Ding, Hui, Zhang, Riquan and Zhang, Jian (2018) *Quantile Estimation for a Hybrid Model of Functional and Varying Coefficient Regressions.* Journal of Statistical Planning and Inference, 196 . pp. 1-18. ISSN 0378-3758.

Quantile Estimation for a Hybrid Model of Functional and Varying Coefficient Regressions *

Hui Ding[1,2], Riquan Zhang[1]†, Jian Zhang[3]
1. School of Statistics
East China Normal University, Shanghai, 200241, P. R. China
2. School of Mathematics and Finance
Chuzhou University, Chuzhou, 239000, P. R. China
3. School of Mathematics, Statistics and Actuarial Science
University of Kent, Canterbury, Kent CT2 7NF, UK

October 27, 2017

Abstract

We consider a hybrid of functional and varying-coefficient regression models for the analysis of mixed functional data. We propose a quantile estimation of this hybrid model as an alternative to the least square approach. Under regularity conditions, we establish the asymptotic normality of the proposed estimator. We show that the estimated slope function can attain the minimax convergence rate as in functional linear regression. A Monte Carlo simulation study and a real data application suggest that the proposed estimation is promising.

Key words and phrases: *Functional data analysis, varying coefficient, partially functional regression, convergence rate, mixed data.*

AMS 2000 Subject Classifications: *Primary* **62G05**, *Secondary* **62G20**.

Short title: *Quantile Estimation of a Hybrid Functional Regression Model*

*The research was supported in part by National Natural Science Foundation of China (11571112), Program of Shanghai Subject Chief Scientist (14XD1401600) and the 111 Project of China (B14019).

†Correspondence address: School of Statistics, East China Normal University, Shanghai, 200241, P. R. China. E-mail: zhangriquan@163.com.

1 Introduction

Over the past two decades, technological innovations in biology, chemistry, medicine, engineering, economics and finance have produced large scale data with functions or images as the units of observation. The analysis of these functional datasets has stimulated extensive research on functional regressions where the response variable or covariates are functions. See Ramsay and Silverman (2005), Morris (2015) and Wang et al. (2015) for systematic reviews on this subject. As the simplest form of functional data analysis, functional linear regression analysis has been intensively studied and applied to solve a wide range of scientific problems. See Cardot et al. (1999, 2003), Yao et al. (2005), Hall and Horowitz (2007), Cai and Hall (2006), Kato (2012) and among others. The functional linear regression aims to model the relationship between a scalar response variable and a functional covariate. But in practice, we often see that a scalar response is related not only to functional covariate, but also to scalar covariates. For example, as we discuss in section 5, the percentage of fat content of finely chopped pure meat depends not only on the spectrometric curve but also on the corresponding percentages of protein content and water content. Functional regression models have been used to handle this problem, where a scalar response variable is regressed to both functional covariates and scalar covariates. The partial functional linear model, a most frequently used mixed data model, has attracted a lot of interests in the literature. For instance, Shin (2009) considered a partial functional linear model, in which both the scalar covariates and the functional covariate are linear. Zhang et al. (2007) introduced a measurement error partial functional linear model. Various extensions of partial functional linear model have been proposed to broaden the applicability of functional regression models with mixed data in the literature. For example, Aneiros-Pérez and Vieu (2006) considered a semi-functional partial linear regression model in which the scalar covariates are the linear component and the functional covariate is nonparametric component. Dabo-Niang and Guillas (2010) proposed a functional semiparametric model. This model is similar to semi-functional partial linear model but with autocorrelated random errors. A hybrid model of functional and varying coefficient regressions, as an important extension of partial functional linear model is becoming popular in the literature. The model is defined in the following form:

$$Y = \alpha_0(U) + \boldsymbol{X}^T\boldsymbol{\alpha}(U) + \int_I \beta(t)Z(t)\mathrm{d}t + \varepsilon, \tag{1}$$

where Y is a scalar variable, $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^T$ are p-dimensional random vector of scalar covariates, U is a univariate scalar variable, $\alpha_0(U)$ is a baseline function and $\boldsymbol{\alpha}(U) = (\alpha_1(U), \alpha_2(U), \ldots, \alpha_p(U))^T$ are unknown varying coefficient functions, $Z(t)$ is a zero mean random functional predictor defined on a compact interval I, $\beta(t)$ is a square-integrable regression slope function, ε is an error term with mean zero and variance σ^2, and $(X, U, Z(t))$ and ε are independent. By the hybrid model, we describe a functional linear relationship plus a varying interaction term between the scalar covariates . It seems to be more sensible to characterize the dynamic feature in the varying interaction term which may exist in the data set. For example, as we discuss in section 5, the fat content in a piece of finely chopped pure meat will depend on the water content, and the dynamical pattern of this relationship is of importance. It would make much more sense to treat the parameters of the protein content as functions of the water content than constants. So, we will employe the above hybrid model to predict the percentage of fat content. The model is concerned about a varying interaction term between the protein and water content. On the other hand, the model is flexible as it takes the classical functional linear regression model and partial functional linear model as special cases if let $\boldsymbol{\alpha}(U) = \mathbf{0}$, $\alpha_0(U) = \alpha_0$ and $\boldsymbol{\alpha}(U) = \boldsymbol{\alpha}, \alpha_0(U) = \alpha_0$ respectively. Due to its flexibility to explore the dynamic features which may exist in the data, the hybrid model of functional and varying coefficient regressions has been investigated intensively. Peng et al. (2015) proposed a least squares-based spline approach to estimating the above hybrid model and provided the asymptotic behavior of their estimation. Feng et al. (2016) proposed a profile least squares estimation of the same model by use of functional principal component analysis and local linear smoothing technique.

The least square estimation procedures in the aforementioned two papers are based on the conditional mean of the response variable for the given set of covariates. As a result, there is lack of information on the response variable at the various quantile values (for example, the lower or upper quantiles). Furthermore, assumptions related to random errors in the least square estimators are not always valid in reality. Even a few outlying data points may introduce undesirable artificial features in the estimated functions. Here, to address these issues, we develop a novel and robust estimation procedure called quantile estimation for the hybrid model, which can be interpreted as the effect of covariates on the response variable at each quantile level. There are few studies on quantile-regression-based estimation procedures for non-hybrid functional regression models. In literature, Cardot et al. (2005) proposed a spline-based estimation for functional linear quantile regression models. Chen and Müller (2012) proposed a method for conditional quantile analysis for the generalized functional regression models. Kato (2012) studied estimation in functional linear quantile regression model and showed that the rate of convergence for slope function estimator was optimal in a minimax sense. Lu et al. (2014); Tang and Cheng (2014) also investigated the quantile estimation of partially functional linear models and the asymptotic performance of the proposed estimator.

In this paper, we focus on quantile estimation of the hybrid model between partially functional linear regression and varying coefficient models. Our contributions to this area are as follows. We develop the quantile estimators for the slope function, the baseline function and the varying coefficients in the above hybrid model with mixed data. Under some regularity conditions, we establish the asymptotic normality of the proposed estimators. We show that the global convergence rates of the proposed slope function estimator can attain the same optimal minimax rate as in functional linear regression. A Monte Carlo simulation study and a real application to spectrometric data show that the proposed estimation procedure has a few advantages over its competitors.

The article is organized as follows. The quantile estimation of the hybrid model between partially functional regression and varying coefficients is developed in Section 2. The asymptotic properties of the proposed quantile estimators are established in Section 3. The finite sample performance of the proposed estimators is presented in Section 4. The proposed method is then applied to the spectrometric data. Technical proofs are delayed to an Appendix.

2 Model and Estimation

2.1 Model

Given quantile level $\tau \in (0, 1)$, we consider the following hybrid quantile model of functional linear regression and varying-coefficients for mixed functional data

$$Y = \alpha_{0\tau}(U) + \boldsymbol{X}^T \boldsymbol{\alpha}_\tau(U) + \int_I \beta_\tau(t) Z(t) \mathrm{d}t + \varepsilon_\tau, \tag{2}$$

where $\alpha_{0\tau}(U)$ is a unknown baseline function and $\boldsymbol{\alpha}_\tau(U) = (\alpha_{1\tau}(U), \alpha_{2\tau}(U), \ldots, \alpha_{p\tau}(U))^T$ are unknown varying coefficient functions to be estimated, $U \in [u_l, u_r]$, $Z(t)$ is zero mean random functional predictor defined on a compact interval I, $\beta_\tau(t)$ is square-integrable regression slope function, ε_τ is a random error whose τth quantile conditional on $(\boldsymbol{X}, U, Z(t))$ being zero.

2.2 Estimation

Suppose that $\{(Y_i, \boldsymbol{X}_i, U_i, Z_i(t)), i = 1, 2, \ldots, n\}$ is a random sample generated from model (2). We estimate slope function $\beta_\tau(t)$, baseline function $\alpha_{0\tau}(U)$ and varying coefficients $\boldsymbol{\alpha}_\tau(U)$ in model (2), by minimizing the following quantile loss function

$$\sum_{i=1}^{n} \rho_\tau \left(Y_i - \alpha_{0\tau}(U_i) - \boldsymbol{X}_i^T \boldsymbol{\alpha}_\tau(U_i) - \int_I \beta_\tau(t) Z_i(t) \mathrm{d}t \right), \tag{3}$$

where $\rho_\tau(s) = s\{\tau - I(s < 0)\}$.

To begin with, we note that $\int_I \beta_\tau(t) Z_i(t) dt$ is simplified by expanding $\beta_\tau(t) = \sum_{k=1}^{\infty} b_{\tau k} \phi_k(t)$, where $\phi_1(t), \phi_2(t), \ldots$ are orthonormal basis of square-integrable function on interval I. The basis $\phi_1(t), \phi_2(t), \ldots$ can be chosen independently of data (e.g., Fourier basis, Spline basis, etc). Here, we adopt a principal component basis, constructed from the covariance function $K_Z(u, v) = \mathrm{Cov}(Z(u), Z(v))$ of the random process $Z(t)$ as follows. The spectral decomposition of $K_Z(u, v)$ is given by

$$K_Z(u, v) = \sum_{k=1}^{\infty} \lambda_k \phi_k(u) \phi_k(v), \tag{4}$$

where the principal component basis $\phi_1(t), \phi_2(t), \ldots$ is a complete orthonormal sequence of eigenfunctions of the transformations K_Z, with respective eigenvalues $\lambda_1 > \lambda_2 > \cdots > 0$. By Karhunen-Loève expansion, we have

$$Z(t) = \sum_{k=1}^{\infty} \xi_k \phi_k(t),$$

where $\xi_k = \int_I Z(t)\phi_k(t)\mathrm{d}t$. Hence, $\int_I \beta_\tau(t)Z(t)\mathrm{d}t = \sum_{k=1}^{\infty} b_{\tau k}\xi_k$. Correspondingly, $\int_I \beta_\tau(t)Z_i(t)\mathrm{d}t = \sum_{k=1}^{\infty} b_{\tau k}\xi_{ik}$, where $\xi_{ik} = \int_I Z_i(t)\phi_k(t)\mathrm{d}t$. However, in practice, the value of ξ_k and ξ_{ik} depend on the value of $\phi_k(t)$, but the scalars λ_k and the functions $\phi_k(t)$ are unknown and must be replaced by estimators in order to produce estimator of $\beta_\tau(t)$. For this purpose, we consider the empirical version of $K_Z(u,v)$ given by

$$\hat{K}_Z(u,v) = \frac{1}{n}\sum_{i=1}^{n} Z_i(u)Z_i(v) = \sum_{k=1}^{n} \hat{\lambda}_k \hat{\phi}_k(u)\hat{\phi}_k(v),$$

where $(\hat{\lambda}_k, \hat{\phi}_k)$ are pairs of eigenvalues and eigenfunctions, ordered such that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \hat{\lambda}_n \geq 0$. We take $(\hat{\lambda}_k, \hat{\phi}_k(t))$ as the estimator of $(\lambda_k, \phi_k(t))$. The functions $\hat{\phi}_1(t), \hat{\phi}_2(t), \ldots, \hat{\phi}_m(t)$ are known, where m is a tuning parameter for "frequency cut-off". By using the approximate expansion $\beta_\tau(t) \approx \sum_{k=1}^{m} b'_{\tau k}\hat{\phi}_k(t)$, we show that $\int_I \beta_\tau(t)Z(t)\mathrm{d}t$ can be properly approximated by $\sum_{k=1}^{m} b'_{\tau k}\hat{\xi}_k$, where $\hat{\xi}_k = \int_I Z(t)\hat{\phi}_k(t)\mathrm{d}t$. Consequently, $\int_I \beta_\tau(t)Z_i(t)\mathrm{d}t \approx \sum_{k=1}^{m} b'_{\tau k}\hat{\xi}_{ik}$, where $\hat{\xi}_{ik} = \int_I Z_i(t)\hat{\phi}_k(t)\mathrm{d}t$.

In order to approximate $\alpha_{0\tau}(U)$ and $\boldsymbol{\alpha}_\tau(U)$ for $U \in [u_l, u_r]$, we construct piecewise polynomial estimators of $\alpha_{0\tau}(U)$ and $\boldsymbol{\alpha}_\tau(U)$ of degree $\tilde{q}$. We divide $[u_l, u_r]$ into N_n subintervals of equal length. Then the length of every subinterval is $2h_0 = (u_r - u_l)/N_n$. Let $I_k = [u_l + 2(k-1)h_0, u_l + 2kh_0)$ for $1 \leq k \leq N_n - 1$ and $I_{N_n} = [u_r - 2h_0, u_r]$. Let u_k denote the centre of the interval I_k and $\chi_k(u)$ denote the indicator function of I_k, i.e.,

$$\chi_k(u) = \begin{cases} 1, & u \in I_k \\ 0, & u \notin I_k \end{cases}.$$

To facilitate the presentation, we need some more notations as follows. Let

$$\begin{aligned} \boldsymbol{B}_k(u) &= \left(1, (u-u_k)/h_0, \ldots, [(u-u_k)/h_0]^{\tilde{q}}\right)^\top, k = 1, \ldots, N_n, \\ \boldsymbol{B}(u) &= (\chi_1(u)\boldsymbol{B}_1(u)^\top, \ldots, \chi_{N_n}(u)\boldsymbol{B}_{N_n}(u)^\top)^\top, \boldsymbol{M}(u) = \mathrm{diag}(\boldsymbol{B}(u)^\top, \ldots, \boldsymbol{B}(u)^\top)^\top_{p\times p}. \end{aligned}$$

Denote $\boldsymbol{\omega}_k = (\omega_{k1}, \ldots, \omega_{k\tilde{q}})^\top$, $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^\top, \ldots, \boldsymbol{\omega}_{N_n}^\top)^\top$, $\boldsymbol{\gamma}_{jk} = (\gamma_{jk1}, \ldots, \gamma_{jk\tilde{q}})^\top$, $\boldsymbol{\gamma}_j = (\boldsymbol{\gamma}_{j1}^\top, \ldots, \boldsymbol{\gamma}_{jN_n}^\top)^\top$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_p^\top)^\top$. We use $\boldsymbol{B}(U)^\top\boldsymbol{\omega}$ and $\boldsymbol{M}(U)^\top\boldsymbol{\gamma}$ to approximate $\alpha_{0\tau}(U)$ and $\boldsymbol{\alpha}_\tau(U)$, respectively. Thus, quantile loss function in (3) can be approximated by the following target function

$$\sum_{i=1}^{n} \rho_\tau\left(Y_i - \boldsymbol{B}(U_i)^\top\boldsymbol{\omega} - \boldsymbol{X}_i^T\boldsymbol{M}(U_i)^\top\boldsymbol{\gamma} - \sum_{k=1}^{m} b'_{\tau k}\hat{\xi}_{ik}\right). \tag{5}$$

The solution to Equation (5) can be obtained numerically by linear programming method (such as Frisch-Newton Interior Point Method or Interior point method with preprocessing). For convenience, let $\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\gamma}}, \hat{b}'_{\tau k}$ be the minimizer of Equation (5). Then, the estimator of $\beta_\tau(t)$ is denoted by $\hat{\beta}_\tau(t) = \sum_{k=1}^m \hat{b}'_{\tau k}\hat{\phi}_k(t)$.

