jun Sun

Robust estimation for partial functional linear regression models based on FPCA and weighted composite quantile regression

Abstract: In this paper, we consider a novel estimation for partial functional linear regression models. The functional principal component analysis method is employed to estimate the slope function and the functional predictive variable, respectively. An efficient estimation based on principal component basis function approximation is used for minimizing the proposed weighted composite quantile regression (WCQR) objective function. Since the proposed WCQR involves a vector of weights, we develop a computational strategy for data-driven selection of the optimal weights. Under some mild conditions, the theoretical properties of the proposed WCQR method are obtained. The simulation study and a real data analysis are provided to illustrate the numerical performance of the resulting estimators.

Keywords: partial functional linear regression, weighted composite quantile regression, functional principal component analysis, asymptotic properties

MSC 2020: 62J05, 62M10

1 Introduction

In the literature of functional data analysis, functional linear regression models (FLRMs) provide an efficient method to analyze the relationship between a functional predictor and a scalar response. Therefore, much effort has been devoted to studying its estimation and other relevant inference problems, see [1–13]. However, in practice, we often see that a scalar response is related not only to functional covariates but also to scalar covariates. Thus, Shin considered partial functional linear regression models (PFLRMs) to balance the flexibility of FLRMs and interpretation of classical linear regression models [14]. Generally, PFLRM has the following form:

$$Y = \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\theta} + \int_{\mathcal{T}} X(t)\beta(t)\mathrm{d}t + \varepsilon, \tag{1}$$

where Y is a real-valued response variable defined on a probability space $(\Omega, \mathcal{B}, P)$, $\boldsymbol{Z}$ is a p-dimensional random vector and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^{\mathrm{T}}$ is an unknown parameter vector, $\{X(t); t \in \mathcal{T}\}$ is a zero mean, second-order stochastic process defined on $(\Omega, \mathcal{B}, P)$ with sample paths in $H = L^2(\mathcal{T})$, the Hilbert space containing

* **Corresponding author: Peng Cao,** School of Mathematics and Statistics, Hunan Normal University, Changsha, 410081, P. R. China; Key Laboratory of Computing and Stochastic Mathematics (Ministry of Education), School of Mathematics and Statistics, Hunan Normal University, Changsha, Hunan 410081, P. R. China, e-mail: cp@hunnu.edu.cn
Jun Sun: School of Mathematics and Statistics, Hunan Normal University, Changsha, 410081, P. R. China; Key Laboratory of Computing and Stochastic Mathematics (Ministry of Education), School of Mathematics and Statistics, Hunan Normal University, Changsha, Hunan 410081, P. R. China

square integrable functions on $\mathcal{T}$ with inner product $\langle x, y \rangle = \int_{\mathcal{T}} x(t)y(t)\mathrm{d}t$, $\forall x, y \in H$ and norm $\|x\| = \langle x, x \rangle^{1/2}$, $\beta(t)$ is an unknown slope function belonging to H, ε is the random error independent of $\mathbf{Z}$ and X, $E(\varepsilon|\mathbf{Z}, X) = 0$. $\mathcal{T}$ denotes the transport operation throughout this paper. Without loss of generality, we further assume $\mathcal{T} = [0, 1]$.

The PFLRMs have been studied by many authors. For example, Shin [14] proposed the estimation method based on functional principal component analysis (FPCA) and investigated the asymptotic properties of the estimators for model (1). Zhou et al. [15] considered the polynomial spline method to estimate model (1) and established the asymptotic normality for the parameter vector and the global convergence rate for the slope function. Furthermore, Kong et al. [16] studied variable selection in model (1) with multiple functional and ultrahigh-dimensional scalar predictors. Ling et al. [17] developed the k-nearest-neighbors estimation for PFLRMs. Kong et al. [18] proposed the partial functional linear cox regression model for censored outcomes. Cao et al. [19] developed a two-step estimation procedure via maximizing the quasi likelihood function in the framework of generalized PFLRMs. Yuan and Zhang [20] investigated the hypothesis test of the parametric components in PFLRMs based on B-spline. Zhu et al. [21] considered estimation and testing problems for PFLRMs when the covariates of the non-functional linear component are measured with additive error. Li et al. [22] proposed two test statistics for series correlation in PFLRMs and derived their asymptotic distributions under the null hypothesis.

Quantile regression (QR) is an effective approach for functional data analysis. This is due to the fact that QR models compared to the mean regression models [23] offer more robust merits especially the observations contain some outliers. For relevant literature, see [24–26]. Hence, QR method may result in an arbitrarily small relative efficiency compared with the mean regression. To conquer this drawback of QR method, Zou and Yuan proposed the composite quantile regression (CQR) method for linear models and showed that the relative efficiency of the CQR estimator is greater than 70% regardless of the error distribution compared with the least square (LS) estimator [27]. In many cases, CQR provides more efficient estimation compared with a single QR or the LS estimator for non-normal error distributions. This method has been deeply investigated in the literature and widely applied in many statistical models. For further references about the CQR method on nonparametric/semi-parametric models see [28–32]. Even though CQR has been well developed, theory and methodology of PFLRMs are rare. Du et al. [33] used the penalized CQR method to study variable selection for parametric part in the PFLRMs, Yu et al. [34] considered composite quantile estimation for the PFLRMs with errors from short-range dependent and strictly stationary linear processes. However, the aforementioned CQR method for PFLRMs is a sum of different QRs with equal weights. Intuitively, using equal weights is not optimal in general, though CQR enjoys great advantages in terms of estimation efficiency. Therefore, Jiang et al. [35] addressed the issue of statistical inference for nonlinear models with large-dimensional covariates by using the weighted composite quantile regression (WCQR) method. Moreover, they pointed out that their method was more robust than some existing methods. Guo et al. [36] investigated group SCAD penalized WCQR estimation for varying coefficient models with diverging number of parameters. Jiang et al. [37] developed a data-driven WCQR for heteroscedastic partially linear varying coefficient models. Similar conclusions of WCQR method have been further confirmed in [6,38,39]. Until now, the WCQR method has not been used in functional regression analysis. Consequently, motivated by this fact, we devote to extending the WCQR method to PFLRMs in this paper. We show that the resulting estimators for both parametric components and slope function are more efficient in the case of asymmetrical error distribution, and as asymptotically efficient as the corresponding LS estimators when the error distribution is normal.

It is well known that many smoothing methods have been proposed for functional regressions, and two popular approaches for estimating the slope function $\beta(\cdot)$ include the FPCA approach based on spectral decompositions of both the covariance of $X(t)$ and its estimator (see [3,7,40]) and the basis function expansion method with both $\beta(\cdot)$ and $X(t)$ approximated by basis functions such as polynomial spline functions in [15]. In this paper, we propose a WCQR estimation method for PFLRMs through weighted composite quantile loss function with principal component basis function approximations. The main contributions of this paper are threefold:

(I) This is the first attempt to provide robust estimate via WCQR and establish the asymptotic properties of the proposed estimators for model (1).

(II) We develop a data-driven weighting strategy that maximizes the efficiency of the WCQR estimators. Practically, the proposed method is more efficient than the CQR method in [33] and rank regression method proposed under asymmetrical error distributed data set in [38].

(III) Theoretically, the proposed WCQR method generalizes the results of the composite quantile estimators in [8]. Particularly, our method reduces to that of Du et al. when the weights are all equal.

The rest of this paper is organized as follows. In Section 2, we describe the implementation details of the proposed algorithm and establish the theoretical properties of the estimators. In Section 3, we discuss the selection of weights and tuning parameters. Simulation study and a real data analysis are conducted in Sections 4 and 5 to evaluate the performance of our proposed approach, and then followed by a short conclusion in Section 6. All technical proofs are collected in the Appendix.

2 Method and main results

In partial functional linear QR model, for a given quantile level $\tau \in (0, 1)$, $(\mathbf{Z}_i, X_i(\cdot), Y_i)$, $i = 1, \ldots, n$ is an independent and identically distributed (i.i.d.) sample from model (1), that is

$$Q_\tau(Y|Z, X(t)) = \mathbf{Z}^{\mathrm{T}}\boldsymbol{\theta}_\tau + \int_{\mathcal{T}} X(t)\beta_\tau(t)\mathrm{d}t + \varepsilon_\tau, \tag{2}$$

where $Q_\tau(Y|Z, X(t))$ is the τth conditional quantile of Y given $(Z, X(t))$, Y is a real-valued random variable defined on a probability space $(\Omega, \mathcal{B}, P)$, Z is a p-dimensional vector of random variables with finite second moments, $\boldsymbol{\theta}_\tau$ is a $p \times 1$ coefficient vector to be estimated, $\beta_\tau(t)$ is an unknown square integrable function on $[0, 1]$, and ε_τ is a random error whose τth quantile conditional on $(Z, X(t))$ being zero. For the convenience of presentation, we will omit τ from $\boldsymbol{\theta}_\tau$, $\beta_\tau(t)$ and ε_τ in model (2) wherever clear from the context, but we should keep in mind that those quantities are τ-specific.

