Empirical likelihood inference for longitudinal data with covariate measurement errors: An application to the LEAN study

Yuexia Zhang
Department of Computer and Mathematical Sciences, University of Toronto
Guoyou Qin
Department of Biostatistics, Fudan University
Zhongyi Zhu
Department of Statistics, Fudan University
and
Jiajia Zhang
Department of Epidemiology and Biostatistics, University of South Carolina

Abstract

Measurement errors usually arise during the longitudinal data collection process. Ignoring the effects of measurement errors will lead to invalid estimates. The Lifestyle Education for Activity and Nutrition (LEAN) study was designed to assess the effectiveness of intervention for enhancing weight loss over nine months. The covariates systolic blood pressure (SBP) and diastolic blood pressure (DBP) were measured at baseline, month 4, and month 9. At each assessment time, there were two replicate measurements for SBP and DBP. The replicate measurement errors of SBP follow different distributions, as does DBP. To account for the distributional difference of replicate measurement errors, a new method for analyzing longitudinal data with replicate covariate measurement errors is developed based on the empirical likelihood method. The asymptotic properties of the proposed estimator are established under some regularity conditions. The confidence region for the parameters of interest can be constructed based on the chi-squared approximation without estimating the covariance matrix. Additionally, the proposed empirical likelihood estimator is asymptotically more efficient than the estimator of Lin et al. (2018). Extensive simulations demonstrate that the proposed method can eliminate the effects of measurement errors in the covariate and has a high estimation efficiency. The proposed method indicates the significant effect of the intervention on BMI in the LEAN study

Keywords: auxiliary random vector; distributional difference; efficiency; replicate measurement errors

1 Introduction

Longitudinal data are commonly seen in various fields, such as psychology, economics, social sciences, and public health, and measurement errors usually arise during the data collection process. The Lifestyle Education for Activity and Nutrition (LEAN) study (Barry et al., 2011) was designed to assess the effectiveness of intervention for enhancing weight loss over nine months in sedentary overweight or obese adults. In this study, 197 men and women between the ages of 18 and 64 who were underactive, overweight, or obese (BMI $\geq$ 25), and had access to the internet were randomly assigned to the standard care group and the intervention group. For each participant, systolic blood pressure (SBP) and diastolic blood pressure (DBP) were measured at baseline, month 4, and month 9. As pointed out by Qin et al. (2016a,b) and Lin et al. (2018), there exist measurement errors in the covariates SBP and DBP.

We denote the surrogate values of SBP as $\text{SBP}_{(1)}$ and $\text{SBP}_{(2)}$, and the corresponding measurement errors as $\xi_{(1)}$ and $\xi_{(2)}$. If we further assume the additive measurement error models for $\text{SBP}_{(1)}$ and $\text{SBP}_{(2)}$, then $\text{cSBP}_{(1)} \triangleq \text{SBP}_{(1)} - (\text{SBP}_{(1)} + \text{SBP}_{(2)})/2 = (\xi_{(1)} - \xi_{(2)})/2$. If $\xi_{(1)}$ and $\xi_{(2)}$ follow the same distribution, then the density function of $\text{cSBP}_{(1)}$ is symmetric. Similarly, we denote one of the centralized surrogate values of DBP as $\text{cDBP}_{(1)}$. Figure 1 displays the density functions of $\text{cSBP}_{(1)}$ and $\text{cDBP}_{(1)}$, which illustrates that the density functions of $\text{cSBP}_{(1)}$ and $\text{cDBP}_{(1)}$ are not symmetric. We further find that the density functions of $\text{cSBP}_{(1)}$ and $\text{cDBP}_{(1)}$ are significantly asymmetric at the significance level of 0.05 based on the D'Agostino skewness test statistic (D'Agostino, 1970). Therefore, the replicate measurement errors of SBP follow different distributions, as does DBP. However, few existing methods have accounted for this distributional difference in measurement errors. The main purpose of this paper is to develop a new longitudinal data analysis method which can account for the distributional difference of replicate measurement errors.

The likelihood-based method and estimating equation method are the most popular methods in longitudinal data analysis (Laird and Ware, 1982; Liang and Zeger, 1986; Diggle, 2002; Zhang et al., 2015; Cheng et al., 2016; Funatogawa and Funatogawa, 2018). The likelihood-based method is generally efficient but sensitive to the distribution misspecification, because

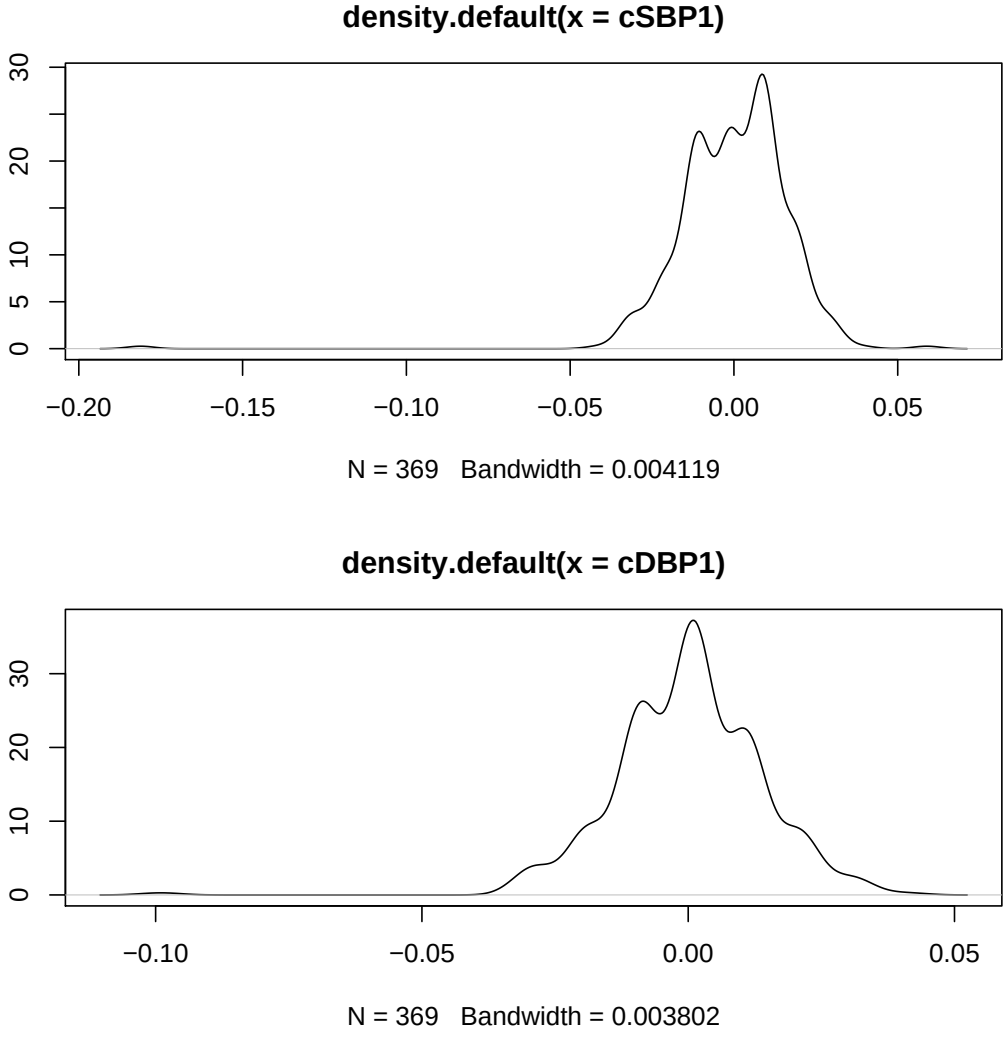

Figure 1: Density functions of $\text{cSBP}_{(1)}$ and $\text{cDBP}_{(1)}$.

it always assumes the joint distribution of repeated observations for each subject and applies the maximum likelihood estimation (MLE) method or restricted maximum likelihood estimation (REML) method to estimate. The estimating equation method avoids making assumptions for the multivariate distribution by specifying the first two moments of response. However, the estimating equation method cannot deal with the problem where the number of estimating functions is larger than the number of parameters. The empirical likelihood method (Owen, 1988), a combination of the likelihood-based method and estimating equation method, has attracted much attention recently (Wang et al., 2010; Qiu and Wu, 2015; Zhao et al., 2019; Hu and Xu, 2022). The empirical likelihood method is nonparametric, distribution-free, and also enjoys some good properties of the parametric likelihood method. For example, the empirical likelihood ratio statistic asymptotically follows a chi-squared distribution (Owen, 1990, 2001). At the same time, it can deal with the problem where there are more estimating functions than parameters. Besides, it can combine the information in the estimating functions in a most efficient way (Qin and Lawless, 1994).

Although there is considerable literature on how to deal with measurement errors using the likelihood-based method and estimating equation method (Wulfsohn and Tsiatis, 1997; Wang and Sullivan, 2000; Wu, 2002; Hsieh et al., 2006; Wang, 2006; Qin et al., 2016a; Li et al., 2019), the empirical likelihood method is not widely used in analysing longitudinal data with measurement errors. Zhao and Xue (2009) investigated the empirical likelihood inference for semiparametric varying-coefficient partially linear error-in-variables models, where they applied the correction for attenuation technique to construct a bias-corrected auxiliary random vector. However, their method needs to make some assumptions about the covariance matrix of measurement errors, which may not be satisfied in practice.

For the LEAN study, Lin et al. (2018) constructed an unbiased estimating equation using the independence between replicate measurement errors to eliminate the effects of measurement errors. Although their method is asymptotically more efficient than the method of Qin et al. (2016a), their method may lose some efficiency if the distributions of replicate measurement errors are different. Therefore, it is important to consider the distributional difference of measurement errors to get a more efficient estimator. In this paper, we propose a new empirical likelihood-based method. The proposed estimator is asymptotically more

efficient than the estimator of Lin et al. (2018). In addition, the proposed method can deal with the problem where there are more than two replicate measurements at each assessment time.

The remainder of this paper is organized as follows. In Section 2, we introduce the mean model and measurement error process. The proposed empirical likelihood-based method is outlined in Section 3 and some asymptotic properties are established in Section 4. We assess the performance of the proposed method with simulation studies in Section 5 and apply the proposed method to the LEAN data set in Section 6. The paper is concluded with a discussion in Section 7. The detailed proofs are given in the appendices. The R codes for simulation studies are available on the RunMyCode website.

2 Model specification

2.1 Mean model

In this paper, we consider a longitudinal study with n subjects and m_i observations over time for the ith subject. Let Y_{ij} be the response variable and $\boldsymbol{X}_{ij} = (X_{ij,1}, \cdots, X_{ij,p})^\top$ be the vector of covariates, where $i = 1, \cdots, n$, $j = 1, \cdots, m_i$, and $\{m_i, i = 1, \cdots, n\}$ are bounded positive integers. Assume the longitudinal data set follows the linear regression model, i.e.,

$$Y_{ij} = \boldsymbol{X}_{ij}^\top \boldsymbol{\beta}_0 + \varepsilon_{ij}, \quad i = 1, \cdots, n, \quad j = 1, \cdots, m_i, \tag{1}$$

where $\boldsymbol{\beta}_0 = (\beta_{01}, \cdots, \beta_{0p})^\top$ is a p-dimensional unknown vector and ε_{ij} is the random error term. In matrix form, we denote $\boldsymbol{Y}_i = (Y_{i1}, \cdots, Y_{im_i})^\top$, $\mathbf{X}_i = (\boldsymbol{X}_{i1}, \cdots, \boldsymbol{X}_{im_i})^\top$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{im_i})^\top$, where $\{\boldsymbol{\varepsilon}_i, i = 1, \ldots, n\}$ are mutually independent with $\mathrm{E}(\boldsymbol{\varepsilon}_i | \mathbf{X}_i) = \mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_i$ for each $i \in \{1, \ldots, n\}$.

2.2 Measurement error process

Let $\boldsymbol{X}_{ij}$ denote the covariate vector measured with error, and $\boldsymbol{W}_{ij}$ denote its observed version. Then, we assume that $\boldsymbol{X}_{ij}$ and $\boldsymbol{W}_{ij}$ follow a classical additive measurement error model, i.e.,

$$\boldsymbol{W}_{ij} = \boldsymbol{X}_{ij} + \boldsymbol{\xi}_{ij},$$

where $\boldsymbol{\xi}_{ij}$ is the measurement error with mean zero, and $\boldsymbol{\xi}_{ij}$ is independent of $\mathbf{X}_i$ and $\boldsymbol{\varepsilon}_i$.

In practice, replicate measurements of $\boldsymbol{X}_{ij}$ are often conducted to get more reliable results. We assume that there exist K ($K \geq 2$) replicate measurements for the error-prone covariate $\boldsymbol{X}_{ij}$, i.e.,

$$\boldsymbol{W}_{ij(k)} = \boldsymbol{X}_{ij} + \boldsymbol{\xi}_{ij(k)}, \quad k = 1, \cdots, K,$$

where $\{\boldsymbol{\xi}_{ij(k)}, k = 1, \cdots, K\}$ are mutually independent and the distributions of measurement errors $\{\boldsymbol{\xi}_{ij(k)}, k = 1, \cdots, K\}$ can be different. For convenience, we denote $\mathbf{W}_{i(k)} = (\boldsymbol{W}_{i1(k)}, \cdots, \boldsymbol{W}_{im_i(k)})^\top$, $\boldsymbol{W}_{ij(k)} = (W_{ij(k),1}, \cdots, W_{ij(k),p})^\top$, and $\boldsymbol{\xi}_{i(k)} = (\boldsymbol{\xi}_{i1(k)}, \cdots, \boldsymbol{\xi}_{im_i(k)})^\top$ for $k = 1, \cdots, K$.

3 Proposed method

When there are two replicate measurements for $\boldsymbol{X}_{ij}$, Lin et al. (2018) proposed the following estimating equation for estimation of $\boldsymbol{\beta}_0$

$$\sum_{i=1}^{n} \left\{ \mathbf{W}_{i(1)}^\top \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \mathbf{W}_{i(2)}\boldsymbol{\beta}) + \mathbf{W}_{i(2)}^\top \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \mathbf{W}_{i(1)}\boldsymbol{\beta}) \right\} = \mathbf{0}.$$

When there are more than two measurements, we can extend Lin et al. (2018)'s method directly and use the following estimating equation

$$\sum_{i=1}^{n} \sum_{k_1 \neq k_2} \mathbf{W}_{i(k_1)}^\top \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \mathbf{W}_{i(k_2)}\boldsymbol{\beta}) = \mathbf{0}, \tag{2}$$

where $k_1, k_2 \in \{1, \ldots, K\}$, and $K > 2$. Denote the solution to (2) as $\hat{\boldsymbol{\beta}}_{LIN}$.

However, the estimating equation (2) does not consider the heterogeneity of different measurement errors, because it gives the same weight to all the estimating functions $\sum_{i=1}^{n} \mathbf{W}_{i(k_1)}^\top \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \mathbf{W}_{i(k_2)}\boldsymbol{\beta})$, where $k_1 \neq k_2$. As a result, $\hat{\boldsymbol{\beta}}_{LIN}$ may not be highly efficient. To improve the estimation efficiency, we propose to estimate $\boldsymbol{\beta}_0$ based on the empirical likelihood method.

