



Smoothed empirical likelihood inference via the modified Cholesky decomposition for quantile varying coefficient models with longitudinal data

Jing Lv¹ · Chaohui Guo² · Jibo Wu³

Received: 29 December 2017 / Accepted: 12 October 2018 / Published online: 23 October 2018
© Sociedad de Estadística e Investigación Operativa 2018

Abstract

It is essential to deal with the within-subject correlation among repeated measures over time to improve statistical inference efficiency. However, it is a challenging task to correctly specify a working correlation in quantile regression with longitudinal data. In this paper, we first develop an adaptive approach to estimate the within-subject covariance matrix of quantile regression by applying a modified Cholesky decomposition. Then, weighted kernel GEE-type quantile estimating equations are proposed for varying coefficient functions. Note that the proposed estimating equations include a discrete indicator function, which results in some problems for computation and asymptotic analysis. Thus, we construct smoothed estimating equations by introducing a bounded kernel function. Furthermore, we develop a smoothed empirical likelihood method to improve the accuracy of interval estimation. Finally, simulation studies and a real data analysis indicate that the proposed method has superior advantages over the existing methods in terms of coverage accuracies and widths of confidence intervals.

Keywords Confidence band · Longitudinal data · Modified Cholesky decomposition · Quantile regression · Robustness and efficiency

Mathematics Subject Classification 62G08 · 62G20 · 62G35

✉ Chaohui Guo
guochaohui2010@126.com

¹ School of Mathematics and Statistics, Southwest University, Chongqing 400715, China

² College of Mathematics Science, Chongqing Normal University, Chongqing 401331, China

³ School of Mathematics and Finances, Chongqing University of Arts and Sciences, Chongqing 402160, China

1 Introduction

Varying coefficient models proposed by Hastie and Tibshirani (1993) are fashionable nonparametric fitting techniques. A salient feature of varying coefficient models is that the response is linear in the covariates, but the regression coefficients are smooth nonparametric functions which depend on certain covariates such as the measurement time. Thus, varying coefficient models can be regarded as a natural extension of multiple linear regression models. Due to its flexibility to explore the dynamic features, many procedures had been proposed to estimate the varying coefficient functions. Wu et al. (1998) obtained kernel estimators of the coefficient functions by minimizing local least-squares criterion and established its asymptotic normality. Combined with a polynomial spline, Huang et al. (2002, 2004) considered the estimation of the coefficient functions by employing the least-squares method. Noh and Park (2010) developed a regularized estimation procedure for variable selection by linearizing the group smoothly clipped absolute deviation penalty. Other related literature includes Wang et al. (2008), Wang and Xia (2009) and Wang and Lin (2015).

An important feature of longitudinal data is that subjects are often measured repeatedly over a given period, which results in possible correlations between repeated measures. Thus, it is important to account for the correlation to improve the efficiency of statistical inference. However, the mentioned references above did not consider the correlations, which will result in a great loss of efficiency. Although it is an interesting topic to estimate the covariance functions in the analysis of longitudinal data, we will meet some challenges. The reasons are as follows. (i) Longitudinal data are frequently collected at irregular and possibly subject-specific time points, which leads to different dimensions of the covariance matrix for each subject. (ii) It involves more parameters to be estimated in the covariance matrix, and the positive definiteness of the covariance matrix also needs to be assured. To overcome these difficulties, the modified Cholesky decomposition approach (Pourahmadi 1999) is developed to decompose the covariance matrix, which has been demonstrated to be effective and attractive. Recently, Ye and Pan (2006) constructed several sets of parametric generalized estimating equations by utilizing the modified Cholesky decomposition. In order to relax the parametric assumption in Ye and Pan (2006), Leng et al. (2010) constructed the joint partially linear mean–covariance model based on the B-spline basis approximation. Liu and Li (2015) studied a class of marginal models with time-varying coefficients for longitudinal data by using the modified Cholesky decomposition.

However, these mentioned articles are focused on longitudinal mean regression. Quantile regression has become a powerful complement to the mean regression. Its main merits include the following two aspects. On the one hand, it can offer a more complete description of the entire distribution of the response variable. On the other hand, it does not require specification of the error distribution, and thus it is a robust regression approach. So far, quantile regression has become a widely used technique for analyzing longitudinal data. Tang and Leng (2011) applied the empirical likelihood (Owen 2001) to account for the within-subject correlations, which yields more efficient estimators than the conventional quantile regression. Leng and Zhang (2014) proposed a new quantile regression approach to account for correlations through combining multiple sets of unbiased estimating equations. Tang et al. (2015) extended this method

to composite quantile regression. Yuan et al. (2017) used a similar strategy to construct a new weighted quantile regression model with longitudinal data. However, the above literature only considered some commonly used working correlation structures (e.g., compound symmetry (CS) and first-order autoregressive [AR(1)]). When the specified working correlation structure is far from the true one, it seriously affects the estimation efficiency. However, it is difficult to justify a correct working correlation matrix in quantile regression with longitudinal data (Leng and Zhang 2014; Fu and Wang 2016). This paper focuses on longitudinal varying coefficient models and proposes a data-driven approach to estimate the covariance matrix through the modified Cholesky decomposition. The proposed method does not need to specify a working correlation structure to improve the efficiency of statistical inference, which is more flexible than the existing methods.

The confidence interval construction for varying coefficient functions is also an interesting issue. Empirical likelihood (Owen 2001) is a popular nonparametric inference method for the interval estimation. The confidence region based on empirical likelihood has many advantages. On the one hand, it does not require estimating either the unknown error density function or any covariance matrix. On the other hand, it does not impose prior constraints on the region shape, and the shape and orientation of confidence regions are determined entirely by the data. Xue and Zhu (2007) applied the local empirical likelihood technique to construct confidence intervals for varying coefficient functions, which is more accurate than the normal approximation and bootstrap methods. However, they did not consider the possible correlation between repeated measures, which will result in rough confidence intervals. Meanwhile, they only consider mean regression rather than quantile regression. Under the framework of longitudinal quantile regression, we propose an efficient smoothed empirical likelihood procedure to construct more accurate confidence intervals for varying coefficient functions.

The rest of this paper is organized as follows. In Sect. 2, we propose the smoothed estimating equations for varying coefficient functions and establish its asymptotic theories. Furthermore, we define an empirical log-likelihood ratio function and prove that its asymptotic distribution is a standard Chi-squared under mild conditions. Section 3 discusses the bandwidth choice and reports the simulation results. In Sect. 4, the proposed approach is illustrated by a longitudinal progesterone data. Section 5 ends the article with a discussion. Finally, the proofs of the main results are provided in “Appendix.”

2 Methodology

We consider the varying coefficient model with longitudinal data

$$Y_{ij} = \mathbf{X}_i^T(t_{ij}) \boldsymbol{\beta}(t_{ij}) + \varepsilon_i(t_{ij}), \quad i = 1, \dots, n, \quad j = 1, \dots, m_i, \quad (1)$$

where $Y_{ij} = Y_i(t_{ij})$, $\mathbf{X}_i(t_{ij}) = (X_{i1}(t_{ij}), \dots, X_{ip}(t_{ij}))^T$, $t_{ij} \in \mathcal{R}$, $\boldsymbol{\beta}(t) = (\beta_1(t), \dots, \beta_p(t))^T$, $\beta_r(t)$ are smooth functions of t and $\beta_r(t) \in \mathcal{R}$ for all

$r = 1, \dots, p$, $\varepsilon_i(t)$ are stochastic processes and independent of $\mathbf{X}_i(t)$. In this paper, we assume that the design points $\{t_{ij}, i = 1, \dots, n, j = 1, \dots, m_i\}$ are random and iid according to an underlying density function $f_T(\cdot)$ with respect to the Lebesgue measure, and $S(f_T)$ is the support of f_T . We assume that the random error $\varepsilon_{ij} = \varepsilon_i(t_{ij})$ satisfies $P(\varepsilon_i(t_{ij}) < 0 | t_{ij}) = \tau$ and with an unspecified density function $f_{ij}(\cdot)$. Then it is easy to show that

$$E\left\{\tau - I\left(Y_{ij} - \mathbf{X}_i^T(t_{ij})\boldsymbol{\beta}(t_{ij}) < 0\right) | t_{ij}\right\} = 0, \quad (2)$$

where $I(\cdot)$ is a indicator function. Note that the random errors are correlated within the same subject but independent across the subjects.

2.1 Efficient estimating equations for regression coefficients $\boldsymbol{\beta}(t)$

Based on the local constant approximation, the standard quantile regression estimator $\hat{\boldsymbol{\beta}}(t; h_1)$ of $\boldsymbol{\beta}(t)$ can be obtained by minimizing

$$\sum_{i=1}^n \sum_{j=1}^{m_i} \rho_\tau\left(Y_{ij} - \mathbf{X}_i^T(t_{ij})\boldsymbol{\beta}(t)\right) K_{h_1}(t - t_{ij})$$

with respect to $\boldsymbol{\beta}(t)$, where $K_{h_1}(\cdot) = K(\cdot/h_1)$, $K(\cdot)$ is a kernel function and h_1 is a bandwidth, $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is the quantile loss function. And equivalently, the resulting estimator $\hat{\boldsymbol{\beta}}(t; h_1)$ is the solution of the following estimating equations

$$\sum_{i=1}^n \mathbf{X}_i^T \mathbf{K}_i(t; h_1) \psi_\tau(\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}(t)) = \mathbf{0}, \quad (3)$$

where $\mathbf{K}_i(t; h_1) = \text{diag}(K_{h_1}(t - t_{i1}), \dots, K_{h_1}(t - t_{im_i}))$, $\mathbf{X}_i = (\mathbf{X}_i(t_{i1}), \dots, \mathbf{X}_i(t_{im_i}))^T$, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im_i})^T$, $\psi_\tau(u) = \tau - I(u < 0)$ is the quantile score function and

$$\psi_\tau(\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}(t)) = \left\{ \psi_\tau\left(Y_{i1} - \mathbf{X}_i^T(t_{i1})\boldsymbol{\beta}(t)\right), \dots, \psi_\tau\left(Y_{im_i} - \mathbf{X}_i^T(t_{im_i})\boldsymbol{\beta}(t)\right) \right\}^T.$$

Estimating equations (3) are obtained under the independent working model; hence, the efficiency of the estimators $\hat{\boldsymbol{\beta}}(t; h_1)$ can be improved if the within correlations are incorporated. To incorporate the correlation within subjects, we can use the estimating equations that take the form

$$\sum_{i=1}^n \mathbf{X}_i^T \mathbf{K}_i(t; h_1) \boldsymbol{\Sigma}_i^{-1} \psi_\tau(\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}(t)) = \mathbf{0}, \quad (4)$$

where $\Sigma_i = \text{Cov}(\psi_\tau(\mathbf{e}_i))$ and $\mathbf{e}_i = (\varepsilon_{i1}, \dots, \varepsilon_{im_i})^\top$. But we cannot directly obtain the estimator of $\beta(t)$ by solving (4) because the estimating equations include the unknown covariance matrix Σ_i .

In this paper, we estimate Σ_i via the modified Cholesky decomposition. That is, $\Sigma_i = \mathbf{L}_i \mathbf{D}_i \mathbf{L}_i^\top$, where \mathbf{L}_i is a lower triangular matrix with 1's on its diagonal and $\mathbf{D}_i = \text{diag}(d_{i1}^2, \dots, d_{im_i}^2)$. Let $\mathbf{e}_i = \mathbf{L}_i^{-1} \psi_\tau(\mathbf{e}_i) = (e_{i1}, \dots, e_{im_i})^\top$ with $e_{ij} = e_i(t_{ij})$, we have

$$\psi_\tau(\varepsilon_{ij}) = \sum_{k=1}^{j-1} l_{ijk} e_{ik} + e_{ij}, \tag{5}$$

where l_{ijk} are the below diagonal entries of \mathbf{L}_i . The sum $\sum_{k=1}^{j-1}$ is interpreted as zero when $j = 1$ throughout this paper. Note that $E(\mathbf{e}_i) = \mathbf{0}$ and $\text{Cov}(\mathbf{e}_i) = \mathbf{D}_i = \text{diag}(d_{i1}^2, \dots, d_{im_i}^2)$ for $i = 1, \dots, n$. Thus we know that $e_{ij}, j = 1, \dots, m_i$ are uncorrelated and $d_{ij}^2 = d^2(t_{ij})$ is called as innovation variance for $j = 1, \dots, m_i$. The merit of this decomposition is that the parameters l_{ijk} and d_{ij}^2 are unconstrained. In the spirit of Pourahmadi (1999), we construct the following regression models to estimate l_{ijk} and d_{ij}^2

$$l_{ijk} = \mathbf{w}_{ijk}^\top \boldsymbol{\gamma}, d_{ij}^2 = g(t_{ij}), \tag{6}$$

where $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^\top$, $g(t)$ is an unknown smooth function of t , and \mathbf{w}_{ijk} is usually taken as a polynomial of time difference $t_{ij} - t_{ik}$. Some authors applied similar parametric models to parsimoniously parameterize the covariance matrix Σ_i when considering the modified Cholesky decomposition (Ye and Pan 2006; Zhang and Leng 2012). Compared with parametric models, we relax the parametric assumption and propose a nonparametric model to estimate innovation variance d_{ij}^2 , which reduce the risk of model misspecification. Eyeballing (6), we can use regression models to describe the within-subject correlation among repeated measures. Furthermore, model-based analysis of l_{ijk} and d_{ij}^2 permits more accessible statistical inference.

