Statistics

A Journal of Theoretical and Applied Statistics

ISSN: 0233-1888 (Print) 1029-4910 (Online) Journal homepage: http://www.tandfonline.com/loi/gsta20

Quantile regression in functional linear semiparametric model

Tang Qingguo & Linglong Kong

STATISTICS, 2017
VOL. 51, NO. 6, 1342–1358
https://doi.org/10.1080/02331888.2017.1300803

Quantile regression in functional linear semiparametric model

Tang Qingguo[a] and Linglong Kong[b]

[a]School of Economics and Management, Nanjing University of Science and Technology, Nanjing, People's Republic of China; [b]Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada

ABSTRACT
This paper proposes nonparametric estimation methods for functional linear semiparametric quantile regression, where the conditional quantile of the scalar responses is modelled by both scalar and functional covariates and an additional unknown nonparametric function term. The slope function is estimated using the functional principal component basis and the nonparametric function is approximated by a piecewise polynomial function. The asymptotic distribution of the estimators of slope parameters is derived and the global convergence rate of the quantile estimator of unknown slope function is established under suitable norm. The asymptotic distribution of the estimator of the unknown nonparametric function is also established. Simulation studies are conducted to investigate the finite-sample performance of the proposed estimators. The proposed methodology is demonstrated by analysing a real data from ADHD-200 sample.

ARTICLE HISTORY
Received 21 March 2016
Accepted 24 February 2017

KEYWORDS
Functional linear semiparametric model; quantile regression; asymptotic distribution; convergence rate; ADHD

1. Introduction

In the last two decades, there has been an increasing interest in regression models for functional variables as more and more data have arisen where the primary unit of observation can be viewed as a curve or in general a function, such as in biology, chemometrics, econometrics, geophysics, the medical sciences, meteorology and neurosciences. As a natural extension of the ordinary regression to the case where predictors include random functions and responses are scalars or functions, functional linear regression analysis provides valuable insights into these problems. The functional linear model has been extensively studied and successfully applied; see [1–10], among many others.

Engle et al. [11] introduced and applied the following semiparametric model:

$$Y = \beta^{\mathrm{T}} Z + f(T) + \varepsilon$$

to study the effect of weather on electricity demand, where β is a vector of unknown parameters, Z is a vector of random variables, $f(\cdot)$ is an unknown function to be estimated, T is a random variable and ε is an unobserved disturbance. This model is much more flexible than the standard linear model since it combines both parametric and nonparametric components. The semiparametric model has been extensively investigated by many authors, such as Heckman [12], Carrol et al. [13], Zhang et al. [14], He et al. [15], Fan and Li [16], Chen and Jin [17] and Tang [18].

In many applications, such as in the analysis of complex neuroimaging data [19] and the foetal heart rate data [20], a response Y is usually related to some scalar covariates and some functional covariates, and in addition the response Y may be nonlinearly related to certain scalar variable.

CONTACT Tang Qingguo tangqguo@163.com School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, People's Republic of China

Motivated by neuroimaging data analysis from ADHD-200 sample [19], we propose the following functional linear semiparametric model:

$$Y = \beta^{\mathrm{T}} Z + \int_0^1 \gamma(s) X(s)\, \mathrm{d}s + f(T) + \varepsilon, \tag{1}$$

where $X(s)$ is a square integrable random function on $[0,1]$, $\gamma(s)$ is a square integrable function on $[0,1]$, $Z = (Z_1, \ldots, Z_d)^{\mathrm{T}}$ is a d-dimensional vector of random variables, β is a d-dimensional vector of regression coefficients, T is a random variable, $f(t)$ is an unknown nonparametric function and ε is a random error. Model (1) generalizes both the functional linear regression model and the semiparametric model which correspond to the cases $\beta = 0$, $f(t) = 0$ and $\gamma(s) = 0$, respectively. Moreover, this model includes the partial functional linear model which corresponds to the case $f(t) = 0$, which has been widely studied by many authors, see, for example, [7,8,21–23].

In this paper, we estimate the conditional quantile of the response instead of the usual conditional mean. Quantile regression has several advantages over mean regression. For example, they can be defined without any moment conditions. Plotting a set of quantiles would give us more complete understanding of the data than plotting just the mean. When the centre of the conditional distribution of Y is of interest, the median regression, a special case of quantile regression, provides more robust estimators than the mean regression. For a complete review on quantile regression, see [24]. Inference procedures based on quantile regression have been considered in the literature; see, for example, [25–27], among others. In this paper, we first use a piecewise polynomial function to approximate the unknown function $f(t)$ and then estimate the quantile slope function $\gamma(s)$ using the functional principal component analysis. Under some regularity conditions, we derive the asymptotic normality of the estimator of β and establish the global convergence rate of the estimator of the quantile slope function $\gamma(s)$. It may not be appealing to use this piecewise polynomial directly as the estimate of the function $f(t)$, because it is only piecewise smooth. In order to estimate the function $f(t)$ more efficiently, in model (1), we replace β and $\gamma(s)$ by their estimators and use local linear kernel smooth method to estimate $f(t)$. Under the general regularity conditions, we derive the asymptotic distribution of the estimator of $f(t)$.

The paper is organized as follows. Section 2 describes the estimation methods. Section 3 presents the asymptotic theory of our estimator. In Section 4, we conduct simulation studies to examine the finite sample performance of the proposed procedures. In Section 5, the proposed methods are illustrated by a real data example from ADHD 200 sample. All proofs are relegated to the appendix.

2. Estimation method

In functional linear semiparametric quantile regression, for a given quantile level $\tau \in (0,1)$, the data $(Z_i, X_i(s), T_i, Y_i)$, $i = 1, \ldots, n$, which are independent and identically distributed (i.i.d.), are generated from the model

$$q_\tau(Y \mid Z, X(s), T) = Z^{\mathrm{T}} \beta_\tau + \int_0^1 \gamma_\tau(s) X(s)\, \mathrm{d}s + f_\tau(T), \tag{2}$$

where $q_\tau(Y \mid Z, X(s), T)$ is the τth conditional quantile of Y given $(Z, X(s), T)$, both Y and T are real-valued random variables defined on a probability space $(\Omega, \mathcal{F}, P)$, Z is a d-dimensional vector of random variables with finite second moments, $\{X(s) : s \in [0,1]\}$ is a zero-mean and second-order (i.e. $EX(s)^2 < \infty$ for all $s \in [0,1]$) stochastic process defined on $(\Omega, \mathcal{F}, P)$ with sample paths in $L_2([0,1])$, the set of all square integrable functions on $[0,1]$, β_τ is a $d \times 1$ coefficient vector to be estimated, $\gamma_\tau(s)$ is an unknown square integrable function on $[0,1]$, $f_\tau(t)$, $t \in [t_1, t_2]$ is an unknown smooth function. Let $< \cdot, \cdot >$ and $\| \cdot \|$ represent, respectively, the $L_2([0,1])$ inner product and norm.

Define the covariance kernel $S(s,t) = \mathrm{Cov}(X(s), X(t))$. Assume that $S(s,t)$ is continuous on $[0,1] \times [0,1]$. Then, Mercer's theorem implies that

$$S(s,t) = \sum_{j=1}^{\infty} \kappa_j \psi_j(s)\psi_j(t), \quad \kappa_1 \geq \kappa_2 \geq \cdots \geq 0, \tag{3}$$

where (κ_j, ψ_j) are (eigenvalue, eigenfunction) pairs for the linear operator with kernel S, and the functions $\psi_1, \psi_2, \cdots$ form an orthonormal basis for $L_2([0,1])$. By the Karhunen–Lo ève representation, $X(s)$ and $\gamma_\tau(s)$ can be expanded in $L_2([0,1])$ as

$$X(s) = \sum_{j=1}^{\infty} \eta_j \psi_j(s), \quad \gamma_\tau(s) = \sum_{j=1}^{\infty} \gamma_{\tau j} \psi_j(s),$$

where the $\eta_j = \int_0^1 X(s)\psi_j(s)\,\mathrm{d}s$ are uncorrelated random variables with mean 0 and variance $E\eta_j^2 = \kappa_j$. Thus (2) can be written as

$$q_\tau(Y \mid Z, X(s), T) = Z^{\mathrm{T}}\beta_\tau + \sum_{j=1}^{\infty} \gamma_{\tau j}\eta_j + f_\tau(T). \tag{4}$$

The sample version of S is $\hat{S}(s,t) = (1/n)\sum_{i=1}^{n} X_i(s)X_i(t)$. Similar to the case of S, $\hat{S}$ has the spectral expansion

$$\hat{S}(s,t) = \sum_{j=1}^{\infty} \hat{\kappa}_j \hat{\psi}_j(s)\hat{\psi}_j(t), \quad \hat{\kappa}_1 \geq \hat{\kappa}_2 \geq \cdots \geq 0.$$

where $(\hat{\kappa}_j, \hat{\psi}_j)$ are (eigenvalue, eigenfunction) pairs and $\{\hat{\psi}_j\}_{j=1}^{\infty}$ is an orthonormal basis for $L_2([0,1])$. We set $(\hat{\kappa}_j, \hat{\psi}_j)$ to be the estimator of (κ_j, ψ_j) and use $\sum_{j=1}^{m} \gamma_j \hat{\psi}_j(s)$ to approximate $\gamma_\tau(s)$, where m is the truncation level that trade-off approximation error against variability, and m typically diverges with n.

In order to approximate $f_\tau(t)$ for $t \in [t_1, t_2]$, we construct piecewise polynomial estimators of $f_\tau(t)$ of degree p. We split equally $[t_1, t_2]$ into M_n subintervals. Then the length of every subinterval is $2h_0 = (t_2 - t_1)/M_n$. Let $I_\nu = [t_1 + 2(\nu - 1)h_0, t_1 + 2\nu h_0,)$ for $1 \leq \nu \leq M_n - 1$ and $I_{M_n} = [t_2 - 2h_0, t_2]$. Let t_ν denote the centre of the interval I_ν and χ_ν denote the indicator function of I_ν, so that $\chi_\nu(t) = 1$ or 0 according to $t \in I_\nu$ or $t \bar{\in} I_\nu$. Denote

$$A_\nu(t) = (1, (t - t_\nu)/h_0, \ldots, [(t - t_\nu)/h_0]^p)^{\mathrm{T}}, \quad \nu = 1, \ldots, M_n,$$
$$A(t) = (\chi_1(t)A_1(t)^{\mathrm{T}}, \ldots, \chi_{M_n}(t)A_{M_n}(t)^{\mathrm{T}})^{\mathrm{T}}.$$

Set $\theta_\nu = (\theta_{\nu 0}, \ldots, \theta_{\nu p})^{\mathrm{T}}$ and $\theta = (\theta_1^{\mathrm{T}}, \ldots, \theta_{M_n}^{\mathrm{T}})^{\mathrm{T}}$. We use $\breve{f}(t) = A^{\mathrm{T}}(t)\theta$ to approximate $f_\tau(t)$. Note that $\breve{f}(t)$ is a piecewise polynomial of degree p. Based on $(Z_i, X_i(t), T_i, Y_i),\ i = 1, \ldots, n$, we solve the following minimization problem:

$$\min \sum_{i=1}^{n} \rho_\tau(Y_i - Z_i^{\mathrm{T}}\beta - \sum_{j=1}^{m} \gamma_j \hat{\eta}_{ij} - A^{\mathrm{T}}(T_i)\theta) \tag{5}$$

with respect to the $\beta, \gamma_j,\ j = 1, \ldots, m$ and θ, where $\rho_\tau(t) = t(\tau - I_{(t<0)})$ is the quantile loss function and $\hat{\eta}_{ij} = \langle X_i, \hat{\psi}_j \rangle$. The solution to Equation (5) can be obtained numerically by linear programming method. The quantile estimators of β_τ and $\gamma_\tau(s)$ are denoted by $\hat{\beta}$ and $\hat{\gamma}(s) = \sum_{j=1}^{m} \hat{\gamma}_j \hat{\psi}_j(s)$, respectively.

