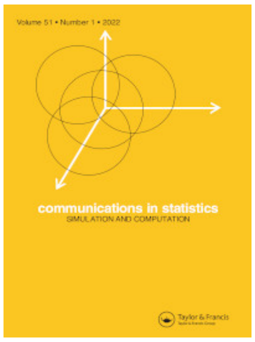

Communications in Statistics - Simulation and Computation

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/lssp20

A methodology for improving efficiency estimation based on conditional mix-GEE models in longitudinal studies

Yanchun Xing, Ma Wenqing & Chunhui Liang

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
2022, VOL. 51, NO. 1, 254–265
https://doi.org/10.1080/03610918.2019.1649423

Taylor & Francis
Taylor & Francis Group

Check for updates

A methodology for improving efficiency estimation based on conditional mix-GEE models in longitudinal studies

Yanchun Xing[a,b], Ma Wenqing[b], and Chunhui Liang[b]

[a]School of Statistics, Jilin University of Finance and Economics, Changchun, Jilin, China; [b]School of Mathematics and Statistics, Northeast Normal University, Changchun, Jilin, China

ABSTRACT
Estimating random effects accurately is crucial since it reflects the subject-specific effect in longitudinal studies. In this paper, we develop a new methodology for improving the efficiency of fixed-effects and random-effects estimation based on conditional mix-GEE models. The advantage of our proposed approach is that the serial correlation over time was accommodated in estimating random effects. Meanwhile, the normality assumption for random effects is not required. In addition, according to the estimates of some mixture proportions, the true working correlation matrix can be identified. The feature of our proposed approach is that the estimators of the regression parameters are more efficient than CCQIF, cmix-GEE and CQIF approaches even if the working correlation structure is not correctly specified. In theory, we show that the proposed method yields a consistent and more efficient estimator than the random-effect estimator that ignores correlation information from longitudinal data. We establish asymptotic results for both fixed-effects and random-effects estimators. Simulation studies confirm the performance of our proposed method.

ARTICLE HISTORY
Received 3 January 2019
Accepted 23 July 2019

KEYWORDS
Conditional mix-GEE; Random effects; Quadratic inference functions; Longitudinal data

1. Introduction

Longitudinal data arise frequently in many studies where repeated measurements from a same subject are correlated. Identifying the true correlation structure of longitudinal data is important for appropriate statistical analysis. For longitudinal data analysis, it is important to incorporate the correlation among repeated measurements since utilizing the true correlation structure can improve the efficiency of regression parameter estimation and reduce the bias of the estimation (Wang 2003). Xu et al. (2012) proposed the mixture generalize estimation equation method (mix-GEE) which views the working correlation matrix by a combination of many working correlation structures. In addition to obtaining more efficient fixed-effects estimates, their method can also identify the true correlation structure. On the other hand, generalized linear mixed-effects model (GLMM) has been widely used to analyze correlated longitudinal data when the subject-specific effect is one of our interests. Standard mixed-effects models assuming normality of random effects (see, e.g. Laird and Ware 1982; Breslow and Clayton 1993; McCulloch

CONTACT Ma Wenqing wenqingma@nenu.edu.cn School of Mathematics and Statistics, Northeast Normal University, Changchun, Jilin 130024, China.

1997; Jiang and Zhang 2001; Vonesh et al. 2002; McCulloch, Searle, and Neuhaus 2008; Diaz et al. 2007; Diaz, Yeh, and Leon 2012; Fang, Zhang, and Sun 2016) are major tools to describe such longitudinal data. However, a normality assumption for random effects could not be satisfied in practice on some occasions. This could affect the bias and efficiency of the fixed-effects estimators. Moreover, even conditional on the random effects, there may exist some serial correlation within the same cluster. To handle these correlations, Wang, Tsai, and Qu (2012) proposed a new approach which is defined as conditional quadratic inference function (CQIF) to estimate both fixed and random effects. However, CQIF approach cannot identify the true correlation structure. Xing et al. (2018) developed a conditional mixture generalized estimation equation (cmix-GEE) to improve the efficiency of the fixed-effects estimators and identify the true working correlation structure for the correlated data. Specifically, estimating random effects accurately is a key step since it reflects the subject-specific effect and affects the efficiency of the fixed-effects estimation. Cho, Wang, and Qu (2017) estimated unobserved subject-specific treatment effects through conditional random-effects modeling and used the results to personalize treatment for longitudinal data. In this paper, based on conditional mix-GEE models, we propose a new methodology for improving efficiency of fixed-effects and random-effects estimation. In theory, we show that the proposed method yields a consistent and more efficient estimator than the random-effect estimator that ignores correlation information from longitudinal data. Simulation results show that the accuracy of the random effects is increased and the efficiency of the fixed effects is greatly improved too. In addition, we can identify the true working correlation structure.

The paper is organized as follows. Section 2 describes the existing cmix-GEE approaches firstly, then introduces the new methodology for improving the efficiency of the fixed effects and random effects estimation and provides the computational algorithm and its asymptotic properties for the proposed method. Simulation studies and conclusions are given in Sec. 3 and Sec. 4. Proofs and necessary conditions are provided in the Appendix.

2. Notation and framework

In this section, the cmix-GEE method for longitudinal data will be firstly described. Then, the methodology for improving the efficiency of the fixed effects and random effects estimation based on the cmix-GEE method will be presented.

2.1. Cmix-GEE method

For longitudinal data, Let y_{ij} be a response variable measured at the jth time point for the ith subject, $i = 1, ..., N, j = 1, ..., T$. $\boldsymbol{X}_i$ be the corresponding known $T \times p$ covariate matrix associated with a p-dimensional vector of fixed effects $\boldsymbol{\beta}$. Then, the observed data can be written as $\boldsymbol{y}_i = (y_{i1}, ..., y_{iT})^T$. When there is a strong indication of individual variations, it is more sensible to apply a random-effects model to analyze longitudinal data. For the generalized linear mixed model (GLMM), given the random effects $\boldsymbol{b}_i$, the conditional mean of the response $E(\boldsymbol{y}_i|\boldsymbol{b}_i) = \boldsymbol{\mu}_i^b$ is a function of the linear predictor $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$, that is, $g(\boldsymbol{\mu}_i^b) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$, where $g(\cdot)$ is a known link function, $\boldsymbol{b}_i$ is a $q \times 1$ vector of random effects, and $\boldsymbol{Z}_i$ is the covariate associated with random effect $\boldsymbol{b}_i$.