Now we give specific details to estimate l_{ijk} and d_{ij}^2 . Based on $\hat{\beta}(t; h_1)$, we can obtain the estimated residuals

$$\hat{\varepsilon}_{ij} = Y_{ij} - \mathbf{X}_i^\top(t_{ij}) \hat{\beta}(t_{ij}; h_1), \quad i = 1, \dots, n, \quad j = 1, \dots, m_i. \tag{7}$$

From (5)–(7), we can obtain the estimator $\hat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ by using the GEE method

$$\sum_{i=1}^n \mathbf{V}_i^\top \tilde{\mathbf{e}}_i = \mathbf{0},$$

where $\tilde{\mathbf{e}}_i = \mathbf{L}_i^{-1} \psi_\tau(\hat{\varepsilon}_i) = (\tilde{e}_{i1}, \dots, \tilde{e}_{im_i})^\top$ with $\tilde{e}_{ij} = \psi_\tau(\hat{\varepsilon}_{ij}) - \sum_{k=1}^{j-1} l_{ijk} \tilde{e}_{ik}$ and $\mathbf{V}_i^\top = \partial \tilde{\mathbf{e}}_i^\top / \partial \boldsymbol{\gamma}$ is a $q \times m_i$ matrix with the first column zero and the j th $j \geq 2$ column

$\partial \tilde{e}_{ij} / \partial \boldsymbol{\gamma} = - \sum_{k=1}^{j-1} [\mathbf{w}_{ijk} \tilde{e}_{ik} + l_{ijk} \partial \tilde{e}_{ik} / \partial \boldsymbol{\gamma}]$. Then, we have $\hat{l}_{ijk} = \mathbf{w}_{ijk}^T \hat{\boldsymbol{\gamma}}$. Please note that $E(e_{ij}^2) = \text{Var}(e_{ij}) + [E(e_{ij})]^2 = \text{Var}(e_{ij}) = d_{ij}^2$ due to $E(e_i) = \mathbf{0}$ and $\text{Cov}(e_i) = \mathbf{D}_i$. Thus, a local polynomial regression technique can be used to estimate the nonparametric function $g(\cdot)$ (Fan and Yao 1998). Specifically, we can obtain the estimator $\hat{d}^2(t) = \hat{a}$ by minimizing the following weighted least-squares problem

$$(\hat{a}, \hat{b}) = \arg \min_{a,b} \sum_{i=1}^n \sum_{j=1}^{m_i} \left\{ \hat{e}_{ij}^2 - a - b(t_{ij} - t) \right\}^2 K_{h_2}(t - t_{ij}),$$

where $\hat{e}_{ij} = \psi_\tau(\hat{e}_{ij}) - \sum_{k=1}^{j-1} \hat{l}_{ijk} \hat{e}_{ik}$. Therefore, we can obtain $\hat{\boldsymbol{\Sigma}}_i = \hat{\mathbf{L}}_i \hat{\mathbf{D}}_i \hat{\mathbf{L}}_i^T$, where $\hat{\mathbf{L}}_i$ is an $m_i \times m_i$ lower triangular matrix with 1's on its diagonal and the below diagonal entries of $\hat{\mathbf{L}}_i$ are $\hat{l}_{ijk} = \mathbf{w}_{ijk}^T \hat{\boldsymbol{\gamma}}$ and $\hat{\mathbf{D}}_i = \text{diag}(\hat{d}_{i1}^2, \dots, \hat{d}_{im_i}^2)$ with $\hat{d}_{ij}^2 = \hat{d}^2(t_{ij})$.

Replacing $\boldsymbol{\Sigma}_i$ by $\hat{\boldsymbol{\Sigma}}_i$, estimating equations (4) can be rewritten as

$$\sum_{i=1}^n \mathbf{X}_i^T \mathbf{K}_i(t; h_1) \hat{\boldsymbol{\Sigma}}_i^{-1} \psi_\tau(Y_i - \mathbf{X}_i \boldsymbol{\beta}(t)) = \mathbf{0}. \tag{8}$$

However, proposed estimating equations (8) are neither smooth nor monotone. In order to overcome the calculation difficulty, we approximate $\psi_\tau(\cdot)$ by a smooth function $\psi_{\tau h}(\cdot)$ based on the idea of Wang and Zhu (2011). Define $G(x) = \int_{u < x} K_1(u) du$ and $G_h(x) = G(x/h)$, where $K_1(\cdot)$ is a kernel function and h is a positive bandwidth parameter. Then, we approximate $\psi_\tau(\cdot)$ by $\psi_{\tau h}(\cdot) = \tau - 1 + G_h(\cdot)$. Therefore, based on the approximation, estimating equations (8) can be replaced by the following smoothed estimating equations

$$\sum_{i=1}^n \mathbf{X}_i^T \mathbf{K}_i(t; h_1) \hat{\boldsymbol{\Sigma}}_i^{-1} \psi_{\tau h}(Y_i - \mathbf{X}_i \boldsymbol{\beta}(t)) = \mathbf{0}, \tag{9}$$

where $\psi_{\tau h}(Y_i - \mathbf{X}_i \boldsymbol{\beta}(t)) = (\psi_{\tau h}\{Y_{i1} - \mathbf{X}_i^T(t_{i1}) \boldsymbol{\beta}(t)\}, \dots, \psi_{\tau h}\{Y_{im_i} - \mathbf{X}_i^T(t_{im_i}) \boldsymbol{\beta}(t)\})^T$. We assume that $\bar{\boldsymbol{\beta}}(t; h_1)$ is the solution of estimating equations (9).

We assume that t_0 is an interior point of $S(f_T)$. Next, we derive the asymptotic distribution of $\hat{\boldsymbol{\gamma}}$ and establish the large sample properties of $\hat{d}^2(t_0)$ and $\bar{\boldsymbol{\beta}}(t_0; h_1)$ at a fixed point t_0 . Let $\rho_\varepsilon(t_0) = \lim_{\delta \rightarrow 0} E\{[\psi_\tau(\varepsilon_1(t_0 + \delta))][\psi_\tau(\varepsilon_1(t_0))]\}$, $\eta_{lr}(t_0) = E[X_{il}(t_{ij}) X_{ir}(t_{ij}) | t_{ij} = t_0]$, $\mu_j = \int u^j K(u) du$ and $\nu_j = \int u^j K^2(u) du$. In this paper, $\dot{d}(t)$ and $\ddot{d}(t)$ stand for the first and second derivative of $d(t)$ respectively.

Theorem 1 *Let $\boldsymbol{\gamma}_0$ be the true value of $\boldsymbol{\gamma}$. Under conditions (C1)–(C8) stated in ‘‘Appendix,’’ the proposed estimator $\hat{\boldsymbol{\gamma}}$ is \sqrt{n} -consistent and asymptotically normal, that is,*

$$\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Xi}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Xi}^{-1}),$$

where \xrightarrow{d} means the convergence in distribution, $\Xi = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{V}_i^T \mathbf{V}_i$ and $\Gamma = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{V}_i^T \mathbf{D}_i \mathbf{V}_i$.

In order to obtain the asymptotic result of $\hat{d}^2(t)$, we rewrite (5) as follows:

$$\psi_\tau(\varepsilon_{ij}) = \sum_{k=1}^{j-1} l_{ijk} d_{ik} \varsigma_{ik} + d_{ij} \varsigma_{ij}, \tag{10}$$

where $e_{ij} = d_{ij} \varsigma_{ij}$, $E(\varsigma_{ij} | t_{ij}) = 0$ and $\text{Var}(\varsigma_{ij} | t_{ij}) = 1$.

Theorem 2 Under the same set of conditions as in Theorem 1, together with $Nh_2 \rightarrow \infty$ as $n \rightarrow \infty$ and $\limsup_{n \rightarrow \infty} Nh_2^5 < \infty$, we have

$$\begin{aligned} & \sqrt{Nh_2} \left\{ \hat{d}^2(t_0) - d^2(t_0) - \frac{1}{2} \mu_2 h_2^2 \ddot{d}^2(t_0) \right\} \\ & \xrightarrow{d} N \left(0, \frac{\nu_0}{f_T(t_0)} \lim_{n \rightarrow \infty} \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{m_i} E \left[(\varsigma_{ij}^2 - 1)^2 | t_{ij} = t_0 \right] d^4(t_0) \right). \end{aligned}$$

Let $b_l(t_0) = \mu_2 h_1^2 \sum_{r=1}^p [\dot{\beta}_r(t_0) \dot{\eta}_{lr}(t_0) f_T(t_0) + \frac{1}{2} \ddot{\beta}_r(t_0) \eta_{lr}(t_0) f_T(t_0) + \dot{\beta}_r(t_0) \eta_{lr}(t_0) \dot{f}_T(t_0)]$ and $\mathbf{b}(t_0) = (b_1(t_0), \dots, b_p(t_0))^T$. Let $\hat{\sigma}^{jj'}$ be the (j, j') th element of $\hat{\Sigma}_i^{-1}$ and $\Phi(t_0)$ be given in condition (C7).

Theorem 3 Suppose that conditions (C1)–(C10) hold, we have

$$\sqrt{Nh_1} \left(\bar{\beta}(t_0; h_1) - \beta(t_0) - \mathbf{B}(t_0) \right) \xrightarrow{d} N \left(\mathbf{0}, \mathbf{\Omega}_1^{-1} (\mathbf{\Omega}_2 + \mathbf{\Omega}_3) \mathbf{\Omega}_1^{-1} \right),$$

where $\mathbf{B}(t_0) = f_T^{-1}(t_0) \Phi^{-1}(t_0) \mathbf{b}(t_0)$, $\mathbf{\Omega}_1 = \lim_{n \rightarrow \infty} \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{j'=1}^{m_i} f_T(t_0) \hat{\sigma}^{jj'} f_{ij'}(0) \Phi(t_0)$,

$$\begin{aligned} \mathbf{\Omega}_2 &= \lim_{n \rightarrow \infty} \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{m_i} \nu_0 f_T(t_0) \\ & \quad \times \left[\sum_{j'=1}^{m_i} (\hat{\sigma}^{jj'})^2 \tau(1-\tau) + \sum_{j_1 \neq j_2} \hat{\sigma}^{j_1 j_1} \hat{\sigma}^{j_2 j_2} \rho_\varepsilon(t_0) \right] \Phi(t_0), \\ \mathbf{\Omega}_3 &= \lim_{n \rightarrow \infty} \frac{1}{N} \sum_{i=1}^n \sum_{j_1 \neq j_2} h_1 f_T^2(t_0) \\ & \quad \times \left[\sum_{j'=1}^{m_i} \hat{\sigma}^{j_1 j'} \hat{\sigma}^{j_2 j'} \tau(1-\tau) + \sum_{j'' \neq j'''} \hat{\sigma}^{j_1 j''} \hat{\sigma}^{j_2 j'''} \rho_\varepsilon(t_0) \right] \Phi(t_0). \end{aligned}$$

2.2 Empirical likelihood method

To construct the empirical likelihood ratio function for $\beta(t)$, based on Owen (2001) and Qin and Lawless (1994), we introduce an auxiliary random vector as follows $Z_{ih}(\beta(t)) = X_i^T K_i(t; h_1) \hat{\Sigma}_i^{-1} \psi_{\tau h}(Y_i - X_i \beta(t))$. Note that $\{Z_{ih}(\beta(t)) : 1 \leq i \leq n\}$ are independent. Invoking (2), it can be shown that $E\{Z_{ih}(\beta(t))\} = o(1)$ when $\beta(t)$ is the true value.

Let p_1, \dots, p_n be nonnegative numbers satisfying $\sum_{i=1}^n p_i = 1$, where $p_i = p_i(t), i = 1, \dots, n$. Using such information, a natural block empirical log-likelihood ratio for $\beta(t)$ is defined as

$$l(\beta(t)) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i Z_{ih}(\beta(t)) = 0 \right\}. \tag{11}$$

Tang and Leng (2011) pointed out that for finite sample, existence of a solution to empirical likelihood for a given $\beta(t)$ requires that 0 is inside the convex hull of the points $\{Z_{ih}(\beta(t)), i = 1, \dots, n\}$. By the Lagrange multiplier method, the optimal value for p_i is given by

$$p_i = n^{-1} \left\{ 1 + \lambda^T Z_{ih}(\beta(t)) \right\}^{-1}, \tag{12}$$

where λ is a p -dimensional Lagrange multiplier satisfying

$$\sum_{i=1}^n \frac{Z_{ih}(\beta(t))}{1 + \lambda^T Z_{ih}(\beta(t))} = 0. \tag{13}$$

By (11) and (12), $l(\beta(t))$ can be represented as

$$l(\beta(t)) = 2 \sum_{i=1}^n \log \left\{ 1 + \lambda^T Z_{ih}(\beta(t)) \right\} \tag{14}$$

with λ satisfying (13). In our numerical studies, we can employ the modified Newton–Raphson algorithm of Chen et al. (2002) to obtain the estimator of λ . We further define the maximum empirical likelihood (EL) estimator of $\beta(t)$ as

$$\hat{\beta}(t)_{EL} = \arg \min_{\beta(t) \in \mathcal{R}^p} l(\beta(t)).$$

Remark 1 From (14) and (A.20), we can prove that the maximum empirical likelihood estimator $\hat{\beta}(t)_{EL}$ is asymptotically the solution of $\sum_{i=1}^n Z_{ih}(\beta(t)) = \mathbf{0}$. Let

$$\hat{V}(t) = \frac{1}{Nh_1} \sum_{i=1}^n X_i^T K_i(t; h_1) \hat{\Sigma}_i^{-1} \tilde{A}_i X_i,$$

and

$$\tilde{\mathbf{A}}_i = \text{diag} \left\{ h^{-1} \mathbf{K}_1 \left[\left(Y_{i1} - \mathbf{X}_i(t_{i1})^T \boldsymbol{\beta}(t) \right) / h \right], \dots, h^{-1} \mathbf{K}_1 \left[\left(Y_{im_i} - \mathbf{X}_i(t_{im_i})^T \boldsymbol{\beta}(t) \right) / h \right] \right\}.$$

If we assume that the matrix $\hat{\mathbf{V}}(t)$ is invertible together with (14), then we can obtain

$$\hat{\boldsymbol{\beta}}(t)_{\text{EL}} - \boldsymbol{\beta}(t) = \hat{\mathbf{V}}(t)^{-1} \left\{ \frac{1}{Nh_1} \sum_{i=1}^n \mathbf{X}_i^T \mathbf{K}_i(t; h_1) \hat{\boldsymbol{\Sigma}}_i^{-1} \psi_{\tau h}(Y_i - \mathbf{X}_i \boldsymbol{\beta}(t)) \right\} + o_p \left((Nh_1)^{-1/2} \right).$$

Then we can see that the leading term is that of $\bar{\boldsymbol{\beta}}(t; h_1)$, seeing the proof of Theorem 3 in ‘‘Appendix.’’ In other words, the $\hat{\boldsymbol{\beta}}(t)_{\text{EL}}$ is asymptotically equivalent to the $\bar{\boldsymbol{\beta}}(t; h_1)$.

Theorem 4 *Suppose that conditions (C1)–(C10) hold. If $\boldsymbol{\beta}(t_0)$ is the true parameter, then $l(\boldsymbol{\beta}(t_0)) \xrightarrow{d} \chi_p^2$, where χ_p^2 is the standard Chi-squared distribution with p degrees of freedom.*

Remark 2 By Theorem 4, we can use the test statistic $l(\boldsymbol{\beta}(t_0))$ to obtain confidence regions for $\boldsymbol{\beta}(t_0)$. Specifically, we define $\{\boldsymbol{\beta}(t) : l(\boldsymbol{\beta}(t)) \leq \chi_{1-\alpha}^2(p)\}$ as the EL confidence regions for $\boldsymbol{\beta}(t_0)$, where $\chi_{1-\alpha}^2(p)$ is the $(1 - \alpha)$ th quantile of χ_p^2 .

