

Generalized empirical likelihood inference in partially linear errors-in-variables models with longitudinal data

Juanfang Liu^{*†}, Liugen Xue[‡] and Ruiqin Tian[§]

Abstract

This article is concerned with estimations for longitudinal partial linear models with covariate that is measured with error. We propose a generalized empirical likelihood method by combining correction attenuation and quadratic inference functions. The method takes into account the within-subject correlation without involving direct estimation of nuisance parameters in the correlation matrix. We define a generalized empirical likelihood-based statistic for the regression coefficients and residual adjusted empirical likelihood for the baseline function. The empirical log-likelihood ratios are proven to be asymptotically chi-squared, and the corresponding confidence regions are then constructed. Compared with methods based on normal approximations, the generalized empirical likelihood does not require consistent estimators for the asymptotic variance and bias. Furthermore, a simulation study is conducted to evaluate the performance of the proposed method.

Keywords: Longitudinal Data, Generalized Empirical Likelihood, Confidence Region, Measurement Error, Partially Linear Model.

Mathematics Subject Classification (2010): 62G05, 62G20

Received: 19.12.2015 *Accepted:* 09.08.2016 *Doi:* 10.15672/HJMS.201614220442

^{*}College of Applied Sciences, Beijing University of Technology and College of Mathematics and Information Science, Henan Normal University, Email: liujuanfang@emails.bjut.edu.cn

[†]Corresponding Author.

[‡]College of Applied Sciences, Beijing University of Technology, Email: lgxue@bjut.edu.cn

[§]Department of Statistics, Zhejiang Agriculture and Forestry University, Email: tianruiqin@emails.bjut.edu.cn

1. Introduction

Longitudinal data analysis has attracted considerable research interest and a large number of inference methods have been proposed in the literature. Consider data from n subjects with n_i observations in the i th subject ($i = 1, \dots, n$) for a total of $N = \sum_{i=1}^n n_i$. Let Y_{ij} and X_{ij}, T_{ij} respectively be the response variable and the covariates of the j th observation ($j = 1, \dots, n_i$) in the i th, where X_{ij} is a $p \times 1$ vector and T_{ij} is a scalar or time. Zeger and Diggle [28] proposed a semiparametric regression model of the form

$$(1.1) \quad Y_{ij} = X_{ij}^T \beta + g(T_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, n, j = 1, \dots, n_i.$$

where β is a $p \times 1$ vector of unknown regression coefficients associated with covariate X_{ij} , $g(t)$ is an unknown smooth function, ε_{ij} is random error with $E(\varepsilon_{ij}|X_{ij}, T_{ij}) = 0$ and $\sigma_\varepsilon^2(t) = E(\varepsilon_{ij}^2|T_{ij} = t)$. We assume, without loss of generality, that T_{ij} are all scaled into closed interval $[0, 1]$. We assume further that the observations from different subjects are independent.

Model (1.1) is especially useful for longitudinal data analysis as the level of response often depends on time in a nonlinear pattern. Many authors have studied models in the form of (1.1), see for example, He et al. [8], You et al. [27] and Xue and Zhu [24], among others. Zeger and Diggle [28] used a semiparametric mixed model to analyse the CD4 cell numbers in HIV seroconverters where $g(t)$ is estimated by a kernel smoother. A major aspect of longitudinal data is the within-subject correlation among the repeated measurements. Ignoring this within-subject correlation causes a loss of efficiency in general problems. Using a working correlation matrix with a small set of nuisance parameters α , the generalized estimating equations (GEE) estimator of the regression coefficients proposed by Liang and Zeger [13] are consistent even when the working correlation structure is misspecified. However, Crowder [4] established that there are difficulties with estimating the nuisance parameters α and that in some simple cases, consistent estimators of α do not always exist. To avoid the drawback, Qu et al. [18] introduced a method of quadratic inference functions (QIF). It avoids estimating the nuisance correlation structure parameters by assuming that the inverse of working correlation matrix can be approximated by a linear combination of several known basis matrices. The QIF can efficiently take the within-cluster correlation into account and is more efficient than the GEE approach when the working correlation is misspecified. Bai et al. [2] extended the QIF method to the semiparametric partial linear model. Dziak et al. [6] gave an overview on QIF approaches for longitudinal data. Owen [15] introduced a nonparametric method of inference - an empirical likelihood (EL) method. The EL uses only the data to determine the shape and orientation of a confidence region and does not use the estimator of the asymptotic covariance. Hence, EL is indeed appealing for the construction of confidence region.

Owen [17] provided a comprehensive account of empirical likelihood and its properties. For longitudinal data, You et al. [27] constructed a block empirical likelihood method for partially linear regression models with longitudinal data. Xue and Zhu [24] considered the same model, and they provided the EL inference for the baseline function as well as the regression coefficients.

In many practical situations, there often exist covariate measurement errors. Some recently related works include Cui and Chen [5], Liang et al. [11], Liu [14], and Zhao and Xue [29], among others. Zhao and Xue [29] investigated empirical likelihood inferences for semiparametric varying-coefficient partially linear EV models. Liu [14] considered the asymptotic normality for the partially linear EV models with longitudinal data. Yang et al. [26] investigated the empirical likelihood of varying coefficient errors-in-variables models with longitudinal. A common feature of these articles is that the data dependence

within each subject is not taken into consideration. To consider the within correlation, Tian et al. [22] proposed a generalized empirical likelihood (GEL) method by combining quadratic inference functions for generalized linear model with longitudinal data. Tian et al. [21] discussed the variable selection for the partial linear EV model with longitudinal data when some covariates are measured with errors.

We propose a modified generalized empirical log-likelihood ratio function for the regression coefficients and a residual-adjusted empirical likelihood for the baseline function, the empirical log-likelihood ratios are proven to be asymptotically chi-squared. The following three desired features are worth mentioning. First, the method directly incorporates within-subject correlation into model building, but does not require estimation of the nuisance parameters associated with the correlation. Second, the modified generalized empirical log-likelihood ratio function eliminate the effects of measurement errors on parameter estimation. Third, by using the residual adjusted EL ratio, undersmoothing for estimating the baseline function is avoided.

The outline of this paper is organized as follows. In Section 2, we define a generalized empirical log-likelihood ratio for regression coefficients and investigate its asymptotic properties. In Section 3, we discuss the empirical likelihood inference for nonparametric function. In Section 4, we conduct a simulation study to compare the finite sample properties of these suggested estimators. We also apply our method to analyze an AIDS clinical trial dataset in Section 5. The proofs of theorems appear in the Appendix.

2. Empirical likelihood for the regression coefficients

2.1. Known measurement error covariance matrix. For model (1.1), the covariates X_{ij} are not always observable without error. If X_{ij} are measured with error, instead of observing X_{ij} , we observe

$$(2.1) \quad W_{ij} = X_{ij} + U_{ij}.$$

where $U_{ij}, i = 1, \dots, n, j = 1, \dots, n_i$, are measurement errors. As in Liang et al. [11], we assume that U_{ij} are independent and identically distributed, independent of $\{Y_{ij}, X_{ij}, T_{ij}, \varepsilon_{ij}\}$. Although this assumption is not the weakest possible condition, it is imposed to facilitate the technical proofs, and it can be satisfied in many applications. We suppose that $E(U_{ij}) = 0$, $\text{cov}(U_{ij}) = \Sigma_u$.