Given random effects $\boldsymbol{b}_i$, if the conditional likelihood of $\boldsymbol{y}_i$ is unknown, quasi-likelihood equation (Wedderburn 1974) was used to obtain fixed effects and random effects estimators. Specifically, to ensure the identification for the fixed effect $\boldsymbol{\beta}$ and random effects $\boldsymbol{b}_i$, a constraint $\boldsymbol{P}_A\boldsymbol{b} = 0$ was imposed in Jiang (1999), where $\boldsymbol{P}_A$ is a known projection matrix. Then the quasi-score equations corresponding to $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$ are

$$\sum_{i=1}^{N}\left(\frac{\partial \boldsymbol{\mu}_i^b}{\partial \boldsymbol{\beta}}\right)^T \boldsymbol{W}_i^{b-1}\left(\boldsymbol{y}_i - \boldsymbol{\mu}_i^b\right) = 0 \tag{1}$$

and

$$\begin{pmatrix} h_1 = \left(\frac{\partial \boldsymbol{\mu}_1^b}{\partial \boldsymbol{b}_1}\right)^T \boldsymbol{W}_1^{b-1}\left(\boldsymbol{y}_1 - \boldsymbol{\mu}_1^{b_1}\right) - \lambda \frac{\partial \boldsymbol{P}_A\boldsymbol{b}}{\partial \boldsymbol{b}_1}\boldsymbol{P}_A\boldsymbol{b} = 0 \\ \vdots \\ h_N = \left(\frac{\partial \boldsymbol{\mu}_N^b}{\partial \boldsymbol{b}_N}\right)^T \boldsymbol{W}_N^{b-1}\left(\boldsymbol{y}_N - \boldsymbol{\mu}_N^{b_N}\right) - \lambda \frac{\partial \boldsymbol{P}_A\boldsymbol{b}}{\partial \boldsymbol{b}_N}\boldsymbol{P}_A\boldsymbol{b} = 0 \end{pmatrix} \tag{2}$$

where $\boldsymbol{W}_i^b$ is a working correlation matrix conditional on the random effects, λ is a Lagrange multiplier. In the PQL approach, $\boldsymbol{W}_i^b$ is diagonal. However, $\boldsymbol{W}_i^b = var(\boldsymbol{y}_i|\boldsymbol{b}_i)$ is not necessarily a diagonal matrix, since conditional on random effects $\boldsymbol{b}_i$, the repeated measurements within a subject may be correlated. Wang, Tsai, and Qu (2012) developed the conditional quadratic inference function (CQIF) method to improve the efficiency of the fixed effects estimators. The main idea of their approach is that the inverse $\boldsymbol{R}^{-1}$ of the working correlation in $\boldsymbol{W}_i^b = \boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{R}\boldsymbol{A}_i^{\frac{1}{2}}$ was approximated by a class of linear combinations of known matrices $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_m$, that is,

$$\boldsymbol{R}^{-1} \approx a_1\boldsymbol{M}_1 + \cdots + a_m\boldsymbol{M}_m \approx \sum_{j=1}^{m} a_j\boldsymbol{M}_j$$

where $\boldsymbol{A}_i = diag(var(y_{i1}|\boldsymbol{b}), \ldots, var(y_{iT}|\boldsymbol{b}))$. The advantage of the CQIF approach is that the objective functions were proposed for inference. Given the random effects, CQIF method considered the serial correlation within the same cluster and the distribution assumption for the random effects was not required. However, CQIF method cannot identify the true correlation structure for the correlated data. Wang and Carey (2003) indicated that when the correlation structure is misspecified, the efficiency of the estimators for regression parameters can be seriously affected. Condition mix-GEE approach was proposed by Xing et al. (2018) to identify the true correlation structure for longitudinal data with unspecified random-effects distributions. The basic idea is that conditional on the random effects, the working correlation matrix $\boldsymbol{R}(\alpha)$ is represented by a finite combination of many working correlation structures, that is,

$$\boldsymbol{R}(\alpha) = \pi_1\boldsymbol{R}^{(1)}(\alpha_1) + \pi_2\boldsymbol{R}^{(2)}(\alpha_2) + \ldots + \pi_L\boldsymbol{R}^{(L)}(\alpha_L) = \sum_{l=1}^{L}\pi_l\boldsymbol{R}^{(l)}(\alpha_l)$$

where the coefficient π_l of working correlation $\boldsymbol{R}^{(l)}$ represents its proportion in the sample from the finite mixture of distributions with different correlation structures. Then, given the random effects $\boldsymbol{b}_i$, the covariance matrix of $\boldsymbol{y}_i$ can be expressed as

$$Cov(\mathbf{y}_i|\boldsymbol{b}_i) = \boldsymbol{A}_i^{\frac{1}{2}}\left[\sum_{l=1}^{L}\pi_l\boldsymbol{R}^l(\alpha_l)\right]\boldsymbol{A}_i^{\frac{1}{2}}. \tag{3}$$

Given the random effects $\boldsymbol{b}_i$ and nuisance parameters π_l, α_l, the estimators of the fixed-effects $\boldsymbol{\beta}$ can be obtained by solving the following equation:

$$\sum_{i=1}^{N}\left(\frac{\partial\boldsymbol{\mu}_i^b}{\partial\boldsymbol{\beta}}\right)^T\boldsymbol{A}_i^{-1/2}\left[\sum_{l=1}^{L}\pi_l\boldsymbol{R}^l(\alpha_l)\right]^{-1}\boldsymbol{A}_i^{-1/2}\left(\boldsymbol{y}_i - \boldsymbol{\mu}_i^b\right) = 0. \tag{4}$$