Denote the covariance function of the process $X(\cdot)$ and its empirical version, respectively, as

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

The covariance function K defines a linear operator which maps a function f to Kf given by $(Kf)(u) = \int K(u, v)f(v)\mathrm{d}v$. We assume that the linear operator with kernel K is positive definite. By Mercer's theorem, then

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

where $\lambda_1 > \lambda_2 > \cdots > 0$ and $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_{n+1} = \cdots = 0$ are, respectively, the ordered eigenvalue sequences of the linear operators with kernels K and $\hat{K}$, and $\{\phi_j\}$ and $\{\hat{\phi}_j\}$ are the corresponding orthonormal eigenfunction sequences. Obviously, the vectors $\{\phi_j\}$ and $\{\hat{\phi}_j\}$ each forms an orthonormal basis in H. According to the Karhunen-Loèvere presentation, then

$$X(t) = \sum_{i=1}^{\infty} \xi_i \phi_i(t), \quad \beta(t) = \sum_{i=1}^{\infty} \gamma_i \phi_i(t), \tag{3}$$

where ξ_i and γ_i are defined by

$$\xi_i = \int_0^1 X(t)\phi_i(t)\mathrm{d}t, \quad \gamma_i = \int_0^1 \beta(t)\phi_i(t)\mathrm{d}t,$$

where ξ_i is referred to as the ith functional principal component score of $X(t)$. It follows that ξ_i are uncorrelated random variables with mean 0 and variance $E(\xi_i^2) = \lambda_i$. For more details see [41]. Substituting (3) into model (2), the model (2) can be written as

$$Y = \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\theta} + \sum_{j=1}^{\infty}\gamma_j\langle\phi_j, X\rangle + \varepsilon. \tag{4}$$

Therefore, the regression model in (4) can be well approximated by

$$Y \approx \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\theta} + \sum_{j=1}^{m}\gamma_j\langle\phi_j, X\rangle + \varepsilon, \tag{5}$$

where $m \leq n$ is the truncation level that trades off approximation error against variability and typically diverges with n. Replace ϕ_j by $\hat{\phi}_j$ for $j = 1, \ldots, m$, model (5) can be rewritten as

$$Y \approx \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\theta} + \boldsymbol{U}^{\mathrm{T}}\boldsymbol{\gamma} + \varepsilon, \tag{6}$$

where $\boldsymbol{U} = \{\langle X, \hat{\phi}_j\rangle\}_{j=1, \ldots, m}$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^{\mathrm{T}}$. Thus, the estimation problem of the parameters and the slope function are transformed into the estimation problem of the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$.

Note that the CQR method used the same weight for different QR models. Intuitively, it will be more effective if different weights are used. According to Jiang et al. [35], the WCQR procedure estimates $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ via minimizing the following loss function:

$$\mathcal{L}_n(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{c}; \boldsymbol{\omega}) = \sum_{k=1}^{q}\omega_k\sum_{i=1}^{n}\rho_{\tau_k}(Y_i - c_{\tau_k} - \boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\theta} - \boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{\gamma}), \tag{7}$$

where $\rho_{\tau_k}(r) = \tau_k r - rI$ $(r < 0)$, $k = 1, \ldots, q$, q is the number of quantiles, $\tau_k \in (0, 1)$. Typically, we use equally spaced quantile position: $\tau_k = k/(1 + q)$, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_q)^{\mathrm{T}}$ is a vector of weight satisfying $\omega_k \geq 0$. c_{τ_k} is $100\tau_k\%$ quantile of ε, $\boldsymbol{c} = (c_{\tau_1}, \ldots, c_{\tau_q})^{\mathrm{T}}$. Let the coefficients $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_m)^{\mathrm{T}}$ and $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)^{\mathrm{T}}$ be the solution of equation (6), then the resulting estimator of the slope function is given by

$$\hat{\beta}(t) = \sum_{i=1}^{m}\hat{\gamma}_i\hat{\phi}_i(t).$$

3 Assumptions and asymptotic results

To establish the asymptotic properties of the proposed estimators, first, we introduce some notations. Let $\boldsymbol{\theta}_0$ and $\beta_0(t)$ be the true values of $\boldsymbol{\theta}$ and $\beta(t)$, respectively, the notation $\|\cdot\|$ is the L^2 norm for a function or the Euclidean norm for a vector. Let C denote a generic constant that might assume different values at different places, $a_n \sim b_n$ means that a_n/b_n is bounded away from 0 and infinity as $n \to \infty$. The following technical assumptions are imposed.

(A1) The random function $X(\cdot)$ satisfies $E\|X(\cdot)\|^4 < \infty$.

(A2) For each j, $E[U_j^4] \leq C\lambda_j^2$. 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 $b > a/2 + 1$.

(A3) The tuning parameter m satisfies $m \sim n^{1/(a+2b)}$.

(A4) The random vector $\boldsymbol{Z}$ has bounded fourth moment, $E\|\boldsymbol{Z}\|^4 < \infty$.

(A5) $\boldsymbol{Z} = \boldsymbol{\eta} + \langle\boldsymbol{g}, X\rangle$, where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)^{\mathrm{T}}$ is zero-mean random variable, $\boldsymbol{g} = (g_1, \ldots, g_p)^{\mathrm{T}}$ with $g_j \in L^2([0, 1])$, $j = 1, \ldots, p$.

(A6) $E[\boldsymbol{\eta}] = 0$ and $E[\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}] = \boldsymbol{\Sigma}$. Furthermore, we need that $\boldsymbol{\Sigma}$ is a positive definite matrix.

(A7) ε has the distribution function $F(\cdot)$ with density $f(\cdot)$. The density function $f(\cdot)$ of ε is positive and continuous at the τ_kth quantiles c_{τ_k}.

Remark 1. Assumptions A1–A3 are required in the classical functional linear regression (see [5,7,24]). Specifically, Assumption A1 ensures the consistency of $\hat{K}(s, t)$. Assumption A2 makes the slope function sufficiently smooth to be approximated [14]. Assumption A3 gives the order of the truncation parameter m

to obtain the convergence rate of the slope function. Assumptions A4–A5 are quite parallel to [24,33]. Assumption A6 is used to establish the asymptotic normality of the parameters in model. Assumption A7 is very common in WCQR, see [35].

Theorem 3.1. *Suppose that Assumptions* A1–A7 *are satisfied, then*

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

(ii) $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to N(0, \sigma^2(\boldsymbol{\omega})\boldsymbol{\Sigma}^{-1})$,

where

$$\sigma^2(\boldsymbol{\omega}) = \left\{\sum_{k=1}^{q} \omega_k f(c_{\tau_k})\right\}^{-2} \sum_{k,k'=1}^{q} \omega_k \omega_{k'} \min(\tau_k, \tau_{k'})(1 - \max(\tau_k, \tau_{k'})).$$

Remark 2. The result of Theorem 3.1(i) indicates that the estimator $\widehat{\beta}(t)$ has the same rate of convergence as for the estimators of [7,20], which is optimal in minimax sense; (ii) indicates that if the tuning parameter m and weight $\boldsymbol{\omega}$ are properly chosen, the WCQR estimator of parameter vector is $\sqrt{n}$-consistent.

Remark 3. When $q = 1$ and $\tau_1 = \tau$, denote the τth QR estimate of $\boldsymbol{\theta}$ and $\beta(t)$ by $\widehat{\boldsymbol{\theta}}^{\mathrm{QR}}$ and $\widehat{\beta}^{\mathrm{QR}}(t)$, respectively. When all ω_k are equal, denote the CQR estimate of $\boldsymbol{\theta}$ and $\beta(t)$ by $\widehat{\boldsymbol{\theta}}^{\mathrm{CQR}}$ and $\widehat{\beta}^{\mathrm{CQR}}(t)$, respectively. Similarly, we denote the LS estimators of $\boldsymbol{\theta}$ and $\beta(t)$ by $\widehat{\boldsymbol{\theta}}^{\mathrm{LS}}$ and $\widehat{\beta}^{\mathrm{LS}}(t)$, respectively. After some simple calculations, we can draw the following conclusions (1)–(3) and generalize the results of the LS estimators in [14], and the single-level quantile estimators in [24] and the composite quantile estimators in [33], respectively.

(1) Under the assumptions of Theorem 3.1, then
   (i) $\|\widehat{\beta}^{\mathrm{LS}}(t) - \beta_0(t)\|^2 = O_p(n^{-(2b-1)/(a+2b)})$,
   (ii) $\sqrt{n}(\widehat{\boldsymbol{\theta}}^{\mathrm{LS}} - \boldsymbol{\theta}_0) \to N(0, \boldsymbol{\Sigma}_{\mathrm{LS}})$;
(2) Under the assumptions of Theorem 3.1, then
   (i) $\|\widehat{\beta}^{\mathrm{QR}}(t) - \beta_0(t)\|^2 = O_p(n^{-(2b-1)/(a+2b)})$,
   (ii) $\sqrt{n}(\widehat{\boldsymbol{\theta}}^{\mathrm{QR}} - \boldsymbol{\theta}_0) \to N(0, \boldsymbol{\Sigma}_{\mathrm{QR}})$;
(3) Under the assumptions of Theorem 3.1, then
   (i) $\|\widehat{\beta}^{\mathrm{CQR}}(t) - \beta_0(t)\|^2 = O_p(n^{-(2b-1)/(a+2b)})$,
   (ii) $\sqrt{n}(\widehat{\boldsymbol{\theta}}^{\mathrm{CQR}} - \boldsymbol{\theta}_0) \to N(0, \boldsymbol{\Sigma}_{\mathrm{CQR}})$,

where $\boldsymbol{\Sigma}_{\mathrm{LS}} = \sigma^2\boldsymbol{\Sigma}^{-1}$ with $\sigma^2 = E(\varepsilon^2)$, $\boldsymbol{\Sigma}_{\mathrm{QR}} = \frac{\tau(1-\tau)}{f^2(c_\tau)}\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Sigma}_{\mathrm{CQR}} = \frac{\sum_{k,k'=1}^{q} \min(\tau_k, \tau_{k'})(1-\max(\tau_k, \tau_{k'}))}{\left(\sum_{k=1}^{q} f(c_{\tau_k})\right)^2}\boldsymbol{\Sigma}^{-1}$.

4 Computational issues

From Theorem 3.1, we find that the asymptotic variance of $\widehat{\boldsymbol{\theta}}$ depends on $\boldsymbol{\omega}$ only through $\sigma^2(\boldsymbol{\omega})$. Thus, the optimal choice of weights for maximizing efficiency of the estimator $\widehat{\boldsymbol{\theta}}$ is $\boldsymbol{\omega}_{\mathrm{opt}} = \arg\min_{\boldsymbol{\omega}} \sigma^2(\boldsymbol{\omega})$. Based on [36], the optimal choice of weights is to solve the following quadratic optimization problem:

$$\boldsymbol{\omega}_{\mathrm{opt}} = \arg\min_{\boldsymbol{\omega}}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Omega}\omega) \text{ subject to } \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{f} = 1, \boldsymbol{\omega} \geqslant 0, \tag{8}$$

where $\boldsymbol{f} = (f(c_{\tau_1}), \ldots, f(c_{\tau_q}))^{\mathrm{T}}$ and $\boldsymbol{\Omega}$ is a $q \times q$ matrix with the (k, k') element $\boldsymbol{\Omega}_{k,k'} = \min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}$.

