Journal of Multivariate Analysis 118 (2013) 115–127

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Quadratic inference functions for partially linear single-index models with longitudinal data

Peng Lai [a,*], Gaorong Li [b], Heng Lian [c]

[a] School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing 210044, China
[b] College of Applied Sciences, Beijing University of Technology, Beijing 100124, China
[c] Division of Mathematical Sciences, SPMS, Nanyang Technological University, Singapore, 637371, Singapore

ARTICLE INFO

Article history:
Received 22 July 2012
Available online 6 April 2013

AMS 2000 subject classifications:
62G05
62G10
62G20

Keywords:
Bias correction
Generalized likelihood ratio
Longitudinal data
Partially linear single-index models
QIF

ABSTRACT

In this paper, we consider the partially linear single-index models with longitudinal data. We propose the bias-corrected quadratic inference function (QIF) method to estimate the parameters in the model by accounting for the within-subject correlation. Asymptotic properties for the proposed estimation methods are demonstrated. A generalized likelihood ratio test is established to test the linearity of the nonparametric part. Under the null hypotheses, the test statistic follows asymptotically a χ^2 distribution. We also evaluate the finite sample performance of the proposed methods via Monte Carlo simulation studies and a real data analysis.

1. Introduction

Partially linear single-index (PLSI) models describe both the linear relationship between a scalar response variable Y and a q-dimensional vector Z and the nonlinear relationship between Y and a p-dimensional vector X in the form

$$Y = Z^\top\theta + g(X^\top\beta) + \varepsilon, \tag{1.1}$$

where $g(\cdot)$ is an unknown link function and $\|\beta\| = 1$ ($\|\cdot\|$ denotes the Euclidean norm here).

Model (1.1) has gained much attention in recent years. For example, Carroll et al. [3] studied the generalized partially linear single-index models, Xia and Härdle [21] proposed the MAVE method to estimate the parameters of PLSI models, Zhu and Xue [28] studied the confidence interval of parameters based on the empirical likelihood method. PLSI models avoid the problem of "curse of dimensionality" and are flexible enough to capture the hidden linear and nonlinear relationships between covariates and the response variable. More recently, the literature on the applications of partially linear single-index models for repeated measurements is available, especially in econometrics, biomedical research, and epidemiology. For example, Tian et al. [19] used the generalized penalized spline least squares method and assumed working correlation matrices to estimate the parameters and the unknown link function. Li et al. [13] constructed confidence intervals or confidence regions for PLSI models with longitudinal data based on empirical likelihood.

* Corresponding author.
E-mail addresses: laipeng@amss.ac.cn, laipengnuist@163.com (P. Lai).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.03.019

Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im_i})^\top$, $1 \le i \le n$, be the vector of outcome values, which is assumed to be related to an $m_i \times p$ covariate matrix $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{im_i})^\top$ and an $m_i \times q$ covariate matrix $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{im_i})^\top$. Define $N = \sum_{i=1}^n m_i$, while $\{m_i\}$ are the bounded sequences of positive integers. This means that n and N have the same order. The partially linear single-index models for longitudinal data take the form

$$Y_{ij} = g(X_{ij}^\top \beta) + Z_{ij}^\top \theta + \varepsilon_{ij}, \quad i = 1, \ldots, n, \; j = 1, \ldots, m_i, \tag{1.2}$$

where $(\beta^\top, \theta^\top)^\top$ is an unknown vector in $\mathbb{R}^p \times \mathbb{R}^q$ with $\|\beta\| = 1$, $g(\cdot)$ is an unknown univariate link function. The random error vector of the ith subject is $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{im_i})^\top$, and $\{\boldsymbol{\varepsilon}_i,\ i = 1, \ldots, n\}$ are mutually independent with $E(\boldsymbol{\varepsilon}_i|\boldsymbol{X}_i, \boldsymbol{Z}_i) = 0$, $\mathrm{Var}(\boldsymbol{\varepsilon}_i) = \Sigma_i$. Denote the true parameters as β_0 and θ_0.

Model (1.2) is flexible enough to cover many important statistical models. For example, when $\theta = 0$, model (1.2) is a longitudinal single-index model. When $p = 1$ and $\beta = 1$, model (1.2) becomes the longitudinal partially linear model. They have been investigated in [25,16,6,23,22,5,12,1,11]. The generalized estimating equations (GEEs) could be used to estimate the parameters. The estimates of regression parameters are consistent even when the working correlation matrices are misspecified. However, the GEE approach requires a consistent estimator of the working correlation matrix for efficient estimation. To overcome the problem, Qu et al. [18] introduced the quadratic inference functions (QIFs) by representing the inverse of the working correlation matrix as a linear combination of basis matrices, that is

$$R^{-1} \approx a_1 M_1 + a_2 M_2 + \cdots + a_k M_k, \tag{1.3}$$

where M_1 is the identity matrix, and $M_2, \ldots, M_k$ are symmetric basis matrices which are determined by the structure of R, and $a_1, \ldots, a_k$ are constant coefficients. The QIF method is a strong competitor to the GEE approach in analyzing longitudinal data. The advantage of this approach is that it does not require estimations of linear coefficients a_i's which can be viewed as nuisance parameters. Qu and Li [17] applied the QIF method to varying coefficient models for longitudinal data. Bai et al. [2, 1] applied the QIF method to partially linear models and single-index models with longitudinal data.

The rest of the paper is organized as follows. In Section 2, we propose the bias-corrected QIF procedure for the partially linear single-index models with longitudinal data and show its asymptotic properties. We propose a test for testing whether the nonparametric part of PLSI models is actually linear in Section 3. In Section 4, we present the results from some simulation studies, and a real dataset is analyzed to illustrate the proposed methods. The proofs of the asymptotic properties are presented in the Appendix.

2. Bias-corrected QIF estimation

For the sake of identifiability, it is assumed that $\|\beta\| = 1$ in model (1.2). For the convenience of computing the estimated parameter, we use the delete-one-component method proposed by Yu and Ruppert [24]. Without loss of generality, we assume that the parameter β has a positive component β_r (otherwise, consider $-\beta$). Let $\beta = (\beta_1, \ldots, \beta_p)^\top$, and let $\beta^{(r)} = (\beta_1, \ldots, \beta_{r-1}, \beta_{r+1}, \ldots, \beta_p)^\top$ be a $(p-1)$-dimensional vector after removing the rth component β_r of β. Then, we may write β as a function of $\beta^{(r)}$

$$\beta(\beta^{(r)}) = (\beta_1, \ldots, \beta_{r-1}, (1 - \|\beta^{(r)}\|^2)^{1/2}, \beta_{r+1}, \ldots, \beta_p)^\top.$$

The true parameter $\beta_0^{(r)}$ must satisfy the constraint $\|\beta_0^{(r)}\| < 1$. β is infinitely differentiable in a neighborhood of $\beta_0^{(r)}$, and the Jacobian matrix is

$$J_{\beta^{(r)}} = \frac{\partial \beta}{\partial \beta^{(r)}} = (b_1, \ldots, b_p)^\top,$$

where $b_s (1 \le s \le p, s \ne r)$ is a $(p-1)$-dimensional unit vector with sth component 1, and $b_r = -(1 - \|\beta^{(r)}\|^2)^{-1/2} \beta^{(r)}$.

