Efficient estimation in the partially linear quantile regression model for longitudinal data

Seonjin Kim

Department of Statistics
Miami University, Oxford, OH 45056
e-mail: kims20@miamioh.edu

and

Hyunkeun Ryan Cho

Department of Biostatistics
University of Iowa, Iowa City, IA 52242
e-mail: hyunkeun-cho@uiowa.edu

Abstract: The focus of this study is efficient estimation in a quantile regression model with partially linear coefficients for longitudinal data, where repeated measurements within each subject are likely to be correlated. We propose a weighted quantile regression approach for time-invariant and time-varying coefficient estimation. The proposed approach can employ two types of weights obtained from an empirical likelihood method to account for the within-subject correlation: the global weight using all observations and the local weight using observations in the neighborhood of the time point of interest. We investigate the influence of choice of weights on asymptotic estimation efficiency and find theoretical results that are counter intuitive; it is essential to use the global weight for both time-invariant and time-varying coefficient estimation. This benefits from the within-subject correlation and prevents an adverse effect due to the weight discordance. For statistical inference, a random perturbation approach is utilized and evaluated through simulation studies. The proposed approach is also illustrated through a Multi-Center AIDS Cohort study.

Contents

1. Introduction

Efficient estimation has been attracting significant attention in parametric quantile regression (QR) models for longitudinal data; see [9, 6, 19, 5, 11, 20, 14, 3]. However, this has not been investigated in semiparametric or nonparametric QR models. In response to this gap in literature, this paper thoroughly investigates estimation efficiency in the partially linear quantile regression model; for $\tau \in (0,1)$, the τth conditional quantile of the jth response from subject i measured at time t_{ij}, denoted by $Y_i(t_{ij})$, is formulated as

$$Q_\tau\{Y_i(t_{ij})|X_i(t_{ij}), Z_i(t_{ij})\} = Z_i(t_{ij})^T\beta_\tau(t_{ij}) + X_i(t_{ij})^T\alpha_\tau \quad (1.1)$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$, where $m_1, \ldots, m_n$ are the number of measurements taken from n subjects and $Z_i(t_{ij}) = \{Z_{i1}(t_{ij}), \ldots, Z_{iq}(t_{ij})\}^T$ and $X_i(t_{ij}) = \{X_{i1}(t_{ij}), \ldots, X_{ip}(t_{ij})\}^T$ are a q-dimensional and p-dimensional covariate vector, respectively. We remark that model (1.1) allows us to explore efficient estimation in various quantile regression frameworks, such as the parametric QR model including only $X_i(t_{ij})^T\alpha_\tau$ and the time-varying coefficient QR model including only $Z_i(t_{ij})^T\beta_\tau(t_{ij})$. To fit the partially linear QR model for longitudinal data, [21] considers B-spline basis functions, yet it does not achieve estimation efficiency due to ignoring the within-subject correlation commonly existing in longitudinal studies.

This paper extends the empirical likelihood based weighted QR approach [19] to the partial linear QR framework. We develop a weighted marginal QR and a weighted local linear QR for estimation of time-varying parameters $\beta_\tau(t)$ and time-invariant parameters α_τ, respectively. Two types of weights are considered from the empirical likelihood: the global weight using correlations of all observations and the local weight using correlations of observations in the neighborhood of the time point of interest. For comparison purposes, we also consider the equal weight where the within-subject correlation is completely ignored in estimation. We have investigated the impact of the choice of weights and have obtained very interesting results that oppose to one's intuition in accommodating the local weight and global weight for efficient estimation of $\beta_\tau(t)$ and α_τ, respectively. According to our theoretical results, the estimator of α_τ obtained based on the aforementioned intuition is not always more efficient than that obtained by completely ignoring the within-subject correlation.

The key finding in our asymptotic analyses is as follows. Asymptotic estimation efficiency of α_τ can be gained when the global weight is used in the weighted

marginal QR for estimation of α_τ. On the other hand, estimation efficiency of $\beta_\tau(t)$ is not asymptotically affected by the choice of the weights in the sense that the limiting variances of the estimator of $\beta_\tau(t)$ are identical with the three weights considered. One may conclude from the foregoing result that the choice of weights in estimation of $\beta_\tau(t)$ is not crucial. However, it has a significant effect on the asymptotic estimation efficiency of α_τ, since a loss of estimation efficiency of α_τ has been found from the discordance between the weights used for estimation of $\beta_\tau(t)$ and α_τ. In summary, the global weight should be used in both estimation of $\beta_\tau(t)$ and α_τ to achieve estimation efficiency of α_τ; the amount of estimation efficiency for α_τ gained from the proposed approach corresponds to that obtained in [19] under the parametric QR model.

For statistical inference on $\beta_\tau(t)$ and α_τ, we implement the random perturbation approach [8]. This does not require estimation of the complicated limiting variance of the estimators. Through extensive simulation studies, the proposed inference approach accomplishes good empirical coverage probabilities very close to the nominal level in all cases under consideration. In addition, the simulation results confirm that estimation efficiency of α_τ can be achieved when the global weight is used for both estimation of $\beta_\tau(t)$ and α_τ.

This article is organized as follows. In Section 2, we introduce how to obtain empirical likelihood weights and propose the weighted QR approach in the partially linear QR model. In Section 3, we investigate asymptotic properties of the estimators of $\beta_\tau(t)$ and α_τ under various combinations of weights and propose the best strategy that leads to estimation efficiency of α_τ asymptotically. Section 4 presents a two-step procedure to implement the proposed approach and a cross-validation approach to choose the optimal bandwidth. In Section 5, we assess the finite sample performance of the proposed approach through simulation studies and apply this to the analysis of real-life data. Regularity conditions and proofs of theoretical results are provided in the Appendix.

2. Estimation procedures with auxiliary information

For the ease of presentation, we denote $Y_{ij} = Y_i(t_{ij})$, $Z_{ij} = Z_i(t_{ij})$, and $X_{ij} = X_i(t_{ij})$ in (1.1) and write $U_{ij} = (t_{ij}, Z_{ij}, X_{ij})$. If repeated measures within a subject are assumed independent, estimators of $\beta_\tau(t)$ and α_τ in (1.1) can be obtained as

$$\{\tilde{\beta}_\tau(t), \tilde{\beta}_\tau^*(t)\} = \operatorname*{argmin}_{\beta, \beta^*} \sum_{i=1}^n \sum_{j=1}^{m_i} \rho_\tau[\tilde{Y}_{ij}^* - Z_{ij}^T\{\beta + \beta^*(t_{ij} - t)\}] K\{(t_{ij} - t)/h\}, \quad (2.1)$$

and

$$\tilde{\alpha}_\tau = \operatorname*{argmin}_{\alpha} \sum_{i=1}^n \sum_{j=1}^{m_i} \rho_\tau\big(\tilde{Y}_{ij}^\# - X_{ij}^T\alpha\big), \quad (2.2)$$

where $\rho_\tau(u) = u\{\tau - \mathbf{1}(u < 0)\}$ is called the check function with an indicator function $\mathbf{1}(\cdot)$, $\tilde{Y}_{ij}^* = Y_{ij} - X_{ij}^T\tilde{\alpha}_\tau$ and $\tilde{Y}_{ij}^\# = Y_{ij} - Z_{ij}^T\tilde{\beta}_\tau(t_{ij})$ are partial quantile residuals, and $K(\cdot)$ is a kernel function with a bandwidth h. However, this

procedure ignores the within-subject correlation and therefore can cause a loss of estimation efficiency.

In order to account for the within-subject correlation, [19] proposed a weighted QR with subject-specific weights obtained from the empirical likelihood in the parametric mean regression model. Since these weights contain auxiliary information with regards to the within-subject correlation, estimation efficiency of regression quantiles is improved. Here, we adopt auxiliary information from the partial linear mean regression model. This is the counterpart of (1.1), $Y_{ij} = Z_{ij}^T \delta(t_{ij}) + X_{ij}^T \gamma + \epsilon_{ij}$, where γ is a p-dimensional vector of time-invariant parameters, $\delta(\cdot)$ is a q-dimensional vector of time-varying functions, and the error ϵ_{ij} satisfies the conditional variance $\sigma^2(U_{ij}) = E(\epsilon_{ij}^2|U_{ij})$ and the conditional mean $E(\epsilon_{ij}|U_{ij}) = 0$.

For efficient estimation of γ and $\delta(t_{ij})$, [13] developed kernel estimating equations and profile estimating equations involving a working correlation matrix of the repeated measures, denoted by R_i. Following the matrix expansion idea of quadratic inference functions [17], i.e., $R_i^{-1} = \sum_{\ell=1}^s b_\ell B_{i\ell}$ where $B_{i1}, \ldots, B_{is}$ are known basis matrices and $b_1, \ldots, b_s$ are unknown coefficients, we extend the ith component in kernel estimating equations as $g_i\{\delta(t|\gamma)\} = (g_{i1}\{\delta(t|\gamma)\}, \ldots, g_{is}\{\delta(t|\gamma)\})^T =$

$$
\begin{pmatrix}
Z_i K_i(t)^{1/2} V_i^{-1/2} B_{i1} V_i^{-1/2} K_i(t)^{1/2} \{Y_i' - Z_i^T \delta(t|\gamma)\} \\
\vdots \\
Z_i K_i(t)^{1/2} V_i^{-1/2} B_{is} V_i^{-1/2} K_i(t)^{1/2} \{Y_i' - Z_i^T \delta(t|\gamma)\}
\end{pmatrix}
\tag{2.3}
$$

and the ith one in profile estimating equations as

$$
h_i(\gamma) =
\begin{pmatrix}
(X_i + Z_i^T \frac{\partial \delta_i}{\partial \gamma}) V_i^{-1/2} B_{i1} V_i^{-1/2} (Y_i'' - X_i^T \gamma) \\
\vdots \\
(X_i + Z_i^T \frac{\partial \delta_i}{\partial \gamma}) V_i^{-1/2} B_{is} V_i^{-1/2} (Y_i'' - X_i^T \gamma)
\end{pmatrix}
=
\begin{pmatrix}
h_{i1}(\gamma) \\
\vdots \\
h_{is}(\gamma)
\end{pmatrix}, \tag{2.4}
$$

where $Y_i' = Y_i - X_i^T \gamma$, $Y_i'' = Y_i - Z_i^T \delta_i$, $Y_i = (Y_{i1}, \ldots, Y_{im_i})^T$, $Z_i = (Z_{i1}, \ldots, Z_{im_i})$, $X_i = (X_{i1}, \ldots, X_{im_i})$, $\delta_i = \{\delta(t_{i1}|\gamma), \ldots, \delta(t_{im_i}|\gamma)\}^T$, $K_i(t) = \text{diag}[K\{(t_{i1} - t)/h_1\}, \ldots, K\{(t_{im_i} - t)/h_1\}]$, and $V_i = \text{diag}\{\sigma^2(U_{i1}), \ldots, \sigma^2(U_{im_i})\}$. In what follows, we omit conditioning γ in $\delta(t|\gamma)$, denoted by $\delta(t)$, if no confusion arises. See Remark 2.1 below for discussion about the choice of basis matrices.

To secure auxiliary information with regards to the within-subject correlation, two types of subject-specific weights can be provided as $L\{\delta(t)\} =$

$$
\max\left\{\prod_{i=1}^n w_i\{\delta(t)\} \,\middle|\, \sum_{i=1}^n w_i\{\delta(t)\} g_i\{\delta(t)\} = 0, \sum_{i=1}^n w_i\{\delta(t)\} = 1, 0 \le w_i\{\delta(t)\}\right\},
\tag{2.5}
$$

and

$$
L(\gamma) = \max\left\{\prod_{i=1}^n w_i(\gamma) \,\middle|\, \sum_{i=1}^n w_i(\gamma) h_i(\gamma) = 0, \sum_{i=1}^n w_i(\gamma) = 1, 0 \le w_i(\gamma)\right\}. \tag{2.6}
$$

Since $L\{\delta(t)\}$ and $L(\gamma)$ consider observations in the neighborhood of the time point of interest and all observations, respectively, their corresponding weights, $w_i\{\delta(t)\}$ and $w_i(\gamma)$, are called the local weight and global weight. Following [16], the optimal local weight is given $w_i\{\hat{\delta}(t)\} = n^{-1}[1 + g_i\{\hat{\delta}(t)\}^T\lambda_{\hat{\delta}(t)}]^{-1}$, where $\hat{\delta}(t) = \operatorname{argmax}_{\delta(t)} L\{\delta(t)\}$ and $\lambda_{\hat{\delta}(t)}$ satisfies $n^{-1}\sum_{i=1}^n g_i\{\hat{\delta}(t)\}[1 + g_i\{\hat{\delta}(t)\}^T \times \lambda_{\hat{\delta}(t)}]^{-1} = 0$. Similarly, the optimal global weight is $w_i(\hat{\gamma}) = n^{-1}\{1 + h_i(\hat{\gamma})^T\lambda_{\hat{\gamma}}\}^{-1}$ with $\hat{\gamma} = \operatorname{argmax}_{\gamma} L(\gamma)$ and $\lambda_{\hat{\gamma}}$ holding $n^{-1}\sum_{i=1}^n h_i(\hat{\gamma})\{1 + h_i(\hat{\gamma})^T\lambda_{\hat{\gamma}}\}^{-1} = 0$.