To estimate the random effects $\boldsymbol{b}_i$, extended scores for random effects was proposed by Xing et al. (2018) as

$$G_N^r = \left\{\left(g_1^I\right)^T, \lambda\boldsymbol{b}_1^T, \ldots, \left(g_N^I\right)^T, \lambda\boldsymbol{b}_N^T, \lambda(\boldsymbol{P}_A\boldsymbol{b})^T\right\}^T \tag{5}$$

where $g_i^I = (\frac{\partial\boldsymbol{\mu}_i^b}{\partial\boldsymbol{b}_i})^T\boldsymbol{A}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i^b)$, the Lagrange multiplier λ is chosen to be log(N), $\boldsymbol{P}_A = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ is a known projection matrix on the null space of $(\boldsymbol{I} - \boldsymbol{P}_X)\boldsymbol{Z}$, and $\boldsymbol{P}_X$ is defined similarly as $\boldsymbol{P}_A$ with covariates $\boldsymbol{X}$ and $\boldsymbol{Z}$ associated with fixed effects and random effects, respectively.

Given fixed-effects $\boldsymbol{\beta}$, the random effects estimators can be obtained by minimizing the objective function

$$\left(G_N^r\right)^T\left(G_N^r\right). \tag{6}$$

Then,the estimators for the fixed-effects $\boldsymbol{\beta}$ and random-effects $\boldsymbol{b}_i$ can be obtained by an iterative algorithm.

2.2. A new methodology for improving efficiency estimation

Cmix-GEE method can accommodate the serial correlation within the same cluster and dose not requires the normality assumption for random effects. The feature of cmix-GEE approach is that the estimators of the regression parameters are more efficient than CQIF even if the working correlation structure is not correctly specified. In addition, according to the estimated proportions, the true correlation structure can be identified (Xing et al. 2018). However, they do not accommodate the serial correlation information from longitudinal observations in obtaining the estimators of $\boldsymbol{b}_i$. It is important to obtain the more accuracy estimators of random effects when the subject-specific effect is one of our interests (Cho, Wang, and Qu 2017). At the same time, the accuracy of the estimators for random effects could affect the efficiency of the estimator of fixed effects $\boldsymbol{\beta}$.

A new methodology was developed to improve the accuracy of the random-effects estimation without the normality assumption for the random effects. The crucial step relies on accurate estimation for the random effects $\boldsymbol{b}_i$ in formulation (6). We formulate the estimating equations corresponding to the random-effects as

$$g_i^C = \left(\frac{\partial\boldsymbol{\mu}_i^b}{\partial\boldsymbol{b}_i}\right)^T\boldsymbol{A}_i^{-1/2}C^{-1}\boldsymbol{A}_i^{-1/2}\left(\boldsymbol{y}_i - \boldsymbol{\mu}_i^b\right)$$

where $C = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{\mu}_i^b)(\boldsymbol{y}_i - \boldsymbol{\mu}_i^b)^T$ is the correlation matrix estimator based on the method of moments. Then the extended scores with constraints of the mean and variance for the random effects $\boldsymbol{b}$ is defined as:

$$G_N^C = \left\{ \left(g_1^C\right)^T, \lambda \boldsymbol{b}_1^T, ..., \left(g_N^C\right)^T, \lambda \boldsymbol{b}_N^T, \lambda (\boldsymbol{P}_A \boldsymbol{b})^T \right\}^T \tag{7}$$

where the Lagrange multiplier λ is chosen to be log(N). Given the fixed-effects $\boldsymbol{\beta}$, the random effects estimators can be obtained by minimizing the inference function

$$\left(G_N^C\right)^T \left(G_N^C\right). \tag{8}$$

Then, the estimators for the fixed-effects $\boldsymbol{\beta}$ and random-effects $\boldsymbol{b}_i$ can be obtained by an iterative algorithm using Eq. (4) and inference function in (8). The proposed approach for obtaining the estimator of random effects $\boldsymbol{b}_i$ could improve the efficiency of fixed-effects and random-effects estimators based on the cmix-GEE models. Our proposed approach can also address unbalanced longitudinal data, the strategy is similar to Xing et al. (2018). So we omit here.

2.3. Implementation and algorithm

In this section, we provide an algorithm to estimate fixed-effects and random-effects parameters. Conditional on the random-effects $\boldsymbol{b}_i$ and fixed-effects $\boldsymbol{\beta}$, an pseudo likelihood expectation maximization (PL-EM) algorithm was used to obtain the estimator of nuisance parameters $\psi = (\alpha_l, \pi_l)_{l=1}^L$ in Eq. (4). The details of estimation can be seen in Sec. 2.4 (Xing et al., 2018). Given the estimated random-effects $\hat{\boldsymbol{b}}_i$ and estimated nuisance parameters $\hat{\psi}$, the regression parameters $\boldsymbol{\beta}$ can be obtained by plugging the estimated working correlation matrix $\hat{\boldsymbol{R}} = \sum_{l=1}^{L} \hat{\pi}_l \boldsymbol{R}^l(\hat{\alpha}_l)$ into (4), which leads to the following iterative algorithm.

Step 1. Set the initial value of random effects as $\hat{\mathbf{b}} = 0$;
Step 2. Obtain the initial $\hat{\boldsymbol{\beta}}^0$ from the generalized linear model assuming independent correlation structure;
Step 3. At the kth step, apply the PL-EM algorithm to obtain $(\hat{\alpha}_l^k, \hat{\pi}_l^k)_{l=1}^L$;
Step 4. Let $\hat{\boldsymbol{R}}^k = \sum_{l=1}^{L} \hat{\pi}_l^k \boldsymbol{R}^l(\hat{\alpha}_l^k)$, compute $\hat{\boldsymbol{\beta}}^k$ by solving the Eq. (4);
Step 5. Update the correlation matrix C using the $\hat{\boldsymbol{\beta}}^k$ and $\hat{\boldsymbol{b}}^{(k-1)}$;
Step 6. Given $\hat{\boldsymbol{R}}^k, \hat{\boldsymbol{\beta}}^k$ and C, $\hat{\boldsymbol{b}}^k$ can be obtained by minimizing (8);
Step 7. Repeat Steps 3-6 until the convergence criterion is reached.