Generalized estimating equations (GEEs) are often used in semiparametric models; see [15,13]. Let

$$\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{im_i})^\top, \qquad \boldsymbol{Z}_i = (Z_{i1}, Z_{i2}, \ldots, Z_{im_i})^\top,$$
$$\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{im_i})^\top, \qquad G(\boldsymbol{X}_i \beta) = (g(X_{i1}^\top \beta), g(X_{i2}^\top \beta), \ldots, g(X_{im_i}^\top \beta))^\top,$$

and $g_1(t) = E[X_{ij}|X_{ij}^\top \beta = t]$ and $g_2(t) = E[Z_{ij}|X_{ij}^\top \beta = t]$. We construct the bias-corrected GEE as follows

$$\sum_{i=1}^n \Lambda_i V_i^{-1} [\boldsymbol{Y}_i - G(\boldsymbol{X}_i \beta) - \boldsymbol{Z}_i \theta] = 0, \tag{2.1}$$

where

$$\Lambda_i = \begin{pmatrix} g'(X_{i1}^\top \beta)(X_{i1} - g_1(X_{i1}^\top \beta))^\top J_{\beta^{(r)}} & (Z_{i1} - g_2(X_{i1}^\top \beta))^\top \\ g'(X_{i2}^\top \beta)(X_{i2} - g_1(X_{i2}^\top \beta))^\top J_{\beta^{(r)}} & (Z_{i2} - g_2(X_{i2}^\top \beta))^\top \\ \vdots & \vdots \\ g'(X_{im_i}^\top \beta)(X_{im_i} - g_1(X_{im_i}^\top \beta))^\top J_{\beta^{(r)}} & (Z_{im_i} - g_2(X_{im_i}^\top \beta))^\top \end{pmatrix}^\top$$

and $g'(\cdot)$ is the derivative of $g(\cdot)$. Here g_1 and g_2 are used for bias correction in the estimating equations [28,13,11]. For the generalized estimating Eq. (2.1), Liang and Zeger [15] assumed $V_i = A_i^{1/2}R(\alpha)A_i^{1/2}$, where A_i is an $m_i \times m_i$ diagonal matrix of marginal variances and $R(\alpha)$ is a common working correlation matrix which involves a small number of nuisance parameters α. However, it is difficult to estimate V_i. If the working correlations are misspecified, the estimator from (2.1) is still consistent but is not optimal. Qu et al. [18] introduced the quadratic inference function (QIF) by representing

$$R^{-1} \approx a_1M_1 + a_2M_2 + \cdots + a_kM_k.$$

In practice, we need to choose the basis for the inverse of the correlation matrix R. Suppose that R is the exchangeable working correlation matrix, then $R^{-1} = a_1M_1 + a_2M_2$, where M_1 is the identity matrix and M_2 is a matrix with 0 on the diagonal and 1 off-diagonal. If the working correlation matrix is AR(1) with $R_{ij} = \alpha^{|i-j|}$, then $R^{-1} = a_1M_1^* + a_2M_2^* + a_3M_3^*$, where M_1^* is the identity matrix, M_2^* has 1 on the sub-diagonal and 0 elsewhere, and M_3^* has 1 on the corners (1, 1) and (m, m) and 0 elsewhere. However, in practice M_3^* can often be dropped out of the model, as removing M_3^* does not affect the efficiency of the estimator too much, and this can simplify the estimation procedure. More details can be found in [18,17].

The bias-corrected GEE (2.1) is a linear combination of elements of the estimating functions

$$\sum_{i=1}^n U_i(\beta^{(r)}, \theta) = \begin{pmatrix} \sum_{i=1}^n \Lambda_i A_i^{-1/2} M_1 A_i^{-1/2}[\mathbf{Y}_i - G(\mathbf{X}_i\beta) - \mathbf{Z}_i\theta] \\ \sum_{i=1}^n \Lambda_i A_i^{-1/2} M_2 A_i^{-1/2}[\mathbf{Y}_i - G(\mathbf{X}_i\beta) - \mathbf{Z}_i\theta] \\ \vdots \\ \sum_{i=1}^n \Lambda_i A_i^{-1/2} M_k A_i^{-1/2}[\mathbf{Y}_i - G(\mathbf{X}_i\beta) - \mathbf{Z}_i\theta] \end{pmatrix}. \tag{2.2}$$

Since the dimension of $U_i(\beta^{(r)}, \theta)$ is $k(p - 1 + q)$, greater than the number of unknown parameters, the bias-corrected GEE method cannot be applied. Qu et al. [18] proposed the quadratic inference functions by extending the method of GMM proposed by Hansen [8]. We follow the idea of QIF and estimate $\beta^{(r)}$ and θ by minimizing the following bias-corrected QIF

$$Q_n(\beta^{(r)}, \theta) = \bar{U}_n^\top(\beta^{(r)}, \theta) C_n^{-1}(\beta^{(r)}, \theta) \bar{U}_n(\beta^{(r)}, \theta), \tag{2.3}$$

where $\bar{U}_n(\beta^{(r)}, \theta) = \frac{1}{n}\sum_{i=1}^n U_i(\beta^{(r)}, \theta)$ and $C_n = \frac{1}{n}\sum_{i=1}^n U_i(\beta^{(r)}, \theta)U_i^\top(\beta^{(r)}, \theta)$.

The bias-corrected QIF (2.3) contains the unknown functions $g(\cdot), g'(\cdot), g_1(\cdot)$ and $g_2(\cdot)$. To obtain the estimators for β and θ, we need to plug in the estimates for the unknown functions. The local linear smoothing method [4] is used here. We can obtain $\hat{g}(t; \beta, \theta)$ and $\hat{g}'(t; \beta, \theta)$ by minimizing

$$\min_{a,b} \sum_{i=1}^n \sum_{j=1}^{m_i} (Y_{ij} - Z_{ij}^\top\theta - a - b(X_{ij}^\top\beta - t))^2 K_h(X_{ij}^\top\beta - t), \tag{2.4}$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ is a kernel function, and the bandwidth $h = h_n$ is tending to zero. Let $(\hat{a}, \hat{b})$ be the solution to the weighted least squares problem (2.4). We set $\hat{g}(t; \beta, \theta) = \hat{a}$ and $\hat{g}'(t; \beta, \theta) = \hat{b}$. Simple calculations yield

$$\hat{g}(t; \beta, \theta) = \sum_{i=1}^n \sum_{j=1}^{m_i} W_{nij}(t; \beta)(Y_{ij} - Z_{ij}^\top\theta), \tag{2.5}$$

and

$$\hat{g}'(t; \beta, \theta) = \sum_{i=1}^n \sum_{j=1}^{m_i} \widetilde{W}_{nij}(t; \beta)(Y_{ij} - Z_{ij}^\top\theta), \tag{2.6}$$

where

$$W_{nij}(t; \beta) = \frac{N^{-1}K_h(X_{ij}^\top\beta - t)[S_{n,2}(t; \beta) - (X_{ij}^\top\beta - t)S_{n,1}(t; \beta)]}{S_{n,0}(t; \beta)S_{n,2}(t; \beta) - S_{n,1}^2(t; \beta)},$$

$$\widetilde{W}_{nij}(t; \beta) = \frac{N^{-1}K_h(X_{ij}^\top\beta - t)[(X_{ij}^\top\beta - t)S_{n,0}(t; \beta) - S_{n,1}(t; \beta)]}{S_{n,0}(t; \beta)S_{n,2}(t; \beta) - S_{n,1}^2(t; \beta)},$$

and

$$S_{n,l}(t; \beta) = \frac{1}{N}\sum_{i=1}^n \sum_{j=1}^{m_i} K_h(X_{ij}^\top\beta - t)(X_{ij}^\top\beta - t)^l, \quad l = 0, 1, 2.$$

Given β, we can also obtain the estimators of $g_1(t) = E[X_{ij}|X_{ij}^\top\beta = t]$ and $g_2(t) = E[Z_{ij}|X_{ij}^\top\beta = t]$, respectively, as

$$\hat{g}_1(t;\beta) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} W_{nij}(t;\beta)X_{ij}, \tag{2.7}$$

and

$$\hat{g}_2(t;\beta) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} W_{nij}(t;\beta)Z_{ij}. \tag{2.8}$$