To incorporate the auxiliary information for estimation of regression quantiles in the partial linear QR model (1.1), we propose weighted quantile regressions for estimation of $\beta_\tau(t)$ and α_τ. With a given α_τ, the weighted local linear QR is formulated to estimate $\beta_\tau(t)$ as $\{\hat{\beta}_\tau(t|\alpha_\tau), \hat{\beta}_\tau^*(t|\alpha_\tau)\} =$

$$
\operatorname*{argmin}_{\beta,\beta^*} \sum_{i=1}^n w_i^n \sum_{j=1}^{m_i} \rho_\tau[Y_{ij}^* - Z_{ij}^T\{\beta + \beta^*(t_{ij} - t)\}]K\left(\frac{t_{ij} - t}{h}\right), \qquad (2.7)
$$

where $Y_{ij}^* = Y_{ij} - X_{ij}^T\alpha_\tau$ and w_i^n are weights used for time-varying coefficient estimation. Here, the superscript n in w_i^n stands for nonparametric estimation. Given $\hat{\beta}_\tau(t|\alpha_\tau)$ from (2.7), an estimator of α_τ is obtained by minimizing the weighted marginal QR as

$$
\hat{\alpha}_\tau = \operatorname*{argmin}_{\alpha} \sum_{i=1}^n w_i^p \sum_{j=1}^{m_i} \rho_\tau\left(Y_{ij}^{\#} - X_{ij}^T\alpha\right), \qquad (2.8)
$$

where $Y_{ij}^{\#} = Y_{ij} - Z_{ij}^T\hat{\beta}_\tau(t_{ij})$, w_i^p are weights used for time-invariant coefficient estimation, and the superscript p in w_i^p stands for parametric estimation.

The performance of our estimation method relies on the choice of two weights, w_i^n and w_i^p, that plays an important role to achieve estimation efficiency. With the local weight $w_i\{\hat{\delta}(t)\}$ and global weight $w_i(\hat{\gamma})$, four combinations for $\left(w_i^n, w_i^p\right)$ can be considered as $\left(w_i(\hat{\gamma}), w_i(\hat{\gamma})\right)$, $\left(w_i\{\hat{\delta}(t)\}, w_i\{\hat{\delta}(t)\}\right)$, $\left(w_i(\hat{\gamma}), w_i\{\hat{\delta}(t)\}\right)$, and $\left(w_i\{\hat{\delta}(t)\}, w_i(\hat{\gamma})\right)$. Intuitively, the last one would be a natural choice because they are obtained from the empirical likelihood for their counterpart in the mean regression model. We note that $w_i^n = w_i^p = 1/n$ is the case, where the within-subject correlation is completely ignored, corresponding to (2.1) and (2.2).

Remark 2.1. Basis matrices $B_{i1}, \ldots, B_{is}$ are determined by the type of working correlation structure. For example, if R_i is assumed a first-order autoregressive model denoted by AR(1), then R_i^{-1} is decomposed as $R_i^{-1} = b_1B_{i1} + b_2B_{i2} + b_3B_{i3}$, where B_{i1} is the identity matrix, B_{i2} is a symmetric matrix with 1 on the sub-diagonal entries and 0 elsewhere, and B_{i3} is a symmetric matrix with 1 in elements $(1,1)$ and (m_i, m_i) and 0 elsewhere. If R_i corresponds to a compound symmetry structure, two basis matrices B_{i1} and B_{i2} are required, where B_{i1} is an identity matrix and B_{i2} is a symmetric matrix with 0 on the diagonal and 1 elsewhere. See [17] for more details.

3. Theoretical studies

In this section, we investigate the impact of auxiliary information on asymptotic properties of $\hat{\beta}_\tau(t|\hat{\alpha}_\tau)$ and $\hat{\alpha}_\tau$ based on three weights: global weight $w_i(\hat{\gamma})$, local weight $w_i\{\hat{\delta}(t)\}$, and equal weight across subjects $w_i^e = 1/n$.

3.1. Asymptotic properties of $\hat{\beta}_\tau(t|\alpha_\tau)$

We assume that $U_i(t) = \{t, Z_i(t), X_i(t)\}$ are independent realizations of the process $U(t) = \{t, Z(t), X(t)\}$ and t_{ij} are independent and identically distributed with a probability density function $p_T(\cdot)$. Let $\xi_{ij}(\tau) = Y_{ij} - Q_\tau(Y_{ij}|X_{ij}, Z_{ij})$ and denote $f_{\xi_\tau}(\cdot|U_{ij})$ and $F_{\xi_\tau}(\cdot|U_{ij})$ by the conditional density and distribution functions of $\xi_{ij}(\tau)$ given U_{ij}, respectively. For simplicity of theoretical derivation, we let $m_i = m$ for all subjects, yet the theorems derived in this paper can be proved with different $m_i < \infty$ in a straightforward manner. Accordingly, we let $N = nm$, simplify R_i^{-1} as $R_i^{-1} = R^{-1} = \sum_{\ell=1}^s b_\ell B_\ell$, where $B_\ell = B_{1\ell} = \cdots = B_{n\ell}$, and denote the (i,j)th component of the ℓth basis matrix B_ℓ by b_{ij}^ℓ. We define $\boldsymbol{\Omega}(t) = \{\boldsymbol{\Gamma}^T(t)\boldsymbol{\Sigma}^{-1}(t)\boldsymbol{\Gamma}(t)\}^{-1}$, $\boldsymbol{\Sigma}(t) = \mathbf{B_1} \otimes \Sigma_\sigma(t)$, $\boldsymbol{\Gamma}(t) = \mathbf{B_2} \otimes \Sigma_\sigma(t)$, where $\mathbf{B_1}$ is a q by q symmetric matrix with $\sum_{j=1}^m b_{jj}^\ell b_{jj}^{\ell'}$ in the element (ℓ, ℓ'), $\mathbf{B_2}$ is a q-dimensional vector with $\sum_{j=1}^m b_{jj}^\ell$ in the ℓth element, $\Sigma_\sigma(t) = E[Z(t)Z(t)^T/\sigma^2\{U(t)\}]$, and $\otimes$ is the Kronecker product. We further define $\mathbf{M} = \boldsymbol{\Sigma}(t)^{-1}\{\mathbf{I} - \boldsymbol{\Gamma}(t)\boldsymbol{\Omega}(t)\boldsymbol{\Gamma}^T(t)\boldsymbol{\Sigma}^{-1}(t)\}$ with the identity matrix $\mathbf{I}$, and $\mathbf{L} = \boldsymbol{\Psi}^{-1}(\gamma_0)\{\mathbf{I} - \boldsymbol{\Upsilon}(\gamma_0)\boldsymbol{\Xi}(\gamma_0)\boldsymbol{\Upsilon}^T(\gamma_0)\boldsymbol{\Psi}^{-1}(\gamma_0)\}$ with $\boldsymbol{\Psi}(\gamma_0) = E\{h(\gamma_0)h(\gamma_0)^T\}$, $\boldsymbol{\Upsilon}(\gamma_0) = E\{\partial h(\gamma_0)/\partial\gamma\}$, and $\boldsymbol{\Xi}(\gamma_0) = \{\boldsymbol{\Upsilon}^T(\gamma_0)\boldsymbol{\Psi}^{-1}(\gamma_0)\boldsymbol{\Upsilon}(\gamma_0)\}^{-1}$, where γ_0 is the true value of γ.

Theorem 1. *Let* $\Sigma_{fZ}(t) = E[Z(t)Z(t)^T f_{\xi_\tau}\{0|U(t)\}]$, $\mu_K = \int_{\mathbb{R}} u^2 K(u)du$, *and* $\varphi_K = \int_{\mathbb{R}} K^2(u)du$. *If* $\hat{\alpha}_\tau$ *is* $\sqrt{n}$ *consistent, under conditions (A1)–(A5) and (B1)–(B3) in the Appendix, for* $\tau \in (0,1)$ *and any interior point t in the support of* $p_T(\cdot)$, *we have*

$$\sqrt{Nh}\left\{\hat{\beta}_\tau(t|\hat{\alpha}_\tau) - \beta_\tau(t) - \frac{\beta_\tau''(t)\mu_k}{2}h^2\right\} \xrightarrow{d} N\left(0, \frac{\varphi_k}{p_T(t)}\Sigma_{fZ}^{-1}(t)S(t)\Sigma_{fZ}^{-1}(t)\right). (3.1)$$

1. *If the equal weight w_i^e is used in (2.7) for estimation of $\beta_\tau(t)$, i.e., $w_i^n = w_i^e$, we have*

$$S(t) = \tau(1-\tau)\Sigma_Z(t), \tag{3.2}$$

    where $\Sigma_Z(t) = E\{Z(t)Z(t)^T\}$.
2. *If the local weight $w_i\{\hat{\delta}(t)\}$ is used in (2.7) for estimation of $\beta_\tau(t)$, i.e., $w_i^n = w_i\{\hat{\delta}(t)\}$, we have*

$$S(t) = \left\{\tau(1-\tau)\Sigma_Z(t) - \boldsymbol{\Lambda}_\tau(t)^T\mathbf{M}\boldsymbol{\Lambda}_\tau(t)\right\} = \tau(1-\tau)\Sigma_Z(t), \tag{3.3}$$

since $\mathbf{\Lambda}_\tau(t)^T \mathbf{M} \mathbf{\Lambda}_\tau(t) = 0$ always holds, regardless of the working correlation structure, where $\mathbf{\Lambda}_\tau(t) = \mathbf{B_2} \otimes \Lambda_\tau(t)$ and $\Lambda_\tau(t) = E[E\{(\tau - \mathbf{1}_{\xi<0})\epsilon|U(t)\}Z(t)Z(t)^T/\sigma^2\{U(t)\}]$.

3. If the global weight $w_i(\hat{\gamma})$ is used in (2.7) for estimation of $\beta_\tau(t)$, i.e., $w_i^n = w_i(\hat{\gamma})$, we have

$$S(t) = \tau(1-\tau)\Sigma_Z(t) - \frac{mhp_T(t)}{\varphi_k}\mathbf{G}(t)^T\mathbf{L}\mathbf{G}(t) \to \tau(1-\tau)\Sigma_Z(t), \quad (3.4)$$

where $\mathbf{G}(t)^T = \{G_1(t)^T, \ldots, G_s(t)^T\}^T$ with $G_\ell^T(t) = E[Z(t)E\{(\tau - \mathbf{1}_{\xi<0})h_{i\ell}(\gamma_0)^T|U(t)\}]$.

   Theorem 1 states that the asymptotic theory of $\hat{\beta}_\tau(t|\hat{\alpha}_\tau)$ does not depend on $\hat{\alpha}_\tau$, if $\hat{\alpha}_\tau$ is $\sqrt{n}$ consistent. Thus, in what follows we omit conditioning α_τ in $\hat{\beta}_\tau(t|\alpha_\tau)$. More importantly, the variances of $\hat{\beta}_\tau(t)$ under three different weights are asymptotically equal. The results also suggest that the auxiliary information through the empirical likelihood does not improve asymptotic estimation efficiency in the time-varying coefficient QR model, which is a simple case of model (1.1). The phenomenon that the additional term $\mathbf{\Lambda}_\tau(t)^T\mathbf{M}\mathbf{\Lambda}_\tau(t)$ in (3.3) is zero can be explained as follows: as bandwidth $h \to 0$, at most one measurement per subject is ultimately used in the estimation of $\beta_\tau(t)$ and hence, the working correlation structure does not make an impact on the asymptotic result. The use of the global weight also fails to yield an asymptotically more efficient estimator of $\beta_\tau(t)$ than the one assuming the independent correlation structure. This is because the convergence rate of the additional term generated by the global weight in (3.4) is faster than that of the variance of $\hat{\beta}_\tau(t)$.

3.2. Asymptotic properties of $\hat{\alpha}_\tau$

According to the asymptotic result as in Theorem 1, the choice of weights among the three considered weights seems to be trivial for estimation of $\beta_\tau(t)$. However, in the semiparametric regression, the asymptotic results of the time-invariant coefficient estimator is often influenced by those of the time-varying coefficient estimator. For example, the invariant coefficient estimator is often biased in the semiparametric regression unless undersmoothing is used [13, 1]. Therefore, the choice of weights for estimation of $\beta_\tau(t)$ may play an important role in the asymptotic properties of $\hat{\alpha}_\tau$. As shown in [19], the natural choice of weights w_i^p in (2.8) for estimation of α_τ is the global weight $w_i(\hat{\gamma})$. Thus, we first investigate asymptotic properties of $\hat{\alpha}_\tau$ with $w_i^p = w_i(\hat{\gamma})$ and discuss the impact of using other weights for estimation of α_τ later.