2.4. Asymptotic properties

In this section, we investigate the asymptotic properties of our proposed approach estimators. Let $\boldsymbol{b}_0 = (\boldsymbol{b}_{01}', ..., \boldsymbol{b}_{0N}')'$ be the true realization of the random effects and let $\hat{b} = (\hat{\boldsymbol{b}}_1', ..., \hat{\boldsymbol{b}}_N')'$ be the corresponding random effects estimators, where $\boldsymbol{b}_{0i}$ and $\hat{b}_i$ are $q \times 1$ vectors of random effects for the ith subject.

Theorem 1. *Under the regularity conditions, the estimator of nuisance parameters* $\hat{\psi} = (\hat{\alpha}_l, \hat{\pi}_l)_{l=1}^L$ *has* $\sqrt{N}(\hat{\psi} - \psi^*) = O_p(1)$. *Where* $\hat{\psi} = \hat{\psi}(\boldsymbol{\beta}|\boldsymbol{b}), \psi^* = \psi^*(\boldsymbol{\beta}|\boldsymbol{b})$.

Proof. Noting that given any random effects $\boldsymbol{b}_i, \hat{\psi}$ and ψ^* are actually functions of $\boldsymbol{\beta}$. The detailed proof of this theorem is similar to Theorem 2.1 in Xing et al. (2018). So, the proof is omitted here. □

Let $\boldsymbol{\beta}_0$ and b_0 be the true value of the fixed effects parameter and random effects respectively. Let $\hat{\boldsymbol{\beta}}_0$ be a solution of the following equation:

$$U\left(\boldsymbol{\beta}, M\left(\hat{\psi}(\boldsymbol{\beta})\right)|\boldsymbol{b}_0\right) = \frac{1}{N}\sum_{i=1}^{N} U_i\left(\boldsymbol{\beta}, M\left(\hat{\psi}(\boldsymbol{\beta})\right)|\boldsymbol{b}_0\right) = 0$$

where $U_i(\boldsymbol{\beta}, M|\boldsymbol{b}) = ((\frac{\partial \boldsymbol{\mu}_i^b}{\partial \boldsymbol{\beta}})^T \boldsymbol{A}_i^{-1/2}\boldsymbol{M}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_i^b))$. For any given random effects $\boldsymbol{b}$, $\hat{\boldsymbol{\beta}}^M$ is the solution of the equation

$$U(\boldsymbol{\beta}, M|\boldsymbol{b}) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial \boldsymbol{\mu}_i^b}{\partial \boldsymbol{\beta}}\right)^T \boldsymbol{A}_i^{-1/2}\boldsymbol{M}^{-1}\boldsymbol{A}_i^{-1/2}\left(\boldsymbol{y}_i - \boldsymbol{\mu}_i^b\right)) = 0$$

where $M = M(\hat{\psi}(\boldsymbol{\beta})) = \sum_{i=1}^{N}\hat{\pi}_l \boldsymbol{R}^l(\hat{\alpha}_l)$.

Then the asymptotic normality properties of regressive parameters were obtain in Theorem 2.

Theorem 2. *Under the regularity conditions provided in the Appendix,*

I. $\sqrt{N}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \boldsymbol{\Omega}_0)$, *where* $\boldsymbol{\Omega}_0 = \lim_{N\to\infty}(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{H}_i)^{-1}(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{G}_i)\ (\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{H}_i)^{-1}$, $\boldsymbol{H}_i = \boldsymbol{D}_i^T\boldsymbol{A}_i^{-1/2}(\boldsymbol{R}^*)^{-1}\boldsymbol{A}_i^{-1/2}\boldsymbol{D}_i$, $\boldsymbol{G}_i = \boldsymbol{D}_i^T\boldsymbol{A}_i^{-1/2}(\boldsymbol{R}^*)^{-1}\tilde{\boldsymbol{R}}(\boldsymbol{R}^*)^{-1}\boldsymbol{A}_i^{-1/2}\boldsymbol{D}_i$. $\tilde{\boldsymbol{R}}$ *is a true correlation matrix of* $\boldsymbol{y}_i$, $\boldsymbol{R}^* = \sum_{l=1}^{L}\pi_l^*\boldsymbol{R}^l(\alpha_l^*)$;

II. $\hat{\boldsymbol{\beta}}^M \xrightarrow{p} \boldsymbol{\beta}_0$;

III. $\sqrt{N}(\hat{\boldsymbol{\beta}}^M - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \boldsymbol{\Omega}_1)$, *where*

$$\boldsymbol{\Omega}_1 = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{H}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i^*\right)\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{H}_i\right)^{-1},$$

$$\boldsymbol{\Sigma}_i^* = E\left[\left(U_i\left(\hat{\boldsymbol{\beta}}^M, M(\hat{\psi})|\boldsymbol{b}_0\right) - U_i\left(\boldsymbol{\beta}_0, M(\hat{\psi})|\boldsymbol{b}_0\right)\right)\left(U_i\left(\hat{\boldsymbol{\beta}}^M, M(\hat{\psi})|\boldsymbol{b}_0\right) - U_i\left(\boldsymbol{\beta}_0, M(\hat{\psi})|\boldsymbol{b}_0\right)\right)^T\right].$$

IV. *If* $\hat{\boldsymbol{b}}$ *is a consistent estimator of* $\boldsymbol{b}_0$, *then* $\boldsymbol{\Omega}_1 \to \boldsymbol{\Omega}_0$, *as* $N \to \infty$.