From the model (1.1), we have $E(Y_{ij}|T_{ij}) = E(X_{ij}^T|T_{ij})\beta + g(T_{ij})$. For the sake of descriptive convenience, we denote $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})^T$, and X_i, U_i, T_i, W_i in a similar fashion. $g(T_i) = (g(T_{i1}), \dots, g(T_{in_i}))^T$, $m_X(t) = E(X_{ij}|T_{ij} = t)$, $m_Y(t) = E(Y_{ij}|T_{ij} = t)$, $m_W(t) = E(W_{ij}|T_{ij} = t)$, $\tilde{Y}_i = \varphi(T_{ij})(Y_i - m_Y(T_i))$, $\tilde{X}_i = \varphi(T_{ij})(X_i - m_X(T_i))$, $\tilde{W}_i = \varphi(T_{ij})(W_i - m_W(T_i))$, where $\varphi(\cdot)$ is a bounded nonnegative weight function with a compact support $[a, b] \in [0, 1]$. Motivated by Liang et al. [11], we construct the modified generalized auxiliary random vectors

$$(2.2) \quad Z_i(\beta) = \tilde{W}_i^T V_i^{-1}(\tilde{Y}_i - \tilde{W}_i\beta) + E(U_i^T V_i^{-1} U_i)\beta,$$

where V_i is an arbitrarily specified working covariance matrix with nuisance parameters α . Following Liang and Zeger [13], the matrix V_i is often modeled as $A_i^{1/2}R(\alpha)A_i^{1/2}$, where $A_i = \text{diag}\{\text{var}(Y_{i1}), \dots, \text{var}(Y_{in_i})\}$, $R(\alpha)$ is some working correlation which involves a small number of nuisance parameters α . By Qu et al. [18], we model the inverse of the working correlation $R^{-1}(\alpha)$ by the class of matrices

$$(2.3) \quad \sum_{k=1}^s a_k M_k,$$

where M_1, \dots, M_s are known matrices and a_1, \dots, a_s are unknown constants. This is a sufficiently rich class that accommodates, or at least approximates, the correlation structures most commonly used. Substituting (2.3) to (2.2) and using the idea of QIF, we need not to find the estimators of parameters $a = (a_1, a_2, \dots, a_s)$ by optimizing some function of the information matrix. Instead, we define the "extend" generalized auxiliary random vectors

$$(2.4) \quad \tilde{Z}_i(\beta) = \begin{pmatrix} \tilde{W}_i^T A_i^{-1/2} M_1 A_i^{-1/2} (\tilde{Y}_i - \tilde{W}_i \beta) + D_i^{(1)} \beta \\ \vdots \\ \tilde{W}_i^T A_i^{-1/2} M_s A_i^{-1/2} (\tilde{Y}_i - \tilde{W}_i \beta) + D_i^{(s)} \beta \end{pmatrix},$$

where $D_i^{(k)} = E(U_i^T A_i^{-1/2} M_k A_i^{-1/2} U_i), k = 1, \dots, s$. Note that $E(\tilde{Z}_i(\beta)) = 0$ if β is the true parameter. Therefore, using such information, we can define a generalized empirical log-likelihood ratio function $l(\beta)$. If β is the true parameter, $l(\beta)$ can be shown to be asymptotically distributed as a chi-square with ps degrees of freedom.

However, the formula above cannot be applied directly, because $m_Y(T_i), m_W(T_i)$ are unknown. Using kernel estimate method, the estimators of $m_Y(T_i)$ and $m_W(T_i)$ are, respectively, defined by

$$(2.5) \quad \hat{m}_W(t) = \sum_{i=1}^n \sum_{j=1}^{n_i} \omega_{ij}(t) W_{ij}, \hat{m}_Y(t) = \sum_{i=1}^n \sum_{j=1}^{n_i} \omega_{ij}(t) Y_{ij}$$

where

$$(2.6) \quad \omega_{ij}(t) = K_h(T_{ij} - t) \Big/ \sum_{k=1}^n \sum_{l=1}^{n_i} K_h(T_{kl} - t)$$

h is a bandwidth, $K(\cdot)$ is a kernel function and $K_h(\cdot) = K(\cdot/h)$. Therefore, an estimator of $Z_i(\beta)$,

$$(2.7) \quad \hat{Z}_i(\beta) = \begin{pmatrix} \hat{W}_i^T A_i^{-1/2} M_1 A_i^{-1/2} (\hat{Y}_i - \hat{W}_i \beta) + D_i^{(1)} \beta \\ \vdots \\ \hat{W}_i^T A_i^{-1/2} M_s A_i^{-1/2} (\hat{Y}_i - \hat{W}_i \beta) + D_i^{(s)} \beta \end{pmatrix},$$

where $\hat{Y}_i = \varphi(T_i)(Y_i - \hat{m}_Y(T_i))$, and $\hat{W}_i = \varphi(T_i)(W_i - \hat{m}_W(T_i))$. Then a modified generalized empirical log-likelihood ratio function for β is defined as

$$(2.8) \quad \hat{l}(\beta) = -2 \max_{p_1, \dots, p_n} \left\{ \sum_{i=1}^n \log(np_i) \Big| p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{Z}_i(\beta) = 0 \right\}.$$

For any given β , a unique value for $\hat{l}(\beta)$ exists, we assume that 0 is inside the convex hull of the points $(\hat{Z}_1(\beta), \dots, \hat{Z}_n(\beta))$ (Owen, [15]). By the Lagrange multiplier method, $\hat{l}(\beta)$ can be represented as

$$(2.9) \quad \hat{l}(\beta) = 2 \sum_{i=1}^n \log(1 + \lambda^T \hat{Z}_i(\beta)),$$

where $\lambda = \lambda(\beta)$ is a $ps \times 1$ vector that solves

$$(2.10) \quad \frac{1}{n} \sum_{i=1}^n \frac{\hat{Z}_i(\beta)}{1 + \lambda^T \hat{Z}_i(\beta)} = 0.$$

The following Theorem 2.1 gives that $\hat{l}(\beta)$ is asymptotically distributed as a chi-square with ps degrees of freedom.

2.1. Theorem. Suppose that the regularity conditions C1 – C6 in the Appendix hold. If β is the true parameter, then

$$\hat{l}(\beta) \xrightarrow{\mathcal{L}} \chi_{ps}^2,$$

where $\xrightarrow{\mathcal{L}}$ represents the convergence in distribution, and χ_{ps}^2 means the chi-square distribution with ps degrees of freedom.

Let $\chi_{ps}^2(1 - \alpha)$ be the $1 - \alpha$ quantile of χ_{ps}^2 for any $0 < \alpha < 1$. By using Theorem 2.1, we obtain an approximate $1 - \alpha$ confidence region for β , defined by

$$C_\alpha(\beta) = \{\beta \mid \hat{l}(\beta) \leq \chi_{ps}^2(1 - \alpha)\}.$$

We may maximize $\{-\hat{l}(\beta)\}$ to obtain an estimator of the parameter β , say $\hat{\beta}$, called as the generalized maximum empirical likelihood estimator (GMELE). Denote

$$(2.11) \quad \Gamma = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E \begin{pmatrix} \tilde{X}_i^T A_i^{-1/2} M_1 A_i^{-1/2} \tilde{X}_i \\ \vdots \\ \tilde{X}_i^T A_i^{-1/2} M_s A_i^{-1/2} \tilde{X}_i \end{pmatrix},$$

$$(2.12) \quad \Sigma = \begin{pmatrix} \Sigma_{11} \cdots \Sigma_{1s} \\ \vdots \ddots \vdots \\ \Sigma_{s1} \cdots \Sigma_{ss} \end{pmatrix},$$

where for $k, m = 1, \dots, s$,

$$\begin{aligned} \Sigma_{km} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \{E[(X_i - m_X(T_i))^T \Psi_i A_i^{-1/2} M_k A_i^{-1/2} (\Psi_i \varepsilon_i - U_i \beta)] \\ &\quad [(X_i - m_X(T_i))^T \Psi_i A_i^{-1/2} M_m A_i^{-1/2} (\Psi_i \varepsilon_i - U_i \beta)]^T \\ &\quad + E(U_i^T A_i^{-1/2} M_k A_i^{-1/2} \Psi_i \varepsilon_i) (U_i^T A_i^{-1/2} M_m A_i^{-1/2} \Psi_i \varepsilon_i)^T \\ &\quad + E(-U_i A_i^{-1/2} M_k A_i^{-1/2} U_i \beta + D_i^{(k)} \beta) \\ &\quad (-U_i A_i^{-1/2} M_m A_i^{-1/2} U_i \beta + D_i^{(m)} \beta)^T. \end{aligned}$$

where $\Psi_i = \text{diag}\{\psi(T_{i1}), \dots, \psi(T_{in_i})\}$. If the matrix Σ and $\Gamma^T \Sigma^{-1} \Gamma$ are invertible, then we can obtain the asymptotic normality for $\hat{\beta}$ in the following Theorem.

2.2. Theorem. Suppose that the regularity conditions C1 – C6 in the Appendix hold. Then when $n \rightarrow \infty$, we have

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N(0, \Sigma_\beta),$$

where $\Sigma_\beta = [\Gamma^T \Sigma^{-1} \Gamma]^{-1}$.

To apply Theorem 2.2 to construct the confidence region of β , we give the consistent estimator of Σ_β , say $\hat{\Sigma}_\beta = [\hat{\Gamma}^T \hat{\Sigma}^{-1} \hat{\Gamma}]^{-1}$, where

$$(2.13) \quad \hat{\Gamma} = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \hat{W}_i^T A_i^{-1/2} M_1 A_i^{-1/2} \hat{W}_i - D_i^{(1)} \\ \vdots \\ \hat{W}_i^T A_i^{-1/2} M_s A_i^{-1/2} \hat{W}_i - D_i^{(s)} \end{pmatrix},$$

$$(2.14) \quad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \hat{Z}_i(\beta) \hat{Z}_i(\beta)^T.$$

Therefore, by Theorem 2.2, we have

$$(2.15) \quad \hat{\Sigma}_\beta^{-1/2} \sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N(0, I_p),$$

where I_p is an identity matrix of order p . Using Theorem 10.2d in Arnold [1] to obtain

$$(2.16) \quad (\hat{\beta} - \beta)^T n \hat{\Sigma}_\beta^{-1} (\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} \chi_p^2.$$

Therefore, the confidence region of β can be constructed by using (2.15) or (2.16).