In addition, we can apply the profile empirical log likelihood ratio test statistic to obtain the confidence region of the subset of $\boldsymbol{\beta}(t_0)$. Let $\boldsymbol{\beta}(t_0) = (\boldsymbol{\beta}^{(1)}(t_0)^T, \boldsymbol{\beta}^{(2)}(t_0)^T)^T$, where $\boldsymbol{\beta}^{(1)}(t_0)$ and $\boldsymbol{\beta}^{(2)}(t_0)$ are $p_1 \times 1$ and $(p - p_1) \times 1$ vectors, respectively. If we are interested in testing $H_0 : \boldsymbol{\beta}^{(1)}(t_0) = \mathbf{b}(t_0)$, where $\mathbf{b}(t_0)$ is some known $p_1 \times 1$ vector. The profile log likelihood ratio test statistic is defined as

$$\bar{l}(\mathbf{b}(t_0)) = l(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t)) - l(\hat{\boldsymbol{\beta}}^{(1)}(t), \hat{\boldsymbol{\beta}}^{(2)}(t)),$$

where $\tilde{\boldsymbol{\beta}}^{(2)}(t)$ minimizes $l(\mathbf{b}(t_0), \boldsymbol{\beta}^{(2)}(t))$ with respect to $\boldsymbol{\beta}^{(2)}(t)$, and $(\hat{\boldsymbol{\beta}}^{(1)T}(t), \hat{\boldsymbol{\beta}}^{(2)T}(t))^T$ are the EL estimators.

Corollary 1 *Under the conditions of Theorem 4 and $H_0 : \boldsymbol{\beta}^{(1)}(t_0) = \mathbf{b}(t_0)$, we have $\bar{l}(\mathbf{b}(t_0)) \xrightarrow{d} \chi_{p_1}^2$ as $n \rightarrow \infty$.*

Remark 3 According to Corollary 1, we can construct the confidence region for $\boldsymbol{\beta}^{(1)}(t_0)$, namely $\{\mathbf{b}(t) : l(\mathbf{b}(t)) \leq \chi_{1-\alpha}^2(p_1)\}$.

3 Simulation studies

In this section, we conduct simulation studies to investigate the finite sample performance of the proposed method. In the following examples, we fix the kernel function to be the Epanechnikov kernel, that is, $K(u) = 0.75(1 - u^2)_+$. We apply the fivefold cross-validation to choose the optimal bandwidths h_1 and h_2 . Taking an example of the h_1 , let $T - T^v$ and T^v be the cross-validation training and test set respectively for $v = 1, \dots, 5$, where T is the full dataset. Note that in model (1), the observations are dependent within each subject. To retain this dependence structure, we treat the observations in each subject as a block. For each h_1 and v , we find the estimator $\hat{\beta}(t; h_1, v)$ of $\beta(t)$ using the training set $T - T^v$; then, we can form the fivefold cross-validation criterion as

$$\text{CV}(h_1) = \sum_{v=1}^5 \sum_{(Y_{ij}, X_{ij}) \in T^v} \rho_\tau \left(Y_{ij} - \mathbf{X}_i^T(t_{ij}) \hat{\beta}(t_{ij}; h_1, v) \right).$$

By using the grid search method, we find $h_{1\text{opt}} = \min_{h_1} \text{CV}(h_1)$. Based on the idea of Wang and Zhu (2011), we smooth the quantile score function by the following second-order ($\nu = 2$) Bartlett kernel

$$K_1(u) = \frac{3}{4\sqrt{5}} \left(1 - u^2/5 \right) I(|u| \leq \sqrt{5}).$$

Wang and Zhu (2011) had proved that the smoothed approach is robust to the bandwidth h . Thus, we fix $h = n^{-0.4}$ which satisfies the theoretical requirement $nh^{2\nu} \rightarrow 0$ with $\nu = 2$.

In addition, we compare the following estimation approaches. (i) The conventional quantile regression estimator $\hat{\beta}(t; h_1)$ without considering possible correlations denoted as $\hat{\beta}_{\text{qr}}$. (ii) The estimator proposed by Xue and Zhu (2007) denoted as $\hat{\beta}_{\text{ls}}$. (iii) The proposed estimator $\bar{\beta}(t; h_1)$ denoted as $\hat{\beta}_{\text{pr}}$. (iv) The proposed empirical likelihood (EL) estimator $\hat{\beta}(t)_{\text{EL}}$ denoted as $\hat{\beta}_{\text{el}}$. For $\hat{\beta}_{\text{ls}}$ and $\hat{\beta}_{\text{el}}$, we construct confidence regions/intervals based on the estimated empirical likelihood (EL) as suggested in Sect. 2.2; for $\hat{\beta}_{\text{qr}}$ and $\hat{\beta}_{\text{pr}}$, we construct confidence regions/intervals based on the normal approximation (NA). To assess the performance of the empirical likelihood, we make a comparison between the EL and NA in terms of coverage accuracies and average widths of the confidence intervals. In the following simulation studies, we focus on $\tau = 0.5$ and 0.75 , and choose $n = 50$ and 200 representing moderate and large sample sizes. The number of simulation replications is set to 500 .

Example 1 For simplicity, we consider model (1) with a time-independent covariate $X_i = (X_{i1}, X_{i2})^T$, where X_i follows a multivariate normal distribution $N(\mathbf{0}, \Sigma_x)$ with $(\Sigma_x)_{k,l} = 0.5^{|k-l|}$ for $1 \leq k, l \leq 2$. The coefficient functions are given by $\beta_1(t) = 6 - 0.2t$, $\beta_2(t) = -4 + (20 - t)^3/2000$. Each subject is measured m_i times with $m_i - 1 \sim \text{Binomial}(9, 0.8)$, which results in different numbers of repeated measurements for each subject. Here, we define $\varepsilon_{ij} = \xi_{ij} - c_\tau$, and c_τ is the τ th

quantile of the random error ξ_{ij} , which implies the corresponding τ th quantile of ε_{ij} is zero. Meanwhile, we consider two error distributions of $\xi_i = (\xi_{i1}, \dots, \xi_{im_i})^T$ for assessing the robustness and effectiveness of the proposed method.

Case 1: correlated normal error, ξ_i , is generated from a multivariate normal distribution $N(\mathbf{0}, \Sigma_i)$, where Σ_i will be listed later.

Case 2: ξ_i is generated from a multivariate t -distribution with degree 3 and covariance matrix Σ_i .

We construct the covariance matrix Σ_i of ε_i as $\Sigma_i = \Delta_i^{-1} \mathbf{B}_i (\Delta_i^T)^{-1}$, where \mathbf{B}_i is an $m_i \times m_i$ diagonal matrix with the j th element $\sin(\pi t_{ij})/3 + 1.5$ and Δ_i is a unit lower triangular matrix with (j, k) element $-\delta_{j,k}^{(i)}$ ($k < j$), $\delta_{j,k}^{(i)} = 0.3 + 0.6(t_{ij} - t_{ik})$ and $t_{ij} \sim U(0, 1)$. Similar to Zhang and Leng (2012), we take the covariate in model (6) as $\mathbf{w}_{ijk} = \{1, t_{ij} - t_{ik}, (t_{ij} - t_{ik})^2, (t_{ij} - t_{ik})^3\}^T$.

Example 2 The coefficient functions are given by $\beta_1(t) = \cos(2\pi t)$, $\beta_2(t) = 4t(1 - t)$. We set the covariance matrix as $\Sigma_i = \mathbf{D}_i^{1/2} \mathbf{R}(\rho) \mathbf{D}_i^{1/2}$, $\mathbf{D}_i = \text{diag}(\sigma_{i1}^2, \dots, \sigma_{im_i}^2)$, with $\sigma_{ij}^2 = \sigma(t_{ij})$, where

$$\sigma(t) = \begin{cases} |50t^2 - 2|, & 0 \leq t \leq 0.3 \\ |2t|, & 0.3 < t \leq 0.6 \\ |4t|, & 0.6 < t \leq 1 \end{cases}$$

In addition, $\mathbf{R}(\rho)$ is AR(1) or compound symmetry structure with correlation coefficient $\rho = 0.9$. Other settings are the same as that in Example 1.

Tables 1, 2, 3 and 4 give the biases, the standard deviations (SD), estimated coverage probabilities (CP) and average lengths (AL) of 95% confidence intervals for different methods at $\tau = 0.5$ and 0.75. Tables 2 and 4 only list the results of $\hat{\beta}_{qr}$, $\hat{\beta}_{pr}$ and $\hat{\beta}_{el}$ at $\tau = 0.75$. The reason is that there is no comparison between the least-squares estimator and the quantile estimation at $\tau = 0.75$. From the four tables, we can obtain the following findings. (1) All methods yield asymptotic unbiased estimators, since the corresponding biases are small for the large sample size $n = 200$. (2) For the point estimation, $\hat{\beta}_{pr}$ and $\hat{\beta}_{el}$ have similar performances for all settings due to similar biases and standard deviations. In other words, the $\hat{\beta}_{pr}$ is asymptotically equivalent to the $\hat{\beta}_{el}$, which is consistent with our theory. (3) For the interval estimation, it is easy to find that $\hat{\beta}_{el}$ has the shortest confidence intervals and achieves the highest empirical coverage probabilities among all methods. Thus, the proposed empirical likelihood method $\hat{\beta}_{el}$ significantly improves the accuracy of the confidence intervals. Finally, Fig. 1 depicts the sample standard deviations of different methods at $\tau = 0.5$ and 0.75 in Example 1, and we can see that $\hat{\beta}_{pr}$ has the smallest sample standard deviations.

Overall, the simulation results show that the proposed estimators $\hat{\beta}_{pr}$ and $\hat{\beta}_{el}$ generally work well and outperform other existing methods.

Table 1 Simulation results of the bias, SD, estimated coverage probabilities (CP) and average lengths (AL) for different methods with $n = 50$, $\tau = 0.5$ at $t_0 = 0.2, 0.7$ in Example 1

	t_0		$\hat{\beta}_1(t_0)$				$\hat{\beta}_2(t_0)$			
			Bias	SD	CP	AL	Bias	SD	CP	AL
Case 1	0.2	$\hat{\beta}_{ls}$	-0.0037	0.1488	0.934	0.5587	-0.0003	0.1510	0.926	0.5600
		$\hat{\beta}_{qr}$	-0.0032	0.1723	0.916	0.6322	0.0045	0.1845	0.912	0.6351
		$\hat{\beta}_{pr}$	-0.0034	0.1459	0.904	0.5384	0.0033	0.1407	0.912	0.5425
		$\hat{\beta}_{el}$	-0.0034	0.1459	0.930	0.5235	0.0033	0.1407	0.930	0.5339
Case 1	0.7	$\hat{\beta}_{ls}$	-0.0051	0.2654	0.896	0.8195	0.0212	0.2590	0.894	0.8234
		$\hat{\beta}_{qr}$	-0.0036	0.2466	0.892	0.8891	0.0234	0.2329	0.924	0.9063
		$\hat{\beta}_{pr}$	0.0079	0.1970	0.894	0.6961	0.0123	0.1781	0.908	0.6833
		$\hat{\beta}_{el}$	0.0078	0.1970	0.892	0.6372	0.0122	0.1781	0.920	0.6370
Case 2	0.2	$\hat{\beta}_{ls}$	-0.0056	0.2449	0.876	0.7834	-0.0078	0.2685	0.894	0.7871
		$\hat{\beta}_{qr}$	-0.0091	0.1823	0.906	0.7359	-0.0009	0.1975	0.910	0.7460
		$\hat{\beta}_{pr}$	-0.0083	0.1580	0.898	0.5968	0.0023	0.1661	0.914	0.6102
		$\hat{\beta}_{el}$	-0.0083	0.1580	0.928	0.5876	0.0024	0.1661	0.942	0.5967
Case 2	0.7	$\hat{\beta}_{ls}$	-0.0032	0.4141	0.757	0.8880	-0.0011	0.4129	0.741	0.8905
		$\hat{\beta}_{qr}$	0.0174	0.2688	0.914	0.9680	-0.0101	0.2516	0.898	0.9876
		$\hat{\beta}_{pr}$	0.0140	0.2134	0.878	0.7491	0.0075	0.1937	0.906	0.7503
		$\hat{\beta}_{el}$	0.0140	0.2134	0.937	0.6962	0.0075	0.1937	0.922	0.6890

Table 2 Simulation results of the bias, SD, estimated coverage probabilities (CP) and average lengths (AL) for different methods with $n = 50$, $\tau = 0.75$ at $t_0 = 0.2, 0.7$ in Example 1

	t_0		$\hat{\beta}_1(t_0)$				$\hat{\beta}_2(t_0)$			
			Bias	SD	CP	AL	Bias	SD	CP	AL
Case 1	0.2	$\hat{\beta}_{qr}$	-0.0035	0.1777	0.902	0.6582	0.0074	0.1747	0.914	0.6850
		$\hat{\beta}_{pr}$	-0.0047	0.1524	0.904	0.5640	0.0124	0.1481	0.912	0.5712
		$\hat{\beta}_{el}$	-0.0047	0.1524	0.926	0.5466	0.0124	0.1481	0.928	0.5478
Case 1	0.7	$\hat{\beta}_{qr}$	0.0022	0.2607	0.906	0.9598	0.0054	0.2721	0.880	0.9458
		$\hat{\beta}_{pr}$	0.0017	0.1917	0.912	0.7573	0.0116	0.2001	0.866	0.7473
		$\hat{\beta}_{el}$	0.0012	0.1919	0.914	0.6662	0.0120	0.2026	0.900	0.6641
Case 2	0.2	$\hat{\beta}_{qr}$	-0.0100	0.2121	0.900	0.7992	0.0053	0.2071	0.904	0.7978
		$\hat{\beta}_{pr}$	-0.0020	0.1856	0.870	0.6485	0.0039	0.1793	0.892	0.6459
		$\hat{\beta}_{el}$	-0.0019	0.1853	0.906	0.6328	0.0032	0.1776	0.918	0.6319
Case 2	0.7	$\hat{\beta}_{qr}$	0.0070	0.2988	0.890	1.1043	0.0114	0.2998	0.914	1.1223
		$\hat{\beta}_{pr}$	0.0057	0.2201	0.888	0.7982	0.0142	0.2157	0.910	0.8171
		$\hat{\beta}_{el}$	0.0059	0.2199	0.914	0.7151	0.0150	0.2163	0.904	0.7222