The above theorem shows that given the efficiency estimated random effects $\hat{\boldsymbol{b}}$, the estimator of fixed effects $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}^M$ are consistent and asymptotic normality. The proof is similar to the Lemma 2.2 and Theorem 2.3 in Xing et al. (2018). So the proofs were omitted here.

Theorem 3. *Under the regularity conditions provided in the Appendix, the estimator* $\hat{\boldsymbol{b}}_i$ *satisfies* $\| \hat{\boldsymbol{b}}_i - \boldsymbol{b}_{oi} \| = O_p(n^{-1/2})$, *where* $\| \cdot \|$ *is the Euclidean norm.*

The advantage of our new methodology for improving the efficiency estimation is that the correlation information for random effects estimation was incorporated. In contrast to the approaches assuming independent working structure for estimating the random effects Xing et al. (2018), and Wang, Tsai, and Qu (2012), the efficiency of random effect estimation is improved greatly. The performance can be seen in Table 2 of simulation studies.

The $\sqrt{n}$-consistency of random-effects estimation is challenging due to the additional serial correlation from repeated measurements conditional on the random effects. Then an L_2-mixingale method (Ortega and Rheinboldt 1973) conditional on the serial correlation for the repeated measurements was imposed. The proof is provided in the Appendix.

3. Simulation studies

In this section, we conducted simulation studies to evaluate the performance of the proposed method in improving efficiency estimation based on conditional mix-GEE models. The conditional correlated responses were generated using the conditional mean and covariance

$$\boldsymbol{\mu}_i^b = \boldsymbol{\beta}_0 + \mathbf{X}_i\boldsymbol{\beta}_1 + b_i, \quad corr(y_i|\mathbf{X}_i, b_i) = R.(i = 1, \ldots, N)$$

where $\boldsymbol{\beta}_0 = 1$, $\boldsymbol{\beta}_1 = -1$, the covariate $\mathbf{X}_i$ are generated from uniform (0.5, 1.5). The random effects b_i is generated from Beta(0.5,0.5) distribution. The true correlation structures are CS and AR(1) with the correlation coefficient $\alpha = 0.7$, 0.3 or a three component mixture of AR(1), CS and MA(1) with the true nuisance parameters $(0.7, 0.7, 0.4, 0.3, 0.3, 0.4)$, $(0.4, 0.4, 0.3, 0.3, 0.3, 0.4)$, $(0.7, 0.7, 0.4, 0.5, 0.5, 0)$, $(0.7, 0.7, 0.4, 0.2, 0, 0.8)$, $(0.4, 0.4, 0.3, 0.2, 0, 0.8)$. The sample size is $N = 20$, 50 or 100. The cluster size is chosen to be $T = 10$. In each simulation study, 200 Monte Carlo samples will be generated. In order to compare the performance of different methods, the mean square error of $\hat{\boldsymbol{b}}$ is defined as

$$MSE(\hat{\boldsymbol{b}}) = \sum_{k=1}^{200}\sum_{i=1}^{N} \| \hat{\boldsymbol{b}}_i^{(k)} - \boldsymbol{b}_i \|^2 / 200N$$

where $\hat{\boldsymbol{b}}_i^{(k)}$ is the estimator of the true parameter $\boldsymbol{b}_i$ from the kth simulation, $\| \cdot \|$ denotes the Euclidean norm.

We compared the new methodology for improving the efficiency of fixed-effects and random-effects estimation based on conditional mix-GEE models (pcmix-GEE) to Xing et al. (2018) conditional mix-GEE models (cmix-GEE), to Cho, Wang, and Qu (2017) accuracy conditional quadratic inference functions (CCQIF) approach and to Wang, Tsai, and Qu (2012) conditional quadratic inference functions approach (CQIF).

3.1. Simulation 1: single component correlation structure from continuous response

In this section, we firstly consider the data from a single correlation structure CS and AR(1). Tables 1 and 2 provide the MSEs of the estimators for the fixed-effects β_0 and β_1 under different values of the nuisance parameter α with different sample sizes.

As seen from Tables 1 and 2, the MSEs of the new methodology pcGEE are smaller than those obtained from the other three CCQIF, cmix-GEE and CQIF approaches, even under the misspecified working correlation structure. At the same time, the MSEs of pcmix-GEE method are smaller as the sample sizes increases.

Table 3 shows the MSEs of random-effects $\hat{b}$ under different approaches. As seen from Table 3, the MSEs for the random effects $\hat{b}$ of pcmix-GEE approach are smaller than those obtained from cmix-GEE (Xing et al. 2018) and CQIF (Wang, Tsai, and Qu 2012) methods. In addition, the MSEs of pcmix-GEE approach could provide the same

Table 1. MSE for the estimator of the intercept $\beta_0 = 1$ with single component correlation structure.

$R(\alpha)$	N	α	pcmixGEE	$CCQIF_{cs}$	$CCQIF_{AR}$	cmixGEE	$CQIF_{cs}$	$CQIF_{AR}$
CS	20	0.7	0.054433	0.062325	0.071364	0.056334	0.064484	0.062349
CS	20	0.3	0.067921	0.071172	0.092051	0.070934	0.089193	0.079959
AR	20	0.7	0.051422	0.074994	0.056061	0.054287	0.091233	0.055656
AR	20	0.3	0.060313	0.072211	0.075600	0.063401	0.124061	0.078787
CS	50	0.7	0.020720	0.022101	0.025506	0.020711	0.021757	0.022711
CS	50	0.3	0.023436	0.025801	0.028807	0.023962	0.026950	0.025281
AR	50	0.7	0.016512	0.024329	0.017115	0.016995	0.026519	0.017250
AR	50	0.3	0.021064	0.022222	0.020904	0.021412	0.032811	0.021338
CS	100	0.7	0.011921	0.012303	0.013738	0.011917	0.012192	0.012640
CS	100	0.3	0.013005	0.013291	0.013916	0.012931	0.013305	0.012761
AR	100	0.7	0.007581	0.012534	0.007638	0.007765	0.013261	0.007678
AR	100	0.3	0.013205	0.014928	0.013827	0.013588	0.017615	0.014204

pcmix-GEE: our proposed new methodology for improving the efficiency of estimation based on conditional mix-GEE models; CCQIF: Cho, Wang, and Qu (2017) accuracy conditional quadratic inference functions; cmix-GEE: Xing et al. (2018) conditional mix-GEE models; CQIF: Wang, Tsai, and Qu (2012) conditional quadratic inference functions approach.