After plugging in the estimators of $g(\cdot)$, $g'(\cdot)$, $g_1(\cdot)$ and $g_2(\cdot)$, we get

$$\hat{\Lambda}_i = \begin{pmatrix} \hat{g}'(X_{i1}^\top\beta;\beta,\theta)(X_{i1} - \hat{g}_1(X_{i1}^\top\beta;\beta))^\top J_{\beta^{(r)}} & (Z_{i1} - \hat{g}_2(X_{i1}^\top\beta;\beta))^\top \\ \hat{g}'(X_{i2}^\top\beta;\beta,\theta)(X_{i2} - \hat{g}_1(X_{i2}^\top\beta;\beta))^\top J_{\beta^{(r)}} & (Z_{i2} - \hat{g}_2(X_{i2}^\top\beta;\beta))^\top \\ \vdots & \vdots \\ \hat{g}'(X_{im_i}^\top\beta;\beta,\theta)(X_{im_i} - \hat{g}_1(X_{im_i}^\top\beta;\beta))^\top J_{\beta^{(r)}} & (Z_{im_i} - \hat{g}_2(X_{im_i}^\top\beta;\beta))^\top \end{pmatrix}^\top,$$

$$\hat{U}_i(\beta^{(r)},\theta) = \begin{pmatrix} \hat{\Lambda}_i A_i^{-1/2} M_1 A_i^{-1/2}[\mathbf{Y}_i - \hat{G}(\mathbf{X}_i\beta;\beta,\theta) - \mathbf{Z}_i\theta] \\ \hat{\Lambda}_i A_i^{-1/2} M_2 A_i^{-1/2}[\mathbf{Y}_i - \hat{G}(\mathbf{X}_i\beta;\beta,\theta) - \mathbf{Z}_i\theta] \\ \vdots \\ \hat{\Lambda}_i A_i^{-1/2} M_k A_i^{-1/2}[\mathbf{Y}_i - \hat{G}(\mathbf{X}_i\beta;\beta,\theta) - \mathbf{Z}_i\theta] \end{pmatrix}$$

and $\bar{\hat{U}}_n(\beta^{(r)},\theta) = \frac{1}{n}\sum_{i=1}^{n}\hat{U}_i(\beta^{(r)},\theta)$, $\hat{C}_n(\beta^{(r)},\theta) = \frac{1}{n}\sum_{i=1}^{n}\hat{U}_i(\beta^{(r)},\theta)\hat{U}_i^\top(\beta^{(r)},\theta)$. Therefore, we can obtain the bias-corrected QIF estimators $\hat{\beta}^{(r)}$ and $\hat{\theta}$ by minimizing

$$\hat{Q}_n(\beta^{(r)},\theta) = \bar{\hat{U}}_n^\top(\beta^{(r)},\theta)\hat{C}_n^{-1}(\beta^{(r)},\theta)\bar{\hat{U}}_n(\beta^{(r)},\theta). \tag{2.9}$$

It is worth pointing out that the objective function defined in (2.9) contains only the parameters $\beta^{(r)}$ and θ, and only the basis matrices from the working correlation structure. In other words, the bias-corrected QIF estimators, $(\hat{\beta}^{(r)\top},\hat{\theta}^\top)^\top = \arg\min_{\beta^{(r)},\theta}\hat{Q}_n(\beta^{(r)},\theta)$ are obtained with no need to estimate the nuisance correlation parameters $a_1,\ldots,a_k$.

Theorem 1. *Suppose that the technical conditions* (C1)–(C9) *hold in the Appendix, and the rth component of β is positive, then we have*

(1) *the bias-corrected QIF estimator $(\hat{\beta}^{(r)\top},\hat{\theta}^\top)^\top$ by minimizing (2.9) converges to $(\beta_0^{(r)\top},\theta_0^\top)^\top$ in probability;*
(2) *$(\hat{\beta}^{(r)\top},\hat{\theta}^\top)^\top$ is asymptotically normal*

$$\sqrt{n}\begin{pmatrix}\hat{\beta}^{(r)} - \beta_0^{(r)} \\ \hat{\theta} - \theta_0\end{pmatrix} \xrightarrow{L} N(0, (\Gamma^\top\Sigma^{-1}(\beta_0^{(r)},\theta_0)\Gamma)^{-1}),$$

where

$$\Sigma(\beta_0^{(r)},\theta_0) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E[U_i(\beta_0^{(r)},\theta_0)U_i^\top(\beta_0^{(r)},\theta_0)]$$

and

$$\Gamma = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\begin{pmatrix}\Lambda_i A_i^{-1/2}M_1A_i^{-1/2}\Lambda_i^\top \\ \Lambda_i A_i^{-1/2}M_2A_i^{-1/2}\Lambda_i^\top \\ \vdots \\ \Lambda_i A_i^{-1/2}M_kA_i^{-1/2}\Lambda_i^\top\end{pmatrix}.$$

Corollary 2.1. *Under the conditions of Theorem 1, we have*

$$\sqrt{n}\begin{pmatrix}\hat{\beta} - \beta_0 \\ \hat{\theta} - \theta_0\end{pmatrix} \xrightarrow{L} N\big(0, P_\beta(\Gamma^\top\Sigma^{-1}(\beta_0^{(r)},\theta_0)\Gamma)^{-1}P_\beta^\top\big),$$

where $P_\beta = \begin{pmatrix} J_{\beta_0^{(r)}} & \mathbf{0} \\ \mathbf{0} & I_q \end{pmatrix}$.

3. Hypothesis tests

In practice, parsimonious models are always desirable to enhance model predictability. For partially linear single-index models, it is interesting to test whether the single-index part can be replaced by a linear expression of the index function. That means we are interested in testing

$$H_0 : g(t) = \alpha_0 + \alpha_1 t \leftrightarrow H_1 : g(t) \neq \alpha_0 + \alpha_1 t, \tag{3.1}$$

where α_0 and α_1 are two unknown parameters. That is

$H_0 : g(t)$ is a linear function $\leftrightarrow H_1 : g(t)$ is a nonlinear function.

We extend the generalized likelihood ratio (GLR) tests to the testing problem (3.1). The GLR tests are proposed by Fan et al. [7] for inferences of nonparametric models. Zhang [26,27] used the GLR tests to test the nonparametric parts on partially linear models and partially linear single-index models, respectively. Here we consider the case for longitudinal data. Noting that under H_1, $g(\cdot)$ is a nonlinear function. Let $\hat{\beta}$, $\hat{\theta}$ and $\hat{g}(\cdot)$ be the estimators of β_0, θ_0 and $g(\cdot)$ based on the bias-corrected QIF. Under H_0, the data $\{Y_{ij}, X_{ij}, Z_{ij},\ i = 1, \ldots, n;\ j = 1, \ldots, m_i\}$ are from the linear model

$$Y_{ij} = \alpha_0 + \alpha_1 X_{ij}^\top \beta + Z_{ij}^\top \theta + \varepsilon_{ij}, \quad \text{with } \|\beta\| = 1.$$

Therefore, we can construct the following generalized likelihood ratio (GLR) statistic

$$T_{\mathrm{GLR}} = \frac{N}{2} \frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}, \tag{3.2}$$

where

$$\mathrm{RSS}_0 = \sum_{i=1}^{n} \sum_{j=1}^{m_i} (Y_{ij} - \tilde{\alpha}_0 - \tilde{\alpha}_1 \tilde{\beta}^\top X_{ij} - \tilde{\theta}^\top Z_{ij})^2$$

is the residual sums of squares under the null hypothesis, $\tilde{\alpha}_0$, $\tilde{\alpha}_1$, $\tilde{\beta}$ and $\tilde{\theta}$ are the QIF estimators of α_0, α_1, β_0 and θ_0 respectively, under H_0 and

$$\mathrm{RSS}_1 = \sum_{i=1}^{n} \sum_{j=1}^{m_i} (Y_{ij} - \hat{g}(\hat{\beta}^\top X_{ij}) - \hat{\theta}^\top Z_{ij})^2$$

is the residual sums of squares under the alternative hypothesis.

Theorem 2. *Suppose the conditions* (C1)–(C9) *hold in the Appendix. Then under the null hypothesis H_0, $\tau_K T_{\mathrm{GLR}}$ has an asymptotic χ^2 distribution with degrees of freedom $\tau_K c_K |\mathcal{T}|/h$, where $\tau_K = \{K(0) - 0.5 \int K^2(u)du\}\{\int (K(u) - 0.5K * K(u))^2 du\}^{-1}$, $c_K = K(0) - 0.5 \int K^2(u)du$, and $K * K$ denotes the convolution of K, $|\mathcal{T}|$ stands for the length of $\mathcal{T}$ and $\mathcal{T}$ is defined in condition* (C1).