First, we introduce an auxiliary random vector as follows

$$\boldsymbol{g}_i(\boldsymbol{\beta}) = \begin{pmatrix} \mathbf{W}_{i(1)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(2)}\boldsymbol{\beta}) \\ \mathbf{W}_{i(2)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(1)}\boldsymbol{\beta}) \\ \vdots \\ \mathbf{W}_{i(K-1)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(K)}\boldsymbol{\beta}) \\ \mathbf{W}_{i(K)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(K-1)}\boldsymbol{\beta}) \end{pmatrix}. \tag{3}$$

Because of the independence between replicate measurement errors, the auxiliary random vector has expectation zero if $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Thus, the effects of measurement errors can be eliminated. However, the elements in $\boldsymbol{g}_i(\boldsymbol{\beta})$ are not functionally independent in all situations. As illustrated in a toy example in Appendix A, there exist some duplicate elements in $\boldsymbol{g}_i(\boldsymbol{\beta})$. Besides, there is an inner relationship among the elements of $\boldsymbol{g}_i(\boldsymbol{\beta})$. Both factors make the matrix $\mathrm{E}\{\boldsymbol{g}_i(\boldsymbol{\beta}_0)\boldsymbol{g}_i(\boldsymbol{\beta}_0)^\top\}$ not positive definite. However, positive definiteness of the matrix $\mathrm{E}\{\boldsymbol{g}_i(\boldsymbol{\beta}_0)\boldsymbol{g}_i(\boldsymbol{\beta}_0)^\top\}$ is one of the necessary conditions for the asymptotic normality of the empirical likelihood estimator, as shown in Section 4. Therefore, we need to eliminate the elements which are functionally dependent or have inner relationships with other elements in $\boldsymbol{g}_i(\boldsymbol{\beta})$ from $\boldsymbol{g}_i(\boldsymbol{\beta})$. Denote the reduced auxiliary random vector as $\boldsymbol{g}_i^*(\boldsymbol{\beta})$ and assume the dimension of $\boldsymbol{g}_i^*(\boldsymbol{\beta})$ is q, where $\boldsymbol{g}_i^*(\boldsymbol{\beta})$ satisfies the condition that $\mathrm{E}\{\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)^\top\}$ is a positive definite matrix. We illustrate how to obtain the reduced auxiliary random vector by using the toy example, which is provided in Appendix A. In general, the reduced auxiliary random vector $\boldsymbol{g}_i^*(\boldsymbol{\beta})$ can be obtained based on Algorithm 1.

Second, following the standard procedure for the empirical likelihood method, we define the profile empirical likelihood ratio function as

$$R(\boldsymbol{\beta}) = \max\left\{ \prod_{i=1}^n (n\pi_i) \;\middle|\; \pi_i \geq 0, \sum_{i=1}^n \pi_i = 1, \sum_{i=1}^n \pi_i \boldsymbol{g}_i^*(\boldsymbol{\beta}) = \mathbf{0} \right\}. \tag{4}$$

Using the Lagrange multiplier method, $R(\boldsymbol{\beta})$ is maximized at

$$\pi_i = \frac{1}{n\{1 + \boldsymbol{\lambda}^\top \boldsymbol{g}_i^*(\boldsymbol{\beta})\}}, \quad i = 1, \cdots, n, \tag{5}$$

where the Lagrange multiplier $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_q)^\top$ satisfies the following condition

$$\frac{1}{n}\sum_{i=1}^n \frac{\boldsymbol{g}_i^*(\boldsymbol{\beta})}{1 + \boldsymbol{\lambda}^\top \boldsymbol{g}_i^*(\boldsymbol{\beta})} = \mathbf{0}. \tag{6}$$

Algorithm 1 The proposed procedure for obtaining the reduced auxiliary random vector $\boldsymbol{g}_i^*(\boldsymbol{\beta})$

1. Write the complete formula of $\boldsymbol{g}_i(\boldsymbol{\beta})$ based on (3).
2. Check whether there are some duplicate elements in $\boldsymbol{g}_i(\boldsymbol{\beta})$. If there are some duplicate elements, then keep the unique elements and eliminate the duplicate elements from $\boldsymbol{g}_i(\boldsymbol{\beta})$. Denote the reduced random vector as $\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta})$; if there is no duplicate element, then let $\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta}) = \boldsymbol{g}_i(\boldsymbol{\beta})$.
3. Write the complete formula of $\mathrm{E}\{\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta}_0)\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta}_0)^\top\}$ based on model assumptions.
4. Check whether there are some elements in $\mathrm{E}\{\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta}_0)\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta}_0)^\top\}$ that can be represented as a linear function of other elements in $\mathrm{E}\{\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta}_0)\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta}_0)^\top\}$. If it is true, then eliminate the corresponding elements in $\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta})$ from $\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta})$. Denote the reduced auxiliary random vector as $\boldsymbol{g}_i^*(\boldsymbol{\beta})$; if it is false, then let $\boldsymbol{g}_i^*(\boldsymbol{\beta}) = \tilde{\boldsymbol{g}}_i(\boldsymbol{\beta})$.

Based on (4) and (5), we have

$$-2\log R(\boldsymbol{\beta}) = -2\log\left[\prod_{i=1}^{n}\left\{1+\boldsymbol{\lambda}^\top \boldsymbol{g}_i^*(\boldsymbol{\beta})\right\}^{-1}\right] = 2\sum_{i=1}^{n}\log\left\{1+\boldsymbol{\lambda}^\top \boldsymbol{g}_i^*(\boldsymbol{\beta})\right\}. \tag{7}$$

The maximum empirical likelihood estimator (MELE) of $\boldsymbol{\beta}_0$, $\hat{\boldsymbol{\beta}}$, can be obtained by maximizing $R(\boldsymbol{\beta})$ or minimizing $-2\log R(\boldsymbol{\beta})$ under the constraint (6).

In general, we can estimate $\boldsymbol{\beta}_0$ based on Algorithm 2.

4 Asymptotic properties

This section shows the asymptotic properties of the proposed estimator. Specially, Theorem 1 presents the asymptotic normality of the proposed estimator and Theorem 2 shows that the proposed estimator is asymptotically more efficient than the estimator of Lin et al. (2018). Theorems 3, 4 and 5 show the properties of statistics, which are obtained from the empirical likelihood ratio function. To establish the asymptotic properties, we introduce the following regularity conditions:

(R.1) The number of replicate measurements for the error-prone covariate $\boldsymbol{X}_{ij}$, K, is a bounded positive integer.

Algorithm 2 The proposed procedure for estimating $\boldsymbol{\beta}_0$

1. Choose an initial value $\boldsymbol{\beta}^{(0)}$, which can be obtained by using the method in Lin et al. (2018) with a working independence correlation matrix. Set $k = 0$.
2. With the value of $\boldsymbol{\beta}^{(k)}$, estimate the covariance matrix $\boldsymbol{\Sigma}_i$ by using the same method as that in Qin et al. (2016b). Denote the estimated value of covariance matrix as $\hat{\boldsymbol{\Sigma}}_i^{(k)}$.
3. Construct the reduced auxiliary random vector $\boldsymbol{g}_i^*(\boldsymbol{\beta}^{(k)}, \hat{\boldsymbol{\Sigma}}_i^{(k)})$ based on Algorithm 1, and solve equation (6) to obtain $\boldsymbol{\lambda}^{(k)}$ by using the modified Newton-Raphson method. Take the value of $\boldsymbol{\lambda}^{(k)}$ into the objective function (7) and obtain the value of $-2\log R(\boldsymbol{\beta}^{(k)})$. Then calculate the new estimated value $\boldsymbol{\beta}^{(k+1)}$ based on an optimization method (e.g., the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm). Set $k = k + 1$.
4. Iterate Step 2 and Step 3 until convergence. The final estimated value of $\boldsymbol{\beta}_0$ is denoted as $\hat{\boldsymbol{\beta}}$.

(R.2) The regression parameter $\boldsymbol{\beta}_0$ is identifiable, i.e., there is a unique $\boldsymbol{\beta}_0 \in \mathscr{B}$ satisfying the model assumption (1) which guarantees $\mathrm{E}\{\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\} = \mathbf{0}$, where $\mathscr{B}$ is a compact parameter space.

(R.3) There exist two positive constants c_1 and c_2 such that

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

where η_{i1} and η_{im_i} denote the smallest and largest eigenvalues of $\boldsymbol{\Sigma}_i$, respectively.

(R.4) $\max_{1\leq i\leq n} \mathrm{E}\|\mathbf{X}_i\|^6 < \infty$, $\max_{1\leq i\leq n, 1\leq k\leq K} \mathrm{E}\|\boldsymbol{\xi}_{i(k)}\|^3 < \infty$, and $\max_{1\leq i\leq n} \sup_{\mathbf{x}} \mathrm{E}(\|\boldsymbol{\varepsilon}_i\|^3 | \mathbf{X}_i = \mathbf{x}) < \infty$, where $\|\cdot\|$ denotes the Euclidean norm.

(R.5) $\mathbf{L}_n/n \to \mathbf{L}$ in probability for some matrix $\mathbf{L}$ and $\mathbf{M}_n/n \to \mathbf{M}$ in probability for some positive definite matrix $\mathbf{M}$, where $\mathbf{L}_n = \sum_{i=1}^n \partial \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)/\partial \boldsymbol{\beta}^\top$ and $\mathbf{M}_n = \sum_{i=1}^n \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)^\top$.

Remark 1. *Condition (R.1) requires the number of replicate measurements for $\boldsymbol{X}_{ij}$ to be bounded, it can ensure $\sum_{i=1}^n \mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)/\sqrt{n}\|^3 \to 0$. This condition is easy to verify in practice. Condition (R.2) assumes the identifiability of the true parameter $\boldsymbol{\beta}_0$. It is not easy to check in practice, but it is a commonly used condition in empirical likelihood literature, see Owen (2001). Condition (R.3) requires the eigenvalues of covariance matrices $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, n$ to*

be bounded away from 0 *and* ∞. *If* $\boldsymbol{\Sigma}_i$ $(i = 1, \ldots, n)$ *are known, Condition (R.3) can be checked directly; if* $\boldsymbol{\Sigma}_i$ $(i = 1, \ldots, n)$ *are unknown, we can first use some methods to estimate them, such as that in Qin et al. (2016b). Then we can check whether the sample covariance matrices satisfy Condition (R.3) or not. Sometimes, we assume the covariance matrices* $\boldsymbol{\Sigma}_i$ $(i = 1, \ldots, n)$ *satisfy some specific structures, such as independent structure, exchangeable structure, or AR(1) structure, then Condition (R.3) can be satisfied naturally. Condition (R.4) contains the moment conditions for the covariate, measurement error, and random error, which can be met under some common distributions, such as normal distribution and exponential distribution. Condition (R.5) assumes the convergence of* $\mathbf{L}_n/n$ *and* $\mathbf{M}_n/n$ *in probability. If the distribution of variables and true models are known, then Conditions (R.4)–(R.5) can be checked; otherwise, it is not easy to check. However, the conditions which are similar to (R.4)–(R.5) can be found be in many references, such as Xue and Zhu (2007) and Zhang et al. (2019).*

Theorem 1. *Assuming that conditions (R.1)–(R.5) hold, we have*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \rightsquigarrow \mathcal{N}\big(\mathbf{0}, (\mathbf{L}^\top \mathbf{M}^{-1} \mathbf{L})^{-1}\big).$$

When $\boldsymbol{g}_i^*(\boldsymbol{\beta}) = \boldsymbol{g}_i(\boldsymbol{\beta})$, according to Theorem 3.6 in Owen (2001), the asymptotic variance of the empirical likelihood estimator $\hat{\boldsymbol{\beta}}$ is at least as small as that of $\hat{\boldsymbol{\beta}}_{LIN}$, because $\sum_{k_1 \neq k_2} \mathbf{W}_{i(k_1)}^\top \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \mathbf{W}_{i(k_2)} \boldsymbol{\beta})$ is a linear combination of $\boldsymbol{g}_i(\boldsymbol{\beta})$. When $\boldsymbol{g}_i^*(\boldsymbol{\beta}) \neq \boldsymbol{g}_i(\boldsymbol{\beta})$, the above conclusion still holds because (1) the asymptotic variance of $\hat{\boldsymbol{\beta}}_{LIN}$ is equal to the asymptotic variance of one certain estimator obtained from the estimating equation $\sum_{i=1}^n \mathbf{B} \boldsymbol{g}_i^*(\boldsymbol{\beta}) = \mathbf{0}$, where $\mathbf{B}$ is a special $p \times q$ matrix and (2) the asymptotic variance of the empirical likelihood estimator $\hat{\boldsymbol{\beta}}$ is at least as small as that of any estimator obtained from the estimating equation $\sum_{i=1}^n \mathbf{C} \boldsymbol{g}_i^*(\boldsymbol{\beta}) = \mathbf{0}$, where $\mathbf{C}$ is an arbitrary $p \times q$ matrix. The proof of this conclusion in the toy example is given in Appendix B, which can be easily extended to other cases. We summarize the conclusion in the following theorem.

Theorem 2. *The asymptotic variance of the empirical likelihood estimator* $\hat{\boldsymbol{\beta}}$ *is at least as small as that of* $\hat{\boldsymbol{\beta}}_{LIN}$. *In other words, the empirical likelihood estimator* $\hat{\boldsymbol{\beta}}$ *is at least as efficient as* $\hat{\boldsymbol{\beta}}_{LIN}$.

In particular, if the replicate measurements for the covariate follow the same distribution or more generally, the models meet some specific moment conditions, then the asymptotic variance of $\hat{\boldsymbol{\beta}}$ is the same as that of $\hat{\boldsymbol{\beta}}_{LIN}$. For example, if there are two replicate measurements, the moment condition is

$$\begin{aligned}\sum_{i=1}^{n} \big[\mathrm{E}\{\mathbf{X}_i^\top \Sigma_i^{-1}\mathrm{cov}(\boldsymbol{\delta}_{i(2)}\boldsymbol{\beta})\Sigma_i^{-1}\mathbf{X}_i\} + \mathrm{E}(\boldsymbol{\delta}_{i(1)}^\top \Sigma_i^{-1}\boldsymbol{\delta}_{i(1)}) \\ + \mathrm{cov}(\boldsymbol{\delta}_{i(1)}^\top \Sigma_i^{-1}\boldsymbol{\delta}_{i(2)}\boldsymbol{\beta}) - \mathrm{E}(\boldsymbol{\delta}_{i(1)}^\top \Sigma_i^{-1}\boldsymbol{\delta}_{i(2)}\boldsymbol{\beta}\boldsymbol{\beta}^\top \boldsymbol{\delta}_{i(1)}^\top \Sigma_i^{-1}\boldsymbol{\delta}_{i(2)})\big] \\ = \sum_{i=1}^{n} \big[\mathrm{E}\{\mathbf{X}_i^\top \Sigma_i^{-1}\mathrm{cov}(\boldsymbol{\delta}_{i(1)}\boldsymbol{\beta})\Sigma_i^{-1}\mathbf{X}_i\} + \mathrm{E}(\boldsymbol{\delta}_{i(2)}^\top \Sigma_i^{-1}\boldsymbol{\delta}_{i(2)}) \\ + \mathrm{cov}(\boldsymbol{\delta}_{i(2)}^\top \Sigma_i^{-1}\boldsymbol{\delta}_{i(1)}\boldsymbol{\beta}) - \mathrm{E}(\boldsymbol{\delta}_{i(2)}^\top \Sigma_i^{-1}\boldsymbol{\delta}_{i(1)}\boldsymbol{\beta}\boldsymbol{\beta}^\top \boldsymbol{\delta}_{i(2)}^\top \Sigma_i^{-1}\boldsymbol{\delta}_{i(1)})\big].\end{aligned}$$

The detailed proof is given in Appendix C. However, if the moment condition is violated, the asymptotic variance of $\hat{\boldsymbol{\beta}}$ is smaller than that of $\hat{\boldsymbol{\beta}}_{LIN}$. In summary, the proposed empirical likelihood estimator is asymptotically more efficient than Lin et al. (2018)'s estimator.