It can be seen from (8) that the optimal weight vector is rather complicated and involves the density of the errors $c_k = F^{-1}(\tau_k)$ and $f(c_k)$, $k = 1, \ldots, q$. In practice, the error density $f(\cdot)$ is generally unknown. Following [36], we propose an estimation procedure as follows.

Step 1. We give the initial estimators $\tilde{\boldsymbol{\theta}}$ and $\tilde{\beta}(t)$ by minimizing the LS objective function and estimate $\sigma^2(\cdot)$ by

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - \mathbf{Z}_i^{\mathrm{T}}\tilde{\boldsymbol{\theta}} - \int_0^1 X_i(t)\tilde{\beta}(t)\mathrm{d}t\right\}^2.$$

Step 2. Compute $\tilde{\varepsilon}_i = \left(Y_i - \mathbf{Z}_i^{\mathrm{T}}\tilde{\boldsymbol{\theta}} - \int_0^1 X_i(t)\tilde{\beta}(t)\mathrm{d}t\right)\Big/\tilde{\sigma}$ and then we can use kernel density estimation to estimate $f(\cdot)$ as follows:

$$\tilde{f}(\cdot) = \frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{\tilde{\varepsilon}_i - \cdot}{h}\right).$$

Similar to Silverman [42], the bandwidth h is taken as

$$h = 0.9 \times \min\left(\mathrm{std}(\tilde{\varepsilon}_1,\ldots,\tilde{\varepsilon}_n), \frac{\mathrm{IQR}(\tilde{\varepsilon}_1,\ldots,\tilde{\varepsilon}_n)}{1.34}\right) \times n^{-1/5}.$$

$K(\cdot)$ is always chosen as the Gaussian kernel. Here $\mathrm{std}(\cdot)$ and $\mathrm{IQR}(\cdot)$ denote the sample standard deviation and sample interquartile, respectively.

Step 3. Estimate $f\{F^{-1}(\tau_k)\}$ by $\tilde{f}\{\tilde{F}^{-1}(\tau_k)\}$, where $\tilde{F}^{-1}(\tau_k)$ is the sample τ_k quantile of $\{\tilde{\varepsilon}_i,\ i = 1,\ldots,n\}$.

The proposed estimators involve the tuning parameter m, which needs to be selected by minimizing some selection criteria such as cross validation (CV), generalized cross validation (GCV), AIC information criterion, BIC information criterion, etc. We use a BIC-type criterion to select m [36], that is

$$\mathrm{BIC}(m) = \log\left\{\sum_{k=1}^{q}\omega_k\sum_{i=1}^{n}\rho_{\tau_k}(Y_i - \hat{c}_{\tau_k}^{(m)} - \mathbf{Z}_i^{\mathrm{T}}\hat{\boldsymbol{\theta}}^{(m)} - \mathbf{U}_i^{\mathrm{T}}\hat{\boldsymbol{\gamma}}^{(m)})\right\} + \frac{\log n}{2n}(m + p + q),$$

where p is the number of components in parameter vector. $\hat{c}_{\tau_k}^{(m)}$, $\hat{\boldsymbol{\theta}}^{(m)}$, and $\hat{\boldsymbol{\gamma}}^{(m)}$ are obtained by minimizing equation (7) with the first m principal components.

5 Simulation study

In this section, we conduct a Monte Carlo simulation example to assess the finite sample performance of the proposed procedures. Moreover, we include four competitors in our comparisons: (1) the LS method [14], (2) the QR method (QR) with $\tau = 0.5$ [24], (3) the rank regression (RR) method [43], (4) the composite quantile regression (CQR) method [33]. Following [30], the performance of CQR estimation does not depend sensitively on the choice of the number of quantiles q. For the sake of simplicity, we only consider $q = 9$ for CQR and WCQR method in our simulation. The data sets are generated from the following PFLRM:

$$Y_i = Z_{i1}\theta_1 + Z_{i2}\theta_2 + \int_0^1 X_i(t)\beta(t)\mathrm{d}t + \varepsilon_i,$$

where both Z_{i1} and Z_{i2} are standard normal with correlation coefficient 0.5, $\theta_1 = 0.8$ and $\theta_2 = 2$. For the functional linear component, we take the same form as [40]. Specifically, the slope function $\beta(t) = \sqrt{2}\sin(\pi t/2) + 3\sqrt{2}\sin(3\pi t/2)$ and $X(t) = \sum_{j=1}^{100}\xi_j v_j(t)$, where the ξ_j are independently distributed as the normal with mean 0 and variances given by $\lambda_j = 10((j - 0.5)\pi)^{-2}$, $v_j(t) = \sqrt{2}\sin((j - 0.5)\pi t)$.

In order to show the robustness of our estimators, the following six different error distributions are considered: standard normal distribution $N(0, 1)$, $t(3)$ distribution that is used to produce heavy-tailed distribution, F distribution $F(4, 6)$, exponential distribution $\mathrm{Exp}(1)$, Log normal distribution $\mathrm{LN}(0, 1)$, and Weibull distribution $\mathrm{Wb}(1, 1.5)$ are used to produce asymmetric error distributions.

To assess the performance of the estimated parameters, the following criteria are considered:

(1) the average value of the corresponding coefficients (MEAN);

(2) the sample standard deviation (SD);

(3) the median of absolute bias (ABISE). To assess the performance of the slope function, we consider the following square root of average square errors (RASE):

$$
\text{RASE} = \left\{ \frac{1}{n_{\text{grid}}} \sum_{k=1}^{n_{\text{grid}}} (\hat{\beta}(t_k) - \beta(t_k))^2 \right\}^{1/2},
$$

where $\{t_k, k = 1, \ldots, n_{\text{grid}}\}$ are the grid points at which the slope function $\beta(\cdot)$ is evaluated. In our simulation experiments, we set $n_{\text{grid}} = 200$, 500 repetitions are carried out with sample size $n = 100, 200, 400$. In addition, the range of the selected m is from 1 to 8. The simulation results are presented in Tables 1–6.

Table 1: Simulation results for $\boldsymbol{\theta}$ with N(0, 1) random error

n	Method	θ_1(MEAN)	θ_1(SD)	θ_1(ABISE)	θ_2(MEAN)	θ_2(SD)	θ_2(ABISE)	$\beta(t)$(RASE)
100	LS	0.8010	0.1003	0.0778	1.9969	0.0989	0.0784	0.4293
	QR	0.5162	0.4613	0.5243	0.4427	0.5346	0.5016	0.5162
	RR	0.4331	0.4403	0.3859	0.3696	0.3931	0.4097	0.4331
	CQR	0.4371	0.4408	0.3882	0.3717	0.3942	0.4072	0.4371
	WCQR	0.4385	0.4445	0.3654	0.3492	0.3625	0.3721	0.4385
200	LS	0.2894	0.3419	0.3935	0.3033	0.3840	0.3712	0.2894
	QR	0.3405	0.3302	0.3791	0.2966	0.3807	0.3419	0.3405
	RR	0.2930	0.3141	0.2762	0.2504	0.2774	0.2909	0.2903
	CQR	0.2940	0.3138	0.2764	0.2500	0.2778	0.2908	0.2940
	WCQR	0.2941	0.3153	0.2587	0.2350	0.2602	0.2645	0.2941
400	LS	0.2007	0.2316	0.2794	0.2133	0.2828	0.2483	0.2007
	QR	0.2273	0.2275	0.2737	0.2078	0.2701	0.2336	0.2273
	RR	0.2025	0.2160	0.1987	0.1776	0.1899	0.1957	0.2025
	CQR	0.2037	0.2162	0.1990	0.1783	0.1910	0.1959	0.2037
	WCQR	0.2045	0.2160	0.1877	0.1676	0.1788	0.1800	0.2045

Table 2: Simulation results for $\boldsymbol{\theta}$ with $t(3)$ random error

n	Method	θ_1(MEAN)	θ_1(SD)	θ_1(ABISE)	θ_2(MEAN)	θ_2(SD)	θ_2(ABISE)	$\beta(t)$(RASE)
100	LS	0.7997	0.1735	0.1323	2.0071	0.1671	0.1311	0.4742
	QR	0.7993	0.1409	0.1119	2.0033	0.1457	0.1147	0.4613
	RR	0.7959	0.1256	0.1007	1.9999	0.1310	0.1017	0.4403
	CQR	0.7961	0.1268	0.1015	1.9994	0.1319	0.1022	0.4408
	WCQR	0.7957	0.1287	0.1015	1.9986	0.1352	0.1052	0.4445
200	LS	0.8124	0.1229	0.0969	1.9922	0.1142	0.0911	0.3419
	QR	0.8090	0.1019	0.0813	2.0029	0.0951	0.0757	0.3302
	RR	0.8102	0.0941	0.0749	1.9984	0.0862	0.0689	0.3141
	CQR	0.8096	0.0935	0.0743	1.9988	0.0860	0.0689	0.3138
	WCQR	0.8096	0.0937	0.0741	2.0004	0.0888	0.0713	0.3153
400	LS	0.7984	0.0831	0.0660	1.9967	0.0836	0.0665	0.2316
	QR	0.8029	0.0696	0.0561	2.0016	0.0696	0.0555	0.2275
	RR	0.8009	0.0623	0.0504	2.0006	0.0642	0.0505	0.2160
	CQR	0.8007	0.0623	0.0502	2.0002	0.0637	0.0501	0.2162
	WCQR	0.8012	0.0623	0.0498	2.0006	0.0634	0.0495	0.2160

