

Quantile estimations via modified Cholesky decomposition for longitudinal single-index models

Jing Lv¹ · Chaohui Guo²

Received: 4 January 2018 / Revised: 15 April 2018 / Published online: 19 June 2018
© The Institute of Statistical Mathematics, Tokyo 2018

Abstract Quantile regression is a powerful complement to the usual mean regression and becomes increasingly popular due to its desirable properties. In longitudinal studies, it is necessary to consider the intra-subject correlation among repeated measures over time to improve the estimation efficiency. In this paper, we focus on longitudinal single-index models. Firstly, we apply the modified Cholesky decomposition to parameterize the intra-subject covariance matrix and develop a regression approach to estimate the parameters of the covariance matrix. Secondly, we propose efficient quantile estimating equations for the index coefficients and the link function based on the estimated covariance matrix. Since the proposed estimating equations include a discrete indicator function, we propose smoothed estimating equations for fast and accurate computation of the index coefficients, as well as their asymptotic covariances. Thirdly, we establish the asymptotic properties of the proposed estimators. Finally, simulation studies and a real data analysis have illustrated the efficiency of the proposed approach.

Keywords B-spline · Longitudinal data · Modified Cholesky decomposition · Quantile regression · Single-index models

1 Introduction

Single-index models are becoming increasingly popular due to its flexibility and interpretability. They also can effectively overcome the problem of “curse of dimen-

✉ 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

sionality” through projecting multivariate covariates onto one-dimensional index variate $\mathbf{x}^T \beta$. Longitudinal data frequently occur in the biomedical, epidemiological, social, and economical fields. For longitudinal data, subjects are often measured repeatedly over a given time period. Thus, observations from the same subject are correlated and those from different subjects are often independent. The technique of generalized estimating equation (GEE) proposed by [Liang and Zeger \(1986\)](#) is widely used for longitudinal data. The GEE method produces consistent estimators for the mean parameters through specifying a working structure. However, the main drawback of GEE is that it may lead to a great loss of efficiency when the working covariance structure is misspecified. Thus, it is an interesting topic to model the covariance structure. Recently, the modified Cholesky decomposition has been demonstrated to be effective for modeling the covariance structure. It not only permits more general forms of the correlation structures, but also leads automatically to positive definite covariance matrix. [Ye and Pan \(2006\)](#) utilized the modified Cholesky decomposition to decompose the inverse of covariance matrix and proposed a joint mean–covariance model for longitudinal data. [Leng et al. \(2010\)](#) constructed a semiparametric mean–covariance model through relaxing the parametric assumption, which is more flexible. [Zhang and Leng \(2012\)](#) used a new Cholesky factor to deal with the within-subject structure by decomposing the covariance matrix rather than its inverse. Other related references include [Mao et al. \(2011\)](#), [Zheng et al. \(2014\)](#), [Liu and Zhang \(2013\)](#), [Yao and Li \(2013\)](#), and [Liu and Li \(2015\)](#).

In recent years, some statistical inference methods have been proposed for longitudinal single-index models. [Xu and Zhu \(2012\)](#) proposed a kernel GEE method. [Lai et al. \(2012\)](#) presented the bias-corrected GEE estimation and variable selection procedure for the index coefficients. [Zhao et al. \(2017\)](#) constructed a robust estimation procedure based on quantile regression and a specific correlation structure (e.g., compound symmetry (CS) or the first-order autoregressive (AR(1))). All these articles used some specific correlation structures when taking into account the within-subject correlation. Thus, these methods may result in a loss of efficiency when the true correlation structure is not correctly specified. Recently, [Lin et al. \(2016\)](#) developed a new efficient estimation procedure for single-index models by combining the modified Cholesky decomposition and the local linear smoothing method. Motivated by [Leng et al. \(2010\)](#), [Guo et al. \(2016\)](#) proposed a two-step estimation procedure for single-index models based on the modified Cholesky decomposition and the GEE method. The above two papers are built on mean regression, which is very sensitive to outliers and heavy tail errors. In contrast with mean regression, quantile regression not only has the ability of describing the entire conditional distribution of the response variable, but also can accommodate non-normal errors. Thus, it has emerged as a powerful complement to the mean regression. Although the modified Cholesky decomposition has been well studied for the mean regression models, it is lack of analyzing longitudinal single-index quantile models. In this paper, we use the modified Cholesky decomposition to parameterize the within-subject covariance structure and construct more efficient estimation procedure for the index coefficients and the link function. Compared with existing research results, the new method has several advantages. Firstly, the proposed method does not need to specify the working correlation structure to improve the estimate efficiency. So our approach not only can

take the within-subject correlation into consideration, but also permits more general forms of the covariance structures, which indicates that it is more flexible than most of the existing methods. Secondly, since the proposed estimating equations include the discrete quantile score function, we construct new smoothed estimating equations for fast and accurate computation of the parameter estimates. Thirdly, the estimators of the index coefficients and the link function are demonstrated to be asymptotically efficient.

The rest of this article is organized as follows: In Sect. 2, within the framework of independent working structure, we propose the quantile score estimating functions for the index coefficients based on “remove-one-component” method, and the corresponding theoretical properties are also given in this section. In Sect. 3, we apply the modified Cholesky decomposition to decompose the within-subject covariance matrix as moving average coefficients and innovation variances, which can be estimated through constructing two estimating equations. In Sect. 4, more efficient quantile estimating functions are derived based on the estimated covariance matrix. In Sect. 5, extensive simulation studies are carried out to evaluate the finite sample performance of the proposed method. In Sect. 6, we illustrate the proposed method through a real data analysis. Finally, all the conditions and the proofs of the main results are provided in “Appendix.”

2 Estimation procedure under the independent structure

A marginal quantile single-index model with longitudinal data has the following structure

$$Y_{ij} = g_{0\tau} \left(X_{ij}^T \beta_{0\tau} \right) + \varepsilon_{ij,\tau}, i = 1, \dots, n, j = 1, \dots, m_i,$$

where $Y_{ij} = Y(t_{ij}) \in \mathbb{R}$ is the j th measurement of the i th subject, $X_{ij} = X(t_{ij}) \in \mathbb{R}^p$, $g_{0\tau}(\cdot)$ is an unknown differentiable univariate link function, $\varepsilon_{ij,\tau}$ is the random error term with an unspecified density function $f_{ij}(\cdot)$ and $P(\varepsilon_{ij,\tau} < 0) = \tau$ for any i, j and $\tau \in (0, 1)$, and $\beta_{0\tau}$ is an unknown parameter vector which belongs to the parameter space

$$\Theta = \left\{ \beta = (\beta_1, \dots, \beta_p)^T : \|\beta\| = 1, \text{ and the } r\text{th component is positive} \right\},$$

where $\|\cdot\|$ is the Euclidean norm. Without loss of generality, we assume that the true vector β has a positive component β_r (otherwise, consider $-\beta$). For simplicity, we omit τ from $\varepsilon_{ij,\tau}$, $\beta_{0\tau}$ and $g_{0\tau}(\cdot)$ in the rest of this article, but we should remember that they are τ -specific. Let $Y_i = (Y_{i1}, \dots, Y_{im_i})^T$, $X_i = (X_{i1}, \dots, X_{im_i})^T$, and $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{im_i})^T$. In this paper, we assume the number of measurements m_i is uniformly bounded for each i , which means that n and N ($N = \sum_{i=1}^n m_i$) have the same order.

2.1 Estimations of $g_0(\cdot)$ and its first derivative $g'_0(\cdot)$

B-spline is commonly used to approximate the nonparametric function for its efficient in function approximation and numerical computation, which can refer to [Ma and He \(2016\)](#), [Guo et al. \(2016\)](#), and [Zhao et al. \(2017\)](#). In this paper, we adopt B-spline basis functions to approximate the unknown link function $g_0(\cdot)$. We assume $X_{ij}^T \beta$ is confined in a compact set $[a, b]$. Consider a knot sequence with N_n interior knots, denoted by $\xi_1 = \dots = a = \xi_q < \xi_{q+1} < \dots < \xi_{q+N_n} < b = \xi_{q+N_n+1} = \dots = \xi_{2q+N_n}$. We set the B-spline basis functions as $B_q(u) = (B_{1,q}(u), \dots, B_{J_n,q}(u))^T$ with the order q ($q \geq 2$) and $J_n = N_n + q$. We approximate the link function $g_0(u)$ by $g_0(u) \approx B_q(u)^T \theta$, where $\theta = (\theta_1, \dots, \theta_{J_n})^T$ is the spline coefficient vector. For a given β , we can obtain the estimator $\hat{\theta}(\beta)$ of θ by minimizing the following objective function

$$L_n(\beta; \theta) = \sum_{i=1}^n \sum_{j=1}^{m_i} \rho_\tau \left(Y_{ij} - B_q \left(X_{ij}^T \beta \right)^T \theta \right), \tag{1}$$

where $\rho_\tau(u) = u \{ \tau - I(u < 0) \}$ is the quantile loss function. Then, the link function $g_0(\cdot)$ is estimated by the spline functions $\hat{g}(u; \beta) = B_q(u)^T \hat{\theta}(\beta)$. Following [Ma and Song \(2015\)](#), the estimator of $g'_0(\cdot)$ is defined by

$$\hat{g}'(u; \beta) = \sum_{s=1}^{J_n} B'_{s,q}(u) \hat{\theta}_s(\beta) = \sum_{s=2}^{J_n} B_{s,q-1}(u) \hat{\omega}_s(\beta),$$

where $\hat{\omega}_s(\beta) = (q - 1) \left\{ \hat{\theta}_s(\beta) - \hat{\theta}_{s-1}(\beta) \right\} / (\xi_{s+q-1} - \xi_s)$ for $2 \leq s \leq J_n$. Thus, we have $\hat{g}'(u; \beta) = B_{q-1}(u)^T D_1 \hat{\theta}(\beta)$, where $B_{q-1}(u) = (B_{s,q-1}(u) : 2 \leq s \leq J_n)^T$ is the $(q - 1)$ th-order B-spline basis and

$$D_1 = (q - 1) = \begin{bmatrix} \frac{-1}{\xi_{q+1}-\xi_2} & \frac{1}{\xi_{q+1}-\xi_2} & 0 & \dots & 0 \\ 0 & \frac{-1}{\xi_{q+2}-\xi_3} & \frac{1}{\xi_{q+2}-\xi_3} & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{-1}{\xi_{N_n+2q-1}-\xi_{N_n+q}} & \frac{1}{\xi_{N_n+2q-1}-\xi_{N_n+q}} \end{bmatrix}_{(J_n-1) \times J_n}$$

2.2 The profile-type estimating equations for β

The parameter space Θ means that β is on the boundary of a unit ball. Therefore, $g_0(X_{ij}^T \beta)$ is non-differential at the point β . However, we must use the derivative of $g_0(X_{ij}^T \beta)$ with respect to β when constructing the profile-type estimating equations. To solve the above problem, we employ “remove-one-component” method ([Cui et al. 2011](#)) to transform the boundary of a unit ball in \mathbb{R}^p to the interior in \mathbb{R}^{p-1} . Specifically, let $\beta^{(r)} = (\beta_1, \dots, \beta_{r-1}, \beta_{r+1}, \dots, \beta_p)^T$ be a $p - 1$ dimensional vector by removing the r th component β_r in β . Then, β can be rewritten as

$\beta = \beta(\beta^{(r)}) = \left(\beta_1, \dots, \beta_{r-1}, \left(1 - \|\beta^{(r)}\|^2\right)^{1/2}, \beta_{r+1}, \dots, \beta_p \right)^T$ and $\beta^{(r)}$ belongs to the parameter space

$$\Theta^{(r)} = \left\{ \beta^{(r)} = (\beta_1, \dots, \beta_{r-1}, \beta_{r+1}, \dots, \beta_p)^T : \|\beta^{(r)}\| < 1 \right\}.$$

So β is infinitely differentiable with respect to $\beta^{(r)}$ and the Jacobian matrix is

$$J_{\beta^{(r)}} = \frac{\partial \beta}{\partial \beta^{(r)}} = (\varsigma_1, \dots, \varsigma_p)^T,$$

where $\varsigma_r = -\left(1 - \|\beta^{(r)}\|^2\right)^{-1/2} \beta^{(r)}$ and ς_s ($1 \leq s \leq p, s \neq r$) is a $(p - 1) \times 1$ unit vector with s th component 1.

Motivated by the idea of GEE (Liang and Zeger 1986), together with the estimators of $g_0(\cdot)$ and $g'_0(\cdot)$, we construct the profile-type estimating equations for $p - 1$ dimensional vector $\beta^{(r)}$

$$R\left(\beta^{(r)}\right) = \sum_{i=1}^n J_{\beta^{(r)}}^T \hat{X}_i^T \hat{G}'\left(X_i \beta; \beta\right) \mathbf{A}_i \psi_{\tau}\left(Y_i - \hat{g}\left(X_i \beta; \beta\right)\right) = \mathbf{0}, \tag{2}$$

where $\hat{G}'\left(X_i \beta; \beta\right) = \text{diag}\left\{\hat{g}'\left(X_{i1}^T \beta; \beta\right), \dots, \hat{g}'\left(X_{im_i}^T \beta; \beta\right)\right\}$, $\psi_{\tau}(u) = I(u < 0) - \tau$ is the quantile score function, $\psi_{\tau}\left(\mathbf{u}_i\right) = \left(\psi_{\tau}\left(u_{i1}\right), \dots, \psi_{\tau}\left(u_{im_i}\right)\right)^T$, $\hat{g}\left(X_i \beta; \beta\right) = \left(\hat{g}\left(X_{i1}^T \beta; \beta\right), \dots, \hat{g}\left(X_{im_i}^T \beta; \beta\right)\right)^T$, $\hat{X}_i = \left(\hat{X}_{i1}, \dots, \hat{X}_{im_i}\right)^T$, $\hat{X}_{ij} = X_{ij} - \hat{E}\left(X_{ij} \mid X_{ij}^T \beta\right)$ and $\hat{E}\left(X_{ij} \mid X_{ij}^T \beta\right)$ is the spline estimator of $E\left(X_{ij} \mid X_{ij}^T \beta_0\right)$. In estimating Eq. (2), the term $\mathbf{A}_i = \text{diag}\left\{f_{i1}\left(0\right), \dots, f_{im_i}\left(0\right)\right\}$ describes the dispersions in ε_{ij} . In some cases when f_{ij} is difficult to estimate, \mathbf{A}_i can be simply treated as an identity matrix with a slight loss of efficiency (Jung 1996). We define the solution of (2) as $\hat{\beta}^{(r)}$ and then use the fact $\beta_r = \sqrt{1 - \|\beta^{(r)}\|^2}$ to obtain $\hat{\beta}$. The asymptotic property of $\hat{\beta}$ is given in Lemma 3 of ‘‘Appendix.’’

2.3 Computational algorithm

Solving estimating Eq. (2) faces some interesting challenges due to the discontinuous indicator function. To overcome the calculation difficulty, we approximate $\psi_{\tau}(\cdot)$ by a smooth function $\psi_{h\tau}(\cdot)$ based on the idea of Wang and Zhu (2011). Define $G(x) = \int_{u < x} K(u)du$ and $G_h(x) = G(x/h)$, where $K(\cdot)$ is a kernel function and h is a positive bandwidth parameter. Then, we approximate $\psi_{\tau}\left(Y_{ij} - \hat{g}\left(X_{ij}^T \beta; \beta\right)\right)$ by $\psi_{h\tau}\left(Y_{ij} - \hat{g}\left(X_{ij}^T \beta; \beta\right)\right) = 1 - G_h\left(Y_{ij} - \hat{g}\left(X_{ij}^T \beta; \beta\right)\right) - \tau$. Therefore, based on the approximation, estimating Eq. (2) can be replaced by the following smoothed estimating equations

$$\tilde{R}(\beta^{(r)}) = \sum_{i=1}^n \mathbf{J}_{\beta^{(r)}}^T \hat{\mathbf{X}}_i^T \hat{\mathbf{G}}'(\mathbf{X}_i \beta; \beta) \mathbf{A}_i \psi_{h\tau}(\mathbf{Y}_i - \hat{g}(\mathbf{X}_i \beta; \beta)) = \mathbf{0}. \tag{3}$$

We define the solution of (3) as $\tilde{\beta}^{(r)}$. Since (3) is nonlinear equation about $\beta^{(r)}$, the Fisher–Scoring iterative algorithm can be used to solve it. Specifically, the iterative algorithm is described as follows:

Step 0 Start with an initial value $\beta^{(0)}$, which is obtained by Ma and He (2016).

Step 1 Set $k = 0$. Use the current estimate $\beta^{(k)}$ and minimize $L_n(\beta^{(k)}; \theta)$ with respect to θ to obtain the estimator $\hat{\theta}^{(k)}$. Then, we can obtain $\hat{g}^{(k)}(u; \beta^{(k)}) = \mathbf{B}_q(u)^T \hat{\theta}^{(k)}$ and $\hat{g}'^{(k)}(u; \beta^{(k)}) = \mathbf{B}_{q-1}(u)^T \mathbf{D}_1 \hat{\theta}^{(k)}$.