After estimating $\beta_\tau(t)$, for a given $u \in [u_l, u_r]$, when U tends to u, we employ the local linear approximation $\alpha_{0\tau}(U) \approx a_{0\tau} + b_{0\tau}(U-u), \boldsymbol{\alpha}_\tau(U) \approx \boldsymbol{a}_\tau + \boldsymbol{b}_\tau(U-u)$. We can obtain estimators of $a_{0\tau}, b_{0\tau}, \boldsymbol{a}_\tau, \boldsymbol{b}_\tau, \boldsymbol{\beta}_\tau$ by minimizing the following local weighted quantile loss function

$$\sum_{i=1}^n \rho_\tau\left(Y_i - a_{0\tau} - b_{0\tau}(U_i-u) - \boldsymbol{X}_i^T\left\{\boldsymbol{a}_\tau + \boldsymbol{b}_\tau(U_i-u)\right\} - \sum_{k=1}^m \hat{b}'_{\tau k}\hat{\xi}_{ik}\right) K_h(U_i-u). \tag{6}$$

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function and h is a bandwidth. The solution to Equation (6) can also be obtained numerically by linear programming method. For convenience, let $\hat{a}_{0\tau}, \hat{b}_{0\tau}, \hat{\boldsymbol{a}}_\tau, \hat{\boldsymbol{b}}_\tau$ be the minimizer of Equation (6). Then, the estimator of $\alpha_{0\tau}(u)$ and $\boldsymbol{\alpha}_\tau(u)$ are denoted by $\hat{a}_{0\tau}$ and $\hat{\boldsymbol{a}}_\tau$, respectively.

The above estimation procedure is summarized as follows:

Step 1: Compute $\hat{\xi}_{ik}$ and $\hat{\phi}_k(t)$ by functional principal component analysis method ($i = 1, \ldots, n; k = 1, \ldots, m$);

Step 2: Obtain $\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\gamma}}, \hat{b}'_{\tau k}$ by minimizing (5), then $\hat{\beta}_\tau(t) = \sum_{k=1}^m \hat{b}'_{\tau k}\hat{\phi}_k(t)$;

Step 3: Obtain $\hat{a}_{0\tau}, \hat{b}_{0\tau}, \hat{\boldsymbol{a}}_\tau, \hat{\boldsymbol{b}}_\tau$ by minimizing (6), then $\hat{\alpha}_{0\tau}(u) = \hat{a}_{0\tau}(u), \hat{\boldsymbol{\alpha}}_\tau(u) = \hat{\boldsymbol{a}}_\tau(u)$.

Remark 1 The proposed estimation is designed for use in situations where functional predictors are measured at a dense grid of regular space time points. For situations where this is not the case it may be feasible to use sparse functional principal components analysis method (see Yao et al., 2005) to produce the estimators $(\hat{\lambda}_k, \hat{\phi}_k)$.

Remark 2 The proposed procedure estimates the functional slope function and varying coefficients by minimizing quantile loss function. In the next section, we show that $\hat{\boldsymbol{\beta}}_\tau$ can result in a slope function estimator which achieve the optimal rate of convergence as in functional linear regression analysis.

2.3 Tuning parameter and bandwidth selection

To implement our estimation method, we need to choose the tuning parameter m, N_n and bandwidth h. Theorem 1 and Theorem 2 imply that the selection of the tuning parameter m, N_n and bandwidth h are of crucial importance. An appropriate choice of m, N_n and h can result in good estimators of the slope function and varying coefficients. We use Bayesian information criterion (BIC) to select m and N_n. The BIC is given by

$$BIC(m, N_n) = \log\left\{\sum_{i=1}^{n}\rho_\tau\left(Y_i - \boldsymbol{B}(U_i)^\top\hat{\boldsymbol{\omega}} - \boldsymbol{X}_i^T\boldsymbol{M}(U_i)^\top\hat{\boldsymbol{\gamma}} - \sum_{k=1}^{m} b'_{\tau k}\hat{\xi}_{ik}\right)\right\} + \frac{(m+N_n)\log n}{n}.$$

The optimal m and N_n are selected by minimizing BIC. The bandwidth h can be selected by leave-one-out cross-validation of the prediction error. More precisely, CV is defined as

$$CV(h) = \sum_{i=1}^{n}\rho_\tau\left(Y_i - \hat{\alpha}_{0\tau}^{(-i)}(U_i) - \boldsymbol{X}_i^T\hat{\boldsymbol{\alpha}}_\tau^{(-i)}(U_i) - \sum_{k=1}^{m}\hat{b}'_{\tau k}\hat{\xi}_{ik}\right),$$

$\hat{\alpha}_{0\tau}^{(-i)}(U_i)$ and $\hat{\boldsymbol{\alpha}}_\tau^{(-i)}(U_i)$ denote that the estimators of $\alpha_{0\tau}(U_i)$ and $\boldsymbol{\alpha}_\tau(U_i)$ computed without observation i. We find the minimizer of $CV(h)$, which is the selected value for h.

3 Asymptotic Properties

In this section we study asymptotic properties of the estimators proposed in Section 2. We first introduce some notations for the brevity of presentation. Let $f_\tau(\cdot|\boldsymbol{x}, u, z(t))$ and $F_\tau(\cdot|\boldsymbol{x}, u, z(t))$ denote the density function and cumulative distribution function of the error ε_τ condition on $(\boldsymbol{X}, U, Z(t)) = (\boldsymbol{x}, u, z(t))$, respectively. Denote the marginal density function of the covariate U by $f_U(\cdot)$. Let $\boldsymbol{G}(u) = \mathrm{E}\{f_\tau(0|\boldsymbol{X}, U, Z(t))(1, \boldsymbol{X}^T)^T(1, \boldsymbol{X}^T)|U = u\}$, $\boldsymbol{H}(u) = \mathrm{E}\{(1, \boldsymbol{X}^T)^T(1, \boldsymbol{X}^T)|U = u\}$. For kernel function $K(\cdot)$, define $\mu_j = \int u^j K(u)\mathrm{d}u$ and $\nu_j = \int u^j K^2(u)\mathrm{d}u$, $j = 0, 1, 2, \ldots$. We use the symbol $a_n \asymp b_n$ to denote that the ratio a_n/b_n is bounded away from zero and infinity. Let the symbol $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ denote inner product and norm.

The following conditions are needed:

(C1) The covariate U has a bounded support Θ and its density function $f_U(\cdot)$ is positive and has a continuous second derivative.

(C2) $K(\cdot)$ is a nonnegative and symmetric density function with bounded support and satisfies a Lipschitz condition.

(C3) $F_\tau(0|\boldsymbol{x}, u, z(t)) = \tau$ for all $(\boldsymbol{x}, u, z(t))$, $f_\tau(\cdot|\boldsymbol{x}, u, z(t))$ is bounded away from zero and has a continuous and uniformly bounded derivative. We also assume that there exist constants c_0 and c_1 such that $0 < c_0 \le f_\tau(0|\boldsymbol{x}, u, z(t)) \le c_1 < \infty$.

(C4) $Z(t)$ is square-integrable random function supported on the compact interval I, and has a zero mean and finite fourth moment. We assume that for each j, $\mathrm{E}(\xi_j^4) \le B_1\lambda_j^2$ for some constant B_1.

(C5) The eigenvalues λ_j in the spectral decomposition (4) satisfy

$$B_2^{-1} j^{-\beta_1} \leq \lambda_j \leq B_2 j^{-\beta_1}, \qquad \lambda_j - \lambda_{j+1} \geq B_2^{-1} j^{-(\beta_1+1)}, \qquad j \geq 1,$$

where $\beta_1 > 1, B_2 > 0$.

(C6) For the Fourier coefficients $b_{\tau j}$ of $\beta_\tau(t)$, there exist constant B_3, $\beta_2 > 8 + \frac{3\beta_1}{2}$ such that $|b_{\tau j}| \leq B_3^{-1} j^{-\beta_2}$.

(C7) The baseline function $\alpha_{0\tau}(u)$ and varying function $\boldsymbol{\alpha}_\tau(u)$ have continuous $\tilde{q}$ derivatives such that $|\alpha_{0\tau}^{(\tilde{q})}(u) - \alpha_{0\tau}^{(\tilde{q})}(u')| \leq B_4|u-u'|^\varsigma$ and $\|\boldsymbol{\alpha}_\tau^{(\tilde{q})}(u) - \boldsymbol{\alpha}_\tau^{(\tilde{q})}(u')\| \leq B_4|u-u'|^\varsigma$ for $u_l \leq u, u' \leq u_r$, where $0 < \varsigma \leq 1$ and B_4 is a positive constant. Think of $q = \tilde{q} + \varsigma$ as a measure of the smoothness of the function $\alpha_{0\tau}(U)$ and $\boldsymbol{\alpha}_\tau(U)$, $q > (3\beta_1 + 6\beta_2 - 2)/4$.

(C8) The tuning parameter m satisfies that $m \asymp n^{1/(\beta_1+2\beta_2)}$ and N_n also satisfies that $N_n \asymp n^{1/(\beta_1+2\beta_2)}$.

(C9) $\mathrm{E}X_j^4 < \infty, j = 1, \ldots, p$.

(C10) $\mathrm{E}(\boldsymbol{X}|U) = 0$, $\mathrm{E}(Z(t)|U, \boldsymbol{X}) = 0$ and $\mathrm{E}(\xi_i \xi_j | U, \boldsymbol{X}) = 0$ for $i \neq j$. For each i, $\mathrm{E}(\xi_i^2|U, \boldsymbol{X}) < B_5 \lambda_i$ for some constant B_5.

(C11) $\boldsymbol{G}(u)$ are non-singular for all $u \in \Theta$.

Remark 1 Conditions C1-C11 are not the weakest possible conditions. They are imposed to facilitate the proof of the following theorems. Conditions C1-C4, C7, C9 and C11 are required in the context of nonfunctional varying coefficient partially linear model (see Kai et al., 2011), while conditions C5, C6 and C8 are needed to cope with linear part corresponding to the functional predictor $Z(t)$ of varying coefficient partially functional linear regression model with mixed data. And conditions C5, C6 and C8 are quite usual in functional linear regression model (see Cai and Hall, 2006; Hall and Horowitz, 2007). Condition C10 is a technical condition for description of the correlation between scalar covariate X and U and functional covariate $Z(t)$.

Theorem 1. *Suppose that the regularity conditions C1-C11 hold, then*

$$\int_I (\hat{\beta}_\tau(t) - \beta_\tau(t))^2 \mathrm{d}t = O_p\left(n^{-\frac{2\beta_2 - 1}{\beta_1 + 2\beta_2}}\right). \tag{7}$$

Theorem 2. *Suppose that the regularity conditions C1-C11 hold. If $h \to 0$, $nh \to \infty$ and $nh/\log(1/h) \to \infty$ as $n \to \infty$, then*

$$\sqrt{nh}\left[\begin{pmatrix} \hat{\alpha}_{0\tau}(u) - \alpha_{0\tau}(u) \\ \hat{\boldsymbol{\alpha}}_\tau(u) - \boldsymbol{\alpha}_\tau(u) \end{pmatrix} - \frac{\mu_2 h^2}{2}\begin{pmatrix} \alpha_{0\tau}''(u) \\ \boldsymbol{\alpha}_\tau''(u) \end{pmatrix}\right] \xrightarrow{\mathscr{L}} N\left(\mathbf{0}, \frac{\nu_0 \tau(1-\tau)}{f_U(u)} \boldsymbol{G}^{-1}(u) \boldsymbol{H}(u) \boldsymbol{G}^{-1}(u)\right). \tag{8}$$

Remark 2 Our results shows that we can obtain the same rate of convergence as for the estimator in functional linear regressionwhich are optimal in the minimax sense (see Hall and Horowitz, 2007). Under the condition about kernel bandwidth h in Theorem 2, we can get the asymptotic normality of estimators of baseline function and varying coefficient functions.

4 Simulation Studies

In this section, we implement simulation studies to investigate the performance of the proposed estimation methods. The data sets are generated from the following model:

$$Y_i = \alpha_1(U_i)X_{1i} + \alpha_2(U_i)X_{2i} + \int_I \beta(t)Z_i(t)\mathrm{d}t + \varepsilon_i, \quad i = 1, 2, \ldots, n.$$

For the functional linear component, we take the same form as Hall and Horowitz (2007), that is, $I = [0,1]$, $\beta(t) = \sum_{k=1}^{50} b_k\phi_k(t)$ and $Z_i(t) = \sum_{k=1}^{50} \vartheta_k W_{ik}\phi_k(t)$, where $\phi_1(t) = 1$, $\phi_k(t) = \sqrt{2}\cos[(k-1)\pi t]$ for $k \geq 2$, $b_1 = 0.3$, $b_k = 4(-1)^{k+1}k^{-2}$ for $k \geq 2$, and $\vartheta_k = (-1)^{k+1}k^{-1}$, W_{ik} are independent and identically distributed uniform random variables on $(-\sqrt{3}, \sqrt{3})$. For the varying coefficient component, we let $\alpha_1(U) = \sin(2\pi U)$, $\alpha_2(U) = \sin(6\pi U)$, the covariate U is uniformly distributed on $(0,1)$, X_1, X_2 are independent and identically distributed normal random variables with mean 0 and variance 1. Furthermore, U and X_1, X_2 are independent.