Table 3 Simulation results of the bias, SD, estimated coverage probabilities (CP) and average lengths (AL) for different methods with $\tau = 0.5$ at $t_0 = 0.7$ in Example 2

n		$\hat{\beta}_1(t_0)$				$\hat{\beta}_2(t_0)$					
		Bias	SD	CP	AL	Bias	SD	CP	AL		
50	Case 1	cs	$\hat{\beta}_{ls}$	0.0482	0.1903	0.908	0.6903	0.1890	0.902	0.6740	
			$\hat{\beta}_{qr}$	-0.0100	0.2312	0.884	0.8637	0.2207	0.902	0.8293	
			$\hat{\beta}_{pr}$	0.0045	0.1697	0.872	0.6248	0.1584	0.894	0.5723	
			$\hat{\beta}_{el}$	0.0045	0.1697	0.908	0.5823	0.1584	0.920	0.5469	
			$\hat{\beta}_{ls}$	0.0355	0.2081	0.882	0.6927	0.1963	0.914	0.6704	
	ar1	$\hat{\beta}_{qr}$	-0.0220	0.2446	0.878	0.8671	0.2342	0.870	0.8012		
		$\hat{\beta}_{pr}$	-0.0078	0.1983	0.888	0.6723	0.1878	0.852	0.6223		
		$\hat{\beta}_{el}$	-0.0077	0.1983	0.912	0.6336	0.1878	0.892	0.5936		
		$\hat{\beta}_{ls}$	0.0711	0.3109	0.850	0.8387	0.2889	0.854	0.8273		
		$\hat{\beta}_{qr}$	0.0165	0.2595	0.901	0.9429	0.2540	0.902	0.8870		
Case 2	cs	cs	$\hat{\beta}_{ls}$	0.0213	0.1848	0.874	0.6614	0.1772	0.879	0.5947	
			$\hat{\beta}_{pr}$	0.0225	0.1841	0.906	0.6137	0.1776	0.903	0.5755	
			$\hat{\beta}_{el}$	0.0615	0.3332	0.850	0.8449	0.3251	0.856	0.8383	
			$\hat{\beta}_{ls}$	-0.0013	0.2685	0.892	0.9393	0.2551	0.896	0.8935	
			$\hat{\beta}_{qr}$	0.0097	0.2019	0.858	0.7133	0.1969	0.874	0.6725	
	ar1	ar1	ar1	$\hat{\beta}_{pr}$	0.0109	0.2033	0.912	0.6618	0.1977	0.902	0.6257
				$\hat{\beta}_{el}$							
				$\hat{\beta}_{ls}$							
				$\hat{\beta}_{qr}$							
				$\hat{\beta}_{pr}$							

Table 3 continued

n		$\hat{\beta}_1(t_0)$				$\hat{\beta}_2(t_0)$					
		Bias	SD	CP	AL	Bias	SD	CP	AL		
200	Case 1	cs	$\hat{\beta}_{ls}$	0.0303	0.1038	0.952	0.4209	-0.0231	0.0979	0.942	0.4077
			$\hat{\beta}_{qr}$	0.0020	0.1295	0.922	0.5045	-0.0123	0.1249	0.912	0.4836
			$\hat{\beta}_{pr}$	0.0113	0.0940	0.912	0.3589	-0.0139	0.0898	0.900	0.3340
		ar1	$\hat{\beta}_{el}$	0.0113	0.0940	0.948	0.3565	-0.0139	0.0898	0.936	0.3383
			$\hat{\beta}_{ls}$	0.0261	0.1172	0.918	0.4203	-0.0162	0.1059	0.934	0.4146
			$\hat{\beta}_{qr}$	0.0047	0.1437	0.890	0.5031	-0.0105	0.1294	0.912	0.4834
	Case 2	cs	$\hat{\beta}_{pr}$	0.0122	0.1127	0.888	0.3886	-0.0121	0.0978	0.912	0.3693
			$\hat{\beta}_{el}$	0.0122	0.1127	0.926	0.3921	-0.0121	0.0978	0.948	0.3720
			$\hat{\beta}_{ls}$	0.0176	0.1795	0.930	0.6412	-0.0176	0.1730	0.928	0.6355
		ar1	$\hat{\beta}_{qr}$	-0.0197	0.1454	0.908	0.5473	-0.0030	0.1363	0.928	0.5308
			$\hat{\beta}_{pr}$	0.0006	0.1017	0.898	0.3754	-0.0093	0.0966	0.926	0.3530
			$\hat{\beta}_{el}$	0.0006	0.1017	0.918	0.3709	-0.0093	0.0966	0.936	0.3555
	ar1	$\hat{\beta}_{ls}$	0.0362	0.1719	0.922	0.6522	-0.0218	0.1872	0.924	0.6526	
		$\hat{\beta}_{qr}$	0.0079	0.1429	0.922	0.5410	-0.0098	0.1396	0.910	0.5191	
		$\hat{\beta}_{pr}$	0.0160	0.1092	0.920	0.4157	-0.0168	0.1013	0.904	0.3944	
		$\hat{\beta}_{el}$	0.0160	0.1092	0.940	0.4141	-0.0168	0.1012	0.936	0.3898	

Table 4 Simulation results of the bias, SD, estimated coverage probabilities (CP) and average lengths (AL) for different methods with $\tau = 0.75$ at $t_0 = 0.7$ in Example 2

n		$\hat{\beta}_1(t_0)$				$\hat{\beta}_2(t_0)$					
		Bias	SD	CP	AL	Bias	SD	CP	AL		
50	Case 1	cs	$\hat{\beta}_{qr}$	0.0210	0.2367	0.886	0.9055	-0.0303	0.2333	0.902	0.8315
				0.0223	0.1668	0.886	0.6429	-0.0346	0.1679	0.874	0.5924
				0.0223	0.1668	0.906	0.5888	-0.0346	0.1679	0.908	0.5462
	Case 2	ar1	$\hat{\beta}_{pr}$	-0.0138	0.2370	0.882	0.8401	-0.0205	0.2320	0.878	0.8241
				0.0098	0.1834	0.880	0.6638	-0.0404	0.1684	0.878	0.6360
				0.0098	0.1834	0.904	0.6245	-0.0404	0.1684	0.900	0.5882
50	Case 1	cs	$\hat{\beta}_{el}$	0.0035	0.2655	0.882	0.9333	-0.0237	0.2474	0.880	0.9083
				0.0269	0.1983	0.852	0.6756	-0.0383	0.1795	0.860	0.6484
				0.0269	0.1983	0.892	0.6151	-0.0383	0.1795	0.892	0.5784
	Case 2	ar1	$\hat{\beta}_{qr}$	0.0060	0.2716	0.872	0.9545	-0.0210	0.2546	0.894	0.9140
				0.0233	0.2031	0.890	0.7279	-0.0376	0.1830	0.886	0.6843
				0.0233	0.2031	0.898	0.6695	-0.0376	0.1830	0.900	0.6377

Table 4 continued

<i>n</i>	$\hat{\beta}_1(t_0)$					$\hat{\beta}_2(t_0)$					
		Bias	SD	CP	AL	Bias	SD	CP	AL		
200	Case 1	cs	$\hat{\beta}_{qr}$	0.0069	0.1385	0.906	0.5190	-0.0155	0.1338	0.914	0.4838
			$\hat{\beta}_{pr}$	0.0173	0.0996	0.904	0.3644	-0.0169	0.0932	0.906	0.3363
			$\hat{\beta}_{el}$	0.0173	0.0996	0.914	0.3620	-0.0167	0.0933	0.930	0.3407
	ar1	$\hat{\beta}_{qr}$	-0.0072	0.1377	0.908	0.5206	-0.0096	0.1356	0.916	0.4990	
		$\hat{\beta}_{pr}$	0.0053	0.1092	0.916	0.3958	-0.0141	0.1010	0.922	0.3777	
		$\hat{\beta}_{el}$	0.0051	0.1093	0.934	0.3972	-0.0141	0.1010	0.950	0.3752	
Case 2	cs	$\hat{\beta}_{qr}$	0.0111	0.1464	0.926	0.5910	-0.0143	0.1472	0.928	0.5584	
		$\hat{\beta}_{pr}$	0.0179	0.0965	0.926	0.4021	-0.0167	0.0995	0.900	0.3686	
		$\hat{\beta}_{el}$	0.0179	0.0965	0.944	0.3987	-0.0167	0.0995	0.948	0.3737	
	ar1	$\hat{\beta}_{qr}$	0.0013	0.1497	0.912	0.5747	-0.0139	0.1468	0.922	0.5596	
		$\hat{\beta}_{pr}$	0.0137	0.1135	0.912	0.4268	-0.0200	0.1103	0.906	0.4100	
		$\hat{\beta}_{el}$	0.0137	0.1135	0.940	0.4275	-0.0200	0.1103	0.940	0.4078	

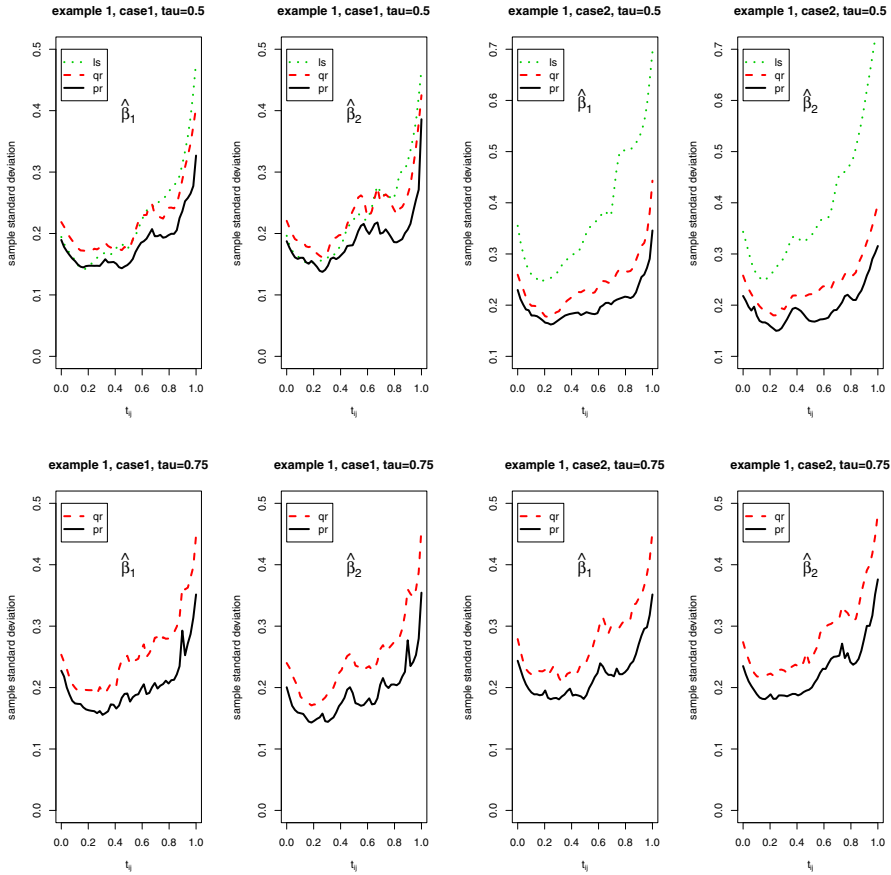


Fig. 1 The sample standard deviations of different methods at $\tau = 0.5$ and $\tau = 0.75$ with $n = 50$ in Example 1. Here “ls” is the estimator proposed by Xue and Zhu (2007); “qr” is the conventional quantile regression estimator $\hat{\beta}(t; h_1)$ without considering possible correlations; “pr” stands for the proposed estimator $\hat{\beta}(t; h_1)$

4 Real data analysis

In this section, we illustrate the proposed method by analyzing a longitudinal progesterone data (Zhang et al. 1998). The dataset used here consists of 492 observations of progesterone level within a menstrual cycle from 34 women clinical participants. Zheng et al. (2013) constructed robust parametric mean-covariance regression model to analyze this dataset. Here we consider the following quantile varying coefficient model

$$Q(Y_{ij}|X_i(t_{ij})) = \beta_0(t_{ij}) + X_{i1}(t_{ij})\beta_1(t_{ij}) + X_{i2}(t_{ij})\beta_2(t_{ij}), \quad i = 1, \dots, 34; \\ j = 1, \dots, m_i,$$

where $Q(Y_{ij}|X_i(t_{ij}))$ is the τ th conditional quantile of Y_{ij} given $X_i(t_{ij})$. Here X_{i1} and X_{i2} represent patient's age and body mass index, t_{ij} is taken as the repeated measurement time, and the log-transformed progesterone level is defined as the response Y_{ij} . Then $\beta_0(t)$ stands for the intercept, and $\beta_1(t)$ and $\beta_2(t)$ describe the effects of woman's age and body mass index on the progesterone level at time t . Before implementing estimation procedures, we normalize all predictor variables and rescale the repeated measurement time t_{ij} into interval $[0,1]$. In this real data analysis, we consider conventional quantile regression estimator $\hat{\beta}_{qr}$ and the proposed estimation approaches $\hat{\beta}_{pr}$ and $\hat{\beta}_{el}$ at quantile levels $\tau = 0.25$ and $\tau = 0.5$. For $\hat{\beta}_{qr}$ and $\hat{\beta}_{pr}$, we use normal approximation to construct the confidence intervals, and the corresponding standard error is calculated by the bootstrap resampling method. The confidence intervals of $\hat{\beta}_{el}$ are constructed by the empirical likelihood method. For covariance model (6), we take the covariate \mathbf{w}_{ijk} as $\mathbf{w}_{ijk} = \{1, t_{ij} - t_{ik}, (t_{ij} - t_{ik})^2, (t_{ij} - t_{ik})^3\}^T$.

Figure 2 displays the estimated varying coefficient functions and their 95% confidence intervals for $\tau = 0.25$ and $\tau = 0.5$. For the estimated values, there is little difference for $\hat{\beta}_{pr}$ and $\hat{\beta}_{el}$, which has been confirmed by simulations given in Sect. 3. So we do not display the estimated curves of $\hat{\beta}_{el}$ in Fig. 2. Obviously, the effects of AGE and BMI are insignificant for all methods due to the large standard errors. These results are consistent with that of Zheng et al. (2013). Eyeballing Fig. 3, it is clear that $\hat{\beta}_{el}$ performs best among all methods because it has the shortest widths of confidence intervals. This indicates that the proposed estimation procedure is more efficient.

5 Conclusion and discussion

This paper considers robust GEE analysis for varying coefficient quantile regression models with longitudinal data based on the modified Cholesky decomposition. Asymptotic normalities for the estimators of the coefficient functions and the parameters in the decomposition of covariance matrix are established. Smoothed quantile score estimating equations are proposed to facilitate computation. We also developed block empirical likelihood-based inference procedures for varying coefficient functions to improve the accuracy of interval estimation. Simulations and an real data analysis have showed that the proposed methods are clear superior to other existing methods.