Theorem 3. *Assuming that conditions (R.1)–(R.5) hold, we have* $-2\log R(\boldsymbol{\beta}_0) \rightsquigarrow \chi^2(q)$, *where* $\chi^2(q)$ *is a chi-squared distribution with* q *degrees of freedom.*

Theorem 4. *The empirical likelihood ratio statistic for the test of* $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ *versus* $H_a : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ *is*

$$W_1(\boldsymbol{\beta}_0) = -2\log\big\{R(\boldsymbol{\beta}_0)/R(\hat{\boldsymbol{\beta}})\big\}.$$

In addition, assuming that conditions (R.1)–(R.5) hold, we have $W_1(\boldsymbol{\beta}_0) \rightsquigarrow \chi^2(p)$ *under* H_0.

Similar to the parametric likelihood method, Theorems 3 and 4 allow us to use the test statistics $-2\log R(\boldsymbol{\beta}_0)$ and $W_1(\boldsymbol{\beta}_0)$ to perform hypothesis testing and construct confidence region for $\boldsymbol{\beta}_0$. Specifically, the $100(1-\alpha)\%$ empirical likelihood confidence region for $\boldsymbol{\beta}_0$ can be constructed as

$$\mathbf{I}_1 = \big\{\boldsymbol{\beta} : -2\log R(\boldsymbol{\beta}) \leq \chi^2_{1-\alpha}(q)\big\} \quad \text{or} \quad \mathbf{I}_2 = \big\{\boldsymbol{\beta} : W_1(\boldsymbol{\beta}) \leq \chi^2_{1-\alpha}(p)\big\},$$

where $\chi^2_{1-\alpha}(d)$ is the $(1-\alpha)$th quantile of $\chi^2(d)$ for any positive integer d.

If we are only interested in a part of elements in the parameter vector $\boldsymbol{\beta}$, we can use a profile empirical likelihood ratio test statistic to perform hypothesis testing and construct

confidence region for the parameters of interest. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are r-dimensional and $(p-r)$-dimensional vectors, respectively. In order to test $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0$ versus $H_a : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_1^0$, we define the profile empirical likelihood ratio test statistic as

$$W_2(\boldsymbol{\beta}_1^0) = -2\log\left\{R(\boldsymbol{\beta}_1^0, \hat{\boldsymbol{\beta}}_2^0)/R(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)\right\},$$

where $\hat{\boldsymbol{\beta}}_2^0$ minimizes $-2\log R(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2)$ with respect to $\boldsymbol{\beta}_2$. Furthermore, we can show the following theorem:

Theorem 5. *Assuming that conditions (R.1)–(R.5) hold, we have* $W_2(\boldsymbol{\beta}_1^0) \rightsquigarrow \chi^2(r)$ *under* H_0.

Thus, an approximate $100(1-\alpha)\%$ confidence region for $\boldsymbol{\beta}_1^0$ is

$$\mathbf{I}_3 = \left\{\boldsymbol{\beta}_1 : W_2(\boldsymbol{\beta}_1) \leq \chi^2_{1-\alpha}(r)\right\}.$$

The proofs of Theorems 1, 3, 4, and 5 are given in Appendix D. When constructing the confidence region for the parameters of interest, compared with the normal approximation-based method, the chi-squared approximation-based method can avoid estimating the asymptotic covariance matrix and does not need to impose prior constraints on the shape of the confidence region. Therefore, the chi-squared approximation-based method is recommended instead of the normal approximation-based method in practice.

5 Simulation studies

In this section, we conduct comprehensive simulations to evaluate the performance of the proposed empirical likelihood (EL) method when there are replicate measurements for the covariate and there may exist measurement errors in the covariate. For comparison, we also present the simulation results of the naive generalized estimating equation (GEE) method (Liang and Zeger, 1986), the naive EL method (Qin and Lawless, 1994), and Lin et al. (2018)'s estimating equation method. When there are measurement errors in the covariate, the naive methods simply replace the unobserved true covariate with the average of replicate surrogate measurements.

5.1 Simulation settings

We consider the following linear regression model

$$Y_{ij} = \beta_{00} + X_{ij,1}\beta_{01} + X_{ij,2}\beta_{02} + \varepsilon_{ij}, \quad i = 1, \cdots, n, \quad j = 1, \cdots, m,$$

where $(\beta_{00}, \beta_{01}, \beta_{02}) = (1, 1, 1)$, $m = 6$, and the number of subjects n is taken to be 50, 100, 200, 300, or 500. The covariates $X_{ij,1}$ and $X_{ij,2}$ are drawn independently from the standard normal distribution $\mathcal{N}(0, 1)$. The random error $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{im})^\top$ is generated from a multivariate normal distribution with mean zero and covariance matrix $\mathbf{R}_i(\rho)\sigma_e^2$, where $\mathbf{R}_i(\rho)$ is the correlation matrix chosen to have an exchangeable structure with $\rho = 0.6$, and $\sigma_e^2 = 0.8$.

Assuming that there are measurement errors in the covariate $X_{ij,1}$, the surrogate values $W_{ij(k),1}$, $k = 1, \cdots, K$ are generated from the following additive measurement error model:

$$W_{ij(k),1} = X_{ij,1} + \xi_{ij(k)}, \quad k = 1, \cdots, K.$$

To investigate the impacts of measurement errors on estimation accuracy and efficiency, we consider the following four cases, respectively.

C1: There are two replicate measurements for $X_{ij,1}$. $\xi_{ij(1)}$ and $\xi_{ij(2)}$ are independently generated from a normal distribution with mean zero and standard deviation 0.6.

C2: There are two replicate measurements for $X_{ij,1}$. $\xi_{ij(1)}$ is generated from a normal distribution with mean zero and standard deviation 0.6, while $\xi_{ij(2)}$ is generated from a t-distribution with 4 degrees of freedom. $\xi_{ij(1)}$ and $\xi_{ij(2)}$ are independent.

C3: There are three replicate measurements for $X_{ij,1}$. $\xi_{ij(1)}$, $\xi_{ij(2)}$, and $\xi_{ij(3)}$ are independently generated from a normal distribution with mean zero and standard deviation 0.6.

C4: There are three replicate measurements for $X_{ij,1}$. $\xi_{ij(1)}$ is generated from a normal distribution with mean zero and standard deviation 0.6, $\xi_{ij(2)}$ is generated from a t-distribution with 4 degrees of freedom, and $\xi_{ij(3)}$ is first generated from an exponential distribution with the rate parameter $\lambda = 2$ and then is centralized by subtracting its expectation 0.5. $\xi_{ij(1)}$, $\xi_{ij(2)}$, and $\xi_{ij(3)}$ are independent.

Under each simulation setting, 1000 replications are conducted.

5.2 Simulation results

For each method, we calculate the bias, standard deviation (SD), mean squared error (MSE), and coverage probability (CP) and mean length (ML) of the 95% confidence interval. For the GEE-based methods, the confidence intervals are constructed based on the asymptotic normality of the estimators. For the EL-based methods, the confidence intervals are constructed based on the asymptotic chi-squared distribution of the profile empirical likelihood ratio test statistic. The simulation results are presented in Tables 1–8.

By comparison, we find that the results of the naive GEE method and the naive EL method are very similar. This is consistent with the statement that the GEE method and EL method are asymptotically equivalent when the number of elements in the reduced auxiliary random vector equals the number of unknown parameters (Qin and Lawless, 1994). When there are measurement errors in the covariate $X_{ij,1}$, the biases and MSEs of $\hat{\beta}_1$ obtained from the two naive methods are all very large. Besides, the CPs for β_{01} are very close to zero based on the two naive methods, which means that the true value always does not fall into the 95% confidence interval. Therefore, the effects of measurement errors cannot be ignored. However, the biases of $\hat{\beta}_1$ based on Lin et al. (2018)'s method and the proposed method are much smaller than those based on the two naive methods, indicating that the biases induced by measurement errors can be eliminated successfully. When the distributions of different replicate measurements are the same (C1 and C3), as the number of subjects increases, Lin et al. (2018)'s method and the proposed method tend to be comparable with similar SD, MSE, and ML. However, when the distributions of different replicate measurements are different (C2 and C4), the proposed method is more efficient than Lin et al. (2018)'s method when the number of subjects is not too small, because the SD and ML of the proposed method are generally smaller than those of Lin et al. (2018)'s method.

Comparing the results when the measurement errors are generated by way of C1 with those when the measurement errors are generated by way of C3, we can find that as the number of replicate measurements for $X_{ij,1}$ increases, all the methods become more efficient if the number of subjects is not too small. From the results when the measurement errors are

generated by ways of C2 and C4, we can get the same conclusion. Therefore, we need to make full use of information from all the replicate measurements. In addition, the CPs of Lin et al. (2018)'s method and the proposed method are all close to the nominal confidence level 95% if the number of subjects is larger than 50, which shows that the confidence intervals obtained from the asymptotic theories are acceptable when the number of subjects is not too small.

It is worth mentioning that when the number of subjects is small and the number of elements in the reduced auxiliary random vector is large, the empirical likelihood-based methods may have some problems. For example, Han (2014) stated that their proposed empirical likelihood-based method might have numerical issues when the sample size (i.e., number of subjects) was small and/or the number of constraints (i.e., number of elements in the reduced auxiliary random vector) was large. This may explain why the SD and MSE of the proposed method are sometimes larger than those of the Lin et al. (2018)'s method when the number of subjects is 50 and the measurement errors are generated by ways of C3 and C4, under which the number of elements in the reduced auxiliary random vector is 11. Tsao (2004) also showed that the least upper bounds on coverage probabilities of the empirical likelihood ratio confidence regions might be surprisingly small when the ratio of the number of subjects and the number of elements in the reduced auxiliary random vector was small. This may be the reason why the CPs of confidence intervals based on the proposed EL estimator are lower than the nominal level 95% when the number of subjects is 50.

In summary, the proposed method performs well when the number of subjects is not too small. Specifically, it can eliminate the effects of measurement errors in the covariate and has a high estimation efficiency.

In order to obtain the simulation results in Tables 1–8, we compute on the Digital Research Alliance of Canada's cluster Graham and use the R software (version 3.6.1). The R codes are available on the RunMyCode website. Table 9 shows the average computation time for one replication. It can be found that the EL-based methods take more time than the GEE-based methods. This is not surprising because we need to perform more optimization calculations for the EL-based methods, especially when we calculate the confidence intervals. Thus, we will lose some computation efficiency to gain estimation efficiency. There is a trade-off between them. If we place more emphasis on the estimation efficiency, then the

proposed method is a good choice.

Table 1: Bias, standard deviation (SD) and mean squared error (MSE) of the estimator when the measurement errors are generated by way of C1

n	$\beta_{00}=1$				$\beta_{01}=1$				$\beta_{02}=1$			
	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed
Bias												
50	-0.49	-0.49	-0.46	-0.34	-14.93	-14.93	0.74	0.69	0.01	0.01	0.04	0.05
100	0.00	0.00	0.01	0.04	-15.27	-15.27	0.19	0.13	0.06	0.06	0.08	0.08
200	-0.03	-0.03	-0.03	-0.03	-15.15	-15.15	0.24	0.22	-0.11	-0.11	-0.12	-0.11
300	0.03	0.03	0.04	0.02	-15.28	-15.28	0.03	0.02	-0.09	-0.09	-0.10	-0.10
500	-0.05	-0.05	-0.05	-0.05	-15.22	-15.22	0.11	0.11	0.02	0.02	0.04	0.04
SD												
50	10.55	10.55	10.61	10.93	4.10	4.10	5.33	5.50	4.34	4.34	4.43	4.63
100	7.32	7.32	7.39	7.53	2.95	2.95	3.78	3.86	3.12	3.12	3.21	3.29
200	5.18	5.18	5.19	5.21	2.00	2.00	2.64	2.67	2.15	2.15	2.18	2.21
300	4.35	4.35	4.37	4.40	1.64	1.64	2.13	2.15	1.70	1.70	1.76	1.78
500	3.34	3.34	3.36	3.37	1.27	1.27	1.67	1.67	1.34	1.34	1.38	1.38
MSE												
50	1.12	1.12	1.13	1.19	2.40	2.40	0.29	0.31	0.19	0.19	0.20	0.21
100	0.54	0.54	0.55	0.57	2.42	2.42	0.14	0.15	0.10	0.10	0.10	0.11
200	0.27	0.27	0.27	0.27	2.33	2.33	0.07	0.07	0.05	0.05	0.05	0.05
300	0.19	0.19	0.19	0.19	2.36	2.36	0.05	0.05	0.03	0.03	0.03	0.03
500	0.11	0.11	0.11	0.11	2.33	2.33	0.03	0.03	0.02	0.02	0.02	0.02

Note: **All the values of simulation results are multiplied by 100**. GEEN: the naive GEE method; ELN: the naive EL method; LIN: Lin et al. (2018)'s estimating equation method; Proposed: the proposed EL method.

Table 2: Coverage probability (CP) and mean length (ML) of the 95% confidence interval when the measurement errors are generated by way of C1

n	$\beta_{00}=1$				$\beta_{01}=1$				$\beta_{02}=1$			
	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed
CP												
50	93.2	93.5	93.1	92.8	3.9	4.4	93.5	91.5	94.0	94.1	94.2	92.2
100	94.9	95.0	95.3	94.5	0.0	0.0	93.9	93.9	94.3	94.4	94.3	94.3
200	95.3	95.5	95.1	95.0	0.0	0.0	93.6	93.8	94.5	94.4	94.4	94.1
300	95.5	95.5	95.2	95.4	0.0	0.0	94.2	94.2	95.6	95.6	95.2	95.7
500	95.1	95.1	95.2	95.1	0.0	0.0	93.4	93.1	95.2	95.3	94.7	94.4
ML												
50	40.8	41.5	41.0	40.5	15.2	15.4	20.0	19.9	16.5	16.8	17.0	16.9
100	29.0	29.2	29.1	29.2	10.9	11.0	14.3	14.3	11.7	11.9	12.1	12.1
200	20.5	20.6	20.6	20.7	7.7	7.7	10.1	10.1	8.4	8.4	8.6	8.6
300	16.9	16.9	16.9	17.0	6.3	6.3	8.3	8.3	6.8	6.8	7.0	7.0
500	13.1	13.1	13.1	13.1	4.9	4.9	6.4	6.4	5.3	5.3	5.5	5.5

Note: **All the values of simulation results are multiplied by 100**. GEEN: the naive GEE method; ELN: the naive EL method; LIN: Lin et al. (2018)'s estimating equation method; Proposed: the proposed EL method.

Table 3: Bias, standard deviation (SD) and mean squared error (MSE) of the estimator when the measurement errors are generated by way of C2

n	$\beta_{00}=1$				$\beta_{01}=1$				$\beta_{02}=1$			
	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed
Bias												
50	-0.57	-0.57	-0.52	-0.41	-36.72	-36.72	1.32	0.78	-0.07	-0.08	-0.07	0.12
100	-0.04	-0.05	-0.06	-0.03	-37.06	-37.08	0.65	0.47	0.01	0.01	0.05	0.03
200	0.02	0.02	0.06	-0.01	-36.90	-36.90	0.50	0.43	-0.06	-0.06	-0.06	-0.08
300	0.01	0.01	0.01	0.00	-36.91	-36.92	0.31	0.30	-0.03	-0.03	-0.04	-0.04
500	-0.01	-0.01	0.01	-0.03	-37.08	-37.11	0.13	0.12	-0.01	-0.00	0.04	0.02
SD												
50	10.88	10.88	11.28	11.27	5.33	5.35	8.16	8.36	5.49	5.50	6.31	5.46
100	7.50	7.51	7.83	7.70	3.86	3.90	5.85	5.73	3.69	3.70	4.17	3.54
200	5.32	5.32	5.47	5.35	2.80	2.81	4.32	4.29	2.65	2.66	3.01	2.49
300	4.47	4.47	4.62	4.47	2.24	2.27	3.35	3.30	1.98	1.98	2.32	1.96
500	3.43	3.43	3.55	3.43	1.88	1.96	2.59	2.57	1.68	1.68	1.92	1.59
MSE												
50	1.18	1.18	1.27	1.27	13.76	13.77	0.68	0.70	0.30	0.30	0.40	0.30
100	0.56	0.56	0.61	0.59	13.88	13.90	0.35	0.33	0.14	0.14	0.17	0.13
200	0.28	0.28	0.30	0.29	13.69	13.70	0.19	0.19	0.07	0.07	0.09	0.06
300	0.20	0.20	0.21	0.20	13.67	13.68	0.11	0.11	0.04	0.04	0.05	0.04
500	0.12	0.12	0.13	0.12	13.79	13.81	0.07	0.07	0.03	0.03	0.04	0.03

Note: **All the values of simulation results are multiplied by 100**. GEEN: the naive GEE method; ELN: the naive EL method; LIN: Lin et al. (2018)'s estimating equation method; Proposed: the proposed EL method.