In our simulation, we consider four cases for error terms ε: $N(0, 0.5^2)$, standard Cauchy, t(3) and mixture of normals $0.9N(0, 0.5^2) + 0.1N(0, 5^2)$. We also consider three choices for the number of samples n = 200, 400 and 600. Each $Z_i(t)$ is observed at 100 equally space points on $[0,1]$. In order to evaluate the performance of estimators of different method, we compare the profile least squares (PLS) method (see Feng et al., 2016) and our quantile regression (QR) method. We focus on τ = 0.25, 0.5 and 0.75 in quantile regression. The Epanechnikov kernel is used in the simulations. We use the BIC criterion and cross-validation procedure as described in section 2.3 to select the tuning parameters N_n, m and bandwidth h. All simulations are replicated for 1000 times.

Performance of estimator of functional slope function $\beta(t)$ is assessed using the square root of the integrated squared errors (RISE) defined as

$$\mathrm{RISE}\left\{\hat{\beta}(t)\right\} = \left\{\int_0^1 \left[\hat{\beta}(t) - \beta(t)\right]^2 \mathrm{d}t\right\}^{\frac{1}{2}},$$

while performance of the estimate of varying coefficient functions $\alpha_1(U)$ and $\alpha_2(U)$ are assessed using the square root of mean average squared errors (RASE) defined as

$$\mathrm{RASE}\left\{\hat{\alpha}_1(U)\right\} = \left\{\frac{1}{100}\sum_{i=1}^{100}\left[\hat{\alpha}_1(u_i) - \alpha_1(u_i)\right]^2\right\}^{\frac{1}{2}}$$

and

$$\mathrm{RASE}\left\{\hat{\alpha}_2(U)\right\} = \left\{\frac{1}{100}\sum_{i=1}^{100}\left[\hat{\alpha}_2(u_i) - \alpha_2(u_i)\right]^2\right\}^{\frac{1}{2}}.$$

where $u_i, i = 1, 2, \cdots, 100$ are 100 equally space points on interval $[0, 1]$.

To save space we only show the BIC and CV scores for different tuning parameters and bandwidth under $t(3)$ distribution error with $n = 400$, $\tau = 0.5$. Table 1 presents BIC scores for different N_n and m. The minimum BIC score is emphasized with boldface font. Table 2 presents CV scores for different bandwidths. The optimal bandwidth h is obtained by leave-one-out cross-validation for given optimal m. The minimum CV score is also emphasized with boldface font.

Table 1: The BIC scores of different tuning parameters with $t(3)$ distribution error for $\tau = 0.5$, $n = 400$

	m									
N_n	1	2	3	4	5	6	7	8	9	10
1	5.974	5.933	5.926	5.937	5.944	5.954	5.964	5.975	5.985	5.995
2	5.960	5.919	5.907	5.917	5.920	5.931	5.941	5.952	5.961	5.972
3	5.933	5.888	5.887	5.897	5.903	5.914	5.925	5.935	5.946	5.954
4	5.949	5.895	5.894	5.904	5.911	5.921	5.932	5.942	5.952	5.962
5	5.928	**5.880**	5.882	5.890	5.900	5.910	5.921	5.931	5.940	5.949
6	5.943	5.896	5.893	5.903	5.907	5.918	5.929	5.939	5.950	5.959
7	5.932	5.885	5.885	5.894	5.903	5.913	5.924	5.934	5.943	5.950
8	5.933	5.881	5.887	5.895	5.904	5.914	5.925	5.935	5.945	5.953
9	5.939	5.891	5.886	5.896	5.905	5.916	5.926	5.937	5.947	5.956
10	5.941	5.895	5.901	5.911	5.921	5.931	5.941	5.952	5.962	5.968

Table 3-6 list RISEs of $\hat{\beta}(t)$ and RASEs of $\hat{\alpha}_1(U)$ and $\hat{\alpha}_2(U)$ under different error terms. There is a general tendency for RISE of $\hat{\beta}(t)$ and RASE of $\hat{\alpha}_1(U)$ and $\hat{\alpha}_2(U)$ to decrease as sample sizes increases. From Table 3, we can see that both PLS estimators and QR estimators have small RISEs and RASEs under normal error terms. QR estimators are slightly worse than PLS estimators as expected. When the error term follows heavy-tailed distributions, Table 4-6 illustrate

Table 2: The CV scores of bandwidths with $t(3)$ distribution error for $\tau = 0.5$, $n = 400$.

	h									
	0.032	0.048	0.064	0.080	0.096	0.111	0.127	0.143	0.159	0.175
CV	0.596	0.584	0.581	0.578	0.578	**0.577**	0.579	0.580	0.580	0.582

that QR estimators is robust and more efficient than PLS estimators. Specifically, when the error follows standard Cauchy distribution, PLS estimators have very large RISEs and RASEs while QR estimators have reasonably small RISEs and RASEs. This is because PLS fails when the error variance is infinite.

Table 3: RISEs and RASEs with standard deviations(in parentheses) with normal distribution error $N(0, 0.5^2)$

n	Method	$\hat{\beta}(t)$	$\hat{\alpha}_1(U)$	$\hat{\alpha}_2(U)$
200	PLS	0.1467(0.0454)	0.0306(0.0232)	0.0387(0.0176)
	QR(0.25)	0.1693(0.0801)	0.0497(0.0375)	0.0621(0.0307)
	QR(0.50)	0.1700(0.0504)	0.0498(0.0169)	0.0696(0.0277)
	QR(0.75)	0.1639(0.0544)	0.0527(0.0351)	0.0701(0.0409)
400	PLS	0.1351(0.0298)	0.0097(0.0044)	0.0181(0.0062)
	QR(0.25)	0.1427(0.0327)	0.0176(0.0076)	0.0296(0.0120)
	QR(0.50)	0.1435(0.0508)	0.0132(0.0065)	0.0250(0.0128)
	QR(0.75)	0.1440(0.0324)	0.0099(0.0067)	0.0272(0.0097)
600	PLS	0.1303(0.0239)	0.0060(0.0025)	0.0148(0.0041)
	QR(0.25)	0.1353(0.0255)	0.0112(0.0042)	0.0224(0.0065)
	QR(0.50)	0.1232(0.0306)	0.0132(0.0045)	0.0155(0.0055)
	QR(0.75)	0.1335(0.0242)	0.0125(0.0052)	0.0194(0.0066)

To evaluate reliability of the estimators, we construct pointwise confidence intervals based on the asymptotic normalities. To save space we describe the construction of confidence intervals of $\alpha_1(u)$ and $\alpha_2(u)$ for $u = 0.2, 0.4, 0.6$ and 0.8 under $t(3)$ distribution only. It follows from (8)that approximate $100(1-\alpha)\%$ confidence intervals for $\alpha_1(u)$ and $\alpha_2(u)$ can be expressed respectively as follows:

$$\hat{\alpha}_1(u) - \frac{\mu_2 h^2}{2}\hat{\alpha}_1''(u) \pm z_{1-\alpha/2}\sqrt{\frac{\nu_0\tau(1-\tau)}{\hat{f}_U(u)nh}}(\hat{\boldsymbol{G}}^{-1}(u)\hat{\boldsymbol{H}}(u)\hat{\boldsymbol{G}}^{-1}(u))_{11}^{1/2}$$

Table 4: RISEs and RASEs with standard deviations(in parentheses) with standard Cauchy distribution error

n	Method	$\hat{\beta}(t)$	$\hat{\alpha}_1(U)$	$\hat{\alpha}_2(U)$
200	PLS	3905.88(48553.82)	1721.01(9514.03)	8141.77(1050.17)
	QR(0.25)	1.3376(0.0561)	0.5719(1.5715)	1.0221(1.5855)
	QR(0.50)	1.3365(0.0515)	0.7491(1.0359)	1.2059(1.4219)
	QR(0.75)	1.3386(0.0462)	0.4610(0.7657)	1.4677(1.4366)
400	PLS	3581.39(3188.14)	1424.86(4157.33)	3406.98(1408.77)
	QR(0.25)	0.3271(0.0252)	0.1171(0.0780)	0.2120(0.0970)
	QR(0.50)	0.3267(0.0243)	0.1128(0.0704)	0.1636(0.0917)
	QR(0.75)	0.3270(0.0254)	0.1052(0.0764)	0.2469(0.1155)
600	PLS	3562.84(2362.09)	1232.55(2475.53)	2467.62(1017.32)
	QR(0.25)	0.3229(0.0188)	0.0649(0.0376)	0.1572(0.0535)
	QR(0.50)	0.3232(0.0191)	0.0658(0.0336)	0.1048(0.0433)
	QR(0.75)	0.3212(0.0194)	0.0651(0.0360)	0.1816(0.0509)

Table 5: RISEs and RASEs with standard deviations(in parentheses) with t(3) distribution error

n	Method	$\hat{\beta}(t)$	$\hat{\alpha}_1(U)$	$\hat{\alpha}_2(U)$
200	PLS	0.4117(0.1256)	0.2841(0.2858)	0.2874(0.2771)
	QR(0.25)	0.3887(0.0970)	0.1988(0.1692)	0.2184(0.1236)
	QR(0.50)	0.3864(0.0952)	0.1221(0.0725)	0.1778(0.0953)
	QR(0.75)	0.3937(0.1027)	0.1728(0.1109)	0.2075(0.0851)
400	PLS	0.3830(0.0693)	0.1024(0.0831)	0.1149(0.0777)
	QR(0.25)	0.3528(0.0673)	0.0678(0.0325)	0.1081(0.0517)
	QR(0.50)	0.3511(0.0626)	0.0483(0.0287)	0.1034(0.0340)
	QR(0.75)	0.3550(0.0687)	0.0833(0.0473)	0.1020(0.0476)
600	PLS	0.3690(0.0573)	0.0577(0.0943)	0.1005(0.0595)
	QR(0.25)	0.2422(0.0517)	0.0324(0.0203)	0.0799(0.0304)
	QR(0.50)	0.2414(0.0518)	0.0298(0.0140)	0.0868(0.0247)
	QR(0.75)	0.2427(0.0534)	0.0294(0.0242)	0.0673(0.0284)

Table 6: RISEs and RASEs with standard deviations(in parentheses) with mixture of normals distribution error $0.9N(0, 0.5^2) + 0.1N(0, 5^2)$

n	Method	$\hat{\beta}(t)$	$\hat{\alpha}_1(U)$	$\hat{\alpha}_2(U)$
200	PLS	0.3453(0.2300)	0.7864(0.5148)	0.4973(0.2728)
	QR(0.25)	0.1762(0.0613)	0.2014(0.2410)	0.3733(0.2463)
	QR(0.50)	0.1727(0.0547)	0.1397(0.1259)	0.2320(0.2606)
	QR(0.75)	0.1738(0.0561)	0.3598(0.2676)	0.2495(0.3166)
400	PLS	0.2208(0.0927)	0.1746(0.1377)	0.1756(0.2161)
	QR(0.25)	0.1411(0.0319)	0.0620(0.0955)	0.0640(0.0871)
	QR(0.50)	0.1410(0.0312)	0.0200(0.0121)	0.0298(0.0150)
	QR(0.75)	0.1422(0.0320)	0.0322(0.0210)	0.0396(0.0226)
600	PLS	0.1957(0.0897)	0.0989(0.0719)	0.0976(0.0574)
	QR(0.25)	0.1354(0.0259)	0.0212(0.0132)	0.0236(0.0100)
	QR(0.50)	0.1349(0.0261)	0.0110(0.0044)	0.0213(0.0071)
	QR(0.75)	0.1356(0.0259)	0.0207(0.0102)	0.0238(0.0121)

and

$$\hat{\alpha}_2(u) - \frac{\mu_2 h^2}{2}\hat{\alpha}_2''(u) \pm z_{1-\alpha/2}\sqrt{\frac{\nu_0\tau(1-\tau)}{\hat{f}_U(u)nh}}(\hat{\boldsymbol{G}}^{-1}(u)\hat{\boldsymbol{H}}(u)\hat{\boldsymbol{G}}^{-1}(u))_{22}^{1/2},$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard Gaussian distribution, $(\hat{\boldsymbol{G}}^{-1}(u)\hat{\boldsymbol{H}}(u)\hat{\boldsymbol{G}}^{-1}(u))_{11}^{1/2}$ and $(\hat{\boldsymbol{G}}^{-1}(u)\hat{\boldsymbol{H}}(u)\hat{\boldsymbol{G}}^{-1}(u))_{22}^{1/2}$ are the $(1,1)$th and $(2,2)$th entries of the matrix $(\hat{\boldsymbol{G}}^{-1}(u)\hat{\boldsymbol{H}}(u)\hat{\boldsymbol{G}}^{-1}(u))^{1/2}$, $\hat{\alpha}_1''(u)$ and $\hat{\alpha}_2''(u)$ are local polynomial estimators of $\alpha_1''(u)$ and $\alpha_2''(u)$, $\hat{f}_U(u)$ is a kernel density estimator of U, and $\hat{\boldsymbol{G}}(u)$ and $\hat{\boldsymbol{H}}(u)$ are estimators of $\boldsymbol{G}(u)$ and $\boldsymbol{H}(u)$. The average coverage probabilities of 90% confidence intervals are listed in Table 7. From Table 7, we can see that the simulation results confirm the asymptotic properties: the coverage probabilities approach to the nominal value as sample size increase. The performance with small sample size may be poor, and the estimation of $\alpha_1''(u)$, $\alpha_2''(u)$, $\hat{\boldsymbol{G}}(u)$ and $\hat{\boldsymbol{H}}(u)$ has a large impact on the performance especially when sample size is small. This is not surprising since some of these quantities are more difficult to estimate than the functions of interest.

We also plot the estimators of $\alpha_1(U)$and $\alpha_2(U)$ and 90% pointwise confidence intervals. To save space we only show results of the estimators and 90% pointwise confidence intervals under $t(3)$ distribution error with $n = 400$, $\tau = 0.5$. Figure 1 shows the true functions of $\alpha_1(U)$ and $\alpha_2(U)$ together with some of their estimators and 90% pointwise confidence intervals under $t(3)$ distribution error with $n = 400$, $\tau = 0.5$. The true functions of $\beta(t)$ and its pointwise medians,

Table 7: Average coverage probabilities of 90% confidence intervals with $t(3)$ distribution error

		$\alpha_1(u)$				$\alpha_2(u)$			
τ	n	$u=0.2$	$u=0.4$	$u=0.6$	$u=0.8$	$u=0.2$	$u=0.4$	$u=0.6$	$u=0.8$
0.25	200	0.763	0.750	0.775	0.752	0.700	0.734	0.705	0.726
	400	0.847	0.842	0.858	0.84	0.832	0.792	0.798	0.813
	600	0.905	0.899	0.900	0.897	0.887	0.883	0.888	0.882
0.5	200	0.765	0.775	0.778	0.772	0.692	0.689	0.750	0.798
	400	0.881	0.861	0.866	0.830	0.870	0.845	0.851	0.888
	600	0.914	0.886	0.887	0.911	0.905	0.923	0.910	0.912
0.75	200	0.751	0.739	0.764	0.744	0.784	0.726	0.735	0.783
	400	0.831	0.847	0.863	0.855	0.819	0.794	0.802	0.841
	600	0.923	0.912	0.902	0.895	0.904	0.876	0.888	0.902

5% and 95% quantiles of the 1000 simulations are also plotted in Figure 1(c). We can see that the estimated curves (dotted line) is close to the true curve (solid line). Overall, Our proposed estimation methods shows better performance even with infinite variance errors. The simulation studies indicate that the proposed estimation procedure in Section 2 is effective in the varying coefficient partially functional linear regression model with mixed data.