This paper does not prove that the proposed method is more efficient than the conventional quantile approach which does not deal with the within-subject correlation among repeated measures over time. The main reason is that we cannot guarantee covariance model (6) is always correct. In practice, all models are wrong since no one knows which model is optimal for estimating l_{ijk} and d_{ij}^2 . Although theoretical results about estimation efficiency cannot be established in this article, the efficiency of our method is investigated by simulations and real data analysis, which indicates that our method is more useful in practice. Recently, Li (2011) utilized the non-

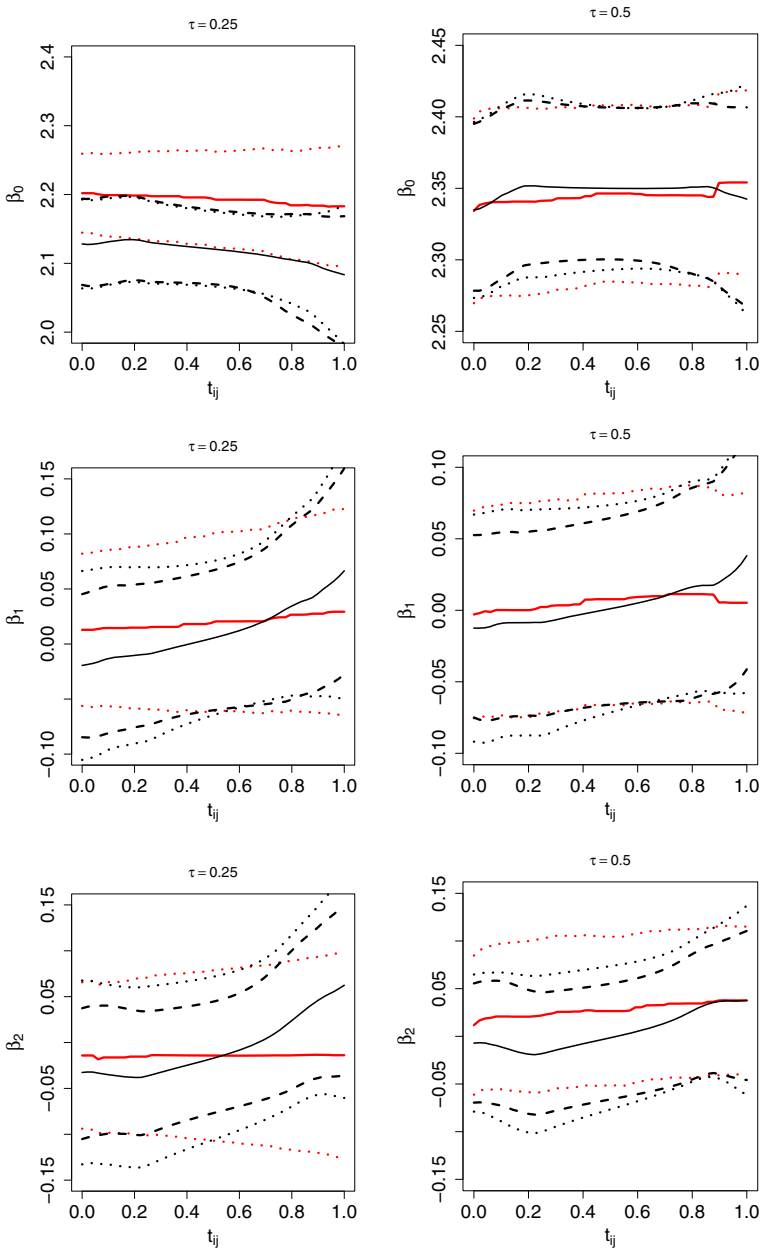


Fig. 2 Estimated nonparametric curves and their 95% confidence intervals of different methods at $\tau = 0.25$ and $\tau = 0.5$. The red and black solid lines stand for the estimated values of $\hat{\beta}_{qr}$ and $\hat{\beta}_{pr}$. The red and black dotted lines represent 95% confidence intervals of $\hat{\beta}_{qr}$ and $\hat{\beta}_{pr}$. The black dashed lines represent 95% confidence intervals of $\hat{\beta}_{el}$

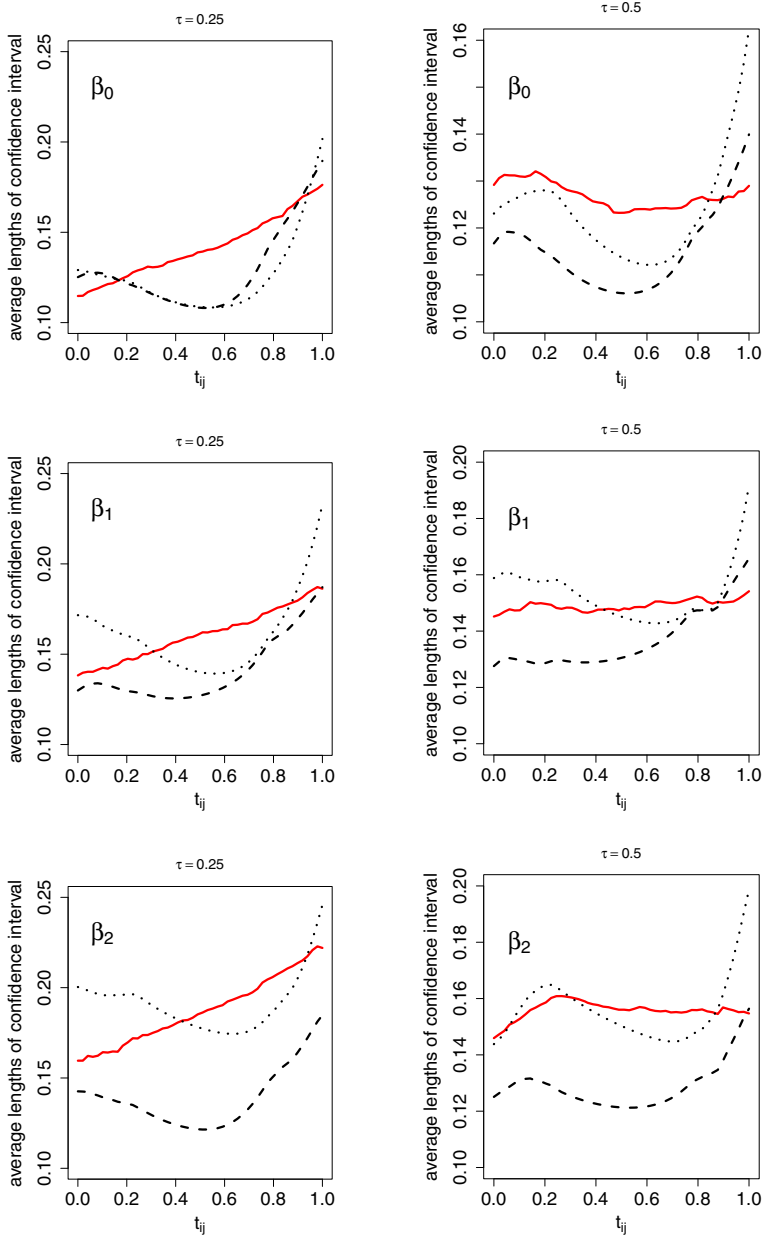


Fig. 3 Average lengths of 95% confidence intervals for different methods at $\tau = 0.25$ and $\tau = 0.5$. The red solid lines, black dotted lines and black dashed lines represent average interval lengths of $\hat{\beta}_{qr}$, $\hat{\beta}_{pr}$ and $\hat{\beta}_{ci}$ respectively

parametric kernel technique to estimate the within-subject covariance matrix. There is no model assumption to estimate the covariance matrix, and thus his method is robust against misspecification of covariance models. Furthermore, he proved that the proposed nonparametric covariance estimation is uniformly consistent. The semi-parametrical efficiency of mean regression coefficients is also established. However, he only focused on mean regression. Therefore, nonparametric covariance estimation procedures for quantile regression should be developed to estimate the covariance matrix Σ_i involved in formula (6), but this is beyond the scope of this paper. We will focus on the study of these aspects in the future.

Acknowledgements The authors are very grateful to the editor and anonymous referees for their detailed comments on the earlier version of the manuscript, which leads to a much improved paper. This work is supported by the National Social Science Fund of China (Grant No. 17CTJ015).

Appendix

To establish the asymptotic properties of the proposed estimators, the following regularity conditions are needed in this paper.

(C1) The bandwidth satisfies $h_1 = N^{-1/5}h_0$ for some constant $h_0 > 0$.

(C2) $\lim_{n \rightarrow \infty} N^{-6/5} \sum_{i=1}^n m_i^2 = \kappa$ for some $0 \leq \kappa < \infty$.

(C3) The kernel function $K(\cdot)$ has a compact support on \mathcal{R} and satisfies $\int K(u) du = 1$, $\int K^2(u) du < \infty$, $\int u^2 K(u) du < \infty$, $\int u K(u) du = 0$ and $\int u^4 K(u) du < \infty$.

(C4) There exists a constant $\delta \in (2/5, 2]$, and we have $\sup_t E \left(|\psi_\tau(\varepsilon_1(t_{ij}))|^{2+\delta} |t_{ij} = t \right) < \infty$ and $\sup_t E \left(|\psi_{\tau h}(\varepsilon_1(t_{ij}))|^{2+\delta} |t_{ij} = t \right) < \infty$ for all $i = 1, \dots, n$, $j = 1, \dots, m_i$, $l = 1, \dots, p$ and $t \in S(f_T)$.

(C5) For all $l, r = 1, \dots, p$, $\beta_r(t)$, $\eta_{lr}(t)$ and $f_T(t)$ have continuous second derivatives at t_0 .

(C6) $\left\{ (Y_{ij}, \mathbf{X}_i^T(t_{ij}))^T, j = 1, \dots, m_i \right\}$ are independently and identically distributed for $i = 1, \dots, n$. We assume that the dimension p of the covariates $\mathbf{X}_i(t_{ij})$ is fixed and there is a positive constant M such that $|X_{il}(t)| \leq M$ for all t and $i = 1, \dots, n$, $l = 1, \dots, p$.

(C7) The covariance function $\rho_\varepsilon(t)$ is continuous at t_0 , and $\Phi(t_0) = \left(\eta_{lr}(t_0)_{l,r=1}^p \right)$ is positive definite matrixes.

(C8) The distribution function $F_{ij}(x) = P(Y_{ij} - \mathbf{X}_i^T(t_{ij})\boldsymbol{\beta}(t_{ij}) < x | t_{ij})$ is absolutely continuous, with continuous densities $f_{ij}(\cdot)$ uniformly bounded, and its first derivative uniformly bounded away from 0 and ∞ at the points 0, $i = 1, \dots, n$, $j = 1, \dots, m_i$.

(C9) $K_1(\cdot)$ is a symmetric density function with a bounded support in \mathcal{R} . For some constant $C_K \neq 0$, $K_1(\cdot)$ is a ν th-order kernel, i.e., $\int u^j K_1(u) du = 1$ if $j = 0$; 0 if $1 \leq j \leq \nu - 1$; C_K if $j = \nu$, where $\nu \geq 2$ is an integer.

(C10) The positive bandwidth parameter h satisfies $nh^{2\nu} \rightarrow 0$.

Lemma 1 *Suppose that conditions (C2)–(C10) hold and that the bandwidth satisfies $\sup \lim_{n \rightarrow \infty} Nh_1^5 < \infty$. If $\boldsymbol{\beta}(t_0)$ is the true parameter, then $\max_{1 \leq i \leq n} \|Z_i(\boldsymbol{\beta}(t_0))\| =$*

$o_p(\sqrt{Nh_1})$ and $\max_{1 \leq i \leq n} \|Z_{ih}(\boldsymbol{\beta}(t_0))\| = o_p(\sqrt{Nh_1})$, where $\|\cdot\|$ is the Euclidean norm.

Proof The proof of Lemma 1 is omitted since it is similar to the proof of Lemma A.1 in Xue and Zhu (2007). \square

Lemma 2 Suppose that conditions (C1)–(C8) hold, we have

$$\sqrt{Nh_1} \left(\hat{\boldsymbol{\beta}}(t_0; h_1) - \boldsymbol{\beta}(t_0) - \mathbf{B}(t_0) \right) \xrightarrow{d} N(\mathbf{0}, \bar{\mathbf{D}}(t_0)),$$

where $\bar{\mathbf{D}}(t_0) = (f_T(t_0)\bar{f}(0))^{-2} v^2(t_0) \boldsymbol{\Phi}^{-1}(t_0)$ and $\bar{f}(0) = \lim_{n \rightarrow \infty} N^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} f_{ij}(0)$, $\mathbf{B}(t_0)$ and $\boldsymbol{\Phi}(t_0)$ are defined in Theorem 3 and condition (C7), $v^2(t_0) = \tau(1-\tau)f_T(t_0)v_0 + \kappa h_0 \rho_\varepsilon(t_0)f_T^2(t_0)$, h_0 and κ are defined by conditions (C1) and (C2).