Table 3: Simulation results for $\boldsymbol{\theta}$ with F(4, 6) random error

n	Method	θ_1(MEAN)	θ_1(SD)	θ_1(ABISE)	θ_2(MEAN)	θ_2(SD)	θ_2(ABISE)	$\beta(t)$(RASE)
100	LS	0.7799	0.2804	0.2112	1.9781	0.2627	0.1940	0.5477
	QR	0.7965	0.2059	0.1661	1.9935	0.1992	0.1591	0.5243
	RR	0.7970	0.0870	0.0684	1.9940	0.0921	0.0731	0.3859
	CQR	0.7956	0.0887	0.0693	1.9931	0.0929	0.0738	0.3882
	WCQR	0.7990	0.0552	0.0431	1.9979	0.0584	0.0459	0.3654
200	LS	0.7998	0.1780	0.1379	1.9956	0.1754	0.1346	0.3935
	QR	0.8030	0.1429	0.1161	1.9984	0.1391	0.1144	0.3791
	RR	0.7998	0.0570	0.0446	1.9966	0.0564	0.0452	0.2762
	CQR	0.7999	0.0567	0.0448	1.9972	0.0566	0.0455	0.2764
	WCQR	0.8007	0.0330	0.0254	1.9985	0.0343	0.0273	0.2587
400	LS	0.8054	0.1368	0.1065	2.0018	0.1212	0.0966	0.2794
	QR	0.8046	0.1078	0.0872	2.0088	0.1154	0.0948	0.2737
	RR	0.7999	0.0426	0.0332	2.0009	0.0371	0.0301	0.1987
	CQR	0.8004	0.0430	0.0337	2.0010	0.0379	0.0305	0.1990
	WCQR	0.8003	0.0213	0.0168	2.0012	0.0211	0.0166	0.1877

Table 4: Simulation results for $\boldsymbol{\theta}$ with Exp(1) random error

n	Method	θ_1(MEAN)	θ_1(SD)	θ_1(ABISE)	θ_2(MEAN)	θ_2(SD)	θ_2(ABISE)	$\beta(t)$(RASE)
100	LS	0.8006	0.1473	0.1171	2.0058	0.1478	0.1165	0.4633
	QR	0.8001	0.1255	0.0956	2.0021	0.1279	0.0993	0.4427
	RR	0.7976	0.0721	0.0565	2.0019	0.0742	0.0578	0.3696
	CQR	0.7975	0.0717	0.0558	2.0017	0.0731	0.0567	0.3717
	WCQR	0.7970	0.0484	0.0371	2.0025	0.0478	0.0370	0.3492
200	LS	0.8014	0.1012	0.0793	1.9995	0.1045	0.0819	0.3033
	QR	0.8026	0.0780	0.0615	1.9981	0.0831	0.0651	0.2966
	RR	0.7983	0.0486	0.0388	2.0022	0.0461	0.0358	0.2504
	CQR	0.7984	0.0481	0.0384	2.0020	0.0460	0.0362	0.2500
	WCQR	0.7997	0.0270	0.0214	2.0012	0.0260	0.0201	0.2350
400	LS	0.7996	0.0730	0.0581	1.9989	0.0671	0.0536	0.2133
	QR	0.8010	0.0565	0.0445	2.0002	0.0512	0.0416	0.2078
	RR	0.8000	0.0308	0.0246	1.9991	0.0336	0.0267	0.1776
	CQR	0.7999	0.0307	0.0246	1.9992	0.0329	0.0264	0.1783
	WCQR	0.8005	0.0157	0.0123	2.0001	0.0157	0.0123	0.1676

Table 5: Simulation results for $\boldsymbol{\theta}$ with LN(0, 1) random error

n	Method	θ_1(MEAN)	θ_1(SD)	θ_1(ABISE)	θ_2(MEAN)	θ_2(SD)	θ_2(ABISE)	$\beta(t)$(RASE)
100	LS	0.8032	0.2684	0.2074	1.9916	0.2784	0.2112	0.5724
	QR	0.8091	0.2067	0.1646	1.9828	0.2042	0.1671	0.5346
	RR	0.8015	0.0937	0.0725	1.9982	0.0906	0.0711	0.3931
	CQR	0.8017	0.0943	0.0736	1.9961	0.0910	0.0718	0.3942
	WCQR	0.8009	0.0597	0.0456	1.9991	0.0574	0.0445	0.3625
200	LS	0.7921	0.1878	0.1469	1.9992	0.1898	0.1473	0.3840
	QR	0.7934	0.1530	0.1265	1.9964	0.1561	0.1271	0.3807
	RR	0.7953	0.0581	0.0460	1.9998	0.0606	0.0481	0.2774
	CQR	0.7949	0.0589	0.0471	1.9998	0.0608	0.0484	0.2778
	WCQR	0.7989	0.0327	0.0261	2.0005	0.0332	0.0259	0.2602
400	LS	0.8036	0.1418	0.1135	1.9964	0.1380	0.1076	0.2828
	QR	0.8014	0.1186	0.0971	1.9953	0.1158	0.0928	0.2701
	RR	0.7992	0.0415	0.0328	1.9975	0.0422	0.0337	0.1899
	CQR	0.7994	0.0425	0.0337	1.9971	0.0430	0.0345	0.1910
	WCQR	0.7997	0.0222	0.0175	2.0003	0.0207	0.0162	0.1788

Table 6: Simulation results for $\boldsymbol{\theta}$ with Wb(1, 1.5) random error

n	Method	θ_1(MEAN)	θ_1(SD)	θ_1(ABISE)	θ_2(MEAN)	θ_2(SD)	θ_2(ABISE)	$\beta(t)$(RASE)
100	LS	0.7913	0.2093	0.1658	2.0096	0.2212	0.1768	0.5117
	QR	0.7972	0.1829	0.1423	2.0069	0.1928	0.1527	0.5016
	RR	0.7916	0.1024	0.0806	2.0000	0.1071	0.0843	0.4097
	CQR	0.7924	0.1011	0.0799	2.0014	0.1039	0.0821	0.4072
	WCQR	0.7958	0.0642	0.0500	2.0030	0.0667	0.0522	0.3721
200	LS	0.7888	0.1497	0.1199	1.9919	0.1450	0.1155	0.3712
	QR	0.7881	0.1246	0.0984	1.9944	0.1175	0.0912	0.3419
	RR	0.7991	0.0660	0.0515	1.9997	0.0718	0.0557	0.2909
	CQR	0.7983	0.0650	0.0508	1.9994	0.0706	0.0550	0.2908
	WCQR	0.7999	0.0381	0.0292	1.9992	0.0402	0.0306	0.2645
400	LS	0.7932	0.1092	0.0878	1.9959	0.1113	0.0869	0.2483
	QR	0.7988	0.0852	0.0661	1.9961	0.0829	0.0637	0.2336
	RR	0.7986	0.0473	0.0375	2.0016	0.0502	0.0403	0.1957
	CQR	0.7984	0.0471	0.0374	2.0012	0.0502	0.0402	0.1959
	WCQR	0.8007	0.0226	0.0177	1.9993	0.0233	0.0181	0.1800

From Tables 1–6, we have the following findings: (1) As the sample size n increases, the ABISE, SD, and RASE values of all estimators decrease. As expected, increasing the sample size leads to a better performance of all estimators. (2) The standard errors become smaller and the estimators of MEANs for the parametric components are very close to the true values $\boldsymbol{\theta}$ as n increases, due to the fact of root-n consistency of the parameter estimators. This result confirms the asymptotic property that the parameter estimators are asymptotically unbiased as given in Theorem 3.1. (3) When the error follows N(0, 1), LS is the best one among the five estimators and WCQR performs nearly as well as CQR and RR. Although the QR method with quantile being 0.5 have similar performances, the WCQR seems to be slightly better than QR in most scenarios. (4) When the error distribution is symmetrical, CQR, RR, and WCQR perform similarly and satisfactorily. For all asymmetric errors: F(4, 6), Exp(1), LN(0, 1), and Wb(1, 1.5), WCQR performs better than others.

6 Application to Tecator data

We illustrate the proposed approach by analyzing a real estate data set which is available at http://lib.stat.cmu.edu/datasets/tecator. The data have been widely used to predict fat content on samples of finely chopped meat [44–46]. For each food sample, the functional data consist of a 100-channel spectrum of absorbances recorded on a Tecator Infratec Food and Feed Analysis working in the wavelength range 850–1,050 nm by the near infrared transmission principle. More details on the data can be found in Ferraty and Vieu [47]. In this section, we construct the following PFLRM model:

$$Y_i = \alpha_1 Z_i + \alpha_2 U_i + \int_{850}^{1050} X_i(t)\beta(t)\mathrm{d}t + e_i,$$

We denote the scale of fat content as Y_i, the protein content as Z_i, the moisture content as U_i, and the absorbances as $X_i(t)$.

To evaluate the performance of the method, the sample is randomly divided into two subsamples: the training sample, $I_1 = \{(X_i, Y_i, Z_i, U_i), |I_1| = 165\}$, where $|I_1|$ denotes the cardinality of I_1 and the remaining is test sample, $I_2 = \{(X_i, Y_i, Z_i, U_i), |I_2| = 50\}$. To verify our robust estimation procedures, we reanalyzed this data set by including some outliers in the response variable. The case is considered through controlling the number of outliers and the values of outliers, we randomly generate κ outliers by increasing the value Y

to $Y + t(\kappa = 20, t = 5)$. The training samples are used to estimate the parameters and the test samples are employed to verify the quality of predictions. For this, we compute the mean square error of prediction MSEP [44], which is defined by

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

where $\hat{Y}_i$ is the predicted value based on the training sample and Var_{I_2} is the variance of response variables from test sample. We use different estimation methods to predict the fat content of a meat sample based on its protein and/or moisture contents and/or its NIT absorbance spectrum, and the corresponding average MSEPs based on 200 times randomly splits are shown in Table 7. Compared with other methods, the MSEP of our proposed method (WCQR) is relatively smaller. Meanwhile, based on random simulations, the average estimated weight is $\boldsymbol{\omega}_{\text{opt}} = (0.4287, 0.07228, 0.0119, 0.0114, 0.0624, 0.0631, 0.0244, 0.0022, 0)$ and the corresponding standard deviation is $(0.1608, 0.0895, 0.0481, 0.0257, 0.0275, 0.0403, 0.0690, 0.1261, 0)$. From Figure 1, we can see that the estimated slope functions $\hat{\beta(t)}$ by our WCQR method and CQR method, respectively, and it demonstrates that the two estimation approaches perform similarly.