Step 2 Utilize the estimators $\hat{g}^{(k)}$ and $\hat{g}'^{(k)}$ obtained by Step 1; $(\beta^{(r)})^{(k)}$ can be updated by

$$(\beta^{(r)})^{(k+1)} = (\beta^{(r)})^{(k)} - \tilde{D}(\beta^{(r)})^{-1} \tilde{R}(\beta^{(r)})|_{\beta^{(r)}=(\beta^{(r)})^{(k)}},$$

where

$$\tilde{D}(\beta^{(r)}) \triangleq \frac{\partial \tilde{R}(\beta^{(r)})}{\partial \beta^{(r)}} = \sum_{i=1}^n \mathbf{J}_{\beta^{(r)}}^T \hat{\mathbf{X}}_i^T \hat{\mathbf{G}}'(\mathbf{X}_i \beta; \beta) \mathbf{A}_i \tilde{\mathbf{A}}_i(\beta) \hat{\mathbf{G}}'(\mathbf{X}_i \beta; \beta) \hat{\mathbf{X}}_i \mathbf{J}_{\beta^{(r)}}$$

and

$$\begin{aligned} \tilde{\mathbf{A}}_i(\beta) &= \text{diag} \left\{ h^{-1} K \left[\left(Y_{i1} - \hat{g}(\mathbf{X}_{i1}^T \beta; \beta) \right) / h \right], \dots, h^{-1} K \left[\left(Y_{im_i} - \hat{g}(\mathbf{X}_{im_i}^T \beta; \beta) \right) / h \right] \right\}. \end{aligned}$$

Step 3 Set $k = k + 1$ and repeat Steps 1 and 2 until convergence.

Step 4 With the final estimators $\tilde{\beta}^{(r)}$ and $\hat{\theta}$ obtained from Step 3, we can get the final estimator of $g_0(u)$ by $\hat{g}(u; \tilde{\beta}) = \mathbf{B}_q(u)^T \hat{\theta}(\tilde{\beta})$, where $\tilde{\beta}$ is obtained by the fact $\beta_r = \sqrt{1 - \|\beta^{(r)}\|^2}$.

Remark 1 If the sum of $\left| (\beta^{(r)})^{(k+1)} - (\beta^{(r)})^{(k)} \right|$ is smaller than a cutoff value (such as 10^{-6}), we stop the iteration. Our simulation studies indicate that the Fisher–Scoring algorithm can find the numerical solution of (3) quickly.

2.4 Asymptotic properties

Let $g_0(u)$ and β_0 be the true values of $g(u)$ and β , respectively. In the following theorems, we need to restrict $\beta \in \Theta_n$, where $\Theta_n = \{\beta \in \Theta : \|\beta - \beta_0\| \leq Cn^{-1/2}\}$ for some positive constant C . Since we anticipate that the estimators of β_0 are root- n consistent, we should look for the solutions of (3) which involve β distant from β_0 by order $n^{-1/2}$. Define

$$\Phi = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{J}_{\beta_0^{(r)}}^T \tilde{\mathbf{X}}_i^T \mathbf{G}'(X_i \beta_0) \mathbf{A}_i \mathbf{A}_i \mathbf{G}'(X_i \beta_0) \tilde{\mathbf{X}}_i \mathbf{J}_{\beta_0^{(r)}}$$

and

$$\Psi = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{J}_{\beta_0^{(r)}}^T \tilde{\mathbf{X}}_i^T \mathbf{G}'(X_i \beta_0) \mathbf{A}_i \Sigma_{\tau i} \mathbf{A}_i \mathbf{G}'(X_i \beta_0) \tilde{\mathbf{X}}_i \mathbf{J}_{\beta_0^{(r)}}$$

where $\Sigma_{\tau i} = Cov(\psi_{\tau}(\epsilon_i))$, $\mathbf{G}'(X_i \beta_0) = \text{diag} \left\{ g'_0(X_{i1}^T \beta_0), \dots, g'_{im_i}(X_{im_i}^T \beta_0) \right\}$, $\tilde{\mathbf{X}}_i = (\tilde{X}_{i1}, \dots, \tilde{X}_{im_i})^T$ and $\tilde{X}_{ij} = X_{ij} - E(X_{ij} | X_{ij}^T \beta_0)$.

Theorem 1 Under conditions (C1)–(C7) in “Appendix,” and the number of knots satisfies $n^{1/(2d+2)} \ll N_n \ll n^{1/4}$, where d is given in the condition (C2) of “Appendix,” we have

$$\sqrt{n} (\tilde{\beta} - \beta_0) \xrightarrow{d} N \left(\mathbf{0}, \mathbf{J}_{\beta_0^{(r)}} \Phi \Phi^{-1} \Psi \Phi^{-1} \mathbf{J}_{\beta_0^{(r)}}^T \right)$$

as $n \rightarrow \infty$, where \xrightarrow{d} denotes convergence in distribution.

Theorem 2 Let θ^0 be the best approximation coefficient of $g_0(u)$ in the B-spline space. When β is a known constant β_0 or estimated to the order $O_p(n^{-1/2})$, under conditions (C1)–(C7) in “Appendix,” and the number of knots satisfies $n^{1/(2d+2)} \ll N_n \ll n^{1/4}$, then (i) $|\hat{g}(u; \beta) - g_0(u)| = O_p(\sqrt{N_n/n} + N_n^{-d})$ uniformly in $u \in [a, b]$; and (ii) under $n^{1/(2d+1)} \ll N_n \ll n^{1/4}$,

$$\sigma_n^{-1}(u) (\hat{g}(u; \beta) - \check{g}(u; \beta)) \xrightarrow{d} N(0, 1),$$

where $\sigma_n^2(u) = \mathbf{B}_q^T(u) \mathbf{V}^{-1}(\beta_0) \sum_{i=1}^n \mathbf{B}_q^T(X_i \beta_0) \mathbf{A}_i \Sigma_{\tau i} \mathbf{A}_i \mathbf{B}_q(X_i \beta_0) \mathbf{V}^{-1}(\beta_0)$
 $\mathbf{B}_q(u)$, $\check{g}(u; \beta) = \mathbf{B}_q^T(u) \theta^0(\beta)$, $\mathbf{B}_q(X_i \beta_0) = (\mathbf{B}_q(X_{i1}^T \beta_0), \dots, \mathbf{B}_q(X_{im_i}^T \beta_0))^T$
 and $\mathbf{V}(\beta_0) = \sum_{i=1}^n \mathbf{B}_q^T(X_i \beta_0) \mathbf{A}_i \mathbf{A}_i \mathbf{B}_q(X_i \beta_0)$.

3 Modeling the within-subject covariance matrix via the modified Cholesky decomposition

To incorporate the correlation within subjects, following the idea of GEE (Liang and Zeger 1986), we can use the estimating equations that take the form

$$\sum_{i=1}^n \left(\begin{matrix} \mathbf{B}_q^T(X_i \beta) \\ \mathbf{J}_{\beta^{(r)}}^T \hat{\mathbf{X}}_i^T \mathbf{G}'(X_i \beta) \end{matrix} \right) \mathbf{A}_i \Sigma_{\tau i}^{-1} \psi_{\tau}(Y_i - \mathbf{B}_q(X_i \beta) \theta) = \mathbf{0}. \tag{4}$$

Unfortunately, estimating Eq. (4) includes an unknown covariance matrix $\Sigma_{\tau i}$. So our primary task is to estimate it. To guarantee the positive definiteness of $\Sigma_{\tau i}$, we firstly apply the modified Cholesky decomposition to decompose $\Sigma_{\tau i}$ as

$$\Sigma_{\tau i} = L_{\tau i} D_{\tau i} L_{\tau i}^T,$$

where $L_{\tau i}$ is a lower triangular matrix with 1's on its diagonal and $D_{\tau i}$ is an $m_i \times m_i$ diagonal matrix. Let $e_{\tau i} = L_{\tau i}^{-1} \psi_{\tau}(\boldsymbol{\varepsilon}_i) = (e_{\tau, i1}, \dots, e_{\tau, im_i})^T$, so we have $Cov(e_{\tau i}) = L_{\tau i}^{-1} \Sigma_{\tau i} (L_{\tau i}^{-1})^T = D_{\tau i} \triangleq \text{diag}(d_{\tau, i1}^2, \dots, d_{\tau, im_i}^2)$, where $d_{\tau, ij}^2$ is called as innovation variance. Furthermore, we assume that the below diagonal entries of $L_{\tau i}$ are $l_{\tau, ijk} (k < j = 2, \dots, m_i)$, and then $e_{\tau i} = L_{\tau i}^{-1} \psi_{\tau}(\boldsymbol{\varepsilon}_i)$ can be rewritten as

$$\psi_{\tau}(\boldsymbol{\varepsilon}_{ij}) = \sum_{k=1}^{j-1} l_{\tau, ijk} e_{\tau, ik} + e_{\tau, ij},$$

where $l_{\tau, ijk}$ are the so-called moving average coefficients. In this paper, we define the notation $\sum_{k=1}^0$ means zero when $j = 1$. The main advantage of the modified Cholesky decomposition is that $l_{\tau, ijk}$ and $d_{\tau, ij}^2$ are unconstrained. In order to estimate the moving average coefficients $l_{\tau, ijk}$ and the innovation variances $d_{\tau, ij}^2$, we construct two generalized linear models as follows:

$$l_{\tau, ijk} = \mathbf{w}_{ijk}^T \boldsymbol{\gamma}_{\tau}, \log(d_{\tau, ij}^2) = \mathbf{z}_{ij}^T \boldsymbol{\lambda}_{\tau}, \tag{5}$$

where $\boldsymbol{\gamma}_{\tau} = (\gamma_{\tau 1}, \dots, \gamma_{\tau p_1})^T$ and $\boldsymbol{\lambda}_{\tau} = (\lambda_{\tau 1}, \dots, \lambda_{\tau p_2})^T$. Based on the idea of Zhang and Leng (2012), the covariates \mathbf{z}_{ij} are those used in regression analysis, and \mathbf{w}_{ijk} is usually adopted as a polynomial of time difference $t_{ij} - t_{ik}$. By adopting the idea of the GEE approach, we construct two estimating equations for $\boldsymbol{\gamma}_{\tau}$ and $\boldsymbol{\lambda}_{\tau}$ by

$$U_1(\boldsymbol{\gamma}_{\tau}) = \sum_{i=1}^n \left(\frac{\partial \mathbf{e}_{\tau i}^T}{\partial \boldsymbol{\gamma}_{\tau}} \right) D_{\tau i}^{-1} \mathbf{e}_{\tau i} = \mathbf{0}, \tag{6}$$

$$U_2(\boldsymbol{\lambda}_{\tau}) = \sum_{i=1}^n \mathbf{z}_i^T D_{\tau i} \mathbf{W}_{\tau i}^{-1} (\mathbf{e}_{\tau i}^2 - d_{\tau i}^2) = \mathbf{0}, \tag{7}$$

where $\partial \mathbf{e}_{\tau i}^T / \partial \boldsymbol{\gamma}_{\tau}$ is a $p_1 \times m_i$ matrix with the first column zero and the j th $j \geq 2$ column $\partial e_{\tau, ij} / \partial \boldsymbol{\gamma}_{\tau} = -\sum_{k=1}^{j-1} [\mathbf{w}_{ijk} e_{\tau, ik} + l_{\tau, ijk} \partial e_{\tau, ik} / \partial \boldsymbol{\gamma}_{\tau}]$, $\mathbf{z}_i = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{im_i})^T$ and $d_{\tau i}^2 = (d_{\tau, i1}^2, \dots, d_{\tau, im_i}^2)^T$. Here, $\mathbf{W}_{\tau i}$ is the covariance matrix of $\mathbf{e}_{\tau i}^2$, namely $\mathbf{W}_{\tau i} = Cov(\mathbf{e}_{\tau i}^2)$. The true $\mathbf{W}_{\tau i}$ is unknown and can be approximated by a sandwich “working” covariance structure $\mathbf{W}_{\tau i} = \mathbf{A}_{\tau i}^{1/2} \mathbf{R}_{\tau i}(\rho) \mathbf{A}_{\tau i}^{1/2}$ (Liu and Zhang 2013), where $\mathbf{A}_{\tau i} = 2 \text{diag}(d_{\tau, i1}^4, \dots, d_{\tau, im_i}^4)$ and $\mathbf{R}_{\tau i}(\rho)$ mimics the correlation between

$e_{\tau,ij}^2$ and $e_{\tau,ik}^2$ ($j \neq k$) with a new parameter ρ . The common structures of $\mathbf{R}_{\tau i}(\rho)$ contain the compound symmetry and the first-order autoregressive. We assume that $\hat{\gamma}_\tau$ and $\hat{\lambda}_\tau$ are the solutions of estimating Eqs. (6) and (7). Liu and Zhang (2013) have pointed out that $\hat{\gamma}_\tau$ and $\hat{\lambda}_\tau$ are not sensitive to the parameter ρ . So, we take $\rho = 0$ in our simulation studies and real data analysis.

Let $\gamma_{\tau 0}$ and $\lambda_{\tau 0}$ be the true values of γ_τ and λ_τ , respectively. Meanwhile, we define the covariance matrix of the function $(\mathbf{U}_1(\gamma_{\tau 0})^T, \mathbf{U}_2(\lambda_{\tau 0})^T)^T / \sqrt{n}$ by $\mathbf{V}_{\tau n} = (\mathbf{v}_{\tau n}^{jl})_{j,l=1,2}$, where $\mathbf{v}_{\tau n}^{jl} = n^{-1} Cov(\mathbf{U}_j, \mathbf{U}_l)$ for $j, l = 1, 2$. Furthermore, we assume that the covariance matrix $\mathbf{V}_{\tau n}$ is positive definite at the true value $(\gamma_{\tau 0}^T, \lambda_{\tau 0}^T)^T$ and

$$\mathbf{V}_{\tau n} = \begin{pmatrix} \mathbf{v}_{\tau n}^{11} & \mathbf{v}_{\tau n}^{12} \\ \mathbf{v}_{\tau n}^{21} & \mathbf{v}_{\tau n}^{22} \end{pmatrix} \xrightarrow{p} \mathbf{V}_\tau = \begin{pmatrix} \mathbf{v}_\tau^{11} & \mathbf{v}_\tau^{12} \\ \mathbf{v}_\tau^{21} & \mathbf{v}_\tau^{22} \end{pmatrix},$$

where \xrightarrow{p} denotes convergence in probability. Then, the proposed estimators $(\hat{\gamma}_\tau^T, \hat{\lambda}_\tau^T)^T$ are \sqrt{n} -consistent and have the following asymptotic distribution

$$\sqrt{n} \begin{pmatrix} \hat{\gamma}_\tau - \gamma_{\tau 0} \\ \hat{\lambda}_\tau - \lambda_{\tau 0} \end{pmatrix} \xrightarrow{d} N \left\{ \mathbf{0}, \begin{pmatrix} \mathbf{v}_\tau^{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{v}_\tau^{22} \end{pmatrix}^{-1} \mathbf{V}_\tau \begin{pmatrix} \mathbf{v}_\tau^{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{v}_\tau^{22} \end{pmatrix}^{-1} \right\}.$$

The proof is omitted since it is similar to the proof of Theorem 1 of Lv et al. (2017). Now, we show that the estimated covariance matrix $\hat{\Sigma}_{\tau i}$ is consistent. For a matrix \mathbf{A} , $\|\mathbf{A}\| = [tr(\mathbf{A}\mathbf{A}^T)]^{1/2}$ denotes its Frobenius norm.