Proof Because

$$\begin{aligned} & \sum_{i=1}^n \mathbf{X}_i^T \mathbf{K}_i(t; h_1) \psi_\tau(\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}(t)) \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} \psi_\tau(\mathbf{Y}_{ij} - \mathbf{X}_i^T(t_{ij}) \boldsymbol{\beta}(t)) \mathbf{X}_i(t_{ij}) K_{h_1}(t - t_{ij}) \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} \psi_\tau(\varepsilon_i(t_{ij}) + \mathbf{X}_i^T(t_{ij}) \boldsymbol{\beta}(t_{ij}) - \mathbf{X}_i^T(t_{ij}) \boldsymbol{\beta}(t)) \mathbf{X}_i(t_{ij}) K_{h_1}(t - t_{ij}) \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{X}_i(t_{ij}) K_{h_1}(t - t_{ij}) \left[\tau - I \left\{ \varepsilon_i(t_{ij}) < \mathbf{X}_i^T(t_{ij}) (\boldsymbol{\beta}(t) - \boldsymbol{\beta}(t_{ij})) \right\} \right] \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{X}_i(t_{ij}) K_{h_1}(t - t_{ij}) \left[\psi_\tau(\varepsilon_i(t_{ij})) + I \left\{ \varepsilon_i(t_{ij}) < 0 \right\} \right. \\ & \quad \left. - I \left\{ \varepsilon_i(t_{ij}) < \mathbf{X}_i^T(t_{ij}) (\boldsymbol{\beta}(t) - \boldsymbol{\beta}(t_{ij})) \right\} \right] \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{X}_i(t_{ij}) K_{h_1}(t - t_{ij}) \psi_\tau(\varepsilon_i(t_{ij})) \\ & \quad - \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{X}_i(t_{ij}) K_{h_1}(t - t_{ij}) \Upsilon_i(t_{ij}) \\ & \triangleq \mathbf{I} - \mathbf{II}, \end{aligned} \tag{A.1}$$

where $\Upsilon_i(t_{ij}) = [I\{\varepsilon_i(t_{ij}) < \mathbf{X}_i^T(t_{ij})(\boldsymbol{\beta}(t) - \boldsymbol{\beta}(t_{ij}))\} - I\{\varepsilon_i(t_{ij}) < 0\}]$. In addition, we define

$$\begin{aligned} \mathbf{II} &= \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{X}_i(t_{ij}) K_{h_1}(t - t_{ij}) E(\Upsilon_i(t_{ij})) \\ &\quad + \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{X}_i(t_{ij}) K_{h_1}(t - t_{ij}) [\Upsilon_i(t_{ij}) - E(\Upsilon_i(t_{ij}))] \\ &\triangleq \mathbf{II}_1 + \mathbf{II}_2. \end{aligned}$$

By condition (C8), we have

$$\begin{aligned} \mathbf{II}_1 &= \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{X}_i(t_{ij}) K_{h_1}(t - t_{ij}) \left[F_{ij}(\mathbf{X}_i^T(t_{ij})(\boldsymbol{\beta}(t) - \boldsymbol{\beta}(t_{ij}))) - F_{ij}(0) \right] \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{X}_i(t_{ij}) K_{h_1}(t - t_{ij}) f_{ij}(0) \{1 + o(1)\} \mathbf{X}_i^T(t_{ij}) \\ &\quad \times (\boldsymbol{\beta}(t) - \boldsymbol{\beta}(t_{ij})). \end{aligned} \tag{A.2}$$

By Cauchy–Schwartz inequality and conditions (C6) and (C8), for all $\boldsymbol{\zeta} \in \mathcal{R}^p$ with $\boldsymbol{\zeta}^T \boldsymbol{\zeta} = 1$,

$$\begin{aligned} E(\boldsymbol{\zeta}^T \mathbf{II})^2 &= \sum_{i=1}^n E \left\| \boldsymbol{\zeta}^T \mathbf{X}_i^T \mathbf{K}_i(t; h_1) [\boldsymbol{\Upsilon}_i - E(\boldsymbol{\Upsilon}_i)] \right\|^2 \\ &\leq \sum_{i=1}^n \boldsymbol{\zeta}^T \mathbf{X}_i^T \mathbf{K}_i(t; h_1) \mathbf{K}_i(t; h_1) \mathbf{X}_i \boldsymbol{\zeta} \\ &\quad \times E \left\{ (\boldsymbol{\Upsilon}_i - E(\boldsymbol{\Upsilon}_i))^T (\boldsymbol{\Upsilon}_i - E(\boldsymbol{\Upsilon}_i)) \right\} \\ &\leq C \max \left\{ \left| \mathbf{X}_i^T(t_{ij})(\boldsymbol{\beta}(t) - \boldsymbol{\beta}(t_{ij})) \right| \right\} \\ &\quad \times \sum_{i=1}^n \boldsymbol{\zeta}^T \mathbf{X}_i^T \mathbf{K}_i(t; h_1) \mathbf{K}_i(t; h_1) \mathbf{X}_i \boldsymbol{\zeta} \\ &= o(Nh_1), \end{aligned} \tag{A.3}$$

where $\boldsymbol{\Upsilon}_i = (\Upsilon_i(t_{i1}), \dots, \Upsilon_i(t_{im_i}))^T$. According to McCullagh (1983), we have

$$\begin{aligned} &\hat{\boldsymbol{\beta}}(t_0; h_1) - \boldsymbol{\beta}(t_0) \\ &= \left\{ \frac{1}{Nh_1} \sum_{i=1}^n \mathbf{X}_i^T \mathbf{K}_i(t_0; h_1) \boldsymbol{\Lambda}_i \mathbf{X}_i \right\}^{-1} \end{aligned}$$

$$\begin{aligned} & \times \left\{ \frac{1}{Nh_1} \sum_{i=1}^n \mathbf{X}_i^T \mathbf{K}_i(t_0; h_1) \psi_\tau(\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}(t_0)) \right\} \\ & + o_p\left((Nh_1)^{-1/2}\right), \end{aligned} \tag{A.4}$$

where $\mathbf{A}_i = \text{diag}(f_{i1}(0), \dots, f_{im_i}(0))$. Then, by the law of large numbers together with (A.1)–(A.4), we have

$$\hat{\boldsymbol{\beta}}(t_0; h_1) - \boldsymbol{\beta}(t_0) = f_T^{-1}(t_0) \bar{f}^{-1}(0) \boldsymbol{\Phi}^{-1}(t_0) \hat{\mathbf{R}}(t_0; h_1) + o_p\left((Nh_1)^{-1/2}\right). \tag{A.5}$$

where $\hat{\mathbf{R}}(t_0; h_1) = \frac{1}{Nh_1} \sum_{i=1}^n \{ \mathbf{X}_i^T \mathbf{K}_i(t_0; h_1) \psi_\tau(\boldsymbol{\varepsilon}_i) + \mathbf{X}_i^T \mathbf{K}_i(t_0; h_1) \mathbf{A}_i \mathbf{X}_i [\boldsymbol{\beta}(t_{ij}) - \boldsymbol{\beta}(t_0)] \}$. It can be shown that the l th element of $\hat{\mathbf{R}}(t_0; h_1)$ can be written as

$$\hat{R}_l(t_0; h_1) = \frac{1}{Nh_1} \sum_{i=1}^n \varphi_{il}(t_0; h_1), \tag{A.6}$$

where $\varphi_{il}(t_0; h_1) = \sum_{j=1}^{m_i} \xi_{il}(t_0, t_{ij}) K_{h_1}(t_0 - t_{ij})$ and

$$\xi_{il}(t_0, t_{ij}) = X_{il}(t_{ij}) \psi_\tau(\varepsilon_i(t_{ij})) + X_{il}(t_{ij}) f_{ij}(0) \sum_{r=1}^p X_{ir}(t_{ij}) [\beta_r(t_{ij}) - \beta_r(t_0)].$$

Then (A.6) implies that $\hat{\mathbf{R}}(t_0; h_1)$ is a sum of independent vectors

$$\hat{\mathbf{R}}(t_0; h_1) = \frac{1}{Nh_1} \sum_{i=1}^n \boldsymbol{\Psi}_i(t_0; h_1),$$

where $\boldsymbol{\Psi}_i(t_0; h_1) = (\varphi_{i1}(t_0; h_1), \dots, \varphi_{ip}(t_0; h_1))^T$. Because $E(\psi_\tau\{\varepsilon_i(t_{ij})\}) = 0$ and the design points $t_{ij}, i = 1, \dots, n, j = 1, \dots, m_i$ are independent, direct calculation and the change of variables show that

$$\begin{aligned} E(\varphi_{il}(t_0; h_1)) &= \sum_{j=1}^{m_i} \int E(\xi_{il}(t_0, t_{ij}) | t_{ij} = v) K\left(\frac{t_0 - v}{h_1}\right) f_T(v) dv \\ &= \sum_{j=1}^{m_i} f_{ij}(0) h_1 \sum_{r=1}^p \int [\beta_r(t_0 - h_1 u) - \beta_r(t_0)] \\ &\quad \times \eta_{lr}(t_0 - h_1 u) f_T(t_0 - h_1 u) K(u) du \\ &= \sum_{j=1}^{m_i} f_{ij}(0) \mu_2 h_1^3 \sum_{r=1}^p [\dot{\beta}_r(t_0) \dot{\eta}_{lr}(t_0) f_T(t_0)] \end{aligned}$$

$$+ \frac{1}{2} \ddot{\beta}_r(t_0) \eta_{lr}(t_0) f_T(t_0) + \dot{\beta}_r(t_0) \eta_{lr}(t_0) \dot{f}_T(t_0) \Big] \{1 + o(1)\}.$$

Then, by (A.6) and assumptions (C1), (C3) and (C6), and taking the Taylor expansions on the right side of the foregoing equation, we have

$$E(\hat{R}_l(t_0; h_1)) = \frac{1}{Nh_1} \sum_{i=1}^n E(\varphi_{il}(t_0; h_1)) \rightarrow \mu_2 h_1^2 \bar{f}(0) \sum_{r=1}^p [\dot{\beta}_r(t_0) \dot{\eta}_{lr}(t_0) f_T(t_0) + \frac{1}{2} \ddot{\beta}_r(t_0) \eta_{lr}(t_0) f_T(t_0) + \dot{\beta}_r(t_0) \eta_{lr}(t_0) \dot{f}_T(t_0)].$$

Therefore,

$$E(\hat{\beta}(t_0; h_1) - \beta(t_0)) = f_T^{-1}(t_0) \Phi^{-1}(t_0) \mathbf{b}(t_0) = \mathbf{B}(t_0).$$

For the covariance of $\sqrt{Nh_1} \hat{\mathbf{R}}(t_0; h_1)$, because

$$\begin{aligned} & \text{Cov}[\sqrt{Nh_1} \hat{R}_l(t_0; h_1), \sqrt{Nh_1} \hat{R}_r(t_0; h_1)] \\ &= E[\sqrt{Nh_1} \hat{R}_l(t_0; h_1) \sqrt{Nh_1} \hat{R}_r(t_0; h_1)] \\ & \quad - E[\sqrt{Nh_1} \hat{R}_l(t_0; h_1)] E[\sqrt{Nh_1} \hat{R}_r(t_0; h_1)], \end{aligned} \tag{A.7}$$

and

$$\begin{aligned} & E[\sqrt{Nh_1} \hat{R}_l(t_0; h_1) \sqrt{Nh_1} \hat{R}_r(t_0; h_1)] \\ &= E\left\{ \left[\frac{1}{\sqrt{Nh_1}} \sum_{i=1}^n \varphi_{il}(t_0; h_1) \right] \left[\frac{1}{\sqrt{Nh_1}} \sum_{i=1}^n \varphi_{ir}(t_0; h_1) \right] \right\} \\ &= \frac{1}{Nh_1} \sum_{i=1}^n E[\varphi_{il}(t_0; h_1) \varphi_{ir}(t_0; h_1)] \\ & \quad + \frac{1}{Nh_1} \sum_{i_1 \neq i_2} E[\varphi_{i_1 l}(t_0; h_1) \varphi_{i_2 r}(t_0; h_1)]. \end{aligned} \tag{A.8}$$

For the first term on the right side of (A.8), we consider the further decomposition

$$\begin{aligned} \varphi_{il}(t_0; h_1) \varphi_{ir}(t_0; h_1) &= \sum_{j=1}^{m_i} \xi_{il}(t_0, t_{ij}) \xi_{ir}(t_0, t_{ij}) \left[K\left(\frac{t_0 - t_{ij}}{h_1}\right) \right]^2 \\ & \quad + \sum_{j_1 \neq j_2} \xi_{il}(t_0, t_{ij_1}) \xi_{ir}(t_0, t_{ij_2}) K\left(\frac{t_0 - t_{ij_1}}{h_1}\right) K\left(\frac{t_0 - t_{ij_2}}{h_1}\right) \end{aligned} \tag{A.9}$$

The change of variables, and the fact that $\psi_\tau \{\varepsilon_i(t_{ij})\}$ is a mean 0 and independent of $X_i(t_{ij})$, it can be shown by direct calculation that, as $n \rightarrow \infty$ and $v \rightarrow t_0$,

$$\begin{aligned} & E(\xi_{il}(t_0, t_{ij}) \xi_{ir}(t_0, t_{ij}) | t_{ij} = v) \\ &= \tau(1 - \tau) E(X_{il}(t_{ij}) X_{ir}(t_{ij}) | t_{ij} = v) \\ &+ \sum_{c=1}^p [\beta_c(v) - \beta_c(t_0)]^2 f_{ij}^2(0) E(X_{il}(t_{ij}) X_{ir}(t_{ij}) X_{ic}^2(t_{ij}) | t_{ij} = v) \\ &+ \sum_{c_1 \neq c_2} [\beta_{c_1}(v) - \beta_{c_1}(t_0)] [\beta_{c_2}(v) - \beta_{c_2}(t_0)] f_{ij}^2(0) \\ &\times E(X_{il}(t_{ij}) X_{ir}(t_{ij}) X_{ic_1}(t_{ij}) X_{ic_2}(t_{ij}) | t_{ij} = v) \\ &\rightarrow \tau(1 - \tau) \eta_{lr}(t_0). \end{aligned}$$

Then, we have

$$\begin{aligned} & E \left[\sum_{j=1}^{m_i} \xi_{il}(t_0, t_{ij}) \xi_{ir}(t_0, t_{ij}) K^2 \left(\frac{t_0 - t_{ij}}{h_1} \right) \right] \\ &= \sum_{j=1}^{m_i} \int \left[E(\xi_{il}(t_0, t_{ij}) \xi_{ir}(t_0, t_{ij}) | t_{ij} = v) K^2 \left(\frac{t_0 - v}{h_1} \right) \right] f_T(v) dv \\ &= m_i h_1 \tau(1 - \tau) \eta_{lr}(t_0) \nu_0 f_T(t_0) + o(m_i h_1). \end{aligned} \quad (\text{A.10})$$

Similarly, it can be shown by direct calculation that as $n \rightarrow \infty$, $v_1 \rightarrow t_0$, $v_2 \rightarrow t_0$

$$\begin{aligned} & E(\xi_{il}(t_0, t_{ij_1}) \xi_{ir}(t_0, t_{ij_2}) | t_{ij_1} = v_1, t_{ij_2} = v_2) \\ &= \rho_\varepsilon(v_1, v_2) E(X_{il}(t_{ij_1}) X_{ir}(t_{ij_2}) | t_{ij_1} = v_1, t_{ij_2} = v_2) \\ &+ \sum_{c=1}^p [\beta_c(v_1) - \beta_c(t_0)] [\beta_c(v_2) - \beta_c(t_0)] f_{ij}^2(0) \\ &\times E(X_{il}(t_{ij_1}) X_{ir}(t_{ij_2}) X_{ic}(t_{ij_1}) X_{ic}(t_{ij_2}) | t_{ij_1} = v_1, t_{ij_2} = v_2) \\ &+ \sum_{c_1 \neq c_2} [\beta_{c_1}(v_1) - \beta_{c_1}(t_0)] [\beta_{c_2}(v_2) - \beta_{c_2}(t_0)] f_{ij}^2(0) \\ &\times E(X_{il}(t_{ij_1}) X_{ir}(t_{ij_2}) X_{ic_1}(t_{ij_1}) X_{ic_2}(t_{ij_2}) | t_{ij_1} = v_1, t_{ij_2} = v_2) \\ &\rightarrow \rho_\varepsilon(t_0) \eta_{lr}(t_0). \end{aligned}$$