2.2. Estimated measurement error covariance matrix. Generally, the covariance matrix Σ_u is unknown and must be estimated. We further assume longitudinal data is a balance data, that is $n_i = m$. For unbalanced data, we can use the multiple groups analysis (Shao et al. [20]). The usual method of doing so (Carroll et al. [3], Ch3) is by partial replication, so that we observe $W_i^{(r)} = X_i + U_i^{(r)}$, $r = 1, \dots, m_i$. For notation convenience, we consider here only the case that $m_i = 2$. Let \bar{W}_i is the sample mean of the replicates, and \bar{U}_i in a similar fashion. Then a consistent, unbiased moments estimator for $D_i^{(k)}$ is

$$(2.17) \quad \hat{D}_i^{(k)} = n^{-1} \sum_{i=1}^n \sum_{r=1}^2 (W_i^{(r)} - \bar{W}_i)^T A_i^{-1/2} M_k A_i^{-1/2} (W_i^{(r)} - \bar{W}_i),$$

The estimator of $Z_i(\beta)$ changes only slightly to accommodate the replicates, becoming

$$(2.18) \quad \hat{Z}_i^*(\beta) = \begin{pmatrix} \hat{W}^T A_i^{-1/2} M_1 A_i^{-1/2} (\hat{Y}_i - \hat{W}\beta) + \hat{D}_i^{(1)} \beta / 2 \\ \vdots \\ \hat{W}^T A_i^{-1/2} M_s A_i^{-1/2} (\hat{Y}_i - \hat{W}\beta) + \hat{D}_i^{(s)} \beta / 2 \end{pmatrix},$$

where $\hat{W} = \psi(T_i)(\bar{W}_i - \hat{m}_{\bar{W}}(T_i))$ and $\hat{m}_{\bar{W}}(T_i)$ is the kernel estimate of $\hat{m}_W(T_i)$ based on the data (\bar{W}_i, T_i) . The empirical likelihood ratio function for β may be defined as

$$(2.19) \quad \hat{l}^*(\beta) = -2 \max_{p_1, \dots, p_n} \left\{ \sum_{i=1}^n \log(np_i) \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{Z}_i^*(\beta) = 0 \right\}.$$

We may maximize $\{-\hat{l}^*(\beta)\}$ to obtain maximum empirical likelihood estimator $\hat{\beta}^*$ of β . Similarly, we have the following theorem.

2.3. Theorem. *Under the general conditions of Theorem 2.1, and β is the true parameter, then*

$$\hat{l}^*(\beta) \xrightarrow{\mathcal{L}} \chi_{ps}^2.$$

2.4. Theorem. *Under the general conditions of Theorem 2.2, the estimator $\hat{\beta}^*$ is consistent and asymptotically normal with covariance matrix $[\Gamma^T \Sigma_*^{-1} \Gamma]^{-1}$,*

$$\Sigma_* = \begin{pmatrix} \Sigma_{*11} & \cdots & \Sigma_{*1s} \\ \vdots & \ddots & \vdots \\ \Sigma_{*s1} & \cdots & \Sigma_{*ss} \end{pmatrix},$$

where for $k, m = 1, \dots, s$,

$$\begin{aligned}
 \Sigma_{*km} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \{E[(X_i - m_X(T_i))^T A_i^{-1/2} M_k A_i^{-1/2} (\varepsilon_i - \bar{U}_i \beta)] \\
 &\quad [(X_i - m_X(T_i))^T A_i^{-1/2} M_m A_i^{-1/2} (\varepsilon_i - \bar{U}_i \beta)]^T \\
 (2.20) \quad &+ E(\bar{U}_i^T A_i^{-1/2} M_k A_i^{-1/2} \varepsilon_i) (\bar{U}_i^T A_i^{-1/2} M_m A_i^{-1/2} \varepsilon_i)^T \\
 &+ E(-\bar{U}_i A_i^{-1/2} M_k A_i^{-1/2} \bar{U}_i \beta + \frac{1}{2} D_i^{(k)} \beta) \\
 &\quad (-\bar{U}_i A_i^{-1/2} M_m A_i^{-1/2} \bar{U}_i \beta + \frac{1}{2} D_i^{(m)} \beta)^T.
 \end{aligned}$$

With the replication data, the standard error estimators of $\hat{\beta}^*$ can also be derived (Liang, [12]). We omit details here.

3. Inference based on empirical likelihood for the nonparametric function

We assume from now on that t_0 is an interior point of $[0, 1]$. Also, we suppose that the time points $T_{ij}, i = 1, \dots, n, j = 1, \dots, n_i$, are independent and have identical distribution with common density function $f(t)$. Introduce the following auxiliary random vectors

$$(3.1) \quad \hat{\eta}_i(g(t_0)) = \sum_{j=1}^{n_i} [Y_{ij} - W_{ij}^T \hat{\beta} - g(t_0)] K_h(T_{ij} - t_0).$$

An estimated empirical log-likelihood ratio function for $g(t_0)$ can be define by

$$(3.2) \quad \tilde{l}(g(t_0)) = -2 \max_{p_1, \dots, p_n} \left\{ \sum_{i=1}^n \log(np_i) \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\eta}_i(g(t_0)) = 0 \right\}.$$

We can also maximize $\{-\tilde{l}(g(t_0))\}$ to obtain the maximum empirical likelihood estimator of $g(t_0)$, says $\hat{g}(t_0)$. It can be proved that

$$(3.3) \quad \hat{g}(t_0) = \sum_{i=1}^n \sum_{j=1}^{n_i} \omega_{ij} [Y_{ij} - W_{ij}^T \hat{\beta}] + o_P((Nh)^{-1/2}),$$

where ω_{ij} is defined in (2.6). If we define that

$$(3.4) \quad b(t_0) = h_0^{5/2} [g'(t_0) f'(t_0) + (1/2) g''(t_0) f(t_0)] \int_{-1}^1 u^2 K(u) du,$$

$$(3.5) \quad v^2(t_0) = \sigma_{\varepsilon, u, \delta}^2(t_0) f(t_0) \int_{-1}^1 K^2(u) du,$$

where $\sigma_{\varepsilon, u, \delta}^2(t_0) = E[(\varepsilon_{ij} - U_{ij}^T \beta)^2 | T_{ij} = t_0]$ and h_0 is the constant satisfying Condition C1 in the Appendix. The following theorem gives the asymptotical property of $\hat{g}(t_0)$.

3.1. Theorem. *Suppose that the regularity conditions C1 – C6 in the Appendix hold, Then*

$$\sqrt{Nh} [\hat{g}(t_0) - g(t_0)] - b(t_0) (f(t_0))^{-1} \xrightarrow{\mathcal{L}} N(0, \sigma^2(t_0)),$$

where $\sigma^2(t_0) = v^2(t_0) (f(t_0))^{-2}$, $b(t_0)$ and $v^2(t_0)$ are defined in (3.4) and (3.5).

Similar to Xue and Zhu [24], we can show that if we substitute Condition C1 in Theorem 3.1, with $Nh^2/\log N \rightarrow \infty$ and $Nh^5 \rightarrow 0$, that is, if undersmoothing is adopted, then the biased term $b(t_0)$ vanished asymptotically. Denote that

$$(3.6) \quad \hat{f}(t_0) = \frac{1}{Nh} \sum_{i=1}^n \sum_{j=1}^{n_i} K_h(T_{ij} - t_0),$$

$$(3.7) \quad \hat{v}^2(t_0) = \frac{1}{Nh} \sum_{i=1}^n \hat{\eta}_i^2(\hat{g}(t_0)).$$

Then, a consistent estimator of $\sigma^2(t_0)$ can be given by $\hat{\sigma}^2(t_0) = \hat{v}^2(t_0)/(\hat{f}(t_0))^2$. If we define that

$$(3.8) \quad \hat{b}(t_0) = \frac{1}{\sqrt{Nh}} \sum_{i=1}^n \sum_{j=1}^{n_i} [\hat{g}(T_{ij}) - \hat{g}(t_0)] K_h(T_{ij} - t_0).$$

from the Lemma 6.7 in Appendix, $\hat{b}(t_0)$ is a consistent estimator $b(t_0)$. Then, an approximate $1 - \alpha$ confidence interval for $g(t_0)$ can be given by

$$\hat{g}(t_0) - (Nh)^{-1/2} \hat{b}(t_0) (\hat{f}(t_0))^{-1} \pm z_{\alpha/2} (Nh)^{-1/2} \hat{\sigma}(t_0),$$

where $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution.