After estimating β_τ and $\gamma_\tau(s)$, for a given $t_0 \in [t_1, t_2]$, for t in the neighbourhood of t_0, we use $a_0 + a_1(t - t_0)$ to approximate the unknown function $f_\tau(t)$. Based on $(Z_i, X_i(s), T_i, Y_i)$, $i = 1, \ldots, n$, we solve the following minimization problem:

$$\min \sum_{i=1}^{n} \rho_\tau(Y_i - Z_i^{\mathrm{T}}\hat{\beta} - \sum_{j=1}^{m} \hat{\gamma}_j\hat{\eta}_{ij} - (a_0 + a_1(T_i - t_0)))K\left(\frac{T_i - t_0}{h}\right) \tag{6}$$

with respect to a_0, a_1, where $K(\cdot)$ is a given kernel function and h is a chosen bandwidth. The solution to Equation (6) can also be obtained numerically by linear programming method. Let $\hat{a}_0, \hat{a}_1$ be the minimizer of Equation (6). Then the estimator $\hat{f}(t_0)$ of $f_\tau(t_0)$ is $\hat{f}(t_0) = \hat{a}_0$.

For the finite-sample case, smoothing parameters m, M_n and h need to be chosen. The parameters m and M_n can be chosen by information criteria BIC. The BIC criteria as a function of m and M_n is given by

$$\mathrm{BIC}(m, M_n) = \log\left\{\sum_{i=1}^{n} \rho_\tau\left(Y_i - Z_i^{\mathrm{T}}\hat{\beta} - \sum_{j=1}^{m}\hat{\gamma}_j\hat{\eta}_{ij} - A^{\mathrm{T}}(T_i)\hat{\theta}\right)\right\} + \frac{(m + M_n)\log n}{n}.$$

Large values of BIC indicate poor fits. The bandwidth h can be selected by leave-one-curve-out cross-validation of the prediction error. Define CV function as

$$\mathrm{CV}(h) = \sum_{i=1}^{n} \rho_\tau\left(Y_i - Z_i^{\mathrm{T}}\hat{\beta}^{-i} - \sum_{j=1}^{m}\hat{\gamma}_j^{-i}\hat{\eta}_{ij} - \hat{f}^{-i}(T_i)\right),$$

where $\hat{\beta}^{-i}$, $\hat{\gamma}_j^{-i}$ and $\hat{f}^{-i}$ are computed by removing the ith subject.

3. Theoretical properties

To establish asymptotic normality and rates of convergence for the proposed estimators, the following assumptions are required.

(1) $E(X \mid T) = 0$ and $E(\eta_k\eta_j \mid T) = 0$ for $k \neq j$. X has finite fourth moment, in that $\int_0^1 E[X^4(s)]\,\mathrm{d}s < \infty$, and for each j, $E(\eta_j^4) < C_1\kappa_j^2$ and $E(\eta_j^2 \mid T) < C_1\kappa_j$ for some constant C_1. The density function $v(t)$ of T is continuous and bounded away from zero and infinity over $[t_1, t_2]$.

(2) The eigenvalues κ_j in the spectral decomposition (3) satisfy

$$C_2^{-1}j^{-a} \leq \kappa_j \leq C_2 j^{-a}, \quad \kappa_j - \kappa_{j+1} \geq C_2 j^{-(a+1)}, \quad j \geq 1,$$

where $a > 1$ and C_2 is a positive constant.

(3) The coefficients $\gamma_{\tau j}$ of $\gamma(s)$ satisfy that $|\gamma_{\tau j}| \leq C_3 j^{-b}$ for all $j \geq 1$, where $b > 1 + a/2$ and C_3 is a positive constant.

(4) $f_\tau(t)$ is a p-times continuously differential function such that $|f_\tau^{(p)}(t') - f_\tau^{(p)}(t)| \leq C_4|t' - t|^{\varsigma}$, for $t_1 \leq t,\ t' \leq t_2$, where $0 < \varsigma \leq 1$ and C_4 is a positive constant. Think of $\tilde{p} = p + \varsigma$ as a measure of the smoothness of the function $f_\tau(t)$, $\tilde{p} > (a + 2b - 1)/2$.

(5) $m = O(n^{1/(a+2b)})$ and $M_n = O(n^{1/(a+2b)})$.

(6) Let $\varepsilon_\tau = Y - Z^{\mathrm{T}}\beta_\tau - \int_0^1 \gamma_\tau(s)X(s)\,\mathrm{d}s - f_\tau(T)$, the conditional density function $g(u \mid X, T)$ of ε_τ has a continuous and uniformly bounded derivative in the neighbourhood of zero and satisfies that $0 < c_0 \leq g(0 \mid X, T) \leq c_1 < \infty$, where c_0 and c_1 are two positive constants. $\varrho(t) = E(g(0 \mid X, T) \mid T = t)$ is continuous in a neighbourhood of t_0.

(7) $E(Z_r^4) < +\infty$ for $r = 1, \ldots, d$.

(8) The bandwidth h satisfies that $h \leq C_5 n^{-1/5}$ for some positive constant C_5 and $(nh)^{-1/2} m^{1/2} \to 0$ as $n \to \infty$.
(9) The kernel $K(\cdot) \geq 0$ is a bounded symmetric function with a compact support $[-M, M]$.

Assumption 2 prevents the spacings among eigenvalues being too small and Assumption 3 requests that the regression weight function is sufficiently smooth relative to covariance kernel S. Assumptions 1–3 are typically used in functional linear regression model (see [2,4]). Assumption 4 is equivalent to suppose that $f_\tau(t)$ is $\tilde{p}$-Hölder continuous and the quantity $\tilde{p}$ is the order of smoothness of the function $f_\tau(t)$. Assumptions 6 is standard assumption used in quantile regression and similar assumption can be found in [28].

One complicating issue for the functional linear semiparametric model is the dependence between Z and X and T. To this end, we assume the relationship

$$Z_{ir} = \sum_{j=1}^{\infty} w_{rj}\eta_{ij} + f_r(T_i) + \epsilon_{ir}, \quad r = 1, \ldots, d, \tag{7}$$

where the w_{rj} satisfying Assumption 3, that is, $|w_{rj}| \leq C_3 j^{-b}$ for all $j \geq 1$, and $f_r(t)$ satisfying assumption 4, the $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{id})^{\mathrm{T}}$ are zero-mean vectors of random variables and independent of the η_{ij}, T_i and $\varepsilon_{\tau i}$. Let $B = E(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^{\mathrm{T}})$. We have the following results.

Theorem 3.1: *Under the assumptions* 1–8, *if B is invertible, then,*

$$\sqrt{n}(\hat{\beta} - \beta_\tau) \to_d N\left(0, \frac{\tau(1-\tau)}{\varpi^2} B^{-1}\right), \tag{8}$$

where $\varpi = E[g(0 \,|\, X, T)]$ and $\to_d$ means convergence in distribution.

Theorem 3.2: *Assume that Assumptions* 1–7 *hold. Then,*

$$\int_0^1 [\hat{\gamma}(s) - \gamma_\tau(s)]^2 \, \mathrm{d}s = O_p(n^{-(2b-1)/(a+2b)}). \tag{9}$$

The result of Theorem 3.2 indicates that the estimator $\hat{\gamma}(s)$ obtain the same rate of convergence as for the estimators of Hall and Horowitz [4] and [28], which are optimal in the minimax sense.

Let $\mu_k = \int u^k K(u)\,\mathrm{d}u$, $\nu_k = \int u^k K^2(u)\,\mathrm{d}u$ for $k = 0, 1, \ldots$. The following Theorem 3.3 gives the asymptotic distribution of the estimator of $f_\tau(t_0)$.

Theorem 3.3: *Suppose that assumptions* 1–9 *hold. If t_0 is an interior point of $[t_1, t_2]$, then, as $n \to \infty$,*

$$\sqrt{nh}\left(\hat{f}(t_0) - f_\tau(t_0) - \frac{\mu_2 h^2}{2\mu_0} f_\tau''(t_0)\right) \to_d N\left(0, \frac{\nu_0 \tau(1-\tau)}{\mu_0^2 \upsilon(t_0)\varrho^2(t_0)}\right), \tag{10}$$

Remark 3.1: In practical application, $X(s)$ is only discretely observed. Without loss of generality, suppose for each $i = 1, \ldots, n$, $X_i(s)$ is observed at n_i discrete points $0 = s_{i1} < \cdots < s_{in_i} = 1$. Typically, $d_n = \max_i \max_{1 \leq j \leq n_i - 1}(s_{i(j+1)} - s_{ij}) \to 0$ as $n \to \infty$ is also assumed. Linear interpolation functions or spline interpolation functions can be used for the estimators of $X_i(s)$. For example, we can use the following linear interpolation function:

$$\hat{X}_i(s) = X_i(s_{ij}) + \frac{(X_i(s_{i(j+1)}) - X_i(s_{ij}))}{s_{i(j+1)} - s_{ij}}(s - s_{ij}), \quad \text{for } s \in [s_{ij}, s_{i(j+1)}], \; j = 0, \ldots, n_i - 1$$

as the estimator of $X_i(s)$. Furthermore, under Assumptions 1–9 and the assumptions that $d_n = O(n^{-1/2})$ and the covariance function $S(s,t)$ is twice continuously differentiable, by arguments similar to those used in the proof of Lemma D.1 of Kato [28], we can prove that the conclusions of Theorems 3.1–3.3 still hold.

Remark 3.2: We note that Li et al. [29] studied statistical inference in functional quantile regression for scalar response and a functional covariate. Under the conditions that the functional covariate $X_i(s)$ is contaminated with noise and the number of eigenvalues of the covariance function $S(s,t)$ is finite, they derived the asymptotic normality of the quantile regression estimators. We assume that the number of eigenvalues of $S(s,t)$ is infinite. The corresponding quantile regression problem is ill-posed and the estimator of the slope function can not achieve the root-n rate as shown by Kato [28]. We instead established the global convergence rate of the slope function estimation in Theorem 3.2.

Remark 3.3: Condition (7) is needed to derive the asymptotic distribution of the estimator of unknown parametric vector β_τ in functional linear semiparametric quantile regression. Similar conditions can be found in [15,21]. In fact, the assumption that $|w_{rj}| \le C_3 j^{-b}$ for $j \ge 1$ is weaker than condition (15) in [21].