5 A real application

In this section, we apply the proposed method to analyze the spectrometric data which are available from `http://lib.stat.cmu.edu/datasets/tecator`. These data are obtained for 215 pieces of pure meat. Each data sample contains fat, protein, water contents and spectrometric curve. The three contents measured in percent, are determined by analytic chemistry. Spectrometric curve consist of 100 wavelengths absorbance spectrum records. Our aim is to predict the fat content Y from the spectrometric curve $Z(t)$ and the corresponding percentages of protein content X and water content U. To capture interaction effect between the corresponding percentages of protein content X and water content U and find more accurately the underlying relationship between the response variable and the covariates, we consider hybrid model between partially functional linear regression with varying coefficients. Specifically, we consider the following model:

$$Y = \alpha_0(U) + X\alpha_1(U) + \int_{850}^{1050} \beta(t)Z(t)\mathrm{d}t + \varepsilon.$$

In order to evaluate the predictive ability of the model, we use only part of the data with data selection performed in the same way as in Aneiros-Pérez and Vieu (2006). We randomly select

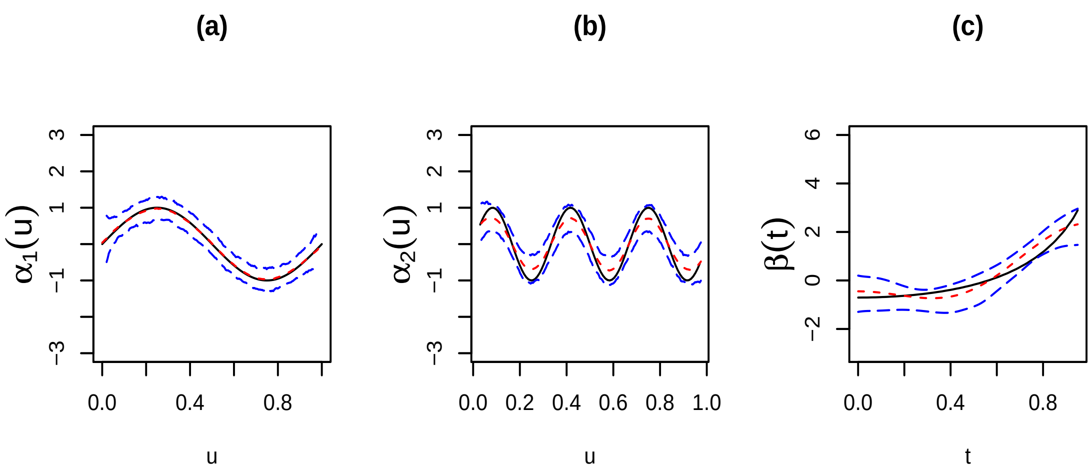

Figure 1: Plots of the true functions and their estimators when the error term follows $t(3)$ distribution, the sample size $n = 400$ and $\tau = 0.5$. Solid lines stand for the true functions. Dotted lines in (a) and (b) correspond to the pointwise estimated of $\alpha_1(U)$ and $\alpha_2(U)$, respectively. Dotted lines in (c) correspond to the pointwise medians of $\beta(t)$. Dashed lines in (a) and (b) correspond to the 90% pointwise confidence intervals of $\alpha_1(U)$ and $\alpha_2(U)$, respectively. Dashed lines in (c) correspond to the pointwise 5% and 95% quantiles of $\beta(t)$.

165 observations as training sample I and the remaining 50 observations as testing sample J. We use three kinds of different models to predict the fat content of a meat sample. One is semi-functional partial linear model (see Aneiros-Pérez and Vieu, 2006), another is partial functional linear regression model (see Shin, 2009) and the third is our model. For semi-functional partial linear model and partial functional linear regression, we employ their methods to the data. For our model, we apply profile least squares estimation method to the data. The criteria used on the test sample J in order to compare the skill of the different models is mean quadratic error of prediction $\frac{1}{50}\sum_{j\in J}(Y_j - \hat{Y}_j)^2/\mathrm{Var}_J(Y)$. The process is replicated for 500 times. The different models used and the corresponding values of this criteria are shown in Table 8.

Table 8: Means and standard errors (in parentheses) of test prediction error for different models

Models	Test prediction error
(i)$Y = U\theta_1 + X\theta_2 + g\left(Z(t)\right) + \varepsilon$	0.0168(0.0063)
(ii)$Y = \mu + U\theta_1 + X\theta_2 + \int_{850}^{1050}\beta(t)Z(t)\mathrm{d}t + \varepsilon$	0.0075(0.0045)
(iii)$Y = \alpha_0(U) + X\alpha_1(U) + \int_{850}^{1050}\beta(t)Z(t)\mathrm{d}t + \varepsilon$	0.0061(0.0035)

We observe that the mean and standard error of the prediction mean quadratic error in model (iii) is the smallest among the three models. The model (iii) improves more than 63.6% upon the model (i) and more than 18.5% upon the model (ii) in terms of prediction mean quadratic error. So, the model (iii) is a competitive one for such data.

Finally, profile least squares estimation method is used in our model to analyze the normality of the residuals. The norm quantile-quantile of the residuals is shown in Figure 2 (a), from which we can see apparently that the residuals cannot follow normal distribution. We also make a Shapiro-Wilk hypothesis test to judge the normality of the residuals. By Shapiro-Wilk test, we find that the p value is less than 5.497×10^{-5}. This reminds us further that the error cannot be normal, and the mean regression based on least square is unsuitable here. So, our quantile regression method with $\tau = 0, 25, 0.5$ and 0.75 is used here to analyze interaction effect between the corresponding percentages of protein content X and water content U. The kernel used in the real analysis is $K(u) = 0.75(1-u^2)I_{[0,1]}(u)$. The bandwidths and tuning parameters are chosen as $h = 5.8$, $N_n = 7$,$m = 13$ for $\tau = 0.25$ and $h = 6.7$, $N_n = 7$, $m = 17$ for $\tau = 0.5$ and $h = 7.2$, $N_n = 3$, $m = 11$ for $\tau = 0.75$. To save space we present results with $\tau = 0.5$. The estimator and 90% pointwise confidence intervals of nonparametric function $\alpha_1(U)$ with $\tau = 0.5$ is presented in Figure 2 (b). Figure 2 (b) indicates that the interaction effect between protein content X and water content U is negative and decreases as the water content U increases, which shows that interaction effect between protein content X and water content U is nonlinear. We also construct pointwise estimated interaction effect function $\alpha_1(U)$ at $\tau = 0.25, 0.5$ and 0.75 and show it in Figure 2(c). Both estimators show similar values and trends. It is apparent that the interaction effect between water content and protein content is negative for small U and then tend to stable for large U when τ increase. For example, the stable point is about 65 for $\tau = 0.5$ and is about 50 for $\tau = 0.75$. These findings are helpful to uncover and understand the underlying interaction relationship between water content and protein content.

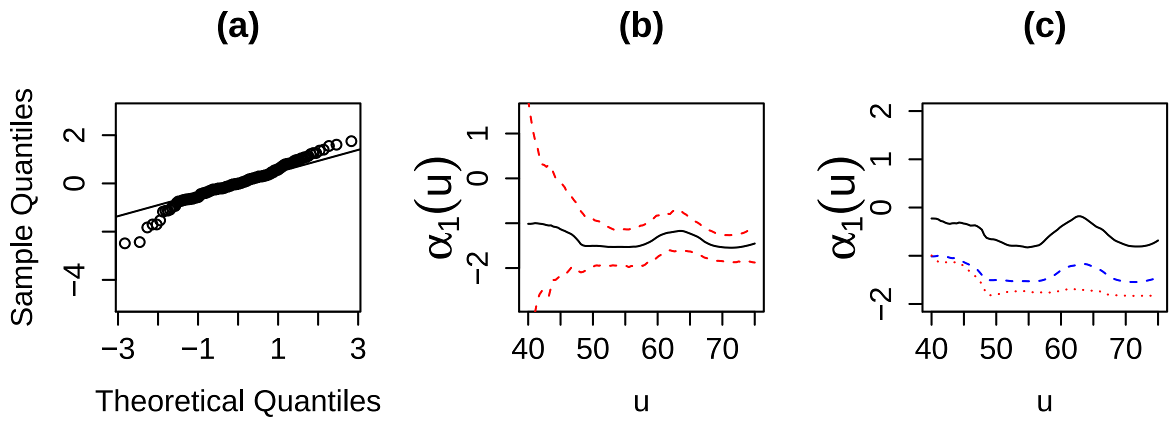

Figure 2: (a) QQ plot of the residual for profile least squares estimation method. (b) Pointwise estimated interaction effect function $\alpha_1(U)$ for $\tau = 0.5$ is shown as solid line. Pointwise 90% confidence intervals are given as dashed lines. (c) Solid, dashed and dotted lines corresponding to the pointwise estimated interaction effect function for $\tau = 0.25$, 0.5 and 0.75, respectively.

6 Conclusion

We have proposed a quantile estimation of a hybrid of functional regression and varying coefficient models for the analysis of the spectrometric data. We have established an asymptotic theory for the proposed estimation. We have conducted a Monte Carlo study to demonstrate the advantage of the proposed procedure over the existing least squares-based approaches.

Appendix

Proof. The proof of Theorem 1 will require some notations and Lemmas. We first introduce some notations. Let $\boldsymbol{A}_i = (\xi_{i1}, \ldots, \xi_{im})^\top$, $\hat{\boldsymbol{A}}_i = (\hat{\xi}_{i1}, \ldots, \hat{\xi}_{im})^\top$, $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$, $\boldsymbol{\beta}_\tau = (b_{\tau 1}, \ldots, b_{\tau m})^\top$, $\boldsymbol{\beta}'_\tau = (b'_{\tau 1}, \ldots, b'_{\tau m})^\top$, $\boldsymbol{P}_n = \sum_{i=1}^n \boldsymbol{M}(U_i)\boldsymbol{X}_i\boldsymbol{X}_i^\top \boldsymbol{M}(U_i)^\top$, $\boldsymbol{V}_{i1} = n^{-1/2}\boldsymbol{\Lambda}^{-1/2}\hat{\boldsymbol{A}}_i$, $\boldsymbol{V}_{i2} = (N_n/n)^{1/2}\boldsymbol{B}(U_i)$, $\quad \boldsymbol{V}_{i3} = \boldsymbol{P}_n^{-1/2}\boldsymbol{M}(U_i)\boldsymbol{X}_i$ and

$$
\begin{aligned}
\boldsymbol{F}_{0\tau} &= \left(\alpha_{0\tau}(u_1), \ldots, h_0^q \alpha_{0\tau}^{(q)}(u_1)/q!, \ldots, \alpha_{0\tau}(u_{N_n}), \ldots, h_0^q \alpha_{0\tau}^{(q)}(u_{N_n})/q!\right)^\top, \\
\boldsymbol{F}_{\tau} &= \left(\boldsymbol{\alpha}_{\tau}(u_1), \ldots, h_0^q \boldsymbol{\alpha}_{\tau}^{(q)}(u_1)/q!, \ldots, \boldsymbol{\alpha}_{\tau}(u_{N_n}), \ldots, h_0^q \boldsymbol{\alpha}_{\tau}^{(q)}(u_{N_n})/q!\right)^\top.
\end{aligned}
$$

Set $\boldsymbol{\theta}_1 = n^{1/2}\boldsymbol{\Lambda}^{1/2}(\boldsymbol{\beta}'_\tau - \boldsymbol{\beta}_\tau)$, $\boldsymbol{\theta}_2 = (n/N_n)^{1/2}(\boldsymbol{\omega} - \boldsymbol{F}_{0\tau})$, $\boldsymbol{\theta}_3 = \boldsymbol{P}_n^{1/2}(\boldsymbol{\gamma} - \boldsymbol{F}_\tau)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, \boldsymbol{\theta}_3^\top)^\top$, $\boldsymbol{V}_i = (\boldsymbol{V}_{i1}^\top, \boldsymbol{V}_{i2}^\top, \boldsymbol{V}_{i3}^\top)^\top$, $W_i = \alpha_{0\tau}(U_i) - \boldsymbol{B}(U_i)^\top \boldsymbol{F}_{0\tau} + \boldsymbol{X}_i^\top\left(\boldsymbol{\alpha}_\tau(U_i) - \boldsymbol{M}(U_i)^\top \boldsymbol{F}_\tau\right) + \sum_{j=m+1}^\infty b_{\tau j}\xi_{ij} + \sum_{j=1}^m b_{\tau j}(\xi_{ij} - \hat{\xi}_{ij})$, $\mathcal{I} = \{(Z_i(t), X_i, U_i)\}$, $\psi_\tau(s) = \tau - I(s < 0)$, $S_{n,i}(\boldsymbol{\theta}) = \rho_\tau(W_i + \varepsilon_{\tau i} - \boldsymbol{V}_i^\top\boldsymbol{\theta}) - \rho_\tau(W_i + \varepsilon_{\tau i})$, $S_n(\boldsymbol{\theta}) = \sum_{i=1}^n S_{n,i}(\boldsymbol{\theta})$, $\Gamma_n(\boldsymbol{\theta}) = \mathrm{E}\{S_{n,i}(\boldsymbol{\theta})|\mathcal{I}\}$, $\Gamma_n(\boldsymbol{\theta}) = \sum_{i=1}^n \Gamma_{n,i}(\boldsymbol{\theta})$, $R_{n,i}(\boldsymbol{\theta}) = S_{n,i}(\boldsymbol{\theta}) - \Gamma_{n,i}(\boldsymbol{\theta}) + \boldsymbol{V}_i^\top\boldsymbol{\theta}\psi_\tau(\varepsilon_{\tau i})$, $R_n(\boldsymbol{\theta}) = \sum_{i=1}^n R_{n,i}(\boldsymbol{\theta})$. For convenience, we use the symbol $\boldsymbol{A}_n = O_p(a_n)$ (or $o_p(a_n)$) to denote that the every element of matrix $\boldsymbol{A}_n$ is $O_p(a_n)$ (or $o_p(a_n)$).