Theorem 3 *Let $\Sigma_{\tau i}$ and $\hat{\Sigma}_{\tau i}$ be the true and estimated covariance matrix within the i th cluster, respectively. Suppose that the regularity conditions in ‘‘Appendix’’ hold. If the covariance matrix has the model structure (5), we have $\|\hat{\Sigma}_{\tau i} - \Sigma_{\tau i}\| = O_p(n^{-1/2})$.*

4 Efficient estimating equations for the index coefficients β and the link function $g(\cdot)$

Based on the discussions in Sect. 3, the covariance matrix $\Sigma_{\tau i}$ can be estimated by $\hat{\Sigma}_{\tau i} = \hat{\mathbf{L}}_{\tau i} \hat{\mathbf{D}}_{\tau i} \hat{\mathbf{L}}_{\tau i}^T$, where $\hat{\mathbf{D}}_{\tau i} = diag(\hat{d}_{\tau,i1}^2, \dots, \hat{d}_{\tau,im_i}^2)$ with $\hat{d}_{\tau,ij}^2 = \exp(z_{ij}^T \hat{\lambda}_\tau)$ and the (j, k) element of $\hat{\mathbf{L}}_{\tau i}$ is $\hat{l}_{\tau,ijk} = \mathbf{w}_{ijk}^T \hat{\gamma}_\tau$ for $k < j = 2, \dots, m_i$. Firstly, for a given β , we construct efficient smoothed estimating equations of θ by

$$\sum_{i=1}^n \mathbf{B}_q^T(\mathbf{X}_i \beta) \mathbf{A}_i \hat{\Sigma}_{\tau i}^{-1} \psi_{h_\tau}(\mathbf{Y}_i - \mathbf{B}_q(\mathbf{X}_i \beta) \theta) = \mathbf{0}. \tag{8}$$

Then, the efficient smoothed estimating equations of $\beta^{(r)}$ are constructed similarly by

$$\sum_{i=1}^n \mathbf{J}_{\beta^{(r)}}^T \tilde{\mathbf{X}}_i^T \bar{\mathbf{G}}' (X_i \beta; \beta) \mathbf{A}_i \tilde{\Sigma}_{\tau_i}^{-1} \psi_{h\tau} (Y_i - \mathbf{B}_q (X_i \beta) \bar{\boldsymbol{\theta}} (\beta)) = \mathbf{0}, \tag{9}$$

where $\bar{\boldsymbol{\theta}} (\beta)$ is the solution of (8), $\bar{\mathbf{G}}' (X_i \beta; \beta) = \text{diag} \left\{ \bar{g}' (X_{i1}^T \beta; \beta), \dots, \bar{g}' (X_{im_i}^T \beta; \beta) \right\}$ and $\bar{g}' (u; \beta) = \sum_{s=1}^{J_n} B'_{s,q} (u) \bar{\theta}_s (\beta)$. Let $\bar{\beta}^{(r)}$ be the resulting estimator of (9). Therefore, the efficient estimator of $g_0 (u)$ can be achieved by $\bar{g} (u; \bar{\beta}) = \mathbf{B}_q (u)^T \bar{\boldsymbol{\theta}} (\bar{\beta})$.

Remark 2 There are two main differences between the proposed estimating Eqs. (8) and (9) with Zhao et al. (2017)'s estimating Eq. (6). On the one hand, we apply a different method to smooth the discontinuous estimating equations. On the other hand, Zhao et al. (2017) applied a working correlation matrix \mathbf{C}_i which need to be specified to improve the estimation efficiency. Therefore, Zhao et al. (2017)'s approach cannot permit more general forms of the correlation structures and results in a loss of efficiency when the \mathbf{C}_i is far from the true correlation structure.

Theorem 4 Under conditions (C1)–(C7) in “Appendix,” and the number of knots satisfies $n^{1/(2d+2)} \ll N_n \ll n^{1/4}$, we have

$$\sqrt{n} (\bar{\beta} - \beta_0) \xrightarrow{d} N \left(\mathbf{0}, \mathbf{J}_{\beta_0^{(r)}} \boldsymbol{\Gamma}^{-1} \mathbf{J}_{\beta_0^{(r)}}^T \right),$$

as $n \rightarrow \infty$, where

$$\boldsymbol{\Gamma} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{J}_{\beta_0^{(r)}}^T \tilde{\mathbf{X}}_i^T \mathbf{G}' (X_i \beta_0) \mathbf{A}_i \Sigma_{\tau_i}^{-1} \mathbf{A}_i \mathbf{G}' (X_i \beta_0) \tilde{\mathbf{X}}_i \mathbf{J}_{\beta_0^{(r)}}$$

and the other symbols are the same as that in Theorem 1.

Let $\boldsymbol{\Upsilon}_i = \boldsymbol{\Phi}^{-1} \mathbf{J}_{\beta_0^{(r)}}^T \tilde{\mathbf{X}}_i^T \mathbf{G}' (X_i \beta_0) \mathbf{A}_i \Sigma_{\tau_i}^{1/2} - \boldsymbol{\Gamma}^{-1} \mathbf{J}_{\beta_0^{(r)}}^T \tilde{\mathbf{X}}_i^T \mathbf{G}' (X_i \beta_0) \mathbf{A}_i \Sigma_{\tau_i}^{-1/2}$. Then,

$$\begin{aligned} \boldsymbol{\Upsilon}_i \boldsymbol{\Upsilon}_i^T &= \boldsymbol{\Phi}^{-1} \mathbf{J}_{\beta_0^{(r)}}^T \tilde{\mathbf{X}}_i^T \mathbf{G}' (X_i \beta_0) \mathbf{A}_i \Sigma_{\tau_i} \mathbf{A}_i \mathbf{G}' (X_i \beta_0) \tilde{\mathbf{X}}_i \mathbf{J}_{\beta_0^{(r)}} \boldsymbol{\Phi}^{-1} \\ &\quad - \boldsymbol{\Phi}^{-1} \mathbf{J}_{\beta_0^{(r)}}^T \tilde{\mathbf{X}}_i^T \mathbf{G}' (X_i \beta_0) \mathbf{A}_i \mathbf{A}_i \mathbf{G}' (X_i \beta_0) \tilde{\mathbf{X}}_i \mathbf{J}_{\beta_0^{(r)}} \boldsymbol{\Gamma}^{-1} \\ &\quad - \boldsymbol{\Gamma}^{-1} \mathbf{J}_{\beta_0^{(r)}}^T \tilde{\mathbf{X}}_i^T \mathbf{G}' (X_i \beta_0) \mathbf{A}_i \mathbf{A}_i \mathbf{G}' (X_i \beta_0) \tilde{\mathbf{X}}_i \mathbf{J}_{\beta_0^{(r)}} \boldsymbol{\Phi}^{-1} \\ &\quad + \boldsymbol{\Gamma}^{-1} \mathbf{J}_{\beta_0^{(r)}}^T \tilde{\mathbf{X}}_i^T \mathbf{G}' (X_i \beta_0) \mathbf{A}_i \Sigma_{\tau_i}^{-1} \mathbf{A}_i \mathbf{G}' (X_i \beta_0) \tilde{\mathbf{X}}_i \mathbf{J}_{\beta_0^{(r)}} \boldsymbol{\Gamma}^{-1}. \end{aligned}$$

Since $\Upsilon_i \Upsilon_i^T$ is nonnegative definite, we have

$$0 \leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Upsilon_i \Upsilon_i^T = \Phi^{-1} \Psi \Phi^{-1} - \Phi^{-1} \Phi \Gamma^{-1} - \Gamma^{-1} \Phi \Phi^{-1} + \Gamma^{-1} \Gamma \Gamma^{-1} = \Phi^{-1} \Psi \Phi^{-1} - \Gamma^{-1}.$$

Thus, $\Gamma^{-1} \leq \Phi^{-1} \Psi \Phi^{-1}$. This implies that $\bar{\beta}$ has smaller asymptotic covariance matrix than that of $\hat{\beta}$. So the proposed estimator $\bar{\beta}$ is asymptotically more efficient than $\hat{\beta}$.

Motivated by (9), a sandwich formula for estimating the covariance of $\bar{\beta}$ is

$$Cov(\bar{\beta}) \approx \mathbf{J}_{\bar{\beta}(r)} \bar{\Gamma}_n^{-1} \bar{\Omega}_n \bar{\Gamma}_n^{-1} \mathbf{J}_{\bar{\beta}(r)}^T, \tag{10}$$

where

$$\begin{aligned} \bar{\Gamma}_n &= \sum_{i=1}^n \mathbf{J}_{\bar{\beta}(r)}^T \hat{\mathbf{X}}_i^T \bar{\mathbf{G}}'(X_i \beta; \beta) \mathbf{A}_i \hat{\Sigma}_{\tau_i}^{-1} \bar{\mathbf{A}}_i(\beta) \bar{\mathbf{G}}'(X_i \beta; \beta) \hat{\mathbf{X}}_i \mathbf{J}_{\bar{\beta}(r)} \Big|_{\beta(r)=\bar{\beta}(r)}, \\ \bar{\Omega}_n &= \sum_{i=1}^n \mathbf{J}_{\bar{\beta}(r)}^T \hat{\mathbf{X}}_i^T \bar{\mathbf{G}}'(X_i \beta; \beta) \mathbf{A}_i \hat{\Sigma}_{\tau_i}^{-1} \bar{S}_{\tau_i}(\beta) \bar{S}_{\tau_i}^T(\beta) \hat{\Sigma}_{\tau_i}^{-1} \mathbf{A}_i \bar{\mathbf{G}}'(\hat{X}_i \beta; \beta) \hat{\mathbf{X}}_i \mathbf{J}_{\bar{\beta}(r)} \Big|_{\beta(r)=\bar{\beta}(r)}, \\ \bar{S}_{\tau_i}(\beta) &= \left\{ \psi_{h\tau} \left(Y_{i1} - \bar{g}(X_{i1}^T \beta; \beta) \right), \dots, \psi_{h\tau} \left(Y_{im_i} - \bar{g}(X_{im_i}^T \beta; \beta) \right) \right\}^T, \end{aligned}$$

and

$$\begin{aligned} \bar{\mathbf{A}}_i(\beta) &= \text{diag} \left\{ h^{-1} K \left(\left(Y_{i1} - \bar{g}(X_{i1}^T \beta; \beta) \right) / h \right), \dots, h^{-1} K \left(\left(Y_{im_i} - \bar{g}(X_{im_i}^T \beta; \beta) \right) / h \right) \right\}. \end{aligned}$$

Theorem 5 When β is a known constant β_0 or estimated to the order $O_p(n^{-1/2})$, under conditions (C1)–(C7) in ‘‘Appendix,’’ and the number of knots satisfies $n^{1/(2d+2)} \ll N_n \ll n^{1/4}$, then (i) $|\bar{g}(u; \beta) - g_0(u)| = O_p(\sqrt{N_n/n} + N_n^{-d})$ uniformly in $u \in [a, b]$; and (ii) under $n^{1/(2d+1)} \ll N_n \ll n^{1/4}$,

$$\sigma_n^{*-1}(u) (\bar{g}(u; \beta) - \check{g}(u; \beta)) \xrightarrow{d} N(0, 1),$$

where $\sigma_n^{*2}(u) = \mathbf{B}_q^T(u) \mathbf{V}^{*-1}(\beta_0) \mathbf{B}_q(u)$, $\mathbf{V}^*(\beta_0) = \sum_{i=1}^n \mathbf{B}_q^T(X_i \beta_0) \mathbf{A}_i \Sigma_{\tau_i}^{-1} \mathbf{A}_i \mathbf{B}_q(X_i \beta_0)$.

Remark 3 We can adopt a similar iterative algorithm given in Sect. 2.3 to find the solutions of estimating Eqs. (8) and (9). Here, we omit it for saving space. Furthermore, we also can prove that $\bar{g}(u; \beta)$ is asymptotically more efficient than $\hat{g}(u; \beta)$ by using the above similar method.

5 Simulation studies

In this section, we conduct simulation studies to compare our approach with some existing methods. The major aim is to show that the proposed method not only can deal with complex correlation structures, but also produces more efficient estimates for the index coefficients and the link function. Specifically, we compare the proposed estimators $\tilde{\beta}$ and \tilde{g} (denoted as $\hat{\beta}_{pr}$ and \hat{g}_{pr}) with the other three types of estimators: (i) the estimator proposed by [Ma and He \(2016\)](#) without using “remove-one-component” method (denoted as $\hat{\beta}_{ma}$ and \hat{g}_{ma}); (ii) the estimators $\tilde{\beta}$ and \tilde{g} (denoted as $\hat{\beta}_{in}$ and \hat{g}_{in}) without considering the within-subject correlation, which are given in Sect. 2.3; (iii) the estimators proposed by [Zhao et al. \(2017\)](#) with the AR(1) working correlation structure (denoted as $\hat{\beta}_{ar}$ and \hat{g}_{ar}) and the compound symmetry structure (denoted as $\hat{\beta}_{cs}$ and \hat{g}_{cs}), which involve a tuning parameter h_1 . We set $h_1 = n^{-1/2}$ in our simulations and real data analysis according to their suggestions. Similar to [Wang and Zhu \(2011\)](#), we smooth the quantile score function by the following second-order ($\nu = 2$) Bartlett kernel

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

In order to achieve good numerical performances, we need to select several parameters appropriately. Firstly, we fix the spline order to be $q = 4$, namely we use cubic B-splines to approximate the nonparametric link function in our numerical examples. Meanwhile, we use equally spaced knots with the number of interior knots $N_n = \lceil n^{1/(2q+1)} \rceil$ which satisfies theoretical requirement. The similar strategy had been adopted by [Ma and Song \(2015\)](#). Secondly, [Wang and Zhu \(2011\)](#) had proved that the smoothed approach is robust to the bandwidth h . Thus, we fix $h = n^{-0.3}$ which satisfies the theoretical requirement $nh^{2\nu} \rightarrow 0$ with $\nu = 2$.

Example 1 Similar to [Ma and He \(2016\)](#), we generate the data from the following longitudinal single-index regression model

$$Y_{ij} = g_0\left(\mathbf{X}_{ij}^T \beta_0\right) + \delta \varepsilon_{ij}, i = 1, \dots, n, j = 1, \dots, m_i,$$

where $\delta = 0.5$, $g_0(u) = \sin\left(\frac{0.2\pi(u-A)}{B-A}\right)$ with $A = \sqrt{3}/2 - 1.645/\sqrt{12}$, $B = \sqrt{3}/2 + 1.645/\sqrt{12}$, $\beta_0 = (\beta_{01}, \beta_{02}, \beta_{03})^T = (3, 2, 0.4)^T / \sqrt{3^2 + 2^2 + 0.4^2}$, and the covariate $\mathbf{X}_{ij} = (X_{ij1}, X_{ij2}, X_{ij3})^T$ follows a multivariate normal distribution $N(0, \Sigma)$ with $(\Sigma)_{k,l} = 0.5^{|k-l|}$ for $1 \leq k, l \leq 3$. 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_j is generated from a multivariate normal distribution $N(\mathbf{0}, \Xi_i)$, where Ξ_i will be listed later.

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

By using a similar strategy of Liu and Zhang (2013), the covariance matrix Ξ_i is constructed by $\Xi_i = L_i D_i L_i^T$, where $D_i = \text{diag} \left(\exp(\mathbf{h}_{i1}^T \boldsymbol{\alpha}), \dots, \exp(\mathbf{h}_{im_i}^T \boldsymbol{\alpha}) \right)$ and L_i is a unit lower triangular matrix with (j, k) element $\omega_{ijk}^T \phi$ ($k < j = 2, \dots, m_i$), where $\phi = (-0.3, 0.2, 0, 0.5)^T$, $\boldsymbol{\alpha} = (-0.3, 0.5, 0.4, 0)^T$, $\mathbf{h}_{ij} = (1, h_{ij1}, h_{ij2}, h_{ij3})^T$, $\omega_{ijk} = (1, (t_{ij} - t_{ik}), (t_{ij} - t_{ik})^2, (t_{ij} - t_{ik})^3)^T$, h_{ijl} follows a standard normal distribution for $l = 1, 2, 3$, and t_{ij} is generated from the standard uniform distribution. In addition, we adopt a similar strategy of Liu and Zhang (2013) to generate unbalanced longitudinal data. Specifically, each subject is measured m_i times with $m_i - 1 \sim \text{Binomial}(11, 0.8)$, which results in different numbers of repeated measurements for each subject. For the covariates \mathbf{z}_{ij} and \mathbf{w}_{ijk} of the covariance model (5), we set $\mathbf{z}_{ij} = \mathbf{h}_{ij}$ and $\mathbf{w}_{ijk} = \omega_{ijk}$.

Example 2 The model setup is similar to that of Example 1. Firstly, we take $\delta = 1$ and $\beta_0 = (\beta_{01}, \beta_{02}, \beta_{03})^T = (3, 2, 1)^T / \sqrt{14}$. Secondly, we define the covariance matrix Ξ_i by $\Xi_i = \mathbf{A}_i^{-1} \mathbf{B}_i (\mathbf{A}_i^T)^{-1}$, where \mathbf{B}_i is an $m_i \times m_i$ diagonal matrix with the j th element $\sin(\pi \zeta_{ij}) / 3 + 0.5$ and $\zeta_{ij} \sim U(0, 2)$, and \mathbf{A}_i is a unit lower triangular matrix with (j, k) element $-\delta_{j,k}^{(i)}$ ($k < j = 2, \dots, m_i$), $\delta_{j,k}^{(i)} = 0.2 + 0.5(t_{ij} - t_{ik})$ and $t_{ij} \sim U(0, 1)$. For the covariates \mathbf{z}_{ij} and \mathbf{w}_{ijk} of the regression model (5), we set $\mathbf{z}_{ij} = (1, t_{ij}, t_{ij}^2, t_{ij}^3)^T$, and \mathbf{w}_{ijk} is the same as that in Example 1. Meanwhile, we set $m_i = 12$, but each element has 20% probability of being missing at random, which leads to unbalanced longitudinal data. Other settings are the same as that in Example 1.