4. Numerical studies

4.1. Simulation studies

In this section, we present some simulation studies to evaluate the finite sample performances of the proposed estimators and the testing method. To study the finite sample properties of our proposed estimator, we compare the proposed estimators $\hat{g}(\cdot)$, $\hat{\beta}$ and $\hat{\theta}$ with the working independence estimators $\hat{g}_I(\cdot)$, $\hat{\beta}_I$ and $\hat{\theta}_I$, where $\hat{\beta}_I$ and $\hat{\theta}_I$ are the solutions of the bias-corrected GEE (2.1) with $V_i = I_{m_i},\ i = 1, \ldots, n$.

To measure the performance of the proposed estimators, the biases and standard errors of the estimators of β and θ are reported, we also define the mean square errors of the estimators of β_0, θ_0 and $g(\cdot)$ as

$$\mathrm{MSE}_{\hat{\beta}.} = \frac{1}{S} \sum_{i=1}^{S} \left(\frac{1}{p} \sum_{k=1}^{p} (\hat{\beta}_{\cdot k}^{(i)} - \beta_{0k})^2 \right), \qquad \mathrm{MSE}_{\hat{\theta}.} = \frac{1}{S} \sum_{i=1}^{S} \left(\frac{1}{q} \sum_{k=1}^{q} (\hat{\theta}_{\cdot k}^{(i)} - \theta_{0k})^2 \right)$$

and

$$\mathrm{MSE}_{\hat{g}.} = \frac{1}{S} \sum_{i=1}^{S} \left(\frac{1}{n} \sum_{j=1}^{n} \left(\frac{1}{m_j} \sum_{k=1}^{m_j} [\hat{g}_{\cdot}^{(i)}(\hat{t}_{jk}^{(i)}) - g(t_{jk}^{(i)})]^2 \right) \right), \quad \hat{t}_{jk}^{(i)} = \hat{\beta}_{\cdot}^{(i)\top} X_{jk}^{(i)},\ t_{jk}^{(i)} = \beta_0^\top X_{jk}^{(i)},$$

where $\hat{\beta}.$, $\hat{\theta}.$ and $\hat{g}.$ denote the bias-corrected QIF estimators or working independence estimators, S denotes the size of the generated datasets.

Table 1
The biases, standard errors and MSE of the proposed estimators for Example 1 with true correlation exchangeable (the values in the parentheses are the corresponding standard errors of the estimators).

n		Independence	Exchangeable QIF	AR(1) QIF
60	β_1	0.0164(0.0878)	0.0169(0.0824)	0.0171(0.0826)
	β_2	−0.0229(0.2032)	−0.0232(0.1998)	−0.0234(0.1998)
	β_3	−0.0050(0.1555)	−0.0058(0.1454)	−0.0061(0.1469)
	θ	0.0043(0.3762)	0.0034(0.2346)	−0.0038(0.2412)
	$\mathrm{MSE}_{\hat\beta}$	0.0251	0.0233	0.0234
	$\mathrm{MSE}_{\hat\theta}$	0.1414	0.0550	0.0581
	$\mathrm{MSE}_{\hat g}$	0.2254	0.1776	0.1798
120	β_1	−0.0013(0.0232)	−0.0005(0.0117)	−0.0007(0.0124)
	β_2	0.0018(0.0345)	0.0004(0.0168)	0.0008(0.0179)
	β_3	0.0003(0.0396)	0.0008(0.0187)	0.0007(0.0200)
	θ	0.0039(0.2447)	−0.0026(0.1240)	−0.0031(0.1318)
	$\mathrm{MSE}_{\hat\beta}$	0.0011	0.0003	0.0003
	$\mathrm{MSE}_{\hat\theta}$	0.0598	0.0154	0.0174
	$\mathrm{MSE}_{\hat g}$	0.0994	0.0758	0.0768

Example 1. We generate 1000 datasets, each consisting of $n = 60, 120$ subjects and $\{m_k \equiv m = 3,\ k = 1, \ldots, n\}$ observations per subject. The data are generated from the following model:

$$Y_{ik} = g(t_{ik}) + Z_{ik}^\top \theta_0 + \varepsilon_{ik}, \quad t_{ik} = X_{ik}^\top \beta_0, \tag{4.1}$$

where $\beta_0 = \frac{1}{\sqrt{14}}(3, 2, 1)^\top$, $\theta_0 = 0.5$ and $g(t) = \cos(\pi t)$, $X_{ik} = (X_{ik1}, \ldots, X_{ikp})^\top$ and Z_{ik} are generated from $N(0, I_p)$ and $U(0, 1)$, respectively, $i = 1, \ldots, n$; $k = 1, \ldots, m$. The error $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im})^\top$, $m = 3$ follows a multivariate normal distribution with mean 0 and covariance Σ_i, $\Sigma_i = \sigma_1^2\Sigma_0, \sigma_1 = 0.5$, $i = 1, \ldots, [n/3]$; $\Sigma_i = \sigma_2^2\Sigma_0, \sigma_2 = 1.0$, $i = [n/3] + 1, \ldots, [2n/3]$; $\Sigma_i = \sigma_3^2\Sigma_0, \sigma_3 = 2.0$, $i = [2n/3] + 1, \ldots, n$, $[x]$ denotes the integer part of x, Σ_0 is the correlation matrix. Here we consider two different types of correlation matrix Σ_0, one is an exchangeable correlation structure with correlation $\rho = 0.6$ and the other is an AR(1) correlation structure with $\rho = 0.5$. In order to compare the performance of the QIF estimators, the exchangeable correlation and AR(1) correlation matrices are used as the working correlation matrices $R(\rho)$ respectively in the simulation. Since A_i can be estimated by the method of moment. Thus Σ_i can be estimated by $\hat A_i^{1/2} R(\rho) \hat A_i^{1/2}$. The kernel function used here is $K(x) = \frac{3}{4}(1 - x^2)$ if $|x| \le 1$, 0 otherwise. The bandwidth is obtained through the leave-one-out cross-validation bandwidth selection method. We report the results in Tables 1–2.

Example 2. We illustrate how the bias-corrected QIF performs for testing whether the single-index part is linear. Similar to Example 1, we make some changes to $g(t)$ and simulate data from

$$Y_{ik} = \xi_0 + \xi_1 t_{ik} + c_0 * \cos(\pi t_{ik}) + Z_{ik}^\top \theta_0 + \varepsilon_{ik}, \quad t_{ik} = X_{ik}^\top \beta_0,$$

where $\xi_0 = 0.3$, $\xi_1 = 1.0$, $0 \le c_0 \le 1$. The true correlation matrix is the exchangeable correlation. To examine the power of the bias-corrected QIF approach when the index function $X_{ik}^\top \beta_0$ deviates from a linear expression, we let c_0 increase from 0 to 0.8 with a step size of 0.025. Three working correlation matrices, exchangeable, AR(1) and independent correlation are considered. To compute the power of our proposed test, we calculate test statistics $\tau_K T_{\mathrm{GLR}}$ from 1000 simulations for each c_0 with $n = 120$ and type I error 0.05, the result is shown in Fig. 1. We also provide quantile–quantile plots under exchangeable and AR(1) working correlation for the power test in Fig. 2.

From Tables 1–2, based on the reported standard errors and MSE of the proposed estimators, the bias-corrected QIF estimators have similar or better performance than the working independence estimators, especially when the working correlation matrix is specified correctly. Tables 1–2 also tell us that with the sample size increasing, the biases, standard errors and MSE of these estimators usually decrease. Fig. 1 shows the power of bias-corrected QIF versus c_0. It shows that when c_0 is small, the test sizes are approximately 0.05 and when c_0 reaches 0.8, the probability of rejection reaches 0.95 power. Fig. 1 also shows that when the assumed working correlation is correctly specified, the test power is better. From Fig. 2, we can find that under H_0 the empirical quantiles of $\tau_K T_{\mathrm{GLR}}$ follow the theoretical chi-squared quantile well.