Therefore, the expectation of the second term on the right side of (A.9) is

$$E \left[\sum_{j_1 \neq j_2} \xi_{il}(t_0, t_{ij_1}) \xi_{ir}(t_0, t_{ij_2}) K \left(\frac{t_0 - t_{ij_1}}{h_1} \right) K \left(\frac{t_0 - t_{ij_2}}{h_1} \right) \right]$$

$$\begin{aligned}
 &= \sum_{j_1 \neq j_2} \left\{ \iint E \left(\xi_{il} (t_0, t_{i j_1}) \xi_{ir} (t_0, t_{i j_2}) \mid t_{i j_1} = v_1, t_{i j_2} = v_2 \right) \right. \\
 &\quad \times K \left(\frac{t_0 - v_1}{h_1} \right) K \left(\frac{t_0 - v_2}{h_1} \right) f_T (v_1) f_T (v_2) dv_1 dv_2 \left. \right\} \\
 &= m_i (m_i - 1) h_1^2 \rho_\varepsilon (t_0) \eta_{lr} (t_0) f_T^2 (t_0) + o \left(m_i (m_i - 1) h_1^2 \right). \quad (\text{A.11})
 \end{aligned}$$

Combining (A.9)–(A.11), it follows immediately that when n is sufficiently large

$$\begin{aligned}
 &\frac{1}{Nh_1} \sum_{i=1}^n E [\varphi_{il} (t_0; h_1) \varphi_{ir} (t_0; h_1)] \\
 &= \tau (1 - \tau) \eta_{lr} (t_0) \nu_0 f_T (t_0) + N^{-1} h_1 \left(\sum_{i=1}^n m_i^2 - N \right) \rho_\varepsilon (t_0) \eta_{lr} (t_0) f_T^2 (t_0) \\
 &\quad + o \left(N^{-1} h_1 \left(\sum_{i=1}^n m_i^2 - N \right) \right) + o (1) \\
 &\rightarrow \tau (1 - \tau) \eta_{lr} (t_0) \nu_0 f_T (t_0) + \kappa h_0 \rho_\varepsilon (t_0) \eta_{lr} (t_0) f_T^2 (t_0) \quad (\text{A.12})
 \end{aligned}$$

because $h_1 = N^{-1/5} h_0$ and $\lim_{n \rightarrow \infty} N^{-6/5} \sum_{i=1}^n m_i^2 = \kappa$, and it is easy to see that as $n \rightarrow \infty$

$$N^{-1} h_1 \left(\sum_{i=1}^n m_i^2 - N \right) = N^{-6/5} \left(\sum_{i=1}^n m_i^2 - N \right) h_0 \rightarrow \kappa h_0.$$

Similar to the proof of (A.13) in Wu et al. (1998), as $n \rightarrow \infty$, we have

$$\begin{aligned}
 &\left| \frac{1}{Nh_1} \sum_{i_1 \neq i_2} E [\varphi_{i_1 l} (t_0; h_1) \varphi_{i_2 r} (t_0; h_1)] \right. \\
 &\quad \left. - E \left[\frac{1}{\sqrt{Nh_1}} \sum_{i=1}^n \varphi_{il} (t_0; h_1) \right] E \left[\frac{1}{\sqrt{Nh_1}} \sum_{i=1}^n \varphi_{ir} (t_0; h_1) \right] \right| \\
 &\rightarrow 0. \quad (\text{A.13})
 \end{aligned}$$

Based on (A.7), (A.8), (A.12) and (A.13), we have

$$\begin{aligned}
 &\text{Cov} \left[\sqrt{Nh_1} \hat{R}_l (t_0; h_1), \sqrt{Nh_1} \hat{R}_r (t_0; h_1) \right] \\
 &= \tau (1 - \tau) \eta_{lr} (t_0) \nu_0 f_T (t_0) + \kappa h_0 \rho_\varepsilon (t_0) \eta_{lr} (t_0) f_T^2 (t_0) + o (1).
 \end{aligned}$$

The multivariate central limit theorem and the Slutsky’s theorem imply that

$$\sqrt{N h_1} \left(\hat{\boldsymbol{\beta}}(t_0; h_1) - \boldsymbol{\beta}(t_0) - \mathbf{B}(t_0) \right) \xrightarrow{d} N \left(\mathbf{0}, v^2(t_0) f_T^{-2}(t_0) \bar{f}^{-2}(0) \boldsymbol{\Phi}(t_0)^{-1} \right).$$

Proof of Theorem 1 According to McCullagh (1983), we have

$$\begin{aligned} \hat{\boldsymbol{y}} - \boldsymbol{y}_0 &= - \left(\frac{1}{n} \sum_{i=1}^n \mathbf{V}_i^T \mathbf{V}_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n \mathbf{V}_i^T \tilde{\boldsymbol{e}}_i + o_p \left(n^{-1/2} \right), \\ \tilde{\boldsymbol{e}}_i &= \mathbf{L}_i^{-1} \psi_\tau(\hat{\boldsymbol{\epsilon}}_i) = \mathbf{L}_i^{-1} \left[\tau \mathbf{1}_{m_i} - I(\boldsymbol{\epsilon}_i < 0) + I(\boldsymbol{\epsilon}_i < 0) - I(\hat{\boldsymbol{\epsilon}}_i < 0) \right] \\ &= \boldsymbol{e}_i - \mathbf{L}_i^{-1} \boldsymbol{\Delta}_i, \end{aligned}$$

where $\mathbf{1}_{m_i}$ is an $m_i \times 1$ vector with all elements being 1 and $\boldsymbol{\Delta}_i = (\Delta_{i1}, \dots, \Delta_{im_i})^T$ with $\Delta_{ij} = \left[I \left\{ \boldsymbol{\epsilon}_i(t_{ij}) < \mathbf{X}_i^T(t_{ij}) \left(\hat{\boldsymbol{\beta}}(t_{ij}; h_1) - \boldsymbol{\beta}(t_{ij}) \right) \right\} - I \left\{ \boldsymbol{\epsilon}_i(t_{ij}) < 0 \right\} \right]$. Because \boldsymbol{e}_i are independent random variables with $E(\boldsymbol{e}_i) = \mathbf{0}$ and $\text{Cov}(\boldsymbol{e}_i) = \mathbf{D}_i$. In addition, $\Delta_{ij} = O_p(1/\sqrt{N h_1} + h_1^2)$ by Lemma 2. The multivariate central limit theorem and the Slutsky’s theorem imply that

$$\sqrt{n} (\hat{\boldsymbol{y}} - \boldsymbol{y}_0) \xrightarrow{d} N \left(\mathbf{0}, \boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Sigma}^{-1} \right).$$

□

Proof of Theorem 2 Following the same line of argument of Theorem 1 of Fan and Yao (1998), we have

$$\hat{d}^2(t) - d^2(t) = \frac{1}{N h_2 f_T(t)} \sum_{i=1}^n \sum_{j=1}^{m_i} K_{h_2}(t - t_{ij}) \left\{ \hat{\boldsymbol{e}}_{ij}^2 - d^2(t) - \dot{d}^2(t)(t_{ij} - t) \right\}.$$

Note that

$$\begin{aligned} \hat{\boldsymbol{e}}_{ij}^2 &= \left(\psi_\tau(\hat{\boldsymbol{\epsilon}}_{ij}) - \sum_{k=1}^{j-1} \hat{l}_{ijk} \hat{\boldsymbol{e}}_{ik} \right)^2 \\ &= \left[\psi_\tau(\boldsymbol{\epsilon}_{ij}) - \left(\Delta_{ij} + \sum_{k=1}^{j-1} \hat{l}_{ijk} \hat{\boldsymbol{e}}_{ik} \right) \right]^2 \\ &= d_{ij}^2 \zeta_{ij}^2 + 2d_{ij} \varsigma_{ij} \left[\sum_{k=1}^{j-1} \left\{ l_{ijk} (\boldsymbol{e}_{ik} - \hat{\boldsymbol{e}}_{ik}) + (l_{ijk} - \hat{l}_{ijk}) \hat{\boldsymbol{e}}_{ik} \right\} - \Delta_{ij} \right] \\ &\quad + \left[\sum_{k=1}^{j-1} \left\{ l_{ijk} (\boldsymbol{e}_{ik} - \hat{\boldsymbol{e}}_{ik}) + (l_{ijk} - \hat{l}_{ijk}) \hat{\boldsymbol{e}}_{ik} \right\} - \Delta_{ij} \right]^2, \end{aligned}$$

where Δ_{ij} is given in the proof of Theorem 1. It follows that

$$\hat{d}^2(t) - d^2(t) = I_1 + I_2 + I_3 + I_4 \{1 + o_p(1)\},$$

where

$$I_1 = \frac{1}{Nh_2 f_T(t)} \sum_{i=1}^n \sum_{j=1}^{m_i} K_{h_2}(t - t_{ij}) \{d^2(t_{ij}) - d^2(t) - \dot{d}^2(t)(t_{ij} - t)\},$$

$$I_2 = \frac{1}{Nh_2 f_T(t)} \sum_{i=1}^n \sum_{j=1}^{m_i} K_{h_2}(t - t_{ij}) \{d^2(t_{ij}) (\varsigma_{ij}^2 - 1)\},$$

$$I_3 = 2 \frac{1}{Nh_2 f_T(t)} \sum_{i=1}^n \sum_{j=1}^{m_i} K_{h_2}(t - t_{ij}) d_{ij} \varsigma_{ij} \times \left[\sum_{k=1}^{j-1} \{l_{ijk}(e_{ik} - \hat{e}_{ik}) + (l_{ijk} - \hat{l}_{ijk}) \hat{e}_{ik}\} - \Delta_{ij} \right],$$

$$I_4 = \frac{1}{Nh_2 f_T(t)} \sum_{i=1}^n \sum_{j=1}^{m_i} K_{h_2}(t - t_{ij}) \left[\sum_{k=1}^{j-1} \{l_{ijk}(e_{ik} - \hat{e}_{ik}) + (l_{ijk} - \hat{l}_{ijk}) \hat{e}_{ik}\} - \Delta_{ij} \right]^2.$$

It is easy to see that Theorem 2 follows directly from statements (a)–(d) below

(a) $I_1 = \frac{1}{2} \mu_2 h_2^2 \ddot{d}^2(t) + o_p(h_2^2),$

(b) $\sqrt{Nh_2} I_2 \rightarrow^d N \left(0, \frac{v_0}{f_T(t)} \lim_{n \rightarrow \infty} \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{m_i} E \left[(\varsigma_{ij}^2 - 1)^2 | t_{ij} = t \right] d^4(t) \right),$

(c) $I_3 = o_p\left(\frac{1}{\sqrt{Nh_2}}\right),$

(d) $I_4 = o_p\left(\frac{1}{\sqrt{Nh_2}}\right).$

It is easy to see that (a) follows from a Taylor expansion. I_2 is asymptotically normal with mean 0 and variance

$$\text{Var}(I_2) = \frac{v_0}{N^2 h_2 f_T(t)} \sum_{i=1}^n \sum_{j=1}^{m_i} E \left[(\varsigma_{ij}^2 - 1)^2 | t_{ij} = t \right] d^4(t).$$

It follows from the definition of I_3 that

$$I_3 = 2 \frac{1}{Nh_2 f_T(t)} \sum_{i=1}^n \sum_{j=1}^{m_i} K_{h_2}(t - t_{ij}) d_{ij} \varsigma_{ij} \sum_{k=1}^{j-1} \{l_{ijk}(e_{ik} - \hat{e}_{ik}) + (l_{ijk} - \hat{l}_{ijk}) \hat{e}_{ik}\} - 2 \frac{1}{Nh_2 f_T(t)} \sum_{i=1}^n \sum_{j=1}^{m_i} K_{h_2}(t - t_{ij}) d_{ij} \varsigma_{ij} \Delta_{ij}$$

$$= I_{31} + I_{32}.$$

By Lemma 2 and condition (C1), together with $E(\varsigma_{ij}|t_{ij}) = 0$, $\text{Var}(\varsigma_{ij}|t_{ij}) = 1$, we have

$$I_{31} = O_p\left(\frac{1}{\sqrt{Nh_1}} + h_1^2\right) O_p\left(\frac{1}{\sqrt{Nh_2}}\right) = o_p\left(\frac{1}{\sqrt{Nh_2}}\right),$$

and

$$I_{32} = O_p\left(\frac{1}{\sqrt{Nh_1}} + h_1^2\right) O_p\left(\frac{1}{\sqrt{Nh_2}}\right) = o_p\left(\frac{1}{\sqrt{Nh_2}}\right).$$

Then $I_3 = o_p(1/\sqrt{Nh_2})$. By the same arguments as proving I_3 , we have $I_4 = o_p(1/\sqrt{Nh_2})$. Under the conditions $Nh_2 \rightarrow \infty$ as $n \rightarrow \infty$ and $\limsup_{n \rightarrow \infty} Nh_2^5 < \infty$, then the proof of Theorem 2 is completed. \square

Proof of Theorem 3 Similar to the proof of Lemma 2, we have

$$\begin{aligned} \bar{\beta}(t_0; h_1) - \beta(t_0) &= \left\{ \frac{1}{Nh_1} \sum_{i=1}^n X_i^T \mathbf{K}_i(t_0; h_1) \hat{\Sigma}_i^{-1} \tilde{\Lambda}_i X_i \right\}^{-1} \\ &\quad \times \left\{ \frac{1}{Nh_1} \sum_{i=1}^n X_i^T \mathbf{K}_i(t_0; h_1) \hat{\Sigma}_i^{-1} \psi_{\tau h}(Y_i - X_i \beta(t_0)) \right\} \\ &\quad + o_p\left((Nh_1)^{-1/2}\right), \end{aligned}$$