4. Simulation studies

To illustrate the numerical performance of the proposed method, some simulation study was conducted. We generated the data sets from the following model:

$$q_\tau(Y_i \mid Z_{i1}, Z_{i2}, X_i(s), T_i) = \beta_1 Z_{i1} + \beta_2 Z_{i2} + \int_0^1 \gamma_\tau(s) X_i(s)\, ds + f_\tau(T_i), \tag{11}$$

where $\beta_1 = 3$, $\beta_2 = -2$, $\gamma_\tau(s) = \sum_{j=1}^{50} \tilde{\gamma}_{\tau j}\psi_j(s)$ with $\tilde{\gamma}_{\tau 1} = 1.5$ and $\tilde{\gamma}_{\tau j} = 2(-1)^{j+1} j^{-2}$ for $j \ge 2$, $X_i(s) = \sum_{j=1}^{50} \eta_{ij}\psi_j(s)$ with $\psi_1(s) \equiv 1$ and $\psi_j(s) = 2^{1/2}\cos((j-1)\pi s)$ for $j \ge 2$ and the ξ_{ij}'s were independent and normal $N(0, j^{-1.5})$. We chose $f_\tau(t) = 1.6t^3 - 2.5t^2 - 2t + 3$. The T_i's were uniformly distributed on $[0,1]$. We let $Z_{i1} = 2\eta_{i1} + \eta_{i2} - T_i + 2 + \epsilon_{i1}$ and $Z_{i2} = \eta_{i2} - 2\eta_{i3} + 2T_i - 1 + \epsilon_{i2}$, where $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2})^{\mathrm{T}}$ are independent and normal $N(0, \Sigma)$ with $\Sigma = (\sigma_{ij})_{2\times 2}$ and $\sigma_{11} = 1$, $\sigma_{12} = \sigma_{21} = 1/2$ and $\sigma_{22} = 2$. $\boldsymbol{\epsilon}_i$ were independent of the η_{ij} and T_i. The errors $\varepsilon_{\tau i} = \varepsilon_i - F^{-1}(\tau)$ with F being the distribution function of ε_i. Here, $F^{-1}(\tau)$ is subtracted from ε_i to make the τth quantile of $\varepsilon_{\tau i}$ zero for identifiability purpose.

We conducted 500 trials from model (11) with sample size $n = 100$ and $n = 200$, respectively. In each trial the estimators of β_1, β_2 and $\gamma_\tau(s)$ were computed by solving minimization problem (5) with $f_\tau(t)$ being approximated by piecewise local linear functions. The tuning parameter m and the number of subintervals M_n were determined by BIC criterion as described in Section 2. The estimators of $f_\tau(t)$ was computed by minimizing (6) and using Epanechnikov kernel. The bandwidth h was chosen by leave-one-curve-out cross-validation. We first consider the case that the error $\varepsilon_i \sim N(0,1)$. Table 1 reports the biases and standard deviations (sd) of the estimators $\hat{\beta}_k$ for $k = 1,2$ and integrated squared biases (Bias^2) and integrated variances (Var) of the estimators $\hat{\gamma}(s)$ and $\hat{f}(t)$ computed on a grid of

Table 1. Simulation results for quantile estimators under normal error.

		$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\gamma}(s)$		$\hat{f}(t)$	
n	τ	bias	sd	bias	sd	Bias^2	Var	Bias^2	Var
100	0.05	−0.0009	0.2386	−0.0034	0.1655	0.0303	0.8442	0.0102	0.3237
	0.25	0.0156	0.1536	−0.0095	0.1120	0.0309	0.3838	0.0013	0.1499
	0.5	−0.0076	0.1381	0.0033	0.1004	0.0302	0.2958	0.0001	0.1171
	0.75	−0.0101	0.1556	0.0038	0.1111	0.0305	0.3496	0.0003	0.1474
	0.95	−0.0120	0.2300	−0.0019	0.1626	0.0305	0.8426	0.0104	0.3048
200	0.05	−0.0051	0.1759	0.0002	0.1154	0.0301	0.4199	0.0031	0.1733
	0.25	0.0036	0.1063	−0.0055	0.0772	0.0306	0.1627	0.0001	0.0704
	0.5	0.0069	0.1027	−0.0022	0.0676	0.0304	0.1510	0.0003	0.0646
	0.75	0.0106	0.1058	−0.0038	0.0740	0.0305	0.1605	0.0009	0.0727
	0.95	−0.0054	0.1541	0.0034	0.1166	0.0304	0.4005	0.0029	0.1561

100 equally spaced points on [0, 1], based on quantile levels $\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$. We see from Table 1 that the standard deviations of the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ and the integrated variances of $\hat{\gamma}(s)$ and $\hat{f}(t)$ decrease as the sample size n increases from 100 to 200, but increase as the quantile level τ departs from 0.5. These indicate that the median estimators are more efficient than the estimators of other quantile levels. Table 1 also shows that changes of n and quantile level τ have little influence on the biases of $\hat{\beta}_1$ and $\hat{\beta}_1$ and the integrated squared biases of $\hat{\gamma}(s)$ and $\hat{f}(t)$.

To investigate the effect of data contamination on the quantile estimators, we consider the situation of model (11) with non-normal errors $\varepsilon_i \sim 0.9N(0, 1) + 0.1N(0, 8^2)$, that is, ε_i come from a contaminated normal distribution. $N(0, 8^2)$ can be interpreted as an outlier distribution. The simulation results under this contaminated normal error are reported in Table 2. We see from Table 2 that the estimators under quantile levels 0.5 and 0.75,0.25 are more efficient than the estimators under quantile levels 0.1,0.05 and 0.9,0.95. Comparing Table 2 with Table 1, we see that outliers in the data have little influence on the estimators under quantile levels 0.5 and 0.75,0.25, but have large influence on the estimators under quantile levels 0.1,0.05 and 0.9,0.95.

In the following, we consider another situation of model (11) with non-normal errors $\varepsilon_i \sim 0.9N(0, 1) + 0.1N(\mu, 1)$. Tables 3 and 4 give the simulation results with $\mu = -10$ and $\mu = 10$, respectively, which correspond to adding outliers on the left end and right end, respectively. Table 3 shows that outliers on the left end have large influence on the estimators under lower quantile levels 0.05 and 0.1, but have little influence on the estimators under larger quantile levels 0.95 and 0.9. This is because of the fact that outliers on the left end mainly focus on the left end of the data, which has little influence on the estimators of large quantile levels. Similarly, Table 4 shows that the estimators under lower quantile levels are affected less by the outliers on the right end than the estimators under larger quantile levels.

Table 2. Simulation results for quantile estimators under contaminated normal error.

		$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\gamma}(s)$		$\hat{f}(t)$	
n	τ	bias	sd	bias	sd	Bias^2	Var	Bias^2	Var
100	0.05	−0.0358	0.8106	0.0325	0.5727	0.0336	9.9817	3.7361	6.7977
	0.1	0.0144	0.3707	−0.0009	0.2612	0.0317	1.9736	0.3121	1.5031
	0.25	−0.0055	0.1842	0.0014	0.1342	0.0303	0.4942	0.0061	0.2138
	0.5	−0.0004	0.1560	−0.0006	0.1071	0.0300	0.3752	0.0001	0.1616
	0.75	−0.0036	0.1861	−0.0054	0.1336	0.0305	0.4871	0.0098	0.2114
	0.9	0.0147	0.3470	−0.0084	0.2731	0.0316	1.9345	0.2390	1.2172
	0.95	0.0515	0.8051	−0.0284	0.5843	0.0491	9.7863	3.5268	5.2692
200	0.05	0.0073	0.5905	0.0008	0.4242	0.0325	5.2496	2.9602	3.9100
	0.1	−0.0004	0.2278	0.0082	0.1564	0.0300	0.7470	0.1341	0.4209
	0.25	0.0005	0.1228	−0.0020	0.0848	0.0303	0.2249	0.0062	0.0926
	0.5	0.0007	0.1037	−0.0061	0.0765	0.0307	0.1659	0.0001	0.0678
	0.75	−0.0026	0.1193	0.0018	0.0885	0.0300	0.2162	0.0065	0.0892
	0.9	−0.0007	0.2218	−0.0063	0.1536	0.0300	0.7472	0.1269	0.4082
	0.95	0.0217	0.5778	0.0295	0.4026	0.0368	5.0164	2.5157	3.4262

Table 3. Simulation results for quantile estimators with outliers on the left end.

	$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\gamma}(s)$		$\hat{f}(t)$	
τ	bias	sd	bias	sd	Bias^2	Var	Bias^2	Var
0.05	−0.1773	1.3133	0.0091	0.9898	0.2274	32.6601	37.3243	9.2052
0.1	−0.1046	1.2599	0.0257	0.9148	0.1074	23.4066	18.9613	9.3544
0.25	−0.0184	0.2547	0.0043	0.1782	0.0311	0.9061	0.1509	0.5515
0.5	−0.0077	0.1626	−0.0071	0.1109	0.0322	0.3654	0.0223	0.1617
0.75	−0.0005	0.1601	−0.0025	0.1168	0.0307	0.3922	0.0119	0.1648
0.9	−0.0024	0.2074	0.0030	0.1493	0.0302	0.6085	0.0140	0.2525
0.95	−0.0105	0.2387	0.0048	0.1707	0.0314	0.8997	0.0227	0.3200

Table 4. Simulation results for quantile estimators with outliers on the right end.

	$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\gamma}(s)$		$\hat{f}(t)$	
τ	bias	sd	bias	sd	Bias^2	Var	Bias^2	Var
0.05	−0.0186	0.2498	−0.0050	0.1825	0.0313	0.9058	0.0355	0.3327
0.1	−0.0018	0.2016	−0.0072	0.1363	0.0319	0.5845	0.0158	0.2465
0.25	0.0130	0.1581	−0.0006	0.1183	0.0321	0.3964	0.0089	0.1650
0.5	0.0124	0.1692	−0.0123	0.1182	0.0311	0.3829	0.0172	0.1556
0.75	0.0009	0.2421	−0.0014	0.1704	0.0307	0.9244	0.1728	0.5553
0.9	0.2046	1.1093	−0.0923	0.8159	0.2466	23.6445	16.9080	6.8469
0.95	0.7120	1.1999	−0.1867	1.0283	2.4930	29.7720	27.6573	6.8513