Table 2. MSE for the estimator of the intercept $\beta_1 = -1$ with single component correlation structure.

$R(\alpha)$	N	α	pcmixGEE	$CCQIF_{cs}$	$CCQIF_{AR}$	cmixGEE	$CQIF_{cs}$	$CQIF_{AR}$
CS	20	0.7	0.017208	0.018194	0.025816	0.019451	0.021183	0.021266
CS	20	0.3	0.047542	0.048422	0.063874	0.050370	0.064022	0.057495
AR	20	0.7	0.023977	0.045190	0.029007	0.024529	0.062207	0.027969
AR	20	0.3	0.053810	0.065141	0.065353	0.056426	0.117585	0.069948
CS	50	0.7	0.008365	0.008578	0.010071	0.008461	0.009025	0.009525
CS	50	0.3	0.019535	0.020228	0.023409	0.020428	0.022700	0.020890
AR	50	0.7	0.009244	0.015890	0.009840	0.009232	0.017722	0.010048
AR	50	0.3	0.019318	0.022216	0.019072	0.019830	0.032824	0.019929
CS	100	0.7	0.004418	0.004452	0.005789	0.004432	0.004487	0.004829
CS	100	0.3	0.013005	0.013291	0.013916	0.012931	0.013305	0.012761
AR	100	0.7	0.007581	0.012534	0.007638	0.007765	0.013261	0.007678
AR	100	0.3	0.008887	0.008959	0.009707	0.008812	0.009042	0.008731

efficiency for the random-effects $\hat{b}$ to the CCQIF method which could not identify the true correlation structure. However, according to the estimated proportions, the pcmix-GEE method can identify the true correlation structure.

Table 4 shows the nuisance parameters estimation of pcmix-GEE method. From Table 4, we can see that the estimated mixture proportions given by pcmix-GEE method correctly identified the true correlation structure as the sample size increase, and the correlation parameters α can even be correctly estimated.

3.2. Simulation 2: three component correlation structures from continuous response

In this section, we consider the case in which the true correlation is given by a three component mixture of AR(1), CS and MA(1) for the longitudinal data. The true values of the correlation parameters and the true mixture proportions are given in vectors in Table 5. We mainly compare the performance of the MSEs for the CCQIF and CQIF under the AR(1) and CS correlation structures and the cmix-GEE method. In general, to all sample sizes, the pcmix-GEE estimators for the slope have the lowest MSEs compared to the MSEs under the CCQIF and CQIF approaches. Meanwhile, as the sample size increases, the efficiency of the pcmix-GEE estimator also improves, as expected. In addition, we also observe that the component proportions are consistently estimated by the pcmix-GEE method. We also compare the performance of the MSEs for the

Table 3. MSE for the estimator of the random effect b with single component correlation structure.

$R(\alpha)$	N	α	*pcmixGEE*	$CCQIF_{cs}$	$CCQIF_{AR}$	*cmixGEE*	$CQIF_{CS}$	$CQIF_{AR}$
CS	20	0.7	0.011083	0.011084	0.011083	0.417760	0.417465	0.417455
CS	20	0.3	0.011162	0.011163	0.011162	0.215643	0.214635	0.216105
AR	20	0.7	0.011034	0.011035	0.011034	0.244917	0.240004	0.242258
AR	20	0.3	0.011140	0.011141	0.011027	0.106387	0.104815	0.106432
CS	50	0.7	0.011285	0.011285	0.011285	0.375658	0.375532	0.375407
CS	50	0.3	0.011206	0.011206	0.011206	0.192019	0.191809	0.192104
AR	50	0.7	0.011187	0.011187	0.011187	0.212031	0.210905	0.211404
AR	50	0.3	0.011244	0.011244	0.011244	0.089845	0.089267	0.089819
CS	100	0.7	0.011189	0.011189	0.011189	0.332674	0.332633	0.332561
CS	100	0.3	0.011281	0.011281	0.011281	0.168972	0.168921	0.168992
AR	100	0.7	0.011237	0.011237	0.011237	0.189025	0.188612	0.188816
AR	100	0.3	0.011237	0.011237	0.011237	0.080115	0.079946	0.080127

Table 4. Nuisance parameters estimation of PCmix-GEE method.

$R(\alpha)$	N	α	$\hat{\pi}_{cs}$	$\hat{\pi}_{ar1}$	$\hat{\pi}_{ma1}$	$\hat{\alpha}_{cs}$	$\hat{\alpha}_{ar1}$	$\hat{\alpha}_{ma1}$
CS	20	0.7	0.97	0.02	0.01	0.70	0.73	0.25
CS	20	0.3	0.86	0.07	0.07	0.33	0.20	0.05
AR	20	0.7	0.05	0.91	0.04	0.63	0.71	0.44
AR	20	0.3	0.16	0.45	0.39	0.14	0.35	0.29
CS	50	0.7	0.98	0.01	0.01	0.70	0.74	0.26
CS	50	0.3	0.91	0.04	0.05	0.33	0.21	0.02
AR	50	0.7	0.03	0.95	0.02	0.64	0.70	0.45
AR	50	0.3	0.15	0.50	0.35	0.15	0.34	0.27
CS	100	0.7	0.99	0.01	0.00	0.70	0.76	0.22
CS	100	0.3	0.92	0.04	0.04	0.32	0.21	0.01
AR	100	0.7	0.02	0.96	0.02	0.65	0.70	0.44
AR	100	0.3	0.11	0.53	0.36	0.14	0.35	0.27

Table 5. MSE for the estimator of the slope $\beta_1 = -1$ with three components mixture correlation structures.