Table 7: MSEPs for different methods

Method	MSEP
LS	0.5684
QR	0.0522
CQR	0.0503
WCQR	0.0495
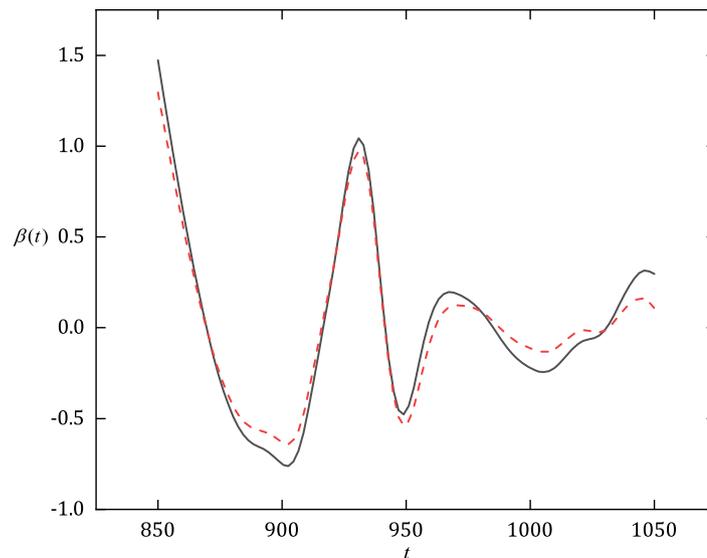
Figure 1: Curves of slope function $\hat{\beta(t)}$ with CQR (solid line) and WCQR (dotted line).

7 Conclusion

In this paper, a novel and robust procedure based on WCQR and principal component basis function approximations is developed for PFLRMs. Theoretical properties of the estimators of both slope function and linear parameters are derived under some mild assumptions. We show that the estimators do not require any specification for error distribution, and thus more robust than those based on the LS, QR and CQR in the case of asymmetric errors. Furthermore, we present an efficient algorithm for the selection of optimal weights and discuss the selection of tuning parameters. Numerical studies and the real data analysis illustrate that our proposed method performs very well for moderate sample size.

In addition, we could extend the proposed procedure to the functional regression models with diverse dimension of covariates. These deserve to be studied further.

Conflict of interest: Authors state no conflict of interest.

References

[1] H. Cardot, F. Ferraty, and P. Sarda, *Functional linear model*, Statist. Probab. Lett. **45** (1999), no. 1, 11–22, DOI: https://doi.org/10.1016/S0167-7152(99)00036-X.

[2] H. Cardot, F. Ferraty, and P. Sarda, *Spline estimators for the functional linear model*, Statist. Sinica **13** (2003), no. 3, 571–591, http://www.jstor.org/stable/24307112.

[3] F. Yao, H. G. Müller, and J. L. Wang, *Functional linear regression analysis for longitudinal data*, Ann. Statist. **33** (2005), no. 6, 2873–2903, DOI: https://doi.org/10.1214/009053605000000660.

[4] J. Ramsay and B. Silverman, *Functional Data Analysis*, 2nd edn., Springer, New York, 2005.

[5] T. Cai and P. Hall, *Prediction in functional linear regression*, Ann. Statist. **34** (2006), no. 5, 2159–2179, DOI: https://doi.org/10.1214/009053606000000830.

[6] H. Liu and H. Yang, *Estimation and variable selection in single-index composite quantile regression*, Comm. Statist. Simulation Comput. **46** (2016), no. 9, 7022–7039.

[7] P. Hall and J. Horowitz, *Methodology and convergence rates for functional linear regression*, Ann. Statist. **35** (2007), no. 1, 70–91, DOI: https://doi.org/10.1214/009053606000000957.

[8] C. Crambes, A. Kneip, and P. Sarda, *Smoothing splines estimators for functional linear regression*, Ann. Statist. **37** (2009), no. 1, 35–72, DOI: https://doi.org/10.1214/07-AOS563.

[9] P. Hall and Y. Yang, *Ordering and selecting components in multivariate or functional data linear prediction*, J. R. Stat. Soc. Ser. B. Stat. Methodol. **72** (2010), no. 1, 93–110.

[10] T. Cai and M. Yuan, *Minimax and adaptive prediction for functional linear regression*, J. Amer. Statist. Assoc. **107** (2012), no. 499, 1201–1216, https://www.jstor.org/stable/23427425.

[11] A. Delaigle and P. Hall, *Methodology and theory for partial least squares applied to functional data*, Ann. Statist. **40** (2012), no. 1, 322–352, DOI: https://doi.org/10.1214/11-AOS958.

[12] E. R. Lee and B. U. Park, *Sparse estimation in functional linear regression*, J. Multivariate Anal. **105** (2012), no. 1, 1–17, DOI: https://doi.org/10.1016/j.jmva.2011.08.005.

[13] L. Huang, H. Wang, and A. Zheng, *The M-estimator for functional linear regression model*, Statist. Probab. Lett. **88** (2014), no. C, 165–173, DOI: https://doi.org/10.1016/j.spl.2014.01.016.

[14] H. J. Shin, *Partial functional linear regression*, J. Statist. Plann. Inference **139** (2009), no. 10, 3405–3418, DOI: https://doi.org/10.1016/j.jspi.2009.03.001.

[15] J. Zhou, Z. Chen, and Q. Peng, *Polynomial spline estimation for partial functional linear regression models*, Comput. Statist. **31** (2016), no. 3, 1107–1129, DOI: https://doi.org/10.1007/s00180-015-0636-0.

[16] D. Kong, K. Xue, F. Yao, and H. Zhang, *Partially functional linear regression in high dimensions*, Biometrika **103** (2016), no. 1, 147–159, DOI: https://doi.org/10.1093/biomet/asv062.

[17] N. Ling, G. Aneiros, and P. Vieu, *kNN estimation in functional partial linear modeling*, Statist. Papers **61** (2020), no. 1, 423–444, DOI: https://doi.org/10.1007/s00362-017-0946-0.

[18] D. Kong, J. Ibrahim, E. Lee, and H. Zhu, *FLCRM: functional linear cox regression model*, Biometrics **74** (2018), no. 1, 109–117.

[19] R. Cao, J. Du, J. Zhou, and T. Xie, *FPCA-based estimation for generalized functional partially linear models*, Statist. Papers **61** (2020), no. 6, 2715–2735, DOI: https://doi.org/10.1007/s00362-018-01066-8.

[20] M. Yuan and Y. Zhang, *Test for the parametric part in partial functional linear regression based on B-spline*, Comm. Statist. Simulation Comput. **50** (2021), no. 1, 1–15, DOI: https://doi.org/10.1080/03610918.2019.1667391.

[21] H. Zhu, R. Zhang, Z. Yu, H. Lian, and Y. Liu, *Estimation and testing for partially functional linear errors-in-variables models*, J. Multivariate Anal. **170** (2019), 296–314, DOI: https://doi.org/10.1016/j.jmva.2018.11.005.

[22] Q. Li, X. Tan, and L. Wang, *Testing for error correlation in partially functional linear regression models*, Comm. Statist. Theory Methods **50** (2021), no. 3, 747–761, DOI: https://doi.org/10.1080/03610926.2019.1642492.

[23] R. Koenker and G. Basset, *Regression quantiles*, Econometrica **46** (1978), no. 1, 33–50, DOI: https://doi.org/10.2307/1913643.

[24] Y. Lu, J. Du, Z. Sun, *Functional partially linear quantile regression model*, Metrika **77** (2014), no. 2, 317–332, DOI: https://doi.org/10.1007/s00184-013-0439-7.

[25] H. Ma, T. Li, H. Zhu, and Z. Zhu, *Quantile regression for functional partially linear model in ultra-high dimensions*, Comput. Statist. Data Anal. **129** (2019), no. 1, 135–147, DOI: https://doi.org/10.1016/j.csda.2018.06.005.

[26] R. Koenker, *Quantile Regression*, Cambridge University Press, Cambridge, 2005.

[27] H. Zou and M. Yuan, *Composite quantile regression and the Oracle model selection theory*, Ann. Statist. **36** (2008), no. 3, 1108–1126, https://www.jstor.org/stable/25464661.

[28] B. Kai, R. Li, and H. Zou, *Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression*, J. R. Stat. Soc. Ser. B Stat. Methodol. **72** (2010), no. 1, 49–69, DOI: https://doi.org/10.1111/j.1467-9868.2009.00725.x.

[29] B. Kai, R. Li, and H. Zou, *New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models*, Ann. Statist. **39** (2011), no. 1, 305–332, DOI: https://doi.org/10.1214/10-AOS842.

[30] J. Guo, M. Tang, M. Tian, and K. Zhu, *Variable selection in high-dimensional partially linear additive models for composite quantile regression*, Comput. Statist. Data Anal. **65** (2013), 56–67, DOI: https://doi.org/10.1016/j.csda.2013.03.017.

[31] R. Jiang, Z. Zhou, W. Qian, and Y. Chen, *Two step composite quantile regression for single-index models*, Comput. Statist. Data Anal. **64** (2013), 180–191, DOI: https://doi.org/10.1016/j.csda.2013.03.014.

[32] R. Jiang, W. Qian, and Z. Zhou, *Test for single-index composite quantile regression*, Hacet. J. Math. Stat. **43** (2014), no. 5, 861–871.

[33] J. Du, D. Xu, and R. Cao, *Estimation and variable selection for partially functional linear models*, J. Korean Statist. Soc. **47** (2018), no. 4, 436–449, DOI: https://doi.org/10.1016/j.jkss.2018.05.002.