Theorem 3.1 together Lemma 6.6 in the Appendix, implies that $\tilde{l}(g(t_0))$ is asymptotically non-central chi-squared if optimal bandwidth is used, and this increases the difficulty of the study. In a manner similar to Xue and Zhu [24], we can adjust the weighted residuals $\hat{\eta}_i\{g(t_0)\}$ and then obtain an adjusted empirical likelihood ratio without undersmoothing. Introduce the auxiliary random vectors

$$\hat{\eta}_i^*\{g(t_0)\} = \hat{\eta}_i\{g(t_0)\} - \sum_{j=1}^{n_i} [\hat{g}(T_{ij}) - \hat{g}(t_0)] K_h(T_{ij} - t_0).$$

A residual-adjusted generalized empirical likelihood ratio can be defined as

$$\tilde{l}^*(g(t_0)) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\eta}_i^*(g(t_0)) = 0 \right\}.$$

Then, the asymptotic result of $\tilde{l}^*(g(t))$ is stated in the following theorem.

3.2. Theorem. *Suppose that the regularity conditions C1 – C6 in the Appendix hold, if $g(t_0)$ is the true value of the baseline function, we have*

$$\tilde{l}^*(g(t_0)) \xrightarrow{\mathcal{L}} \chi_1^2.$$

Applying Theorem 3.2, the approximate $1 - \alpha$ confidence interval for $g(t)$ is defined as $\tilde{I}_\alpha(g(t_0)) = \{g(t) \mid \tilde{l}^*(g(t_0)) \leq \chi_1^2(1 - \alpha)\}$.

4. Simulation studies

We simulated data from the semiparametric regression model

$$Y_{ij} = X_{1ij}\beta_1 + X_{2ij}\beta_2 + \sin(\pi T_{ij}/2 + \pi/2) + \varepsilon_{ij}, \quad i = 1, \dots, n; j = 1, \dots, 5,$$

where $\beta_1 = \beta_2 = 1$, $n = 100$, $X_{1ij} \sim N(1, 1)$, $X_{2ij} \sim N(2, 1)$, $T_{ij} \sim U(-1, 1)$, and error vector $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{i5})^T \sim N(0, \sigma^2 \text{corr}(\varepsilon_i, \rho))$, where $\sigma^2 = 0.6$, $\rho = 0.5$ and $\text{corr}(\varepsilon_i, \rho)$ is a known correlation matrix with parameter ρ used to determine the strength of within-subject dependence. Here we consider ε_{ij} has the compound symmetry (CS) correlation. Considering the measurement error models $W_{1ij} = X_{1ij} + U_{1ij}$, $W_{2ij} = X_{2ij} + U_{2ij}$, where $U_{1ij} \sim N(0, 0.04)$, $U_{2ij} \sim N(0, 0.04)$.

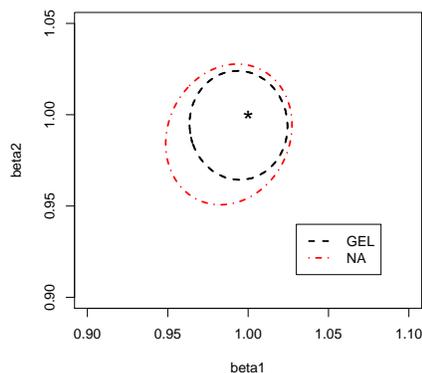
For each simulated dataset, we computed the empirical likelihood ratio and the estimators of β and $g(t)$. The kernel function was taken to be $K(u) = 0.75(1 - u^2)_+$ and the cross-validation bandwidth h_{CV} was obtained by minimizing

$$CV(h) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{n_i} \{Y_{ij} - X_{ij}^T \hat{\beta}_{[i]} - \hat{g}_{[i]}(t_{ij})\}^2,$$

where $\hat{g}_{[i]}(\cdot)$ and $\hat{\beta}_{[i]}$ are estimators of $g(\cdot)$ and β which are computed with all of the measurements but not the i th subject. We experimented with bandwidths around the selected values, and the results did not change significantly. To estimate the variance of \bar{U}_{ij} , we generated duplicate samples of W_{ij} .

FIGURE 1

Figure 1: The averages of 95% confidence regions for (β_1, β_2) .



For the confidence region of β , two methods were compared: the generalized empirical likelihood (GEL) and the normal approximation (NA) in terms of coverage accuracy and area of the confidence region with 1000 simulation runs. The simulation results are presented in Figure 1. Figure 1 shows that the GEL gives smaller confidence region than the NA method. The coverage probability for the GEL is 0.943, while that for the NA is 0.939. This also shows the GEL has higher accuracy than the NA for the confidence region.

FIGURE 2

Figure 2: The 95% confidence intervals and coverage probabilities for $g(t)$. The left panel shows the pointwise confidence intervals and the true curve (solid curve) for $g(t)$. The right panel is for coverage probabilities.

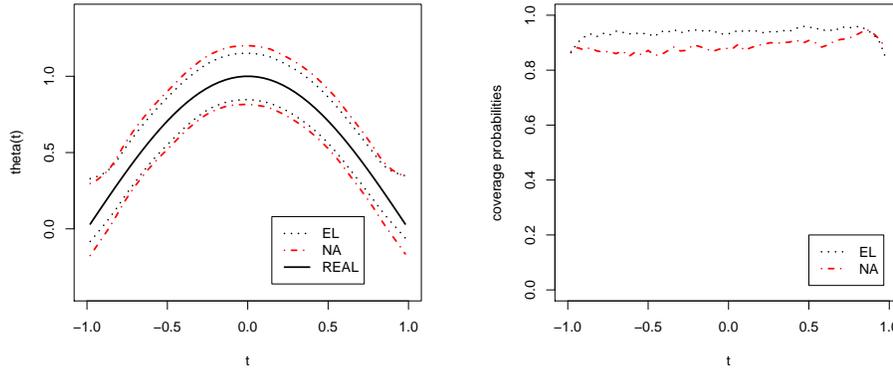


Figure 2 depicts the performance of the residual-adjusted GEL and the NA in terms of 95% pointwise confidence intervals. From Figure 2, the residual-adjusted GEL clearly performs better than the NA because the associated confidence intervals have uniformly higher coverage accuracies and shorter average lengths.

5. A real example

We now illustrate the proposed procedures in this paper through analysis of a data set from the Multi-Center AIDS Cohort study. The data set contains the human immunodeficiency virus (HIV) status of 283 homosexual men who were infected with HIV during a follow-up period between 1984 and 1991. The original design was to collect the measurements for all individuals semiannually. More details of the study design and medical implications can be found in Kaslow et al. [10]. Some authors have analyzed the same dataset using varying coefficient and semiparametric models; see for example Wu et al. [23], Huang et al. [9] and Fan and Li [7]. Their analysis aimed to describe the trend of the mean CD4 percentage depletion over time and to evaluate the effects of cigarette smoking, pre-HIV infection CD4 percentage and age at HIV infection on the mean CD4 percentage after the infection. The results of the hypothesis testing of Huang et al. [9] indicate that, at significance level 0.05, only the baseline function varies over time and preCD4 has a constant effect over time; neither smoking nor age has a significant impact on mean CD4 percentage. This motivates us to use model (1.1) for this dataset.