Lemma A.1. *Suppose $\{\ell_n(\theta) : \theta \in \Theta\}$ is a sequence of convex function and can be written as $\frac{1}{2}\theta^\top F\theta + U_n^\top\theta + G_n + r_n(\theta)$, where F is symmetric and positive definite, U_n is stochastically bounded sequence of random vectors, G_n is arbitrary sequence, and $r_n(\theta)$ tends to zero in probability for each θ. Let θ_n be the argmin of $\ell_n(\theta)$, then θ_n is only $o_p(1)$ away from $\gamma_n = -F^{-1}U_n$, the argmin of $\frac{1}{2}\theta^\top F\theta + U_n^\top\theta + G_n$. If also $U_n \xrightarrow{\mathscr{L}} U$, then $\theta_n \xrightarrow{\mathscr{L}} -F^{-1}U$.*

Proof. This lemma comes from the result by Hjort and Pollard (2011).

Lemma A.2. *Let $(\boldsymbol{X}_1, Y_1), \cdots, (\boldsymbol{X}_n, Y_n)$ be independent and identically distributed random vectors, where $\boldsymbol{X}_i$ and Y_i are scalar random variables. Assume that $\mathrm{E}(|Y|^m) < \infty$ and that $\sup_{\boldsymbol{x}} \int |y|^m f(\boldsymbol{x}, y)\mathrm{d}y < \infty$, where f denote the joint density of $(\boldsymbol{X}, Y)$. Let $K(\cdot)$ be a bounded positive function with a bounded support and satisfying a Lipschitz condition. Then*

$$
\sup_{x\in\Omega}\left|\frac{1}{nh}\sum_{i=1}^n [K(h^{-1}(\boldsymbol{X}_i - \boldsymbol{x}))Y_i - \mathrm{E}K(h^{-1}(\boldsymbol{X}_i - \boldsymbol{x}))Y_i]\right| = O_p\left(\frac{\log^{1/2}(1/h)}{\sqrt{nh}}\right),
$$

provide that $n^{2\varepsilon-1}h \to \infty$ for some $\varepsilon < 1 - r^{-1}$.

Proof. This lemma comes from the result by Mack and Silverman (1982).

Lemma A.3. *Let $X_1, \ldots, X_n$ be arbitrary scalar random variables such that* $\max_{1\le i\le n} \mathrm{E}(|X_i|^r) <$

∞ *for some* $r \geq 1$. *Then, we have*

$$\mathrm{E}(\max_{1\leq i\leq n} |X_i|) \leq C_r n^{1/r},$$

where C_r *is a constant depending only on* r *and* $\max_{1\leq i\leq n} \mathrm{E}(|X_i|^r)$.

Proof. This lemma comes from lemma 2.2.2 Van Der Vaart and Wellner (1996).

Lemma A.4. *There exist positive constants* κ_1 *and* κ_2 *such that, except on an event whose probability tends to zero, all the eigenvalues of* $\sum_{i=1}^n \boldsymbol{V}_{i2}\boldsymbol{V}_{i2}^\top$ *fall between* κ_1 *and* κ_2.

Proof. Observe that $\sum_{i=1}^n \boldsymbol{V}_{i2}\boldsymbol{V}_{i2}^\top$ can be denoted by $\mathrm{diag}(\Psi_1, \ldots, \Psi_{N_n})$, where $\Psi_{N_n} = (\upsilon_{kij})_{(q+1)\times(q+1)}$, $\upsilon_{kij} = (N_n/n)\sum_{s=1}^n [(U_s - u_k)/h_0]^{i+j} I_{|U_s-u_k|\leq h_0}$, $i,j = 1,\ldots,q; k = 1,\ldots,N_n$. Let $\tilde{\Psi}_{N_n} = (\tilde{\upsilon}_{kij})_{(q+1)\times(q+1)}$, $\tilde{\upsilon}_{kij} = ((u_r - u_l)/2)\int_{|u|\leq 1} u^{i+j} k(u_k + h_0 u)\mathrm{d}u$. For any $\epsilon > 0$, there exist constant $c_5 > 0$, such that

$$\sum_{n=1}^{\infty}\sum_{k=1}^{N_n} P\{|\upsilon_{kij} - \tilde{\upsilon}_{kij}| > \epsilon\} \leq c_5 \sum_{n=1}^{\infty} (nN_n^4 + n^2 N_n^3)/(\epsilon^4 n^4) < \infty,$$

By Borel-Cantelli lemma, we have

$$\upsilon_{kij} - \tilde{\upsilon}_{kij} \to 0 \quad a.s. \quad i,j = 1,\ldots,q;\ k = 1,\ldots,N_n$$

Let $\hat{\Psi}_{N_n} = (\hat{\upsilon}_{ij})_{(q+1)\times(q+1)}$ with $\hat{\upsilon}_{ij} = \int_{|u|\leq 1} u^{i+j}\mathrm{d}u$. It is easy to prove that $\hat{\Psi}_{N_n}$ is positive definite. Thus, there exist positive constants κ_1 and κ_2 such that all the eigenvalues of $\sum_{i=1}^n \boldsymbol{V}_{i2}\boldsymbol{V}_{i2}^\top$ fall between κ_1 and κ_2.

Lemma A.5. *Under assumptions C4-C9, it holds that* $m^{1/2}(\log n)\max_i \|\boldsymbol{V}_i\| = o_p(1)$.

Proof. Note that

$$\|\boldsymbol{V}_{i1}\| = n^{-1/2}\sqrt{\sum_{j=1}^m \lambda_j^{-1}\hat{\xi}_{ij}^2} \leq n^{-1/2}\sqrt{2\sum_{j=1}^m \lambda_j^{-1}\xi_{ij}^2 + 2\sum_{j=1}^m \lambda_j^{-1}\langle Z_i, \hat{\phi}_j - \phi_j\rangle^2}. \tag{9}$$

Using Lemma A.3 and $\mathrm{E}(\lambda_j^{-1}\xi_{ij}^2) = 1$, we deduce that $\max_i \lambda_j^{-1}\xi_{ij}^2 = O_p(n^{1/2})$. By Lemma 5.1 of Hall and Horowitz (2007), we have

$$\langle Z_i, \hat{\phi}_j - \phi_j\rangle = \sum_{k\neq j}(\hat{\lambda}_j - \lambda_k)^{-1}\xi_{ik}\int \Delta\hat{\phi}_j\phi_k + \xi_{ij}\int(\hat{\phi}_j - \phi_j)\phi_j$$

where $\Delta = \hat{K}_Z - K_Z$ and $\int fg$ denotes $\int f(t)g(t)\mathrm{d}t$. Thus,

$$\langle Z_i, \hat{\phi}_j - \phi_j\rangle^2 \leq 2\left(\sum_{k\neq j}(\hat{\lambda}_j - \lambda_k)^{-2}\xi_{ik}^2\right)\left\|\int \Delta\hat{\phi}_j\right\|^2 + 2\xi_{ij}^2\left(\int(\hat{\phi}_j - \phi_j)\phi_j\right)^2 \tag{10}$$

Since $\sup_{j\ge 1}|\hat\lambda_j-\lambda_j|\le |||\Delta|||=O_p(n^{-1/2})$, we deduce that $|\hat\lambda_j-\lambda_k|\le 2|\lambda_j-\lambda_k|(1+o_p(1))$, where $o_p(1)$ uniformly for $1\le j\le m$ and $k\ne j$. Hence

$$\sum_{j=1}^m \lambda_j^{-1}\left(\sum_{k\ne j}(\hat\lambda_j-\lambda_k)^{-2}\xi_{ik}^2\right)\left\|\int\Delta\hat\phi_j\right\|^2\le 4|||\Delta|||^2\sum_{j=1}^m\lambda_j^{-1}\left(\sum_{k\ne j}(\lambda_j-\lambda_k)^{-2}\xi_{ik}^2\right)(1+o_p(1))$$

Since

$$\begin{aligned}|||\Delta|||^2\sum_{j=1}^m\lambda_j^{-1}\left(\sum_{k\ne j}(\lambda_j-\lambda_k)^{-2}\mathrm{E}\left(\max_i\xi_{ik}^2\right)\right) &\le Cn^{1/2}|||\Delta|||^2\sum_{j=1}^m\lambda_j^{-1}\left(\sum_{k\ne j}(\lambda_j-\lambda_k)^{-2}\lambda_k\right)\\ &= O(n^{-1/2}m^{2\beta_1+3}),\end{aligned}$$

we deduce that

$$\sum_{j=1}^m\lambda_j^{-1}\left(\sum_{k\ne j}(\hat\lambda_j-\lambda_k)^{-2}\max_i\xi_{ik}^2\right)\left\|\int\Delta\hat\phi_j\right\|^2=O(n^{-1/2}m^{2\beta_1+3}). \tag{11}$$

Using (5.27) of Cai and Hall (2006), we have $\int(\hat\phi_j-\phi_j)\phi_j=O_p(n^{-1}j^2)$ uniformly for $1\le j\le m$. By Lemma A.3 and assumptions C5 and C6, we have

$$\sum_{j=1}^m\lambda_j^{-1}\max_i\xi_{ij}^2\left(\int(\hat\phi_j-\phi_j)\phi_j\right)^2=O_p\left(n^{1/2}n^{-2}\sum_{j=1}^m j^4\right)=O_p(n^{-3/2}m^5). \tag{12}$$

Combining (10)-(12), we can get

$$\max_i\sum_{j=1}^m\lambda_j^{-1}\langle Z_i,\hat\phi_j-\phi_j\rangle^2=O(n^{-1/2}m^{2\beta_1+3}+n^{-3/2}m^5).$$

Combining (9) and (12), we obtain that

$$\max_i\|\boldsymbol{V}_{i1}\|=O_p\left(n^{-1/2}(n^{1/4}m^{1/2}+n^{-1/4}m^{\beta_1+3/2}+n^{-3/4}m^{5/2})\right). \tag{13}$$

Since $\|\boldsymbol{B}(U_i)\|$ and $\|\boldsymbol{M}(U_i)\|$ are bounded, we have $\max_i\|\boldsymbol{V}_{i2}\|=O_p\left(N_n^{1/2}n^{-1/2}\right)$. By assumption C9 and Lemma A.3, we have $\max_i\|\boldsymbol{X}_i\|=O_p(n^{1/4})$. Thus, $\max_i\|\boldsymbol{M}(U_i)\boldsymbol{X}_i\|=O_p(n^{1/4})$. Since $\boldsymbol{P}_n/n\to_P \mathrm{E}\{\boldsymbol{M}(U)\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{M}(U)^\top\}$, we have $\max_i\|\boldsymbol{V}_{i3}\|=P_n^{-1/2}\max_i\|\boldsymbol{M}(U_i)\boldsymbol{X}_i\|=O_p\left(n^{-1/4}\right)$. Hence, by Assumption C8, we obtain

$$\begin{aligned}m^{1/2}(\log n)\max_i\|\boldsymbol{V}_i\| &\le m^{1/2}(\log n)\left(\max_i\|\boldsymbol{V}_{i1}\|+\max_i\|\boldsymbol{V}_{i2}\|+\max_i\|\boldsymbol{V}_{i3}\|\right)\\ =\ & O_p\left(\log n\left(n^{-1/4}m+n^{-3/4}m^{\beta_1+2}+n^{-5/4}m^3+N_n^{1/2}n^{-1/2}m^{1/2}+n^{-1/4}m^{1/2}\right)\right)=o_p(1).\end{aligned}$$

Lemma A.6. *Under assumptions C4-C9, it holds that* $\max_i |W_i| = o_p(1)$.

Proof. Using Lemma A.3 and $\mathrm{E}(\lambda_j^{-1}\xi_{ij}^2) = 1 < \infty$, we deduce that $\max_i |\xi_{ij}| = O_p(\lambda_j^{1/2} n^{1/4})$. By Assumptions C5, C6 and C8, we have

$$\max_i | \sum_{j=m+1}^{\infty} b_{\tau j}\xi_{ij}| \leq \sum_{j=m+1}^{\infty} b_{\tau j} \max_i |\xi_{ij}| = O_p(\sum_{j=m+1}^{\infty} j^{-\beta_2} j^{-\beta_1/2} n^{1/4}) = o_p(1).$$

Using Lemma A.3 and assumption C4, we have $\max_i \|Z_i\| = O_p(n^{1/4})$. Using (5.21) and (5.22) of Cai and Hall (2006), we have $\|\hat{\phi}_j - \phi_j\|^2 = O_p(n^{-1}j^2)$ uniformly for $1 \leq j \leq m$. By Assumptions C6 and C8, we have

$$\begin{aligned}
\max_i | \sum_{j=1}^{m} b_{\tau j}(\xi_{ij} - \hat{\xi}_{ij})| &\leq \sum_{j=1}^{m} b_{\tau j} \max_i \langle Z_i, \hat{\phi}_j - \phi_j \rangle \leq \sum_{j=1}^{m} b_{\tau j} \max_i \|Z_i\| \|\hat{\phi}_j - \phi_j\| \\
&= O_p\left(n^{-1/4} \sum_{j=1}^{m} j^{1-\beta_2} \right) = o_p(1).
\end{aligned}$$

By Assumptions C7, it holds that $\max_i |\alpha_{0\tau}(U_i) - \boldsymbol{B}(U_i)^\top \boldsymbol{F}_{0\tau}| = O(h_0^q) = O(N_n^{-q}) = o_p(1)$, $\max_i |\boldsymbol{X}_i^\top \left(\boldsymbol{\alpha}_\tau(U_i) - \boldsymbol{M}(U_i)^\top \boldsymbol{F}_\tau \right)| \leq \max_i \|\boldsymbol{X}_i\| \max_i \|\boldsymbol{\alpha}_\tau(U_i) - \boldsymbol{M}(U_i)^\top \boldsymbol{F}_\tau\| = O(h_0^q n^{1/4}) = o_p(1)$. Thus, we have $\max_i |W_i| = \max_i |\alpha_{0\tau}(U_i) - \boldsymbol{B}(U_i)^\top \boldsymbol{F}_{0\tau}| + \max_i |\boldsymbol{X}_i^\top \left(\boldsymbol{\alpha}_\tau(U_i) - \boldsymbol{M}(U_i)^\top \boldsymbol{F}_\tau \right)| + \max_i | \sum_{j=1}^{m} b_{\tau j}(\xi_{ij} - \hat{\xi}_{ij})| + \max_i | \sum_{j=m+1}^{\infty} b_{\tau j}\xi_{ij}| = o_p(1)$.