5. Application

We apply our proposed method to a dataset on attention deficit hyperactivity disorder (ADHD) from the ADHD-200 Sample Initiative Project. Yu et al. [19] described the dataset in detail. ADHD is the most commonly diagnosed behavioural disorder of childhood, and can continue through adolescence and adulthood. The symptoms include lack of attention, hyperactivity, and impulsive behaviour. The dataset we use is the filtered preprocessed resting state data from New York University (NYU) Child Study Center using the Anatomical Automatic Labeling (AAL) atlas. AAL contains 116 Regions of Interests (ROI) fractionated into functional space using nearest-neighbour interpolation. After cleaning the raw data that failed in quality control or has missing data, we include 120 individuals in the analysis. The response of interest Y is the ADHD index, Conners' parent rating scale-revised, long version (CPRS-LV), a continuous behaviour score reflecting the severity of the ADHD disease. In the AAL atalas data, the mean of the grey scale in each region is calculated for 172 equally spaced time points. We consider the most important part of the brain cerebelum which contain at least 4 ROIs. The functional predictor $X(s)$ is computed by taking the average grey scale of the ROIs corresponding to this part. The scalar covariates of primary interest include gender (Z_1), age (T), handedness (Z_2), continuous between -1 and 1, diagnosis status (Z_3, Z_4, categorical with 3 levels: ADHD-combined, ADHD-inattentative and typically developing children), medication status (Z_5), Verbal IQ (Z_6), Full4 IQ (Z_7) and Performance IQ (Z_8). Yu et al. [19] used partial functional linear quantile regression to analyse these data. We find that there is no explicit linear relation between the ADHD index and age and it is very likely that Y is related nonlinearly to T. We then take age (T) as the nonlinear variable and construct the following functional linear semiparametric quantile regression:

$$q_\tau(Y_i \mid Z_i, X_i(s), T_i) = \sum_{j=1}^{8} Z_{ij}\beta_{\tau j} + \int_0^1 \gamma_\tau(s)X_i(s)\,\mathrm{d}s + f_\tau(T_i), \tag{12}$$

where $Z_{i1} = 1$ for female and $Z_{i1} = 0$ for male, $Z_{i2} = -1$ denotes totally left-handed and $Z_{i2} = 1$ denotes totally right-handed, $Z_{i3} = 0, Z_{i4} = 0$ for typically developing children, $Z_{i3} = 1, Z_{i4} = 0$ for ADHD-combined and $Z_{i3} = 0, Z_{i4} = 1$ for ADHD-inattentative, $Z_{i5} = 1$ for medication naïve and $Z_{i5} = 0$ for not medication naïve. Since the value of functional predictor cerebelum is too small, $X_i(s)$ is chosen as the standardization of functional predictor cerebelum so that $E[X_i(s)] = 0$ and $E(\|X_i(s)\|^2) = 1$.

In order to examine the impacts of the covariates on the high or low ADHD index, we study quantile regression (12) under $\tau = 0.25, 0.5$ and 0.75. The quantile estimators of unknown parameters and functions in model (12) are computed by the procedure given in Section 2. Table 5 displays the estimated quantile coefficient parameters and Figure 1 shows the estimated quantile curves. We see from Table 5 that the scalar covariate effects on ADHD index under different τ's appear quite different in magnitude, but mostly are of the same signs. The ADHD index for male is greater than that for female. The ADHD index for typically developing children is much less than that for ADHD-combined and ADHD-inattentative children and the ADHD index for ADHD-inattentative children is less than that

Table 5. The parametric estimators for model (12).

τ	$\beta_{\tau 1}$	$\beta_{\tau 2}$	$\beta_{\tau 3}$	$\beta_{\tau 4}$	$\beta_{\tau 5}$	$\beta_{\tau 6}$	$\beta_{\tau 7}$	$\beta_{\tau 8}$
0.25	−3.6872	−1.6626	26.8228	22.2464	2.4665	0.3136	−0.6275	0.3445
0.5	−2.9848	1.2263	30.1245	23.4713	2.5050	0.2178	−0.4732	0.2705
0.75	−4.6426	−2.9909	33.2756	25.7632	5.8309	0.2859	−0.6324	0.4690

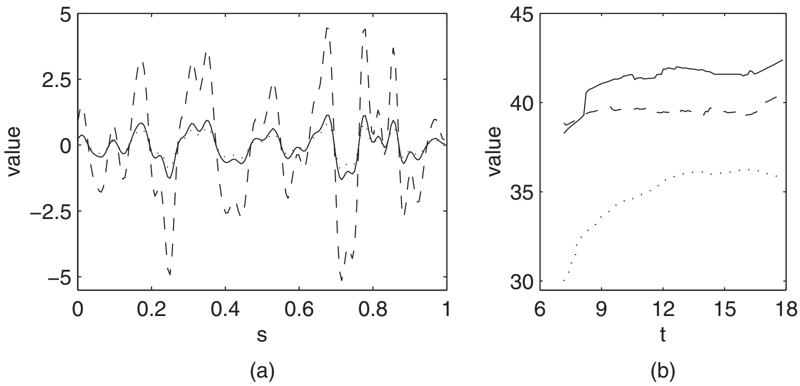

Figure 1. (a) is the estimated curve for $\gamma_\tau(s)$ and (b) is the estimated curve for $f_\tau(t)$ in model (12). - - - is the estimated curve for $\tau = 0.25$, — is the estimated curve for $\tau = 0.5$ and ... is the estimated curve for $\tau = 0.75$.

for ADHD-combined children. The ADHD index for medication naïve is greater than that for not medication naïve. Verbal IQ and Performance IQ have positive effects, while Full4 IQ has negative effect. Figure 1 shows that the slope curve $\gamma_\tau(s)$ and the curve $f_\tau(t)$ under different τ's appear similar shapes, but are different in magnitude. We see from Figure 1(b) that the ADHD index through the function $f_\tau(t)$ increases with age for small t and then tend to stable for large t, which shows that the ADHD index is nonlinearly related age. Figure 1(b) also shows that the stable point tends to increase as the quantile τ increases, for example, the stable point is about 9 for $\tau = 0.5$ and is about 12.5 for $\tau = 0.75$. These findings are helpful to uncover and understand the underlying relationship of the ADHD index with gender, age and so on.

To assess how well the model (12) fits the data, we consider the following model assessment tool by comparing the empirical distribution of Y with the simulated distribution from this model. We first generate τ from $U(0, 1)$. We randomly choose a observation from the data, denote by Y^* the Y of this observation. Let $\hat{\beta}^*_\tau$, $\hat{\gamma}^*_\tau(s)$ and $\hat{f}^*_\tau(t)$ be the estimated τth quantile estimators. The simulated

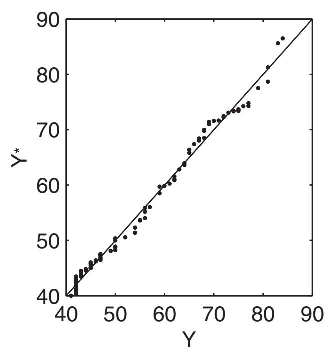

Figure 2. The Q–Q plot of the empirical sample against the model-based simulated data. The diagonal line is $y = x$.

Y^* is obtained by substituting the corresponding values of the chosen observation and the estimated quantile estimators into the model under assessment. Repeating this procedure many times, we can obtain a simulated sample. If the model fits data well, the marginal distribution of the simulated Y^* should match that of the observed Y . Figure 2 shows the Q–Q plots of the empirical Y and simulated Y^*. The Q–Q plot shows that the method proposed performs well and model (12) fits the data well.

Acknowledgments

We would like to thank an Associate Editor and two anonymous referees for constructive comments that helped to improve the quality of the paper. Dr Kong wanted to thank the support of the Canadian Statistical Sciences Institute Collaborative Research Team (CANSSI-CRT), the Program on Challenges in Computational Neuroscience (CCNS) at the Statistical and Applied Mathematical Sciences Institute (SAMSI) during his visit in 2016.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by National Social Science Foundation of China (16BTJ019) and Jiangsu Natural Science Foundation of China (BK20151481). Research of L. Kong was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Statistical Sciences Institute Collaborative Research Team (CANSSI-CRT). He also wanted to the support of the Program on Challenges in Computational Neuroscience (CCNS) at the Statistical and Applied Mathematical Sciences Institute (SAMSI) during his visit in 2016.

References

[1] Brunel E, Roche A. Penalized contrast estimation in functional linear models with circular data. Statistics. 2015;49:1298–1321.
[2] Cai TT, Hall P. Prediction in functional linear regression. Ann Statist. 2006;34:2159–2179.
[3] Cardot H, Crambes C, Sarda P. Quantile regression when the covariates are functions. J Nonparametric Statist. 2005;17:841–856.
[4] Hall P, Horowitz JL. Methodology and convergence rates for functional linear regression. Ann Statist. 2007;35:70–91.
[5] Ramsay JO, Silverman BW. Applied functional data analysis: methods and case studies. New York: Springer; 2002.
[6] Ramsay JO, Silverman BW. Functional data analysis. New York: Springer; 2005.
[7] Reiss PT, Ogden RT. Functional generalized linear models with images as predictors. Biometrics. 2010;66:61–69.
[8] Shin H. Partial functional linear regression. J Statist Plann Inference. 2009;139:3405–3418.
[9] Yao F, Müller H-G, Wang J-L. Functional linear regression analysis for longitudinal data. Ann Statist. 2005;33:2873–2903.
[10] Tang Q. Estimation for semi-functional linear regression. Statistics. 2015;49:1262–1278.
[11] Engle RF, Granger CWJ, Rice J, Weiss A. Semiparametric estimates of the relation between weather and electricity sales. J Amer Statist Assoc. 1986;81:310–320.
[12] Heckman NE. Spline smoothing in a partly linear model. J R Statist Soc B. 1986;48:244–248.
[13] Carroll RJ, Fan J, Gijbels I, Wand M. Generalized partially linear single-index models. J Amer Statist Assoc. 1997;92:477–489.
[14] Zhang D, Lin X, Raz J, Sowers M. Semiparametric stochastic mixed models for longitudinal data. J Amer Statist Assoc. 1998;93:710–719.
[15] He X, Zhu Z, Fung W. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. Biometrika. 2002;89:579–590.
[16] Fan J, Li R. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. J Amer Statist Assoc. 2004;99:710–723.
[17] Chen K, Jin Z. Partial linear regression models for clustered data. J Amer Statist Assoc. 2006;101:195–204.
[18] Tang Q. L_1-estimation in a semiparametric model with longitudinal data. J Statist Plann Infer. 2010;140:393–405.
[19] Yu D, Kong L, Mizera I. Partial functional linear quantile regression for neuroimaging data analysis. Neurocomputing. 2016;195:74–87.
[20] Ratcliffe SJ, Leader LR, Heller GZ. Functional data analysis with application to periodically stimulated foetal heart rate data. I: Functional regression. Stat Med. 2002;21:1103–1114.
[21] Shin H, Lee MH. On prediction rate in partial functional linear regression. J Multi Anal. 2012;103:93–106.
[22] Tang Q, Cheng L. Partial functional linear quantile regression. Sci China Math. 2014;57(12):2589–2608.

[23] Zhang D, Lin X, Sowers MF. Two-stage functional mixed models for evaluating the effect of longitudinal covariate profiles on a scalar outcome. Biometrics. 2007;63:351–362.
[24] Koenker R. Quantile regression. Cambridge: Cambridge University Press; 2005.
[25] Cai Z, Xu X. Nonparametric quantile estimations for dynamic smooth coefficient models. J Amer Statist Assoc. 2009;104:371–383.
[26] Chen K, Müller H-G. Conditional quantile analysis when covariates are functions, with application to growth data. J R Statist Soc B. 2012;74:67–89.
[27] Kai B, Li R, Zou H. New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. Ann Statist. 2011;39:305–332.
[28] Kato K. Estimation in functional linear quantile regression. Ann Statist. 2012;40:3108–3136.
[29] Li M, Wang A, Maity A, Staicu A-M, Inference in functional linear quantile regression. (2016). arXiv:1602.0879.