Example 3 For a clear comparison, we adopt a similar strategy of Zhao et al. (2017) to construct the covariance matrix Ξ_i . We define $\Xi_i = \mathbf{B}_i^{1/2} \mathbf{H}_i \mathbf{B}_i^{1/2}$, where \mathbf{B}_i is given in Example 2 and \mathbf{H}_i follows either the compound symmetry structure (cs) or the AR(1) structure (ar1) with the correlation coefficient $\rho = 0.85$. In addition, we take $\delta = 1$ and $\beta_0 = (\beta_{01}, \beta_{02}, \beta_{03})^T = (3, 2, 0)^T / \sqrt{13}$. The scheme of generating unbalanced longitudinal data, and \mathbf{z}_{ij} and \mathbf{w}_{ijk} are the same as that in Example 2. Other settings are the same as that in Example 1.

Tables 1, 2, 3, 4, and 5 give the biases and the standard deviations (SD) of $\hat{\beta}_{ma}$, $\hat{\beta}_{cs}$, $\hat{\beta}_{ar}$, $\hat{\beta}_{in}$, and $\hat{\beta}_{pr}$ at $\tau = 0.5, 0.75$ for $n = 100$ and 400 . We can derive the following several observations from these tables. Firstly, it is easy to find that all methods yield unbiased estimators for the index coefficients β , since the corresponding biases are small. Furthermore, $\hat{\beta}_{pr}$ has smaller bias in most cases. Secondly, the estimator $\hat{\beta}_{in}$ performs better than $\hat{\beta}_{ma}$, which indicates “remove-one-component” method leads to more efficient estimators. Thirdly, the proposed smoothed estimator $\hat{\beta}_{pr}$ performs best (with smallest SD) among all methods. It is not surprised that the SDs of $\hat{\beta}_{cs}$ and $\hat{\beta}_{ar}$ are bigger than that of $\hat{\beta}_{pr}$ for Examples 1 and 2, because $\hat{\beta}_{cs}$ and $\hat{\beta}_{ar}$ use misspecified correlation structures for Examples 1 and 2, which results in a loss of efficiency. Fourthly, as far as we know, the correlation structure of $\psi_\tau(\boldsymbol{\epsilon}_i)$ also has the

Table 1 Simulation results of the bias ($\times 10^{-2}$) and SD ($\times 10^{-2}$) for $\beta = (\beta_1, \beta_2, \beta_3)^T$ with $n = 100, 400$ and Case 1 in Example 1

n	$\tau = 0.5$						$\tau = 0.75$					
	β_1		β_2		β_3		β_1		β_2		β_3	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
100												
$\hat{\beta}_{ma}$	-0.312	3.176	-0.105	4.918	0.264	4.797	-0.716	3.987	0.286	6.127	0.158	5.432
$\hat{\beta}_{cs}$	-0.304	2.987	0.018	4.661	-0.106	4.473	-0.568	3.692	0.209	5.752	-0.077	5.066
$\hat{\beta}_{ar}$	-0.287	2.988	-0.012	4.667	-0.098	4.507	-0.560	3.666	0.204	5.709	-0.077	5.051
$\hat{\beta}_{in}$	-0.235	2.906	-0.051	4.553	-0.146	4.304	-0.395	3.213	0.094	5.048	-0.089	4.588
$\hat{\beta}_{pr}$	-0.239	2.638	0.028	4.140	-0.145	3.935	-0.423	3.064	0.168	4.782	-0.015	4.415
400												
$\hat{\beta}_{ma}$	-0.004	1.595	-0.153	2.537	0.105	2.485	-0.141	1.888	-0.017	2.964	0.249	2.691
$\hat{\beta}_{cs}$	0.015	1.533	-0.162	2.433	0.059	2.412	-0.130	1.785	0.001	2.814	0.171	2.563
$\hat{\beta}_{ar}$	0.019	1.537	-0.169	2.439	0.061	2.410	-0.132	1.779	0.005	2.809	0.171	2.558
$\hat{\beta}_{in}$	0.022	1.464	-0.154	2.322	0.019	2.309	-0.118	1.617	0.022	2.561	0.118	2.320
$\hat{\beta}_{pr}$	-0.003	1.368	-0.098	2.158	0.015	2.118	-0.098	1.536	0.004	2.416	0.119	2.234

Table 2 Simulation results of the bias ($\times 10^{-2}$) and SD ($\times 10^{-2}$) for $\beta = (\beta_1, \beta_2, \beta_3)^T$ with $n = 100, 400$ and Case 2 in Example 1

n	$\tau = 0.5$						$\tau = 0.75$						
	β_1		β_2		β_3		β_1		β_2		β_3		
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	
100													
	$\hat{\beta}_{ma}$	-0.271	3.528	-0.372	5.649	0.646	5.176	-0.707	4.749	-0.145	7.333	0.062	7.420
	$\hat{\beta}_{cs}$	-0.243	3.372	-0.271	5.362	0.278	4.883	-0.592	4.394	-0.096	6.890	-0.306	6.944
	$\hat{\beta}_{ar}$	-0.219	3.342	-0.311	5.325	0.320	4.881	-0.569	4.414	-0.150	6.926	-0.262	6.987
	$\hat{\beta}_{in}$	-0.262	3.359	-0.227	5.281	0.283	4.774	-0.443	3.914	-0.116	6.136	-0.261	6.233
	$\hat{\beta}_{pr}$	-0.238	3.174	-0.190	4.982	0.214	4.551	-0.410	3.813	-0.106	5.995	-0.279	5.926
400													
	$\hat{\beta}_{ma}$	-0.211	1.706	0.105	2.705	0.293	2.574	-0.180	2.269	-0.016	3.614	0.029	3.572
	$\hat{\beta}_{cs}$	-0.177	1.672	0.073	2.659	0.225	2.500	-0.168	2.171	0.001	3.458	-0.035	3.423
	$\hat{\beta}_{ar}$	-0.172	1.661	0.063	2.643	0.248	2.498	-0.163	2.173	-0.009	3.469	-0.020	3.431
	$\hat{\beta}_{in}$	-0.180	1.599	0.086	2.544	0.233	2.454	-0.169	1.979	0.042	3.160	-0.019	3.153
	$\hat{\beta}_{pr}$	-0.160	1.512	0.080	2.415	0.197	2.286	-0.156	1.940	0.034	3.083	-0.015	3.016

Table 3 Simulation results of the bias ($\times 10^{-2}$) and SD ($\times 10^{-2}$) for $\beta = (\beta_1, \beta_2, \beta_3)^T$ with $n = 100, 400$ and Case 1 in Example 2

n	$\tau = 0.5$						$\tau = 0.75$					
	β_1		β_2		β_3		β_1		β_2		β_3	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
100												
$\hat{\beta}_{ma}$	-0.552	5.079	-0.598	7.935	0.188	7.297	-1.277	6.107	-0.397	9.450	0.776	8.779
$\hat{\beta}_{cs}$	-0.333	4.455	-0.527	7.054	-0.063	6.579	-0.632	5.222	-0.652	8.591	0.080	8.078
$\hat{\beta}_{ar}$	-0.275	4.621	-0.757	7.281	0.086	6.755	-0.743	5.485	-0.622	8.941	0.182	8.083
$\hat{\beta}_{in}$	-0.249	4.669	-0.657	7.330	-0.179	6.618	-0.804	5.310	-0.465	8.435	0.304	7.897
$\hat{\beta}_{pr}$	-0.112	3.810	-0.553	6.016	-0.079	5.512	-0.565	4.190	-0.270	6.987	0.161	6.661
400												
$\hat{\beta}_{ma}$	-0.135	2.450	-0.195	3.956	0.132	3.714	-0.340	3.214	-0.030	5.188	-0.019	4.633
$\hat{\beta}_{cs}$	-0.044	2.236	-0.187	3.580	-0.040	3.384	-0.202	2.838	-0.107	4.638	-0.048	4.114
$\hat{\beta}_{ar}$	-0.062	2.362	-0.194	3.792	-0.018	3.422	-0.237	2.907	-0.023	4.753	-0.160	4.240
$\hat{\beta}_{in}$	-0.102	2.267	-0.121	3.697	-0.023	3.413	-0.215	2.907	-0.083	4.727	-0.090	4.179
$\hat{\beta}_{pr}$	-0.078	1.908	-0.061	3.081	-0.050	2.916	-0.128	2.315	-0.070	3.825	-0.050	3.281

Table 4 Simulation results of the bias ($\times 10^{-2}$) and SD ($\times 10^{-2}$) for $\beta = (\beta_1, \beta_2, \beta_3)^T$ with $n = 100, 400$ and Case 2 in Example 2

n	$\tau = 0.5$						$\tau = 0.75$																																																																																																																																																				
	β_1		β_2		β_3		β_1		β_2		β_3																																																																																																																																																
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD																																																																																																																																															
100													$\hat{\beta}_{ma}$	-1.318	5.951	0.024	9.052	0.458	8.089	-2.088	7.692	-0.835	12.11	1.317	11.90	$\hat{\beta}_{cs}$	-0.753	5.342	-0.148	8.208	-0.169	7.021	-1.254	6.779	-0.537	10.94	-0.401	10.62	$\hat{\beta}_{ar}$	-0.743	5.392	-0.263	8.437	-0.150	7.393	-1.164	6.891	-0.964	11.38	-0.123	10.84	$\hat{\beta}_{in}$	-0.966	5.579	0.021	8.571	-0.199	7.617	-1.284	6.676	-0.919	10.89	0.551	10.45	$\hat{\beta}_{pr}$	-0.565	4.416	-0.120	6.884	-0.068	6.330	-0.938	5.412	-0.228	8.839	-0.114	8.540	400													$\hat{\beta}_{ma}$	-0.244	2.685	0.137	4.347	-0.332	3.997	-0.263	3.724	-0.379	5.951	0.054	5.523	$\hat{\beta}_{cs}$	-0.118	2.482	0.020	4.076	-0.372	3.715	-0.201	3.353	-0.358	5.424	0.085	5.029	$\hat{\beta}_{ar}$	-0.121	2.516	0.004	4.066	-0.340	3.780	-0.207	3.431	-0.377	5.610	0.078	5.092	$\hat{\beta}_{in}$	-0.169	2.525	0.073	4.097	-0.355	3.867	-0.156	3.507	-0.402	5.655	-0.034	5.045	$\hat{\beta}_{pr}$	-0.071	2.173	0.012	3.490	-0.333	3.303	-0.077	2.752	-0.337	4.417	0.102	3.969
$\hat{\beta}_{ma}$	-1.318	5.951	0.024	9.052	0.458	8.089	-2.088	7.692	-0.835	12.11	1.317	11.90																																																																																																																																															
$\hat{\beta}_{cs}$	-0.753	5.342	-0.148	8.208	-0.169	7.021	-1.254	6.779	-0.537	10.94	-0.401	10.62																																																																																																																																															
$\hat{\beta}_{ar}$	-0.743	5.392	-0.263	8.437	-0.150	7.393	-1.164	6.891	-0.964	11.38	-0.123	10.84																																																																																																																																															
$\hat{\beta}_{in}$	-0.966	5.579	0.021	8.571	-0.199	7.617	-1.284	6.676	-0.919	10.89	0.551	10.45																																																																																																																																															
$\hat{\beta}_{pr}$	-0.565	4.416	-0.120	6.884	-0.068	6.330	-0.938	5.412	-0.228	8.839	-0.114	8.540																																																																																																																																															
400													$\hat{\beta}_{ma}$	-0.244	2.685	0.137	4.347	-0.332	3.997	-0.263	3.724	-0.379	5.951	0.054	5.523	$\hat{\beta}_{cs}$	-0.118	2.482	0.020	4.076	-0.372	3.715	-0.201	3.353	-0.358	5.424	0.085	5.029	$\hat{\beta}_{ar}$	-0.121	2.516	0.004	4.066	-0.340	3.780	-0.207	3.431	-0.377	5.610	0.078	5.092	$\hat{\beta}_{in}$	-0.169	2.525	0.073	4.097	-0.355	3.867	-0.156	3.507	-0.402	5.655	-0.034	5.045	$\hat{\beta}_{pr}$	-0.071	2.173	0.012	3.490	-0.333	3.303	-0.077	2.752	-0.337	4.417	0.102	3.969																																																																														
$\hat{\beta}_{ma}$	-0.244	2.685	0.137	4.347	-0.332	3.997	-0.263	3.724	-0.379	5.951	0.054	5.523																																																																																																																																															
$\hat{\beta}_{cs}$	-0.118	2.482	0.020	4.076	-0.372	3.715	-0.201	3.353	-0.358	5.424	0.085	5.029																																																																																																																																															
$\hat{\beta}_{ar}$	-0.121	2.516	0.004	4.066	-0.340	3.780	-0.207	3.431	-0.377	5.610	0.078	5.092																																																																																																																																															
$\hat{\beta}_{in}$	-0.169	2.525	0.073	4.097	-0.355	3.867	-0.156	3.507	-0.402	5.655	-0.034	5.045																																																																																																																																															
$\hat{\beta}_{pr}$	-0.071	2.173	0.012	3.490	-0.333	3.303	-0.077	2.752	-0.337	4.417	0.102	3.969																																																																																																																																															

Table 5 Simulation results of the bias ($\times 10^{-2}$) and SD ($\times 10^{-2}$) for $\beta = (\beta_1, \beta_2, \beta_3)^T$ with $n = 100$ in Example 3

n	$\tau = 0.5$						$\tau = 0.75$						
	β_1		β_2		β_3		β_1		β_2		β_3		
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	
<i>Case 1</i>													
CS													
	$\hat{\beta}_{ma}$	-0.745	4.042	0.317	6.052	-0.087	5.949	-0.709	4.544	0.080	6.774	0.440	6.490
	$\hat{\beta}_{cs}$	-0.242	2.035	0.156	3.047	-0.054	3.080	-0.212	2.658	-0.036	3.995	-0.324	4.034
	$\hat{\beta}_{ar}$	-0.263	2.397	0.105	3.595	-0.208	3.667	-0.311	3.124	-0.001	4.698	-0.116	4.482
	$\hat{\beta}_{in}$	-0.614	3.686	0.254	5.519	-0.307	5.433	-0.371	3.792	-0.123	5.707	0.122	5.330
	$\hat{\beta}_{pr}$	-0.229	1.920	0.163	2.876	-0.063	2.826	-0.219	2.336	0.067	3.507	-0.098	3.363
	arl												
	$\hat{\beta}_{ma}$	-0.593	4.130	0.096	6.123	0.117	5.771	-0.507	4.788	-0.285	7.157	0.325	6.451
	$\hat{\beta}_{cs}$	-0.294	3.021	0.028	4.493	-0.076	4.069	-0.221	3.587	-0.266	5.421	-0.175	4.905
	$\hat{\beta}_{ar}$	-0.173	2.373	0.002	3.521	-0.109	3.248	-0.194	3.089	-0.167	4.614	-0.332	4.457
	$\hat{\beta}_{in}$	-0.481	3.666	0.085	5.443	-0.108	5.243	-0.352	4.147	-0.264	6.163	0.044	5.706
	$\hat{\beta}_{pr}$	-0.133	2.180	-0.018	3.246	-0.069	2.967	-0.137	2.668	-0.130	4.011	-0.134	3.756
<i>Case 2</i>													
CS													
	$\hat{\beta}_{ma}$	-0.581	4.391	-0.054	6.589	0.393	6.298	-0.824	5.786	-0.411	8.648	0.425	8.589
	$\hat{\beta}_{cs}$	-0.017	2.370	-0.264	3.581	0.010	3.690	-0.188	3.504	-0.285	5.265	-0.579	4.751
	$\hat{\beta}_{ar}$	-0.163	2.792	-0.145	4.153	-0.093	4.268	-0.154	3.865	-0.508	5.909	-0.498	5.635
	$\hat{\beta}_{in}$	-0.545	4.121	0.003	6.188	0.043	5.912	-0.370	4.939	-0.629	7.449	-0.128	7.161
	$\hat{\beta}_{pr}$	-0.107	2.335	-0.098	3.515	0.089	3.317	-0.136	2.989	-0.216	4.489	-0.416	4.164

Table 5 continued

<i>n</i>	$\tau = 0.5$											
	β_1				β_2				β_3			
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
ar1												
$\hat{\beta}_{ma}$	-0.538	4.440	-0.134	6.660	0.509	6.313	-0.743	5.746	-0.386	8.404	0.336	7.893
$\hat{\beta}_{cs}$	-0.200	3.216	-0.205	4.872	0.003	4.690	-0.469	4.513	-0.226	6.614	-0.343	6.228
$\hat{\beta}_{ar}$	-0.330	2.746	0.136	4.067	-0.265	3.941	-0.360	3.908	-0.156	5.680	-0.463	5.422
$\hat{\beta}_{in}$	-0.329	4.199	-0.310	6.214	0.280	5.727	-0.582	5.176	-0.316	7.545	-0.094	6.936
$\hat{\beta}_{pr}$	-0.185	2.521	-0.016	3.750	-0.025	3.485	-0.179	3.394	-0.264	5.022	-0.038	4.731

exchangeable structure if the correlation structure of $\boldsymbol{\varepsilon}_i$ is exchangeable. Thus, $\hat{\beta}_{cs}$ use a correct working correlation structure under the compound symmetry (cs) structure in Table 5. However, $\hat{\beta}_{pr}$ also has slighter advantage than $\hat{\beta}_{cs}$. The main reasons are as follows. On the one hand, we adopt another smoothed approach that is different from that of Zhao et al. (2017) to construct the smoothed estimating equations. On the other hand, the moment approach does not work well and yields inaccuracy estimator of the correlation coefficient for unbalanced longitudinal data with large m_i . This leads to a bad estimator of C_i (given in Sect. 2.3 of Zhao et al. 2017). Finally, the correlation structure of $\psi_\tau(\boldsymbol{\varepsilon}_i)$ does not possess the AR(1) correlation structure when $\boldsymbol{\varepsilon}_i$ has the AR(1) correlation structure. Therefore, $\hat{\beta}_{cs}$ and $\hat{\beta}_{ar}$ use the misspecification of correlation structure under the AR(1) correlation structure in Example 3. Therefore, $\hat{\beta}_{cs}$ and $\hat{\beta}_{ar}$ perform worse than $\hat{\beta}_{pr}$. In addition, for the nonparametric link function $g_0(\cdot)$, we apply the mean squared error (MSE) to evaluate the accuracy of the estimator, which is defined as

$$MSE(g) = \frac{1}{500} \sum_{k=1}^{500} \left[\frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{m_i} \left(g^{(k)}(u_{ij}) - g_0(u_{ij}) \right)^2 \right],$$

where $g^{(k)}(u_{ij})$ is the k th estimated value of $g_0(u_{ij})$. From Table 6, the proposed \hat{g}_{pr} achieves the smallest MSE among all methods in general, which indicates that \hat{g}_{pr} outperforms the existing approaches. Overall, the proposed estimators $\hat{\beta}_{pr}$ and \hat{g}_{pr} can achieve better efficiency than the existing methods.