Lemma A.7. *Under assumptions C4-C9, for any sufficient large L, it holds that*

$$\sup_{\|\boldsymbol{\theta}\| \leq L} |R_n(m^{1/2}\boldsymbol{\theta})| = o_p(1).$$

Proof. Note that

$$\begin{aligned}
R_{n,i}(\boldsymbol{\theta}) &= S_{n,i}(\boldsymbol{\theta}) - \Gamma_{n,i}(\boldsymbol{\theta}) + \boldsymbol{V}_i^\top \boldsymbol{\theta} \psi_\tau(\varepsilon_{\tau i}) = \int_{W_i}^{W_i - m^{1/2}\boldsymbol{V}_i^\top \boldsymbol{\theta}} [\psi_\tau(\varepsilon_{\tau i} + t) - \psi_\tau(\varepsilon_{\tau i})] \,\mathrm{d}t \\
&- \mathrm{E}\left\{ \int_{W_i}^{W_i - m^{1/2}\boldsymbol{V}_i^\top \boldsymbol{\theta}} [\psi_\tau(\varepsilon_{\tau i} + t) - \psi_\tau(\varepsilon_{\tau i})] \,\mathrm{d}t | \mathcal{I} \right\}
\end{aligned}$$

Let $M_n = \sup_{\|\boldsymbol{\theta}\| \leq L} |R_n(m^{1/2}\boldsymbol{\theta})|$. Using Lemma A.5, we deduce that

$$(\log n) M_n \leq 4Lm^{1/2} (\log n) \max_i \|\boldsymbol{V}_i\| = o_p(1).$$

Using Lemma A.5 and Lemma A.6, we have $\max_i \sup_{\|\boldsymbol{\theta}\| \leq L} (|W_i| + m^{1/2} |\boldsymbol{V}_i^\top \boldsymbol{\theta}|) = o_p(1)$. Then, we deduce that

$$\begin{aligned}
\sum_{i=1}^{n} \mathrm{V}ar(R_{n,i}(\boldsymbol{\theta})|\mathcal{I}) &\leq \sum_{i=1}^{n} \mathrm{E}\left(\left\{\int_{W_i}^{W_i - m^{1/2}\boldsymbol{V}_i^\top\boldsymbol{\theta}} [\psi_\tau(\varepsilon_{\tau i}+t)-\psi_\tau(\varepsilon_{\tau i})]\,\mathrm{d}t\right\}^2 |\mathcal{I}\right) \\
&\leq \sum_{i=1}^{n} m^{1/2}|\boldsymbol{V}_i^\top\boldsymbol{\theta}| \int_{W_i - m^{1/2}|\boldsymbol{V}_i^\top\boldsymbol{\theta}|}^{W_i + m^{1/2}|\boldsymbol{V}_i^\top\boldsymbol{\theta}|} \mathrm{E}\left([\psi_\tau(\varepsilon_{\tau i}+t)-\psi_\tau(\varepsilon_{\tau i})]^2 |\mathcal{I}\right)\mathrm{d}t \\
&\leq \sum_{i=1}^{n} m^{1/2}|\boldsymbol{V}_i^\top\boldsymbol{\theta}| \int_{W_i - m^{1/2}|\boldsymbol{V}_i^\top\boldsymbol{\theta}|}^{W_i + m^{1/2}|\boldsymbol{V}_i^\top\boldsymbol{\theta}|} \mathrm{E}\left(I(-t<\varepsilon_{\tau i}<t)|\mathcal{I}\right)\mathrm{d}t \\
&\leq 2m^{1/2}\max_i|\boldsymbol{V}_i^\top\boldsymbol{\theta}| \sum_{i=1}^{n} f_\tau(0|\boldsymbol{X}_i, U_i, Z_i(t))[W_i^2 + m(\boldsymbol{V}_i^\top\boldsymbol{\theta})^2][1+o_p(1)].
\end{aligned}$$

Since $\sup_{j\geq 1}|\hat{\lambda}_j - \lambda_j| \leq |||\Delta||| = O_p(n^{-1/2})$, we deduce that

$$\frac{1}{2}\lambda_j[(1+o_p(1)] \leq \hat{\lambda}_j \leq \frac{3}{2}\lambda_j[(1+o_p(1)], \quad j=1,\ldots,m. \tag{14}$$

By Lemma A.4, we deduce that

$$\sum_{i=1}^{n}(\boldsymbol{V}_{i2}^\top\boldsymbol{\theta}_2)^2 \leq \kappa_2\sum_{j=1}^{m}\theta_{2j}^2,$$

where $\boldsymbol{\theta} = (\theta_{1,1},\ldots,\theta_{1,m},\theta_{2,1},\ldots,\theta_{2,2qN_n},\theta_{3,1},\ldots,\theta_{3,2pqN_n})^\top$. By assumption C1, there exist constant C_1 such that

$$\begin{aligned}
\sum_{i=1}^{n} f_\tau(0|\boldsymbol{X}_i, U_i, Z_i(t))(\boldsymbol{V}_i^\top\boldsymbol{\theta})^2 &\leq 3c_1\sum_{i=1}^{n}(\boldsymbol{V}_{i1}^\top\boldsymbol{\theta}_1)^2 + 3c_1\sum_{i=1}^{n}(\boldsymbol{V}_{i2}^\top\boldsymbol{\theta}_2)^2 + 3c_1\sum_{i=1}^{n}(\boldsymbol{V}_{i3}^\top\boldsymbol{\theta}_3)^2 \\
&\leq 3c_1\sum_{j=1}^{m}\lambda_j^{-1}\hat{\lambda}_j\theta_{1j}^2 + 3c_1\kappa_2\sum_{j=1}^{m}\theta_{2j}^2 + 3c_1\sum_{j=1}^{m}\theta_{3j}^2 \\
&\leq C_1\|\boldsymbol{\theta}\|^2.
\end{aligned}$$

By Assumptions C5, C6 and C8, we have

$$\mathrm{E}\sum_{i=1}^{n}\left(\sum_{j=m+1}^{\infty} b_{\tau j}\xi_{ij}\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} b_{\tau j}^2\lambda_j = O(nm^{1-\beta_1-2\beta_2})$$

and

$$\begin{aligned}
\mathrm{E}\sum_{i=1}^{n}\left(\sum_{j=1}^{m} b_{\tau j}(\xi_{ij}-\hat{\xi}_{ij})\right)^2 &\leq \sum_{i=1}^{n} m\sum_{j=1}^{m} b_{\tau j}^2(\xi_{ij}-\hat{\xi}_{ij})^2 \leq \sum_{i=1}^{n} m\sum_{j=1}^{m} b_{\tau j}^2\max_i\|Z_i\|^2\|\hat{\phi}_j - \phi_j\|^2 \\
&= O(m).
\end{aligned}$$

By Assumptions C7, it holds that $\sum_{i=1}^{n}(\alpha_{0\tau}(U_i)-\boldsymbol{B}(U_i)^\top\boldsymbol{F}_{0\tau})^2 = O(nh_0^{2q}) = O(nN_n^{-2q}) = O_p(m)$, $\sum_{i=1}^{n}[\boldsymbol{X}_i^\top\left(\boldsymbol{\alpha}_\tau(U_i)-\boldsymbol{M}(U_i)^\top\boldsymbol{F}_\tau\right)]^2 \le \max_i\|\boldsymbol{X}_i\|^2\max_i\|\boldsymbol{\alpha}_\tau(U_i)-\boldsymbol{M}(U_i)^\top\boldsymbol{F}_\tau\|^2 = O(N_n^{-2q}n^{3/2}) = O_p(m)$. Thus, we have $\sum_{i=1}^{n}W_i^2 = O_p(m)$. Let

$$D_n = \sum_{i=1}^{n}\sup_{\|\boldsymbol{\theta}\|\le L}\mathrm{V}ar(R_{n,i}(\boldsymbol{\theta})|\mathcal{I}).$$

There exist a constant C_2 such that

$$D_n \le C_2 m^{3/2}\max_i|\boldsymbol{V}_i^\top\boldsymbol{\theta}|[1+o_p(1)] \le C_2 L m^{3/2}\max_i\|\boldsymbol{V}_i\|[1+o_p(1)].$$

Let $|\boldsymbol{c}| = \max_{1\le im}|c_i|$ for a vector $\boldsymbol{c} = (c_1,\ldots,c_m)^\top$. Set $G = \{\boldsymbol{\theta},\|\boldsymbol{\theta}\|\le L\}$. Let G be divided into J_n disjoint parts $G_1,\ldots,G_{J_n}$ such that for any $\boldsymbol{g}_k\in G_k$, $1\le k\le J_n$ and any sufficient small $\epsilon>0$, except on an event whose probability tends to zero,

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in G_k}|R_n(m^{1/2}\boldsymbol{\theta})-R_n(m^{1/2}\boldsymbol{g}_k)| &\le 4c_0^{-1}\sup_{\boldsymbol{\theta}\in G_k}\sum_{i=1}^{n}f_\tau(0|\boldsymbol{X}_i,U_i,Z_i(t))m^{1/2}|\boldsymbol{V}_i^\top(\boldsymbol{\theta}-\boldsymbol{g}_k)| \\ &\le C_3\sup_{\boldsymbol{\theta}\in G_k}\sum_{i=1}^{n}m^{1/2}|\boldsymbol{V}_i^\top(\boldsymbol{\theta}-\boldsymbol{g}_k)| \le C_3\sup_{\boldsymbol{\theta}\in G_k}m^{1/2}n^{1/2}\|\boldsymbol{\theta}-\boldsymbol{g}_k\| \\ &\le C_3\sup_{\boldsymbol{\theta}\in G_k}m^{1/2}n^{1/2}|\boldsymbol{\theta}-\boldsymbol{g}_k| < \epsilon/2.\end{aligned}$$

where C_3 is a constant. This can be done with $J_n = (4C_3Ln^{1/2}m/\epsilon)^{m+(\tilde{q}+1)(p+1)N_n}$. Using Bernstein inequality, we have

$$\begin{aligned}P\left(\sup_{\boldsymbol{\theta}\in G_k}m^{-1}|R_n(m^{1/2}\boldsymbol{\theta})|\ge\epsilon|\mathcal{I}\right) &\le \sum_{k=1}^{J_n}P\left(|R_n(m^{1/2}\boldsymbol{g}_k)|\ge m\epsilon/2|\mathcal{I}\right) \\ &\le 2J_n\exp\left(-\epsilon^2m^2/(8D_n+4m\epsilon N_n)\right) = o_p(1).\end{aligned}$$

Therefore,

$$P\left(\sup_{\boldsymbol{\theta}\in G_k}m^{-1}|R_n(m^{1/2}\boldsymbol{\theta})|\ge\epsilon\right) = o_p(1).$$

We complete the proof of Lemma A.7.

Lemma A.8. *Suppose that assumptions C1-C11 hold, we have* $\|\hat{\boldsymbol{\theta}}\| = O_p(m^{1/2})$.

Proof. Note that

$$
\begin{aligned}
& \Gamma_n(m^{1/2}\boldsymbol{\theta}) \\
= \; & \sum_{i=1}^{n} \int_{W_i}^{W_i - m^{1/2}\boldsymbol{V}_i^{\top}\boldsymbol{\theta}} \mathrm{E}\left[\psi_\tau(\varepsilon_{\tau i}+t)|\mathcal{I}\right] \mathrm{d}t \\
= \; & \frac{1}{2}\sum_{i=1}^{n} f_\tau(0|\boldsymbol{X}_i, U_i, Z_i(t))[(W_i - m^{1/2}\boldsymbol{V}_i^{\top}\boldsymbol{\theta})^2 - W_i^2][1+o_p(1)] \\
\geq \; & c_0 m \left(\frac{1}{4}\|\boldsymbol{\theta}\|^2 + \sum_{i=1}^{n}\left(\boldsymbol{V}_{i1}^{\top}\boldsymbol{\theta}_1\boldsymbol{V}_{i2}^{\top}\boldsymbol{\theta}_2 + \boldsymbol{V}_{i1}^{\top}\boldsymbol{\theta}_1\boldsymbol{V}_{i3}^{\top}\boldsymbol{\theta}_3 + \boldsymbol{V}_{i2}^{\top}\boldsymbol{\theta}_2\boldsymbol{V}_{i3}^{\top}\boldsymbol{\theta}_3\right) - m^{-1}\sum_{i=1}^{n} W_i^2 \right)[1+o_p(1)].
\end{aligned}
$$

Set $\tilde{\boldsymbol{V}}_{i1} = n^{-1/2}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{A}_i$. By Assumptions 1 and 5 and the fact that $\|\boldsymbol{B}(U_i)\|$ is bounded, there exist constant $C_4 > 0$ such that

$$
\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{n}\tilde{\boldsymbol{V}}_{i1}^{\top}\boldsymbol{\theta}_1\boldsymbol{V}_{i2}^{\top}\boldsymbol{\theta}_2\right)^2 &= \mathrm{E}\sum_{i=1}^{n}\left(\tilde{\boldsymbol{V}}_{i1}^{\top}\boldsymbol{\theta}_1\boldsymbol{V}_{i2}^{\top}\boldsymbol{\theta}_2\right)^2 + 2\mathrm{E}\sum_{i\neq j}\left(\mathrm{E}(\tilde{\boldsymbol{V}}_{i1}|U,\boldsymbol{X})^{\top}\boldsymbol{\theta}_1\boldsymbol{V}_{j2}^{\top}\boldsymbol{\theta}_2\right) \\
&= \sum_{i=1}^{n}\mathrm{E}\left(\tilde{\boldsymbol{V}}_{i1}^{\top}\boldsymbol{\theta}_1\boldsymbol{V}_{i2}^{\top}\boldsymbol{\theta}_2\right)^2 \leq N_n\|\boldsymbol{\theta}_1\|^2\|\boldsymbol{\theta}_2\|^2\mathrm{E}\left(\|\tilde{\boldsymbol{V}}_{i1}\|^2\|\boldsymbol{B}(U_i)\|^2\right) \\
&\leq C_3 n^{-1} m N_n = o(1), \\
\mathrm{E}\left(\sum_{i=1}^{n}\tilde{\boldsymbol{V}}_{i1}^{\top}\boldsymbol{\theta}_1\boldsymbol{V}_{i3}^{\top}\boldsymbol{\theta}_3\right)^2 &= \mathrm{E}\sum_{i=1}^{n}\left(\tilde{\boldsymbol{V}}_{i1}^{\top}\boldsymbol{\theta}_1\boldsymbol{V}_{i3}^{\top}\boldsymbol{\theta}_3\right)^2 + 2\sum_{i\neq j}\mathrm{E}\left(\mathrm{E}(\tilde{\boldsymbol{V}}_{i1}|U,\boldsymbol{X})^{\top}\boldsymbol{\theta}_1\boldsymbol{V}_{j3}^{\top}\boldsymbol{\theta}_3\right) \\
&= \sum_{i=1}^{n}\mathrm{E}\left(\tilde{\boldsymbol{V}}_{i1}^{\top}\boldsymbol{\theta}_1\boldsymbol{V}_{i3}^{\top}\boldsymbol{\theta}_3\right)^2 \leq \sum_{i=1}^{n}\|\boldsymbol{\theta}_1\|^2\|\boldsymbol{\theta}_2\|^2\mathrm{E}\left(\|\tilde{\boldsymbol{V}}_{i1}\|^2\|\boldsymbol{V}_{i3}\|^2\right) \\
&\leq C_3 n^{-1/2} m = o(1), \\
\mathrm{E}\left(\sum_{i=1}^{n}\boldsymbol{V}_{i2}^{\top}\boldsymbol{\theta}_2\boldsymbol{V}_{i3}^{\top}\boldsymbol{\theta}_3\right)^2 &= \mathrm{E}\sum_{i=1}^{n}\left(\boldsymbol{V}_{i2}^{\top}\boldsymbol{\theta}_2\boldsymbol{V}_{i3}^{\top}\boldsymbol{\theta}_3\right)^2 + 2\sum_{i\neq j}\mathrm{E}\left(\tilde{\boldsymbol{V}}_{i2}^{\top}\boldsymbol{\theta}_2\mathrm{E}(\boldsymbol{V}_{i3}|U)^{\top}\boldsymbol{\theta}_3\right) \\
&= \sum_{i=1}^{n}\mathrm{E}\left(\boldsymbol{V}_{i2}^{\top}\boldsymbol{\theta}_2\boldsymbol{V}_{i3}^{\top}\boldsymbol{\theta}_3\right)^2 \leq \sum_{i=1}^{n}\|\boldsymbol{\theta}_2\|^2\|\boldsymbol{\theta}_3\|^2\mathrm{E}\left(\|\boldsymbol{V}_{i2}\|^2\|\boldsymbol{V}_{i3}\|^2\right) \\
&\leq C_3 n^{-1/2} N_n = o(1).
\end{aligned}
$$