Table 4: Coverage probability (CP) and mean length (ML) of the 95% confidence interval when the measurement errors are generated by way of C2

n	$\beta_{00}=1$				$\beta_{01}=1$				$\beta_{02}=1$			
	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed
CP												
50	93.5	93.7	93.2	92.4	0.0	0.0	95.1	92.3	93.5	94.0	93.8	91.1
100	95.0	95.1	95.2	94.6	0.0	0.0	95.2	94.3	95.1	95.2	94.2	93.9
200	94.6	94.6	94.9	94.7	0.0	0.0	95.2	93.8	94.2	94.3	95.4	95.1
300	95.2	95.0	95.5	95.4	0.0	0.0	95.3	94.2	95.7	95.8	95.6	96.1
500	95.6	95.5	94.9	95.4	0.0	0.0	95.2	95.0	94.8	94.7	94.9	94.5
ML												
50	42.1	42.8	43.6	41.5	19.1	19.3	33.0	31.9	19.9	20.2	23.1	19.2
100	29.9	30.2	30.9	29.8	14.0	14.2	23.2	22.8	14.1	14.2	16.3	13.7
200	21.2	21.3	21.9	21.1	10.2	10.5	16.4	16.1	10.0	10.1	11.5	9.7
300	17.4	17.4	17.9	17.3	8.4	8.6	13.3	13.1	8.2	8.3	9.4	7.9
500	13.5	13.5	13.9	13.4	6.8	6.9	10.3	10.1	6.4	6.4	7.3	6.1

Note: **All the values of simulation results are multiplied by 100**. GEEN: the naive GEE method; ELN: the naive EL method; LIN: Lin et al. (2018)'s estimating equation method; Proposed: the proposed EL method.

Table 5: Bias, standard deviation (SD) and mean squared error (MSE) of the estimator when the measurement errors are generated by way of C3

n	$\beta_{00}=1$				$\beta_{01}=1$				$\beta_{02}=1$			
	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed
Bias												
50	-0.46	-0.46	-0.43	-0.22	-10.59	-10.59	0.35	0.20	0.03	0.03	0.05	0.10
100	0.04	0.04	0.04	0.15	-10.72	-10.72	0.12	-0.01	0.06	0.06	0.06	0.08
200	-0.01	-0.01	-0.01	0.02	-10.75	-10.75	0.02	-0.01	-0.06	-0.06	-0.07	-0.04
300	0.03	0.03	0.03	0.02	-10.74	-10.74	0.01	0.00	-0.07	-0.07	-0.07	-0.06
500	-0.05	-0.05	-0.05	-0.06	-10.70	-10.70	0.04	0.05	0.02	0.02	0.02	0.02
SD												
50	10.50	10.50	10.53	11.78	4.01	4.01	4.67	5.21	4.21	4.21	4.24	4.91
100	7.31	7.31	7.35	7.80	2.82	2.82	3.25	3.47	2.89	2.89	2.95	3.15
200	5.16	5.16	5.17	5.29	2.01	2.01	2.38	2.45	2.04	2.04	2.06	2.11
300	4.31	4.31	4.33	4.39	1.56	1.56	1.83	1.87	1.63	1.63	1.66	1.68
500	3.32	3.32	3.33	3.37	1.24	1.24	1.45	1.46	1.27	1.27	1.30	1.30
MSE												
50	1.10	1.10	1.11	1.39	1.28	1.28	0.22	0.27	0.18	0.18	0.18	0.24
100	0.53	0.53	0.54	0.61	1.23	1.23	0.11	0.12	0.08	0.08	0.09	0.10
200	0.27	0.27	0.27	0.28	1.20	1.20	0.06	0.06	0.04	0.04	0.04	0.04
300	0.19	0.19	0.19	0.19	1.18	1.18	0.03	0.03	0.03	0.03	0.03	0.03
500	0.11	0.11	0.11	0.11	1.16	1.16	0.02	0.02	0.02	0.02	0.02	0.02

Note: **All the values of simulation results are multiplied by 100**. GEEN: the naive GEE method; ELN: the naive EL method; LIN: Lin et al. (2018)'s estimating equation method; Proposed: the proposed EL method.

Table 6: Coverage probability (CP) and mean length (ML) of the 95% confidence interval when the measurement errors are generated by way of C3

n	$\beta_{00}=1$				$\beta_{01}=1$				$\beta_{02}=1$			
	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed
CP												
50	93.8	93.8	93.5	87.4	21.9	22.9	93.4	86.3	92.8	93.1	92.8	84.8
100	94.7	94.9	94.4	93.1	2.5	2.9	94.0	91.8	94.4	94.6	94.4	91.9
200	95.3	95.3	95.1	94.9	0.0	0.0	94.0	93.4	94.7	94.6	94.3	93.5
300	95.1	95.2	95.2	94.9	0.0	0.0	95.2	94.6	94.9	95.0	94.8	94.2
500	94.7	94.6	94.6	94.6	0.0	0.0	94.3	94.3	94.9	94.9	94.6	95.0
ML												
50	40.5	41.2	40.6	36.8	14.8	15.1	17.4	15.7	15.8	16.0	16.0	14.5
100	28.8	29.0	28.8	28.3	10.6	10.7	12.4	12.2	11.2	11.3	11.4	11.1
200	20.4	20.5	20.4	20.4	7.5	7.5	8.8	8.8	7.9	8.0	8.1	8.0
300	16.8	16.8	16.8	16.8	6.2	6.2	7.2	7.2	6.5	6.5	6.6	6.6
500	13.0	13.0	13.0	13.0	4.8	4.8	5.6	5.6	5.0	5.1	5.1	5.1

Note: **All the values of simulation results are multiplied by 100**. GEEN: the naive GEE method; ELN: the naive EL method; LIN: Lin et al. (2018)'s estimating equation method; Proposed: the proposed EL method.

Table 7: Bias, standard deviation (SD) and mean squared error (MSE) of the estimator when the measurement errors are generated by way of C4

n	$\beta_{00}=1$				$\beta_{01}=1$				$\beta_{02}=1$			
	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed
Bias												
50	-0.51	-0.51	-0.47	-0.29	-22.09	-22.09	0.81	0.41	0.05	0.05	0.05	0.15
100	-0.01	-0.02	-0.01	0.12	-22.52	-22.52	0.22	0.02	0.16	0.16	0.19	0.19
200	0.03	0.03	0.05	0.01	-22.43	-22.43	0.13	0.08	-0.09	-0.09	-0.09	-0.12
300	0.01	0.01	0.01	-0.03	-22.37	-22.38	0.11	0.04	-0.02	-0.02	-0.03	-0.00
500	-0.02	-0.03	-0.01	-0.05	-22.45	-22.47	0.09	0.04	0.05	0.05	0.08	0.08
SD												
50	10.65	10.65	10.79	12.00	4.63	4.63	5.47	5.30	4.69	4.68	5.02	4.92
100	7.35	7.35	7.47	7.61	3.34	3.34	4.02	3.61	3.14	3.14	3.35	3.15
200	5.23	5.23	5.28	5.34	2.46	2.47	2.98	2.57	2.32	2.32	2.42	2.21
300	4.36	4.36	4.40	4.35	1.99	1.99	2.37	2.01	1.86	1.86	2.00	1.79
500	3.36	3.36	3.40	3.36	1.56	1.64	1.80	1.52	1.47	1.48	1.56	1.36
MSE												
50	1.14	1.14	1.16	1.44	5.09	5.09	0.31	0.28	0.22	0.22	0.25	0.24
100	0.54	0.54	0.56	0.58	5.18	5.18	0.16	0.13	0.10	0.10	0.11	0.10
200	0.27	0.27	0.28	0.29	5.09	5.09	0.09	0.07	0.05	0.05	0.06	0.05
300	0.19	0.19	0.19	0.19	5.04	5.05	0.06	0.04	0.03	0.03	0.04	0.03
500	0.11	0.11	0.12	0.11	5.06	5.08	0.03	0.02	0.02	0.02	0.02	0.02

Note: **All the values of simulation results are multiplied by 100**. GEEN: the naive GEE method; ELN: the naive EL method; LIN: Lin et al. (2018)'s estimating equation method; Proposed: the proposed EL method.

Table 8: Coverage probability (CP) and mean length (ML) of the 95% confidence interval when the measurement errors are generated by way of C4

n	$\beta_{00}=1$				$\beta_{01}=1$				$\beta_{02}=1$			
	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed	GEEN	ELN	LIN	Proposed
CP												
50	93.5	93.5	93.0	87.9	0.2	0.2	95.6	86.9	94.6	95.0	93.9	84.9
100	94.9	94.9	95.1	93.2	0.0	0.0	95.8	91.5	95.3	95.2	94.9	91.7
200	94.8	94.5	94.3	93.6	0.0	0.0	94.2	92.6	94.5	94.6	95.0	94.3
300	95.3	95.3	95.3	94.9	0.0	0.0	93.9	93.9	94.5	94.7	94.9	94.3
500	94.9	94.7	94.9	94.4	0.0	0.0	95.0	94.9	94.4	94.3	93.9	94.9
ML												
50	41.3	41.9	41.7	37.2	17.4	17.7	22.3	16.6	17.7	18.0	18.9	14.8
100	29.3	29.6	29.6	28.3	12.6	12.8	15.8	12.8	12.5	12.7	13.4	11.4
200	20.8	20.9	21.0	20.4	9.1	9.3	11.2	9.2	8.9	9.0	9.5	8.2
300	17.0	17.1	17.2	16.8	7.5	7.6	9.1	7.6	7.3	7.3	7.7	6.8
500	13.2	13.2	13.3	13.0	5.9	6.0	7.1	5.9	5.7	5.7	6.0	5.3

Note: **All the values of simulation results are multiplied by 100**. GEEN: the naive GEE method; ELN: the naive EL method; LIN: Lin et al. (2018)'s estimating equation method; Proposed: the proposed EL method.

Table 9: Average computation time (seconds) for one replication

Way	Method	50	100	200	300	500
C1	GEEN	0.02	0.05	0.08	0.12	0.21
	ELN	96.89	132.54	186.17	234.11	379.63
	LIN	0.03	0.06	0.11	0.15	0.26
	Proposed	134.72	169.73	228.88	302.45	475.25
C2	GEEN	0.02	0.03	0.08	0.12	0.20
	ELN	83.43	120.92	187.40	255.87	387.77
	LIN	0.02	0.05	0.10	0.15	0.26
	Proposed	119.65	147.34	219.67	300.61	452.90
C3	GEEN	0.02	0.04	0.08	0.13	0.21
	ELN	88.98	116.06	167.73	213.84	336.93
	LIN	0.07	0.13	0.26	0.41	0.69
	Proposed	254.58	327.10	470.43	668.72	1097.78
C4	GEEN	0.02	0.04	0.08	0.12	0.21
	ELN	91.90	114.59	166.00	224.82	383.31
	LIN	0.07	0.13	0.25	0.38	0.68
	Proposed	263.33	311.48	457.86	634.20	1094.20

Note: GEEN: the naive GEE method; ELN: the naive EL method; LIN: Lin et al. (2018)'s estimating equation method; Proposed: the proposed EL method.

6 Application to the LEAN study

We apply the proposed method to the Lifestyle Education for Activity and Nutrition (LEAN) study (Barry et al., 2011). The intervention strategy was providing the participants with a group-based behavioural weight loss program or/and the SenseWear platform which could help improve lifestyle self-monitoring (Shuger et al., 2011). The data of age, gender, race, and education level were collected at baseline. Body weight, height, systolic blood pressure (SBP), and diastolic blood pressure (DBP) were measured at baseline, month 4 and month 9. Among the 197 participants, 74 participants failed to complete the month 4 or/and month 9 assessments. For convenience, we exclude them from the current study.

The main interest of this study is to assess whether the intervention strategy is effective in reducing the BMI value at months 4 and 9. The response variable BMI is calculated by $\log(\text{weight}/\text{height}^2 \times 703)$. The main exposure variable is intervention, which is denoted as "group". Besides, it takes the value of 1 for the intervention group and takes the value of 0 for the standard care group. Two dummy variables t_1 and t_2 are introduced to represent the assessment time. That is, $t_1 = t_2 = 0$ for baseline, $t_1 = 1$ and $t_2 = 0$ for month 4, and $t_1 = 0$ and $t_2 = 1$ for month 9. To reveal the effects of different groups at different time points, we include the interaction terms between group and time as covariates. Other covariates considered are SBP, DBP, age, gender (female, 1; male, 0), race (African American, 1; others, 0), and education level (four-year college or higher, 1; others, 0). To make the covariates have similar scales, the values of SBP, DBP, and age are divided by 100. Before analysis, all the variables are centralized by subtracting their mean values. The following linear regression model is adopted to fit the LEAN data set:

$$\begin{aligned} Y =& \text{SBP}\,\beta_1 + \text{DBP}\,\beta_2 + \text{AGE}\,\beta_3 + \text{gender}\,\beta_4 + \text{race}\,\beta_5 + \text{education}\,\beta_6 \\ &+ \text{group}\,\beta_7 + t_1\beta_8 + t_2\beta_9 + (\text{group} \times t_1)\,\beta_{10} + (\text{group} \times t_2)\,\beta_{11} + \varepsilon. \end{aligned} \tag{8}$$

As mentioned in Section 1, there exist measurement errors in the covariates SBP and DBP, and the replicate measurement errors of SBP follow different distributions, as does DBP. Therefore, it is necessary to consider the distributional difference of measurement errors to obtain an efficient estimator. Table 10 displays the estimates of regression coefficients and confidence intervals by using different methods, including the naive GEE method

(Liang and Zeger, 1986), the naive EL method (Qin and Lawless, 1994), Lin et al. (2018)'s estimating equation method, and the proposed EL method. It can be found that the effect of intervention at month 9 is significantly different from zero at the significance level of 0.05 based on the naive GEE method, Lin et al. (2018)'s estimating equation method, and the proposed EL method. Besides, all these three methods show that intervention at month 9 is negatively related to BMI, which means that the intervention strategy can significantly reduce BMI at month 9. This conclusion is consistent with the finding of LEAN's study group (Shuger et al., 2011). In addition, all the methods show that SBP is significantly positively related to BMI. Besides, we find that the result of the naive GEE method is similar to that of the naive EL method. But the differences between these two naive methods and the proposed EL method are relatively large, which may be due to the effects of measurement errors. There are some differences in the estimates of Lin et al. (2018)'s estimating equation method and the proposed EL method. For example, the effect of t_2 is significantly different from zero at the significance level of 0.05 based on the proposed EL method, while it is not based on Lin et al. (2018)'s estimating equation method. This is perhaps because the confidence intervals for these two methods are constructed in different ways and the confidence intervals of the proposed method are shorter. In general, the proposed method has the shortest average length of confidence intervals. Therefore, the result of the proposed method is recommended.