4.2. Application to CD4 data

We now apply the proposed procedure to the CD4 data from the Multi-Center AIDS Cohort Study. This dataset consists of 283 homosexual males who were HIV positive between 1984 and 1991. All individuals were scheduled to have their measurements made during semiannual visits. But it often happened that patients missed appointments or rescheduled their appointments. Therefore, each patient had a different number of repeated measurements. Details of the study design, methods and the medical implications can be found in Kaslow et al. [10]. Huang et al. [9], Fan and Li [6], Qu and Li [17]

Table 2
The biases, standard errors and MSE of the proposed estimators for Example 1 with true correlation AR(1) (the values in the parentheses are the corresponding standard errors of the estimators).

n		Independence	Exchangeable QIF	AR(1) QIF
60	β_1	0.0007(0.0150)	0.0007(0.0146)	0.0006(0.0144)
	β_2	−0.0010(0.0456)	−0.0008(0.0454)	−0.0008(0.0455)
	β_3	−0.0001(0.0254)	−0.0002(0.0245)	−0.0001(0.0242)
	θ	−0.0049(0.1403)	−0.0037(0.1329)	−0.0034(0.1309)
	$\mathrm{MSE}_{\hat\beta}$	0.0010	0.0009	0.0010
	$\mathrm{MSE}_{\hat\theta}$	0.0197	0.0177	0.0171
	$\mathrm{MSE}_{\hat g}$	0.8166	0.8141	0.8137
120	β_1	−0.0004(0.0073)	−0.0004(0.0066)	−0.0003(0.0064)
	β_2	0.0005(0.0107)	0.0005(0.0096)	0.0005(0.0093)
	β_3	0.0007(0.0123)	0.0001(0.0111)	0.0001(0.0108)
	θ	−0.0004(0.0908)	−0.0004(0.0841)	0.0002(0.0827)
	$\mathrm{MSE}_{\hat\beta}$	0.0001	0.0001	0.0001
	$\mathrm{MSE}_{\hat\theta}$	0.0082	0.0071	0.0068
	$\mathrm{MSE}_{\hat g}$	0.0324	0.0319	0.0318

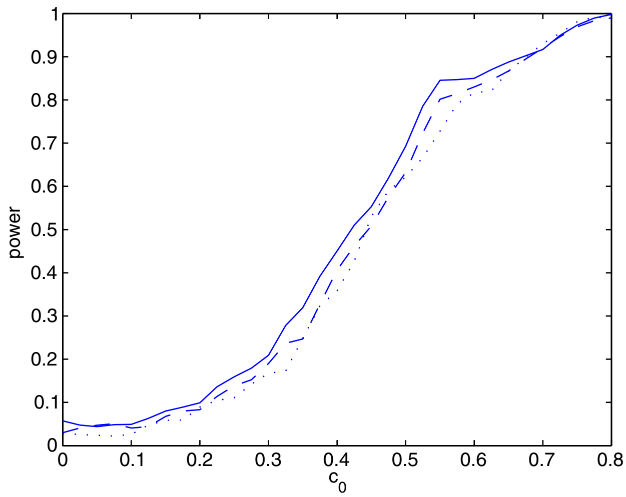

Fig. 1. Power of bias-corrected QIF against c_0 for testing whether g is a linear function. The solid line is the power using exchangeable correlation, the dash line is the power using AR(1) correlation, the dotted line is the power using independent structure. The true correlation structure is exchangeable.

Table 3
Estimates and confidence intervals of nonzero parameters.

	Estimates	Confidence interval		Estimates	Confidence interval
β_{X_1}	0.8321	[0.8296, 0.8346]	β_{X_2}	0.0567	[0.0535, 0.0600]
β_{X_3}	−0.5518	[−0.5552, −0.5483]	θ_Z	0.0453	[0.0431, 0.0474]

analyzed the same dataset using varying coefficient models. Here we apply the partially linear single-index model to this dataset.

It is known that HIV destroys CD4 cells, so by measuring CD4 cell count and percentages in blood, doctors can monitor progression of the disease. Therefore, the response variable Y is the CD4 percentage over time. Four covariates are also collected: X_1, patient's measuring time; X_2, patient's age; X_3, the CD4 cell percentage before their infection; Z, the individual's smoking status, which takes binary values 1 or 0, according to whether an individual is a smoker or nonsmoker. We use the partially linear single-index model:

$$Y_{ij} = Z_{ij}\theta + g(X_{ij}^{\top}\beta) + \varepsilon_{ij}, \qquad X_{ij} = (X_{ij1}, X_{ij2}, X_{ij3})^{\top}.$$

Under the exchangeable working correlation assumption, we obtain the bias-corrected QIF estimates and 95% confidence intervals (obtained similarly as in [3]) in Table 3.

Fig. 3 provides fitted curves (the solid line) for the unknown link function $g(\cdot)$. We also provide 95% pointwise confidence intervals (dotted lines) for g. To test whether the single-index part is linearly correlated with the response variable, we

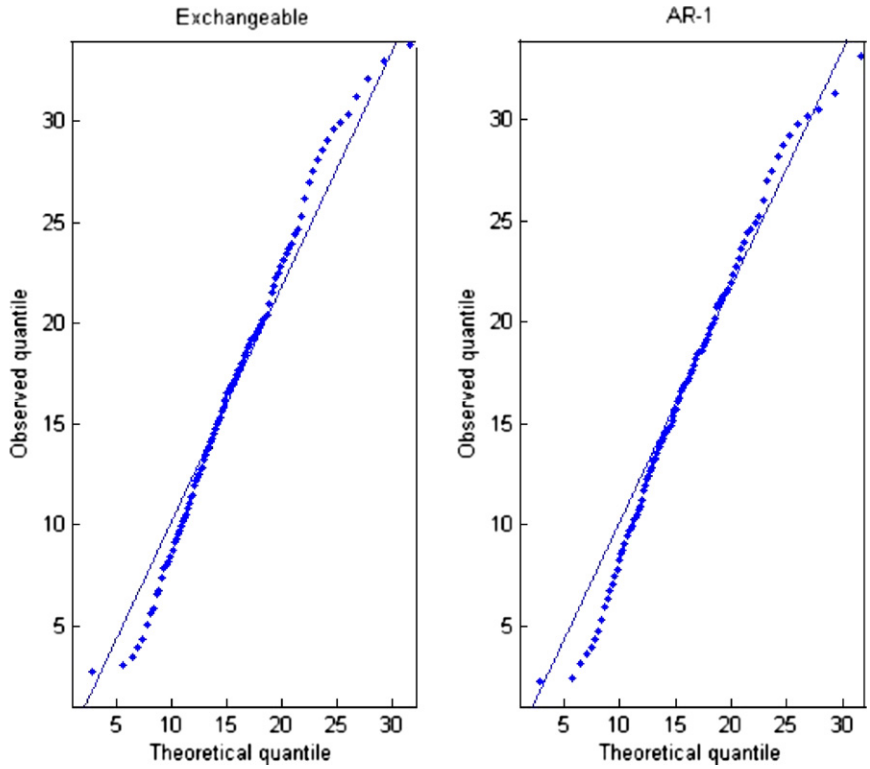

Fig. 2. Quantile–quantile plots for test statistics versus χ^2_d under H_0 from 1000 simulations: assume exchangeable working correlation and AR(1) working correlation respectively.

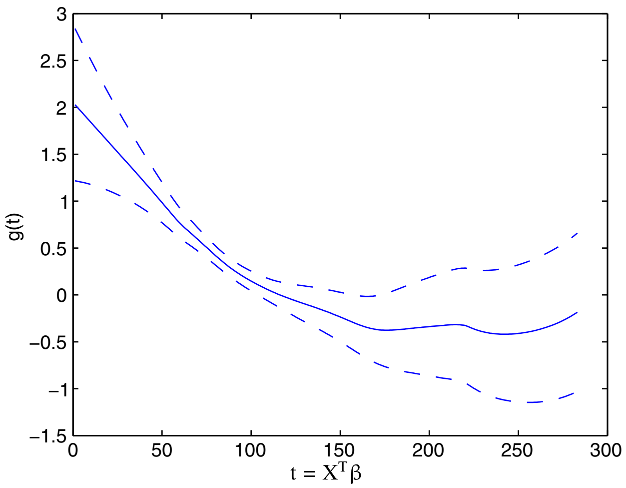

Fig. 3. The estimated curve of link function $g(t)$, $t = X^\top \beta$. The solid line is fitted curve and dash lines are 95% pointwise confidence intervals.

compute the test statistic $\gamma_K T_{\mathrm{GLR}} = 19.0824$. The asymptotic p-value from the chi-squared test is less than 0.015. This indicates that the single-index part is nonlinearly related to the response variable.