   To appreciate how the choice of weights w_i^n in (2.7) for estimation of $\beta_\tau(t)$ influences estimation efficiency of α_τ, its estimator is denoted by $\hat{\alpha}_\tau(w_i^n)$ in this section. We define $\mathcal{U} = E(X_i\mathcal{C}_iX_i^T)$, where $\mathcal{C}_i = (\mathcal{C}_{i,jk})_{j,k=1}^p$ is the conditional correlation matrix for $\mathcal{C}_{i,jk} = E(\tau - \mathbf{1}_{\xi_{ij}<0}, \tau - \mathbf{1}_{\xi_{ik}<0}|X_i)$, $H_{ij} = D(t_{ij})\Sigma_{fZ}^{-1}(t_{ij})Z_{ij}$, $D(t) = E[f_{\xi_\tau}\{0|U(t)\}X(t)Z(t)^T]$, $\Sigma_{fX} = \sum_{j=1}^m E\{f_{\xi_\tau}(0|U_{ij})X_{ij}X_{ij}^T\}$, $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_s)$, $\mathcal{A}_\ell = \sum_{j=1}^m E[(X_{ij} - H_{ij})\text{Cov}\{\tau - \mathbf{1}_{\xi_{ij}<0},$

$h_{i\ell}(\gamma_0)|U_{ij}\}]$, $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_s)$, and $\mathcal{B}_\ell = \sum_{j=1}^m E\big[H_{ij}\mathrm{Cov}\{\tau - \mathbf{1}_{\xi_{ij}<0},$ $h_{i\ell}(\gamma_0)|U_{ij}\}\big]$.

Theorem 2. *Assume that conditions (A1)–(A5) and (B1)–(B3) in the Appendix hold and the global weight $w_i(\hat\gamma)$ is used in (2.8) for estimation of α_τ, i.e., $w_i^p = w_i(\hat\gamma)$.*

1. *If the global weight $w_i(\hat\gamma)$ is used in (2.7) for estimation of $\beta_\tau(t)$, an asymptotic distribution of $\hat\alpha_\tau\{w_i(\hat\gamma)\}$ at $\tau \in (0,1)$ is*

$$\sqrt{n}\left[\hat\alpha_\tau\{w_i(\hat\gamma)\} - \alpha_\tau - \frac{h^2\mu_k}{2}\mathbf{b}_{\alpha_\tau}\right] \xrightarrow{d} N(0, \Sigma_{fX}^{-1}V\Sigma_{fX}^{-1}), \quad (3.5)$$

where $\mathbf{b}_{\alpha_\tau} = \sum_{j=1}^m D(t_{ij})\beta_\tau''(t_{ij})$, and $V = \tau(1-\tau)\mathcal{U} - \mathcal{A}\mathbf{L}\mathcal{A}$.

2. *If the local weight $w_i\{\hat\delta(t)\}$ or the equal weight w_i^e is used as w_i^n in (2.7) to estimate $\beta_\tau(t)$, oth $\hat\alpha_\tau[w_i\{\hat\delta(t)\}]$ and $\hat\alpha_\tau(w_i^e)$ at $\tau \in (0,1)$ satisfies*

$$\sqrt{n}\left\{\hat\alpha_\tau(w_i^n) - \alpha_\tau - \frac{h^2\mu_k}{2}\mathbf{b}_{\alpha_\tau}\right\} \xrightarrow{d} N(0, \Sigma_{fX}^{-1}(V + \mathcal{B}\mathbf{L}\mathcal{B})\Sigma_{fX}^{-1}).$$

Theorem 2 shows that the choice of weights for estimation of $\beta_\tau(t)$ has a significant influence on the limiting variance of the estimator of α_τ. An asymptotic variance of $\hat\alpha_\tau\{w_i(\hat\gamma)\}$ cannot be larger than that of $\hat\alpha_\tau[w_i\{\hat\delta(t)\}]$ and $\hat\alpha_\tau(w_i^e)$, due to a nonnegative definite matrix $\mathbf{L}$. In general, $\mathbf{L}$ is most likely to be positive definite, since the number of estimating equations in (2.4) is larger than the dimensionality of γ. It is important to note that use of the global weight in both estimation of $\beta_\tau(t)$ and α_τ does not only improve estimation efficiency of α_τ, but also prevents efficiency loss from the discordance between w_i^n and w_i^p.

If $X(t)$ and $Z(t)$ are uncorrelated across time t with $E\{X(t)\}_t = 0$ or $E\{Z(t)\}_t = 0$ and homogeneous errors ξ_{ij}, then $\{D(t)\}_t = 0$ holds. Consequently, it follows from $\mathcal{B} = \mathbf{0}$ and the bias term $\mathbf{b}_{\alpha_\tau} = 0$ that $\hat\alpha_\tau(w_i^n)$ satisfies a $\sqrt{n}$-consistency and its asymptotic variance is identical regardless of the choice of w_i^n in estimation of $\beta_\tau(t)$. However, the foregoing conditions are too strong to satisfy in practice, since longitudinal data often involve heteroscedasticity and the covariates from observational studies are more likely to be correlated. To remove the bias of $\hat\alpha_\tau$ caused from nonparametric estimation, undersmoothing is typically needed [13, 1].

When the equal weight is used for estimation of $\beta_\tau(t)$ and α_τ, the limiting variance of $\hat\alpha_\tau$ is $\tau(1-\tau)\Sigma_{fX}^{-1}\mathcal{U}\Sigma_{fX}^{-1}$, excluding $\Sigma_{fX}^{-1}\mathcal{A}\mathbf{L}\mathcal{A}\Sigma_{fX}^{-1}$ from $\Sigma_{fX}^{-1}V\Sigma_{fX}^{-1}$ in (3.5). This causes a loss of parametric estimation efficiency. In addition, use of the global weight, only for estimation of α_τ, might lead to a less efficient estimator than the one obtained from completely ignoring within-subject dependence ($w_i^p = w_i^n = w_i^e$), since $\mathcal{A}\mathbf{L}\mathcal{A} - \mathcal{B}\mathbf{L}\mathcal{B}$ is not always nonnegative definite. Therefore, it is essential to use the global weight for time-invariant coefficient estimation as well as time-varying coefficient estimation to obtain the desired efficient estimator for α_τ.

3.3. Inference with a perturbation approach

The conventional statistical inference on $\beta_\tau(t)$ and α_τ that plugs a consistent estimate of the limiting variance function into (3.1) and (3.5) can be very challenging in practice. A random perturbation approach provides a viable alternative [8]. We generate independent positive random variables $\{v_i\}_{i=1}^n$ from a distribution having both mean and variance 1, such as an exponential distribution with mean 1. Then we obtain $\hat\beta_{\tau,v}(t)$ and $\hat\alpha_{\tau,v}$ by minimizing two weighted objective functions iteratively,

$$\sum_{i=1}^n v_i w_i(\hat\gamma) \sum_{j=1}^m \rho_\tau [Y_{ij}^* - Z_{ij}^T\{\beta(t) + \beta^*(t_{ij}-t)\}]K\{(t_{ij}-t)/h\},$$

and

$$\sum_{i=1}^n v_i w_i(\hat\gamma) \sum_{j=1}^m \rho_\tau \{Y_{ij}^\# - X_{ij}^T\alpha\}.$$

Theorem 3. *Assume that conditions (A1)–(A5) and (B1)–(B3) hold. Given the data $\{Y_{ij}, U_{ij}\}$, $\hat\beta_{\tau,v}(t)$ and $\hat\alpha_{\tau,v}$ satisfy that for $\tau \in (0,1)$,*

$$\sqrt{Nh}\Big\{\hat\beta_{\tau,v}(t) - \beta_\tau(t) - \frac{\beta_\tau^{''}(t)\mu_k}{2}h^2\Big\} \xrightarrow{d} N\Big(0, \frac{\tau(1-\tau)\varphi_k}{p_T(t)}\Sigma_{fZ}^{-1}(t)\Sigma_Z(t)\Sigma_{fZ}^{-1}(t)\Big),$$

and

$$\sqrt{n}\left\{\hat\alpha_{\tau,v} - \alpha_\tau - \frac{h^2\mu_k}{2}\mathbf{b}_{\alpha_\tau}\right\} \xrightarrow{d} N(0, \Sigma_{fX}^{-1}V\Sigma_{fX}^{-1}).$$

Theorem 3 ensures that the asymptotic distributions of $\hat\beta_{\tau,v}(t)$ and $\hat\alpha_{\tau,v}$ are same as those of $\hat\beta(t)$ and $\hat\alpha_\tau$ as shown in Theorem 2. This confirms that the empirical distributions of $\hat\beta(t)$ and $\hat\alpha_\tau$ can be approximated by those of repeated evaluations of $\hat\beta_{\tau,v}(t)$ and $\hat\alpha_{\tau,v}$. Specifically, for statistical inference about $\beta(t)$ and α_τ, the covariance matrix of $\hat\beta(t)$ and $\hat\alpha_\tau$ can be evaluated by the sample covariance matrix constructed from a large collection of $\hat\beta_{\tau,v}(t)$ and $\hat\alpha_{\tau,v}$.

4. Implementation

After presenting a procedure that finds the local weight $w_i\{\hat{\hat\delta}(t)\}$ and the global weight $w_i(\hat\gamma)$, we implement estimation of $\beta_\tau(t)$ and α_τ. In the first step, we obtain two weights, $w_i\{\hat{\hat\delta}(t)\}$ and $w_i(\hat\gamma)$, via the following iterative estimating procedure:

(1-1) Set the time-invariant parameter γ varies over time and fit a time-varying coefficient model, $Y_{ij} = X_{ij}^T\gamma(t_{ij}) + Z_{ij}^T\delta(t_{ij}) + \epsilon_{ij}$, using the local linear least squares to obtain an initial value of $\hat{\hat\delta}(t)$.

(1-2) Substitute $\hat{\hat\delta}(t)$ into (2.4), estimate γ by maximizing (2.6), and compute $w_i(\hat\gamma)$.

(1-3) Replace γ in (2.3) with $\hat\gamma$ and obtain $\hat\delta(t)$ and $w_i\{\hat\delta(t)\}$ that maximizes (2.5).

(1-4) Iterate steps (1-2) and (1-3) until the convergence criterion is reached.

In the second step, we replace w_i^n in (2.7) and w_i^p in (2.8) with $w_i(\hat\gamma)$ and estimate $\beta_\tau(t)$ and α_τ as follows:

(2-1) Set an initial value of $\hat\beta_\tau(t)$ through the time-varying coefficient QR model, $Q_\tau(Y_{ij}|X_{ij}, Z_{ij}) = X_{ij}^T\alpha_\tau(t_{ij}) + Z_{ij}^T\beta_\tau(t_{ij})$, using the weighted local linear QR in (2.7).

(2-2) Update $\hat\beta_\tau(t)$ in (2.8) and obtain $\hat\alpha_\tau$ that minimizes (2.8).

(2-3) Given $\hat\alpha_\tau$ from step (2-2), estimate $\beta_\tau(t)$ by minimizing (2.7).

(2-4) Iterate steps (2-2) and (2-3) until the convergence criterion is met.

When we estimate $\delta(t)$ and γ in the first step, both $g_i\{\delta(t)\}$ and $h_i(\gamma)$ contain the diagonal variance matrix V_i, which can be readily estimated using the method of moments [12]. Moreover, we need to construct a consistent estimator of $\partial\hat\delta(t|\gamma)/\partial\gamma$ in $h_i(\gamma)$. Following [16], we have

$$\hat\delta(t|\gamma) - \delta(t) = \frac{1}{Nhp_T(t)}\boldsymbol{\Omega}(t)\boldsymbol{\Gamma}(t)\boldsymbol{\Sigma}(t)^{-1}\sum_{i=1}^{n}g_i\{\delta(t)\} + o_p(1).$$

Then, we can easily show that

$$\frac{\partial\hat\delta(t|\gamma)}{\partial\gamma} \xrightarrow{p} -\Sigma_\sigma(t)^{-1}E[Z(t)X(t)^T/\sigma^2\{U(t)\}].$$

Since $\Sigma_\sigma(t)$ and $E[Z(t)X(t)^T/\sigma^2\{U(t)\}]$ can be estimated by

$$(Nh_2)^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m}Z_{ij}Z_{ij}^T K\{(t_{ij}-t)/h_2\}/\hat\sigma^2(U_{ij})$$

and

$$(Nh_2)^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m}Z_{ij}X_{ij}^T K\{(t_{ij}-t)/h_2\}/\hat\sigma^2(U_{ij}),$$

respectively, an estimator of $\partial\hat\delta(t|\gamma)/\partial\gamma$ is

$$-\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}Z_{ij}X_{ij}^T K\{(t_{ij}-t)/h_2\}}{\sum_{i=1}^{n}\sum_{j=1}^{m}Z_{ij}Z_{ij}^T K\{(t_{ij}-t)/h_2\}}. \tag{4.1}$$

Estimation of $\beta_\tau(t)$ and α_τ relies on the choice of bandwidths. We adopt a leave-one-subject-out cross-validation approach [18, 23]. Given fixed bandwidths h_1 in (2.3) and h_2 in (4.1), we fit a model to all observations except the ith subject and obtain $\hat\delta^{-i}(t)$ and $\hat\gamma^{-i}$ via the above first step. The cross validation is defined as

$$\text{CV}(h_1, h_2) = \sum_{i=1}^{n}\sum_{j=1}^{m}\left\{Y_{ij} - X_{ij}^T\hat\gamma^{-i} - Z_{ij}^T\hat\delta^{-i}(t_{ij})\right\}^2.$$