Appendix. Proofs

Let $C > 0$ denote a generic constant of which the value may change from line to line. Let $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{im})^{\mathrm{T}}$, $\hat{\boldsymbol{\eta}}_i = (\hat{\eta}_{i1}, \dots, \hat{\eta}_{im})^{\mathrm{T}}$, $\boldsymbol{\gamma}_\tau = (\gamma_{\tau 1}, \dots, \gamma_{\tau m})^{\mathrm{T}}$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m)^{\mathrm{T}}$, $D_i = \sum_{j=m+1}^{\infty} \gamma_{\tau j} \eta_{ij}$ and

$$\boldsymbol{F}_\tau = (f_\tau(t_1), \dots, h_0^p f_\tau^{(p)}(t_1)/p!, \dots, f_\tau(t_{M_n}), \dots, h_0^p f_\tau^{(p)}(t_{M_n})/p!)^{\mathrm{T}},$$
$$\boldsymbol{F}_r = (f_r(t_1), \dots, h_0^p f_r{}^{(p)}(t_1)/p!, \dots, f_r(t_{M_n}), \dots, h_0^p f_r{}^{(p)}(t_{M_n})/p!)^{\mathrm{T}}, \quad r = 1, \dots, d.$$

Set $\boldsymbol{F} = (\boldsymbol{F}_1, \dots, \boldsymbol{F}_d)^{\mathrm{T}}$, $\boldsymbol{W}_{rm} = (w_{r1}, \dots, w_{rm})^{\mathrm{T}}$, $\boldsymbol{W}_m = (\boldsymbol{W}_{1m}, \dots, \boldsymbol{W}_{dm})^{\mathrm{T}}$, $S_{ir} = \sum_{j=m+1}^{\infty} w_{rj}\eta_{ij}$, $\boldsymbol{S}_i = (S_{i1}, \dots, S_{id})^{\mathrm{T}}$ and $\boldsymbol{A}_i = A(T_i)$, $f_\tau^*(t) = f_\tau(t) - A^{\mathrm{T}}(t)\boldsymbol{F}_\tau$, $f_r^*(t) = f_r(t) - A^{\mathrm{T}}(t)\boldsymbol{F}_r$, $\boldsymbol{F}_i^* = (f_1^*(T_i), \dots, f_d^*(T_i))^{\mathrm{T}}$. Then Equation (5) can be written as

$$\min \sum_{i=1}^{n} \rho_\tau((\boldsymbol{\epsilon}_i + \boldsymbol{W}_m \boldsymbol{\eta}_i + \boldsymbol{F}\boldsymbol{A}_i + \boldsymbol{S}_i + \boldsymbol{F}_i^*)^{\mathrm{T}}(\beta_\tau - \beta) + \boldsymbol{\eta}_i^{\mathrm{T}} \boldsymbol{\gamma}_\tau - \hat{\boldsymbol{\eta}}_i^{\mathrm{T}} \boldsymbol{\gamma} + \boldsymbol{A}_i^{\mathrm{T}}(\boldsymbol{F}_\tau - \theta) + D_i + f_\tau^*(T_i) + \varepsilon_{\tau i}).$$

Let $E_n = \sum_{i=1}^n g(0 \mid X_i, T_i)\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^{\mathrm{T}}$, $H = \mathrm{diag}(\kappa_1, \dots, \kappa_m)$, $\alpha_1 = E_n^{1/2}(\beta - \beta_\tau)$, $\alpha_2 = n^{1/2}H^{1/2}[(\boldsymbol{\gamma} - \boldsymbol{\gamma}_\tau) + \boldsymbol{W}_m^{\mathrm{T}}(\beta - \beta_\tau)]$, $\alpha_3 = (n/M_n)^{1/2}[(\theta - \boldsymbol{F}_\tau) + \boldsymbol{F}^{\mathrm{T}}(\beta - \beta_\tau)]$, $V_{i1} = E_n^{-1/2}\boldsymbol{\epsilon}_i$, $V_{i2} = n^{-1/2}H^{-1/2}\hat{\boldsymbol{\eta}}_i$, $V_{i3} = (M_n/n)^{1/2}\boldsymbol{A}_i$, and $\alpha = (\alpha_1^{\mathrm{T}}, \alpha_2^{\mathrm{T}}, \alpha_3^{\mathrm{T}})^{\mathrm{T}}$, $V_i = (V_{i1}^{\mathrm{T}}, V_{i2}^{\mathrm{T}}, V_{i3}^{\mathrm{T}})^{\mathrm{T}}$, $e_i = (\boldsymbol{\eta}_i - \hat{\boldsymbol{\eta}}_i)^{\mathrm{T}}\boldsymbol{\gamma}_\tau + (\boldsymbol{\eta}_i - \hat{\boldsymbol{\eta}}_i)^{\mathrm{T}}\boldsymbol{W}_m^{\mathrm{T}}(\beta_\tau - \beta) + D_i + (\boldsymbol{S}_i + \boldsymbol{F}_i^*)^{\mathrm{T}}(\beta_\tau - \beta) + f_\tau^*(T_i)$. Then we have the following new optimization problem:

$$\hat{\alpha} = \mathrm{Argmin}_\alpha \sum_{i=1}^{n} [\rho_\tau(e_i + \varepsilon_{\tau i} - V_i^{\mathrm{T}}\alpha) - \rho_\tau(e_i + \varepsilon_{\tau i})]. \tag{A1}$$

Obviously,

$$\hat{\alpha}_1 = E_n^{1/2}(\hat{\beta} - \beta_\tau), \quad \hat{\alpha}_2 = n^{1/2}H^{1/2}[(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_\tau) + \boldsymbol{W}_m^{\mathrm{T}}(\hat{\beta} - \beta_\tau)]. \tag{A2}$$

Lemma A.1: *There are positive constants K_1 and K_2 such that, except on an event whose probability tends to zero, all the eigenvalues of $\sum_{i=1}^n V_{i3}V_{i3}^{\mathrm{T}}$ fall between K_1 and K_2, and consequently, $\sum_{i=1}^n V_{i3}V_{i3}^{\mathrm{T}}$ is invertible.*

Proof: Observe that $\sum_{i=1}^n V_{i3}V_{i3}^{\mathrm{T}}$ can be denoted by $\mathrm{diag}(\Omega_1, \dots, \Omega_{M_n})$, where $\Omega_\nu = (\phi_{\nu kl})_{(p+1)\times(p+1)}$, $\phi_{\nu kl} = (M_n/n)\sum_{i=1}^n [(T_i - t_\nu)/h_0]^{k+l} I_{\{|T_i - t_\nu| \le h_0\}}$, $k, l = 0, 1, \dots, p$; $\nu = 1, 2, \dots, M_n$. Let $\tilde{\Omega}_\nu = (\tilde{\phi}_{\nu kl})_{(p+1)\times(p+1)}$, $\tilde{\phi}_{\nu kl} = ((t_2 - t_1)/2)\int_{|t|\le 1} t^{k+l} v(t_\nu + h_0 t)\,\mathrm{d}t$. Since for any $\varepsilon > 0$, by Assumption 5

$$\sum_{n=1}^{\infty}\sum_{\nu=1}^{m} P\{|\phi_{\nu kl} - \tilde{\phi}_{\nu kl}| > \varepsilon\} \le C\sum_{n=1}^{\infty}(nM_n^4 + n^2M_n^3)/(\varepsilon^4 n^4) < +\infty,$$

so by Borel–Cantelli 's lemma, it holds that have

$$\phi_{\nu kl} - \tilde{\phi}_{\nu kl} \to 0 \quad \text{a.s.,} \quad \nu = 1, \dots, M_n; \ k, l = 0, \dots, p. \tag{A3}$$

Let $\check{\Omega} = (\check{\phi}_{kl})_{(p+1)\times(p+1)}$ with $\check{\phi}_{kl} = \int_{|t|\le 1} t^{k+l}\,\mathrm{d}t$. It is easy to prove that $\check{\Omega}$ is positive definite. Hence, by assumption 1, there exist two positive constants K_1 and K_2 such that all the eigenvalues of $\tilde{\Omega}_\nu$, $\nu = 1, 2, \dots, M_n$ fall between K_1 and K_2. So Lemma A.1 follows from Equation (A3). ■

Lemma A.2: *Under Assumptions 1, 2, 5 and 7, it holds that*

$$m^{1/2}(\log n)\max_i \|V_i\| = o_p(1).$$

Proof: By Assumptions 1 and 7 and Lemma E.1 of Kato [28], it holds that

$$\max_i \|X_i\| = O_p(n^{1/4}), \quad \max_i \|\boldsymbol{\epsilon}_i\| = O_p(n^{1/4}), \quad \max_i |\eta_{ij}| = O_p(\kappa_j^{1/2} n^{1/4}) \tag{A4}$$

uniformly for $j = 1, 2, \ldots$. Since $E_n/n \to_P \varpi B$, then

$$\max_i \| V_{i1} \| = E_n^{-1/2} \max_i \|\boldsymbol{\epsilon}_i\| = O_p(n^{-1/4}). \tag{A5}$$

Note that

$$\| V_{i2} \|^2 \le 2n^{-1} \left[\sum_{j=1}^{m} \kappa_j^{-1} \eta_{ij}^2 + \sum_{j=1}^{m} \kappa_j^{-1} \langle X_i, \hat{\psi}_j - \psi_j \rangle^2 \right].$$

By (5.21) and (5.22) of Hall and Horowitz [4], it holds that $\|\hat{\psi}_j - \psi_j\|^2 = O_p(n^{-1}j^2)$ uniformly for $1 \le j \le m$. Hence, by Assumption 2 and (A4), we get $\max_i \| V_{i2} \| = O_p(m^{1/2}n^{-1/4} + m^{\alpha/2+3/2}n^{-3/4})$. Since $\|\boldsymbol{A}_i\|$ is bounded, we have $\max_i \| V_{i3} \| = O_p(M_n^{1/2} n^{-1/2})$. Hence, by Assumption 5, we obtain

$$m^{1/2}(\log n) \max_i \| V_i \| \le m^{1/2}(\log n) \max_i(\| V_{i1} \| + \| V_{i2} \| + \| V_{i3} \|) = o_p(1).$$

The proof of Lemma A.2 is finished. ∎

Denote $\mathcal{A} = \{(Z_i, X_i, T_i), \quad i = 1, \ldots, n\}$, $G_{ni}(\alpha) = \rho_\tau(e_i + \varepsilon_{\tau i} - V_i^{\mathrm{T}}\alpha) - \rho_\tau(e_i + \varepsilon_{\tau i})$, $G_n(\alpha) = \sum_{i=1}^n G_{ni}(\alpha)$, $\Lambda_{ni}(\alpha) = E(G_{ni}(\alpha) \mid \mathcal{A})$, $\Lambda_n(\alpha) = \sum_{i=1}^n \Lambda_{ni}(\alpha)$ and $\Upsilon_{ni}(\alpha) = G_{ni}(\alpha) - \Lambda_{ni}(\alpha) + V_i^{\mathrm{T}}\alpha\varphi_\tau(\varepsilon_{\tau i})$, where $\varphi_\tau(t) = \tau - I_{(t<0)}$ is the derivative of $\rho_\tau(t)$, $\Upsilon_n(\alpha) = \sum_{i=1}^n \Upsilon_{ni}(\alpha)$. Then, we have

$$G_n(\alpha) = \Lambda_n(\alpha) - \sum_{i=1}^{n} V_i^{\mathrm{T}}\alpha\varphi_\tau(\varepsilon_{\tau i}) + \Upsilon_n(\alpha). \tag{A6}$$