Acknowledgments

Peng Lai's research was supported by National Natural Science Foundation of China (11226222, 11201317), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (12KJB110016) and Natural Science Foundation of Jiangsu Province of China (BK2012459). Gaorong Li's research was supported by the National Nature Science Foundation of China (11101014), the Specialized Research Fund for the Doctoral Program of Higher Education of China (20101103120016), Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality (PHR20110822), Training Programme Foundation for the Beijing Municipal Excellent Talents (2010D005015000002), the Fundamental Research Foundation of Beijing University of Technology (X4006013201101) and Program for JingHua Talents in Beijing University of Technology. Heng Lian's research was supported by Singapore MOE Tier 1 RG 62/11.

Appendix. Proof of the theorems

In order to study the asymptotic behavior of the estimator, the following conditions are supposed.

C1 For any $i = 1, \ldots, n, j = 1, \ldots, m_i$, the density function of $X_{ij}^\top \beta$ is bounded away from zero and infinity on $\mathcal{T}$, and satisfies the Lipschitz condition of order 1 on $\mathcal{T}$, where $\mathcal{T} = \{t = X_{ij}^\top \beta : X_{ij} \in A, i = 1, \ldots, n, j = 1, \ldots, m_i\}$ and A is a compact support set of X_{ij}.

C2 $g(t)$ has two bounded and continuous derivatives on $\mathcal{T}$; $g_{1s}(t)$ and $g_{2k}(t)$ satisfy the local Lipschitz condition of order 1, where $g_{1s}(t)$ and $g_{2k}(t)$ are the sth and kth components of $g_1(t)$ and $g_2(t)(1 \le s \le p, 1 \le k \le q)$ respectively.

C3 The kernel $K(u)$ is a bounded and symmetric probability density function, and satisfies

$$\int_{-\infty}^{\infty} u^2 K(u)du \neq 0, \qquad \int_{-\infty}^{\infty} |u|^i K(u)du < \infty, \quad i = 1, 2, \ldots.$$

C4 There exists a positive constant M, such that $\max_{1\le i\le n, 1\le j\le m_i} \sup_{x,z} E(\varepsilon_{ij}^4 | X_{ij} = x, Z_{ij} = z) \le M < \infty$ and $\max_{1\le i\le n, 1\le j\le m_i} \sup_x E(\varepsilon_{ij}^4 | X_{ij} = x) \le M < \infty$.

C5 When $n \to \infty$, the bandwidth h satisfies that $h \to 0,\ n^2h^7 \to \infty,\ nh^8 \to 0$.

C6 There exist two positive constants c_1 and c_2 such that

$$0 < c_1 \le \min_{1\le i\le n} \lambda_{i1} \le \max_{1\le i\le n} \lambda_{im_i} \le c_2 < \infty,$$

where λ_{i1} and λ_{im_i} denote the smallest and largest eigenvalues of Σ_i, respectively.

C7 There exist positive constants c_3 and c_4 such that

$$0 < c_3 \le \min_{1\le i\le n} \lambda'_{i1} \le \max_{1\le i\le n} \lambda'_{im_i} \le c_4 < \infty,$$

where λ'_{i1} and λ'_{im_i} denote the smallest and largest eigenvalues of V_i, respectively.

C8 There exists a positive constant M that satisfies, for all i, j, $\sup_{t\in\mathcal{T}} E(\|Z_{ij}\|^2 | X_{ij}^\top \beta = t) \le M < \infty$.

C9 The matrix $\frac{1}{n}\sum_{i=1}^n U_i(\beta^{(r)}, \theta)U_i^\top(\beta^{(r)}, \theta)$ converges almost surely to an invertible positive definite matrix $\Sigma(\beta^{(r)}, \theta)$, where

$$\Sigma(\beta^{(r)}, \theta) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E[U_i(\beta^{(r)}, \theta)U_i^\top(\beta^{(r)}, \theta)].$$

Proof of Theorem 1. By (2.9), it follows that

$$\bar{\hat{U}}_n(\beta_0^{(r)}, \theta_0) = \frac{1}{n}\sum_{i=1}^n U_i(\beta_0^{(r)}, \theta_0) + \frac{1}{n}\sum_{i=1}^n (\hat{U}_i(\beta_0^{(r)}, \theta_0) - U_i(\beta_0^{(r)}, \theta_0)). \tag{A.1}$$

Noting that $B_1 = \frac{1}{n}\sum_{i=1}^n (\hat{U}_i(\beta_0^{(r)}, \theta_0) - U_i(\beta_0^{(r)}, \theta_0))$ is a $(p + q - 1) \times k$ vector, we consider the first $p + q - 1$ components of B_1 denoted as B_{11}, that is

$$\begin{aligned} B_{11} = {} & \frac{1}{n}\sum_{i=1}^n \Big[(\hat{\Lambda}_i - \Lambda_i)A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}} \boldsymbol{\varepsilon}_i + (\hat{\Lambda}_i - \Lambda_i)A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}} (\hat{G}(\boldsymbol{X}_i\beta_0) - G(\boldsymbol{X}_i\beta_0)) \\ & - \Lambda_i A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}} (\hat{G}(\boldsymbol{X}_i\beta_0) - G(\boldsymbol{X}_i\beta_0)) \Big] \\ := {} & B_{11_1} + B_{11_2} - B_{11_3}. \end{aligned} \tag{A.2}$$

Since the first column of $(\hat{\Lambda}_i - \Lambda_i)$ is $\binom{B_{11_1,\Lambda}}{g_2(X_{i1}^\top\beta_0) - \hat{g}_2(X_{i1}^\top\beta_0)}$, where

$$\begin{aligned} B_{11_1,\Lambda} = {} & g'(X_{i1}^\top\beta_0) J_{\beta_0^{(r)}}^\top (g_1(X_{i1}^\top\beta_0) - \hat{g}_1(X_{i1}^\top\beta_0)) + (\hat{g}'(X_{i1}^\top\beta_0) - g'(X_{i1}^\top\beta_0)) \\ & \times J_{\beta_0^{(r)}}^\top (X_{i1} - g_1(X_{i1}^\top\beta_0)) + (\hat{g}'(X_{i1}^\top\beta_0) - g'(X_{i1}^\top\beta_0)) J_{\beta_0^{(r)}}^\top (g_1(X_{i1}^\top\beta_0) - \hat{g}_1(X_{i1}^\top\beta_0)). \end{aligned}$$

By Lemmas A.1–A.3 in [13], it is easy to show that

$$\sqrt{n}\bar{\hat{U}}_i(\beta_0^{(r)}, \theta_0) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} U_i(\beta_0^{(r)}, \theta_0)\right) + O_p\left(h + \left(\frac{1}{n^2h^7}\right)^{1/4}\right). \tag{A.3}$$

Because of $n^2h^7 \to \infty$, $h \to 0$, using the central limit theorem, we have

$$\sqrt{n}\bar{\hat{U}}_i(\beta_0^{(r)}, \theta_0) \xrightarrow{L} N(0, \Sigma(\beta_0^{(r)}, \theta_0)), \tag{A.4}$$

where

$$\Sigma(\beta_0^{(r)}, \theta_0) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E(U_i(\beta_0^{(r)}, \theta_0)U_i^\top(\beta_0^{(r)}, \theta_0)).$$

Similar to Lemma A.5 in [13], some simple calculations yield that $\hat{C}_n(\beta_0^{(r)}, \theta_0) = \frac{1}{n}\sum_{i=1}^{n} U_i(\beta_0^{(r)}, \theta_0)U_i^\top(\beta_0^{(r)}, \theta_0) + o_p(1)$. By condition (C9), we have

$$\hat{C}_n(\beta_0^{(r)}, \theta_0) \xrightarrow{P} \Sigma(\beta_0^{(r)}, \theta_0). \tag{A.5}$$

Since $(\hat{\beta}^{(r)\top}, \hat{\theta}^\top)^\top = \arg\min_{\beta^{(r)},\theta} \hat{Q}_n(\beta^{(r)}, \theta)$, then obviously, $\hat{Q}_n(\hat{\beta}^{(r)}, \hat{\theta}) \leq \hat{Q}_n(\beta_0^{(r)}, \theta_0)$. Therefore, together with (A.4)–(A.5), we have

$$\hat{Q}_n(\beta_0^{(r)}, \theta_0) = \bar{\hat{U}}_n^\top(\beta_0^{(r)}, \theta_0)\hat{C}_n^{-1}(\beta_0^{(r)}, \theta_0)\bar{\hat{U}}_n(\beta_0^{(r)}, \theta_0) = O_p(n^{-1}) = o_p(1),$$

which implies that $\hat{Q}_n(\hat{\beta}^{(r)}, \hat{\theta}) \xrightarrow{P} 0$. Similar to Lemma 4 in [2], we can obtain that the bias-corrected QIF estimator $(\hat{\beta}^{(r)\top}, \hat{\theta}^\top)^\top$ converges to $(\beta_0^{(r)\top}, \theta_0^\top)^\top$ in probability.