[34] P. Yu, T. Li, Z. Zhu, and Z. Zhang, *Composite quantile estimation in partial functional linear regression model with dependent errors*, Metrika **82** (2019), no. 6, 633–656, DOI: https://doi.org/10.1007/s00184-018-0699-3.

[35] X. Jiang, J. Jiang, and X. Song, *Oracle model selection for nonlinear models based on weighted composite quantile regression*, Statist. Sinica **22** (2012), no. 4, 1479–1506, DOI: https://doi.org/10.5705/ss.2010.203.

[36] C. Guo, H. Yang, and J. Lv, *Robust variable selection in high-dimensional varying coefficient models based on weighted composite quantile regression*, Statist. Papers **58** (2017), no. 4, 1009–1033, DOI: https://doi.org/10.1007/s00362-015-0736-5.

[37] R. Jiang, W. Qian, and Z. Zhou, *Weighted composite quantile regression for partially linear varying coefficient models*, Comm. Statist. Theory Methods **47** (2018), no. 16, 3987–4005, DOI: https://doi.org/10.1080/03610926.2017.1366522.

[38] J. Sun, Y. Gai, and L. Lin, *Weighted local linear composite quantile estimation for the case of general error distributions*, J. Statist. Plann. Inference **143** (2013), no. 6, 1049–1063, DOI: https://doi.org/10.1016/j.jspi.2013.01.002.

[39] R. Jiang, W. Qian, and Z. Zhou, *Weighted composite quantile regression for single-index models*, J. Multivariate Anal. **148** (2016), 34–48, DOI: https://doi.org/10.1016/j.jmva.2016.02.015.

[40] P. Yu, Z. Zhang, and J. Du, *A test of linearity in partial functional linear regression*, Metrika **79** (2016), no. 8, 953–969, DOI: https://doi.org/10.1007/s00184-016-0584-x.

[41] L. Horváth and P. Kokoszka, *Inference for Functional Data with Applications*, Springer, New York, 2012.

[42] B. Silverman, *Density Estimation*, Chapman and Hall, London, 1986.

[43] J. Sun and L. Lin, *Local rank estimation and related test for varying-coefficient partially linear models*, J. Nonparametr. Stat. **26** (2014), no. 1, 187–206, DOI: https://doi.org/10.1080/10485252.2013.841910.

[44] G. Aneiros-Pérez and P. Vieu, *Semi-functional partial linear regression*, Statist. Probab. Lett. **76** (2006), no. 11, 1102–1110, DOI: https://doi.org/10.1016/j.spl.2005.12.007.

[45] G. Aneiros-Pérez and P. Vieu, *Partial linear modelling with multi-functional covariates*, Comput. Statist. **30** (2015), no. 3, 647–671, DOI: https://doi.org/10.1007/s00180-015-0568-8.

[46] P. Yu, J. Du, and Z. Zhan, *Single-index partially functional linear regression model*, Statist. Papers **61** (2020), no. 3, 1107–1123, DOI: https://doi.org/10.1007/s00362-018-0980-6.

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

[48] C. Crambes, A. Kneip, and P. Sarda, *Smoothing splines estimators for functional linear regression*, Ann. Statist. **37** (2009), no. 1, 35–72, DOI: https://doi.org/10.1214/07-AOS563.

[49] K. Knight, *Limiting distributions for L_1 regression estimators under general conditions*, Ann. Statist. **26** (1998), no. 2, 755–770, https://www.jstor.org/stable/120050.

Appendix

Proof of Theorem 3.1(i)

Let $\delta_n = n^{-\frac{2b-1}{2(a+2b)}}$, $\boldsymbol{u}_n = \delta_n^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, $\boldsymbol{s}_n = \delta_n^{-1}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$, $v_{nk} = \delta_n^{-1}(\hat{c}_{\tau_k} - c_{\tau_k})$, $\boldsymbol{v}_n = (v_{n1}, \ldots, v_{nq})^{\mathrm{T}}$, $R_{ni} = \int_0^1 X_i(t)\beta_0(t)\mathrm{d}t - \boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{\gamma}_0$, $\boldsymbol{\Gamma}_n = \boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}}$, $\boldsymbol{\Lambda}_n = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_m)$, $\mathcal{N} = (X_1(t), \ldots, X_n(t), Z_1^{(1)}, \ldots, Z_n^{(1)}, \ldots, Z_1^{(p)}, \ldots, Z_n^{(p)})^{\mathrm{T}}$, and $\mathcal{F}_n = \{(\boldsymbol{u}_n, \boldsymbol{s}_n, \boldsymbol{v}_n): \|(\boldsymbol{u}_n^{\mathrm{T}}, \boldsymbol{s}_n^{\mathrm{T}}, \boldsymbol{v}_n^{\mathrm{T}})^{\mathrm{T}}\| = C\}$, where C is a large enough constant, our aim is to show that for any given $\varrho > 0$, there is a large constant C such that, for large n, we have

$$P\left\{\inf_{(\boldsymbol{u}_n, \boldsymbol{s}_n, \boldsymbol{v}_n) \in \mathcal{F}_n} \mathcal{L}_n(\boldsymbol{\theta}_0 + \delta_n\boldsymbol{u}_n, \boldsymbol{\gamma}_0 + \delta_n\boldsymbol{s}_n, \boldsymbol{c} + \delta_n\boldsymbol{v}_n, \boldsymbol{\omega}) > \mathcal{L}_n(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{c}, \boldsymbol{\omega})\right\} \geq 1 - \varrho. \tag{A.1}$$

This implies that, with probability tending to one, there is local minimum $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\gamma}}$ in the ball $\{(\boldsymbol{\theta}_0 + \delta_n\boldsymbol{u}_n, \boldsymbol{\gamma}_0 + \delta_n\boldsymbol{s}_n, \boldsymbol{c} + \delta_n\boldsymbol{v}_n) : \|(\boldsymbol{u}_n^{\mathrm{T}}, \boldsymbol{s}_n^{\mathrm{T}}, \boldsymbol{v}_n^{\mathrm{T}})^{\mathrm{T}}\| \leq C\}$ such that $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(\delta_n)$ and $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| = O_p(\delta_n)$.

First, by $\|\phi_j - \hat{\phi}_j\|^2 = O_p(n^{-1}j^2)$ (see Crambes et al. [48]), one has

$$|R_{ni}|^2 = \left|\int_0^1 X_i(t)\beta_0(t)\mathrm{d}t - \boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{\gamma}_0\right|^2$$

$$\leq 2\left|\sum_{j=1}^m \langle X_i, \hat{\phi}_j - \phi_j\rangle\gamma_{0j}\right|^2 + 2\left|\sum_{j=m+1}^\infty \langle X_i, \phi_j\rangle\gamma_{0j}\right|^2$$

$$\triangleq 2A_1 + 2A_2.$$

For A_1, by Assumptions A1–A2 and the Hölder inequality, it is obtained

$$A_1 = \left|\sum_{j=1}^m \langle X_i, \phi_j - \hat{\phi}_j\rangle\gamma_{0j}\right|^2 \leq cm\sum_{j=1}^m \|\phi_j - \hat{\phi}_j\|^2|\gamma_{0j}|^2 \leq cm\sum_{j=1}^m O_p(n^{-1}j^{2-2b}) = O_p(n^{-\frac{a+4b-4}{a+2b}}) = o_p(\delta_n^2).$$

As for A_2, due to

$$E\left\{\sum_{j=m+1}^\infty \langle X_i, \phi_j\rangle\gamma_{0j}\right\} = 0,$$

$$\mathrm{Var}\left\{\sum_{j=m+1}^\infty \langle X_i, \phi_j\rangle\gamma_{0j}\right\} = \sum_{j=m+1}^\infty \lambda_j\gamma_{0j}^2 \leq c\sum_{j=m+1}^\infty j^{-(a+2b)} = O\left(n^{-\frac{a+2b-1}{a+2b}}\right),$$

one has $A_2 = O_p\left(n^{-\frac{a+2b-1}{a+2b}}\right) = o_p(\delta_n^2)$. Taking these together, then $|R_{ni}|^2 = O_p\left(n^{-\frac{a+2b-1}{a+2b}}\right) = o_p(\delta_n^2)$.

$$S_n = \mathcal{L}_n(\boldsymbol{\theta}_0 + \delta_n\boldsymbol{u}_n, \boldsymbol{\gamma}_0 + \delta_n\boldsymbol{s}_n, \boldsymbol{c} + \delta_n\boldsymbol{v}_n, \boldsymbol{\omega}) - \mathcal{L}_n(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{c}, \boldsymbol{\omega}),$$

By the study of Knight [49],

$$|z - y| - |z| = -y\mathrm{sgn}(z) + 2(y - z)\{I(0 < z < y) - I(y < z < 0)\},$$

then

$$\rho_\tau(r - s) - \rho_\tau(r) = s(I(r < 0) - \tau) + \int_0^s [I(r \leq t) - I(r \leq 0)]\mathrm{d}t,$$

we can rewrite S_n as

$$
S_n = \sum_{k=1}^{q} \omega_k \sum_{i=1}^{n} \delta_n (\boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{u}_n + \boldsymbol{U}_i^{\mathrm{T}} \boldsymbol{s}_n + v_{nk})[I(\varepsilon_i < R_{ni} + c_{\tau_k}) - \tau_k]
$$

$$
+ \sum_{k=1}^{q} \omega_k \sum_{i=1}^{n} \int_{0}^{\delta_n(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{u}_n + \boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{s}_n + v_{nk})} [I(\varepsilon_i \le x + R_{ni} + c_{\tau_k}) - I(\varepsilon_i \le R_{ni} + c_{\tau_k})]\mathrm{d}x \tag{A.2}
$$

$$
= \sqrt{n}\,\delta_n(\boldsymbol{A}_n^{\mathrm{T}}\boldsymbol{u}_n + \boldsymbol{C}_n^{\mathrm{T}}\boldsymbol{s}_n + \boldsymbol{D}_n^{\mathrm{T}}\boldsymbol{v}_n) + \sum_{k=1}^{q} \omega_k \boldsymbol{B}_n^{(k)},
$$

where

$$
\boldsymbol{B}_n^{(k)} = \sum_{i=1}^{n} \int_{0}^{\delta_n(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{u}_n + \boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{s}_n + v_{nk})} [I(\varepsilon_i \le x + R_{ni} + c_{\tau_k}) - I(\varepsilon_i \le R_{ni} + c_{\tau_k})]\mathrm{d}x,
$$

$$
\boldsymbol{A}_n = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{Z}_i \sum_{k=1}^{q} \omega_k[I(\varepsilon_i < R_{ni} + c_{\tau_k}) - \tau_k],
$$

$$
\boldsymbol{C}_n = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{U}_i \sum_{k=1}^{q} \omega_k[I(\varepsilon_i < R_{ni} + c_{\tau_k}) - \tau_k],
$$

$$
D_{n,k} = n^{-1/2} \sum_{i=1}^{n} \omega_k[I(\varepsilon_i < R_{ni} + c_{\tau_k}) - \tau_k], \quad \boldsymbol{D}_n = (D_{n,1}, \ldots, D_{n,q})^{\mathrm{T}}.
$$