In order to evaluate the accuracy of the sandwich formula (10), we give the ratio of sample standard deviation (SD) and the estimated standard error (SE). For brevity, we only list the results of Example 2. In Table 7, ‘‘SD’’ represents the sample standard deviation of 500 estimators of the parameters. It can be taken as the true standard deviation of the resulting estimators. ‘‘SE’’ represents the sample average of 500 estimated standard errors by utilizing the sandwich formula (10). Table 7 indicates that the sandwich formula (10) works well for different situations, especially for large simple size ($n = 400$), since the value of SD/SE is very close to one. Compared with the method of Zhao et al. (2017), our method provides more accurate variance estimation. In addition, we use $P_{0.95}$ to stand for the coverage probabilities of 95% confidence intervals over 500 repetitions. From Table 7, the proposed estimator $\hat{\beta}_{pr}$ consistently achieves higher coverage probabilities and it is closer to its nominal level.

Finally, it is an interesting question to test whether the proposed estimates $\bar{\beta}$ and $\bar{g}(\cdot)$ are sensitive to the dimensions of the covariates \mathbf{z}_{ij} and \mathbf{w}_{ijk} . For brevity, we only present the results of $\bar{\beta}$ and $\bar{g}(\cdot)$ for Case 1 in Example 2. We set $\mathbf{w}_{ijk} = \{1, t_{ij} - t_{ik}, \dots, (t_{ij} - t_{ik})^{p_1-1}\}^T$ and $\mathbf{z}_{ij} = (1, t_{ij}, \dots, t_{ij}^{p_2-1})^T$, where (p_1, p_2) are given in Fig. 1. From Fig. 1, we can see that $\bar{\beta}$ and $\bar{g}(\cdot)$ are not very sensitive to the dimensions (p_1, p_2) .

Table 6 Simulation results of $MSE(\times 10^{-3})$ for $\tau = 0.5$ with $n = 100, 400$ in Examples 1–3

	$n = 100$				$n = 400$					
	\hat{g}_{ma}	\hat{g}_{cs}	\hat{g}_{ar}	\hat{g}_{in}	\hat{g}_{pr}	\hat{g}_{ma}	\hat{g}_{cs}	\hat{g}_{ar}	\hat{g}_{in}	\hat{g}_{pr}
Example 1										
Case 1	3.420	3.025	3.032	3.287	2.424	0.953	0.892	0.887	0.924	0.729
Case 2	4.662	4.218	4.222	4.541	3.582	1.123	1.055	1.048	1.091	0.881
Example 2										
Case 1	10.67	8.740	9.187	10.27	5.513	2.567	2.228	2.291	2.485	1.524
Case 2	13.25	11.22	11.70	12.89	8.097	3.212	2.842	2.917	3.141	2.044
Example 3										
Case 1-cs	8.994	5.741	6.274	8.803	5.340	2.378	1.634	1.799	2.322	1.501
Case 1-ar1	7.896	5.801	4.819	7.783	5.776	2.067	1.565	1.365	2.001	1.445
Case 2-cs	10.66	6.967	7.671	10.47	6.759	2.607	1.771	1.932	2.574	1.691
Case 2-ar1	9.394	6.905	5.976	9.132	6.970	2.362	1.795	1.527	2.296	1.917

Table 7 Simulation results of the SE/SD and $P_{0.95}$ for $\beta = (\beta_1, \beta_2, \beta_3)^T$ with $n = 100, 400$ in Example 2

n	$\tau = 0.5$						$\tau = 0.75$					
	β_1		β_2		β_3		β_1		β_2		β_3	
	SE/SD	$P_{0.95}$	SE/SD	$P_{0.95}$	SE/SD	$P_{0.95}$	SE/SD	$P_{0.95}$	SE/SD	$P_{0.95}$	SE/SD	$P_{0.95}$
<i>Case 1</i>												
100												
$\hat{\beta}_{cs}$	0.882	0.930	0.923	0.924	0.914	0.924	0.864	0.918	0.859	0.920	0.848	0.910
$\hat{\beta}_{ar}$	0.871	0.896	0.918	0.902	0.908	0.900	0.838	0.886	0.836	0.872	0.857	0.872
$\hat{\beta}_{pr}$	0.930	0.920	0.968	0.944	0.963	0.908	0.971	0.940	0.941	0.914	0.905	0.920
400												
$\hat{\beta}_{cs}$	0.952	0.956	0.992	0.938	0.952	0.944	0.877	0.928	0.888	0.924	0.921	0.928
$\hat{\beta}_{ar}$	0.921	0.928	0.957	0.944	0.965	0.942	0.877	0.914	0.885	0.908	0.916	0.922
$\hat{\beta}_{pr}$	0.998	0.958	1.012	0.948	0.975	0.950	0.940	0.944	0.925	0.940	0.994	0.946
<i>Case 2</i>												
100												
$\hat{\beta}_{cs}$	0.825	0.922	0.886	0.910	0.927	0.928	0.809	0.898	0.804	0.888	0.763	0.892
$\hat{\beta}_{ar}$	0.837	0.886	0.878	0.888	0.895	0.900	0.815	0.852	0.797	0.840	0.769	0.842
$\hat{\beta}_{pr}$	0.930	0.938	0.968	0.942	0.952	0.920	0.925	0.918	0.910	0.910	0.858	0.878
400												
$\hat{\beta}_{cs}$	0.929	0.940	0.939	0.938	0.941	0.936	0.912	0.926	0.937	0.934	0.930	0.934
$\hat{\beta}_{ar}$	0.942	0.918	0.967	0.922	0.953	0.922	0.913	0.912	0.930	0.906	0.941	0.908
$\hat{\beta}_{pr}$	0.979	0.938	0.994	0.936	0.960	0.938	0.974	0.942	0.996	0.936	1.009	0.934

Table 8 Estimated regression coefficients and their standard errors (SE), and PMSE for progesterone data

n	$\tau = 0.5$				$\tau = 0.75$			
	β_1		β_2		β_1		β_2	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
$\hat{\beta}_{cs}$	0.2978	0.0750	0.9546	0.0234	0.3079	0.0686	0.9514	0.0222
$\hat{\beta}_{ar}$	0.3059	0.0750	0.9521	0.0241	0.3181	0.0682	0.9480	0.0229
$\hat{\beta}_{in}$	0.3146	0.0850	0.9492	0.0282	0.3236	0.0618	0.9462	0.0212
$\hat{\beta}_{32}$	0.3594	0.0719	0.9332	0.0277	0.3377	0.0600	0.9412	0.0216
$\hat{\beta}_{23}$	0.3561	0.0721	0.9345	0.0275	0.3560	0.0607	0.9344	0.0231
$\hat{\beta}_{44}$	0.3458	0.0757	0.9383	0.0279	0.3340	0.0712	0.9426	0.0252
								PMSE
								0.1711
								0.1564
								0.1234
								0.1238
								0.1150
								0.1207

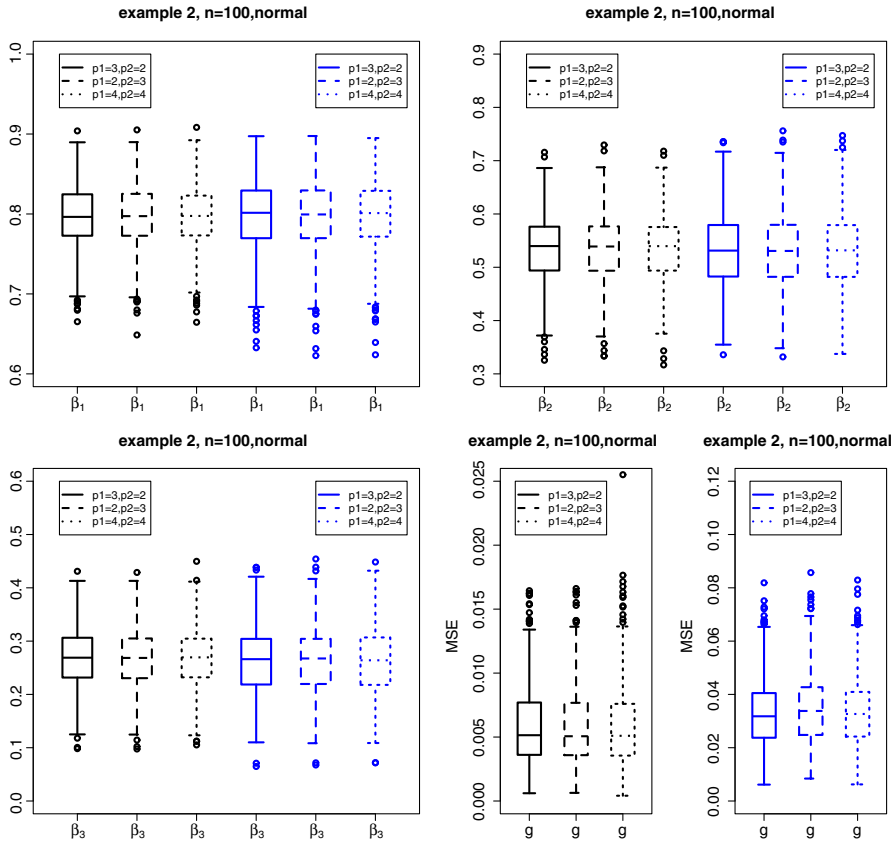


Fig. 1 Estimators of $\beta = (\beta_1, \beta_2, \beta_3)^T$ and $g(\cdot)$ at $\tau = 0.5$ (black) and $\tau = 0.75$ (blue) for Case 1 in Example 2 (color figure online)

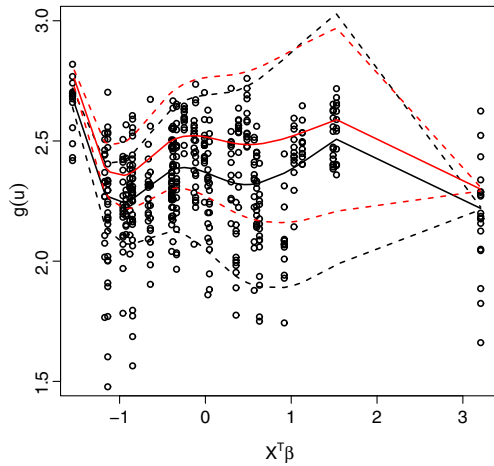
6 Real data analysis

In this section, we illustrate the proposed estimation method through an empirical example which has been studied by Zhang et al. (1998). These data include 34 women whose urine samples were collected in one menstrual cycle and whose urinary progesterone was assayed on alternate days. These women were measured 11–28 times, and it involves a total of 492 observations. Our goal is to explore the relationship between progesterone level and the following covariates: patient’s age and body mass index. Therefore, we define the log-transformed progesterone level as the response (Y), age, and body mass index are taken as the covariates. We use the longitudinal single-index quantile regression model to analyze this data set

$$Y_{ij} = g(\beta_1 X_{ij1} + \beta_2 X_{ij2}) + \varepsilon_{ij},$$

where $(Y_{ij}, X_{ij1}, X_{ij2})$ is the j th observed value at the time t_{ij} for the i th woman, X_{ij1} and X_{ij2} are the standardized variables of age and body mass index, respectively.

Fig. 2 Scatter plots of the response and the estimated link functions (solid curve) with 95% confidence intervals (dashed curve) for $\tau = 0.5$ (black) and $\tau = 0.75$ (red) (color figure online)



Meanwhile, repeated measurement time t_{ij} is rescaled into interval $[0, 1]$. For the covariance model (5), we take the corresponding covariates as

$$w_{ijk} = \left\{ 1, t_{ij} - t_{ik}, \dots, (t_{ij} - t_{ik})^{p_1-1} \right\}^T, z_{ij} = \left(1, t_{ij}, \dots, t_{ij}^{p_2-1} \right)^T.$$

We consider different (p_1, p_2) for this data set. Six estimators are considered: $\hat{\beta}_{cs}$, $\hat{\beta}_{ar}$, and $\hat{\beta}_{in}$ are given in Sect. 5, and $\hat{\beta}_{32}$, $\hat{\beta}_{23}$, and $\hat{\beta}_{44}$ represent the proposed estimators with $(p_1 = 3, p_2 = 2)$, $(p_1 = 2, p_2 = 3)$, and $(p_1 = 4, p_2 = 4)$, respectively. The leave-one-out cross-validation procedure is used to evaluate the forecasting accuracy of the estimators. Specifically, we investigate the forecasting accuracy of different methods by using the prediction mean squared error (PMSE), which is defined as

$$PMSE = \frac{1}{n} \sum_{i=1}^n \left\| Y_i - g \left(X_i \beta_{(-i)} \right) \right\|^2,$$

where $Y_i = (Y_{i1}, \dots, Y_{im_i})^T$, $X_i = (X_{i1}, \dots, X_{im_i})^T$, $X_{ij} = (X_{ij1}, X_{ij2})^T$, and $\beta_{(-i)}$ is the estimator which is obtained based on the data of the other 33 subjects except the i th subject. In Table 8, we report the PMSE, the estimated regression coefficients, and the corresponding standard errors which are obtained by the sandwich formula (10). Compared with the method of Zhao et al. (2017), our proposed method possesses smaller standard errors in general. Meanwhile, we see that our method has smaller PMSE, which indicates that the forecasting accuracy of our method is better. In addition, scatter plots of the response and the estimated link functions with 95% confidence intervals for $\tau = 0.5$ and 0.75 are displayed in Fig. 2. It is clear that there is a nonlinear trend. Thus, using a nonlinear term in the regression is perhaps more appropriate than using a linear term.

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

Appendix

In the proofs, C denotes a positive constant that might assume different values at different places. For any matrix $A = (A_{ij})_{i=1, j=1}^{s, t}$, denote $\|A\|_\infty = \max_{1 \leq i \leq s} \sum_{j=1}^t |A_{ij}|$. To establish the asymptotic properties of the proposed estimators, the following regularity conditions are needed in this paper.

(C1) Let $\mathcal{U} = \left\{ u : u = X_{ij}^T \beta, X_{ij} \in A, i = 1, \dots, n, j = 1, \dots, m_i \right\}$ and A be the support of X_{ij} which is assumed to be compact. Suppose that the density function $f_{X_{ij}^T \beta}(u)$ of $X_{ij}^T \beta$ is bounded away from zero and infinity on \mathcal{U} and satisfies the Lipschitz condition of order 1 on \mathcal{U} for β in a neighborhood of β_0 .

(C2) The function $g_0(\cdot)$ has the d th bounded and continuous derivatives for some $d \geq 2$ and $g_{1s}(\cdot)$ satisfies the Lipschitz condition of order 1, where $g_{1s}(u)$ is the s th component of $\mathbf{g}_1(u) = E(X_{ij} | X_{ij}^T \beta_0 = u)$, $s = 1, \dots, p$.

(C3) Let the distance between neighboring knots be $H_i = \xi_i - \xi_{i-1}$ and $H = \max_{1 \leq i \leq N_{n+1}} \{H_i\}$. Then, there exists a constant C_0 such that $\frac{H}{\min_{1 \leq i \leq N_{n+1}} \{H_i\}} < C_0$, $\max_{1 \leq i \leq N_n} \{H_{i+1} - H_i\} = o(N_n^{-1})$.