Using the fact that $\|\hat{\phi}_j - \phi_j\|^2 = O_p(n^{-1}j^2)$ uniformly for $1 \le j \le m$, we deduce that

$$
\begin{aligned}
& |\sum_{i=1}^n (\boldsymbol{V}_{i1} - \tilde{\boldsymbol{V}}_{i1})^\top \boldsymbol{\theta}_1 \boldsymbol{V}_{i2}^\top \boldsymbol{\theta}_2| \\
\le\ & \left(\sum_{i=1}^n \|(\boldsymbol{V}_{i1} - \tilde{\boldsymbol{V}}_{i1})\|^2\right)^{1/2} \|\boldsymbol{\theta}_1\| \left(\sum_{i=1}^n (\boldsymbol{V}_{i2}^\top \boldsymbol{\theta}_2)^2\right)^{1/2} \\
\le\ & C_4 \left(n^{-1} m \sum_{i=1}^n \sum_{j=1}^m \lambda_j^{-1} (\xi_{ij} - \hat{\xi}_{ij})^2\right)^{1/2} \|\boldsymbol{\theta}_1\| N_n^{1/2} \max_i \|\boldsymbol{B}(U_i)\| \|\boldsymbol{\theta}_2\| \\
\le\ & C_4 \left(n^{-1} m \sum_{i=1}^n \sum_{j=1}^m \lambda_j^{-1} \max_i \|\boldsymbol{X}_i\|^2 \|\hat{\phi}_j - \phi_j\|^2\right)^{1/2} \|\boldsymbol{\theta}_1\| N_n^{1/2} \max_i \|\boldsymbol{B}(U_i)\| \|\boldsymbol{\theta}_2\| \\
=\ & O_p\left(n^{-1/2} m^{4+\beta_1} N_n^{1/2}\right) = o_p(1),
\end{aligned}
$$

where C_4 is a constant. Similarly,

$$
|\sum_{i=1}^n (\boldsymbol{V}_{i1} - \tilde{\boldsymbol{V}}_{i1})^\top \boldsymbol{\theta}_1 \boldsymbol{V}_{i3}^\top \boldsymbol{\theta}_3| = O_p\left(n^{-1/2} m^{4+\beta_1} n^{1/4}\right) = o_p(1),
$$

Thus, $\sum_{i=1}^n \boldsymbol{V}_{i1}^\top \boldsymbol{\theta}_1 \boldsymbol{V}_{i2}^\top \boldsymbol{\theta}_2 = o_p(1)$, $\sum_{i=1}^n \boldsymbol{V}_{i1}^\top \boldsymbol{\theta}_1 \boldsymbol{V}_{i3}^\top \boldsymbol{\theta}_3 = o_p(1)$. Observe that $\sum_{i=1}^n \boldsymbol{V}_{i2}^\top \boldsymbol{\theta}_2 \boldsymbol{V}_{i3}^\top \boldsymbol{\theta}_3 = n^{-1/2} N_n^{1/2} \boldsymbol{P}_n^{-1/2} \sum_{i=1}^n \boldsymbol{B}(U_i)^\top \boldsymbol{\theta}_2 \boldsymbol{X}_i^\top \boldsymbol{M}(U_i)^\top \boldsymbol{\theta}_3$ and $\mathrm{E}\left(\sum_{i=1}^n \boldsymbol{B}(U_i)^\top \boldsymbol{\theta}_2 \boldsymbol{X}_i^\top \boldsymbol{M}(U_i)^\top \boldsymbol{\theta}_3\right)^2 = O(n)$. Thus, $\sum_{i=1}^n \boldsymbol{V}_{i2}^\top \boldsymbol{\theta}_2 \boldsymbol{V}_{i3}^\top \boldsymbol{\theta}_3 = O_p(n^{-1/2} N_n^{1/2}) = o_p(1)$. Hence, for sufficient large L, we have

$$
\inf_{\|\boldsymbol{\theta}\| = L} \Gamma_n(m^{1/2}\boldsymbol{\theta}) \ge \frac{1}{4} c_0 m L^2 [1 + o_p(1)].
$$

Note that

$$
\begin{aligned}
\mathrm{E}\left\{\left(m^{1/2} \sum_{i=1}^n \tilde{\boldsymbol{V}}_i^\top \boldsymbol{\theta} \psi(\varepsilon_{\tau i})\right)^2 |\mathcal{I}\right\} &\le 3m \left\{\sum_{i=1}^n (\boldsymbol{V}_{i1}^\top \boldsymbol{\theta}_1)^2 + \sum_{i=1}^n (\boldsymbol{V}_{i2}^\top \boldsymbol{\theta}_2)^2 + \sum_{i=1}^n (\boldsymbol{V}_{i3}^\top \boldsymbol{\theta}_3)^2\right\} \\
&\le C_5 m \|\boldsymbol{\theta}\|^2 [1 + o_p(1)].
\end{aligned}
$$

where C_5 is a constant. Thus, $\sup_{\|\boldsymbol{\theta}\| \le L} |m^{1/2} \sum_{i=1}^n \tilde{\boldsymbol{V}}_i^\top \boldsymbol{\theta} \psi(\varepsilon_{\tau i})| = O_p(m^{1/2})$. For sufficient large L, we deduce that

$$
\inf_{\|\boldsymbol{\theta}\| = L} S_n(m^{1/2}\boldsymbol{\theta}) = \inf_{\|\boldsymbol{\theta}\| = L} \left(\Gamma_n(\boldsymbol{\theta}) - m^{1/2} \sum_{i=1}^n \boldsymbol{V}_i^\top \boldsymbol{\theta} \psi_\tau(\varepsilon_{\tau i}) + R_n(\boldsymbol{\theta})\right) \ge \frac{1}{4} c_0 m L^2 [1 + o_p(1)]. \quad (15)
$$

Note that the minimizer of Equation (5) is also the minimizer of the following function

$$
\sum_{i=1}^n \left\{\rho_\tau\left(\varepsilon_{\tau i} + \boldsymbol{V}_i^\top \boldsymbol{\theta} + W_i\right) - \rho_\tau\left(\varepsilon_{\tau i} + W_i\right)\right\}. \quad (16)
$$

By the convexity of ρ_τ and (15), we deduce that

$$P\left\{\inf_{\|\boldsymbol{\theta}\|\le L}\left(\sum_{i=1}^{n}\left[\rho_\tau\left(\varepsilon_{\tau i}+\boldsymbol{V}_i^\top\boldsymbol{\theta}+W_i\right)-\rho_\tau\left(\varepsilon_{\tau i}+W_i\right)\right]\right)>0\right\}\to 1.$$

Thus, $P(\|\hat{\boldsymbol{\theta}}\|\le Lm^{1/2})\to 1$, that is, $\|\hat{\boldsymbol{\theta}}\|=O_p(m^{1/2})$. We complete the proof of Lemma A.8.

Proof of Theorem 1. Let

$$T_1=\int_I\left\{\sum_{k=1}^{m}(\hat{b}'_{\tau k}-b_{\tau k})\hat{\phi}_k(t)\right\}^2\mathrm{d}t,\qquad T_2=\int_I\left\{\sum_{k=1}^{m}b_{\tau k}\left(\hat{\phi}_k(t)-\phi_k(t)\right)\right\}^2\mathrm{d}t$$

and

$$T_3=\int_I\left(\sum_{k=m+1}^{\infty}b_{\tau k}\phi_k(t)\right)^2\mathrm{d}t.$$

We can get

$$\begin{aligned}\int_I\left(\hat{\beta}_\tau(t)-\beta_\tau(t)\right)^2\mathrm{d}t &= \int_I\left(\sum_{k=1}^{m}\hat{b}'_{\tau k}\hat{\phi}_k(t)-\sum_{k=1}^{\infty}b_{\tau k}\phi_k(t)\right)^2\mathrm{d}t\\ &= \int_I\left(\sum_{k=1}^{m}\hat{b}'_{\tau k}\hat{\phi}_k(t)-\sum_{k=1}^{m}b_{\tau k}\phi_k(t)-\sum_{k=m+1}^{\infty}b_{\tau k}\phi_k(t)\right)^2\mathrm{d}t\\ &\le 4T_1+4T_2+2T_3. \qquad (17)\end{aligned}$$

Furthermore, using the fact $\int_I\{\phi_k(t)-\hat{\phi}_k(t)\}^2\mathrm{d}t=O_p(n^{-1}k^2)$ uniformly for $k=1,\cdots,m$, we have

$$\begin{aligned}T_2 &\le m\int_I\sum_{k=1}^{m}b_{\tau k}^2\left(\hat{\phi}_k(t)-\phi_k(t)\right)^2\mathrm{d}t=O_p\left(m\sum_{k=1}^{m}b_{\tau k}^2n^{-1}k^2\right)\le O_p\left(n^{-1}m\sum_{k=1}^{m}k^{2-2\beta_2}\right)\\ &= O_p\left(n^{-\frac{\beta_1+2\beta_2-1}{\beta_1+2\beta_2}}\right), \qquad (18)\end{aligned}$$

$$T_3=\sum_{k=m+1}^{\infty}b_{\tau k}^2\le\sum_{k=m+1}^{\infty}B_3^{-2}k^{-2\beta_2}=O_p\left(n^{-\frac{2\beta_2-1}{\beta_1+2\beta_2}}\right). \qquad (19)$$

By Lemma A.8, we obtain $\|\hat{\boldsymbol{\theta}}_1\|=O_p(m^{1/2})$. Thus,

$$\begin{aligned}T_1=\int_I\left\{\sum_{k=1}^{m}(\hat{b}'_{\tau k}-b_{\tau k})\hat{\phi}_k(t)\right\}^2\mathrm{d}t &= \sum_{k=1}^{m}(\hat{b}'_{\tau k}-b_{\tau k})^2\le n^{-1}\lambda_m^{-1}\|n^{1/2}\boldsymbol{\Lambda}^{1/2}(\hat{\boldsymbol{\beta}}'_\tau-\boldsymbol{\beta}_\tau)\|^2\\ &= O_p\left(n^{-1}m^{1+\beta_1}\right)=O_p\left(n^{-\frac{2\beta_2-1}{\beta_1+2\beta_2}}\right). \qquad (20)\end{aligned}$$

Conjoining (17)-(20), we deduce that

$$\begin{aligned}\int_I \left(\hat{\beta}_\tau(t)-\beta_\tau(t)\right)^2 \mathrm{d}t &= O_p\left(n^{-\frac{2\beta_2-1}{\beta_1+2\beta_2}}\right)+O_p\left(n^{-\frac{\beta_1+2\beta_2-1}{\beta_1+2\beta_2}}\right)+O_p\left(n^{-\frac{2\beta_2-1}{\beta_1+2\beta_2}}\right)\\ &= O_p\left(n^{-\frac{2\beta_2-1}{\beta_1+2\beta_2}}\right).\end{aligned}$$

This completes the proof of Theorem 1.

Proof of Theorem 2. Let $r_i = I(\varepsilon_{i\tau}\le 0)-\tau$ and $r_i^*(u) = I(\varepsilon_{i\tau}\le -s_i(u)-\varphi_i)-\tau$, where $s_i(u) = \alpha_{0\tau}(U_i)-\alpha_{0\tau}(u)-\alpha_{0\tau}'(u)(U_i-u)+\boldsymbol{X}_i^T\{\boldsymbol{\alpha}_\tau(U_i)-\boldsymbol{\alpha}_\tau(u)-\boldsymbol{\alpha}_\tau'(u)(U_i-u)\}$ and $\varphi_i=\sum_{k=m+1}^{\infty} b_{\tau k}\xi_{ik}+\sum_{k=1}^{m} b_{\tau k}(\xi_{ik}-\hat{\xi}_{ik})+\sum_{k=1}^{m}(b_{\tau k}-\hat{b}_{\tau k}')\hat{\xi}_{ik}$. Denote $K_i(u)=K\{(U_i-u)/h\}$, $\boldsymbol{\theta}^*=\sqrt{nh}\{a_{0\tau}-\alpha_{0\tau}(u),\{\boldsymbol{a}_\tau-\boldsymbol{\alpha}_\tau(u)\}^T,h\{b_{0\tau}-\alpha_{0\tau}'(u)\},h\{\boldsymbol{b}_\tau-\boldsymbol{\alpha}_\tau'(u)\}^T\}^T$, and $\boldsymbol{X}_i^*(u)=\{1,\boldsymbol{X}_i^T,\frac{U_i-u}{h},\boldsymbol{X}_i^T\frac{U_i-u}{h}\}^T$. Seen from (2), we deduce that

$$\begin{aligned}& Y_i-a_{0\tau}-b_{0\tau}(U_i-u)-\boldsymbol{X}_i^T\{\boldsymbol{a}_\tau+\boldsymbol{b}_\tau(U_i-u)\}-\sum_{k=1}^{m}\hat{b}_{\tau k}'\hat{\xi}_{ik}\\ =\ & \alpha_{0\tau}(U_i)+\boldsymbol{X}_i^T\boldsymbol{\alpha}_\tau(U_i)+\varphi_i+\varepsilon_{\tau i}-a_{0\tau}-b_{0\tau}(U_i-u)-\boldsymbol{X}_i^T\{\boldsymbol{a}_\tau+\boldsymbol{b}_\tau(U_i-u)\}\\ =\ & \varphi_i+\varepsilon_{\tau i}+s_i(u)-\eta_i(u).\end{aligned}$$

where $\eta_i(u)=\{\boldsymbol{X}_i^*(u)\}^T\boldsymbol{\theta}^*/\sqrt{nh}$. Then, $\boldsymbol{\theta}^*$ is also the minimizer of function

$$L_n(\boldsymbol{\theta}^*)=\sum_{i=1}^{n}\left(\rho_\tau\{\varepsilon_{\tau i}+\varphi_i+s_i(u)-\eta_i(u)\}-\rho_\tau\{\varepsilon_{i\tau}+\varphi_i+s_i(u)\}\right)K_i(u).$$