Note that, by Assumptions A2 and A4, respectively, we have $\|\boldsymbol{\Gamma}_n\| = O(1)$, $\|\boldsymbol{\Lambda}_n\| = O(1)$, since ε_i is independent of $\boldsymbol{Z}_i$ and $X_i(t)$, it have that $E(\boldsymbol{A}_n^{\mathrm{T}}\boldsymbol{u}_n) = 0$, $E\{(\boldsymbol{A}_n^{\mathrm{T}}\boldsymbol{u}_n)^2\} = \boldsymbol{u}_n^{\mathrm{T}} E(\boldsymbol{A}_n \boldsymbol{A}_n^{\mathrm{T}})\boldsymbol{u}_n = O(\|\boldsymbol{u}_n\|^2)$. Then we have, $\boldsymbol{A}_n^{\mathrm{T}}\boldsymbol{u}_n = O(\|\boldsymbol{u}_n\|)$. Similarly, we get $\boldsymbol{C}_n^{\mathrm{T}}\boldsymbol{s}_n = O(\|\boldsymbol{s}_n\|)$. This combined with (A.2) leads to

$$
S_n = \sum_{k=1}^{q} \omega_k \boldsymbol{B}_n^{(k)} + o_p(n\delta_n^2)\|\boldsymbol{u}_n\| + o_p(n\delta_n^2)\|\boldsymbol{s}_n\|. \tag{A.3}
$$

Invoking Assumption A7, a simple calculation yields

$$
E(\boldsymbol{B}_n^{(k)}|\mathcal{N}) = E\left[\sum_{i=1}^{n} \int_{0}^{\delta_n(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{u}_n + \boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{s}_n + v_{nk})} [I(\varepsilon_i \le x + R_{ni} + c_{\tau_k}) - I(\varepsilon_i \le R_{ni} + c_{\tau_k})]\mathrm{d}x \Big| \mathcal{N}\right]
$$

$$
= \sum_{i=1}^{n} \int_{0}^{\delta_n(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{u}_n + \boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{s}_n + v_{nk})} [F(x + R_{ni} + c_{\tau_k}) - F(R_{ni} + c_{\tau_k})]\mathrm{d}x
$$

$$
= \sum_{i=1}^{n} \int_{0}^{\delta_n(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{u}_n + \boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{s}_n + v_{nk})} [f(R_{ni} + c_{\tau_k})x(1 + o_p(1))]\mathrm{d}x \tag{A.4}
$$

$$
\le \frac{3f(c_{\tau_k})}{2} n\delta_n^2(v_{nk}^2 + \boldsymbol{u}_n^{\mathrm{T}}\boldsymbol{\Gamma}_n\boldsymbol{u}_n + \boldsymbol{s}_n^{\mathrm{T}}\boldsymbol{\Lambda}_n\boldsymbol{s}_n)(1 + o_p(1)).
$$

Similarly,

$$
\mathrm{Var}(\boldsymbol{B}_n^{(k)}|\mathcal{N}) = \mathrm{Var}\left[\sum_{i=1}^{n} \int_{0}^{\delta_n(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{u}_n + \boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{s}_n + v_{nk})} [I(\varepsilon_i \le x + R_{ni} + c_{\tau_k}) - I(\varepsilon_i \le R_{ni} + c_{\tau_k})]\mathrm{d}x \Big| \mathcal{N}\right]
$$

$$
\le \sum_{i=1}^{n} E\left[\left(\int_{0}^{\delta_n(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{u}_n + \boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{s}_n + v_{nk})} [I(\varepsilon_i \le x + R_{ni} + c_{\tau_k}) - I(\varepsilon_i \le R_{ni} + c_{\tau_k})]\mathrm{d}x\right)^2 \Big| \mathcal{N}\right] \tag{A.5}
$$

$$\leq \sum_{i=1}^{n} \int_{0}^{|\delta_n(\mathbf{Z}_i^{\mathrm{T}}\mathbf{u}_n + \mathbf{U}_i^{\mathrm{T}}\mathbf{s}_n + v_{nk})|} \int_{0}^{|\delta_n(\mathbf{Z}_i^{\mathrm{T}}\mathbf{u}_n + \mathbf{U}_i^{\mathrm{T}}\mathbf{s}_n + v_{nk})|}$$

$$\times [F(|\delta_n(\mathbf{Z}_i^{\mathrm{T}}\mathbf{u}_n + \mathbf{U}_i^{\mathrm{T}}\mathbf{s}_n + v_{nk})| + R_{ni} + c_{\tau_k}) - F(R_{ni} + c_{\tau_k})]\mathrm{d}x_1\mathrm{d}x_2$$

$$\leq o\left(\sum_{i=1}^{n}|\delta_n(\mathbf{Z}_i^{\mathrm{T}}\mathbf{u}_n + \mathbf{U}_i^{\mathrm{T}}\mathbf{s}_n + v_{nk})|^2\right)$$

$$= o_p(n\delta_n^2)(\|\mathbf{u}_n\|^2 + \|\mathbf{s}_n\|^2 + \|\mathbf{v}_n\|^2).$$

Hence,

$$\mathbf{B}_n^{(k)} \leq \frac{3f(c_{\tau_k})}{2}n\delta_n^2(v_{nk}^2 + \mathbf{u}_n^{\mathrm{T}}\mathbf{\Gamma}_n\mathbf{u}_n + \mathbf{s}_n^{\mathrm{T}}\mathbf{\Lambda}_n\mathbf{s}_n)(1 + o_p(1)).$$

Then, we can obtain

$$S_n \leq \frac{3}{2}n\delta_n^2\sum_{k=1}^{q}\omega_k f(c_{\tau_k})(v_{nk}^2 + \mathbf{u}_n^{\mathrm{T}}\mathbf{\Gamma}_n\mathbf{u}_n + \mathbf{s}_n^{\mathrm{T}}\mathbf{\Lambda}_n\mathbf{s}_n)(1 + o_p(1)) + o_p(n\delta_n^2)(\|\mathbf{u}_n\|^2 + \|\mathbf{s}_n\|^2 + \|\mathbf{v}_n\|^2). \quad\text{(A.6)}$$

It follows from (A.2) to (A.6), we can obtain that S_n is dominated by the positive quadratic term $n\delta_n^2\sum_{k=1}^{q}\omega_k f(c_{\tau_k})(v_{nk}^2 + \mathbf{u}_n^{\mathrm{T}}\mathbf{\Gamma}_n\mathbf{u}_n + \mathbf{s}_n^{\mathrm{T}}\mathbf{\Lambda}_n\mathbf{s}_n)$ as long as C is large enough and there exists local minimizer $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\gamma}}$ such that

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(\delta_n), \quad \|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| = O_p(\delta_n). \quad\text{(A.7)}$$

Observe that

$$\|\hat{\beta}(t) - \beta_0(t)\|^2 = \left\|\sum_{j=1}^{m}\hat{\gamma}_j\hat{\phi}_j - \sum_{j=1}^{\infty}\gamma_{0j}\phi_j\right\|^2$$

$$\leq 2\left\|\sum_{j=1}^{m}\hat{\gamma}_j\hat{\phi}_j - \sum_{j=1}^{m}\gamma_{0j}\phi_j\right\|^2 + 2\left\|\sum_{j=m+1}^{\infty}\gamma_{0j}\phi_j\right\|^2$$

$$\leq 4\left\|\sum_{j=1}^{m}(\hat{\gamma}_j - \gamma_{0j})\hat{\phi}_j\right\|^2 + 4\left\|\sum_{j=1}^{m}\gamma_{0j}(\hat{\phi}_j - \phi_j)\right\|^2 + 2\sum_{j=m+1}^{\infty}\gamma_{0j}^2$$

$$\triangleq 4J_1 + 4J_2 + 2J_3.$$

By Assumption A2, the orthogonality of $\{\hat{\phi}_j\}$ and $\|\phi_j - \hat{\phi}_j\|^2 = O_p(n^{-1}j^2)$, one has

$$J_1 = \left\|\sum_{j=1}^{m}(\hat{\gamma}_j - \gamma_{0j})\hat{\phi}_j\right\|^2 \leq \sum_{j=1}^{m}|\hat{\gamma}_j - \gamma_{0j}|^2 = \|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|^2 = O_p(\delta_n^2), \quad\text{(A.8)}$$

$$J_2 = \left\|\sum_{j=1}^{m}\gamma_{0j}(\hat{\phi}_j - \phi_j)\right\|^2 \leq m\sum_{j=1}^{m}\|\hat{\phi}_j - \phi_j\|^2\gamma_{0j}^2 \leq \frac{m}{n}O_p\left(\sum_{j=1}^{m}j^2\gamma_{0j}^2\right)$$

$$= O_p\left(n^{-1}m\sum_{j=1}^{m}j^{2-2b}\right) = O_p(n^{-1}m) = o_p(n^{-\frac{2b-1}{a+2b}}) = o_p(\delta_n^2), \quad\text{(A.9)}$$

and

$$J_3 = \sum_{j=m+1}^{\infty}\gamma_{0j}^2 \leq C\sum_{j=m+1}^{\infty}j^{-2b} = O(n^{-\frac{2b-1}{a+2b}}) = O(\delta_n^2). \quad\text{(A.10)}$$

Then, combining (A.8)–(A.10), we can complete the proof of Theorem 3.1(i).