By the law of large numbers, we have

$$\frac{1}{Nh_1} \sum_{i=1}^n X_i^T \mathbf{K}_i(t_0; h_1) \hat{\Sigma}_i^{-1} \tilde{\Lambda}_i X_i \xrightarrow{p} \Omega_1,$$

where $\tilde{\Lambda}_i = \text{diag} \left\{ \frac{1}{h} K_1\left(\frac{Y_{i1} - X_i^T(t_{i1})\beta(t_0)}{h}\right), \dots, \frac{1}{h} K_1\left(\frac{Y_{im_i} - X_i^T(t_{im_i})\beta(t_0)}{h}\right) \right\}$. Using the conditions (C6), (C9) and (C10), similar to Lemma 3 (k) of Horowitz (1998), we obtain

$$\begin{aligned} &\frac{1}{Nh_1} \sum_{i=1}^n X_i^T \mathbf{K}_i(t_0; h_1) \hat{\Sigma}_i^{-1} \psi_{\tau h}(Y_i - X_i \beta(t_0)) \\ &= \frac{1}{Nh_1} \sum_{i=1}^n X_i^T \mathbf{K}_i(t_0; h_1) \hat{\Sigma}_i^{-1} \psi_{\tau}(Y_i - X_i \beta(t_0)) + o_p\left((Nh_1)^{-1/2}\right) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{Nh_1} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{j'=1}^{m_i} \mathbf{X}_i(t_{ij}) f_{ij'}(0) K\left(\frac{t_0 - t_{ij}}{h_1}\right) \hat{\sigma}^{jj'} \mathbf{X}_i^T(t_{ij'}) [\boldsymbol{\beta}(t_{ij'}) - \boldsymbol{\beta}(t_0)] \\
 &+ \frac{1}{Nh_1} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{j'=1}^{m_i} \mathbf{X}_i(t_{ij}) K\left(\frac{t_0 - t_{ij}}{h_1}\right) \hat{\sigma}^{jj'} \psi_\tau(\varepsilon_i(t_{ij'})) + o_p\left((Nh_1)^{-1/2}\right).
 \end{aligned}$$

Similar to the proof of Lemma 2, we have

$$E\left(\bar{\boldsymbol{\beta}}(t_0; h_1) - \boldsymbol{\beta}(t_0)\right) = f_T^{-1}(t_0) \boldsymbol{\Phi}^{-1}(t_0) \mathbf{b}(t_0) = \mathbf{B}(t_0),$$

and

$$\text{Var}\left[\frac{1}{\sqrt{Nh_1}} \sum_{i=1}^n \mathbf{X}_i^T \mathbf{K}_i(t_0; h_1) \hat{\boldsymbol{\Sigma}}_i^{-1} \psi_{\tau h}(Y_i - \mathbf{X}_i \boldsymbol{\beta}(t_0))\right] = \boldsymbol{\Omega}_2 + \boldsymbol{\Omega}_3 + o(1).$$

The multivariate central limit theorem and the Slutsky's theorem imply that

$$\sqrt{Nh_1} \left(\bar{\boldsymbol{\beta}}(t_0; h_1) - \boldsymbol{\beta}(t_0) - \mathbf{B}(t_0)\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Omega}_1^{-1} (\boldsymbol{\Omega}_2 + \boldsymbol{\Omega}_3) \boldsymbol{\Omega}_1^{-1}\right)$$

Therefore, we complete the proof of Theorem 3. □

Proof of Theorem 4 Let $\zeta(\boldsymbol{\beta}(t_0)) = \frac{1}{\sqrt{Nh_1}} \sum_{i=1}^n \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0))$. By Theorem 3, we have $E[\zeta(\boldsymbol{\beta}(t_0))] = o(1)$ and $\text{Cov}[\zeta(\boldsymbol{\beta}(t_0))] = \boldsymbol{\Omega}_2 + \boldsymbol{\Omega}_3 + o(1)$. By Lemma 1, we know that $\zeta(\boldsymbol{\beta}(t_0))$ satisfies the conditions of the Cramer–Wold theorem (cf. Serfling 1980, theorem in sec. 1.5.2) and the Lindeberg condition (cf. Serfling 1980, theorem in sec. 1.9.2). Hence,

$$\frac{1}{\sqrt{Nh_1}} \sum_{i=1}^n \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}_2 + \boldsymbol{\Omega}_3), \tag{A.14}$$

and

$$\frac{1}{Nh_1} \sum_{i=1}^n \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) \mathbf{Z}_{ih}^T(\boldsymbol{\beta}(t_0)) \xrightarrow{p} \boldsymbol{\Omega}_2 + \boldsymbol{\Omega}_3. \tag{A.15}$$

From (A.14), (A.15) and Lemma 1, and using the same arguments that are used in the proof of (2.14) in Owen (1990), we can prove that

$$\lambda = O_p\left((Nh_1)^{-1/2}\right), \tag{A.16}$$

where λ is defined in (13). Applying the Taylor expansion to (14) and invoking (A.14)–(A.16) and Lemma 1, we obtain

$$l(\boldsymbol{\beta}(t_0)) = 2 \sum_{i=1}^n \left\{ \lambda^T \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) - \left[\lambda^T \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) \right]^2 / 2 \right\} + o_p(1). \quad (\text{A.17})$$

By (13), it follows that

$$\begin{aligned} 0 &= \sum_{i=1}^n \frac{\mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0))}{1 + \lambda^T \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0))} \\ &= \sum_{i=1}^n \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) - \sum_{i=1}^n \lambda^T \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) \\ &\quad + \sum_{i=1}^n \frac{\mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) \left[\lambda^T \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) \right]^2}{\left[1 + \lambda^T \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) \right]}. \end{aligned}$$

The application of Lemma 1 and (A.14)–(A.16) again yields

$$\sum_{i=1}^n \left[\lambda^T \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) \right]^2 = \sum_{i=1}^n \lambda^T \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) + o_p(1), \quad (\text{A.18})$$

and

$$\lambda = \left[\sum_{i=1}^n \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) \mathbf{Z}_{ih}^T(\boldsymbol{\beta}(t_0)) \right]^{-1} \sum_{i=1}^n \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) + o_p\left((Nh_1)^{-1/2}\right). \quad (\text{A.19})$$

Substituting (A.18) and (A.19) into (A.17), we obtain

$$\begin{aligned} l(\boldsymbol{\beta}(t_0)) &= \left[\frac{1}{\sqrt{Nh_1}} \sum_{i=1}^n \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) \right]^T \left[\frac{1}{Nh_1} \sum_{i=1}^n \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) \mathbf{Z}_{ih}^T(\boldsymbol{\beta}(t_0)) \right]^{-1} \\ &\quad \times \left[\frac{1}{\sqrt{Nh_1}} \sum_{i=1}^n \mathbf{Z}_{ih}(\boldsymbol{\beta}(t_0)) \right] + o_p(1). \end{aligned} \quad (\text{A.20})$$

Based on (A.14), (A.15) and (A.20), we can prove Theorem 4. \square

Proof of Corollary 1 Let $Z_{ih}^{(2)}(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t)) = \partial Z_{ih}(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t)) / \partial \tilde{\boldsymbol{\beta}}^{(2)}(t)$ and $\tilde{\boldsymbol{\lambda}} = \boldsymbol{\lambda}(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t))$, then $\tilde{\boldsymbol{\lambda}}$ and $\tilde{\boldsymbol{\beta}}^{(2)}(t)$ satisfy

$$Q_1(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t), \tilde{\boldsymbol{\lambda}}) = \sum_{i=1}^n \frac{Z_{ih}(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t))}{1 + \tilde{\boldsymbol{\lambda}}^T Z_{ih}(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t))} = 0,$$

and

$$Q_2(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t), \tilde{\boldsymbol{\lambda}}) = \sum_{i=1}^n \frac{\tilde{\boldsymbol{\lambda}}^T Z_{ih}^{(2)}(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t))}{1 + \tilde{\boldsymbol{\lambda}}^T Z_{ih}(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t))} = 0.$$

Expanding $Q_1(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t), \tilde{\boldsymbol{\lambda}})$ and $Q_2(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t), \tilde{\boldsymbol{\lambda}})$ at $(\mathbf{b}(t_0), \boldsymbol{\beta}^{(2)}(t_0), 0)$, we have

$$\tilde{\boldsymbol{\lambda}} = (\mathbf{I} - \mathbf{P}) \boldsymbol{\Sigma}_n^{-1} \tilde{\mathbf{Z}} + o_p((Nh_1)^{-1/2}),$$

and

$$\tilde{\boldsymbol{\beta}}^{(2)}(t) - \boldsymbol{\beta}^{(2)}(t_0) = -(\tilde{\mathbf{Z}}^{(2)T} \boldsymbol{\Sigma}_n^{-1} \tilde{\mathbf{Z}}^{(2)})^{-1} \tilde{\mathbf{Z}}^{(2)T} \boldsymbol{\Sigma}_n^{-1} \tilde{\mathbf{Z}} + o_p((Nh_1)^{-1/2}),$$

where $\tilde{\mathbf{Z}}^{(2)} = \sum_{i=1}^n Z_{ih}^{(2)}(\mathbf{b}(t_0), \boldsymbol{\beta}^{(2)}(t_0))$, $\mathbf{P} = \boldsymbol{\Sigma}_n^{-1} \tilde{\mathbf{Z}}^{(2)} (\tilde{\mathbf{Z}}^{(2)T} \boldsymbol{\Sigma}_n^{-1} \tilde{\mathbf{Z}}^{(2)})^{-1} \tilde{\mathbf{Z}}^{(2)T}$, $\tilde{\mathbf{Z}} = \sum_{i=1}^n Z_{ih}(\mathbf{b}(t_0), \boldsymbol{\beta}^{(2)}(t_0))$ and $\boldsymbol{\Sigma}_n = \sum_{i=1}^n Z_{ih}(\mathbf{b}(t_0), \boldsymbol{\beta}^{(2)}(t_0)) Z_{ih}^T(\mathbf{b}(t_0), \boldsymbol{\beta}^{(2)}(t_0))$. Because

$$\begin{aligned} l(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t)) &= 2 \sum_{i=1}^n \log \left\{ 1 + \tilde{\boldsymbol{\lambda}}^T Z_{ih}(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t)) \right\} \\ &= 2 \sum_{i=1}^n \tilde{\boldsymbol{\lambda}}^T Z_{ih}(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t)) \\ &\quad - \sum_{i=1}^n \left\{ \tilde{\boldsymbol{\lambda}}^T Z_{ih}(\mathbf{b}(t_0), \tilde{\boldsymbol{\beta}}^{(2)}(t)) \right\}^2 + o_p(1) \\ &= \tilde{\mathbf{Z}}^T \boldsymbol{\Sigma}_n^{-1/2} (\mathbf{I} - \boldsymbol{\Sigma}_n^{1/2} \mathbf{P} \boldsymbol{\Sigma}_n^{-1/2}) \boldsymbol{\Sigma}_n^{-1/2} \tilde{\mathbf{Z}} + o_p(1). \end{aligned}$$

Similar to the proof of Theorem 4, we have $\boldsymbol{\Sigma}_n^{-1/2} \tilde{\mathbf{Z}} \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$ and $\boldsymbol{\Sigma}_n^{1/2} \mathbf{P} \boldsymbol{\Sigma}_n^{-1/2}$ is symmetric and idempotent, with trace equal to $p - p_1$. Because $Z_{ih} l(\hat{\boldsymbol{\beta}}^{(1)}(t), \hat{\boldsymbol{\beta}}^{(2)}(t)) = 0$. Hence the empirical likelihood ratio statistic $\bar{l}(\mathbf{b}(t_0))$ converges to $\chi_{p_1}^2$. \square

References

- Chen J, Sitter RR, Wu C (2002) Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika* 89:230–237
- Fan J, Yao Q (1998) Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* 85:645–660
- Fu L, Wang YG (2016) Efficient parameter estimation via Gaussian copulas for quantile regression with longitudinal data. *J Multivar Anal* 143:492–502
- Hastie T, Tibshirani R (1993) Varying-coefficient models. *J R Stat Soc Ser B* 55:757–796
- Horowitz JL (1998) Bootstrap methods for median regression models. *Econometrica* 66:1327–1351
- Huang JZ, Wu CO, Zhou L (2002) Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* 89:111–128
- Huang JZ, Wu CO, Zhou L (2004) Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Stat Sin* 14:763–788
- Leng C, Zhang W, Pan J (2010) Semiparametric mean–covariance regression analysis for longitudinal data. *J Am Stat Assoc* 105:181–193
- Leng C, Zhang W (2014) Smoothing combined estimating equations in quantile regression for longitudinal data. *Stat Comput* 24:123–136
- Li Y (2011) Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. *Biometrika* 98:355–370
- Liu S, Li G (2015) Varying-coefficient mean–covariance regression analysis for longitudinal data. *J Stat Plann Inference* 160:89–106
- McCullagh P (1983) Quasi-likelihood functions. *Ann Stat* 11:59–67
- Noh HS, Park BU (2010) Sparse varying coefficient models for longitudinal data. *Stat Sin* 20:1183–1202
- Owen AB (1990) Empirical likelihood ratio confidence regions. *Ann Stat* 18:90–120
- Owen AB (2001) Empirical likelihood. Chapman and Hall, New York
- Pourahmadi M (1999) Joint mean–covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika* 86:677–690
- Qin J, Lawless J (1994) Empirical likelihood and general estimating equations. *Ann Stat* 22:300–325
- Serfling R (1980) Approximation theorems of mathematical statistics. Wiley, New York
- Tang C, Leng C (2011) Empirical likelihood and quantile regression in longitudinal data analysis. *Biometrika* 98:1001–1006
- Tang Y, Wang Y, Li J, Qian W (2015) Improving estimation efficiency in quantile regression with longitudinal data. *J Stat Plan Inference* 165:38–55
- Wang H, Xia Y (2009) Shrinkage estimation of the varying coefficient model. *J Am Stat Assoc* 104:747–757
- Wang HJ, Zhu Z (2011) Empirical likelihood for quantile regression models with longitudinal data. *J Stat Plan Inference* 141:1603–1615
- Wang K, Lin L (2015) Simultaneous structure estimation and variable selection in partial linear varying coefficient models for longitudinal data. *J Stat Comput Simul* 85:1459–1473
- Wang L, Li H, Huang J (2008) Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *J Am Stat Assoc* 103:1556–1569
- Wu CO, Chiang CT, Hoover DR (1998) Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J Am Stat Assoc* 93:1388–1402
- Xue L, Zhu L (2007) Empirical likelihood for a varying coefficient model with longitudinal data. *J Am Stat Assoc* 102:642–654
- Ye H, Pan J (2006) Modelling of covariance structures in generalised estimating equations for longitudinal data. *Biometrika* 93:927–941
- Yuan X, Lin N, Dong X, Liu T (2017) Weighted quantile regression for longitudinal data using empirical likelihood. *Sci China Math* 60:147–164
- Zhang DW, Lin XH, Raz J, Sowers MF (1998) Semiparametric stochastic mixed models for longitudinal data. *J Am Stat Assoc* 93:710–719
- Zhang W, Leng C (2012) A moving average Cholesky factor model in covariance modeling for longitudinal data. *Biometrika* 99:141–150
- Zheng X, Fung WK, Zhu Z (2013) Robust estimation in joint mean–covariance regression model for longitudinal data. *Ann Inst Stat Math* 65:617–638