We choose optimal bandwidths h_1 and h_2 that minimize $\mathrm{CV}(h_1, h_2)$. To choose a bandwidth h in (2.7), we first evaluate the bandwidth h_{ls} through the local linear least squares as

$$h_{ls} = \underset{h}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{m} \left\{ Y_{ij} - X_{ij}^T \tilde{\gamma}^{-i} - Z_{ij}^T \tilde{\delta}^{-i}(t_{ij}) \right\}^2,$$

where $\tilde{\gamma}^{-i} = \operatorname{argmin}_{\gamma} \sum_{k \neq i}^{n} w_i(\hat{\gamma}^{-i}) \sum_{j=1}^{m} \left(Y_{kj}^p - X_{kj}^T \gamma \right)^2$, $Y_{kj}^p = Y_{kj} - Z_{kj}^T \tilde{\delta}^{-i}(t_{kj})$,

$$\{\tilde{\delta}^{-i}(t), \tilde{\delta}^{*-i}(t)\} =$$
$$\underset{\delta, \delta^*}{\operatorname{argmin}} \sum_{k \neq i}^{n} w_i(\hat{\gamma}^{-i}) \sum_{j=1}^{m} \left[Y_{kj}^n - Z_{kj}^T \{\delta + \delta^*(t_{kj} - t)\} \right]^2 K\left(\frac{t_{kj} - t}{h_{ls}} \right),$$

and $Y_{kj}^n = Y_{kj} - X_{kj}^T \tilde{\gamma}^{-i}$. Similar to the proposed iterative estimation procedure above, $\tilde{\delta}^{-i}(t)$ and $\tilde{\gamma}^{-i}$ can be estimated. Following [25], the bandwidth h is computed as $h = h_{ls}[\tau(1 - \tau)/\phi^2 \{\Phi^{-1}(\tau)\}]^{1/5}$, where ϕ and Φ are standard normal density and distribution functions, respectively. To eliminate the bias terms in Theorem 1-3, at least theoretically, $\beta_\tau(t)$ should be undersmoothed, and thus the bandwidth is adjusted as $hn^{-2/15}$ [13].

5. Numerical studies

5.1. Simulation studies

We investigate the finite sample performance of the proposed approach ($w_i^p = w_i^n = w_i(\hat{\gamma})$) and its inference at $\tau = 0.5$, comparing it to procedures with other combinations of w_i^p and w_i^n with 500 simulation runs and a sample size of 100. We design the ith subject's measurement times as $t_{ij} = 0.5(j - 1) + U(0, 0.5)$ for $j = 1, \ldots, 5$. This leads to five unequally spaced time points between 0 and 2.5. Each data set is generated from the partial linear model

$$Y_{ij} = \beta_0(t_{ij}) + \beta_1(t_{ij})Z_{1,ij} + \alpha_1 X_{1,ij} + \alpha_2 X_{2,ij} + \epsilon_{ij},$$

where three covariates, $Z_{1,ij}$, $X_{1,ij}$, and $X_{2,ij}$, are generated from a $U(0, 1)$ with $\operatorname{cor}(Z_{1,ij}, X_{1,ij}) = \operatorname{cor}(Z_{1,ij}, X_{2,ij}) = \operatorname{cor}(X_{1,ij}, X_{2,ij}) = \rho_1$ being either 0 or 0.5, $\alpha_1 = \alpha_2 = 1$, and two time-varying coefficient curves were followed as $\beta_0(t) = (t - 2)^2/3$ and $\beta_1(t) = -\cos(t\pi/2)$. Here, we consider three distributions for the random errors, $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{i5})^T$:
Case 1 (Homogeneous errors) : $\epsilon_i \sim N(0, \Sigma)$, where Σ is an AR(1) correlation structure with a correlation coefficient ρ_0 of either 0.3 or 0.7.
Case 2 (Covariates dependent heteroscedastic errors): $\epsilon_i = (0.7 + 0.2Z_{1,i} + 0.3X_{1,i}^2)\zeta_i$, where $\zeta_i \sim N(0, \Sigma)$ and Σ is defined in Case 1.

Case 3 (Time dependent heteroscedastic errors): $\epsilon_i = (0.8 + 0.1t_i)\zeta_i$, where ζ_i is specified in *Case* 2.

We first obtain a global weight and a local weight using the empirical likelihood under the AR(1) working correlation structure. Next, we use the global weight in (2.8) for estimation of $\alpha_{1,0.5}$ and $\alpha_{2,0.5}$ at $\tau = 0.5$ and investigate how the choice of weights in (2.7) for estimation of $\beta_{0,0.5}(t)$ and $\beta_{1,0.5}(t)$ influences the time-invariant estimation based on three different weights of w_i^n: the global weight, the local weight, and the equal weight. For comparison, the conventional quantile regression using the equal weight in both time-invariant and time-varying coefficient estimation is also investigated. The standard Gaussian kernel is used and the bandwidths are selected as shown in Section 4. To evaluate estimation efficiency of the proposed approach, we compute the mean squared error for time-invariant coefficients, $\mathrm{MSE}(\hat{\alpha}) = \sum_{i=1}^{2} \sum_{j=1}^{500} (\hat{\alpha}_{i,0.5}^{(j)} - \alpha_i)^2/1000$, and the mean integrated squared error for time-varying coefficients, $\mathrm{MISE}\{\hat{\beta}(t)\} = \sum_{i=1}^{2} \sum_{j=1}^{500} \sum_{k=1}^{16} \{\hat{\beta}_{i,0.5}^{(j)}(t_k) - \beta_i(t_k)\}^2/16000$, where $\hat{\alpha}_{i,0.5}^{(j)}$ and $\hat{\beta}_{i,0.5}^{(j)}(t)$ are estimates of $\alpha_{i,0.5}$ and $\beta_{i,0.5}(t)$ from the jth simulated data set and $t_1, \ldots, t_{16}$ are evenly space time points on $[0.5, 2.0]$.

TABLE 1
Empirical relative efficiency (relative to the proposed estimator with $(w_i(\hat{\gamma}), w_i(\hat{\gamma}))$) for Case 1-3.

Case	ρ_0	w_i^n	w_i^p	$\rho_1 = 0$		$\rho_1 = 0.5$	
				$\beta(t)$	α	$\beta(t)$	α
1	0.7	$w_i\{\hat{\delta}(t)\}$	$w_i(\hat{\gamma})$	1.04	1.13	0.98	1.21
		w_i^e	$w_i(\hat{\gamma})$	1.01	1.13	1.00	1.05
		w_i^e	w_i^e	1.02	1.55	0.99	1.40
	0.3	$w_i\{\hat{\delta}(t)\}$	$w_i(\hat{\gamma})$	1.02	1.10	1.07	1.07
		w_i^e	$w_i(\hat{\gamma})$	1.01	1.10	0.99	1.02
		w_i^e	w_i^e	0.98	1.09	0.99	1.08
2	0.7	$w_i\{\hat{\delta}(t)\}$	$w_i(\hat{\gamma})$	1.04	1.21	0.99	1.18
		w_i^e	$w_i(\hat{\gamma})$	1.01	1.19	0.99	1.08
		w_i^e	w_i^e	1.03	1.45	0.98	1.52
	0.3	$w_i\{\hat{\delta}(t)\}$	$w_i(\hat{\gamma})$	1.02	1.03	0.98	1.09
		w_i^e	$w_i(\hat{\gamma})$	1.01	1.06	1.00	1.05
		w_i^e	w_i^e	1.00	1.13	1.00	1.12
3	0.7	$w_i\{\hat{\delta}(t)\}$	$w_i(\hat{\gamma})$	1.07	1.35	1.02	1.20
		w_i^e	$w_i(\hat{\gamma})$	0.98	1.08	1.00	1.03
		w_i^e	w_i^e	0.99	1.41	1.00	1.39
	0.3	$w_i\{\hat{\delta}(t)\}$	$w_i(\hat{\gamma})$	1.03	1.00	1.07	1.01
		w_i^e	$w_i(\hat{\gamma})$	0.98	1.02	0.98	1.00
		w_i^e	w_i^e	0.97	1.08	0.99	1.06

Table 1 reports a relative efficiency of the proposed estimator using $\left(w_i^n, w_i^p\right) = \left(w_i(\hat{\gamma}), w_i(\hat{\gamma})\right)$ to the one using either $\left(w_i\{\hat{\delta}(t)\}, w_i(\hat{\gamma})\right)$, $\left(w_i^e, w_i(\hat{\gamma})\right)$, or $\left(w_i^e, w_i^e\right)$, defined by the ratio of the mean squared errors and the mean integrated squared errors. The larger the value of the relative efficiency, the more efficient the proposed estimator. When the correlation coefficient ρ_0 is 0.7, the approaches utilizing the global weight for time-invariant coefficient estimation outperform the

one ignoring the within-subject correlation in terms of smaller values of the mean squared errors. On the other hand, the relative efficiencies are closer to one as the within-subject correlation becomes weaker. This suggests that accommodating an informative correlation can improve the parametric estimation efficiency when the within-subject correlation is presented. When the results of the approaches using the global weight for time-invariant coefficient estimation are compared, the proposed approach using the global weight for time-varying coefficient estimation yields a more efficient estimator than the others, using either the local weight or the equal weight in all cases. This confirms that the additional parametric estimation efficiency gain can be achieved when the global weight is used for both time-invariant and time-varying coefficient estimation. On the other hand, the relative efficiencies of the estimated time-varying coefficients are all close to one, as expected from the asymptotic results of $\hat{\beta}_\tau(t)$ in Section 3.

We further investigate the statistical inference based on the proposed random perturbation approach using the global weight on both estimations. We evaluate the empirical standard error with 200 repetitions based on an exponential distribution with mean 1 and compute coverage probabilities of pointwise confidence intervals at the nominal 95% level. Since the coverage probabilities of varying coefficients are measured at sixteen time points, we average them out and provide those in Table 2. The results support that the proposed approach is effective on the statistical inference in that the coverage probabilities are close to the nominal level in all cases under consideration.

TABLE 2
Empirical coverage probabilities of 95% confidence intervals for the proposed estimator.

Case	ρ_0	$\rho_1 = 0$				$\rho_1 = 0.5$			
		$\beta_1(t)$	$\beta_2(t)$	α_1	α_2	$\beta_1(t)$	$\beta_2(t)$	α_1	α_2
1	0.7	95.2	95.6	97.2	97.6	95.0	95.2	98.0	97.6
	0.3	94.2	95.0	94.4	95.0	94.2	95.4	95.6	96.0
2	0.7	96.0	95.6	96.6	97.0	94.4	95.4	97.8	97.2
	0.3	94.4	94.6	94.0	95.4	94.2	95.2	95.0	95.0
3	0.7	94.4	94.8	96.8	97.0	94.4	94.8	96.4	97.0
	0.3	93.6	94.6	94.2	96.8	94.8	95.0	95.0	95.4

5.2. Application to real data

In this section, we apply the proposed approach to longitudinal data from the Multi-Center AIDS Cohort study and evaluate the effects of pre-HIV infection CD4 cell count, smoking, and age at HIV infection on CD4 cell count. In general, CD4 cell count is considered a biomarker indicating the health status of HIV infected patients. The objective of this study is to explore the dynamic change of covariate effects on CD4 cell count over time and determine the trend of CD4 cell count depletion after HIV infection at different quantiles. This dataset consists of 283 homosexual men who became infected by HIV between 1984 and 1991. Although each patient was supposed to be repeatedly measured every 6 months,

many patients missed some of their scheduled visits. Thus, unequal numbers of repeated measurements, which range from 1 to 14, are followed at different times. [10] illustrated the study design, methods, and medical implications in more detail.

[7, 4] confirmed that the baseline effect and the pre-infection CD4 cell count effect on the mean response may vary over time, yet neither smoking nor age has a significant impact on the mean response. Therefore, we postulate the quantile regression model with partially linear time-varying coefficients, $Q_\tau\{Y(t)\} = \beta_{0,\tau}(t) + \beta_{1,\tau}(t)Z_1 + \alpha_{1,\tau}X_1 + \alpha_{2,\tau}X_2$, where $Q_\tau\{Y(t)\}$ is the τth conditional quantile of CD4 cell count at year t, Z_1 is the centered pre-infection CD4 cell count, X_1 is 1 if the subject smoked after infection and 0 otherwise, and X_2 is the individual's centered age at infection. Here, the covariates Z_1 and X_2 are computed by subtracting the averages of CD4 cell counts and age at infection from the subject pre-infection CD4 cell count and age at infection, respectively. This allows us to facilitate the biological interpretations, since $\beta_{0,\tau}(t)$ corresponds to the τth conditional quantile of CD4 cell count at year t for a nonsmoker with an average pre-infection CD4 cell count and an average age at infection.

We assume an AR(1) working correlation structure and assess the covariate effects on the post-infection CD4 cell count over time at $\tau = 0.25$, 0.5, and 0.75. We apply the proposed approach with the global weight to the estimation of both time-invariant and time-varying coefficients and use the cross validation approach in Section 4 for bandwidth selection under the standard Gaussian kernel. For statistical inference, the proposed perturbation approach with 200 repetitions and an exponential distribution with mean 1 is implemented to obtain the empirical distributions of the estimators of α_τ and $\beta_\tau(t)$.