Next we will prove the asymptotic normality. Noting that

$$\hat{Q}_n(\beta^{(r)}, \theta) = \bar{\hat{U}}_n^\top(\beta^{(r)}, \theta)\hat{C}_n^{-1}(\beta^{(r)}, \theta)\bar{\hat{U}}_n(\beta^{(r)}, \theta),$$

let $\beta^{(r)} = \beta_0^{(r)} + \frac{u_1}{\sqrt{n}}$ and $\theta = \theta_0 + \frac{u_2}{\sqrt{n}}$. Let $\eta = (\beta^{(r)\top}, \theta^\top)^\top$ and $\eta_0 = (\beta_0^{(r)\top}, \theta_0^\top)^\top$. Thus, $\eta = \eta_0 + \frac{u}{\sqrt{n}}$, $u = (u_1^\top, u_2^\top)^\top$ and

$$\hat{Q}_n\left(\beta_0^{(r)} + \frac{u_1}{\sqrt{n}}, \theta_0 + \frac{u_2}{\sqrt{n}}\right) = \hat{Q}_n\left(\eta_0 + \frac{u}{\sqrt{n}}\right) := \hat{Q}_n^*(u).$$

Let $\hat{u} = \arg\min_u \hat{Q}_n^*(u)$; then $\hat{\eta} = \eta_0 + \frac{\hat{u}}{\sqrt{n}}$ or $\hat{u} = \sqrt{n}(\hat{\eta} - \eta_0)$. Furthermore, we have

$$\hat{Q}_n^*(u) = \left[\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i\left(\eta_0 + \frac{u}{\sqrt{n}}\right)\right]^\top \hat{C}_n^{-1}\left(\eta_0 + \frac{u}{\sqrt{n}}\right)\left[\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i\left(\eta_0 + \frac{u}{\sqrt{n}}\right)\right]$$

$$\hat{Q}_n^*(0) = \left[\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i(\eta_0)\right]^\top \hat{C}_n^{-1}(\eta_0)\left[\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i(\eta_0)\right]. \tag{A.6}$$

We compute $\hat{Q}_n^*(u) - \hat{Q}_n^*(0) = V_a^{(n)}(u)$. By Lemmas A.1–A.3 in [13], similar to the proof of Lemma A.4 in [13], we can obtain

$$\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i\left(\eta_0 + \frac{u}{\sqrt{n}}\right) = \frac{1}{n}\sum_{i=1}^{n} \tilde{U}_i\left(\eta_0 + \frac{u}{\sqrt{n}}\right) + o_p(1) := \frac{1}{n}\sum_{i=1}^{n} \tilde{U}_i(u) + o_p(1), \tag{A.7}$$

where

$$\frac{1}{n}\sum_{i=1}^{n} \tilde{U}_i(u) = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} \tilde{\Lambda}_i(u)A_i^{-\frac{1}{2}}M_1A_i^{-\frac{1}{2}}(\boldsymbol{Y}_i - G(\boldsymbol{X}_i\beta(u)) - \boldsymbol{Z}_i\theta(u)) \\ \cdots \\ \tilde{\Lambda}_i(u)A_i^{-\frac{1}{2}}M_kA_i^{-\frac{1}{2}}(\boldsymbol{Y}_i - G(\boldsymbol{X}_i\beta(u)) - \boldsymbol{Z}_i\theta(u)) \end{pmatrix},$$

$\beta(u) = (\beta_1(u), \dots, \beta_p(u))^\top$, $\beta^{(r)}(u) = \beta_0^{(r)} + \frac{u_1}{\sqrt{n}}$, $\beta_r(u) = \sqrt{1 - \|\beta^{(r)}(u)\|}$, $\theta(u) = \theta_0 + \frac{u_2}{\sqrt{n}}$, and

$$\tilde{\Lambda}_i(u) = \begin{pmatrix} g'(X_{i1}^\top \beta(u))(X_{i1} - g_1(X_{i1}^\top \beta(u)))^\top J_{\beta_0^{(r)}}, & (Z_{i1} - g_2(X_{i1}^\top \beta(u)))^\top \\ \vdots & \vdots \\ g'(X_{im_i}^\top \beta(u))(X_{im_i} - g_1(X_{im_i}^\top \beta(u)))^\top J_{\beta_0^{(r)}}, & (Z_{im_i} - g_2(X_{im_i}^\top \beta(u)))^\top \end{pmatrix}^\top. \tag{A.8}$$

Similarly, we have

$$\hat{Q}_n^*(u) = \left[\frac{1}{n}\sum_{i=1}^n \tilde{U}_i(u)\right]^\top C_n^{-1}(\eta_0)\left[\frac{1}{n}\sum_{i=1}^n \tilde{U}_i(u)\right] + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{A.9}$$

and

$$\hat{Q}_n^*(0) = \left[\frac{1}{n}\sum_{i=1}^n \tilde{U}_i(0)\right]^\top C_n^{-1}(\eta_0)\left[\frac{1}{n}\sum_{i=1}^n \tilde{U}_i(0)\right] + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{A.10}$$

Therefore,

$$\begin{aligned} V_a^{(n)}(u) &= \left[\frac{1}{n}\sum_{i=1}^n (\tilde{U}_i(u) - \tilde{U}_i(0))\right]^\top C_n^{-1}(\eta_0)\left[\frac{1}{n}\sum_{i=1}^n (\tilde{U}_i(u) - \tilde{U}_i(0))\right] \\ &\quad + \left[\frac{1}{n}\sum_{i=1}^n \tilde{U}_i(0)\right]^\top C_n^{-1}(\eta_0)\left[\frac{1}{n}\sum_{i=1}^n (\tilde{U}_i(u) - \tilde{U}_i(0))\right] \\ &\quad + \left[\frac{1}{n}\sum_{i=1}^n (\tilde{U}_i(u) - \tilde{U}_i(0))\right]^\top C_n^{-1}(\eta_0)\left[\frac{1}{n}\sum_{i=1}^n \tilde{U}_i(0)\right] + o_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{A.11}$$

Applying the Taylor expansion to $\varphi(X_{ij}^\top \beta(u))$, it follows

$$\varphi(X_{ij}^\top \beta(u)) = \varphi(X_{ij}^\top \beta_0) + \varphi'(X_{ij}^\top \beta_0) X_{ij}^\top J_{\beta_0^{(r)}}\left(\beta^{(r)}(u) - \beta_0^{(r)}\right) + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{A.12}$$

where $\varphi(\cdot)$ denotes any of $g(\cdot)$, $g'(\cdot)$, $g_1(\cdot)$ and $g_2(\cdot)$.