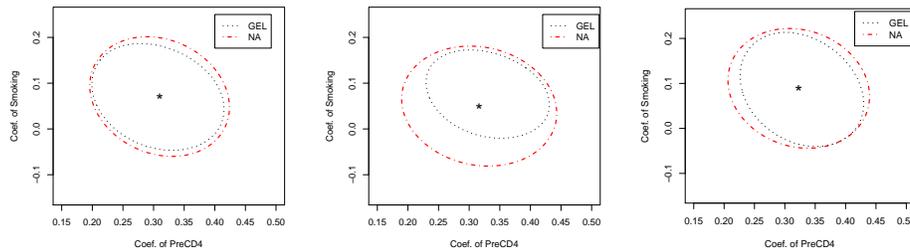
We considered two covariates: X_{1ij} , the individual's smoking status, which is taken to be 1 if the individual ever smoked cigarettes or 0 if never smoked cigarettes after HIV infection; and X_{2ij} , the centered variable for pre-infection CD4 percentage. For the purpose of demonstration and simplicity, the possible effects of other available covariates are omitted. The response variable $Y(t_{ij})$ is the individual's CD4 percentage measured, and both X_{1ij} and X_{2ij} are independent. We assume that observation times are independent of covariates because they are not significantly related to the two covariates. The X_{2ij} are measured with error (Liang et al. [11]), We consider the following semiparametric regression model:

$$Y(t_{ij}) = \beta_1 X_{1ij} + \beta_2 X_{2ij} + g(t_{ij}) + \varepsilon(t_{ij}), W_{2ij} = X_{2ij} + U_{2ij},$$

where W_{2ij} are the observed CD4 cell counts, and $g(t_{ij})$, the baseline CD4 percentage, represents the mean CD4 percentage t years after infection for a non-smoker with average pre-infection CD4 percentage, and β_1 and β_2 describe the effects for cigarette smoking and pre-infection CD4 percentage, respectively, on the post-infection CD4 percentage. We assumed that the measurement errors U_{2ij} are independent and normally distributed with mean zero and variance σ_{uu}^2 . as in Yang et al.[25], we conduct a sensitivity analysis by taking $\sigma_{uu}^2 = 0$, which naively ignores measurement error, $\sigma_{uu}^2 = 0.068$ and $\sigma_{uu}^2 = 0.154$.

FIGURE 3

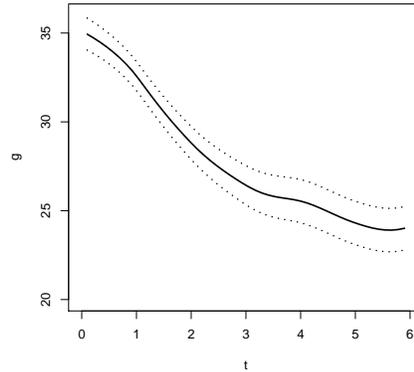
Figure 3:AIDS study: Estimates of (β_1, β_2) and the averages of 95% confidence regions based on the normal approximation (NA) and residual-adjusted generalized empirical likelihood (GEL). The left panel for $\sigma_{uu}^2 = 0$, the middle panel for $\sigma_{uu}^2 = 0.068$ and the right panel for $\sigma_{uu}^2 = 0.135$.



We computed the generalized empirical likelihood ratios for (β_1, β_2) under first-order autoregressive correlation matrix, and the estimators for $g(t)$ by using the Epanechnikov kernel and the cross-validated bandwidth $h_{cv} = 0.38$. The 95% confidence regions for (β_1, β_2) reported in Figure 3, and it shows that the generalized empirical likelihood again works better than the normal approximation. For $\sigma_{uu}^2 = 0$, $\sigma_{uu}^2 = 0.068$ and $\sigma_{uu}^2 = 0.135$, the estimated values of β_2 are 0.3063, 0.3248, 0.3456, respectively. As expected, we find a somewhat stronger positive association between the pre-infection CD4 percentage and the percentage of CD4 cells when the possibility of measurement error is taken into account.

FIGURE 4

Figure 4: AIDS study: The solid curve represents the estimated curve of the baseline function and the dotted line indicate the 95% pointwise confidence intervals for baseline function, based on the residual-adjusted empirical likelihood.



The curve of the estimated baseline function and the corresponding 95% pointwise confidence intervals for the case of $\sigma_{uu}^2 = 0$ is shown in Figure 4. The results for the other cases of $\sigma_{uu}^2 = 0.143$ and $\sigma_{uu}^2 = 0.194$ are similar and are therefore not shown. From Figure 4, we find that the mean baseline CD4 percentage for the population decreases rather quickly at the beginning of HIV infection, but the rate of decrease appears to be slowing down four years after the infection. The findings basically agree with that which was discovered by the local linear fitting method of Fan and Li [7] and Xue and Zhu [24].

Acknowledgments. The first author is supported by the National Natural Science Foundation of China (11701156), Statistical Science Research Project of China (2016LY25) and the Doctoral Startup Foundation of Henan Normal University (qd17127). The second author is supported by the National Natural Science Foundation of China (11571025, Key grant: 11331011), and the Beijing Natural Science Foundation (1182002). The third author is supported by the National Natural Science Foundation of China (11526188), the Zhejiang Provincial Natural Science Foundation of China (LQ15A010008), Startup Foundation for Talents in Zhejiang Agriculture and Forestry University (2014FR085) and Statistical Research Project of Zhejiang Province.

6. Appendix.

For convenience and simplicity, let c denote a positive constant that may be different at each appearance throughout this paper. Before we state one of the main results, we note the following regularity conditions.

- C1. The bandwidth satisfies $h = h_0 N^{-1/5}$ for some constant $h_0 > 0$.
- C2. The kernel $K(\cdot)$ is a symmetric probability density function, and is twice continuously differentiable on its support set $[-1, 1]$.
- C3. $\sup_{x, 0 \leq t \leq 1} E(\varepsilon_{ij}^4 | X_{ijr} = x, T_{ij} = t) < \infty$, $\sup_{0 \leq t \leq 1} E(U_{ij}^4 | T_{ij} = t) < \infty$, $E(X_{ijr}^4) < \infty$ for $i = 1, \dots, n$, $j = 1, \dots, n_i$, and $r = 1, \dots, p$, where X_{ijr} is the r th component of X_{ij} .
- C4. The density function of T_{ij} , $f(t)$ is bounded away from zero and infinity uniformly over $[0, 1]$, and is twice continuously differentiable on $(0, 1)$.

- C5. $g(t)$ and $m_{X,r}(t)$ are twice continuously differentiable on $(0, 1)$ for all $r = 1, \dots, p$, where $m_{X,r}(t)$ is the r th component of $m_X(t)$.
- C6. The variance function $\sigma_{\varepsilon,u}^2(t)$ is continuous at t_0 .

Remark. C1 – C6 are the common conditions used in the literature. C1 ensures that undersmoothing \hat{g} is not needed so that we can use data-driven approach to select the bandwidth. In C2, the compaction by using kernels with small tails; for example, the standard Gaussian kernel. C3 is a necessary moment condition. Smooth conditions C4 and C5 are standard conditions for nonparametric. C6 is a regularity condition. The proofs of Theorems 2.1 and 2.2 rely on the following some lemmas.

6.1. Lemma. *Suppose that conditions C1-C6 hold. Then, for any constants a and b with $0 < a < b < 1$, we have*

$$(6.1) \quad \begin{aligned} \sup_{a \leq t \leq b} E[\|m_W(T_{ij}) - \hat{m}_W(T_{ij})\|^2 | T_{ij} = t] &= O(n^{-1}h^{-1} + h^4) \\ \sup_{a \leq t \leq b} E[\|g(T_{ij}) - \hat{g}_*(T_{ij})\|^2 | T_{ij} = t] &= O(n^{-1}h^{-1} + h^4), \end{aligned}$$

where $\hat{g}_*(t) = \hat{m}_Y(t) - \hat{m}_W^T(t)\beta$.

The proof of Lemma 6.1 is similar to that of Xue and Zhu [24] and we omit the details.

6.2. Lemma. *Suppose that the regularity conditions C1-C6 hold. If β is the true parameter, then*

$$(6.2) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{Z}_i(\beta) \xrightarrow{\mathcal{L}} N(0, \Sigma),$$

where Σ is defined by (2.12).