(C4) The distribution function $F_{ij}(t) = p(Y_{ij} - g_0(X_{ij}^T \beta_0) \leq t)$ is absolutely continuous, with continuous densities $f_{ij}(\cdot)$ uniformly bounded, and its first derivative $f'_{ij}(\cdot)$ uniformly bounded away from 0 and ∞ at the point 0, $i = 1, \dots, n, j = 1, \dots, m_i$.

(C5) The eigenvalues of Σ_{τ_i} are uniformly bounded and bounded away from zero.

(C6) $K(\cdot)$ is bounded and compactly supported on $[-1, 1]$. For some constant $C_K \neq 0$, $K(\cdot)$ is a ν th-order kernel, i.e., $\int u^j K(u) du = 1$ if $j = 0$; 0 if $1 \leq j \leq \nu - 1$; C_K if $j = \nu$, where $\nu \geq 2$.

(C7) The positive bandwidth parameter h satisfies $nh^{2\nu} \rightarrow 0$.

Lemma 1 Under conditions (C1)–(C7), and $N_n \rightarrow \infty$ and $nN_n^{-1} \rightarrow \infty$, as $n \rightarrow \infty$, we have (i) $|\hat{g}(u; \beta_0) - g_0(u)| = O_p\left(\sqrt{N_n/n} + N_n^{-d}\right)$ uniformly for any $u \in [a, b]$; and (ii) under $N_n \rightarrow \infty$ and $nN_n^{-3} \rightarrow \infty$, as $n \rightarrow \infty$, $|\hat{g}'(u; \beta_0) - g'_0(u)| = O_p\left(\sqrt{N_n^3/n} + N_n^{-d+1}\right)$ uniformly for any $u \in [a, b]$.

Proof Suppose $g^0(u) = \mathbf{B}_q(u)^T \boldsymbol{\theta}^0$ is the best approximating spline function for $g_0(u)$. According to the result on page 149 of de Boor (2001) for $g_0(u)$ satisfying condition (C2), we have

$$\sup_{u \in [a, b]} |g_0(u) - g^0(u)| = O(N_n^{-d}). \tag{11}$$

Let $\alpha_n = N_n n^{-1/2} + N_n^{-d+1/2}$ and set $\|\mathbf{u}_n\| = C$, where C is a large enough constant. Our aim is to show that for any given $\delta > 0$, there is a large constant C such that, for large n , we have

$$P \left\{ \inf_{\|\mathbf{u}_n\|=C} L_n(\boldsymbol{\beta}_0; \boldsymbol{\theta}^0 + \alpha_n \mathbf{u}_n) > L_n(\boldsymbol{\beta}_0; \boldsymbol{\theta}^0) \right\} \geq 1 - \delta. \tag{12}$$

This implies that there is local minimum $\hat{\boldsymbol{\theta}}$ in the ball $\{\boldsymbol{\theta}^0 + \alpha_n \mathbf{u}_n : \|\mathbf{u}_n\| \leq C\}$ with probability tending to one, such that $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\| = O_p(\alpha_n)$. Define $\Delta_{ij} = \mathbf{B}_q(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0)^T \boldsymbol{\theta}^0 - g_0(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0)$. Applying the identity

$$\rho_\tau(r - v) - \rho_\tau(r) = -v(\tau - I(r < 0)) + \int_0^v [I(r \leq t) - I(r \leq 0)] dt,$$

we have

$$\begin{aligned} & L_n(\boldsymbol{\beta}_0; \boldsymbol{\theta}^0 + \alpha_n \mathbf{u}_n) - L_n(\boldsymbol{\beta}_0; \boldsymbol{\theta}^0) \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} \rho_\tau(\varepsilon_{ij} - \alpha_n \mathbf{B}_q(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0)^T \mathbf{u}_n - \Delta_{ij}) - \sum_{i=1}^n \sum_{j=1}^{m_i} \rho_\tau(\varepsilon_{ij} - \Delta_{ij}) \\ &= - \sum_{i=1}^n \sum_{j=1}^{m_i} \alpha_n \mathbf{B}_q(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0)^T \mathbf{u}_n (\tau - I(\varepsilon_{ij} - \Delta_{ij} < 0)) \\ &\quad + \sum_{i=1}^n \sum_{j=1}^{m_i} \int_0^{\alpha_n \mathbf{B}_q(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0)^T \mathbf{u}_n} [I(\varepsilon_{ij} - \Delta_{ij} \leq t) - I(\varepsilon_{ij} - \Delta_{ij} \leq 0)] dt \\ &\triangleq I + II. \end{aligned}$$

The observed covariates vector is written as $\mathcal{D} = \{\mathbf{X}_{11}^T, \dots, \mathbf{X}_{1m_1}^T, \dots, \mathbf{X}_{n1}^T, \dots, \mathbf{X}_{nm_n}^T\}^T$. Moreover, we have

$$\begin{aligned} & E(II \mid \mathcal{D}) \\ &= E \left(\sum_{i=1}^n \sum_{j=1}^{m_i} \int_0^{\alpha_n \mathbf{B}_q(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0)^T \mathbf{u}_n} [I(\varepsilon_{ij} - \Delta_{ij} \leq t) - I(\varepsilon_{ij} - \Delta_{ij} \leq 0)] dt \mid \mathcal{D} \right) \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} \int_0^{\alpha_n \mathbf{B}_q(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0)^T \mathbf{u}_n} [F_{ij}(\Delta_{ij} + t) - F_{ij}(\Delta_{ij})] dt \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} \int_0^{\alpha_n \mathbf{B}_q(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0)^T \mathbf{u}_n} f_{ij}(\Delta_{ij}) t (1 + o(1)) dt \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{m_i} f_{ij}(\Delta_{ij}) \left(\alpha_n \mathbf{B}_q(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0)^T \mathbf{u}_n \right)^2 (1 + o(1)) \end{aligned}$$

$$= \frac{1}{2} \alpha_n^2 \mathbf{u}_n^T \left(\sum_{i=1}^n \sum_{j=1}^{m_i} f_{ij}(0) \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right) \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right)^T \right) \mathbf{u}_n + o_p \left(n \alpha_n^2 \|\mathbf{u}_n\|^2 / N_n \right),$$

and

$$\begin{aligned} & \text{Var} (II \mid \mathcal{D}) \\ &= \text{Var} \left(\sum_{i=1}^n \sum_{j=1}^{m_i} \int_0^{\alpha_n \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right)^T \mathbf{u}_n} [I(\varepsilon_{ij} - \Delta_{ij} \leq t) - I(\varepsilon_{ij} - \Delta_{ij} \leq 0)] dt \mid \mathcal{D} \right) \\ &\leq \sum_{i=1}^n E \left[\left(\sum_{j=1}^{m_i} \int_0^{\alpha_n \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right)^T \mathbf{u}_n} [I(\varepsilon_{ij} - \Delta_{ij} \leq t) - I(\varepsilon_{ij} - \Delta_{ij} \leq 0)] dt \right)^2 \mid \mathcal{D} \right] \\ &\leq \sum_{i=1}^n m_i \sum_{j=1}^{m_i} E \left[\left(\int_0^{\alpha_n \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right)^T \mathbf{u}_n} [I(\varepsilon_{ij} - \Delta_{ij} \leq t) - I(\varepsilon_{ij} - \Delta_{ij} \leq 0)] dt \right)^2 \mid \mathcal{D} \right] \\ &\leq \sum_{i=1}^n m_i \sum_{j=1}^{m_i} \int_0^{|\alpha_n \mathbf{B}_q^T \left(\mathbf{X}_{ij}^T \beta_0 \right) \mathbf{u}_n|} \int_0^{|\alpha_n \mathbf{B}_q^T \left(\mathbf{X}_{ij}^T \beta_0 \right) \mathbf{u}_n|} \left[F_{ij} \left(\Delta_{ij} + |\alpha_n \mathbf{B}_q^T \left(\mathbf{X}_{ij}^T \beta_0 \right) \mathbf{u}_n| \right) - F_{ij} \left(\Delta_{ij} \right) \right] dt_1 dt_2 \\ &\leq o_p \left(n \alpha_n^2 \|\mathbf{u}_n\|^2 / N_n \right). \end{aligned}$$

In addition,

$$I = E(I) + O_p \left(\sqrt{\text{Var}(I)} \right).$$

Moreover, the condition that ε_{ij} has the τ th quantile zero implies $E(\psi_\tau(\varepsilon_{ij})) = 0$. By (11) and condition (C4), we have $E(I) = o(1)$ and

$$\begin{aligned} \text{Var}(I \mid \mathcal{D}) &\leq \sum_{i=1}^n E \left[\left(\alpha_n \sum_{j=1}^{m_i} (\tau - I(\varepsilon_{ij} - \Delta_{ij} \leq 0)) \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right)^T \mathbf{u}_n \right)^2 \mid \mathcal{D} \right] \\ &\leq C \mathbf{u}_n^T \alpha_n^2 \sum_{i=1}^n m_i \sum_{j=1}^{m_i} \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right) \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right)^T \mathbf{u}_n \\ &= O_p \left(n \alpha_n^2 \|\mathbf{u}_n\|^2 / N_n \right) \end{aligned}$$

implies that $I = O_p \left(\sqrt{n \alpha_n^2 / N_n} \right) \|\mathbf{u}_n\|$. Based on all the above, $L_n(\beta_0; \theta^0 + \alpha_n \mathbf{u}_n) - L_n(\beta_0; \theta^0)$ is dominated by $\frac{1}{2} \alpha_n^2 \mathbf{u}_n^T \left(\sum_{i=1}^n \sum_{j=1}^{m_i} f_{ij}(0) \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right) \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right)^T \right) \mathbf{u}_n$ by choosing a sufficiently large $\|\mathbf{u}_n\| = C$. Therefore, (12) holds and there exists a local minimizer $\hat{\theta}$ such that

$$\|\hat{\theta} - \theta^0\| = O_p(\alpha_n) = O_p \left(N_n^{-d+1/2} + N_n n^{-1/2} \right). \tag{13}$$

Since $\| \mathbf{B}_q(u) \mathbf{B}_q(u)^T \| = O(1/N_n)$, together with (13), we have

$$\begin{aligned} & |\hat{g}(u; \beta_0) - g^0(u)|^2 \\ &= \left| \mathbf{B}_q(u)^T \hat{\boldsymbol{\theta}} - \mathbf{B}_q(u)^T \boldsymbol{\theta}^0 \right|^2 \\ &\leq \| \mathbf{B}_q(u) \mathbf{B}_q(u)^T \| \| \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 \|^2 = O_p(N_n^{-2d} + N_n n^{-1}). \end{aligned} \tag{14}$$

By the triangle inequality, $|\hat{g}(u; \beta_0) - g_0(u)| \leq |\hat{g}(u; \beta_0) - g^0(u)| + |g^0(u) - g_0(u)|$. Therefore, by (11) and (14), we have $|\hat{g}(u; \beta_0) - g_0(u)| = O_p(N_n^{-d} + \sqrt{N_n/n})$ uniformly for every $u \in [a, b]$.

Since $\hat{g}'(u; \beta_0) = \mathbf{B}_{q-1}(u)^T \mathbf{D}_1 \hat{\boldsymbol{\theta}}(\beta_0)$, where $\mathbf{B}_{q-1}(u) = \{B_{s,q}(u) : 2 \leq s \leq J_n\}^T$ is the $(q - 1)$ th-order B-spline basis and \mathbf{D}_1 is defined in Sect. 2.1. It is easy to prove that $\| \mathbf{D}_1 \|_\infty = O(N_n)$. Then, employing similar techniques to that used in the proof of $\hat{g}(u; \beta_0)$, we obtain that

$$|\hat{g}'(u; \beta_0) - g'_0(u)| = O_p\left(\sqrt{N_n^3/n} + N_n^{-d+1}\right)$$

uniformly for any $u \in [a, b]$. This completes the proof. □

Lemma 2 Under conditions (C1)–(C7), and the number of knots satisfies $n^{1/(2d+1)} \ll N_n \ll n^{1/4}$, then for any $J_n \times 1$ -vector \mathbf{c}_n whose components are not all 0, we have

$$\bar{\sigma}_n^{-1}(u) \mathbf{c}_n^T \left(\hat{\boldsymbol{\theta}}(\beta_0) - \boldsymbol{\theta}^0(\beta_0) \right) \xrightarrow{d} N(0, 1),$$

where $\bar{\sigma}_n^2(u) = \mathbf{c}_n^T \mathbf{V}^{-1}(\beta_0) \sum_{i=1}^n \mathbf{B}_q^T(X_i \beta_0) \boldsymbol{\Lambda}_i \boldsymbol{\Sigma}_{\tau_i} \boldsymbol{\Lambda}_i \mathbf{B}_q(X_i \beta_0) \mathbf{V}^{-1}(\beta_0) \mathbf{c}_n$ and the definition of $\mathbf{V}(\beta_0)$ is given in subsection 2.4.

Proof When $\beta = \beta_0$, the minimizer $\hat{\boldsymbol{\theta}}$ of (1) satisfies the score equations

$$\sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{B}_q(X_{ij}^T \beta_0) f_{ij}(0) \left[I(Y_{ij} - \mathbf{B}_q(X_{ij}^T \beta_0)^T \boldsymbol{\theta} < 0) - \tau \right] = \mathbf{0}. \tag{15}$$

Then, the left-hand side of Eq. (15) becomes

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{B}_q(X_{ij}^T \beta_0) f_{ij}(0) [I(\varepsilon_{ij} < 0) - I(\varepsilon_{ij} < 0) + I(\varepsilon_{ij} < \zeta_{ij}) - \tau] \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{B}_q(X_{ij}^T \beta_0) f_{ij}(0) [I(\varepsilon_{ij} < 0) - \tau] \\ &+ \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{B}_q(X_{ij}^T \beta_0) f_{ij}(0) [F_{ij}(0) - F_{ij}(\zeta_{ij})] \end{aligned}$$

$$\begin{aligned}
 & - \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right) f_{ij} (0) \left[F_{ij} (0) - F_{ij} (\zeta_{ij}) \right. \\
 & \left. - \{ I (\varepsilon_{ij} < \zeta_{ij}) - I (\varepsilon_{ij} < 0) \} \right] \\
 & \triangleq I + II + III,
 \end{aligned} \tag{16}$$

where $\zeta_{ij} = \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right)^T \boldsymbol{\theta} - g_0 \left(\mathbf{X}_{ij}^T \beta_0 \right)$. By (11), taking Taylor’s explanation for $F_{ij} (\zeta_{ij})$ at 0 gives

$$\begin{aligned}
 II &= - \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right) f_{ij}^2 (0) \zeta_{ij} (1 + o (1)) \\
 &= - \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right) f_{ij}^2 (0) \\
 &\quad \times \left[\mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right)^T \left(\boldsymbol{\theta} - \boldsymbol{\theta}^0 \right) + \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right)^T \boldsymbol{\theta}^0 - g_0 \left(\mathbf{X}_{ij}^T \beta_0 \right) \right] (1 + o (1)) \\
 &= - \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right) \mathbf{B}_q \left(\mathbf{X}_{ij}^T \beta_0 \right)^T f_{ij}^2 (0) \left(\boldsymbol{\theta} - \boldsymbol{\theta}^0 \right) (1 + o (1)).
 \end{aligned} \tag{17}$$

By direct calculation of the mean and variance, we can show that $III = o_p \left(\sqrt{n/N_n} \right)$. This combined with (15)–(17) leads to

$$\left(\boldsymbol{\theta} - \boldsymbol{\theta}^0 \right) (1 + o_p (1)) = \mathbf{V}^{-1} \left(\beta_0 \right) \sum_{i=1}^n \mathbf{B}_q^T \left(\mathbf{X}_i \beta_0 \right) \mathbf{A}_i \psi_\tau \left(\boldsymbol{\varepsilon}_i \right).$$

It is easy to derive that $I = \sum_{i=1}^n \mathbf{B}_q^T \left(\mathbf{X}_i \beta_0 \right) \mathbf{A}_i \psi_\tau \left(\boldsymbol{\varepsilon}_i \right)$ is a sum of independent vector, $E (I) = 0$ and $Cov (I) = \sum_{i=1}^n \mathbf{B}_q^T \left(\mathbf{X}_i \beta_0 \right) \mathbf{A}_i \boldsymbol{\Sigma}_{\tau i} \mathbf{A}_i \mathbf{B}_q \left(\mathbf{X}_i \beta_0 \right)$. By the multivariate central limit theorem and the Slutsky’s theorem, we can complete the proof. \square

Lemma 3 Under conditions (C1)–(C7) and the number of knots satisfies $n^{1/(2d+2)} \ll N_n \ll n^{1/4}$, we have

$$\sqrt{n} \left(\hat{\beta} - \beta_0 \right) \xrightarrow{d} N \left(\mathbf{0}, \mathbf{J}_{\beta_0^{(r)}} \boldsymbol{\Phi}^{-1} \boldsymbol{\Psi} \boldsymbol{\Phi}^{-1} \mathbf{J}_{\beta_0^{(r)}}^T \right).$$