TABLE 3
Estimates and their standard errors (se) of α_1 and α_2.

	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
$\hat{\alpha}_{1(\text{se})}$	$1.336_{(1.06)}$	$0.217_{(1.35)}$	$-1.250_{(1.19)}$
$\hat{\alpha}_{2(\text{se})}$	$-0.051_{(0.08)}$	$-0.096_{(0.09)}$	$-0.056_{(0.08)}$

Table 3 reports estimates of $\alpha_{1,\tau}$ and $\alpha_{2,\tau}$ along with their standard errors at $\tau = 0.25$, 0.5, and 0.75. The results confirm that age at infection is negatively associated with the CD4 cell count regardless of quantiles, while the sign of smoking effect at $\tau = 0.75$ is different from that of the smoking effect at $\tau = 0.25$ and 0.5. However, both smoking and age effects are not statistically significant in all quantiles under consideration. Similar results are reported with the conditional mean model in [7, 4].

Figure 1 provides the estimated time-varying coefficient curves of the baseline and the pre-CD4 cell count and their pointwise confidence intervals at the nominal 95% level. This figure indicates that the baseline CD4 cell count deteriorates over time at all quantiles under consideration, yet the decreasing rate of CD4 cell count at $\tau = 0.25$ is higher than the one at $\tau = 0.75$. In addition, the effect of the pre-infection CD4 cell count is statistically significant with pos-

itive estimates overall at all quantiles. The pre-infection CD4 effect appears to be constant over time at $\tau = 0.75$, while this effect decreases as time goes on at $\tau = 0.25$ and 0.5. This finding at the median is comparable with the ones illustrated by kernel smoothing methods [22, 23], an empirical likelihood [24], and a profile weighted least square [4] for the conditional mean regression.
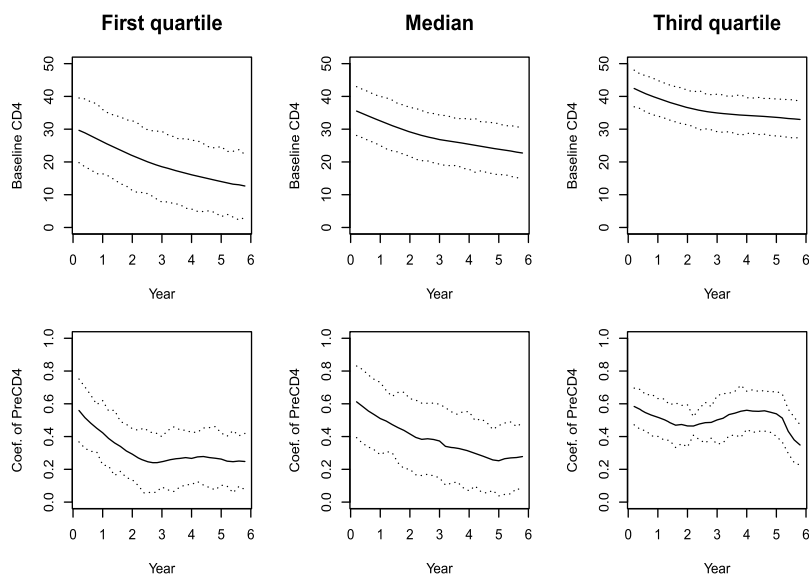

FIG 1. *Estimated varying-coefficient curves (solid) for the baseline and the pre-infection CD4 cell count along with the 95% confidence interval (dotted) at the 25th percentile (the first column), 50th percentile (the second column), and 75th percentile (the last column).*

Appendix: Proofs

Write $x_n \asymp y_n$ if $x_n/y_n \to 1$, $x_n = O(y_n)$ if $\sup_n |x_n/y_n| < \infty$, and $x_n = o(y_n)$ if $x_n/y_n \to 0$. The bandwidths h_1 in (2.3) and h in (2.7) are different. However, since $h_1 = O(h)$, for the ease of presentation we abuse notation $h = h_1$ in the proof. We also omit τ from $\xi_{ij}(\tau)$. The following conditions are needed for our asymptotic results.

(A1) $f_\xi(\cdot|U_{ij})$ is bounded, positive, and twice continuously differentiable on $\{v : 0 < F_\xi(v|U_{ij}) < 1\}$. $f_\xi\{0|U(t)\}$ and $p_T(t)$ are continuously differentiable in a neighborhood of t.

(A2) $X(t)$ and $Z(t)$ are bounded. $\Sigma_Z(t) = E\{Z(t)Z(t)^T\}$ and $\Sigma_{fZ}(t) = E[f_{\xi_\tau}\{0|U(t)\}Z(t)Z(t)^T]$ are positive definite and differentiable.

(A3) Let $m < \infty$, $nh \to \infty$ and $nh^5 < \infty$.

(A4) $K(\cdot)$ is symmetric with bounded support and bounded derivative.

(A5) $\beta_\tau(\cdot)$ is twice continuously differentiable in a neighborhood of t.

(A1)–(A5) are the standard regularity conditions in studying nonparametric quantile regression [2] and the following conditions are considered in the empirical likelihood [16].

(B1) $\delta(\cdot)$ is twice continuously differentiable in a neighborhood of t.

(B2) There exists γ_0 such that $E\{h_i(\gamma_0)\} = 0$, $\boldsymbol{\Psi}(\gamma_0)$ is positive definite, and $\partial h_i(\gamma)/\partial\gamma$ is continuous in a neighborhood of γ_0 and is of full rank. There exists $\delta(t)$ such that $E[g_i\{\delta(t)\}] = 0$ and $\partial g_i(\delta)/\partial\delta$ is continuous in a neighborhood of $\delta(t)$ and is of full rank.

(B3) $\Sigma_\sigma(t)$ and $\Lambda_\tau(t)$ are positive definite and differentiable. $\mathbf{G}_\ell^T(t)$ is differentiable.

(B2) and (B3) are not required for the quantile regression, yet these are quite standard for the marginal mean regression [12, 17], and thereby the proposed method is generally applicable for longitudinal data.

Without loss of generality, the inverse of the correlation matrix is decomposed as $R^{-1} = a_0\mathbf{I} + \mathbf{C}$, where a_0 is an unknown coefficient and $\mathbf{C}$ is the matrix containing off-diagonal elements in R^{-1}. Accordingly, $g_i\{\delta(t)\}$ in (2.3) is rewritten as

$$\left(\begin{array}{c} g_{i1}\{\delta(t)\} \\ g_{i2}\{\delta(t)\} \end{array} \right) = \left(\begin{array}{c} Z_i K_i(t)^{1/2} V_i^{-1/2} \mathbf{I} V_i^{-1/2} K_i(t)^{1/2} \{Y_i' - Z_i^T\delta(t)\} \\ Z_i K_i(t)^{1/2} V_i^{-1/2} \mathbf{C} V_i^{-1/2} K_i(t)^{1/2} \{Y_i' - Z_i^T\delta(t)\} \end{array} \right), \quad (5.1)$$

and use (5.1) to prove Lemma 1 and Theorem 1.

Lemma 1. *Under conditions (B1)–(B3), we have*

$$\frac{1}{Nh} \sum_{i=1}^n g_i\{\delta(t)\} g_i\{\delta(t)\}^T \quad \overset{p}{\to} \quad \varphi_K p_T(t) \mathbf{C_1} \otimes \Sigma_\sigma(t), \quad (5.2)$$

$$\frac{1}{Nh} \sum_{i=1}^n \frac{\partial g_i\{\delta(t)\}}{\partial\delta(t)} \quad \overset{p}{\to} \quad -p_T(t) \mathbf{C_2} \otimes \Sigma_\sigma(t), \quad (5.3)$$

where

$$\mathbf{C_1} = \left(\begin{array}{cc} 1 & \frac{1}{m}\sum_{j=1}^m c_{jj} \\ \frac{1}{m}\sum_{j=1}^m c_{jj} & \frac{1}{m}\sum_{j=1}^m c_{jj}^2 \end{array} \right), \qquad \mathbf{C_2} = \left(\begin{array}{c} 1 \\ \frac{1}{m}\sum_{j=1}^m c_{jj} \end{array} \right)$$

and c_{ij} is the (i,j)th component of $\mathbf{C}$.

Proof of Lemma 1. Write $e_i = Y_i' - Z_i^T\delta(t)$ and $K_{ij} = K\{(t_{ij} - t)/h\}$. Recall $\sigma_{ij}^2 = E(\epsilon_{ij}^2|U_{ij})$, where $\epsilon_{ij} = Y_{ij} - X_{ij}^T\gamma - Z_{ij}^T\delta(t_{ij})$. Simple linear algebra gives

$$g_i\{\delta(t)\} = \left(\begin{array}{c} g_{i1}\{\delta(t)\} \\ g_{i2}\{\delta(t)\} \end{array} \right) = \left(\begin{array}{c} \sum_{j=1}^m \sum_{j'=1}^m Z_{ij} K_{ij}^{1/2} K_{ij'}^{1/2} e_{ij'}/(\sigma_{ij}\sigma_{ij'}) \\ \sum_{j=1}^m \sum_{j'=1}^m Z_{ij} K_{ij}^{1/2} c_{jj'} K_{ij'}^{1/2} e_{ij'}/(\sigma_{ij}\sigma_{ij'}) \end{array} \right)$$

where e_{ij} is the jth element of e_i. To prove (5.2) and (5.3), it is sufficient to show the convergence of block matrices of $(Nh)^{-1}\sum_{i=1}^n g_i\{\delta(t)\} g_i\{\delta(t)\}^T$ and

$(Nh)^{-1} \sum_{i=1}^{n} \partial g_i \{\delta(t)\} / \partial \delta(t)$ in probability. Here we show the convergence of the (2,2)th block matrix of $(Nh)^{-1} \sum_{i=1}^{n} g_i \{\delta(t)\} g_i \{\delta(t)\}^T$ and the 2nd block matrix of $(Nh)^{-1} \sum_{i=1}^{n} \partial g_i \{\delta(t)\} / \partial \delta(t)$ in probability. The convergence of other block matrices can be similarly shown.

Since the kernel function K has bounded support, it is sufficient to consider $|t_{ij} - t| = O(h)$. By the boundedness of Z_{ij} and Taylor's expansion $\delta(t_{ij}) - \delta(t) = (t_{ij} - t) \delta'(t) + o(h) = O(h)$, we have

$$
\begin{aligned}
E(Z_{ij} Z_{ij}^T e_{ij}^2 / \sigma_{ij}^4 | t_{ij}) &= E\{Z_{ij} Z_{ij}^T / \sigma_{ij}^4 E(e_{ij}^2 | U_{ij}) | t_{ij}\} \\
&= E\{Z_{ij} Z_{ij}^T / \sigma_{ij}^4 E([\epsilon_{ij} + Z_{ij}^T \{\delta(t_{ij}) - \delta(t)\}]^2 | U_{ij}) | t_{ij}\} \\
&= E(Z_{ij} Z_{ij}^T \sigma_{ij}^2 / \sigma_{ij}^4 | t_{ij}) + O(h) \\
&\asymp E(Z_{ij} Z_{ij}^T / \sigma_{ij}^2 | t_{ij}) = \Sigma_\sigma(t_{ij}).
\end{aligned}
\tag{5.4}
$$

By the fact that m is bounded, $E(K_{ij} K_{ij'}) = O(h^2)$ for $j \neq j'$, $E(K_{ij} K_{ij'}^{1/2} K_{ij''}^{1/2}) = O(h^3)$ for $j \neq j' \neq j''$ and $E(K_{ij}^{1/2} K_{ij'}^{1/2} K_{ij''}^{1/2} K_{ij'''}^{1/2}) = O(h^4)$ for $j \neq j' \neq j'' \neq j'''$, (5.4) and condition (B3), we can obtain

$$
\begin{aligned}
&\frac{1}{Nh} \sum_{i=1}^{n} E[g_{i2}\{\delta(t)\} g_{i2}\{\delta(t)\}^T] \\
={}& \frac{1}{Nh} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{j'=1}^{m} \sum_{k=1}^{m} \sum_{k'=1}^{m} E\{Z_{ij} Z_{ik}^T K_{ij}^{1/2} K_{ij'}^{1/2} K_{ik}^{1/2} K_{ik'}^{1/2} \\
&\hspace{4.5cm} \times c_{jj'} c_{kk'} e_{ij'} e_{ik'} / (\sigma_{ij} \sigma_{ij'} \sigma_{ij} \sigma_{ik'})\} \\
={}& \frac{1}{Nh} \sum_{i=1}^{n} \sum_{j=1}^{m} E(Z_{ij} Z_{ij}^T K_{ij}^2 e_{ij}^2 / \sigma_{ij}^4) c_{jj}^2 + o(1) \\
={}& \frac{1}{Nh} \sum_{i=1}^{n} \sum_{j=1}^{m} E\{K_{ij}^2 E(Z_{ij} Z_{ij}^T e_{ij}^2 / \sigma_{ij}^4 | t_{ij})\} c_{jj}^2 + o(1) \\
={}& \frac{1}{Nh} \sum_{i=1}^{n} \sum_{j=1}^{m} E\{K_{ij}^2 \Sigma_\sigma(t_{ij})\} c_{jj}^2 + o(1) \\
\rightarrow{}& \varphi_K p_T(t) \Sigma_\sigma(t) \frac{1}{m} \sum_{j=1}^{m} c_{jj}^2,
\end{aligned}
\tag{5.5}
$$

and also

$$
\begin{aligned}
&\frac{1}{Nh} \sum_{i=1}^{n} E\left[\frac{\partial g_{i2}\{\delta(t)\}}{\partial \delta(t)}\right] \\
={}& -\frac{1}{Nh} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{j'=1}^{m} E\{Z_{ij} Z_{ij'}^T K_{ij}^{1/2} K_{ij'}^{1/2} c_{jj'} / (\sigma_{ij} \sigma_{ij'})\} \\
={}& -\frac{1}{Nh} \sum_{i=1}^{n} \sum_{j=1}^{m} E(K_{ij} Z_{ij} Z_{ij}^T / \sigma_{ij}^2) c_{jj} + o(1) \\
\rightarrow{}& -p_T(t) \Sigma_\sigma(t) \frac{1}{m} \sum_{j=1}^{m} c_{jj}.
\end{aligned}
$$