Proof. Consider the k th ($k = 1, \dots, s$) block of $\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{Z}_i(\beta)$:

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{W}_i^T A_i^{-1/2} M_k A_i^{-1/2} (\hat{Y}_i - \hat{W}_i \beta) + D_i^{(k)} \beta \\ = & \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \tilde{X}_i^T A_i^{-1/2} M_k A_i^{-1/2} (\Psi_i \varepsilon_i - U_i \beta) + U_i^T A_i^{-1/2} M_k A_i^{-1/2} \Psi_i \varepsilon_i \right. \\ & \left. - U_i A_i^{-1/2} M_k A_i^{-1/2} U_i \beta + D_i^{(k)} \beta \right\} \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \tilde{X}_i^T A_i^{-1/2} M_k A_i^{-1/2} \Psi_i [g(T_i) - \hat{g}_*(T_i)] \right\} \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ [m_W(T_i) - \hat{m}_W(T_i)]^T \Psi_i A_i^{-1/2} M_k A_i^{-1/2} \Psi_i \varepsilon_i \right\} \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ U_i^T A_i^{-1/2} M_k A_i^{-1/2} \Psi_i [g(T_i) - \hat{g}_*(T_i)] \right\} \\ & - \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ [m_W(T_i) - \hat{m}_W(T_i)]^T \Psi_i A_i^{-1/2} M_k A_i^{-1/2} U_i \beta \right\} \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ [m_W(T_i) - \hat{m}_W(T_i)]^T \Psi_i A_i^{-1/2} M_k A_i^{-1/2} \Psi_i [(g(T_i) - \hat{g}_*(T_i))] \right\} \end{aligned}$$

$$\equiv J_1 + J_2 + J_3 + J_4 + J_5 + J_6.$$

We first deal with J_1 . Denote $J_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_{ik}$, it is easy to obtain $E\zeta_{ik} = 0$ and

$$\begin{aligned} \text{Cov}(\zeta_{ik}) &= E(\tilde{X}_i^T A_i^{-1/2} M_k A_i^{-1/2} (\Psi_i \varepsilon_i - U_i \beta))^{\otimes 2} \\ &\quad + E(U_i^T A_i^{-1/2} M_k A_i^{-1/2} \Psi_i \varepsilon_i)^{\otimes 2} \\ &\quad + E(-U_i A_i^{-1/2} M_k A_i^{-1/2} U_i \beta + D_i^{(k)} \beta)^{\otimes 2} + o_p(1) \\ &= \Sigma_{kk} + o_p(1). \end{aligned}$$

Next, we need to prove $J_v \xrightarrow{P} 0$, $v = 2, 3, 4, 5, 6$. We first deal with J_2 . Let c_{ik}^{jv} denote the (j, v) th element of $A_i^{-1/2} M_k A_i^{-1/2}$. Similar to the proof of (6.1), we can get that

$$(6.3) \quad \sup_{a \leq t \leq b} E \left[\left(g(T_{ij}) - \sum_{k=1}^n \sum_{l=1}^{n_k} \omega_{kl}(T_{ij}) g(T_{kl}) \right)^2 \middle| T_{ij} = t \right] = O(n^{-1}h + h^4),$$

Let $J_{2,k}$ and $\tilde{X}_{ij,k}$ denote the k th ($k = 1, \dots, p$) component of J_2 and \tilde{X}_{ij} . From (6.3) and Conditions C1 – C4 we have

$$\begin{aligned} E(J_{2,k}^2) &\leq 2n^{-1} E \left[\sum_{k=1}^n \sum_{l=1}^{n_k} \left(\sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{v=1}^{n_i} \psi(T_{ij}) \omega_{kl}(T_{ij}) \tilde{X}_{ij,k} c_{ik}^{jv} \right) \varepsilon_{kl} \right]^2 + 2n^{-1} \\ &\quad E \left[\sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{v=1}^{n_i} \psi(T_{ij}) \tilde{X}_{ij,k} c_{ik}^{jv} \left(g(T_{ij}) - \sum_{k=1}^n \sum_{l=1}^{n_k} \omega_{kl}(T_{ij}) g(T_{kl}) \right) \right]^2 \\ &\leq c[(nh)^{-1} + h^4] \rightarrow 0. \end{aligned}$$

Hence, we have $J_2 \rightarrow 0$. Then, we consider J_3 .

$$\begin{aligned} J_3 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ [m_W(T_i) - \hat{m}_W(T_i)]^T \Psi_i A_i^{-1/2} M_k A_i^{-1/2} \Psi_i \varepsilon_i \right\} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{v=1}^{n_i} \psi^2(T_{ij}) [m_W(T_{ij}) - \hat{m}_W(T_{ij})]^T c_{ik}^{jv} \varepsilon_{ij}. \end{aligned}$$

From Lemma 6.1 and C5 we have

$$\begin{aligned} E(\|J_3\|^2) &\leq cn^{-1} \sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{v=1}^{n_i} E \left\{ \psi(T_{ij}) (c_{ik}^{jv})^2 E(\|m_w(T_{ij}) - \hat{m}_w(T_{ij})\|^2 | T_{ij}) \right\} \\ &\leq c[(nh)^{-1} + h^4] \rightarrow 0. \end{aligned}$$

Similarly, we can get $J_4 \xrightarrow{P} 0, J_5 \xrightarrow{P} 0$. From Lemma 6.1 and Cauchy-Schwarz inequality, we have

$$\begin{aligned} E(\|J_6\|) &\leq \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{v=1}^{n_i} E \left\{ \psi(T_{ij}) c_{ik}^{jv} [E(\|m_w(T_{ij}) - \hat{m}_w(T_{ij})\|^2 | T_{ij})]^{1/2} \right. \\ &\quad \left. [E(\|g(T_{ij}) - \hat{g}_*(T_{ij})\|^2 | T_{ij})]^{1/2} \right\} \\ &\leq c\sqrt{n}[(nh)^{-1} + h^4] \rightarrow 0. \end{aligned}$$

This implies $J_6 \xrightarrow{P} 0$. So we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{Z}_i(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_i + o(1),$$

where $\varsigma_i = (\zeta_{i1}^T, \dots, \zeta_{is}^T)^T$. Obviously, $E(\zeta_i) = 0$ and

$$\frac{1}{n} \sum_{i=1}^n \text{Cov}(\zeta_i) = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1s} \\ \vdots & \ddots & \vdots \\ \Sigma_{s1} & \cdots & \Sigma_{ss} \end{pmatrix}.$$

The proof of Lemma 6.2 is completed. \square

6.3. Lemma. *Suppose that Conditions C1-C6 holds. If β is the true parameter, then*

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \hat{Z}_i(\beta) \hat{Z}_i(\beta)^T \xrightarrow{\mathcal{P}} \Sigma.$$

Proof. We also use the notations in the proof of Lemma 6.2, and denote $\hat{Z}_{i,k}(\beta)$ is the k th ($k = 1, \dots, s$) block of $\hat{Z}_i(\beta)$. A simple calculation yields

$$\hat{Z}_{i,k}(\beta) = \zeta_{ik} + I_{1k} + I_{2k}, \hat{Z}_{i,m}(\beta) = \zeta_{im} + I_{1m} + I_{2m},$$

where

$$\begin{aligned} I_{1k} &= (\tilde{X}_i + U_i)^T A_i^{-1/2} M_k A_i^{-1/2} \Psi_i [g(T_i) - \hat{g}_*(T_i)], \\ I_{2k} &= [m_W(T_i) - \hat{m}_W(T_i)]^T \Psi_i A_i^{-1/2} M_k A_i^{-1/2} \\ &\quad [(\Psi_i \varepsilon_i - U_i \beta) + \Psi_i (g(T_i) - \hat{g}_*(T_i))], \\ I_{1m} &= (\tilde{X}_i + U_i)^T A_i^{-1/2} M_m A_i^{-1/2} \Psi_i [g(T_i) - \hat{g}_*(T_i)], \\ I_{2m} &= [m_W(T_i) - \hat{m}_W(T_i)]^T \Psi_i A_i^{-1/2} M_m A_i^{-1/2} \\ &\quad [(\Psi_i \varepsilon_i - U_i \beta) + \Psi_i (g(T_i) - \hat{g}_*(T_i))]. \end{aligned}$$

Then, consider the (k, m) th block of $\hat{\Sigma}$, $k, m = 1, \dots, s$,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \hat{Z}_{i,k}(\beta) \hat{Z}_{i,m}(\beta)^T &= \frac{1}{n} \sum_{i=1}^n \zeta_{ik} \zeta_{im}^T + \frac{1}{n} \sum_{i=1}^n I_{1k} I_{1m}^T + \frac{1}{n} \sum_{i=1}^n I_{2k} I_{1m}^T \\ &\quad + \frac{1}{n} \sum_{i=1}^n I_{1k} I_{2m}^T + \frac{1}{n} \sum_{i=1}^n I_{2k} I_{2m}^T + \frac{1}{n} \sum_{i=1}^n \zeta_{ik} I_{1m}^T \\ &\quad + \frac{1}{n} \sum_{i=1}^n \zeta_{ik} I_{2m}^T + \frac{1}{n} \sum_{i=1}^n I_{1k} \zeta_{im}^T + \frac{1}{n} \sum_{i=1}^n I_{2k} \zeta_{im}^T \\ &= U_1 + U_2 + U_3 + U_4 + U_5 + U_6 + U_7 + U_8 + U_9. \end{aligned}$$

By the law of large numbers, we can derive that $U_1 \xrightarrow{\mathcal{P}} \Sigma_{km}$. Thus, if we can prove $U_\nu \xrightarrow{\mathcal{P}} 0$, $\nu = 2, \dots, 9$. For U_2 , let $U_{2,rq}$ denote its (r, q) element, and I_{ikr}, I_{imr} denote the r th component of I_{1k}, I_{1m} , respectively. We may use the Cauchy-Schwarz inequality to get

$$|U_{2,rq}| \leq \left(\frac{1}{n} \sum_{i=1}^n I_{ikr}^2 \right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^n I_{imq}^2 \right)^{1/2}.$$