Lemma A.3: *Assume that Assumptions* 1–7 *hold, then, for any sufficient large L, we have*

$$\sup_{\|\alpha\| \le L} m^{-1}|\Upsilon_n(m^{1/2}\alpha)| = o_p(1).$$

Proof: Let $U_n = \sup_{\|\alpha\| \le L} |\Upsilon_{ni}(m^{1/2}\alpha)|$, then by Lemma A.2, it holds that

$$(\log n)U_n \le C(\log n)m^{1/2} \max_i |V_i^{\mathrm{T}}\alpha| \le \mathrm{CL}(\log n)m^{1/2} \max_i \| V_i \| = o_p(1). \tag{A7}$$

By Assumption 6, we get

$$\begin{aligned}
\sum_{i=1}^{n} \mathrm{Var}(\Upsilon_{ni}(m^{1/2}\alpha) \mid \mathcal{A}) &\le \sum_{i=1}^{n} E\left(\left\{\int_{e_i}^{e_i - m^{1/2}V_i^{\mathrm{T}}\alpha} [\varphi_\tau(\varepsilon_{\tau i} + t) - \varphi_\tau(\varepsilon_{\tau i})]\,\mathrm{d}t\right\}^2 |\mathcal{A}\right) \\
&\le \sum_{i=1}^{n} m^{1/2}|V_i^{\mathrm{T}}\alpha| \int_{e_i - m^{1/2}|V_i^{\mathrm{T}}\alpha|}^{e_i + m^{1/2}|V_i^{\mathrm{T}}\alpha|} E(I_{(-|t|<\varepsilon_{\tau i}<|t|)}|\mathcal{A})\,\mathrm{d}t \\
&\le Cm^{1/2} \max_i |V_i^{\mathrm{T}}\alpha| \sum_{i=1}^{n} g(0 \mid X_i, T_i)[e_i^2 + m(V_i^{\mathrm{T}}\alpha)^2][1 + o_p(1)].
\end{aligned} \tag{A8}$$

Let $\Delta = \hat{S} - S$. Since $\sup_{j \ge 1} |\hat{\kappa}_j - \kappa_j| \le |||\Delta||| = O_p(n^{-1/2})$, then by Assumption 2, we have

$$\tfrac{1}{2}\kappa_j[1 + o_p(1)] \le \hat{\kappa}_j \le \tfrac{3}{2}\kappa_j[1 + o_p(1)], \quad j = 1, \ldots, m. \tag{A9}$$

Using Assumption 6, Lemma A.1 and Equation (A9), we get

$$\sum_{i=1}^{n} g(0 \mid X_i, T_i)(V_i^{\mathrm{T}}\alpha)^2 \le 3\sum_{i=1}^{n} g(0|X_i, T_i)[(V_{i1}^{\mathrm{T}}\alpha_1)^2 + (V_{i2}^{\mathrm{T}}\alpha_2)^2 + (V_{i3}^{\mathrm{T}}\alpha_3)^2] \le C\|\alpha\|^2[1 + o_p(1)]. \tag{A10}$$

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

$$\sum_{i=1}^{n}[(\boldsymbol{\eta}_i - \hat{\boldsymbol{\eta}}_i)^{\mathrm{T}}\boldsymbol{\gamma}_\tau]^2 \le m\sum_{i=1}^{n}\sum_{j=1}^{m}(\eta_{ij} - \hat{\eta}_{ij})^2\gamma_{\tau j}^2$$

$$\le m\sum_{i=1}^{n}\|X_i\|^2\sum_{j=1}^{m}\|\psi_j - \hat{\psi}_j\|^2\gamma_{\tau j}^2 = O_p(m). \tag{A11}$$

Since $\beta - \beta_\tau = E_n^{-1/2}\alpha_1$ and $E_n^{-1/2} = O_p(n^{-1/2})$, similar to the proof of Equation (A11), we have

$$\sum_{i=1}^{n}[(\boldsymbol{\eta}_i - \hat{\boldsymbol{\eta}}_i)^{\mathrm{T}}\boldsymbol{W}_m^{\mathrm{T}}(\beta_\tau - \beta)]^2 = O_p(n^{-1}m) = o_p(m). \tag{A12}$$

By Assumptions 2, 3 and 5, we have $E(\sum_{i=1}^n D_i^2) = n\sum_{j=m+1}^{\infty}\gamma_{\tau j}^2\kappa_j \le Cnm^{-(a+2b)+1} \le Cm$ and $\sum_{i=1}^n[\boldsymbol{S}_i^{\mathrm{T}}(\beta_\tau - \beta)]^2 = O_p(n^{-1}m) = o_p(m)$. By Assumptions 4 and 5 and using the fact that $\tilde{p} \ge (a+2b-1)/2$, it holds that $\sum_{i=1}^n f_\tau^{*2}(T_i) = O_p(nM_n^{-2\tilde{p}}) = O_p(m)$ and $\sum_{i=1}^n[\boldsymbol{F}_i^{*\mathrm{T}}(\beta_\tau - \beta)]^2 = O_p(n^{-1}m) = o_p(m)$. Hence, by Assumption 6 and Equations (A11) and (A12), we get

$$\sum_{i=1}^{n}g(0 \mid X_i, T_i)e_i^2 = O_p(m). \tag{A13}$$

Set $K_n = \sum_{i=1}^n \sup_{\|\alpha\|\le L} Var(\Upsilon_{ni}(m^{1/2}\alpha) \mid \mathcal{A})$. Using Equations (A8), (A10) and (A13), we conclude that

$$K_n \le Cm^{3/2}\max_i |V_i^{\mathrm{T}}\alpha|[1 + o_p(1)] \le CLm^{3/2}\max_i \|V_i\|[1 + o_p(1)]. \tag{A14}$$

Set $\mathcal{D} = \{\alpha : \|\alpha\| \le L\}$. Let $|c| = \max_{1\le i\le m}|c_i|$ for a vector $c = (c_1, \ldots, c_m)^{\mathrm{T}}$. Let $\mathcal{D}$ be divided into J_n disjoint parts $\mathcal{D}_1, \ldots, \mathcal{D}_{J_n}$ such that for any $d_k \in \mathcal{D}_k$, $1 \le k \le J_n$ and any sufficient small $\epsilon > 0$, except on an event whose probability tends to zero,

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{D}_k}|\Upsilon_n(m^{1/2}\alpha) - \Upsilon_n(m^{1/2}d_k)| \\
&= \sup_{\alpha\in\mathcal{D}_k}\left|\sum_{i=1}^{n}\left(\int_{e_i - m^{1/2}V_i^{\mathrm{T}}d_k}^{e_i - m^{1/2}V_i^{\mathrm{T}}\alpha}[\varphi_\tau(\varepsilon_{\tau i} + t) - \varphi_\tau(\varepsilon_{\tau i})]\,\mathrm{d}t\right.\right. \\
&\quad \left.\left. - E\left(\int_{e_i - m^{1/2}V_i^{\mathrm{T}}d_k}^{e_i - m^{1/2}V_i^{\mathrm{T}}\alpha}[\varphi_\tau(\varepsilon_{\tau i} + t) - \varphi_\tau(\varepsilon_{\tau i})]\,\mathrm{d}t \mid \mathcal{A}\right)\right)\right| \\
&\le 4c_0^{-1}\sup_{\alpha\in\mathcal{D}_k}\sum_{i=1}^{n}m^{1/2}g(0 \mid X_i, T_i)|V_i^{\mathrm{T}}(\alpha - d_k)| \\
&\le C\sup_{\alpha\in\mathcal{D}_k}m^{1/2}n^{1/2}\left(\sum_{i=1}^{n}f(0|X_i)(V_i^{\mathrm{T}}(\alpha - d_k))^2\right)^{1/2} \\
&\le C\sup_{\alpha\in\mathcal{D}_k}m^{1/2}n^{1/2}\|\alpha - d_k\| \le C\sup_{\alpha\in\mathcal{D}_k}mn^{1/2}|\alpha - d_k| < \epsilon/2,
\end{aligned}$$

where the third inequality follows from Equation (A10). This can be done with $J_n = (4CLn^{1/2}m/\epsilon)^{d+m+(p+1)M_n}$. Using Assumption 5, Equations (A7), (A14), Lemma A2 and the Bernstein inequality, we have

$$\begin{aligned}
P\left(\sup_{\|\alpha\|\le L}m^{-1}|\Upsilon_n(m^{1/2}\alpha)| \ge \epsilon \mid \mathcal{A}\right) &\le \sum_{k=1}^{J_n}P(|\Upsilon_n(m^{1/2}d_k)| \ge m\epsilon/2|\mathcal{A}) \\
&\le 2J_n\exp(-\epsilon^2m^2/(8K_n + 4m\epsilon U_n)) = o_p(1).
\end{aligned}$$

Therefore

$$P\left(\sup_{\|\alpha\|\le L}m^{-1}|\Upsilon_n(m^{1/2}\alpha)| \ge \epsilon\right) = o(1).$$

This finishes the proof of Lemma A.3. ■

Lemma A.4: *Suppose that assumptions* 1–8 *hold, then we have*

$$\|\hat{\alpha}\| = O_p(m^{1/2}).$$

Proof: By Assumption 6, Equation (A9) and Lemma A.1, we have

$$\begin{aligned}\Lambda_n(m^{1/2}\alpha) &= \sum_{i=1}^{n}\int_{e_i}^{e_i - m^{1/2}V_i^{\mathrm{T}}\alpha} E(\varphi_\tau(\varepsilon_{\tau i}+t)\mid \mathcal{A})\,\mathrm{d}t \\ &= \frac{1}{2}\sum_{i=1}^{n} g(0\mid X_i, T_i)[(e_i - m^{1/2}V_i^{\mathrm{T}}\alpha)^2 - e_i^2][1+o_p(1)] \\ &\geq c_0 m[c_0^*\|\alpha\|^2 + \sum_{i=1}^{n}(V_{i1}^{\mathrm{T}}\alpha_1 V_{i2}^{\mathrm{T}}\alpha_2 + V_{i1}^{\mathrm{T}}\alpha_1 V_{i3}^{\mathrm{T}}\alpha_3 \\ &\quad + V_{i2}^{\mathrm{T}}\alpha_2 V_{i3}^{\mathrm{T}}\alpha_3) - m^{-1}\sum_{i=1}^{n} e_i^2][1+o_p(1)], \end{aligned} \tag{A15}$$

where c_0^* is a positive constant. Let $\bar{V}_{i2} = n^{-1/2}H^{-1/2}\boldsymbol{\eta}_i$. Using Assumptions 1 and 5 and the fact that $\|\boldsymbol{A}_i\|$ is bounded, we get

$$E\left(\left[\sum_{i=1}^{n}\bar{V}_{i2}^{\mathrm{T}}\alpha_2 V_{i3}^{\mathrm{T}}\alpha_3\right]^2\right) \leq M_n\|\alpha_2\|^2\|\alpha_3\|^2 E(\|\bar{V}_{i2}\|^2\|\boldsymbol{A}_i\|^2) \leq Cn^{-1}mM_n = o(1).$$