7 Conclusion and discussion

In this paper, we propose a new method for analysis of longitudinal data with replicate covariate measurement errors based on the empirical likelihood estimator, where we use the independence between replicate measurement errors to construct an unbiased auxiliary random vector. When some elements in the original auxiliary random vector are functionally dependent or have some inner relationships, the reduced auxiliary random vector is introduced to define the profile empirical likelihood ratio function. The proposed method has the following advantages. Firstly, the proposed empirical likelihood estimator is asymptotically at least as efficient as the estimator of Lin et al. (2018), and under some moment conditions,

Table 10: Estimates of regression coefficients and the 95% confidence intervals in the analysis of the LEAN data set

	GEEN				ELN				LIN				Proposed			
	Coef	Lower	Upper	CL	Coef	Lower	Upper	CL	Coef	Lower	Upper	CL	Coef	Lower	Upper	CL
SBP	10.22*	2.90	17.53	14.63	9.57*	2.34	17.30	14.96	8.43*	0.78	16.08	15.30	9.16*	3.20	11.29	8.09
DBP	5.92	-4.86	16.70	21.56	5.32	-5.48	16.19	21.67	4.09	-7.56	15.73	23.29	2.47	-4.43	4.51	8.94
AGE	-2.47	-26.65	21.71	48.36	-2.32	-26.40	22.18	48.58	-2.08	-26.43	22.27	48.71	-5.67	-29.29	18.05	47.34
gender	0.68	-5.55	6.91	12.46	0.62	-5.93	6.69	12.62	0.53	-5.71	6.76	12.47	2.08	-3.52	7.55	11.07
race	3.90	-2.05	9.85	11.90	3.97	-1.98	9.95	11.93	4.11	-1.87	10.08	11.95	5.08	-1.09	10.80	11.89
education	-5.95	-11.94	0.05	11.99	-5.98	-11.88	0.26	12.14	-6.05	-12.10	0.00	12.10	-6.07	-12.73	0.54	13.27
group	-2.03	-8.42	4.35	12.76	-2.04	-8.55	4.20	12.75	-2.05	-8.47	4.37	12.84	-2.42	-8.65	3.85	12.50
t1	-1.62	-3.28	0.03	3.32	−1.62*	-3.76	-0.25	3.51	−1.62*	-3.22	-0.03	3.19	−1.13*	-1.75	-0.01	1.74
t2	-1.62	-3.71	0.47	4.18	-1.63	-4.26	0.30	4.56	-1.64	-3.71	0.43	4.15	−0.93*	-2.64	-0.12	2.52
group*t1	-0.92	-2.91	1.07	3.98	-0.95	-2.76	1.41	4.17	-1.00	-2.94	0.94	3.89	-1.43	-2.07	0.28	2.35
group*t2	−2.73*	-5.39	-0.08	5.31	-2.75	-5.38	0.24	5.62	−2.78*	-5.43	-0.13	5.30	−3.24*	-5.38	-2.14	3.24

Note: **All the values of results are multiplied by 100**. GEEN: the naive GEE method; ELN: the naive EL method; LIN: Lin et al. (2018)'s estimating equation method; Proposed: the proposed EL method; Coef: the estimate of regression coefficient; Lower: the lower bound of confidence interval; Upper: the upper bound of confidence interval; CL: the length of confidence interval; An asterisk (*) indicates that the effect is significant at the level of $\alpha = 0.05$.

the proposed estimator is strictly more efficient. Secondly, it provides a method that can deal with the problem where there are more than two replicate measurements at each assessment time. Thirdly, it enjoys all the good properties of the empirical likelihood method. According to the simulation results, the proposed method is not sensitive to the measurement errors in the covariate. In addition, it is more efficient than Lin et al. (2018)'s method when the number of subjects is not too small. Due to the flexibility in accounting for the distributional difference of measurement errors in replicate measurements, we recommend using the proposed method in practical problems.

Further extension of the proposed method may be of interest. One may consider how to extend this method to other models, such as the partially linear model and non-linear model. One may also consider how to improve the performance of the proposed method when the number of subjects is very small (Chen et al., 2008).

Acknowledgements

We would like to thank the Co-Editor, Dr. Byeong U. Park, an Associate Editor, and two referees for very helpful suggestions, which led to substantial improvements of the paper. We gratefully acknowledge Dr. Xuemei Sui of the University of South Carolina for providing the LEAN data set. This work was supported by the National Natural Science Foundation of China [11871164, 11671096, 11731011, 12071087].

References

Barry, V. W., McClain, A. C., Shuger, S., Sui, X., Hardin, J. W., Hand, G. A., Wilcox, S., and Blair, S. N. (2011). Using a technology-based intervention to promote weight loss in sedentary overweight or obese adults: a randomized controlled trial study design. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, 4:67–77.

Chen, J., Variyath, A. M., and Abraham, B. (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, 17(2):426–443.

Cheng, M.-Y., Honda, T., Li, J., et al. (2016). Efficient estimation in semivarying coefficient models for longitudinal/clustered data. *The Annals of Statistics*, 44(5):1988–2017.

D'Agostino, R. B. (1970). Transformation to normality of the null distribution of g_1. *Biometrika*, 57(3):679–681.

Diggle, P. (2002). *Analysis of longitudinal data*. Oxford University Press.

Funatogawa, I. and Funatogawa, T. (2018). Longitudinal data and linear mixed effects models. In *Longitudinal data analysis*, pages 1–26. Springer.

Han, P. (2014). Multiply robust estimation in regression analysis with missing data. *Journal of the American Statistical Association*, 109(507):1159–1173.

Hsieh, F., Tseng, Y.-K., and Wang, J.-L. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 62(4):1037–1043.

Hu, S. and Xu, J. (2022). An efficient and robust inference method based on empirical likelihood in longitudinal data analysis. *Communications in Statistics-Theory and Methods*, 51(4):994–1010.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974.

Li, M., Ma, Y., and Li, R. (2019). Semiparametric regression for measurement error model with heteroscedastic error. *Journal of Multivariate Analysis*, 171:320–338.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using general linear models. *Biometrika*, 73(1):13–22.

Lin, H., Qin, G., Zhang, J., and Zhu, Z. (2018). Analysis of longitudinal data with covariate measurement error and missing responses: An improved unbiased estimating equation. *Computational Statistics & Data Analysis*, 121:104–112.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Owen, A. B. (2001). *Empirical likelihood*. Chapman and Hall/CRC.

Qin, G., Zhang, J., and Zhu, Z. (2016a). Simultaneous mean and covariance estimation of partially linear models for longitudinal data with missing responses and covariate measurement error. *Computational Statistics & Data Analysis*, 96:24–39.

Qin, G., Zhang, J., Zhu, Z., and Fung, W. (2016b). Robust estimation of partially linear models for longitudinal data with dropouts and measurement error. *Statistics in Medicine*, 35(29):5401–5416.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general etimating equations. *The Annals of Statistics*, 22(1):300–325.

Qiu, J. and Wu, L. (2015). A moving blocks empirical likelihood method for longitudinal data. *Biometrics*, 71(3):616–624.

Shuger, S. L., Barry, V. W., Sui, X., Mcclain, A., Hand, G. A., Wilcox, S., Meriwether, R. A., Hardin, J. W., and Blair, S. N. (2011). Electronic feedback in a diet- and physical activity-based lifestyle intervention for weight loss: a randomized controlled trial. *International Journal of Behavioral Nutrition & Physical Activity*, 8(1):41–41.

Tsao, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *Annals of Statistics*, 32(3):1215–1221.

Wang, C. Y. (2006). Corrected score estimator for joint modeling of longitudinal and failure time data. *Statistica Sinica*, 16(1):235–253.

Wang, C. Y. and Sullivan, P. M. (2000). Expected estimating equations to accommodate covariate measurement error. *Journal of the Royal Statistical Society: Series B*, 62(3):509–524.

Wang, S., Qian, L., and Carroll, R. J. (2010). Generalized empirical likelihood methods for analyzing longitudinal data. *Biometrika*, 97(1):79–93.

Wu, L. (2002). A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to aids studies. *Journal of the American Statistical Association*, 97(460):955–964.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1):330–339.

Xue, L. and Zhu, L. (2007). Empirical likelihood for a varying coefficient model with longitudinal data. *Journal of the American Statistical Association*, 102(478):642–654.

Zhang, W., Leng, C., and Tang, C. Y. (2015). A joint modelling approach for longitudinal studies. *Journal of the Royal Statistical Society: Series B*, 77(1):219–238.

Zhang, Y., Qin, G., Zhu, Z., and Xu, W. (2019). A novel robust approach for analysis of longitudinal data. *Computational Statistics & Data Analysis*, 138:83–95.

Zhao, P. and Xue, L. (2009). Empirical likelihood inferences for semiparametric varying-coefficient partially linear errors-in-variables models with longitudinal data. *Journal of Nonparametric Statistics*, 21(7):907–923.

Zhao, P., Zhou, X., Wang, X., and Huang, X. (2019). A new orthogonality empirical likelihood for varying coefficient partially linear instrumental variable models with longitudinal data. *Communications in Statistics-Simulation and Computation*, 48:1–17.

Appendix A Construction of the reduced auxiliary random vector in a toy example

Now we use a toy example to introduce how to construct the reduced auxiliary random vector. Assume there are two covariates $X_{ij,1}$ and $X_{ij,2}$, where $X_{ij,1}$ is measured with error and there are three replicate measurements of $X_{ij,1}$. Denote the surrogate values of $X_{ij,1}$ as $W_{ij(k),1}$, $k = 1, 2, 3$. We further assume the response variable and covariates follow the following linear regression model:

$$Y_{ij} = \beta_{00} + X_{ij,1}\beta_{01} + X_{ij,2}\beta_{02} + \varepsilon_{ij}, \quad i = 1, \cdots, n, \quad j = 1, \cdots, m_i.$$

Therefore, $\boldsymbol{W}_{ij(k)} = (1, W_{ij(k),1}, X_{ij,2})^\top$ and $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01}, \beta_{02})^\top$. According to the definition in Section 3,

$$\boldsymbol{g}_i(\boldsymbol{\beta}) = \begin{pmatrix} \mathbf{W}_{i(1)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(2)}\boldsymbol{\beta}) \\ \mathbf{W}_{i(2)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(1)}\boldsymbol{\beta}) \\ \mathbf{W}_{i(1)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(3)}\boldsymbol{\beta}) \\ \mathbf{W}_{i(3)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(1)}\boldsymbol{\beta}) \\ \mathbf{W}_{i(2)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(3)}\boldsymbol{\beta}) \\ \mathbf{W}_{i(3)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(2)}\boldsymbol{\beta}) \end{pmatrix},$$

where

$$\mathbf{W}_{i(k_1)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(k_2)}\boldsymbol{\beta}) = \begin{pmatrix} \mathbf{1}_i^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(k_2)}\boldsymbol{\beta}) \\ \boldsymbol{W}_{i(k_1),1}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(k_2)}\boldsymbol{\beta}) \\ \boldsymbol{X}_{i,2}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(k_2)}\boldsymbol{\beta}) \end{pmatrix}, \quad k_1 \neq k_2,$$

with $\mathbf{1}_i$ being an m_i-dimensional vector with all the elements being one, $\boldsymbol{X}_{i,2} = (X_{i1,2}, \ldots, X_{im_i,2})^\top$, and $\boldsymbol{W}_{i(k_1),1} = (W_{i1(k_1),1}, \ldots, W_{im_i(k_1),1})^\top$. It is obvious that for a fixed value of k_2, the values of $\mathbf{1}_i^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(k_2)}\boldsymbol{\beta})$ are the same for different choices of k_1, it is also the case for $\boldsymbol{X}_{i,2}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(k_2)}\boldsymbol{\beta})$. As a result, the matrix $\mathrm{E}\{\boldsymbol{g}_i(\boldsymbol{\beta}_0)\boldsymbol{g}_i(\boldsymbol{\beta}_0)^\top\}$ is not invertible and the estimated value of $\boldsymbol{\beta}_0$ is unavailable. To solve this problem, we need to eliminate the duplicate elements in $\boldsymbol{g}_i(\boldsymbol{\beta})$. Besides, there is also a potential inner relationship among

$\left\{\boldsymbol{W}_{i(k_1),1}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(k_2)}\boldsymbol{\beta}), k_1\neq k_2\right\}$. Denote

$$\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta})=\begin{pmatrix}\boldsymbol{W}_{i(1),1}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(2)}\boldsymbol{\beta})\\ \boldsymbol{W}_{i(2),1}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(1)}\boldsymbol{\beta})\\ \boldsymbol{W}_{i(1),1}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(3)}\boldsymbol{\beta})\\ \boldsymbol{W}_{i(3),1}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(1)}\boldsymbol{\beta})\\ \boldsymbol{W}_{i(2),1}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(3)}\boldsymbol{\beta})\\ \boldsymbol{W}_{i(3),1}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(2)}\boldsymbol{\beta})\end{pmatrix},$$

and $\mathbf{A}=\mathrm{E}\{\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta}_0)\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta}_0)^{\top}\}$. Let $\boldsymbol{A}_j$ denote the jth row vector of $\mathbf{A}$, $j=1,\cdots,6$. By calculation, we can find that $\boldsymbol{A}_1+\boldsymbol{A}_4+\boldsymbol{A}_5=\boldsymbol{A}_2+\boldsymbol{A}_3+\boldsymbol{A}_6$. Therefore, $\mathbf{A}$ is not invertible. Through performing row transformation on the matrix $\mathrm{E}\{\boldsymbol{g}_i(\boldsymbol{\beta}_0)\boldsymbol{g}_i(\boldsymbol{\beta}_0)^{\top}\}$, $\mathbf{A}$ can be one block of the transformed matrix. As a result, if $\mathbf{A}$ is not invertible, then $\mathrm{E}\{\boldsymbol{g}_i(\boldsymbol{\beta}_0)\boldsymbol{g}_i(\boldsymbol{\beta}_0)^{\top}\}$ will also be not invertible. Therefore, to make $\mathrm{E}\{\boldsymbol{g}_i(\boldsymbol{\beta}_0)\boldsymbol{g}_i(\boldsymbol{\beta}_0)^{\top}\}$ invertible, we need to eliminate one element in $\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta})$ from $\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta})$. After eliminating the duplicate elements in $\boldsymbol{g}_i(\boldsymbol{\beta})$ and one element in $\tilde{\boldsymbol{g}}_i(\boldsymbol{\beta})$, we can get the reduced auxiliary random vector $\boldsymbol{g}_i^*(\boldsymbol{\beta})$ and

$$\boldsymbol{g}_i^*(\boldsymbol{\beta})=\begin{pmatrix}\mathbf{1}_i^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(2)}\boldsymbol{\beta})\\ \boldsymbol{W}_{i(1),1}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(2)}\boldsymbol{\beta})\\ \boldsymbol{X}_{i,2}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(2)}\boldsymbol{\beta})\\ \mathbf{1}_i^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(1)}\boldsymbol{\beta})\\ \boldsymbol{W}_{i(2),1}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(1)}\boldsymbol{\beta})\\ \boldsymbol{X}_{i,2}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(1)}\boldsymbol{\beta})\\ \mathbf{1}_i^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(3)}\boldsymbol{\beta})\\ \boldsymbol{W}_{i(1),1}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(3)}\boldsymbol{\beta})\\ \boldsymbol{X}_{i,2}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(3)}\boldsymbol{\beta})\\ \boldsymbol{W}_{i(3),1}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(1)}\boldsymbol{\beta})\\ \boldsymbol{W}_{i(2),1}^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(3)}\boldsymbol{\beta})\end{pmatrix}.$$

In this example, the number of elements in $\boldsymbol{g}_i^*(\boldsymbol{\beta})$ is 11.