By Lemma 6.1, we can derive that $\frac{1}{n} \sum_{i=1}^n I_{ikr}^2 \xrightarrow{\mathcal{P}} 0, \frac{1}{n} \sum_{i=1}^n I_{imq}^2 \xrightarrow{\mathcal{P}} 0$. Then, we get

$U_2 \xrightarrow{\mathcal{P}} 0$. Similarly, we can prove that $U_\nu \xrightarrow{\mathcal{P}} 0, \nu = 3, \dots, 9$. This completes the proof of Lemma 6.3. \square

6.4. Lemma. *Suppose that the regularity conditions C1 – C6 hold. If β is the true parameter, then*

$$(6.4) \quad \max_{1 \leq i \leq n} \|\hat{Z}_i(\beta)\| = o_P(n^{1/2}),$$

$$(6.5) \quad \|\lambda\| = O_P(n^{-1/2}).$$

Proof. According to the definition of $\hat{Z}_i(\beta)$ and Lemma 6.2, For the k th ($k = 1, \dots, s$) block of $\hat{Z}_i^{(k)}(\beta)$ we have

$$\begin{aligned} \max_{1 \leq i \leq n} \|\hat{Z}_i^{(k)}(\beta)\| &\leq \max_{1 \leq i \leq n} \|\tilde{X}_i^T A_i^{-1/2} M_k A_i^{-1/2} (\Psi_i \varepsilon_i - U_i \beta) + U_i^T A_i^{-1/2} M_k \\ &\quad A_i^{-1/2} \Psi_i \varepsilon_i - U_i A_i^{-1/2} M_k A_i^{-1/2} U_i \beta + D_i^{(k)} \beta\| \\ &\quad + \max_{1 \leq i \leq n} \|(\tilde{X}_i + U_i)^T A_i^{-1/2} M_k A_i^{-1/2} \Psi_i [g(T_i) - \hat{g}_*(T_i)]\| \\ &\quad + \max_{1 \leq i \leq n} \|[m_W(T_i) - \hat{m}_W(T_i)]^T \Psi_i A_i^{-1/2} M_k A_i^{-1/2} \\ &\quad [(\Psi_i \varepsilon_i - U_i \beta) + \Psi_i (g(T_i) - \hat{g}_*(T_i))]\| \\ &\equiv M_1 + M_2 + M_3. \end{aligned}$$

From Lemma 11.2 in Owen [17], we can obtain that $M_1 = o_P(n^{1/2})$. By Lemma 6.1 and Cauchy-Schwarz inequality, $M_v = o_P(1)$, $v = 2, 3$. This proves the first equation.

By Lemma 6.2, 6.3 and using the same arguments that are used in the proof of (2.14) in Owen [16], we can prove the second equation. Then the proof follows. \square

Proof. Proof of the Theorem 2.1. Applying the Taylor expansion to (2.9), and invoking lemmas, we get that

$$(6.6) \quad \hat{l}(\beta) = 2 \sum_{i=1}^n \{\lambda^T \hat{Z}_i(\beta) - [\lambda^T \hat{Z}_i(\beta)]^2 / 2\} + o_P(1).$$

by (2.10), it follows that

$$\begin{aligned} 0 &= \sum_{i=1}^n \frac{\hat{Z}_i(\beta)}{1 + \lambda^T \hat{Z}_i(\beta)} \\ &= \sum_{i=1}^n \hat{Z}_i(\beta) - \sum_{i=1}^n \hat{Z}_i(\beta) \hat{Z}_i(\beta)^T \lambda + \sum_{i=1}^n \frac{\hat{Z}_i(\beta) [\lambda^T \hat{Z}_i(\beta)]^2}{1 + \lambda^T \hat{Z}_i(\beta)}. \end{aligned}$$

This together with Lemma 6.2-6.4 proves that

$$\sum_{i=1}^n [\lambda^T \hat{Z}_i(\beta)]^2 = \sum_{i=1}^n \lambda^T \hat{Z}_i(\beta) + o_P(1).$$

and

$$\lambda = \left\{ \sum_{i=1}^n \hat{Z}_i(\beta) \hat{Z}_i(\beta)^T \right\}^{-1} \sum_{i=1}^n \hat{Z}_i(\beta) + o_P(n^{-1/2}).$$

therefore, we have

$$(6.7) \quad \hat{l}(\beta) = \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{Z}_i(\beta) \right\}^T \hat{\Sigma}^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{Z}_i(\beta) \right\} + o_P(1).$$

This together with lemma 6.2 and 6.3 proves Theorem 2.1. \square

Proof. Proof of the Theorem 2.2. Applying the same argument as in the proof of Theorem 2.2 in Tian and Xue [22], we can prove that

$$\hat{\beta} - \beta = -[\hat{\Gamma}^T \hat{\Sigma}^{-1} \hat{\Gamma}]^{-1} \hat{\Gamma}^T \hat{\Sigma}^{-1} n^{-1} \sum_{i=1}^n \hat{Z}_i(\beta) + o_P(n^{-1/2}),$$

Similarly to the proof of Lemma 6.3, we can obtain $\hat{\Gamma} \xrightarrow{P} \Gamma$. Together this with Lemma 6.2, 6.4 and Slutsky’s Theorem, we can prove Theorem 2.2. \square

The proofs of Theorems 2.3 and 2.4 are similar to the proofs of Theorems 2.1 and 2.2, therefore, we omit their proofs.

6.5. Lemma. *Suppose that the regularity conditions C1 – C6 hold. If $g(t_0)$ is the true value of the baseline function, then*

$$\frac{1}{\sqrt{Nh}} \sum_{i=1}^n \hat{\eta}_i(g(t_0)) - b(t_0) \xrightarrow{\mathcal{L}} N(0, v^2(t_0)).$$

Proof. It is easy to see that

$$(6.8) \quad \frac{1}{\sqrt{Nh}} \sum_{i=1}^n \hat{\eta}_i(g(t_0)) - b(t_0) = S_1(t_0) + S_2(t_0) + S_3(t_0),$$

where

$$\begin{aligned} S_1(t_0) &= \frac{1}{\sqrt{Nh}} \sum_{i=1}^n \sum_{j=1}^{n_i} (\varepsilon_{ij} - U_{ij}^T \beta) K_h(T_{ij} - t_0), \\ S_2(t_0) &= \frac{1}{\sqrt{Nh}} \sum_{i=1}^n \sum_{j=1}^{n_i} [g(T_{ij}) - g(t_0)] K_h(T_{ij} - t_0), \\ S_3(t_0) &= \frac{1}{\sqrt{Nh}} \sum_{i=1}^n \sum_{j=1}^{n_i} K_h(T_{ij} - t_0) W_{ij}^T (\beta - \hat{\beta}). \end{aligned}$$

It is not difficult to prove $E[S_1(t_0)] = 0$ and $\text{var}[S_1(t_0)] = v^2(t_0) + o(1)$. We can check that $S_1(t_0)$ satisfies the conditions of the Cramer-Wold theorem and the Lindeberg condition (Serfling, [19]). Therefore, we get

$$(6.9) \quad S_1(t_0) \xrightarrow{\mathcal{L}} N(0, v^2(t_0)).$$

We can also prove that $\text{var}(S_2(t_0)) = o(1)$. Thus

$$(6.10) \quad S_2(t_0) \xrightarrow{P} 0.$$

By Lemma 6.1 and 6.2 we can get $S_3(t_0) = O_P(h^{1/2})$. This together with (6.8)-(6.10) proves Lemma 6.5. \square

6.6. Lemma. *Suppose that the regularity conditions C1 – C6 hold. If $g(t)$ is the true value of the baseline function, then*

$$\begin{aligned} \frac{1}{Nh} \sum_{i=1}^n \hat{\eta}_i^2(g(t_0)) &\xrightarrow{P} v^2(t_0), \\ \max_{1 \leq i \leq n} \|\hat{\eta}_i(g(t_0))\| &= o_P((Nh)^{1/2}). \end{aligned}$$

Using some arguments similar to those used in the proof of Lemma 6.3 and 6.4, we can prove Lemma 6.6. The proof is omitted.