N	*pcmixGEE*	$CCQIF_{cs}$	$CCQIF_{AR}$	*cmixGEE*	$CQIF_{cs}$	$CQIF_{AR}$	π_{cs}	π_{ar1}	π_{ma1}
			$(\alpha_{cs},\alpha_{ar1},\alpha_{ma1},\pi_{cs},\pi_{ar1},\pi_{ma1})$=(0.7,0.7,0.4,0.3,0.3,0.4)						
20	0.029570	0.041589	0.033207	0.029471	0.060050	0.036324	0.19	0.60	0.21
50	0.011603	0.015958	0.012910	0.011458	0.017486	0.013011	0.19	0.63	0.18
100	0.006440	0.009362	0.007050	0.006749	0.009814	0.007025	0.19	0.64	0.17
			$(\alpha_{cs},\alpha_{ar1},\alpha_{ma1},\pi_{cs},\pi_{ar1},\pi_{ma1})$=(0.4,0.4,0.3,0.3,0.3,0.4)						
20	0.036658	0.046134	0.043606	0.037314	0.081722	0.045130	0.29	0.40	0.31
50	0.017095	0.020561	0.019660	0.017014	0.026297	0.019316	0.30	0.41	0.29
100	0.009124	0.010153	0.009684	0.009270	0.011304	0.009538	0.31	0.42	0.27
			$(\alpha_{cs},\alpha_{ar1},\alpha_{ma1},\pi_{cs},\pi_{ar1},\pi_{ma1})$=(0.7,0.7,0.4,0.5,0.5,0)						
20	0.025063	0.032493	0.028467	0.025356	0.042281	0.027925	0.36	0.61	0.03
50	0.008195	0.011353	0.009207	0.008370	0.012176	0.008838	0.33	0.65	0.02
100	0.004326	0.005437	0.004861	0.004533	0.005279	0.004434	0.34	0.65	0.01
			$(\alpha_{cs},\alpha_{ar1},\alpha_{ma1},\pi_{cs},\pi_{ar1},\pi_{ma1})$=(0.7,0.7,0.4,0.2,0,0.8)						
20	0.038685	0.056486	0.047476	0.039039	0.084081	0.049370	0.12	0.30	0.58
50	0.017014	0.021308	0.017355	0.016795	0.028163	0.017832	0.10	0.31	0.59
100	0.00783	0.010748	0.008181	0.007863	0.011988	0.008423	0.11	0.29	0.60
			$(\alpha_{cs},\alpha_{ar1},\alpha_{ma1},\pi_{cs},\pi_{ar1},\pi_{ma1})$=(0.4,0.4,0.3,0.2,0,0.8)						
20	0.049664	0.059033	0.056486	0.049997	0.115850	0.062242	0.22	0.32	0.46
50	0.019819	0.022883	0.020424	0.019554	0.029064	0.020365	0.21	0.33	0.47
100	0.010902	0.013617	0.011046	0.011291	0.015092	0.011492	0.21	0.34	0.45

random-effects $\hat{b}$ among different methods. The results are similar to the single component correlation structure as expected. So we do not list here.

4. Conclusions

In this paper, a new methodology based on the cmix-GEE method was proposed to improve the efficiency of the fixed-effects and random-effects estimation. A major advantage of the proposed strategy is that we can significantly improve the accuracy of random effects estimation since our proposed approach utilizes the serial correlation information for repeated measurements in estimating random effects. At the same time, the efficiency of fixed effects estimation can be significantly improved because of the accuracy of the random effects estimation. In addition, normality assumption for the random effects in our proposed approach is not required. According to the estimated mixture proportions, the true correlation structure can be correctly identified in our proposed approach. Our approach can not only be applied for correlated data but also for the individual random effect for longitudinal data, which is also our interest. Our simulation studies show that the MSEs of the pcmix-GEE method are smaller than those obtained from CCQIF, CQIF and cmix-GEE approaches, even for the misspecified working correlation structure. Moreover, as the sample size increases, the efficiency of the pcmix-GEE estimator is also improved, which coincided with our expectation. In addition, the estimates of the component proportions are consistent as the sample size tends to infinity.

For the case where the sample size is finite small, especially, when the sample size is smaller than 10, there maybe some difficult in identifying the parameters with our proposed method. Westgate (2012, 2013) has shown that the use of an unstructured working correlation with GEE or the empirical approach with QIF can impact the validity of small-sample inference. In our proposed approach, the correlation matrix $C = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{\mu}_i^b)(\boldsymbol{y}_i - \boldsymbol{\mu}_i^b)^T$ is an unstructured working correlation matrix, this maybe the reason that affects the inference of small samples. So, our future work is to correct the covariance estimators to improve the inference of our proposed method when the sample size is finite small.

Funding

This work was supported by the National Natural Science Foundation of China under Grant 11690012; National Natural Science Foundation of China under Grant 11631003; Education Department Science and Technology Project of Jilin Province under Grant JJKH20190728KJ; and the Key Program of Jilin University of Finance and Economics under Grant 2018Z01.

Appendix: Proof of Theorem 3

We need the following regularity conditions and assumptions:

C1. The parameter space is compact.

C2. There exists $H(\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{b}) = O_p(1)$ such that $|\partial\hat{\psi}/\partial\boldsymbol{\beta}| \leq H(\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{b}) = O_p(1)$ and $|\partial\hat{\psi}/\partial\boldsymbol{b}| \leq H(\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{b}) = O_p(1)$.