B Proof of Theorem 3.1(ii)

According to Theorem 3.1(i), we know that, as $n \to \infty$, with probability tending to 1, $\mathcal{L}_n(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{c}, \boldsymbol{\omega})$ attains the minimal value at $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{c}}, \boldsymbol{\omega})$. Then, we have the following score equations:

$$n^{-1} \sum_{k=1}^{q} \omega_k \sum_{i=1}^{n} \boldsymbol{Z}_i^{\mathrm{T}} \psi_{\tau_k}(Y_i - \hat{c}_{\tau_k} - \boldsymbol{Z}_i^{\mathrm{T}}\hat{\boldsymbol{\theta}} - \boldsymbol{U}_i^{\mathrm{T}}\hat{\boldsymbol{\gamma}}) = 0, \tag{A.11}$$

$$n^{-1} \sum_{k=1}^{q} \omega_k \sum_{i=1}^{n} \boldsymbol{U}_i^{\mathrm{T}} \psi_{\tau_k}(Y_i - \hat{c}_{\tau_k} - \boldsymbol{Z}_i^{\mathrm{T}}\hat{\boldsymbol{\theta}} - \boldsymbol{U}_i^{\mathrm{T}}\hat{\boldsymbol{\gamma}}) = 0, \tag{A.12}$$

where $\psi_{\tau_k}(u) = \rho'_{\tau_k}(u) = \tau_k - I(u < 0)$. By equations (A.11) and (A.12), we have

$$n^{-1} \sum_{k=1}^{q} \omega_k \sum_{i=1}^{n} \boldsymbol{Z}_i^{\mathrm{T}} \psi_{\tau_k}(\varepsilon_i - c_{\tau_k} - R_{ni} - \boldsymbol{Z}_i^{\mathrm{T}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \boldsymbol{U}_i^{\mathrm{T}}(\hat{\boldsymbol{y}} - \boldsymbol{y}_0)) = 0, \tag{A.13}$$

$$n^{-1} \sum_{k=1}^{q} \omega_k \sum_{i=1}^{n} \boldsymbol{U}_i^{T} \psi_{\tau_k}(\varepsilon_i - c_{\tau_k} - R_{ni} - \boldsymbol{Z}_i^{\mathrm{T}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \boldsymbol{U}_i^{\mathrm{T}}(\hat{\boldsymbol{y}} - \boldsymbol{y}_0)) = 0. \tag{A.14}$$

Moreover, we can write equation (A.13) as

$$(\text{A.13}) = -\sum_{i=1}^{n} H_n + \sum_{k=1}^{q} \omega_k B_{n1}^{(k)} + \sum_{k=1}^{q} \omega_k B_{n2}^{(k)},$$

where

$$H_n = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{Z}_i \sum_{k=1}^{q} \omega_k [I(\varepsilon_i < R_{ni} + c_{\tau_k}) - \tau_k],$$

$$B_{n1}^{(k)} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{Z}_i [F(R_{ni} + c_{\tau_k}) - F(c_{\tau_k} + R_{ni} + \boldsymbol{Z}_i^{\mathrm{T}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{U}_i^{\mathrm{T}}(\hat{\boldsymbol{y}} - \boldsymbol{y}_0))],$$

$$B_{n2}^{(k)} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{Z}_i \{ [I(\varepsilon_i < R_{ni} + c_{\tau_k}) - I(\varepsilon_i < c_{\tau_k} + R_{ni} + \boldsymbol{Z}_i^{\mathrm{T}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{U}_i^{\mathrm{T}}(\hat{\boldsymbol{y}} - \boldsymbol{y}_0))]$$

$$- [F(R_{ni} + c_{\tau_k}) - F(c_{\tau_k} + R_{ni} + \boldsymbol{Z}_i^{\mathrm{T}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{U}_i^{\mathrm{T}}(\hat{\boldsymbol{y}} - \boldsymbol{y}_0))] \}.$$

Invoking Taylor expansion, a simple calculation yields

$$B_{n1}^{(k)} = -\frac{1}{n} \sum_{i=1}^{n} f(c_{\tau_k}) [\boldsymbol{Z}_i \boldsymbol{Z}_i^{\mathrm{T}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{Z}_i \boldsymbol{U}_i^{\mathrm{T}}(\hat{\boldsymbol{y}} - \boldsymbol{y}_0)](1 + o(1)).$$

By direct calculation of the mean and variance, we can show, as in the study by Jiang et al. (2012), that $B_{n2}^{(k)} = o_p(\delta_n)$. Then, we have

$$-\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{Z}_i \sum_{k=1}^{q} \omega_k [I(\varepsilon_i < R_{ni} + c_{\tau_k}) - \tau_k] = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{q} \omega_k f(c_{\tau_k}) [\boldsymbol{Z}_i \boldsymbol{Z}_i^{\mathrm{T}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{Z}_i \boldsymbol{U}_i^{\mathrm{T}}(\hat{\boldsymbol{y}} - \boldsymbol{y}_0)](1 + o(1)). \tag{A.15}$$

Similarly, we have

$$-\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{U}_i \sum_{k=1}^{q} \omega_k [I(\varepsilon_i < R_{ni} + c_{\tau_k}) - \tau_k]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{q} \omega_k f(c_{\tau_k}) [\boldsymbol{U}_i \boldsymbol{Z}_i^{\mathrm{T}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{U}_i \boldsymbol{U}_i^{\mathrm{T}}(\hat{\boldsymbol{y}} - \boldsymbol{y}_0)](1 + o(1)). \tag{A.16}$$

Let $\boldsymbol{\Upsilon}_n = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{U}_i [\tau_k - I(\varepsilon_i < R_{ni} + c_{\tau_k})]$, $\boldsymbol{\Phi}_n = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{U}_i \boldsymbol{U}_i^{\mathrm{T}}$, $\boldsymbol{\Psi}_n = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{U}_i \boldsymbol{Z}_i^{\mathrm{T}}$. By equation (A.16), we have

$$\hat{\boldsymbol{y}} - \boldsymbol{y}_0 = (\boldsymbol{\Phi}_n + o_p(1))^{-1} [\boldsymbol{\Upsilon}_n + \boldsymbol{\Psi}_n(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})]. \tag{A.17}$$

Substituting equation (A.17) into equation (A.15), we can obtain that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{q}\omega_k f(c_{\tau_k})\mathbf{Z}_i[\mathbf{Z}_i - \mathbf{\Psi}_n^{\mathrm{T}}\mathbf{\Phi}_n^{-1}\mathbf{U}_i]^{\mathrm{T}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{q}\omega_k \mathbf{Z}_i([\tau_k - I(\varepsilon_i < R_{ni} + c_{\tau_k})] - f(c_{\tau_k})\mathbf{U}_i^{\mathrm{T}}\mathbf{\Phi}_n^{-1}\mathbf{\Upsilon}_n). \tag{A.18}$$

Note that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{q}\omega_k f(c_{\tau_k})\mathbf{\Psi}_n^{\mathrm{T}}\mathbf{\Phi}_n^{-1}\mathbf{U}_i[\mathbf{Z}_i - \mathbf{\Psi}_n^{\mathrm{T}}\mathbf{\Phi}_n^{-1}\mathbf{U}_i]^{\mathrm{T}} = 0, \tag{A.19}$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{q}\omega_k f(c_{\tau_k})\mathbf{\Psi}_n^{\mathrm{T}}\mathbf{\Phi}_n^{-1}\mathbf{U}_i\{[\tau_k - I(\varepsilon_i < R_{ni} + c_{\tau_k})] - \mathbf{U}_i^{\mathrm{T}}\mathbf{\Phi}_n^{-1}\mathbf{\Upsilon}_n\} = 0, \tag{A.20}$$

According to equation (A.18)–(A.20), it is easy to show that

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left(\sum_{i=1}^{n}\sum_{k=1}^{q}\omega_k f(c_{\tau_k})\widetilde{\mathbf{Z}}_i\widetilde{\mathbf{Z}}_i^{\mathrm{T}}\right)^{-1}\left(\sum_{i=1}^{n}\sum_{k=1}^{q}\omega_k[I(\varepsilon_i < R_{ni} + c_{\tau_k}) - \tau_k]\widetilde{\mathbf{Z}}_i\right),$$

where $\widetilde{\mathbf{Z}}_i = \mathbf{Z}_i - \mathbf{\Psi}_n^{\mathrm{T}}\mathbf{\Phi}_n^{-1}\mathbf{U}_i$. According to Lemma 1 in the study by Yu et al. (2016) and Assumptions A5–A6, as $n \to \infty$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{\mathbf{Z}}_i\widetilde{\mathbf{Z}}_i^{\mathrm{T}} \overset{p}{\to} \mathbf{\Sigma}.$$

Note

$$\left(\sum_{i=1}^{n}\sum_{k=1}^{q}\omega_k f(c_{\tau_k})\right)^{-1}\left(\sum_{i=1}^{n}\sum_{k=1}^{q}\omega_k[I(\varepsilon_i < R_{ni} + c_{\tau_k}) - \tau_k]\widetilde{\mathbf{Z}}_i\right)$$

is normal with mean 0 and covariance matrix $\sigma^2(\boldsymbol{\omega})\mathbf{\Sigma}$, then we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to N(0, \sigma^2(\boldsymbol{\omega})\mathbf{\Sigma}^{-1}).$$

We complete the proof of Theorem 3.1(ii).