6.7. Lemma. *Suppose that the regularity conditions C1-C6 hold. Then $\hat{b}(t_0) \xrightarrow{P} b(t_0)$.*

Proof. Denote $\varphi_{ij}(t_0) = [g(T_{ij}) - gt_0]K_h(T_{ij} - t_0)$ and $\hat{\varphi}_{ij}(t_0) = [\hat{g}(T_{ij}) - \hat{g}(t_0)]K_h(T_{ij} - t_0)$. Then, we have

$$\begin{aligned}\hat{b}(t_0) - b(t_0) &= \frac{1}{\sqrt{Nh}} \sum_{i=1}^n \sum_{j=1}^{n_i} [\hat{\varphi}_{ij}(t_0) - \varphi_{ij}(t_0)] \\ &\quad + \frac{1}{\sqrt{Nh}} \sum_{i=1}^n \sum_{j=1}^{n_i} [\varphi_{ij}(t_0) - (h/N)^{1/2}b(t_0)] \\ &= M_1(t_0) + M_2(t_0).\end{aligned}$$

From Conditions C1 – C4 and the Taylor expansion, we have

$$\hat{\varphi}_{ij}(t_0) - \varphi_{ij}(t_0) = \{[\hat{g}'(t_0) - g'(t_0)](T_{ij} - t_0) + o_P(|(T_{ij} - t_0)|^2)\}K_h(T_{ij} - t_0).$$

Using Conditions C1 and C2, we can prove that

$$\frac{1}{\sqrt{Nh}} \sum_{i=1}^n \sum_{j=1}^{n_i} (T_{ij} - t_0)^l K_h(T_{ij} - t_0) = O_P(1), l = 1, 2,$$

and $\hat{g}'(t_0) - g'(t_0) \xrightarrow{P} 0$. Therefore, we have $M_1(t_0) \xrightarrow{P} 0$. It is easy to prove $M_2(t_0) \xrightarrow{P} 0$. The proof of Lemma 6.7 is completed. \square

We now turn to prove Theorems 3.1 and 3.2.

Proof. Proof of the Theorem 3.1. By direct calculation, we can obtain

$$\sqrt{Nh}(\hat{g}(t_0) - g(t_0)) = \frac{1}{\sqrt{Nh}} \sum_{i=1}^n \hat{\eta}_i(g(t_0)) \Big/ \hat{f}(t_0) + o_P(1).$$

Note that $\hat{f}(t_0) \rightarrow f(t)$, almost surely. This together with Lemma 6.5 proves Theorem 3.1. \square

Proof. Proof of the Theorem 3.2. It can be shown by Lemma 6.7 and direction calculation that

$$(6.11) \quad \frac{1}{\sqrt{Nh}} \sum_{i=1}^n \hat{\eta}_i^* \{g(t_0)\} = \frac{1}{\sqrt{Nh}} \sum_{i=1}^n \hat{\eta}_i \{g(t_0)\} - b(t_0) + o_P(1),$$

$$(6.12) \quad \frac{1}{Nh} \sum_{i=1}^n \hat{\eta}_i^{*2} \{g(t_0)\} = \frac{1}{Nh} \sum_{i=1}^n \hat{\eta}_i^2 \{g(t_0)\} + o_P(1),$$

$$(6.13) \quad \max_{1 \leq i \leq n} |\hat{\eta}_i^* \{g(t_0)\}| = o_P\{(Nh)^{1/2}\}.$$

Similar to the proof of the Theorem 2.1, Theorem 3.2 can be proved by (6.11)-(6.13) and Lemma 6.6. \square

References

- [1] Arnold, S. F. *The Theory of Linear Models and Multivariate Analysis*, John Wiley and Sons, New York, 1981.
- [2] Bai, Y. and Zhu, Z. Y. and Fung, W. K. *Partial linear models for longitudinal data based on quadratic inference function*, Scandinavian Journal of Statistics, **35**(1), 104-118, 2008.
- [3] Carroll, R. J. and Ruppert, D. and Stefanski, L.A. *Measurement Error in Nonlinear Models*, Chapman and Hall, New York, 1995.
- [4] Crowder, M. *On the use of a working correlation matrix in using generalised linear models for repeated measures*, Biometrika, **82**(2), 407-410, 1995.
- [5] Cui, H. J. and Chen, S. X. *Empirical likelihoods confidence region for parameter in the errors-in-variables*, Journal of Multivariate Analysis, **84**(1), 101-115, 2003.

- [6] Dziak, J. J. and Li, R. Z. and Qu, A. *An overview on quadratic inference function approaches for longitudinal data*, New Developments in Biostatistic and Bioinformatics, 49-72, 2009.
- [7] Fan, J. Q. and Li, R. Z. *New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis*, Publications of the American Statistical Association, **99**(467), 710-723, 2004.
- [8] He, X. M. and Zhu, Z. Y. and Fung, W. K. *Estimation in a semiparametric model for longitudinal data with unspecified dependence structure*, Biometrika, **89**(3), 579-590, 2002.
- [9] Huang, J. Z. and Wu, C. O. and Zhou, L. *Varying-coefficient models and basis function approximations for the analysis of repeated measurements*, Biometrika, **89**(1), 111-128, 2002.
- [10] Kaslow, R. A. and Ostrow, D. G. and Detels, R. and Phair, J. P. and Polk, B. F. and Rinaldo, C. R. *The multicenter AIDS cohort study: rationale, organization and selected characteristics of the participants*, American Journal of Epidemiology, **126**(2), 310-318, 1987.
- [11] Liang, H. and Hardle, H. and Carrloo, R. J. *Estimation in a semiparametric partially linear error-in-variables model*, The Annals of Statistics, **27**(5), 1519-1535, 1999.
- [12] Liang, H. and Wang, S. and Carroll, R. J. *Partially linear models with missing response variables and error-prone covariates*, Biometrika, **94**(1), 185-198, 2007.
- [13] Liang, K-Y. and Zeger, S. L. *Longitudinal data analysis using generalized linear models*, Biometrika, **73**(1), 13-22, 1986.
- [14] Liu, Q. *Estimation of the linear EV model with censored data*, Journal of Statistical Planning and Inference, **141**(7), 2463-2471, 2011.
- [15] Owen, A. B. *Empirical likelihood ratio confidence intervals for a single functional*, Biometrika, **75**(2), 237-249, 1988.
- [16] Owen, A. B. *Empirical likelihood confidence regions*, The Annals of Statistics, **18**, 90-120, 1990.
- [17] Owen, A. B. *Empirical Likelihood*, Chapman and Hall, New York, 2001.
- [18] Qu, A. and Lindsay, B. G. and Li, B. *Improving generalised estimating equations using quadratic inference functions*, Biometrika, **87**(4), 823-836, 2000.
- [19] Serfling, R. *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.
- [20] Shao, J. and Xiao, Z. and Xu, R. *Estimation with unbalanced panel data having covariate measurement error*, Journal of Statistical Planning and Inference, **141**(2), 800-808, 2011.
- [21] Tian, R. Q. and Xue, L. G. *Variable selection for semiparametric errors-in variables regression model with longitudinal data*, Journal of Statistical Computation and Simulation, **84**(8), 1654-1669, 2014.
- [22] Tian, R. Q. and Xue, L. G. *Generalized empirical likelihood inference in generalized linear models for longitudinal data*, Communication in Statistics-Theory and Methods, **43**(18), 3893-3904, 2014.
- [23] Wu, C. O. and Chiang, C. T. and Hoover, D. R. *Asymptotic confidence regions for kernel smoothing of a varying coefficient model with longitudinal data*, Journal of the American Statistical Association, **93**(444), 1388-1402, 1998.
- [24] Xue, L. G. and Zhu, L. X. *Empirical likelihood semiparametric regression analysis for longitudinal data*, Biometrika, **94**(4), 921-937, 2007.
- [25] Yang, Y. P. and Xue, L. G. and Cheng, W. H. *Two-step estimators in partial linear models with missing response variables and error-prone covariates*, Journal of Systems Science and Complexity, **24**(6), 1165-1182, 2011.
- [26] Yang, Y. P. and Li, G. R. and Peng, H. *Empirical likelihood of varying coefficient errors-in-variables models with longitudinal data*, Journal of Multivariate Analysis, **127**(3), 1-18, 2014.
- [27] You, J. and Chen, G. and Zhou, Y. *Block empirical likelihood for longitudinal partially linear regression models*, Canadian Journal of Statistics, **34**(1), 79-96, 2006.
- [28] Zeger, S. L. and Diggle, P. J. *Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters*, Biometrics, **50**(3), 689-699, 1994.
- [29] Zhao, P. X. and Xue, L. G. *Empirical likelihood inferences for semiparametric varying-coefficient partially linear error-prone models with longitudinal data*, Journal of Nonparametric Statistics, **21**(7), 907-923, 2009.