Proof Define $\mathbf{H}_i^T = \mathbf{J}_{\beta^{(r)}}^T \hat{\mathbf{X}}_i^T \hat{\mathbf{G}}' \left(\mathbf{X}_i \beta; \beta \right)$, $\mathbf{S}_i = \left(S_{i1}, \dots, S_{im_i} \right)^T$ with $S_{ij} = S_{ij} \left(\beta \right) = \psi_\tau \left(Y_{ij} - \hat{g} \left(\mathbf{X}_{ij}^T \beta; \beta \right) \right) = I \left(Y_{ij} - \hat{g} \left(\mathbf{X}_{ij}^T \beta; \beta \right) \leq 0 \right) - \tau$ being a discontinuous function, then $R \left(\beta^{(r)} \right) = \sum_{i=1}^n \mathbf{H}_i^T \mathbf{A}_i \mathbf{S}_i \left(\beta \right)$. Let $\bar{R} \left(\beta^{(r)} \right) = \sum_{i=1}^n \mathbf{H}_i^T \mathbf{A}_i \mathbf{P}_i \left(\beta \right)$, where $\mathbf{P}_i = \left(P_{i1}, \dots, P_{im_i} \right)^T$ with $P_{ij} = P_{ij} \left(\beta \right) =$

$p \left(Y_{ij} - \hat{g} \left(\mathbf{X}_{ij}^T \beta; \beta \right) \leq 0 \right) - \tau$. For any $\beta^{(r)}$ satisfying $\left\| \beta^{(r)} - \beta_0^{(r)} \right\| \leq Cn^{-1/2}$, we have

$$\begin{aligned} & R \left(\beta^{(r)} \right) - R \left(\beta_0^{(r)} \right) \\ &= \sum_{i=1}^n \mathbf{H}_i^T \left(\beta^{(r)} \right) \mathbf{A}_i \mathbf{S}_i \left(\beta \right) - \sum_{i=1}^n \mathbf{H}_i^T \left(\beta_0^{(r)} \right) \mathbf{A}_i \mathbf{S}_i \left(\beta_0 \right) \\ &= \sum_{i=1}^n \mathbf{H}_i^T \left(\beta^{(r)} \right) \mathbf{A}_i \left\{ \mathbf{S}_i \left(\beta \right) - \mathbf{S}_i \left(\beta_0 \right) \right\} \\ &\quad + \sum_{i=1}^n \left\{ \mathbf{H}_i \left(\beta^{(r)} \right) - \mathbf{H}_i \left(\beta_0^{(r)} \right) \right\}^T \mathbf{A}_i \mathbf{S}_i \left(\beta_0 \right). \end{aligned}$$

At first, the first term can be written as

$$\begin{aligned} & \sum_{i=1}^n \mathbf{H}_i^T \left(\beta^{(r)} \right) \mathbf{A}_i \left\{ \mathbf{S}_i \left(\beta \right) - \mathbf{S}_i \left(\beta_0 \right) \right\} \\ &= \sum_{i=1}^n \mathbf{H}_i^T \left(\beta^{(r)} \right) \mathbf{A}_i \mathbf{P}_i \left(\beta^{(r)} \right) + \sum_{i=1}^n \mathbf{H}_i^T \left(\beta^{(r)} \right) \mathbf{A}_i \left(\mathbf{S}_i \left(\beta \right) - \mathbf{S}_i \left(\beta_0 \right) - \mathbf{P}_i \left(\beta \right) \right) \\ &= \bar{R} \left(\beta^{(r)} \right) + \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{h}_{ij} f_{ij} \left(0 \right) \left[I \left(Y_{ij} - \hat{g} \left(\mathbf{X}_{ij}^T \beta; \beta \right) \leq 0 \right) \right. \\ &\quad \left. - I \left(Y_{ij} - \hat{g} \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) \leq 0 \right) - p \left(Y_{ij} - \hat{g} \left(\mathbf{X}_{ij}^T \beta; \beta \right) \leq 0 \right) + \tau \right] \\ &\triangleq \bar{R} \left(\beta^{(r)} \right) + \gamma, \end{aligned}$$

where $\mathbf{H}_i^T = \left(\mathbf{h}_{i1}, \dots, \mathbf{h}_{imi} \right)$ and \mathbf{h}_{ij} is a $(p-1) \times 1$ vector. According to Lemma 3 in Jung (1996) and Lemma 1, we have $\sup |\gamma| = o_p \left(\sqrt{n} \right)$. Then, the first term becomes

$$\sum_{i=1}^n \mathbf{H}_i^T \left(\beta^{(r)} \right) \mathbf{A}_i \left\{ \mathbf{S}_i \left(\beta \right) - \mathbf{S}_i \left(\beta_0 \right) \right\} = \bar{R} \left(\beta^{(r)} \right) + o_p \left(\sqrt{n} \right).$$

By the law of large numbers (Pollard 1990), together with Lemma 1, the second term becomes

$$\begin{aligned} & \sum_{i=1}^n \left\{ \mathbf{H}_i \left(\beta^{(r)} \right) - \mathbf{H}_i \left(\beta_0^{(r)} \right) \right\}^T \mathbf{A}_i \mathbf{S}_i \left(\beta_0 \right) \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} \left(\mathbf{h}_{ij} \left(\beta^{(r)} \right) - \mathbf{h}_{ij} \left(\beta_0^{(r)} \right) \right) f_{ij} \left(0 \right) \\ &\quad \times \left[I \left(Y_{ij} - \hat{g} \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) \leq 0 \right) - p \left(Y_{ij} - g_0 \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) \leq 0 \right) \right] \\ &= o_p \left(\sqrt{n} \right). \end{aligned}$$

Therefore, $R \left(\beta^{(r)} \right) - R \left(\beta_0^{(r)} \right) = \bar{R} \left(\beta^{(r)} \right) + o_p \left(\sqrt{n} \right)$. By Taylor’s expansion of $\bar{R} \left(\beta^{(r)} \right)$, we can obtain

$$R \left(\beta^{(r)} \right) - R \left(\beta_0^{(r)} \right) = \left[\partial \bar{R} \left(\beta^{(r)} \right) / \partial \beta^{(r)} \right] \Big|_{\beta^{(r)} = \beta_0^{(r)}} \left(\beta^{(r)} - \beta_0^{(r)} \right) + o_p \left(\sqrt{n} \right).$$

Because $R(\hat{\beta}^{(r)}) = 0$ and $\hat{\beta}^{(r)}$ is in the $n^{-1/2}$ neighborhood of $\beta_0^{(r)}$, we have

$$\sqrt{n}(\hat{\beta}^{(r)} - \beta_0^{(r)}) = -\left(\frac{1}{n}D(\beta_0^{(r)})\right)^{-1} \frac{1}{\sqrt{n}}R(\beta_0^{(r)}) + o_p(1), \tag{18}$$

where $R(\beta_0^{(r)}) = \sum_{i=1}^n \mathbf{J}_{\beta^{(r)}}^T \hat{\mathbf{X}}_i^T \hat{\mathbf{G}}'(X_i\beta; \beta) \mathbf{A}_i \mathbf{S}_i(\beta)|_{\beta^{(r)}=\beta_0^{(r)}}$, $\mathbf{S}_i(\beta) = (S_{i1}(\beta), \dots, S_{im_i}(\beta))^T$ with $S_{ij}(\beta) = I(Y_{ij} - \hat{g}(X_{ij}^T\beta_0; \beta) \leq 0) - \tau$ and

$$D(\beta_0^{(r)}) = \sum_{i=1}^n \mathbf{J}_{\beta^{(r)}}^T \hat{\mathbf{X}}_i^T \hat{\mathbf{G}}'(X_i\beta; \beta) \mathbf{A}_i \mathbf{A}_i \hat{\mathbf{G}}'(X_i\beta; \beta) \hat{\mathbf{X}}_i \mathbf{J}_{\beta^{(r)}}|_{\beta^{(r)}=\beta_0^{(r)}}.$$

Thus, we have

$$\begin{aligned} S_{ij}(\beta_0) &= S_{ij}(\beta_0) - S_{ij}^*(\beta_0) + S_{ij}^*(\beta_0) \\ &= I(Y_{ij} - \hat{g}(X_{ij}^T\beta_0; \beta_0) \leq 0) - I(Y_{ij} - g_0(X_{ij}^T\beta_0) \leq 0) \\ &\quad + I(Y_{ij} - g_0(X_{ij}^T\beta_0) \leq 0) - \tau, \end{aligned}$$

where $S_{ij}^*(\beta_0) = I(Y_{ij} - g_0(X_{ij}^T\beta_0) \leq 0) - \tau$. Moreover, $| \hat{g}(X_{ij}^T\beta_0; \beta_0) - g_0(X_{ij}^T\beta_0) | = O_p(\sqrt{N_n/n} + N_n^{-d})$ and $E\{I(Y_{ij} - g_0(X_{ij}^T\beta_0) \leq 0)\} = p(Y_{ij} - g_0(X_{ij}^T\beta_0) \leq 0) = \tau$, we have $E(S_{ij}(\beta_0)) = o(1)$. Therefore, we have $E(\mathbf{S}_i(\beta_0)) = o(1)$ and

$$\begin{aligned} \text{Var}\left(\frac{1}{\sqrt{n}}R(\beta_0^{(r)})\right) &= \frac{1}{n} \sum_{i=1}^n \mathbf{J}_{\beta^{(r)}}^T \hat{\mathbf{X}}_i^T \hat{\mathbf{G}}'(X_i\beta; \beta) \mathbf{A}_i \Sigma_{\tau i} \mathbf{A}_i \hat{\mathbf{G}}'(X_i\beta; \beta) \hat{\mathbf{X}}_i \mathbf{J}_{\beta^{(r)}}|_{\beta^{(r)}=\beta_0^{(r)}} (1 + o(1)). \end{aligned}$$

Based on Lemma 1, together with $\mathbf{S}_i(\beta_0)$ are the independent random variables, the multivariate central limit theorem implies that

$$\frac{1}{\sqrt{n}}R(\beta_0^{(r)}) \xrightarrow{d} N(\mathbf{0}, \Psi). \tag{19}$$

By the law of large numbers and Lemma 1, we have

$$\frac{1}{n}D(\beta_0^{(r)}) \xrightarrow{p} \Phi. \tag{20}$$

Then, combine (18)–(20) and use the Slutsky’s theorem; it follows that

$$\sqrt{n}(\hat{\beta}^{(r)} - \beta_0^{(r)}) \xrightarrow{d} N(\mathbf{0}, \Phi^{-1}\Psi\Phi^{-1}). \tag{21}$$

According to the result of (21) and the multivariate delta method, we have

$$\sqrt{n} \left(\hat{\beta} - \beta_0 \right) \xrightarrow{d} N \left(\mathbf{0}, \mathbf{J}_{\beta_0^{(r)}} \Phi^{-1} \Psi \Phi^{-1} \mathbf{J}_{\beta_0^{(r)}}^T \right).$$

□

Proof of Theorem 1 Using conditions (C4), (C6), and (C7), similar to Lemma 3 (k) of Horowitz (1998), we obtain $n^{-1/2} \tilde{R} \left(\beta_0^{(r)} \right) = n^{-1/2} R \left(\beta_0^{(r)} \right) + o_p(1)$. In order to prove the asymptotic normality of $\tilde{\beta}^{(r)}$, we need to prove $n^{-1} \left\{ \tilde{D} \left(\beta_0^{(r)} \right) - D \left(\beta_0^{(r)} \right) \right\} \xrightarrow{p} 0$, where

$$\begin{aligned} \frac{1}{n} \tilde{D} \left(\beta_0^{(r)} \right) &\triangleq \frac{1}{n} \frac{\partial \tilde{R}(\beta^{(r)})}{\partial \beta^{(r)}} \Big|_{\beta^{(r)} = \beta_0^{(r)}} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{J}_{\beta_0^{(r)}}^T \hat{\mathbf{X}}_i^T \hat{\mathbf{G}}' \left(\mathbf{X}_i; \beta_0; \beta_0 \right) \mathbf{A}_i \tilde{\mathbf{A}}_i \left(\beta_0 \right) \hat{\mathbf{G}}' \left(\mathbf{X}_i; \beta_0; \beta_0 \right) \hat{\mathbf{X}}_i \mathbf{J}_{\beta_0^{(r)}}. \end{aligned}$$

It is easy to get that

$$\begin{aligned} E \left\{ \tilde{D} \left(\beta_0^{(r)} \right) \right\} - D \left(\beta_0^{(r)} \right) &= \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{h}_{ij} \left(\beta_0^{(r)} \right) \\ &\times f_{ij} \left(0 \right) \left\{ h^{-1} E \left[K \left(\left(Y_{ij} - \hat{g} \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) \right) / h \right) \right] - f_{ij} \left(0 \right) \right\} \mathbf{h}_{ij}^T \left(\beta_0^{(r)} \right), \end{aligned}$$

where \mathbf{h}_{ij} is given in the proof of Lemma 3. Because

$$\begin{aligned} &\left| h^{-1} E \left[K \left(\left(Y_{ij} - \hat{g} \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) \right) / h \right) \right] - f_{ij} \left(0 \right) \right| \\ &= \left| h^{-1} \int_{-\infty}^{+\infty} K \left(\left(\epsilon + g_0 \left(\mathbf{X}_{ij}^T \beta_0 \right) - \hat{g} \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) \right) / h \right) f_{ij} \left(\epsilon \right) d\epsilon - f_{ij} \left(0 \right) \right| \\ &= \left| \int_{-\infty}^{+\infty} K \left(t \right) f_{ij} \left(\hat{g} \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) - g_0 \left(\mathbf{X}_{ij}^T \beta_0 \right) + ht \right) dt - f_{ij} \left(0 \right) \right| \\ &= \left| \int_{-\infty}^{+\infty} K \left(t \right) \left\{ f_{ij} \left(\hat{g} \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) - g_0 \left(\mathbf{X}_{ij}^T \beta_0 \right) \right) + ht f'_{ij} \left(\varsigma_t \right) \right\} dt - f_{ij} \left(0 \right) \right| \\ &\leq \left| f_{ij} \left(\hat{g} \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) - g_0 \left(\mathbf{X}_{ij}^T \beta_0 \right) \right) - f_{ij} \left(0 \right) \right| + h \int_{-\infty}^{+\infty} \left| K \left(t \right) t f'_{ij} \left(\varsigma_t \right) \right| dt, \end{aligned}$$

where ς_t is between $\hat{g} \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) - g_0 \left(\mathbf{X}_{ij}^T \beta_0 \right)$ and $ht + \hat{g} \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) - g_0 \left(\mathbf{X}_{ij}^T \beta_0 \right)$. By condition (C4), $f'_{ij} \left(\cdot \right)$ is uniformly bounded; hence, there exists a constant M satisfying $\left| f'_{ij} \left(\varsigma_t \right) \right| \leq M$. Combining $\left| \hat{g} \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) - g_0 \left(\mathbf{X}_{ij}^T \beta_0 \right) \right| = O_p \left(\sqrt{N_n/n} + N_n^{-r} \right)$ with conditions (C4), (C6), and (C7), we have

$$\begin{aligned} &\left| h^{-1} E \left[K \left(\left(Y_{ij} - \hat{g} \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) \right) / h \right) \right] - f_{ij} \left(0 \right) \right| \\ &\leq \left| f_{ij} \left(\hat{g} \left(\mathbf{X}_{ij}^T \beta_0; \beta_0 \right) - g_0 \left(\mathbf{X}_{ij}^T \beta_0 \right) \right) - f_{ij} \left(0 \right) \right| + hC \rightarrow 0. \end{aligned}$$

So we can obtain $\left|n^{-1} \left\{E \left\{\tilde{D}\left(\beta_0^{(r)}\right)\right\}-D\left(\beta_0^{(r)}\right)\right\}\right| \rightarrow 0$. By the strong law of large number, we have $n^{-1} \tilde{D}\left(\beta_0^{(r)}\right) \rightarrow E\left(n^{-1} \tilde{D}\left(\beta_0^{(r)}\right)\right)$. Using the triangle inequality, we have

$$\begin{aligned} & \left|n^{-1} \left\{\tilde{D}\left(\beta_0^{(r)}\right)-D\left(\beta_0^{(r)}\right)\right\}\right| \\ & \leq \left|n^{-1} \left\{\tilde{D}\left(\beta_0^{(r)}\right)-E\left(\tilde{D}\left(\beta_0^{(r)}\right)\right)\right\}\right|+\left|n^{-1} \left\{E\left(\tilde{D}\left(\beta_0^{(r)}\right)\right)-D\left(\beta_0^{(r)}\right)\right\}\right| \rightarrow 0 . \end{aligned}$$