Similarly, the variance of each component of $(Nh)^{-1}\sum_{i=1}^{n}g_i\{\delta(t)\}g_i\{\delta(t)\}^T$ and $(Nh)^{-1}\sum_{i=1}^{n}\partial g_i\{\delta(t)\}/\partial\delta(t)$ converges to 0. This implies the assertion of Lemma 1. □

For convenience, we recall important results in [16]:

(C1) By a similar argument in the proof of Theorem 1 in [16] and Lemma 1, we have

$$\lambda_{\hat{\delta}(t)} = \frac{1}{\varphi_K p_T(t)}\mathbf{M}\bar{g}_n\{\delta(t)\} + o_p\{(Nh)^{-1/2}\},$$

where $\bar{g}_n\{\delta(t)\} = (Nh)^{-1}\sum_{i=1}^{n}g_i\{\delta(t)\}$ and $\mathbf{M}$ are defined in Section 3.1, and

$$\lambda_{\hat{\gamma}} = \mathbf{L}\bar{h}_n(\gamma_0) + o_p(n^{-1/2})$$

where $\bar{h}_n(\gamma_0) = n^{-1}\sum_{i=1}^{n}h_i(\gamma_0)$ and $\mathbf{L}$ are defined in Section 3.1.

(C2) $w_i\{\hat{\delta}(t)\} = n^{-1}\Big[1 + g_i\{\hat{\delta}(t)\}^T\lambda_{\hat{\delta}(t)}\Big]^{-1} = n^{-1}\Big[1 - g_i\{\hat{\delta}(t)\}^T\lambda_{\hat{\delta}(t)}\{1 + o_p(1)\}\Big]$

and $w_i(\hat{\gamma}) = n^{-1}\Big[1 + h_i(\hat{\gamma})^T\lambda_{\hat{\gamma}}\Big]^{-1} = n^{-1}\Big[1 - h_i(\hat{\gamma})^T\lambda_{\hat{\gamma}}\{1 + o_p(1)\}\Big]$ uniformly for i.

(C3) $\max_{1\leq i\leq n}|g_i\{\hat{\delta}(t)\}^T\lambda_{\hat{\delta}(t)}| = o_p(1)$ and $\max_{1\leq i\leq n}|h_i(\hat{\gamma})^T\lambda_{\hat{\gamma}}| = o_p(1)$.

Proof of Theorem 1. When $\hat{\alpha}_\tau$ is a $\sqrt{n}$ consistent estimator of α_τ, we have

$$
\begin{aligned}
&\sqrt{Nh}\{\hat{\beta}_\tau(t|\hat{\alpha}_\tau) - \beta_\tau(t)\}\\
=\ &\sqrt{Nh}\{\hat{\beta}_\tau(t|\hat{\alpha}_\tau) - \hat{\beta}_\tau(t|\alpha_\tau)\} + \sqrt{Nh}\{\hat{\beta}_\tau(t|\alpha_\tau) - \beta_\tau(t)\}\\
=\ &\sqrt{mh}\left\{\frac{\partial\hat{\beta}(t|\alpha_\tau)}{\partial\alpha_\tau^T}\right\}\sqrt{n}(\hat{\alpha}_\tau - \alpha_\tau) + \sqrt{Nh}\{\hat{\beta}_\tau(t|\alpha_\tau) - \beta_\tau(t)\} + o_p(1)\\
=\ &\sqrt{Nh}\{\hat{\beta}_\tau(t|\alpha_\tau) - \beta_\tau(t)\} + o_p(1)
\end{aligned}
$$

since $\partial\hat{\beta}_\tau(t|\alpha_\tau)/\partial\alpha_\tau^T = \Sigma_{fZ}^{-1}(t)E[f_\xi\{0|U(t)\}Z(t)X(t)^T] + o_p(1) = O_p(1)$ due to (5.13). As a result, the asymptotic distribution of $\hat{\beta}_\tau(t|\alpha_\tau)$ is the same as that of $\hat{\beta}_\tau(t|\hat{\alpha}_\tau)$, and thus we assume that α_τ is known. Let $\Theta = \sqrt{Nh}[\{\beta - \beta_\tau(t)\}^T, h\{\beta^* - \beta_\tau'(t)\}^T]^T$ and $J_{ij} = \{Z_{ij}^T, h^{-1}(t_{ij} - t)Z_{ij}^T\}^T$. Recall $K_{ij} = K\{(t_{ij} - t)/h\}$, $Y_{ij}^* = Y_{ij} - X_{ij}^T\alpha_\tau$ and $\xi_{ij} = Y_{ij}^* - Z_{ij}^T\beta_\tau(t_{ij})$. Then we can write

$$Y_{ij}^* - Z_{ij}^T\{\beta + \beta^*(t_{ij} - t)\} = d_{ij} + \xi_{ij} - J_{ij}^T\Theta/\sqrt{Nh}, \tag{5.6}$$

where $d_{ij} = Z_{ij}^T[\beta_\tau(t_{ij}) - \beta_\tau(t) - \beta_\tau'(t)(t_{ij} - t)]$. Since $\{\hat{\beta}_\tau(t)^T, \hat{\beta}_\tau^*(t)^T\}^T$ minimizes the criterion function in (2.7), by (5.6) $\hat{\Theta} = \sqrt{Nh}[\{\hat{\beta}_\tau(t) - \beta_\tau(t)\}^T, h\{\hat{\beta}_\tau^*(t) - \beta_\tau'(t)\}^T]^T$ minimizes the following re-parameterized function of Θ:

$$\mathcal{L}(\Theta) = \sum_{i=1}^{n}w_i^n\sum_{j=1}^{m}\left\{\rho_\tau\Big(d_{ij} + \xi_{ij} - J_{ij}^T\Theta/\sqrt{Nh}\Big) - \rho_\tau\Big(d_{ij} + \xi_{ij}\Big)\right\}K_{ij}.$$

Let $\theta_{ij} = J_{ij}^T \Theta / \sqrt{Nh}$. Applying Knight's identity $\rho_\tau(u - v) - \rho_\tau(u) = -v(\tau - \mathbf{1}_{u<0}) + \int_0^v (\mathbf{1}_{u \leq s} - \mathbf{1}_{u \leq 0}) ds$, we can write $\mathcal{L}(\Theta) = -A_n \Theta + I_n$, where

$$
\begin{aligned}
A_n &= \frac{1}{\sqrt{Nh}} \sum_{i=1}^n w_i^n \sum_{j=1}^m (\tau - \mathbf{1}_{d_{ij} + \xi_{ij} < 0}) K_{ij} J_{ij}^T, \\
I_n &= \sum_{i=1}^n w_i^n \sum_{j=1}^m K_{ij} \eta_{ij}, \quad \eta_{ij} = \int_0^{\theta_{ij}} (\mathbf{1}_{d_{ij} + \xi_{ij} \leq s} - \mathbf{1}_{d_{ij} + \xi_{ij} \leq 0}) ds.
\end{aligned}
$$

Consider I_n first. By (C2) and (C3), we have

$$
nI_n = nI_{n1} - nI_{n2} := \sum_{i=1}^n \sum_{j=1}^m K_{ij} \eta_{ij} - \sum_{i=1}^n o_p(1) \sum_{j=1}^m K_{ij} \eta_{ij}.
$$

Recall $|t_{ij} - t| = O(h)$ due to the bounded support of K. Then, by the boundedness of Z_{ij}, $|\theta_{ij}| \leq c_1 / \sqrt{Nh}$ and $|d_{ij}| \leq c_1 h^2$ for some constant c_1. Applying the inequality $| \int_0^\theta (\mathbf{1}_{u \leq s} - \mathbf{1}_{u \leq 0}) ds | \leq |\theta| \mathbf{1}_{|u| \leq |\theta|}$, we have

$$
|\eta_{ij}| \leq |\theta_{ij}| \mathbf{1}_{-|\theta_{ij}| \leq \xi_{ij} + d_{ij} \leq |\theta_{ij}|} \leq \frac{c_1}{\sqrt{Nh}} \mathbf{1}_{-c_1 \rho_n \leq \xi_{ij} \leq c_1 \rho_n}, \quad \rho_n = \frac{1}{\sqrt{Nh}} + h^2,
$$

and $E(K_{ij}^2 \eta_{ij}^2) = O(\rho_n / N)$. By the Cauchy-Schwarz inequality and condition (A3), we have $\text{var}(nI_{n1}) =$

$$
\sum_{i=1}^n \text{var}\left(\sum_{j=1}^m K_{ij} \eta_{ij} \right) \leq \sum_{i=1}^n \left\{ m \sum_{j=1}^m E(K_{ij}^2 \eta_{ij}^2) \right\} = O\left(m^2 \sum_{i=1}^n \rho_n / N \right) \to 0. (5.7)
$$

Because $d_{ij} = O(h^2)$, the Taylor's expansion yields

$$
E(\eta_{ij}|U_{ij}) = \int_0^{\theta_{ij}} \left\{ F_\xi(s - d_{ij}|U_{ij}) - F_\xi(-d_{ij}|U_{ij}) \right\} ds \asymp \frac{\theta_{ij}^2}{2} f_\xi(0|U_{ij}), \quad (5.8)
$$

uniformly for all (i, j). Then we have

$$
\sum_{i=1}^n \sum_{j=1}^m E\{K_{ij} \theta_{ij}^2 f_\xi(0|U_{ij})\} \to p_T(t) \Theta^T \mathcal{W}(t) \Theta, \qquad (5.9)
$$

where $\mathcal{W}(t) = \text{diag}\{\Sigma_{fZ}(t), \mu_K \Sigma_{fZ}(t)\}$ is a block diagonal matrix. By a similar argument with (5.7)–(5.9), it is easy to show that

$$
nI_{n2} = o_p(1) \sum_{i=1}^n \sum_{j=1}^m K_{ij} \eta_{ij} = o_p(1). \qquad (5.10)
$$

Therefore, by (5.7)–(5.10), we have the convergence in probability:

$$
nI_n = E(nI_n) + o_p(1) = \sum_{i=1}^n \sum_{j=1}^m E\{K_{ij} E(\eta_{ij}|U_{ij})\} + o_p(1) \xrightarrow{p} \frac{p_T(t)}{2} \Theta^T \mathcal{W}(t) \Theta.
$$

By the convexity lemma [15], $\hat{\Theta}$ has the Bahadur representation:

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \left\{ - A_n \Theta + \frac{p_T(t)}{2n} \Theta^T \mathcal{W}(t) \Theta \right\} + o_p(1) = \frac{n}{p_T(t)} \mathcal{W}(t)^{-1} A_n^T + o_p(1).$$

From the first component of $\hat{\Theta}$, we have the asymptotic Bahadur representation:

$$\begin{aligned} &\hat{\beta}_\tau(t|\alpha_\tau) - \beta_\tau(t|\alpha_\tau) \\ =\ & \frac{\Sigma_{fZ}^{-1}(t)}{p_T(t)} \frac{n}{Nh} \sum_{i=1}^n w_i^n \sum_{j=1}^m (\tau - \mathbf{1}_{\xi_{ij}<0} + \zeta_{ij}) K_{ij} Z_{ij} + o_p\{(Nh)^{-1/2}\}, \quad (5.11) \end{aligned}$$

where $\zeta_{ij} = \mathbf{1}_{\xi_{ij}<0} - \mathbf{1}_{d_{ij}+\xi_{ij}<0}$. Let $R_n = (Nh)^{-1} n \sum_{i=1}^n w_i^n \sum_{j=1}^m \zeta_{ij} K_{ij} Z_{ij}$. Similar to nI_n above, the convergence of $R_{n1} = (Nh)^{-1} \sum_{i=1}^n \sum_{j=1}^m \zeta_{ij} K_{ij} Z_{ij}$ in probability should be shown, since $R_{n2} = (Nh)^{-1} \sum_{i=1}^n o_p(1) \sum_{j=1}^m \zeta_{ij} K_{ij} Z_{ij}$ is negligible by (C3). By the arguments in (5.7)–(5.8) and Taylor's expansion $d_{ij} = (t_{ij} - t)^2 Z_{ij}^T \beta_\tau''(t)/2 + o(h^2)$,

$$\begin{aligned} E(R_{n1}) &= \frac{1}{Nh} \sum_{i=1}^n \sum_{j=1}^m E\{K_{ij} Z_{ij} E(\zeta_{ij}|U_{ij})\} \\ &= \frac{p_T(t)\mu_K}{2} \Sigma_{fZ}(t) \beta_\tau''(t) h^2 + o(h^2), \quad (5.12) \end{aligned}$$

and $\operatorname{var}(R_{n1}) = o\{(Nh)^{-1/2}\}$. Thus, it follows from (5.11)–(5.12) that

$$\begin{aligned} &\sqrt{Nh} \left\{ \hat{\beta}_\tau(t|\alpha_\tau) - \beta_\tau(t|\alpha_\tau) - \frac{\beta_\tau''(t)\mu_K}{2} h^2 \right\} \\ =\ & \frac{\Sigma_{fZ}^{-1}(t)}{p_T(t)} \frac{n}{\sqrt{Nh}} \sum_{i=1}^n w_i^n \varrho_i + o_p(1) \quad (5.13) \end{aligned}$$

where $\varrho_i = \sum_{j=1}^m \varrho_{ij}$ with $\varrho_{ij} = (\tau - \mathbf{1}_{\xi_{ij}<0}) K_{ij} Z_{ij}$.