By the identity of Knight (1998)

$$\rho_\tau(u-v)-\rho_\tau(u)=v\{I(u\le 0)-\tau\}+\int_0^v\{I(u\le t)-I(u\le 0)\}\,\mathrm{d}t, \tag{21}$$

we can get

$$\begin{aligned}L_n(\boldsymbol{\theta}^*) &= \sum_{i=1}^{n}\Big\{\eta_i(u)\left[I(\varepsilon_{i\tau}\le -s_i(u)-\varphi_i)-\tau\right]+\int_0^{\eta_i(u)}\{I(\varepsilon_{i\tau}\le -s_i(u)-\varphi_i+t)\\ &- I(\varepsilon_{i\tau}\le -s_i(u)-\varphi_i)\}\mathrm{d}t\Big\}K_i(u)\\ &= \sum_{i=1}^{n}\left\{\eta_i(u)r_i^*(u)+\int_0^{\eta_i(u)}\{I(\varepsilon_{i\tau}\le -s_i(u)-\varphi_i+t)-I(\varepsilon_{i\tau}\le -s_i(u)-\varphi_i)\}\mathrm{d}t\right\}K_i(u)\\ &= \{\boldsymbol{S}_n^*(u)\}^T\boldsymbol{\theta}^*+R_n^*(\boldsymbol{\theta}^*),\end{aligned}$$

where

$$\boldsymbol{S}_n^*(u)=\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} r_i^*(u)\boldsymbol{X}_i^*(u)K_i(u)$$

and

$$R_n^*(\boldsymbol{\theta}^*) = \sum_{i=1}^n K_i(u) \int_0^{\eta_i(u)} \{I(\varepsilon_{i\tau} \le -s_i(u) - \varphi_i + t) - I(\varepsilon_{i\tau} \le -s_i(u) - \varphi_i)\} \mathrm{d}t.$$

By assumptions C5-C8 and Theorem 1, there exist constant C_6 such that

$$\mathrm{E}|\sum_{k=m+1}^{\infty} b_{\tau k}\xi_{ik}| \le \sum_{k=m+1}^{\infty} |b_{\tau k}|\mathrm{E}|\xi_{ik}| \le \sum_{k=m+1}^{\infty} |b_{\tau k}| \left\{\mathrm{E}\xi_{ik}^2\right\}^{1/2} \le \sum_{k=m+1}^{\infty} C_6 k^{-(\beta_2+\frac{\beta_1}{2})} = C_6 n^{-\frac{\beta_1+2\beta_2-2}{2(\beta_1+2\beta_2)}},$$

$$\begin{aligned} |\sum_{k=1}^m b_{\tau k}(\xi_{ik} - \hat{\xi}_{ik})| &\le \sum_{k=1}^m b_{\tau k}\langle Z_i, \hat{\phi}_k - \phi_k\rangle \le \sum_{k=1}^m b_{\tau k}\|Z_i\|\|\hat{\phi}_k - \phi_k\| \\ &= O_p\left(n^{-1/2}\sum_{k=1}^m k^{1-\beta_2}\right) = O_p\left(n^{-\frac{1}{2}}\right) \end{aligned}$$

and

$$\begin{aligned} |\sum_{k=1}^m (b_{\tau k} - \hat{b}'_{\tau k})\hat{\xi}_{ik}| &\le \sum_{k=1}^m |b_{\tau k} - \hat{b}'_{\tau k}||\langle Z_i, \hat{\phi}_k\rangle \le m^{1/2}(\sum_{k=1}^m |b_{\tau k} - \hat{b}'_{\tau k}|^2)^{1/2} \\ &= O_p\left(n^{-\frac{\beta_2-1}{\beta_1+2\beta_2}}\right). \end{aligned}$$

Thus,

$$\varphi_i = O_p\left(n^{-\frac{\beta_2-1}{\beta_1+2\beta_2}}\right)$$

uniformly for $i = 1, \ldots, n$.

Consider the conditional expectation of $R_n^*(\boldsymbol{\theta}^*)$, we have

$$\begin{aligned} &\mathrm{E}\left\{R_n^*(\boldsymbol{\theta}^*)|\boldsymbol{X}, U, \tilde{\boldsymbol{Z}}\right\} \\ = &\sum_{i=1}^n K_i(u) \int_0^{\eta_i(u)} \{F_\tau(-s_i(u) - \varphi_i + t|\boldsymbol{X}, U, Z(t)) - F_\tau(-s_i(u) - \varphi_i|\boldsymbol{X}, U, Z(t))\}\mathrm{d}t \\ = &\frac{1}{2}\boldsymbol{\theta}^{*T}\left\{\frac{1}{nh_0} f_\tau(-s_i(u) - \varphi_i|\boldsymbol{X}, U, Z(t)) \sum_{i=1}^n K_i(u)\{\boldsymbol{X}_i^*(u)\}\{\boldsymbol{X}_i^*(u)\}^T\right\}\boldsymbol{\theta}^* + o_p(1). \end{aligned}$$

Using similar calculations, we can get $\mathrm{Var}\left\{R_n^*(\boldsymbol{\theta}^*)|\boldsymbol{X}, U, \tilde{\boldsymbol{Z}}\right\} = o_p(1)$. Therefore, we obtain

$$\begin{aligned} R_n^*(\boldsymbol{\theta}^*) &= \mathrm{E}\left\{R_n^*(\boldsymbol{\theta}^*)|\boldsymbol{X}, U, \tilde{\boldsymbol{Z}}\right\} + o_p(1) \\ &= \frac{1}{2}\boldsymbol{\theta}^{*T}\left\{\frac{1}{nh}\sum_{i=1}^n K_i(u) f_\tau(-s_i(u) - \varphi_i|\boldsymbol{X}, U, Z(t))\{\boldsymbol{X}_i^*(u)\}\{\boldsymbol{X}_i^*(u)\}^T\right\}\boldsymbol{\theta}^* + o_p(1) \\ &= \frac{1}{2}\boldsymbol{\theta}^{*T}\boldsymbol{Q}_n(u)\boldsymbol{\theta}^* + o_p(1), \end{aligned}$$

where $\boldsymbol{Q}_n(u) = \frac{1}{nh}\sum_{i=1}^n K_i(u) f_\tau(-s_i(u) - \varphi_i|\boldsymbol{X}, U, Z(t))\{\boldsymbol{X}_i^*(u)\}\{\boldsymbol{X}_i^*(u)\}^T$. By Lemma A.2, we have

$$\begin{aligned}
\boldsymbol{Q}_n(u) &= \mathrm{E}\{\boldsymbol{Q}_n(u)\} + O_p(\log^{1/2}(1/h)/\sqrt{nh}) \\
&= \mathrm{E}\left\{\mathrm{E}\left\{f_\tau(0|\boldsymbol{X}, U, Z(t))\{\boldsymbol{X}_i^*(u)\}\{\boldsymbol{X}_i^*(u)\}^T|U = u\right\}\right\} + O_p(\log^{1/2}(1/h)/\sqrt{nh}) \\
&= f_U(u)\boldsymbol{G}(u) + O_p\left(h^2 + n^{-\frac{\beta_2 - 1}{\beta_1 + 2\beta_2}} + \log^{1/2}(1/h)/\sqrt{nh}\right).
\end{aligned}$$

Thus, $L_n(\boldsymbol{\theta}^*)$ can be written as

$$L_n(\boldsymbol{\theta}^*) = \{\hat{\boldsymbol{S}}_n(u)\}^T\boldsymbol{\theta}^* + \frac{1}{2} f_U(u)\boldsymbol{\theta}^{*T}\boldsymbol{G}(u)\boldsymbol{\theta}^* + O_p\left(h^2 + n^{-\frac{\beta_2 - 1}{\beta_1 + 2\beta_2}} + \log^{1/2}(1/h)/\sqrt{nh}\right),$$

where $\hat{\boldsymbol{S}}_n(u) = \frac{1}{\sqrt{nh}}\sum_{i=1}^n r_i^*(u)\boldsymbol{X}_i^*(u)K_i(u)$. Since $L_n(\boldsymbol{\theta}^*)$ is convex function, following from Lemma A.1, the minimizer of $L_n(\boldsymbol{\theta}^*)$ can be written as

$$\hat{\boldsymbol{\theta}}^* = -f_U^{-1}(u)\{\boldsymbol{G}(u)\}^{-1}\hat{\boldsymbol{S}}_n(u) + O_p\left(h^2 + n^{-\frac{\beta_2 - 1}{\beta_1 + 2\beta_2}} + \log^{1/2}(1/h)/\sqrt{nh}\right), \tag{22}$$

where $\hat{\boldsymbol{\theta}}^* = \sqrt{nh}\left\{\hat{a}_{0\tau} - \alpha_{0\tau}(u), \{\hat{\boldsymbol{a}}_\tau - \boldsymbol{\alpha}_\tau(u)\}^T\right\}^T$.

Let $\boldsymbol{S}_{n1}(u) = \frac{1}{\sqrt{nh}}\sum_{i=1}^n K_i(u) r_i \left\{1, \boldsymbol{X}_i^T\right\}^T$. By simple calculations, It is easy to show that $\mathrm{E}\{\boldsymbol{S}_{n1}(u)\} = \boldsymbol{0}$ and $\mathrm{Var}\{\boldsymbol{S}_{n1}(u)\} = \tau(1-\tau) f_U(u)\nu_0\boldsymbol{H}(u)$. By the Cramér-Wald device, it is quite easy to show that the central limit theorem for $\boldsymbol{S}_{n1}(u)$ holds. According to the central limit theorem, we have

$$\boldsymbol{S}_{n1}(u) \xrightarrow{\mathscr{L}} N\left(\boldsymbol{0}, \tau(1-\tau) f_U(u)\nu_0\boldsymbol{H}(u)\right).$$

Furthermore, we have

$$\begin{aligned}
&\mathrm{Var}\{\boldsymbol{S}_{n1}(u) - \hat{\boldsymbol{S}}_n(u)|\boldsymbol{X}, U, Z(t)\} \\
=\ &\mathrm{Var}\left\{\frac{1}{\sqrt{nh}}\sum_{i=1}^n K_i(u)(r_i^*(u) - r_i)\left\{1, \boldsymbol{X}_i^T\right\}^T\right\} \\
\leq\ &\frac{1}{nh}\sum_{i=1}^n (K_i(u))^2\left(1, \boldsymbol{X}_i^T\right)^T\left(1, \boldsymbol{X}_i^T\right)\{F_\tau(|s_i(u) + \varphi_i||\boldsymbol{X}, U, Z(t)) - F_\tau(0|\boldsymbol{X}, U, Z(t))\} \\
=\ &o_p(1).
\end{aligned}$$

Thus, $\mathrm{Var}\{\boldsymbol{S}_{n1}(u) - \hat{\boldsymbol{S}}_n(u)\} = o(1)$. By Slutsky's theorem, we can obtain

$$\hat{\boldsymbol{S}}_n(u) - \mathrm{E}\left(\hat{\boldsymbol{S}}_n(u)\right) \xrightarrow{\mathscr{L}} N\left(\boldsymbol{0}, \tau(1-\tau) f_U(u)\nu_0\boldsymbol{H}(u)\right). \tag{23}$$

Now consider the expectation of $\hat{\boldsymbol{S}}_n(u)$. Note that

$$
\begin{aligned}
& \mathrm{E}\left\{\frac{1}{\sqrt{nh}}\hat{\boldsymbol{S}}_n(u)|\boldsymbol{X},U,Z(t)\right\} \\
= & \frac{1}{nh}\sum_{i=1}^{n}\{F_\tau(-s_i(u)-\varphi_i|\boldsymbol{X},U,Z(t))-F_\tau(0|\boldsymbol{X},U,Z(t))\}K_i(u)\left(1,\boldsymbol{X}_i^T\right)^T \\
= & -\frac{1}{nh}\sum_{i=1}^{n} f_\tau(0|\boldsymbol{X},U,Z(t))K_i(u)s_i(u)\{1+o(1))\left(1,\boldsymbol{X}_i^T\right)^T.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& \mathrm{E}\left\{\frac{1}{\sqrt{nh}}\hat{\boldsymbol{S}}_n(u)\right\} \\
= & \mathrm{E}\left\{\mathrm{E}\left\{\frac{1}{\sqrt{nh}}\hat{\boldsymbol{S}}_n(u)|\boldsymbol{X},U,Z(t)\right\}\right\} \\
= & -\frac{\mu_2 h^2}{2} f_U(u)\boldsymbol{G}(u)\begin{pmatrix}\alpha''_{0\tau}(u)\\ \boldsymbol{\alpha}''_\tau(u)\end{pmatrix}+o_p(h^2).
\end{aligned}
\tag{24}
$$

Seen from (22), (23) and (24), (8) holds. This completes the proof of Theorem 2.

References

Aneiros-Pérez, G. and Vieu, P. (2006). Semi-functional partial linear regression. *Statistics & Probability Letters*, **76**, 1102-1110.

Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics*, **34**, 2159-2179.

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

Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, **45**, 11-22.

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

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

Dabo-Niang, S. and Guillas, S. (2010). Functional semiparametric partially linear model with autoregressive errors. *Journal of multivariate analysis*, **101**, 307-315.

Feng, S. and Xue, L. (2016). Partially functional linear varying coefficient model. *Statistics*, 1-16.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, **35**, 70-91.

Hjort, N. L. and Pollard, D. (2011). Asymptotics for minimisers of convex processes. *arXiv preprint arXiv*:1107.3806.

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

Kato, K. (2013). Estimation in functional linear quantile regression. *The Annals of Statistics*, **40**, 3108-3136.

Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *The Annals of Statistics*, **26**, 755-770.

Lu, Y., Du, J. and Sun, Z. M. (2014). Functional partially linear quantile regression model. *Metrika*, **77**, 317-332.

Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Probability Theory and Related Fields*, **61(3)**, 405-415.

Peng, Q. Y., Zhou, J. J. and Tang, N. S. (2015). Varying coefficient partially functional linear regression models. *Stat Papers*, 1-15.

Shin, H. (2009). Partial functional linear regression. *Journal of Statistical Planning and Inference*, **139**, 3405-3418.

Tang, Q. and Cheng, L. S. (2014). Partial functional linear quantile regression. *Science China Mathematics*, **57**, 2589-2608.

Van Der Vaart, A. W. and Wellner, J. A. (1996). Weak convergence. In Weak Convergence and Empirical Processes (pp. 16-28). Springer New York.

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

Zhang, D. W., Lin, X. H. and Sowers, M. F. (2007). Two-stage functional mixed models for evaluating the effect of longitudinal covariate profiles on a scalar outcome. *Biometrics*, **63**, 351-362.