Appendix B Efficiency of the empirical likelihood estimator in the toy example

In the toy example given in Appendix A, the estimating equation for Lin et al. (2018)'s method is

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{U}_i(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{k_1 \neq k_2} \mathbf{W}_{i(k_1)}^{\top} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \mathbf{W}_{i(k_2)} \boldsymbol{\beta}) = \mathbf{0},$$

where k_1 and k_2 are the elements of $\{1, 2, 3\}$. Let

$$\boldsymbol{h}_i(\boldsymbol{\beta}) = [\{\boldsymbol{g}_i^*(\boldsymbol{\beta})\}^{\top}, \boldsymbol{W}_{i(3),1}^{\top} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \mathbf{W}_{i(2)} \boldsymbol{\beta})]^{\top},$$

then $\boldsymbol{U}_i(\boldsymbol{\beta}) = \mathbf{B}_1 \boldsymbol{h}_i(\boldsymbol{\beta})$, where

$$\mathbf{B}_1 = \begin{pmatrix} 2 & 0 & 0 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 0 & 0 & 2 & 0 & 0 & 2 & 0 & 0 & 0 \end{pmatrix}.$$

For any estimator $\hat{\theta}$, we denote the asymptotic variance of $\hat{\theta}$ as $\mathrm{asyVar}(\hat{\theta})$. According to Theorem 2 in Lin et al. (2018), the asymptotic variance of $\hat{\boldsymbol{\beta}}_{LIN}$ is

$$\begin{aligned} &\mathrm{asyVar}(\hat{\boldsymbol{\beta}}_{LIN}) \\ &\quad = \left[\sum_{i=1}^{n} \mathrm{E}\Big\{ \frac{\partial \boldsymbol{U}_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^{\top}} \Big\} \right]^{-1} \left[\sum_{i=1}^{n} \mathrm{E}\big\{ \boldsymbol{U}_i(\boldsymbol{\beta}_0) \boldsymbol{U}_i(\boldsymbol{\beta}_0)^{\top} \big\} \right] \left[\sum_{i=1}^{n} \mathrm{E}\Big\{ \frac{\partial \boldsymbol{U}_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^{\top}} \Big\} \right]^{-1.\top} \\ &= \left[\sum_{i=1}^{n} \mathbf{B}_1 \mathrm{E}\Big\{ \frac{\partial \boldsymbol{h}_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^{\top}} \Big\} \right]^{-1} \left[\sum_{i=1}^{n} \mathbf{B}_1 \mathrm{E}\big\{ \boldsymbol{h}_i(\boldsymbol{\beta}_0) \boldsymbol{h}_i(\boldsymbol{\beta}_0)^{\top} \big\} \mathbf{B}_1^{\top} \right] \left[\sum_{i=1}^{n} \mathbf{B}_1 \mathrm{E}\Big\{ \frac{\partial \boldsymbol{h}_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^{\top}} \Big\} \right]^{-1.\top}. \end{aligned}$$

By calculation,

$$\mathrm{E}\left\{ \frac{\partial \boldsymbol{h}_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^{\top}} \right\} = \mathbf{B}_2 \mathrm{E}\left\{ \frac{\partial \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^{\top}} \right\},$$

and

$$\mathrm{E}\big\{ \boldsymbol{h}_i(\boldsymbol{\beta}_0) \boldsymbol{h}_i(\boldsymbol{\beta}_0)^{\top} \big\} = \mathbf{B}_2 \mathrm{E}\big\{ \boldsymbol{g}_i^*(\boldsymbol{\beta}_0) \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)^{\top} \big\} \mathbf{B}_2^{\top},$$

where

$$\mathbf{B}_2 = \begin{pmatrix} \mathbf{I}_{11 \times 11} \\ \boldsymbol{B}_3^{\top} \end{pmatrix},$$

with $\boldsymbol{B}_3 = (0,1,0,0,-1,0,0,-1,0,1,1)^\top$ and $\mathbf{I}_{p\times p}$ being the $p \times p$ identity matrix for any positive integer p. Let $\mathbf{B} = \mathbf{B}_1\mathbf{B}_2$, then

$$\begin{aligned}
&\text{asyVar}(\hat{\boldsymbol{\beta}}_{LIN}) \\
&= \left[\sum_{i=1}^{n} \mathbf{B}\mathrm{E}\Big\{\frac{\partial \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^\top}\Big\}\right]^{-1} \left[\sum_{i=1}^{n} \mathbf{B}\mathrm{E}\{\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)^\top\}\mathbf{B}^\top\right] \left[\sum_{i=1}^{n} \mathbf{B}\mathrm{E}\Big\{\frac{\partial \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^\top}\Big\}\right]^{-1.\top}.
\end{aligned}$$

Therefore, the asymptotic variance of $\hat{\boldsymbol{\beta}}_{LIN}$ is equal to the asymptotic variance of the estimator obtained from the estimating equation $\sum_{i=1}^{n} \mathbf{B}\boldsymbol{g}_i^*(\boldsymbol{\beta}) = \mathbf{0}$. According to Theorem 3.6 in Owen (2001), the asymptotic variance of the empirical likelihood estimator $\hat{\boldsymbol{\beta}}$ is at least as small as that of any estimator obtained from $\sum_{i=1}^{n} \mathbf{C}\boldsymbol{g}_i^*(\boldsymbol{\beta}) = \mathbf{0}$, where $\mathbf{C}$ is an arbitrary $p \times q$ matrix. Therefore, the asymptotic variance of the empirical likelihood estimator $\hat{\boldsymbol{\beta}}$ is at least as small as that of $\hat{\boldsymbol{\beta}}_{LIN}$.

Appendix C Comparison of the asymmetric variances

Without loss of generality, we consider the case where there are two replicate measurements for the covariate. For Lin et al. (2018)'s method, the estimating equation is

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{U}_i(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{W}_{i(1)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(2)}\boldsymbol{\beta}) + \mathbf{W}_{i(2)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(1)}\boldsymbol{\beta}) = \mathbf{0}.$$

The asymptotic variance of $\hat{\boldsymbol{\beta}}_{LIN}$ is

$$\begin{aligned}
&\text{asyVar}(\hat{\boldsymbol{\beta}}_{LIN}) \\
&= \left[\sum_{i=1}^{n} \mathrm{E}\Big\{\frac{\partial \boldsymbol{U}_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^\top}\Big\}\right]^{-1} \left[\sum_{i=1}^{n} \mathrm{E}\{\boldsymbol{U}_i(\boldsymbol{\beta}_0)\boldsymbol{U}_i(\boldsymbol{\beta}_0)^\top\}\right] \left[\sum_{i=1}^{n} \mathrm{E}\Big\{\frac{\partial \boldsymbol{U}_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^\top}\Big\}\right]^{-1.\top} \\
&= \left\{\sum_{i=1}^{n} \mathrm{E}\big(\mathbf{X}_i^\top \Sigma_i^{-1} \mathbf{X}_i\big)\right\}^{-1} \left[\frac{1}{4}\sum_{i=1}^{n} \mathrm{E}\{\boldsymbol{U}_i(\boldsymbol{\beta}_0)\boldsymbol{U}_i(\boldsymbol{\beta}_0)^\top\}\right] \left\{\sum_{i=1}^{n} \mathrm{E}\big(\mathbf{X}_i^\top \Sigma_i^{-1} \mathbf{X}_i\big)\right\}^{-1}, \quad \text{(A1)}
\end{aligned}$$

where $\sum_{i=1}^{n} \mathrm{E}\{\boldsymbol{U}_i(\boldsymbol{\beta}_0)\boldsymbol{U}_i(\boldsymbol{\beta}_0)^\top\}/4 = \mathbf{D} + (\mathbf{E}_1 + \mathbf{E}_2 + \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{G}_1 + \mathbf{G}_2 + \mathbf{G}_3 + \mathbf{G}_3^\top)/4$ with $\mathbf{D} = \sum_{i=1}^{n} \mathrm{E}(\mathbf{X}_i^\top \Sigma_i^{-1} \mathbf{X}_i)$, $\mathbf{E}_1 = \sum_{i=1}^{n} \mathrm{E}\big\{\mathbf{X}_i^\top \Sigma_i^{-1} \mathrm{cov}(\boldsymbol{\delta}_{i(2)}\boldsymbol{\beta}_0) \Sigma_i^{-1} \mathbf{X}_i\big\}$, $\mathbf{E}_2 = \sum_{i=1}^{n} \mathrm{E}\big\{\mathbf{X}_i^\top \Sigma_i^{-1} \mathrm{cov}(\boldsymbol{\delta}_{i(1)}\boldsymbol{\beta}_0) \Sigma_i^{-1} \mathbf{X}_i\big\}$, $\mathbf{F}_1 = \sum_{i=1}^{n} \mathrm{E}(\boldsymbol{\delta}_{i(1)}^\top \Sigma_i^{-1} \boldsymbol{\delta}_{i(1)})$, $\mathbf{F}_2 = \sum_{i=1}^{n} \mathrm{E}(\boldsymbol{\delta}_{i(2)}^\top \Sigma_i^{-1} \boldsymbol{\delta}_{i(2)})$, $\mathbf{G}_1 = \sum_{i=1}^{n} \mathrm{cov}(\boldsymbol{\delta}_{i(1)}^\top \Sigma_i^{-1} \boldsymbol{\delta}_{i(2)}\boldsymbol{\beta}_0)$, $\mathbf{G}_2 = \sum_{i=1}^{n} \mathrm{cov}(\boldsymbol{\delta}_{i(2)}^\top \Sigma_i^{-1} \boldsymbol{\delta}_{i(1)}\boldsymbol{\beta}_0)$, and $\mathbf{G}_3 = \sum_{i=1}^{n} \mathrm{E}(\boldsymbol{\delta}_{i(1)}^\top \Sigma_i^{-1} \boldsymbol{\delta}_{i(2)}\boldsymbol{\beta}_0\boldsymbol{\beta}_0^\top \boldsymbol{\delta}_{i(1)}^\top \Sigma_i^{-1} \boldsymbol{\delta}_{i(2)})$.

For the proposed empirical likelihood method, the auxiliary random vector is

$$\boldsymbol{g}_i(\boldsymbol{\beta}) = \begin{pmatrix} \mathbf{W}_{i(1)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(2)}\boldsymbol{\beta}) \\ \mathbf{W}_{i(2)}^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \mathbf{W}_{i(1)}\boldsymbol{\beta}) \end{pmatrix}.$$

According to Theorem 1, the asymptotic variance of $\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned} &\mathrm{asyVar}(\hat{\boldsymbol{\beta}}) \\ &= \left[\left\{ \sum_{i=1}^n \mathrm{E}\Big(\frac{\partial \boldsymbol{g}_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^\top} \Big) \right\}^\top \left\{ \sum_{i=1}^n \mathrm{E}\big(\boldsymbol{g}_i(\boldsymbol{\beta}_0) \boldsymbol{g}_i(\boldsymbol{\beta}_0)^\top \big) \right\}^{-1} \left\{ \sum_{i=1}^n \mathrm{E}\Big(\frac{\partial \boldsymbol{g}_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^\top} \Big) \right\} \right]^{-1} \\ &= \left\{ \sum_{i=1}^n \mathrm{E}\big(\mathbf{X}_i^\top \Sigma_i^{-1} \mathbf{X}_i \big) \right\}^{-1} \left[\begin{pmatrix} \mathbf{I}_{p\times p} & \mathbf{I}_{p\times p} \end{pmatrix} \left\{ \sum_{i=1}^n \mathrm{E}\big(\boldsymbol{g}_i(\boldsymbol{\beta}_0) \boldsymbol{g}_i(\boldsymbol{\beta}_0)^\top \big) \right\}^{-1} \begin{pmatrix} \mathbf{I}_{p\times p} \\ \mathbf{I}_{p\times p} \end{pmatrix} \right]^{-1} \\ &\quad \times \left\{ \sum_{i=1}^n \mathrm{E}\big(\mathbf{X}_i^\top \Sigma_i^{-1} \mathbf{X}_i \big) \right\}^{-1}, \end{aligned} \tag{A2}$$

where

$$\sum_{i=1}^n \mathrm{E}\big\{ \boldsymbol{g}_i(\boldsymbol{\beta}_0) \boldsymbol{g}_i(\boldsymbol{\beta}_0)^\top \big\} = \begin{pmatrix} \mathbf{D} + \mathbf{E}_1 + \mathbf{F}_1 + \mathbf{G}_1 & \mathbf{D} + \mathbf{G}_3 \\ \mathbf{D} + \mathbf{G}_3^\top & \mathbf{D} + \mathbf{E}_2 + \mathbf{F}_2 + \mathbf{G}_2 \end{pmatrix}.$$

By calculation,

$$\begin{aligned} \Delta \triangleq \frac{1}{4} \sum_{i=1}^n \mathrm{E}\big\{ \boldsymbol{U}_i(\boldsymbol{\beta}_0) \boldsymbol{U}_i(\boldsymbol{\beta}_0)^\top \big\} - \left[\begin{pmatrix} \mathbf{I}_{p\times p} & \mathbf{I}_{p\times p} \end{pmatrix} \left\{ \sum_{i=1}^n \mathrm{E}\big(\boldsymbol{g}_i(\boldsymbol{\beta}_0) \boldsymbol{g}_i(\boldsymbol{\beta}_0)^\top \big) \right\}^{-1} \begin{pmatrix} \mathbf{I}_{p\times p} \\ \mathbf{I}_{p\times p} \end{pmatrix} \right]^{-1} \\ = \frac{1}{4}(\mathbf{E}_1 + \mathbf{E}_2 + \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{G}_1 + \mathbf{G}_2 + \mathbf{G}_3 + \mathbf{G}_3^\top) - \mathbf{E}_1 - \mathbf{F}_1 - \mathbf{G}_1 \\ + (\mathbf{E}_1 + \mathbf{F}_1 + \mathbf{G}_1 - \mathbf{G}_3)(\mathbf{E}_1 + \mathbf{F}_1 + \mathbf{G}_1 - \mathbf{G}_3 + \mathbf{E}_2 + \mathbf{F}_2 + \mathbf{G}_2 - \mathbf{G}_3^\top)^{-1}(\mathbf{E}_1 + \mathbf{F}_1 + \mathbf{G}_1 - \mathbf{G}_3^\top). \end{aligned}$$

Denote $\mathbf{H}_1 = \mathbf{E}_1 + \mathbf{F}_1 + \mathbf{G}_1 - \mathbf{G}_3$ and $\mathbf{H}_2 = \mathbf{E}_2 + \mathbf{F}_2 + \mathbf{G}_2 - \mathbf{G}_3^\top$, then it is easy to verify that if $\mathbf{H}_1 = \mathbf{H}_2$, then $\Delta = 0$. Furthermore, according to (A1) and (A2), $\mathrm{asyVar}(\hat{\boldsymbol{\beta}}_{LIN}) = \mathrm{asyVar}(\hat{\boldsymbol{\beta}})$. Note that the condition of $\mathbf{H}_1 = \mathbf{H}_2$ can be easily satisfied if the two replicate measurements for the same covariate follow the same distribution.