(i) Suppose that $w_i^n = w_i^e = 1/n$ is used. Note that for $j \ne j'$, $E(\varrho_{ij}\varrho_{ij'}^T) = O(h^2)$. Thus,

$$\begin{aligned} \operatorname{var}\left(\frac{1}{\sqrt{Nh}} \sum_{i=1}^n \varrho_i \right) &= \frac{1}{Nh} \sum_{i=1}^n \sum_{j=1}^m \operatorname{var}\{(\tau - \mathbf{1}_{\xi_{ij}<0}) K_{ij} Z_{ij}\} + \frac{1}{Nh} \sum_{i=1}^n O(m^2 h^2) \\ &\to \tau(1-\tau) p_T(t) \varphi_K \Sigma_Z(t). \end{aligned}$$

The desired result in (3.2) then easily follows from (5.13) and the independence of $\varrho_1, \ldots, \varrho_n$.

(ii) Suppose that local weight $w_i^n = w_i\{\hat{\delta}(t)\}$ is used. By (C2), we have

$$\frac{n}{\sqrt{Nh}} \sum_{i=1}^n w_i\{\hat{\delta}(t)\} \varrho_i = \frac{1}{\sqrt{Nh}} \sum_{i=1}^n \sum_{j=1}^m \varrho_{ij} \left[1 - g_i\{\hat{\delta}(t)\}^T \lambda_{\hat{\delta}(t)} \{1 + o_p(1)\} \right]$$

Recall $g_i(\cdot)$ in (5.1). Since K has bounded support and $\hat{\delta}(t) - \delta(t) = O_p\{(Nh)^{-1/2}\}$, by a similar argument with (5.5), we have

$$E\left[\frac{1}{Nh}\sum_{i=1}^{n}\sum_{j=1}^{m}\varrho_{ij}g_{i2}\{\hat{\delta}(t)\}^{T}\right]$$

$$= E\left[\frac{1}{Nh}\sum_{i=1}^{n}\sum_{j=1}^{m}(\tau-\mathbf{1}_{\xi_{ij}<0})K_{ij}Z_{ij}\sum_{j'=1}^{m}\right.$$

$$\left.\times\sum_{j''=1}^{m}Z_{ij'}^{T}K_{ij'}^{1/2}K_{ij''}^{1/2}c_{j'j''}e_{ij''}/(\sigma_{ij'}\sigma_{ij''})\right]+o(1)$$

$$= \frac{1}{Nh}\sum_{i=1}^{n}\sum_{j=1}^{m}E\left\{(\tau-\mathbf{1}_{\xi_{ij}<0})\epsilon_{ij}K_{ij}^{2}Z_{ij}Z_{ij}^{T}c_{jj}/\sigma_{ij}^{2}\right\}+o(1)$$

$$= \frac{1}{Nh}\sum_{i=1}^{n}\sum_{j=1}^{m}E\left[E\{(\tau-\mathbf{1}_{\xi_{ij}<0})\epsilon_{ij}|U_{ij}\}K_{ij}^{2}Z_{ij}Z_{ij}^{T}/\sigma_{ij}^{2}\right]c_{jj}+o(1)$$

$$\to p_{T}(t)\varphi_{K}\Lambda_{\tau}(t)\frac{1}{m}\sum_{j=1}^{m}c_{jj}$$

and $E\left[(Nh)^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m}\varrho_{ij}g_{i1}\{\hat{\delta}(t)\}^{T}\right]\to p_{T}(t)\varphi_{K}\Lambda_{\tau}(t)$. By the law of large numbers,

$$\frac{1}{Nh}\sum_{i=1}^{n}\sum_{j=1}^{m}\varrho_{ij}g_{i}\{\hat{\delta}(t)\}^{T}\xrightarrow{p}p_{T}(t)\varphi_{K}\mathbf{\Lambda}_{\tau}(t)^{T}. \tag{5.14}$$

By (5.14), (C1) and Lemma 1, using steps similar to those in the proof of Theorem 1 in [19], we have

$$\operatorname{var}\left[\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m}\varrho_{ij}g_{i}\{\hat{\delta}(t)\}^{T}\lambda_{\hat{\delta}(t)}\right]$$

$$= \operatorname{var}\left\{\sqrt{Nh}p_{T}(t)\varphi_{K}\mathbf{\Lambda}_{\tau}(t)^{T}\lambda_{\hat{\delta}(t)}\right\}\{1+o(1)\}$$

$$= \operatorname{var}\left[\sqrt{Nh}\mathbf{\Lambda}_{\tau}(t)^{T}\mathbf{M}\bar{g}_{n}\{\delta(t)\}\right]\{1+o(1)\}$$

$$= \mathbf{\Lambda}_{\tau}(t)^{T}\mathbf{M}E\left[\frac{1}{Nh}\sum_{i=1}^{n}g_{i}\{\delta(t)\}g_{i}\{\delta(t)\}^{T}\right]\mathbf{M}\mathbf{\Lambda}_{\tau}(t)\{1+o(1)\}$$

$$= p_{T}(t)\varphi_{K}\mathbf{\Lambda}_{\tau}(t)^{T}\mathbf{M}\mathbf{\Sigma}(t)\mathbf{M}\mathbf{\Lambda}_{\tau}(t)+o(1)$$

$$\to p_{T}(t)\varphi_{K}\mathbf{\Lambda}_{\tau}(t)^{T}\mathbf{M}\mathbf{\Lambda}_{\tau}(t). \tag{5.15}$$

Similarly, we can obtain

$$\operatorname{cov}\left(\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m}\varrho_{ij}g_{i}\{\hat{\delta}(t)\}^{T}\lambda_{\hat{\delta}(t)},\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m}\varrho_{ij}\right)$$

$$= \operatorname{cov}\left(\mathbf{\Lambda}_{\tau}(t)^{T}\mathbf{M}\lambda_{\hat{\delta}(t)},\sum_{i=1}^{n}\sum_{j=1}^{m}\varrho_{ij}\right)+o(1)$$

$$\to p_{T}(t)\varphi_{K}\mathbf{\Lambda}_{\tau}(t)^{T}\mathbf{M}\mathbf{\Lambda}_{\tau}(t).$$

Thus, we have

$$\text{var}\left[\frac{n}{\sqrt{Nh}}\sum_{i=1}^{n}w_i\{\hat\delta(t)\}\varrho_i\right] \to p_T(t)\varphi_K\Big\{\tau(1-\tau)\Sigma_Z(t) - \mathbf{\Lambda}_\tau(t)^T\mathbf{M}\mathbf{\Lambda}_\tau(t)\Big\}.$$

However, it is easy to show that $\mathbf{\Lambda}_\tau(t)^T\mathbf{M}\mathbf{\Lambda}_\tau(t) = \mathbf{0}$ regardless of the dependence matrix $\mathbf{C}$ in (5.1) because $\mathbf{C}_2^T\mathbf{C}_1^{-1}\{\mathbf{I} - \mathbf{C}_2\{\mathbf{C}_2^T\mathbf{C}_1^{-1}\mathbf{C}_2\}^{-1}\mathbf{C}_2^T\mathbf{C}_1^{-1}\}\mathbf{C}_2 = \mathbf{0}$. The desired result in (3.3) follows from (5.13) and the independence of $\varrho_1,\dots,\varrho_n$.

(iii) Suppose that global weight $w_i^n = w_i(\hat\gamma)$ is used. By (C2), we have

$$\frac{n}{\sqrt{Nh}}\sum_{i=1}^{n}w_i(\hat\gamma)\varrho_i = \frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m}\varrho_{ij}\big[1 - h_i(\hat\gamma)^T\lambda_{\hat\gamma}\{1 + o_p(1)\}\big].$$

Recall $h_{i\ell}(\cdot)$ defined in (2.4).

We have

$$E\left\{\frac{1}{Nh}\sum_{i=1}^{n}\sum_{j=1}^{m}\varrho_{ij}h_{i\ell}(\hat\gamma)^T\right\}$$

$$= E\left\{\frac{1}{Nh}\sum_{i=1}^{n}\sum_{j=1}^{m}K_{ij}Z_{ij}(\tau - \mathbf{1}_{\xi_{ij}<0})h_{i\ell}(\gamma_0)^T\right\} + o(1)$$

$$= E\left[\frac{1}{Nh}\sum_{i=1}^{n}\sum_{j=1}^{m}K_{ij}Z_{ij}E\{(\tau - \mathbf{1}_{\xi_{ij}<0})h_{i\ell}(\gamma_0)^T|U_{ij}\}\right] + o(1) \to p_T(t)G_\ell(t)^T.$$

By the law of large numbers, we have

$$\frac{1}{Nh}\sum_{i=1}^{n}\sum_{j=1}^{m}\varrho_{ij}h_i(\hat\gamma)^T \xrightarrow{p} p_T(t)\mathbf{G}(t)^T. \tag{5.16}$$

By (5.16) and (C1) using steps similar to (5.15),

$$\begin{aligned}
\text{var}&\left\{\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m}\varrho_{ij}h_i(\hat\gamma)^T\lambda_{\hat\gamma}\right\}\\
&= \text{var}\Big\{\sqrt{Nh}p_T(t)\mathbf{G}(t)^T\lambda_{\hat\gamma}\Big\}\{1 + o(1)\}\\
&= \text{var}\Big\{\sqrt{Nh}p_T(t)\mathbf{G}(t)^T\mathbf{L}\bar h_n(\gamma_0)\Big\}\{1 + o(1)\}\\
&= mhp_T^2(t)\mathbf{G}(t)^T\mathbf{L}E\left[\frac{1}{n}\sum_{i=1}^{n}h_i(\gamma_0)h_i(\gamma_0)^T\right]\mathbf{L}\mathbf{G}(t)\{1 + o(1)\}\\
&= mhp_T^2(t)\mathbf{G}(t)^T\mathbf{L}\mathbf{G}(t) + o(1) = o(1). \tag{5.17}
\end{aligned}$$

Similarly, we can obtain

$$\text{cov}\Big(\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m}\varrho_{ij}h_i(\hat\gamma)^T\lambda_{\hat\gamma}, \frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m}\varrho_{ij}\Big) = o(1).$$

Thus, we have

$$\mathrm{var}\left\{\frac{n}{\sqrt{Nh}}\sum_{i=1}^{n}w_i(\hat{\gamma})\varrho_i\right\} \to \tau(1-\tau)p_T(t)\varphi_K\Sigma_Z(t).$$

The desired result in (3.4) follows from (5.13) and the independence of $\varrho_1, \ldots, \varrho_n$. $\square$

Proof of Theorem 2. Let $\theta = \sqrt{n}(\alpha - \alpha_\tau)$. Then we can write

$$Y_{ij}^{\#} - X_{ij}^T\alpha = Y_{ij} - Z_{ij}^T\hat{\beta}_\tau(t_{ij}|\alpha_\tau) - X_{ij}^T\alpha = \xi_{ij} - X_{ij}^T\theta/\sqrt{n} - d_{ij},$$

where $d_{ij} = Z_{ij}^T\{\hat{\beta}_\tau(t_{ij}|\alpha_\tau) - \beta_\tau(t_{ij}|\alpha_\tau)\}$. Similar to the proof of Theorem 1, we can find the asymptotic Bahadur representation:

$$
\begin{aligned}
\sqrt{n}(\hat{\alpha}_\tau - \alpha_\tau) &= \frac{n\Sigma_{fX}^{-1}}{\sqrt{n}}\sum_{i=1}^{n}w_i(\hat{\gamma})\sum_{j=1}^{m}(\tau - \mathbf{1}_{\xi_{ij} - d_{ij} < 0})X_{ij} + o_p(1) \\
&= B_n - C_n + o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
B_n &= \frac{n\Sigma_{fX}^{-1}}{\sqrt{n}}\sum_{i=1}^{n}w_i(\hat{\gamma})\sum_{j=1}^{m}(\tau - \mathbf{1}_{\xi_{ij} < 0})X_{ij}, \\
C_n &= \frac{n\Sigma_{fX}^{-1}}{\sqrt{n}}\sum_{i=1}^{n}w_i(\hat{\gamma})\sum_{j=1}^{m}(\mathbf{1}_{\xi_{ij} - d_{ij} < 0} - \mathbf{1}_{\xi_{ij} < 0})X_{ij}.
\end{aligned}
$$