Through some simple calculations, (A.8) and (A.12) imply that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n (\tilde{U}_i(u) - \tilde{U}_i(0)) &= -\frac{1}{n}\sum_{i=1}^n \begin{pmatrix} \tilde{\Lambda}_i(0) A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}} \tilde{\Lambda}_{i*}^\top(0) \\ \vdots \\ \tilde{\Lambda}_i(0) A_i^{-\frac{1}{2}} M_k A_i^{-\frac{1}{2}} \tilde{\Lambda}_{i*}^\top(0) \end{pmatrix} \frac{u}{\sqrt{n}} + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= -\Gamma_{1n}\frac{u}{\sqrt{n}} + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \tag{A.13}$$

where

$$\tilde{\Lambda}_{i*}^\top(0) = \begin{pmatrix} g'(X_{i1}^\top \beta_0) X_{i1}^\top J_{\beta_0^{(r)}} & Z_{i1}^\top \\ \vdots & \vdots \\ g'(X_{im_i}^\top \beta_0) X_{im_i}^\top J_{\beta_0^{(r)}} & Z_{im_i}^\top \end{pmatrix}.$$

Furthermore, note that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \tilde{U}_i(0)\right) := W \xrightarrow{L} N(0, \Sigma(\beta_0^{(r)}, \theta_0)), \tag{A.14}$$

and

$$\Gamma_{1n} \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n} E\left[\begin{pmatrix} \tilde{\Lambda}_i(0)A_i^{-\frac{1}{2}}M_1A_i^{-\frac{1}{2}}\tilde{\Lambda}_{i*}^{\top}(0) \\ \vdots \\ \tilde{\Lambda}_i(0)A_i^{-\frac{1}{2}}M_kA_i^{-\frac{1}{2}}\tilde{\Lambda}_{i*}^{\top}(0) \end{pmatrix}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E\left[\begin{pmatrix} \tilde{\Lambda}_i(0)A_i^{-\frac{1}{2}}M_1A_i^{-\frac{1}{2}}\tilde{\Lambda}_{i}^{\top}(0) \\ \vdots \\ \tilde{\Lambda}_i(0)A_i^{-\frac{1}{2}}M_kA_i^{-\frac{1}{2}}\tilde{\Lambda}_{i}^{\top}(0) \end{pmatrix}\right] = \Gamma. \tag{A.15}$$

Therefore, by (A.5), (A.13)–(A.14), (A.11) reduces to

$$V_a^{(n)}(u) = \frac{1}{n}u^{\top}\Gamma^{\top}\Sigma^{-1}(\beta_0^{(r)},\theta_0)\Gamma u - \frac{2}{n}u^{\top}\Gamma^{\top}\Sigma^{-1}(\beta_0^{(r)},\theta_0)W + o_p\left(\frac{1}{n}\right). \tag{A.16}$$

By Slutsky's theorem, we see that

$$\hat{Q}_n^*(u) - \hat{Q}_n^*(0) \xrightarrow{L} V_a(u) \quad \text{for every } u,\ \|u\| \le M,$$

and the unique minimum of $V_a(u)$ is $u = (\Gamma^{\top}\Sigma^{-1}(\beta_0^{(r)},\theta_0)\Gamma)^{-1}\Gamma^{\top}\Sigma^{-1}(\beta_0^{(r)},\theta_0)W$. Noting that $\hat{u}$ minimizes $\hat{Q}_n^*(u)$, using Lemma 4 in [2], Argmax theorem and its corollary in [20], we can conclude that

$$\hat{u} \xrightarrow{L} \left(\Gamma^{\top}\Sigma^{-1}(\beta_0^{(r)},\theta_0)\Gamma\right)^{-1}\Gamma^{\top}\Sigma^{-1}(\beta_0^{(r)},\theta_0)W,$$

thus,

$$\hat{u} \xrightarrow{L} N\big(0, (\Gamma^{\top}\Sigma^{-1}(\beta_0^{(r)},\theta_0)\Gamma)^{-1}\big).$$

With $\hat{u} = \sqrt{n}(\hat{\eta} - \eta_0)$, it follows

$$\sqrt{n}\begin{pmatrix} \hat{\beta}^{(r)} - \beta_0^{(r)} \\ \hat{\theta} - \theta_0 \end{pmatrix} \xrightarrow{L} N\big(0, (\Gamma^{\top}\Sigma^{-1}(\beta_0^{(r)},\theta_0)\Gamma)^{-1}\big). \quad \square$$

Proof of Theorem 2. The proof of this theorem is similar to the proof of Theorem 3.1 in [26], the theorem in [27] and Theorem 5 in [14]. Hence, the proof details are omitted here. □

References

[1] Y. Bai, W. Fung, Z. Zhu, Penalized quadratic inference functions for single-index models with longitudinal data, Journal of Multivariate Analysis 100 (2009) 152–161.
[2] Y. Bai, Z. Zhu, W. Fung, Partially linear models for longitudinal data based on quadratic inference functions, Scandinavian Journal of Statistics 35 (2008) 104–118.
[3] R. Carroll, J. Fan, I. Gijbels, M. Wand, Generalized partially linear single-index models, Journal of the American Statistical Association 92 (1997) 477–489.
[4] J. Fan, I. Gijbels, Local Polynomial Modelling and its Applications, Volume 66, Chapman & Hall/CRC, 1996.
[5] J. Fan, T. Huang, R. Li, Analysis of longitudinal data with semiparametric estimation of covariance function, Journal of the American Statistical Association 102 (2007) 632–641.
[6] J. Fan, R. Li, New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis, Journal of the American Statistical Association 99 (2004) 710–723.
[7] J. Fan, C. Zhang, J. Zhang, Generalized likelihood ratio statistics and Wilks phenomenon, The Annals of Statistics 29 (2001) 153–193.
[8] L. Hansen, Large sample properties of generalized method of moments estimators, Econometrica: Journal of the Econometric Society 50 (1982) 1029–1054.
[9] J. Huang, C. Wu, L. Zhou, Varying-coefficient models and basis function approximations for the analysis of repeated measurements, Biometrika 89 (2002) 111–128.
[10] R. Kaslow, D. Ostrow, R. Detels, J. Phair, B. Polk, C. Rinaldo, The Multicenter AIDS Cohort Study: rationale, organization and selected characteristics of the participants, American Journal of Epidemiology 126 (1987) 310–318.
[11] P. Lai, Q. Wang, H. Lian, Bias-corrected GEE estimation and smooth-threshold GEE variable selection for single-index models with clustered data, Journal of Multivariate Analysis 105 (2012) 422–432.
[12] G. Li, P. Tian, L. Xue, Generalized empirical likelihood inference in semiparametric regression model for longitudinal data, Acta Mathematica Sinica 24 (2008) 2029–2040.
[13] G. Li, L. Zhu, L. Xue, S. Feng, Empirical likelihood inference in partially linear single-index models for longitudinal data, Journal of Multivariate Analysis 101 (2010) 718–732.
[14] H. Liang, X. Liu, R. Li, C. Tsai, Estimation and testing for partially linear single-index models, Annals of Statistics 38 (2010) 3811–3836.
[15] K. Liang, S. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1986) 13–22.

[16] X. Lin, R. Carroll, Semiparametric regression for clustered data using generalized estimating equations, Journal of the American Statistical Association 96 (2001) 1045–1056.
[17] A. Qu, R. Li, Quadratic inference functions for varying-coefficient models with longitudinal data, Biometrics 62 (2006) 379–391.
[18] A. Qu, B. Lindsay, B. Li, Improving generalised estimating equations using quadratic inference functions, Biometrika 87 (2000) 823–836.
[19] P. Tian, L. Yang, L. Xue, Asymptotic properties of estimators in partially linear single-index model for longitudinal data, Acta Mathematica Scientia 30 (2010) 677–687.
[20] A. van der Vaart, Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics), 2000.
[21] Y. Xia, W. Härdle, Semi-parametric estimation of partially linear single-index models, Journal of Multivariate Analysis 97 (2006) 1162–1184.
[22] L. Xue, L. Zhu, Empirical likelihood semiparametric regression analysis for longitudinal data, Biometrika 94 (2007) 921–937.
[23] J. You, G. Chen, Y. Zhou, Statistical inference of partially linear regression models with heteroscedastic errors, Journal of Multivariate Analysis 98 (2007) 1539–1557.
[24] Y. Yu, D. Ruppert, Penalized spline estimation for partially linear single-index models, Journal of the American Statistical Association 97 (2002) 1042–1054.
[25] S. Zeger, P. Diggle, Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters, Biometrics 50 (1994) 689–699.
[26] R. Zhang, Inferences on nonparametric component for partially linear models, Communications in Statistics—Theory and Methods 36 (2007) 549–562.
[27] R. Zhang, Tests for nonparametric parts on partially linear single index models, Science in China Series A: Mathematics 50 (2007) 439–449.
[28] L. Zhu, L. Xue, Empirical likelihood confidence regions in a partially linear single-index model, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68 (2006) 549–570.