Similar to the proof of Equation (A11) and using Assumption 6, we get

$$\left|\sum_{i=1}^{n}(V_{i2}-\bar{V}_{i2})^{\mathrm{T}}\alpha_2 V_{i3}^{\mathrm{T}}\alpha_3\right| \leq \left(\sum_{i=1}^{n}\|V_{i2}-\bar{V}_{i2}\|^2\right)^{1/2}\|\alpha_2\|\left[\sum_{i=1}^{n}(V_{i3}^{\mathrm{T}}\alpha_3)^2\right]^{1/2} = o_p(1).$$

Hence $\sum_{i=1}^{n} V_{i2}^{\mathrm{T}}\alpha_2 V_{i3}^{\mathrm{T}}\alpha_3 = o_p(1)$. Similarly, $\sum_{i=1}^{n} V_{i1}^{\mathrm{T}}\alpha_1 V_{i2}^{\mathrm{T}}\alpha_2 = o_p(1)$. Note that $\sum_{i=1}^{n} V_{i1}^{\mathrm{T}}\alpha_1 V_{i3}^{\mathrm{T}}\alpha_3 = n^{-1/2}M_n^{1/2}E_n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^{\mathrm{T}}\alpha_1\boldsymbol{A}_i^{\mathrm{T}}\alpha_3$ and $E([\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^{\mathrm{T}}\alpha_1\boldsymbol{A}_i^{\mathrm{T}}\alpha_3]^2) = O(n)$. Hence $\sum_{i=1}^{n} V_{i1}^{\mathrm{T}}\alpha_1 V_{i3}^{\mathrm{T}}\alpha_3 = O_p(n^{-1/2}M_n^{1/2}) = o_p(1)$ and by Equations (A15), (A13), for sufficient large L, it holds that

$$\inf_{\|\alpha\|=L}\Lambda_n(m^{1/2}\alpha) \geq c_0c_0^*L^2m[1+o_p(1)]. \tag{A16}$$

Since

$$\begin{aligned} E\left(\left(m^{1/2}\sum_{i=1}^{n}(V_i^{\mathrm{T}}\alpha)\varphi_\tau(\varepsilon_{\tau i})\right)^2 \mid \mathcal{A}\right) &\leq 3m[\sum_{i=1}^{n}[(V_{i1}^{\mathrm{T}}\alpha_1)^2 + (V_{i2}^{\mathrm{T}}\alpha_2)^2 + (V_{i3}^{\mathrm{T}}\alpha_3)^2] \\ &\leq Cm\|\alpha\|^2[1+o_p(1)]. \end{aligned}$$

Hence, $\sup_{\|\alpha\|\leq L} m^{1/2}|\sum_{i=1}^{n}(V_i^{\mathrm{T}}\alpha)\varphi_\tau(\varepsilon_{\tau i})| = O_p(m^{1/2})$. Combining Equations (A6), (A16) and Lemma A.3, for sufficiently large L, we have

$$\inf_{\|\alpha\|=L} G_n(m^{1/2}\alpha) \geq c_0c_0^*L^2m[1+o_p(1)],$$

which implies, by the convexity of ρ_τ, that

$$P\left(\inf_{\|\alpha\|\geq L}\left(\sum_{i=1}^{n}[\rho_\tau(e_i+\varepsilon_{\tau i} - m^{1/2}V_i^{\mathrm{T}}\alpha) - \rho_\tau(e_i+\varepsilon_{\tau i})]\right) > 0\right) \to 1.$$

Hence, $P(\|\hat{\alpha}\| \leq Lm^{1/2}) \to 1$. This completes the proof of Lemma A.4. ■

***Proof of Theorem* 3.1.:** Let $\tilde{\alpha}_1 = \sum_{i=1}^{n} V_{i1}\varphi_\tau(\varepsilon_{\tau i})$. By central limit theorem, we have $\tilde{\alpha}_1 \to_d N(0, (\tau(1-\tau)/\varpi)I_d)$, where I_d is a $d\times d$ identity matrix. By Equation (A2), to prove Equation (8), it suffices to prove that for any $\epsilon > 0$,

$P(\|\hat{\alpha}_1 - \tilde{\alpha}_1\| < \epsilon) \to 1$. We need only to show that

$$P\left\{ \inf_{\|\alpha_1-\tilde{\alpha}_1\|\geq\epsilon} \sum_{i=1}^{n} (\rho_\tau(e_i + \varepsilon_{\tau i} - V_{i1}^{\mathrm{T}}\alpha_1 - V_{i2}^{\mathrm{T}}\hat{\alpha}_2 - V_{i3}^{\mathrm{T}}\hat{\alpha}_3) \right.$$
$$\left. - \rho_\tau(e_i + \varepsilon_{\tau i} - V_{i1}^{\mathrm{T}}\tilde{\alpha}_1 - V_{i2}^{\mathrm{T}}\hat{\alpha}_2 - V_{i3}^{\mathrm{T}}\hat{\alpha}_3)) > 0 \right\} \to 1.$$

Set $\tilde{G}_n(\alpha_1, \alpha_2, \alpha_3) = \sum_{i=1}^n (\rho_\tau(e_i + \varepsilon_{\tau i} - V_{i1}^{\mathrm{T}}\alpha_1 - V_{i2}^{\mathrm{T}}\alpha_2 - V_{i3}^{\mathrm{T}}\alpha_3) - \rho_\tau(e_i + \varepsilon_{\tau i} - V_{i2}^{\mathrm{T}}\alpha_2 - V_{i3}^{\mathrm{T}}\alpha_3))$, By Lemma A.4, $\|\hat{\alpha}_2\| = O_p(m^{1/2})$ and $\|\hat{\alpha}_3\| = O_p(m^{1/2})$. Since $\|\tilde{\alpha}_1\| = O_p(1)$, so by the convexity of ρ_τ, it suffices to show that for any sufficient large $L > 0, L' > 0$

$$P\left(\left\{ \inf_{\|\alpha_1-\tilde{\alpha}_1\|=\epsilon, \|\alpha_2\|\leq Lm^{1/2}, \|\alpha_3\|\leq Lm^{1/2}} (\tilde{G}_n(\alpha_1, \alpha_2, \alpha_3) - \tilde{G}_n(\tilde{\alpha}_1, \alpha_2, \alpha_3)) > 0 \right\} \cap \{\|\tilde{\alpha}_1\| \leq L'\}\right) \to 1. \tag{A17}$$

Set $\tilde{\Lambda}_n(\alpha_1, \alpha_2, \alpha_3) = E(\tilde{G}_n(\alpha_1, \alpha_2, \alpha_3) \mid \mathcal{A})$, and $\tilde{\Upsilon}_n(\alpha_1, \alpha_2, \alpha_3) = \tilde{G}_n(\alpha_1, \alpha_2, \alpha_3) - \tilde{\Lambda}_n(\alpha_1, \alpha_2, \alpha_3) + \sum_{i=1}^n V_{i1}^{\mathrm{T}}\alpha_1\varphi_\tau(\varepsilon_{\tau i})$. Then

$$\tilde{G}_n(\alpha_1, \alpha_2, \alpha_3) = \tilde{\Lambda}_n(\alpha_1, \alpha_2, \alpha_3) - \sum_{i=1}^{n} V_{i1}^{\mathrm{T}}\alpha_1\varphi_\tau(\varepsilon_{\tau i}) + \tilde{\Upsilon}_n(\alpha_1, \alpha_2, \alpha_3). \tag{A18}$$

Using arguments similar to those used in the proof of Equation (A16), we get

$$\begin{aligned}\tilde{\Lambda}_n(\alpha_1, \alpha_2, \alpha_3) &= \sum_{i=1}^{n} g(0 \mid X_i, T_i) \left[\frac{1}{2}(V_{i1}^{\mathrm{T}}\alpha_1)^2 + V_{i1}^{\mathrm{T}}\alpha_1(V_{i2}^{\mathrm{T}}\alpha_2 + V_{i3}^{\mathrm{T}}\alpha_3 - e_i)\right][1 + o_p(1)] \\ &= \frac{1}{2}\|\alpha_1\|^2 + o_p(1).\end{aligned} \tag{A19}$$

By Equations (A18) and (A19), we obtain

$$\tilde{G}_n(\alpha_1, \alpha_2, \alpha_3) = \tfrac{1}{2}(\alpha_1 - \tilde{\alpha}_1)^{\mathrm{T}}(\alpha_1 - \tilde{\alpha}_1) - \tfrac{1}{2}\tilde{\alpha}_1^{\mathrm{T}}\tilde{\alpha}_1 + \tilde{\Upsilon}_n(\alpha_1, \alpha_2, \alpha_3) + o_p(1). \tag{A20}$$

By Equation (A20), we have $\tilde{G}_n(\tilde{\alpha}_1, \alpha_2, \alpha_3) = -\frac{1}{2}\tilde{\alpha}_1^{\mathrm{T}}\tilde{\alpha}_1 + \tilde{\Upsilon}_n(\tilde{\alpha}_1, \alpha_2, \alpha_3) + o_p(1)$. Hence, using the fact that $\|\alpha_1 - \tilde{\alpha}_1\| = \epsilon$, we obtain that

$$\tilde{G}_n(\alpha_1, \alpha_2, \alpha_3) - \tilde{G}_n(\tilde{\alpha}_1, \alpha_2, \alpha_3) \geq \frac{1}{2}\epsilon^2 - 2 \sup_{\|\alpha_1\|\leq L', \|\alpha_2\|\leq Lm^{1/2}, \|\alpha_3\|\leq Lm^{1/2}} |\tilde{\Upsilon}_n(\alpha_1, \alpha_2, \alpha_3)| + o_p(1).$$

Using arguments similar to those in the proof of Lemma A.3, it holds that

$$\sup_{\|\alpha_1\|\leq L', \|\alpha_2\|\leq Lm^{1/2}, \|\alpha_3\|\leq Lm^{1/2}} |\tilde{\Upsilon}_n(\alpha_1, \alpha_2, \alpha_3)| = o_p(1).$$