By a similar argument in (5.8) and the result in (5.13), we have $C_n =$

$$
\begin{aligned}
&\frac{n\Sigma_{fX}^{-1}}{\sqrt{n}}\sum_{i=1}^{n}w_i(\hat{\gamma})\sum_{j=1}^{m}f_\xi(0|U_{ij})Z_{ij}^T\left\{\hat{\beta}_\tau(t_{ij}|\alpha_\tau) - \beta_\tau(t_{ij}|\alpha_\tau)\right\}X_{ij} + o_p(1) \\
=\ &\frac{n\Sigma_{fX}^{-1}}{\sqrt{n}}\sum_{i=1}^{n}w_i(\hat{\gamma})\sum_{j=1}^{m}f_\xi(0|U_{ij})Z_{ij}^T\left\{\frac{\Sigma_{fZ}^{-1}(t_{ij})}{p_T(t_{ij})}\frac{n}{Nh}\sum_{i'=1}^{n}w_i^n\right. \\
&\qquad\qquad \times\sum_{j'=1}^{m}(\tau - \mathbf{1}_{\xi_{i'j'} < 0})K_{i'j'}Z_{i'j'} + \left.\frac{h^2\beta_\tau''(t_{ij})\mu_K}{2}\right\}X_{ij} + o_p(1) \\
=\ &\frac{n\Sigma_{fX}^{-1}}{\sqrt{n}}\sum_{i'=1}^{n}w_i^n\sum_{j'=1}^{m}(\tau - \mathbf{1}_{\xi_{i'j'} < 0})\left\{\frac{n}{Nh}\sum_{i=1}^{n}w_i(\hat{\gamma})\right. \\
&\qquad\qquad \times\sum_{j=1}^{m}\frac{f_\xi(0|U_{ij})}{p_T(t_{ij})}K\left(\frac{t_{ij} - t_{i'j'}}{h}\right)X_{ij}Z_{ij}^T\Sigma_{fZ}^{-1}(t_{ij})\left.\right\}Z_{i'j'} \\
&\quad +\frac{\sqrt{n}h^2\mu_K}{2}\Sigma_{fX}^{-1}\left\{\frac{n}{n}\sum_{i=1}^{n}w_i(\hat{\gamma})\sum_{j=1}^{m}f_\xi(0|U_{ij})X_{ij}Z_{ij}^T\beta_\tau''(t_{ij})\right\} + o_p(1)
\end{aligned}
$$

$$= \frac{n\Sigma_{fX}^{-1}}{\sqrt{n}} \sum_{i=1}^{n} w_i^n \sum_{j=1}^{m} (\tau - \mathbf{1}_{\xi_{ij}<0}) H_{ij} + \frac{\sqrt{n}h^2\mu_K}{2} \Sigma_{fX}^{-1} \mathbf{b}_{\alpha_\tau} + o_p(1).$$

It follows that $\sqrt{n}(\hat{\alpha}_\tau - \alpha_\tau) =$

$$\frac{n\Sigma_{fX}^{-1}}{\sqrt{n}} \left\{ \sum_{i=1}^{n} w_i(\hat{\gamma}) \sum_{j=1}^{m} (\tau - \mathbf{1}_{\xi_{ij}<0})(X_{ij} - H_{ij}) \right.$$

$$\left. - \sum_{i=1}^{n} \{w_i^n - w_i(\hat{\gamma})\} \sum_{j=1}^{m} (\tau - \mathbf{1}_{\xi_{ij}<0}) H_{ij} \right\} + \frac{\sqrt{n}h^2\mu_K}{2} \Sigma_{fX}^{-1} \mathbf{b}_{\alpha_\tau} + o_p(1).$$

(i) Suppose that global weight $w_i^n = w_i(\hat{\gamma})$ is used. By Theorem 1 in [19],

$$\sqrt{n} \left(\hat{\alpha}_\tau - \alpha_\tau - \frac{h^2\mu_k}{2} \mathbf{b}_{\alpha_\tau} \right) \xrightarrow{d} N(0, \Sigma_{fX}^{-1} V \Sigma_{fX}^{-1}).$$

(ii) Suppose that $w_i^n = w_i^e = 1/n$ is used. Then by (C2), $n\{w_i^n - w_i(\hat{\gamma})\} = h_i(\hat{\gamma})^T \lambda_{\hat{\gamma}} + o_p(1)$. Let $\mathcal{I}_{i1} = \sum_{j=1}^{m} (\tau - \mathbf{1}_{\xi_{ij}<0}) X_{ij}$ and $\mathcal{I}_{i2} = \sum_{j=1}^{m} (\tau - \mathbf{1}_{\xi_{ij}<0}) H_{ij}$. We can write

$$\frac{n}{\sqrt{n}} \sum_{i=1}^{n} w_i(\hat{\gamma}) \sum_{j=1}^{m} (\tau - \mathbf{1}_{\xi_{ij}<0})(X_{ij} - H_{ij})$$

$$- \frac{n}{\sqrt{n}} \sum_{i=1}^{n} \{w_i^n - w_i(\hat{\gamma})\} \sum_{j=1}^{m} (\tau - \mathbf{1}_{\xi_{ij}<0}) H_{ij}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{1 - h_i(\hat{\gamma})^T \lambda_{\hat{\gamma}}\}(\mathcal{I}_{i1} - \mathcal{I}_{i2}) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_i(\hat{\gamma})^T \lambda_{\hat{\gamma}} \mathcal{I}_{i2} + o_p(1).$$

Since $\text{cov}\{n^{-1/2} \sum_{i=1}^{n} h_i(\hat{\gamma})^T \lambda_{\hat{\gamma}} \mathcal{I}_{ik}, n^{-1/2} \sum_{i=1}^{n} h_i(\hat{\gamma})^T \lambda_{\hat{\gamma}} \mathcal{I}_{ik'}\} = \text{cov}\{n^{-1/2} \sum_{i=1}^{n} \mathcal{I}_{ik}, n^{-1/2} \sum_{i=1}^{n} h_i(\hat{\gamma})^T \lambda_{\hat{\gamma}} \mathcal{I}_{ik'}\}$ for $k = 1, 2$ and $k' = 1, 2$, we have $\text{cov}[n^{-1/2} \sum_{i=1}^{n} \{1 - h_i(\hat{\gamma})^T \lambda_{\hat{\gamma}}\}(\mathcal{I}_{i1} - \mathcal{I}_{i2}), n^{-1/2} \sum_{i=1}^{n} h_i(\hat{\gamma})^T \lambda_{\hat{\gamma}} \mathcal{I}_{i2}] = 0$. Using a similar step to (5.15), we have

$$\text{var} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_i(\hat{\gamma})^T \lambda_{\hat{\gamma}} \mathcal{I}_{i2} \right\} = \mathcal{B}\mathbf{L}\mathcal{B}.$$

Thus,

$$\sqrt{n} \left\{ \hat{\alpha} - \alpha_\tau - \frac{h^2\mu_k}{2} \mathbf{b}_{\alpha_\tau} \right\} \xrightarrow{d} N(0, \Sigma_{fX}^{-1}(V + \mathcal{B}\mathbf{L}\mathcal{B})\Sigma_{fX}^{-1}).$$

(iii) Suppose that local weight $w_i^n = w_i\{\hat{\delta}(t)\}$ is used. Then by (C2), $n\{w_i^n - w_i(\hat{\gamma})\} = h_i(\hat{\gamma})^T \lambda_{\hat{\gamma}} - g_i\{\hat{\delta}(t)\}^T \lambda_{\hat{\alpha}(t)} + o_p(1)$. We can write

$$\frac{n}{\sqrt{n}} \sum_{i=1}^{n} w_i(\hat{\gamma}) \sum_{j=1}^{m} (\tau - \mathbf{1}_{\xi_{ij}<0})(X_{ij} - H_{ij})$$

$$- \frac{n}{\sqrt{n}} \sum_{i=1}^{n} \{w_i^n - w_i(\hat{\gamma})\} \sum_{j=1}^{m} (\tau - \mathbf{1}_{\xi_{ij}<0})H_{ij}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{1 - h_i(\hat{\gamma})^T \lambda_{\hat{\gamma}}\}(\mathcal{I}_{i1} - \mathcal{I}_{i2})$$

$$- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [h_i(\hat{\gamma})^T \lambda_{\hat{\gamma}} - g_i\{\hat{\delta}(t)\}^T \lambda_{\hat{\alpha}(t)}]\mathcal{I}_{i2} + o_p(1).$$

Similar to (5.17), $\mathrm{cov}[n^{-1/2} \sum_{i=1}^{n} \mathcal{I}_{ik}, n^{-1/2} \sum_{i=1}^{n} g_i\{\hat{\delta}(t)\}^T \lambda_{\hat{\alpha}(t)}\mathcal{I}_{i2}] = O(h)$ and $\mathrm{cov}[n^{-1/2} \sum_{i=1}^{n} h_i(\hat{\gamma})^T \lambda_{\hat{\gamma}}\mathcal{I}_{ik}, n^{-1/2} \sum_{i=1}^{n} g_i\{\hat{\delta}(t)\}^T \lambda_{\hat{\alpha}(t)}\mathcal{I}_{i2}] = O(h)$ for $k = 1, 2$ and $\mathrm{var}[n^{-1/2} \sum_{i=1}^{n} g_i\{\hat{\delta}(t)\}^T \lambda_{\hat{\alpha}(t)}\mathcal{I}_{i2}] = O(h)$. Then the rest of the proof is the same as (ii) above. □

Proof of Theorem 3. The proof of Theorem 3 follows from almost the same argument as given in [8, 19], and thus it is omitted here but available from the authors upon request. □

Acknowledgements

References

[1] CAI, Z. AND XIAO, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics* **167**, 413–425. MR2892085

[2] CAI, Z. AND XU, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association* **103**, 1595–1608. MR2504207

[3] CHO, H., HONG, H. G. AND KIM, M. (2016). Efficient quantile marginal regression for longitudinal data with dropouts. *Biostatistics* **17**, 561–575. MR3603954

[4] FAN, J., HUANG, T. AND LI, R. Z. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of American Statistical Association* **35**, 632–641. MR2370857

[5] FU, L. AND WANG, Y. G. (2012). Quantile regression for longitudinal data with a working correlation model. *Computational Statistics and Data Analysis* **56**, 2526–2538. MR2910067

[6] HE, X., FU, B. AND FUNG, W. K. (2003). Median regression for longitudinal data. *Statistics in Medicine* **22**, 3655–3669.

[7] Huang, J. Z., Wu, C. O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111–128. MR1888349

[8] Jin, Z., Ying, Z. and Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381–390. MR1844838

[9] Jung, S. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association* **91**, 251–257. MR1394079

[10] Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F. and Rinaldo, C. R. (1987). The multicenter AIDS cohort study: rationale, organization and selected characteristics of the participants. *American Journal of Epidemiology* **126**, 310–318.

[11] Leng, C. and Zhang, W. (2014). Smoothing combined estimating equations in quantile regression for longitudinal data. *Statistics and Computing* **1**, 123–136. MR3147703

[12] Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 12–22. MR0836430

[13] Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association* **96**, 1045–1056. MR1947252

[14] Lu, X. and Fan, Z. (2015). Weighted quantile regression for longitudinal data. *Computational Statistics* **30**, 569–592. MR3357076

[15] Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7**, 186–199. MR1128411

[16] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics* **22**, 300–325. MR1272085

[17] Qu, A., Lindsay, B. G. and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–836. MR1813977

[18] Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B* **53**, 233–243. MR1094283

[19] Tang, C. Y. and Leng, C. (2011). Empirical likelihood and quantile regression in longitudinal data analysis. *Biometrika* **98**, 1001–1006. MR2860340

[20] Tang, Y., Wang, Y., Li, J. and Qian, W. (2015). Improving estimation efficiency in quantile regression with longitudinal data. *Journal of Statistical Planning and Inference* **165**, 38–55. MR3350257

[21] Wang, H. J., Zhu, Z. and Zhou, J. (2009). Quantile regression in partially linear varying coefficient models. *The Annals of Statistics* **37**, 3841–3866. MR2572445

[22] Wu, C. O., Chiang, C. T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel model smoothing of a varying-coefficient with longitudinal data. *Journal of the American Statistical Association* **93**, 1458–1476. MR1666635

[23] Wu, C. O. and Chiang, C. T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statistica Sinica* **10**, 433–456. MR1769751

[24] Xue, L. and Zhu, L. (2007). Empirical likelihood for a varying coefficient model with longitudinal data. *Journal of the American Statistical Association* **102**, 642–654. MR2370858

[25] Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association* **93**, 228–237. MR1614628