Furthermore, by the Taylor series expansion of $\tilde{R}\left(\beta^{(r)}\right)$ around $\beta_0^{(r)}$, we have

$$\tilde{R}\left(\beta^{(r)}\right)-\tilde{R}\left(\beta_0^{(r)}\right)=\tilde{D}\left(\beta^*\right)\left(\beta^{(r)}-\beta_0^{(r)}\right),$$

where β^* lies between $\beta^{(r)}$ and $\beta_0^{(r)}$. Let $\beta^{(r)}=\tilde{\beta}^{(r)}$, we have

$$\sqrt{n}\left(\tilde{\beta}^{(r)}-\beta_0^{(r)}\right)=-\left(\frac{1}{n} \tilde{D}\left(\beta^*\right)\right)^{-1} n^{-1 / 2} \tilde{R}\left(\beta_0^{(r)}\right)$$

because of $\tilde{R}\left(\tilde{\beta}^{(r)}\right)=0$. Since $\tilde{\beta}^{(r)} \rightarrow \beta_0^{(r)}$, we can obtain $\beta^* \rightarrow \beta_0^{(r)}$ and $\tilde{D}^{-1}\left(\beta^*\right) \rightarrow \tilde{D}^{-1}\left(\beta_0^{(r)}\right)$. Since $n^{-1}\left\{\tilde{D}\left(\beta_0^{(r)}\right)-D\left(\beta_0^{(r)}\right)\right\} \xrightarrow{p} 0$ and $n^{-1 / 2} \tilde{R}\left(\beta_0^{(r)}\right)=n^{-1 / 2} R\left(\beta_0^{(r)}\right)+o_p(1)$, we have

$$\sqrt{n}\left(\tilde{\beta}^{(r)}-\beta_0^{(r)}\right)=-\left(\frac{1}{n} D\left(\beta_0^{(r)}\right)\right)^{-1} n^{-1 / 2} R\left(\beta_0^{(r)}\right)+o_p(1) .$$

Next, similar to the proof of Lemma 3, we can complete the proof of Theorem 1. \square

Proof of Theorem 2 Since $\left\|\tilde{\beta}-\beta_0\right\|=O_p\left(n^{-1 / 2}\right)$, Theorem 2 (i) follows from this result and Lemma 1. Based on Lemma 2, we have

$$\bar{\sigma}_n^{-1}(u) \mathbf{c}_n^T\left(\hat{\boldsymbol{\theta}}\left(\beta_0\right)-\boldsymbol{\theta}^0\left(\beta_0\right)\right) \xrightarrow{d} N(0,1) .$$

By the definition of $\hat{g}(u ; \beta)$ and $\check{g}(u ; \beta)$, choosing $\mathbf{c}_n=\mathbf{B}_q(u)$ yields

$$\sigma_n^{-1}(u)\left(\hat{g}(u ; \beta_0)-\check{g}(u ; \beta_0)\right) \xrightarrow{d} N(0,1) .$$

Thus, when β is a known constant β_0 or estimated to the order $O_p\left(n^{-1 / 2}\right)$, we can complete the proof of Theorem 2 (ii). \square

Proof of Theorem 3 Let $\boldsymbol{\Sigma}_{\tau i}=\left(\sigma_{i j j'}\right)_{j, j'=1}^{m_i}$ and $\hat{\boldsymbol{\Sigma}}_{\tau i}=\left(\hat{\sigma}_{i j j'}\right)_{j, j'=1}^{m_i}$ for $i=1, \dots, n$. Based on the modified Cholesky decomposition, the diagonal elements of $\hat{\boldsymbol{\Sigma}}_{\tau i}$ are $\hat{\sigma}_{i j j}=\hat{d}_{\tau, i j}^2+\sum_{k=1}^{j-1} \hat{l}_{\tau, i j k}^2 \hat{d}_{\tau, i k}^2$ for $j=1, \dots, m_i$, and the elements under the diagonal

are $\hat{\sigma}_{ijk} = \hat{l}_{\tau,ijk} \hat{d}_{\tau,ik}^2 + \sum_{k'=1}^{k-1} \hat{l}_{\tau,ijk'} \hat{l}_{\tau,ikk'} \hat{d}_{\tau,ik'}^2$ for $j = 2, \dots, m_i, k = 1, \dots, j - 1$. Similarly, the diagonal elements of $\Sigma_{\tau i}$ are $\sigma_{ijj} = d_{\tau,ij}^2 + \sum_{k=1}^{j-1} l_{\tau,ijk}^2 d_{\tau,ik}^2$ for $j = 1, \dots, m_i$, and the elements under the diagonal are $\sigma_{ijk} = l_{\tau,ijk} d_{\tau,ik}^2 + \sum_{k'=1}^{k-1} l_{\tau,ijk'} l_{\tau,ikk'} d_{\tau,ik'}^2$ for $j = 2, \dots, m_i, k = 1, \dots, j - 1$. Since $(\hat{\gamma}_{\tau}^T, \hat{\lambda}_{\tau}^T)^T$ are \sqrt{n} -consistent estimators, together with $\hat{d}_{\tau,ij}^2 = \exp(z_{ij}^T \hat{\lambda}_{\tau})$ and $\hat{l}_{\tau,ijk} = \mathbf{w}_{ijk}^T \hat{\gamma}_{\tau}$ for $k < j = 2, \dots, m_i$, we have

$$\hat{d}_{\tau,ij}^2 - d_{\tau,ij}^2 = O_p(n^{-1/2}), \hat{l}_{\tau,ijk} - l_{\tau,ijk} = O_p(n^{-1/2}).$$

Therefore, for $j = 1, \dots, m_i$, we have

$$\begin{aligned} \hat{\sigma}_{ijj} - \sigma_{ijj} &= (\hat{d}_{\tau,ij}^2 - d_{\tau,ij}^2) + \sum_{k=1}^{j-1} (\hat{l}_{\tau,ijk}^2 \hat{d}_{\tau,ik}^2 - l_{\tau,ijk}^2 d_{\tau,ik}^2) \\ &= (\hat{d}_{\tau,ij}^2 - d_{\tau,ij}^2) + \sum_{k=1}^{j-1} [\hat{l}_{\tau,ijk}^2 (\hat{d}_{\tau,ik}^2 - d_{\tau,ik}^2) + (\hat{l}_{\tau,ijk}^2 - l_{\tau,ijk}^2) d_{\tau,ik}^2] \\ &= O_p(n^{-1/2}), \end{aligned}$$

and

$$\begin{aligned} \hat{\sigma}_{ijk} - \sigma_{ijk} &= (\hat{l}_{\tau,ijk} - l_{\tau,ijk}) \hat{d}_{\tau,ik}^2 + l_{\tau,ijk} (\hat{d}_{\tau,ik}^2 - d_{\tau,ik}^2) \\ &\quad + \sum_{k'=1}^{k-1} \left\{ \hat{l}_{\tau,ijk'} \hat{l}_{\tau,ikk'} (\hat{d}_{\tau,ik'}^2 - d_{\tau,ik'}^2) \right. \\ &\quad \left. + [(\hat{l}_{\tau,ijk'} - l_{\tau,ijk'}) \hat{l}_{\tau,ikk'} + l_{\tau,ijk'} (\hat{l}_{\tau,ikk'} - l_{\tau,ikk'})] d_{\tau,ik'}^2 \right\} \\ &= O_p(n^{-1/2}) \end{aligned}$$

for $j = 2, \dots, m_i, k = 1, \dots, j - 1$. This completes the proof. □

Proof of Theorem 4 Similar to the proof of Lemma 3, we have

$$\sqrt{n} (\bar{\beta}^{(r)} - \beta_0^{(r)}) = -(n^{-1} \bar{D}(\beta_0^{(r)}))^{-1} n^{-1/2} \bar{R}(\beta_0^{(r)}) + o_p(1), \tag{22}$$

where $\bar{S}_i(\beta) = (\bar{S}_{i1}(\beta), \dots, \bar{S}_{im_i}(\beta))^T$ with $\bar{S}_{ij}(\beta) = \psi_{h\tau}(Y_{ij} - \bar{g}(X_{ij}^T \beta; \beta))$, $\bar{R}(\beta_0^{(r)}) = \sum_{i=1}^n \mathbf{J}_{\beta_0^{(r)}}^T \hat{X}_i^T \bar{G}'(X_i \beta; \beta) \Lambda_i \hat{\Sigma}_{\tau i}^{-1} \bar{S}_i(\beta) \Big|_{\beta_0^{(r)}}$ and

$$\bar{D}(\beta_0^{(r)}) = \sum_{i=1}^n \mathbf{J}_{\beta_0^{(r)}}^T \hat{X}_i^T \bar{G}'(X_i \beta; \beta) \Lambda_i \hat{\Sigma}_{\tau i}^{-1} \bar{\Lambda}_i(\beta) \bar{G}'(X_i \beta; \beta) \hat{X}_i \mathbf{J}_{\beta_0^{(r)}} \Big|_{\beta_0^{(r)}}.$$

Using conditions (C4), (C6), and (C7), similar to Lemma 3 (k) of Horowitz (1998), we obtain

$$n^{-1/2} \bar{R} \left(\beta_0^{(r)} \right) = n^{-1/2} R^* \left(\beta_0^{(r)} \right) + o_p(1), \tag{23}$$

where $S_i^*(\beta) = \left(S_{i1}^*(\beta), \dots, S_{im_i}^*(\beta) \right)^T$ with $S_{ij}^*(\beta) = \psi_\tau \left(Y_{ij} - \bar{g} \left(X_{ij}^T \beta; \beta \right) \right)$ and $R^* \left(\beta_0^{(r)} \right) = \sum_{i=1}^n J_{\beta^{(r)}}^T \hat{X}_i^T \bar{G}' \left(X_i \beta; \beta \right) \Lambda_i \hat{\Sigma}_{\tau i}^{-1} S_i^* \left(\beta \right) \Big|_{\beta^{(r)} = \beta_0^{(r)}}$. Similar to the proof of Theorem 1, we have

$$n^{-1} \left\{ \bar{D} \left(\beta_0^{(r)} \right) - D^* \left(\beta_0^{(r)} \right) \right\} \xrightarrow{p} 0, \tag{24}$$

where $D^* \left(\beta_0^{(r)} \right) = \sum_{i=1}^n J_{\beta^{(r)}}^T \hat{X}_i^T \bar{G}' \left(X_i \beta; \beta \right) \Lambda_i \hat{\Sigma}_{\tau i}^{-1} \Lambda_i \bar{G}' \left(X_i \beta; \beta \right) \hat{X}_i J_{\beta^{(r)}} \Big|_{\beta^{(r)} = \beta_0^{(r)}}$. By (22)–(24), we have

$$\sqrt{n} \left(\bar{\beta}^{(r)} - \beta_0^{(r)} \right) = - \left(n^{-1} D^* \left(\beta_0^{(r)} \right) \right)^{-1} n^{-1/2} R^* \left(\beta_0^{(r)} \right) + o_p(1). \tag{25}$$

Similar to the proof of Lemma 1, we have

$$\begin{aligned} & \left| \bar{g}(u; \beta_0) - g_0(u) \right| \\ & = O_p \left(\sqrt{N_n/n} + N_n^{-d} \right), \quad \left| \bar{g}'(u; \beta_0) - g_0'(u) \right| = O_p \left(\sqrt{N_n^3/n} + N_n^{-d+1} \right) \end{aligned} \tag{26}$$

uniformly for any $u \in [a, b]$. Because $S_i^*(\beta_0)$ are the independent random variables, together with (26), we have $E \left(S_i^*(\beta_0) \right) = o(1)$ and

$$\begin{aligned} \text{Var} \left(\frac{1}{\sqrt{n}} R^* \left(\beta_0^{(r)} \right) \right) & = \frac{1}{n} \sum_{i=1}^n J_{\beta^{(r)}}^T \hat{X}_i^T \bar{G}' \left(X_i \beta; \beta \right) \Lambda_i \hat{\Sigma}_{\tau i}^{-1} \Sigma_{\tau i} \hat{\Sigma}_{\tau i}^{-1} \Lambda_i \bar{G}' \left(X_i \beta; \beta \right) \hat{X}_i J_{\beta^{(r)}} \Big|_{\beta^{(r)} = \beta_0^{(r)}} \left(1 + o(1) \right). \end{aligned}$$

By the use of the following property (see Lemma 2 in Li 2011), let A_n be a sequence of random matrices converging to an invertible matrix A , and then $A_n^{-1} = A^{-1} - A^{-1} (A_n - A) A^{-1} + O_p \left(\|A_n - A\|^2 \right)$. This together with Theorem 3, we have $\hat{\Sigma}_{\tau i}^{-1} - \Sigma_{\tau i}^{-1} = O_p \left(n^{-1/2} \right)$ uniformly for all i . By the law of large numbers, (26), and the consistency of $\hat{\Sigma}_{\tau i}^{-1}$, we have

$$n^{-1} D^* \left(\beta_0^{(r)} \right) \xrightarrow{p} \Gamma, \quad \text{Var} \left(n^{-1/2} R^* \left(\beta_0^{(r)} \right) \right) \xrightarrow{p} \Gamma.$$

By the multivariate central limit theorem and the Slutsky’s theorem, together with (25), we can complete the proof. □

Proof of Theorem 5 Similar to the proof of Theorem 2, together with the consistency $\hat{\Sigma}_{\tau_i}^{-1} - \Sigma_{\tau_i}^{-1} = O_p(n^{-1/2})$ of $\Sigma_{\tau_i}^{-1}$, when β is a known constant β_0 or estimated to the order $O_p(n^{-1/2})$, we can complete the proof of Theorem 5. \square

References

- Cui, X., Härdle, W. K., Zhu, L. (2011). The EFM approach for single-index models. *The Annals of Statistics*, 39, 1658–688.
- de Boor, C. (2001). *A practical guide to splines*. New York: Springer.
- Guo, C., Yang, H., Lv, J., Wu, J. (2016). Joint estimation for single index mean-covariance models with longitudinal data. *Journal of the Korean Statistical Society*, 45, 526–543.
- Horowitz, J. L. (1998). Bootstrap methods for median regression models. *Econometrica*, 66, 1327–1351.
- Jung, S. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association*, 91, 251–257.
- Lai, P., Wang, Q., Lian, H. (2012). Bias-corrected GEE estimation and smooth-threshold GEE variable selection for single-index models with clustered data. *Journal of Multivariate Analysis*, 105, 422–432.
- Leng, C., Zhang, W., Pan, J. (2010). Semiparametric mean-covariance regression analysis for longitudinal data. *Journal of the American Statistical Association*, 105, 181–193.
- Li, Y. (2011). Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. *Biometrika*, 98(2), 355–370.
- Liang, K. Y., Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22.
- Lin, H., Zhang, R., Shi, J., Liu, J., Liu, Y. (2016). A new local estimation method for single index models for longitudinal data. *Journal of Nonparametric Statistics*, 28, 644–658.
- Liu, S., Li, G. (2015). Varying-coefficient mean-covariance regression analysis for longitudinal data. *Journal of Statistical Planning and Inference*, 160, 89–106.
- Liu, X., Zhang, W. (2013). A moving average Cholesky factor model in joint mean-covariance modeling for longitudinal data. *Science China Mathematics*, 56, 2367–2379.
- Lv, J., Guo, C., Yang, H., Li, Y. (2017). A moving average Cholesky factor model in covariance modeling for composite quantile regression with longitudinal data. *Computational Statistics and Data Analysis*, 112, 129–144.
- Ma, S., He, X. (2016). Inference for single-index quantile regression models with profile optimization. *The Annals of Statistics*, 44, 1234–1268.
- Ma, S., Song, P. X.-K. (2015). Varying index coefficient models. *Journal of the American Statistical Association*, 110, 341–356.
- Mao, J., Zhu, Z., Fung, W. K. (2011). Joint estimation of mean-covariance model for longitudinal data with basis function approximations. *Computational Statistics and Data Analysis*, 55, 983–992.
- Pollard, D. (1990). *Empirical processes: Theories and applications*. Hayward, CA: Institute of Mathematical Statistics.
- Wang, H., Zhu, Z. (2011). Empirical likelihood for quantile regression models with longitudinal data. *Journal of Statistical Planning and Inference*, 141, 1603–1615.
- Xu, P., Zhu, L. (2012). Estimation for a marginal generalized single-index longitudinal model. *Journal of Multivariate Analysis*, 105, 285–299.
- Yao, W., Li, R. (2013). New local estimation procedure for a non-parametric regression function for longitudinal data. *Journal of the Royal Statistical Society: Series B*, 75, 123–138.
- Ye, H., Pan, J. (2006). Modelling of covariance structures in generalised estimating equations for longitudinal data. *Biometrika*, 93, 927–941.
- Zhang, D., Lin, X., Raz, J., Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, 93, 710–719.
- Zhang, W., Leng, C. (2012). A moving average Cholesky factor model in covariance modeling for longitudinal data. *Biometrika*, 99, 141–150.
- Zhao, W., Lian, H., Liang, H. (2017). GEE analysis for longitudinal single-index quantile regression. *Journal of Statistical Planning and Inference*, 187, 78–102.
- Zheng, X., Fung, W. K., Zhu, Z. (2014). Variable selection in robust joint mean and covariance model for longitudinal data analysis. *Statistica Sinica*, 24, 515–531.