However, the condition of $\mathbf{H}_1 = \mathbf{H}_2$ may be violated if the distributions of two replicate measurements for the same covariate are different. If $\mathbf{H}_1 \neq \mathbf{H}_2$, denote $\mathbf{H}_1 - \mathbf{H}_2$ as $\boldsymbol{\Lambda}$. For convenience, assume $\mathbf{G}_3 = \mathbf{G}_3^\top$, then

$$\begin{aligned} \Delta(\boldsymbol{\Lambda}) &= \mathbf{H}_1(\mathbf{H}_1 + \mathbf{H}_2)^{-1}\mathbf{H}_1 + \frac{1}{4}\mathbf{H}_2 - \frac{3}{4}\mathbf{H}_1 \\ &= \mathbf{H}_1(2\mathbf{H}_1 - \boldsymbol{\Lambda})^{-1}\mathbf{H}_1 - \frac{1}{2}\mathbf{H}_1 - \frac{1}{4}\boldsymbol{\Lambda}. \end{aligned}$$

Obviously, $\Delta(\mathbf{0}) = \mathbf{0}$. Besides, for any $\mathbf{\Lambda} \neq \mathbf{0}$, based on the Woodbury matrix identity,

$$\begin{aligned}\Delta(\mathbf{\Lambda}) - \Delta(\mathbf{0}) &= \mathbf{H}_1\big\{(2\mathbf{H}_1 - \mathbf{\Lambda})^{-1} - (2\mathbf{H}_1)^{-1}\big\}\mathbf{H}_1 - \frac{1}{4}\mathbf{\Lambda} \\ &= \frac{1}{4}\big\{\mathbf{\Lambda}^{-1} - (2\mathbf{H}_1)^{-1}\big\}^{-1} - \frac{1}{4}\mathbf{\Lambda} \\ &= \frac{1}{4}\mathbf{\Lambda}(2\mathbf{H}_1 - \mathbf{\Lambda})^{-1}\mathbf{\Lambda} \\ &= \frac{1}{4}\mathbf{\Lambda}(\mathbf{H}_1 + \mathbf{H}_2)^{-1}\mathbf{\Lambda}.\end{aligned}$$

Note that

$$\mathbf{H}_1 + \mathbf{H}_2 = \begin{pmatrix}\mathbf{I}_{p\times p} & -\mathbf{I}_{p\times p}\end{pmatrix}\left\{\sum_{i=1}^{n}\mathrm{E}(\boldsymbol{g}_i(\boldsymbol{\beta}_0)\boldsymbol{g}_i(\boldsymbol{\beta}_0)^\top)\right\}^{-1}\begin{pmatrix}\mathbf{I}_{p\times p} \\ -\mathbf{I}_{p\times p}\end{pmatrix},$$

then

$$\begin{aligned}&\mathrm{asyVar}(\hat{\boldsymbol{\beta}}_{LIN}) - \mathrm{asyVar}(\hat{\boldsymbol{\beta}}) \\ =&\left\{\sum_{i=1}^{n}\mathrm{E}\big(\mathbf{X}_i^\top\Sigma_i^{-1}\mathbf{X}_i\big)\right\}^{-1}\Delta(\mathbf{\Lambda})\left\{\sum_{i=1}^{n}\mathrm{E}\big(\mathbf{X}_i^\top\Sigma_i^{-1}\mathbf{X}_i\big)\right\}^{-1} \\ =&\frac{1}{4}\left\{\sum_{i=1}^{n}\mathrm{E}\big(\mathbf{X}_i^\top\Sigma_i^{-1}\mathbf{X}_i\big)\right\}^{-1}\mathbf{\Lambda}\left[\begin{pmatrix}\mathbf{I}_{p\times p} & -\mathbf{I}_{p\times p}\end{pmatrix}\left\{\sum_{i=1}^{n}\mathrm{E}\big(\boldsymbol{g}_i(\boldsymbol{\beta}_0)\boldsymbol{g}_i(\boldsymbol{\beta}_0)^\top\big)\right\}^{-1}\begin{pmatrix}\mathbf{I}_{p\times p} \\ -\mathbf{I}_{p\times p}\end{pmatrix}\right]^{-1} \\ &\times\mathbf{\Lambda}\left\{\sum_{i=1}^{n}\mathrm{E}\big(\mathbf{X}_i^\top\Sigma_i^{-1}\mathbf{X}_i\big)\right\}^{-1}.\end{aligned}$$

Since $\sum_{i=1}^{n}\mathrm{E}\{\boldsymbol{g}_i(\boldsymbol{\beta}_0)\boldsymbol{g}_i(\boldsymbol{\beta}_0)^\top\}$ is a positive definite matrix, then $\mathrm{asyVar}(\hat{\boldsymbol{\beta}}_{LIN}) - \mathrm{asyVar}(\hat{\boldsymbol{\beta}}) > \mathbf{0}$.

In summary, if the model satisfies the condition of $\mathbf{H}_1 = \mathbf{H}_2$, then the asymptotic variance of Lin et al. (2018)'s estimator is the same as that of the proposed estimator. This condition can be satisfied naturally if the two replicate measurements for the same covariate follow the same distribution. However, if this condition is violated, the asymptotic variance of Lin et al. (2018)'s estimator becomes larger than that of the proposed estimator. In general, the proposed estimator is asymptotically more efficient than Lin et al. (2018)'s estimator. This conclusion can be easily extended to the general case where there are more than two replicate measurements.

Appendix D Proof of asymptotic properties

Appendix D.1 Lemmas

In order to prove Theorems 1, 3, 4, and 5, we first introduce the following two lemmas.

Lemma D.1. *Assuming that conditions (R.1)–(R.5) hold, we have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{g}_i^*(\boldsymbol{\beta}_0) \rightsquigarrow \mathcal{N}(\mathbf{0},\mathbf{M}).$$

Proof. We first prove that $\sum_{i=1}^{n}\mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)/\sqrt{n}\|^3 \to 0$. Note that

$$\sum_{i=1}^{n}\mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)/\sqrt{n}\|^3 = \sum_{i=1}^{n} n^{-3/2}\mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^3,$$

and

$$\begin{aligned}
\mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^3 \le \mathrm{E}\|\boldsymbol{g}_i(\boldsymbol{\beta}_0)\|^3 &= \mathrm{E}\left[\left\{\sum_{k_1\neq k_2}\|\mathbf{W}_{i(k_1)}^\top\Sigma_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(k_2)}\boldsymbol{\beta}_0)\|^2\right\}^{\frac{3}{2}}\right]\\
&\le \big\{K(K-1)\big\}^{\frac{1}{2}}\sum_{k_1\neq k_2}\mathrm{E}\big\|\mathbf{W}_{i(k_1)}^\top\Sigma_i^{-1}(\boldsymbol{Y}_i-\mathbf{W}_{i(k_2)}\boldsymbol{\beta}_0)\big\|^3\\
&= \big\{K(K-1)\big\}^{\frac{1}{2}}\sum_{k_1\neq k_2}\mathrm{E}\big\|(\mathbf{X}_i+\boldsymbol{\xi}_{i(k_1)})^\top\Sigma_i^{-1}(\boldsymbol{\varepsilon}_i-\boldsymbol{\xi}_{i(k_2)}\boldsymbol{\beta}_0)\big\|^3\\
&\le \big\{K(K-1)\big\}^{\frac{1}{2}}\sum_{k_1\neq k_2}\mathrm{E}\big(\|\mathbf{X}_i+\boldsymbol{\xi}_{i(k_1)}\|^3\|\Sigma_i^{-1}\|^3\|\boldsymbol{\varepsilon}_i-\boldsymbol{\xi}_{i(k_2)}\boldsymbol{\beta}_0\|^3\big).
\end{aligned}$$

By conditions (R.1)–(R.4) and the Markov's inequality, we have $\mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^3 < \infty$. Furthermore, $\sum_{i=1}^{n}\mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)/\sqrt{n}\|^3 \to 0$. Based on condition (R.2), $\mathrm{E}\{\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)/\sqrt{n}\} = \mathbf{0}$. By the law of large numbers and condition (R.5), $\sum_{i=1}^{n}\mathrm{cov}\{\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)/\sqrt{n}\} \to \mathbf{M}$. Therefore, according to Lyapunov central limit theorem, we have $\sum_{i=1}^{n}\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)/\sqrt{n} \rightsquigarrow \mathcal{N}(\mathbf{0},\mathbf{M})$. □

Lemma D.2. *Assuming that conditions (R.1)–(R.5) hold, we have*

(a) $\max_{1\le i\le n}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\| = o_p(n^{1/2})$,

(b) $\sum_{i=1}^{n}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^3/n = o_p(n^{1/2})$, *and*

(c) $\|\boldsymbol{\lambda}(\boldsymbol{\beta}_0)\| = O_p(n^{-1/2})$.

Proof. As pointed out in the proof of Lemma D.1, there exists a positive constant δ such that $\mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^{2+\delta} < \infty$, therefore, $\mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^2 < \infty$. According to the Markov's inequality, $\sum_{i=1}^n \mathrm{P}(\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^2 > n) \leq \sum_{i=1}^n \mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^2/n < \infty$. Hence, by the Borel-Cantelli Lemma, $\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\| > n^{1/2}$ finitely often with probability 1. Let $\max_{1\leq i\leq n} \|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\| = Z_n$, then $Z_n > n^{1/2}$ finitely often. By the same argument, $Z_n > An^{1/2}$ finitely often for any $A > 0$. Therefore, $\limsup_{n\to\infty} Z_n n^{-1/2} \leq A$ with probability 1. So $Z_n = \max_{1\leq i\leq n} \|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\| = o_p(n^{1/2})$.

Note that

$$\sum_{i=1}^n \|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^3/n = \sum_{i=1}^n \|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^2/n \leq Z_n \times \sum_{i=1}^n \|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^2/n.$$

By the law of large numbers, $\sum_{i=1}^n \|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^2/n - \sum_{i=1}^n \mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^2/n \xrightarrow{P} 0$. Since $\mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^2 < \infty$, then $\sum_{i=1}^n \|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^2/n = O_p(1)$. Furthermore, $\sum_{i=1}^n \|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^3/n = o_p(n^{1/2})$.

For simplicity, we denote $\boldsymbol{\lambda}(\boldsymbol{\beta}_0)$ as $\boldsymbol{\lambda}_0$. Write $\boldsymbol{\lambda}_0 = \rho\boldsymbol{\theta}$, where $\rho \geq 0$ and $\|\boldsymbol{\theta}\| = 1$. Since $\pi_i = 1/[n\{1 + \boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\}]$, $\pi_i \geq 0$ and $\sum_{i=1}^n \pi_i = 1$, we have $1 + \boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0) > 0$. As we know, $\boldsymbol{\lambda}_0$ should satisfy the condition of $\mathbf{0} = \sum_{i=1}^n \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)/[n\{1 + \boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\}] \triangleq \boldsymbol{\ell}(\boldsymbol{\lambda}_0)$. Besides,

$$\begin{aligned}
0 =& \|\boldsymbol{\ell}(\rho\boldsymbol{\theta})\| = \|\boldsymbol{\ell}(\rho\boldsymbol{\theta})\|\|\boldsymbol{\theta}\| \geq |\boldsymbol{\theta}^\top \boldsymbol{\ell}(\rho\boldsymbol{\theta})| \\
=& \frac{1}{n}\Big|\boldsymbol{\theta}^\top \Big\{ \sum_{i=1}^n \boldsymbol{g}_i^*(\boldsymbol{\beta}_0) - \rho \sum_{i=1}^n \frac{\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\boldsymbol{\theta}^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}{1 + \rho\boldsymbol{\theta}^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)} \Big\}\Big| \\
\geq& \frac{\rho}{n}\boldsymbol{\theta}^\top \sum_{i=1}^n \frac{\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)^\top}{1 + \rho\boldsymbol{\theta}^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}\boldsymbol{\theta} - \frac{1}{n}\Big|\sum_{j=1}^q \boldsymbol{e}_j^\top \sum_{i=1}^n \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\Big| \\
\geq& \frac{\rho\boldsymbol{\theta}^\top \mathbf{M}_n \boldsymbol{\theta}}{n(1 + \rho Z_n)} - \frac{1}{n}\Big|\sum_{j=1}^q \boldsymbol{e}_j^\top \sum_{i=1}^n \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\Big|,
\end{aligned} \tag{A3}$$

where $\boldsymbol{e}_j$ is a q-dimensional vector with its jth element being 1 and all its other elements being 0. Let σ_1 denote the smallest eigenvalue of $\mathbf{M}$. Since $\mathbf{M}$ is positive definite, then $\sigma_1 > 0$. Under condition (R.5), $\boldsymbol{\theta}^\top \mathbf{M}_n \boldsymbol{\theta}/n \geq \sigma_1 + o_p(1)$. Since $\sum_{i=1}^n \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)/\sqrt{n} \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{M})$, then $|\sum_{j=1}^q \boldsymbol{e}_j^\top \sum_{i=1}^n \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)|/n = O_p(n^{-1/2})$. According to (A3), $\rho/(1 + \rho Z_n) = O_p(n^{-1/2})$. Furthermore, since $Z_n = o_p(n^{1/2})$, we can obtain that $\rho = \|\boldsymbol{\lambda}_0\| = O_p(n^{-1/2})$. □

Appendix D.2 Proof of Theorem 1

Define the empirical log-likelihood ratio as

$$l_E(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log\{1 + \boldsymbol{\lambda}(\boldsymbol{\beta})^\top \boldsymbol{g}_i^*(\boldsymbol{\beta})\}.$$

Denote $\boldsymbol{\beta} = \boldsymbol{\beta}_0 + \boldsymbol{u}n^{-1/3}$, for $\boldsymbol{\beta} \in \{\boldsymbol{\beta} \mid \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| = n^{-1/3}\}$, where $\|\boldsymbol{u}\| = 1$. Similar to the proof of Lemma D.1, under conditions (R.1)–(R.4), $\mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta})\|^3 < \infty$. Similar to the proof of Owen (1990), when $\mathrm{E}\|\boldsymbol{g}_i^*(\boldsymbol{\beta})\|^3 < \infty$ and $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le n^{-1/3}$, we have

$$\begin{aligned} \boldsymbol{\lambda}(\beta) =& \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{g}_i^*(\boldsymbol{\beta}) \boldsymbol{g}_i^*(\boldsymbol{\beta})^\top \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{g}_i^*(\boldsymbol{\beta}) \right\} + o(n^{-1/3}) \quad (a.s.) \\ =& O(n^{-1/3}) \quad (a.s.), \end{aligned} \tag{A4}$$

uniformly about $\theta \in \{\boldsymbol{\beta} \mid \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le n^{-1/3}\}$.

By (A4) and the Taylor expansion, we have (uniformly for $\boldsymbol{u}$),

$$\begin{aligned} l_E(\boldsymbol{\beta}) =& \sum_{i=1}^{n} \boldsymbol{\lambda}(\boldsymbol{\beta})^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}) - \frac{1}{2} \sum_{i=1}^{n} \{\boldsymbol{\lambda}(\boldsymbol{\beta})^\top \boldsymbol{g}_i^*(\boldsymbol{\beta})\}^2 + o(n^{1/3}) \quad (a.s.) \\ =& \frac{n}{2} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{g}_i^*(\boldsymbol{\beta}) \right\}^\top \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{g}_i^*(\boldsymbol{\beta}) \boldsymbol{g}_i^*(\boldsymbol{\beta})^\top \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{g}_i^*(\boldsymbol{\beta}) \right\} \\ & + o(n^{1/3}) \quad (a.s.) \\ =& \frac{n}{2} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{g}_i^*(\boldsymbol{\beta}_0) + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^\top} \boldsymbol{u} n^{-1/3} \right\}^\top \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{g}_i^*(\boldsymbol{\beta}) \boldsymbol{g}_i^*(\boldsymbol{\beta})^\top \right\}^{-1} \\ & \times \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{g}_i^*(\boldsymbol{\beta}_0) + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^\top} \boldsymbol{u} n^{-1/3} \right\} + o(n^{1/3}) \quad (a.s.) \\ =& \frac{n}{2} \left[O\{n^{-1/2} (\log\log n)^{1/2}\} + \mathbf{L}\boldsymbol{u} n^{-1/3} \right]^\top \times \mathbf{M}^{-1} \\ & \times \left[O\{n^{-1/2} (\log\log n)^{1/2}\} + \mathbf{L}\boldsymbol{u} n^{-1/3} \right] + o(n^{1/3}) \quad (a.s.) \\ \ge& (c - \varepsilon) n^{1/3}, \quad a.s., \end{aligned}$$

where $c - \varepsilon > 0$ and c is the smallest eigenvalue of $\mathbf{L}^\top \mathbf{M}^{-1} \mathbf{L}$.