Now Equation (A17) follows and the proof of Theorem 3.1 is finished. ■

***Proof of Theorem* 3.2.:** Using Lemma A.4, we obtain $\|\hat{\alpha}_1\| = O_p(m^{1/2})$ and $\|\hat{\alpha}_2\| = O_p(m^{1/2})$. Using Equation (A2), we get $\|\hat{\beta} - \beta_\tau\| = O_p(n^{-1/2}m^{1/2})$. By assumption 2, we have $\|H^{1/2}\boldsymbol{W}_m^{\mathrm{T}}\| = O_p(1)$. Hence, $n^{1/2}\|H^{1/2}\boldsymbol{W}_m^{\mathrm{T}}(\hat{\beta} - \beta_\tau)\| = O_p(m^{1/2})$. So by Equation (A2), it holds that $n^{1/2}\|H^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_\tau)\| = O_p(m^{1/2})$. Using assumptions 2, 3 and 5 and using the fact that $\|\hat{\psi}_j - \psi_j\|^2 = O_p(n^{-1}j^2)$ uniformly for $1 \leq j \leq m$, we obtain that

$$\begin{aligned}\int_0^1 [\hat{\gamma}(s) - \gamma_\tau(s)]^2\,\mathrm{d}s &\leq 3\sum_{j=1}^{m}(\hat{\gamma}_j - \gamma_{\tau j})^2 + 3\int_0^1 \left[\sum_{j=1}^{m} \gamma_{\tau j}(\hat{\psi}_j(s) - \psi_j(s))\right]^2 \mathrm{d}s + 3\sum_{j=m+1}^{\infty} \gamma_{\tau j}^2 \\ &\leq 3n^{-1}\kappa_m^{-1}\|H^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_\tau)\|^2 + 3m\sum_{j=1}^{m}\gamma_{\tau j}^2\|\hat{\psi}_j - \psi_j\|^2 + 3\sum_{j=m+1}^{\infty}\gamma_{\tau j}^2 \\ &= O_p(n^{-1}m^{a+1} + n^{-1}m + m^{-2b+1}) = O_p(n^{-(2b-1)/(a+2b)}).\end{aligned}$$

This completes the proof of Theorem 3.2. ■

***Proof of Theorem* 3.3.:** By Taylor expansion, we have $f_\tau(T_i) = f_\tau(t_0) + f_\tau'(t_0)(T_i - t_0) + \frac{1}{2}f_\tau''(\xi_i)(T_i - t_0)^2$ for $|T_i - t_0| \leq Mh$, where $|\xi_i - t_0| < |T_i - t_0|$. Let $R_i = (nh)^{-1/2}(1, h^{-1}(T_i - t_0))^{\mathrm{T}}$, $\boldsymbol{u} = (nh)^{1/2}(a_0 - f_\tau(t_0), h(a_1 - f_\tau'(t_0)))^{\mathrm{T}}$,

$\tilde{V}_{i1} = n^{-1/2}Z_i$, $q_1 = n^{1/2}(\beta - \beta_\tau)$, $q_2 = n^{1/2}H^{1/2}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_\tau)$ and $\tilde{e}_i = (\boldsymbol{\eta}_i - \hat{\boldsymbol{\eta}}_i)^{\mathrm{T}}\boldsymbol{\gamma}_\tau + D_i + \frac{1}{2}f_\tau''(\xi_i)(T_i - t_0)^2$. We consider the following new optimization problem:

$$\min_{\boldsymbol{u}} \sum_{i=1}^{n}[\rho(\varepsilon_{\tau i} + \tilde{e}_i - R_i^{\mathrm{T}}\boldsymbol{u} - \tilde{V}_{i1}^{\mathrm{T}}\hat{q}_1 - V_{i2}^{\mathrm{T}}\hat{q}_2) - \rho(\varepsilon_{\tau i} + \tilde{e}_i - \tilde{V}_{i1}^{\mathrm{T}}\hat{q}_1 - V_{i2}^{\mathrm{T}}\hat{q}_2)]K\left(\frac{T_i - t_0}{h}\right).$$

We then have

$$\hat{\boldsymbol{u}} = (nh)^{1/2}(\hat{a}_0 - f_\tau(t_0), h(\hat{a}_1 - f_\tau'(t_0)))^{\mathrm{T}}. \tag{A21}$$

Let $G_{ni}^*(\boldsymbol{u}, q_1, q_2) = \rho(\varepsilon_{\tau i} + \tilde{e}_i - R_i^{\mathrm{T}}\boldsymbol{u} - \tilde{V}_{i1}^{\mathrm{T}}q_1 - V_{i2}^{\mathrm{T}}q_2) - \rho(\varepsilon_{\tau i} + \tilde{e}_i - \tilde{V}_{i1}^{\mathrm{T}}q_1 - V_{i2}^{\mathrm{T}}q_2)]K((T_i - t_0)/h)$, $G_n^*(\boldsymbol{u}, q_1, q_2) = \sum_{i=1}^{n} G_{ni}^*(\boldsymbol{u}, q_1, q_2)$ and $\Lambda_n^*(\boldsymbol{u}, q_1, q_2) = \sum_{i=1}^{n} E(G_{ni}^*(\boldsymbol{u}, q_1, q_2) \mid Z_i, X_i, T_i)$. Using arguments similar to those used in the proof of Equation (A16), we have

$$\begin{aligned}\Lambda_n^*(\boldsymbol{u}, q_1, q_2) = \sum_{i=1}^{n} g(0 \mid X_i, T_i)\Bigg[&\frac{1}{2}(R_i^{\mathrm{T}}\boldsymbol{u})^2 \\ &+ R_i^{\mathrm{T}}\boldsymbol{u}(\tilde{V}_{i1}^{\mathrm{T}}q_1 + V_{i2}^{\mathrm{T}}q_2 - \tilde{e}_i)\Bigg]K\left(\frac{T_i - t_0}{h}\right) + o_p(1).\end{aligned} \tag{A22}$$

Since

$$\sum_{i=1}^{n}|(\boldsymbol{\eta}_i - \hat{\boldsymbol{\eta}}_i)^{\mathrm{T}}\boldsymbol{\gamma}_\tau| \le m^{1/2}\sum_{i=1}^{n}\|X_i\|\left(\sum_{j=1}^{m}\|\psi_j - \hat{\psi}_j\|^2\gamma_{\tau j}^2\right)^{1/2} = O_p(m^{1/2}),$$

then by assumption 8, it holds that $\sum_{i=1}^{n} g(0 \mid X_i, T_i)(\boldsymbol{\eta}_i - \hat{\boldsymbol{\eta}}_i)^{\mathrm{T}}\boldsymbol{\gamma}_\tau R_i K((T_i - t_0)/h) = o_p(1)$. By assumptions 5 and 7, we obtain

$$(nh)^{-1/2}E\left|\sum_{i=1}^{n} g(0 \mid X_i, T_i)D_i\right| \le C(nh)^{-1/2}n(E(D_i^2))^{1/2} \le C(nh)^{-1/2}nm^{-(a+2b)+1} = o(1).$$

Hence, by the law of large numbers, we get

$$\sum_{i=1}^{n} g(0 \mid X_i, T_i)(R_i^{\mathrm{T}}\boldsymbol{u})^2 K\left(\frac{T_i - t_0}{h}\right) = v(t_0)\varrho(t_0)\boldsymbol{u}^{\mathrm{T}}\Gamma\boldsymbol{u} + o_p(1), \tag{A23}$$

$$\sum_{i=1}^{n} g(0|X_i, T_i)R_i^{\mathrm{T}}\boldsymbol{u}\tilde{e}_i K\left(\frac{T_i - t_0}{h}\right) = \frac{1}{2}(nh)^{1/2}h^2 v(t_0)\varrho(t_0)f_\tau''(t_0)\Theta^{\mathrm{T}}\boldsymbol{u} + o_p(1), \tag{A24}$$

where $\Gamma = \mathrm{diag}(\mu_0, \mu_2)$ and $\Theta = (\mu_2, 0)^{\mathrm{T}}$. By assumption 1, we have

$$\begin{aligned}&E\left[\left|\sum_{i=1}^{n} g(0 \mid X_i, T_i)R_i^{\mathrm{T}}\boldsymbol{u}\bar{V}_{i2}^{\mathrm{T}}q_2 K\left(\frac{T_i - t_0}{h}\right)\right|\right] \\ &\quad \le C(nh)^{-1/2}\sum_{i=1}^{n}\|\boldsymbol{u}\|E\left[(E((\bar{V}_{i2}^{\mathrm{T}}q_2)^2|T_i))^{1/2}K\left(\frac{T_i - u_0}{h}\right)\right] = O(h^{1/2}) = o(1).\end{aligned}$$

Since $\|V_{i2} - \bar{V}_{i2}\| \le n^{-1/2}\|X_i\|(\sum_{j=1}^{m}\kappa_j^{-1}\|\hat{\psi}_j - \psi_j\|^2)^{1/2}$, then by Assumption 5, we get

$$\begin{aligned}&\left|\sum_{i=1}^{n} g(0|X_i, T_i)R_i^{\mathrm{T}}\boldsymbol{u}(V_{i2} - \bar{V}_{i2})^{\mathrm{T}}q_2 K\left(\frac{T_i - t_0}{h}\right)\right| \\ &\quad \le Cn^{-1}h^{-1/2}\|\boldsymbol{u}\|\|q_2\|\left(\sum_{j=1}^{m}\kappa_j^{-1}\|\hat{\psi}_j - \psi_j\|^2\right)^{1/2}\sum_{i=1}^{n}\|X_i\|K\left(\frac{T_i - t_0}{h}\right) = o_p(1).\end{aligned}$$

Hence, $\sum_{i=1}^{n} g(0 \mid X_i, T_i) R_i^{\mathrm{T}} \boldsymbol{u} V_{i2}^{\mathrm{T}} q_2 K((T_i - t_0)/h) = o_p(1)$. Using the fact that

$$\sum_{i=1}^{n} g(0 \mid X_i, T_i) R_i^{\mathrm{T}} \boldsymbol{u} V_{i1}^{\mathrm{T}} q_1 K\left(\frac{T_i - t_0}{h}\right) = O_p(h^{1/2}) = o_p(1)$$

and combining Equations (A22)–(A24), we conclude that

$$\Lambda_n^*(\boldsymbol{u}, q_1, q_2) = \tfrac{1}{2} v(t_0)\varrho(t_0)\boldsymbol{u}^{\mathrm{T}}\Gamma\boldsymbol{u} - \tfrac{1}{2} n^{1/2}h^{5/2} v(t_0)\varrho(t_0) f_\tau''(t_0)\Theta^{\mathrm{T}}\boldsymbol{u} + o_p(1).$$

Let $\boldsymbol{u}^* = \frac{1}{2} n^{1/2}h^{5/2} f_\tau''(t_0)\Theta^*$ and $\tilde{\boldsymbol{u}} = \boldsymbol{u}^* + (1/v(t_0)\varrho(t_0))\Gamma^{-1}\sum_{i=1}^{n} R_i\varphi(\varepsilon_{\tau i})K((T_i - t_0)/h)$, where $\Theta^* = (\mu_0^{-1}\mu_2, 0)^{\mathrm{T}}$. Similar to the proof of Theorem 3.1, we have $\hat{\boldsymbol{u}} - \tilde{\boldsymbol{u}} = o_p(1)$. By central limit theorem, we have $\sum_{i=1}^{n} R_i\varphi(\varepsilon_{\tau i})K((T_i - t_0)/h) \to_d N(0, \tau(1-\tau)v(t_0)\tilde{\Gamma})$ with $\tilde{\Gamma} = \mathrm{diag}(\nu_0, \nu_2)$. Now Equation (10) follows from Equation (A21). The proof of Theorem 3.3 is finished. ■