Similarly,

$$\begin{aligned} l_E(\boldsymbol{\beta}_0) =& \frac{n}{2}\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\right\}^\top \left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)^*\right\}^{-1} \\ & \times\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\right\} + o(1) \quad (a.s.) \\ =& O(\log\log n), \quad (a.s.). \end{aligned}$$

Since $l_E(\boldsymbol{\beta})$ is a continuous function about $\boldsymbol{\beta}$ as $\boldsymbol{\beta}$ belongs to the ball $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| \le n^{-1/3}$, then, as $n\to\infty$, $l_E(\boldsymbol{\beta})$ attains its minimum value at some point $\hat{\boldsymbol{\beta}}$ in the interior of this ball with probability 1. Besides, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}(\hat{\boldsymbol{\beta}})$ satisfy $\boldsymbol{Q}_{1n}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\lambda}}) = \mathbf{0}$ and $\boldsymbol{Q}_{2n}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\lambda}}) = \mathbf{0}$, where

$$\boldsymbol{Q}_{1n}(\boldsymbol{\beta},\boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\boldsymbol{g}_i^*(\boldsymbol{\beta})}{1+\boldsymbol{\lambda}^\top\boldsymbol{g}_i^*(\boldsymbol{\beta})},$$

and

$$\boldsymbol{Q}_{2n}(\boldsymbol{\beta},\boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\boldsymbol{\lambda}^\top\boldsymbol{g}_i^*(\boldsymbol{\beta})}\left(\frac{\partial\boldsymbol{g}_i^*(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^\top}\right)^\top\boldsymbol{\lambda}.$$

Expanding $\boldsymbol{Q}_{1n}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\lambda}})$ and $\boldsymbol{Q}_{2n}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\lambda}})$ at $(\boldsymbol{\beta}_0,\mathbf{0})$, by conditions (R.2)–(R.5), we have

$$\mathbf{0} = \boldsymbol{Q}_{1n}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\lambda}}) = \boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0,\mathbf{0}) + \frac{\partial\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0,\mathbf{0})}{\partial\boldsymbol{\beta}^\top}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + \frac{\partial\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0,\mathbf{0})}{\partial\boldsymbol{\lambda}^\top}(\hat{\boldsymbol{\lambda}}-\mathbf{0}) + o_p(\delta_n),$$

and

$$\mathbf{0} = \boldsymbol{Q}_{2n}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\lambda}}) = \boldsymbol{Q}_{2n}(\boldsymbol{\beta}_0,\mathbf{0}) + \frac{\partial\boldsymbol{Q}_{2n}(\boldsymbol{\beta}_0,\mathbf{0})}{\partial\boldsymbol{\beta}^\top}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + \frac{\partial\boldsymbol{Q}_{2n}(\boldsymbol{\beta}_0,\mathbf{0})}{\partial\boldsymbol{\lambda}^\top}(\hat{\boldsymbol{\lambda}}-\mathbf{0}) + o_p(\delta_n),$$

where $\delta_n = \|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\| + \|\hat{\boldsymbol{\lambda}}\|$. Therefore,

$$\begin{pmatrix}\hat{\boldsymbol{\lambda}}\\ \hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\end{pmatrix} = \mathbf{S}_n^{-1}\begin{pmatrix}-\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0,\mathbf{0}) + o_p(\delta_n)\\ o_p(\delta_n)\end{pmatrix}, \tag{A5}$$

where

$$\mathbf{S}_n = \begin{pmatrix}\frac{\partial\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0,\mathbf{0})}{\partial\boldsymbol{\lambda}^\top} & \frac{\partial\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0,\mathbf{0})}{\partial\boldsymbol{\beta}^\top}\\ \frac{\partial\boldsymbol{Q}_{2n}(\boldsymbol{\beta}_0,\mathbf{0})}{\partial\boldsymbol{\lambda}^\top} & \mathbf{0}\end{pmatrix} \xrightarrow{P} \mathbf{S} = \begin{pmatrix}-\mathbf{M} & \mathbf{L}\\ \mathbf{L}^\top & \mathbf{0}\end{pmatrix}. \tag{A6}$$

From Lemma D.1, we have $\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0,\mathbf{0}) = \sum_{i=1}^{n}\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)/n = O_p(n^{-1/2})$. Besides, from (A5) and (A6), we know that $\delta_n = O_p(n^{-1/2})$ and

$$\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = (\mathbf{L}^\top\mathbf{M}\mathbf{L})^{-1}\mathbf{L}^\top\mathbf{M}^{-1}\sqrt{n}\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0,\mathbf{0}) + o_p(1).$$

Furthermore, from Lemma D.1, we have $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \rightsquigarrow \mathcal{N}\big(\mathbf{0},(\mathbf{L}^\top\mathbf{M}^{-1}\mathbf{L})^{-1}\big)$.

Appendix D.3 Proof of Theorem 3

Applying the Taylor expansion to the formula (7), we have

$$-2\log R(\boldsymbol{\beta}_0) = 2\sum_{i=1}^{n}\Big[\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0) - \frac{1}{2}\big\{\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\big\}^2\Big] + \epsilon_n,$$

where $\epsilon_n \le c\sum_{i=1}^n |\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)|^3$ in probability for some bounded positive constant c. Because of the fact that $\sum_{i=1}^n \|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^3/n = o_p(n^{1/2})$ and $\|\boldsymbol{\lambda}_0\| = O_p(n^{-1/2})$, we have

$$\epsilon_n \le c\sum_{i=1}^{n}|\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)|^3 = O_p(n^{-3/2})o_p(n^{3/2}) = o_p(1).$$

So

$$-2\log R(\boldsymbol{\beta}_0) = 2\sum_{i=1}^{n}\Big[\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0) - \frac{1}{2}\big\{\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\big\}^2\Big] + o_p(1). \tag{A7}$$

Note that

$$\begin{aligned}
\mathbf{0} =& \frac{1}{n}\sum_{i=1}^{n}\frac{\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}{1+\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)} \\
=& \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\Big[1-\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0) + \frac{\big\{\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\big\}^2}{1+\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}\Big] \\
=& \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i^*(\boldsymbol{\beta}_0) - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)^\top\boldsymbol{\lambda}_0 + \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\frac{\big\{\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\big\}^2}{1+\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}.
\end{aligned} \tag{A8}$$

By Lemma D.2,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\frac{\big\{\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\big\}^2}{1+\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}\right\| &\le \frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\|^3\|\boldsymbol{\lambda}_0\|^2|1+\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)|^{-1} \\
&= o_p(n^{1/2})O_p(n^{-1})O_p(1) = o_p(n^{-1/2}).
\end{aligned} \tag{A9}$$

Therefore, from (A8), (A9), and condition (R.5), we can obtain that

$$\boldsymbol{\lambda}_0 = \Big\{\sum_{i=1}^{n}\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)^\top\Big\}^{-1}\sum_{i=1}^{n}\boldsymbol{g}_i^*(\boldsymbol{\beta}_0) + o_p(n^{-1/2}).$$

Substituting the value of $\boldsymbol{\lambda}_0$ into (A7), we have

$$\begin{aligned}
-2\log R(\boldsymbol{\beta}_0) &= \sum_{i=1}^{n}\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0) + o_p(1) \\
&= \Big\{\frac{\sum_{i=1}^n \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}{\sqrt{n}}\Big\}^\top\Big\{\frac{\sum_{i=1}^n \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)^\top}{n}\Big\}^{-1}\Big\{\frac{\sum_{i=1}^n \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}{\sqrt{n}}\Big\} + o_p(1) \\
&= \Big\{\mathbf{M}^{-1/2}\frac{\sum_{i=1}^n \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}{\sqrt{n}}\Big\}^\top\big(\mathbf{M}^{-1/2}\mathbf{M}\mathbf{M}^{-1/2}\big)^{-1}\Big\{\mathbf{M}^{-1/2}\frac{\sum_{i=1}^n \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}{\sqrt{n}}\Big\} + o_p(1).
\end{aligned} \tag{A10}$$

Since $\mathbf{M}^{-1/2}\sum_{i=1}^{n}\boldsymbol{g}_i^*(\boldsymbol{\beta}_0)/\sqrt{n} \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{I}_{q\times q})$ and the rank of $\mathbf{M}$ is q, we have $-2\log R(\boldsymbol{\beta}_0) \rightsquigarrow \chi^2(q)$.

Appendix D.4 Proof of Theorem 4

According to (7), the empirical likelihood ratio test statistic is

$$W_1(\boldsymbol{\beta}_0) = 2\left[\sum_{i=1}^{n}\log\left\{1+\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\right\} - \sum_{i=1}^{n}\log\left\{1+\hat{\boldsymbol{\lambda}}^\top \boldsymbol{g}_i^*(\hat{\boldsymbol{\beta}})\right\}\right].$$

By (7), (A10), and the definition of $\boldsymbol{Q}_{1n}(\boldsymbol{\beta}, \boldsymbol{\lambda})$, we know that

$$\sum_{i=1}^{n}\log\left\{1+\boldsymbol{\lambda}_0^\top \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)\right\} = \frac{n}{2}\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0, \mathbf{0})^\top \mathbf{M}^{-1}\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0, \mathbf{0}) + o_p(1).$$

By the Taylor expansion, we can obtain that

$$\sum_{i=1}^{n}\log\left\{1+\hat{\boldsymbol{\lambda}}^\top \boldsymbol{g}_i^*(\hat{\boldsymbol{\beta}})\right\} = \frac{n}{2}\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0, \mathbf{0})^\top \mathbf{V}\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0, \mathbf{0}) + o_p(1),$$

where $\mathbf{V} = \mathbf{M}^{-1}\left\{\mathbf{I}_{q\times q} - \mathbf{L}(\mathbf{L}^\top\mathbf{M}^{-1}\mathbf{L})^{-1}\mathbf{L}^\top\mathbf{M}^{-1}\right\}$. Therefore,

$$\begin{aligned}
W_1(\boldsymbol{\beta}_0) =& n\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0, \mathbf{0})^\top(\mathbf{M}^{-1} - \mathbf{V})\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0, \mathbf{0}) + o_p(1)\\
=& n\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0, \mathbf{0})^\top\mathbf{M}^{-1}\mathbf{L}(\mathbf{L}^\top\mathbf{M}^{-1}\mathbf{L})^{-1}\mathbf{L}^\top\mathbf{M}^{-1}\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0, \mathbf{0}) + o_p(1)\\
=& \left\{\mathbf{M}^{-1/2}\sqrt{n}\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0, \mathbf{0})\right\}^\top\left\{\mathbf{M}^{-1/2}\mathbf{L}(\mathbf{L}^\top\mathbf{M}^{-1}\mathbf{L})^{-1}\mathbf{L}^\top\mathbf{M}^{-1/2}\right\}\\
&\times\left\{\mathbf{M}^{-1/2}\sqrt{n}\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0, \mathbf{0})\right\} + o_p(1).
\end{aligned}$$

By Lemma D.1, $\mathbf{M}^{-1/2}\sqrt{n}\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0, \mathbf{0}) \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{I}_{q\times q})$. By calculation, $\mathbf{M}^{-1/2}\mathbf{L}(\mathbf{L}^\top\mathbf{M}^{-1}\mathbf{L})^{-1}\mathbf{L}^\top\mathbf{M}^{-1/2}$ is a symmetric and idempotent matrix with the rank of p. Hence the empirical likelihood ratio statistic $W_1(\boldsymbol{\beta}_0) \rightsquigarrow \chi^2(p)$.

Appendix D.5 Proof of Theorem 5

It is obvious that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^\top} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \boldsymbol{g}_i^*(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)}{\boldsymbol{\beta}_1^\top}, \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \boldsymbol{g}_i^*(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)}{\partial \boldsymbol{\beta}_2^\top}\right).$$

By condition (R.5), $n^{-1}\sum_{i=1}^{n}\partial \boldsymbol{g}_i^*(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}^\top \xrightarrow{P} \mathbf{L}$. Thus, $n^{-1}\sum_{i=1}^{n}\partial \boldsymbol{g}_i^*(\boldsymbol{\beta}_1^0,\boldsymbol{\beta}_2^0)/\partial\boldsymbol{\beta}_1^\top \xrightarrow{P} \mathbf{L}_1$ and $n^{-1}\sum_{i=1}^{n}\partial \boldsymbol{g}_i^*(\boldsymbol{\beta}_1^0,\boldsymbol{\beta}_2^0)/\partial\boldsymbol{\beta}_2^\top \xrightarrow{P} \mathbf{L}_2$, where $\mathbf{L}_1$ and $\mathbf{L}_2$ are the corresponding components of $\mathbf{L}$. By the Taylor expansion, we can obtain that

$$\begin{aligned}
W_2 =& -2\log\left\{R(\boldsymbol{\beta}_1^0,\hat{\boldsymbol{\beta}}_2^0)\right\} + 2\log\left\{R(\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2)\right\} \\
=& \left\{\mathbf{M}^{-1/2}\sqrt{n}\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0,\mathbf{0})\right\}^\top \mathbf{M}^{-1/2}\left\{\mathbf{L}(\mathbf{L}^\top\mathbf{M}^{-1}\mathbf{L})^{-1}\mathbf{L}^\top - \mathbf{L}_2(\mathbf{L}_2^\top\mathbf{M}^{-1}\mathbf{L}_2)^{-1}\mathbf{L}_2^\top\right\} \\
& \times \mathbf{M}^{-1/2}\left\{\mathbf{M}^{-1/2}\sqrt{n}\boldsymbol{Q}_{1n}(\boldsymbol{\beta}_0,\mathbf{0})\right\} + o_p(1).
\end{aligned}$$

Note that

$$\begin{aligned}
& \mathbf{L}(\mathbf{L}^\top\mathbf{M}^{-1}\mathbf{L})^{-1}\mathbf{L}^\top \\
=& \begin{pmatrix}\mathbf{L}_1 & \mathbf{L}_2\end{pmatrix}\left\{\begin{pmatrix}\mathbf{L}_1^\top \\ \mathbf{L}_2^\top\end{pmatrix}\mathbf{M}^{-1}\begin{pmatrix}\mathbf{L}_1 & \mathbf{L}_2\end{pmatrix}\right\}^{-1}\begin{pmatrix}\mathbf{L}_1^\top \\ \mathbf{L}_2^\top\end{pmatrix} \\
\geq& \begin{pmatrix}\mathbf{L}_1 & \mathbf{L}_2\end{pmatrix}\begin{pmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{L}_2^\top\mathbf{M}^{-1}\mathbf{L}_2\right)^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{L}_1^\top \\ \mathbf{L}_2^\top\end{pmatrix} \\
=& \mathbf{L}_2\left(\mathbf{L}_2^\top\mathbf{M}^{-1}\mathbf{L}_2\right)^{-1}\mathbf{L}_2^\top.
\end{aligned}$$

Therefore, $\mathbf{L}(\mathbf{L}^\top\mathbf{M}^{-1}\mathbf{L})^{-1}\mathbf{L}^\top - \mathbf{L}_2(\mathbf{L}_2^\top\mathbf{M}^{-1}\mathbf{L}_2)^{-1}\mathbf{L}_2^\top$ is non-negative definite. Since the matrices of $\mathbf{M}^{-1/2}\mathbf{L}(\mathbf{L}^\top\mathbf{M}^{-1}\mathbf{L})^{-1}\mathbf{L}^\top\mathbf{M}^{-1/2}$ and $\mathbf{M}^{-1/2}\mathbf{L}_2(\mathbf{L}_2^\top\mathbf{M}^{-1}\mathbf{L}_2)^{-1}\times\mathbf{L}_2^\top\mathbf{M}^{-1/2}$ are symmetric and idempotent, with the ranks of p and $p-r$, respectively, then the empirical likelihood ratio statistic $W_2 \rightsquigarrow \chi^2(r)$.