C3. $\left[\frac{\partial U_i^2(\boldsymbol{\beta}, M(\hat{\psi}(\boldsymbol{\beta}))|\boldsymbol{b})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right]$ is continuous at $\boldsymbol{\beta}_0$ with probability one, and $E\left[sup_{\boldsymbol{\beta}\in\mathcal{N}}\frac{\partial U_i^2(\boldsymbol{\beta}, M(\hat{\psi}(\boldsymbol{\beta}))|\boldsymbol{b})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right] < \infty$, where $\mathcal{N}$ denotes the neighborhood of $\boldsymbol{\beta}$.

C4. $\left[\frac{\partial U_i^2(\boldsymbol{\beta}, M(\psi(\boldsymbol{\beta}))|\boldsymbol{b})}{\partial\psi\partial\psi^T}\right]$ is continuous at ψ^* with probability one, and $E\left[sup_{\psi\in\mathcal{N}}\frac{\partial U_i^2(\boldsymbol{\beta}, M(\psi(\boldsymbol{\beta}))|\boldsymbol{b})}{\partial\psi\partial\psi^T}\right] < \infty$, where $\mathcal{N}$ denotes the neighborhood of ψ.

C5. Conditional on the random effects $\boldsymbol{b}_0$, the parameter $\boldsymbol{\beta}$ is identifiable, i.e,

$$E\left\{U_i\left(\boldsymbol{\beta}_0, M\left(\psi^*(\boldsymbol{\beta}_0)|\boldsymbol{b}_0\right)\right)\right\} = 0.$$

C6. $E\{E\{U_i(\boldsymbol{\beta}_0, M(\psi^*(\boldsymbol{\beta}_0)|\hat{\boldsymbol{b}}))\}\} \xrightarrow{P} 0$.
C7. There is a neighborhood $\mathcal{N}$ such that $E\left[\frac{\partial U_i(\boldsymbol{\beta}, M(\psi)|\boldsymbol{b})}{\partial \boldsymbol{\beta}}\right]$ is bounded with probability one.
C8. Assuming that $U_i(\boldsymbol{\beta}, M(\psi(\boldsymbol{\beta}))|\boldsymbol{b})$ is continuous and differentiable with respect to variable $\boldsymbol{\beta}$ and $\boldsymbol{b}$, and $\dot{U}_{i,b}(\boldsymbol{\beta}, M(\psi(\boldsymbol{\beta}))|\boldsymbol{b})$ is bounded in probability one.
C9. If $e_{ij} = y_{ij} - \boldsymbol{\mu}_{ij}(\boldsymbol{\beta}|\boldsymbol{b}_i)$ is the residual for the jth observation of subject i, the residuals within the same subjects $(e_{i1}, \ldots, e_{iT})$ satisfy $\|E(e_{ij}|e_{i,j-m})\|_2 \le c_j\varphi_m$, for $j = 1, \ldots, T$ and $m = 1, \ldots, j-1$, $\|e_{ij} - E(e_{ij}|e_{i,j+m})\| \le c_j\varphi_{m+1}$, $j = 1, \ldots, T, m = 1, \ldots, n-j$, where $\|\cdot\|_2$ is the L_2 norm, φ_m are some non-negative constants such that $\varphi_m \to 0$ as $m \to \infty$, and the $c_j, j \ge 1$ satisfy $\overline{lim}_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} c_j < \infty$, or $\{c_j\}$ can be given by $\{\|e_{ij}\|_2\}$.

Proof. For convenience, let $Q_i = A_i^{-1/2}C^{-1}A_i^{-1/2}$, $\|\cdot\|^1$ denotes the sum of all matrix entries absolute values, n is the cluster size. Since each element of Q_i is bounded in probability, the order of the $\|Q_i\|^1$ is between n and n^2. Then, $g_i^C = (\frac{\partial \boldsymbol{\mu}_i^b}{\partial \boldsymbol{b}_i})^T \boldsymbol{A}_i^{\frac{1}{2}} C^{-1} \boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{y}_i - \boldsymbol{\mu}_i^b) = (\frac{\partial \boldsymbol{\mu}_i^b}{\partial \boldsymbol{b}_i})^T Q_i(\boldsymbol{y}_i - \boldsymbol{\mu}_i^b) = \sum_{k=1}^{n}\sum_{j=1}^{n} c_{ijk}\dot{\boldsymbol{\mu}}_{ik,b_i}e_{ij}$, where $\left(\frac{\partial \boldsymbol{\mu}_i^b}{\partial \boldsymbol{b}_i}\right)^T = (\dot{\boldsymbol{\mu}}_{i1,b_i}, \ldots, \dot{\boldsymbol{\mu}}_{in,b_i})^T$, c_{ijk} is the $k \times j$th component of Q_i, and $e_{ij} = y_{ij} - \boldsymbol{\mu}_{ij}^b(\boldsymbol{\beta}|b_i)$. The estimator $\hat{\boldsymbol{b}}_i$ is obtained by solving $g_i^C(\hat{\boldsymbol{\beta}}^M|b_i) = 0$. By the Taylor expansion of $g_i^C(\hat{\boldsymbol{\beta}}^M|b_i)$ at b_0, we have

$$0 = g_i^C\left(\hat{\boldsymbol{\beta}}^M|\boldsymbol{b}_i\right) = g_i^C\left(\hat{\boldsymbol{\beta}}^M|\boldsymbol{b}_0\right) + \dot{g}_{i,\boldsymbol{b}_i}^C\left(\hat{\boldsymbol{\beta}}^M|\tilde{\boldsymbol{b}}_i\right)\left(\hat{\boldsymbol{b}}_i - \boldsymbol{b}_{0i}\right),$$

where $\tilde{b}_i$ is between $\hat{\boldsymbol{b}}_i$ and $\boldsymbol{b}_{0i}$, $\dot{g}_{i,\boldsymbol{b}_i}^C(\hat{\boldsymbol{\beta}}^M|\tilde{\boldsymbol{b}}_i) = \frac{\partial}{\partial \boldsymbol{b}_i} g_i^C(\hat{\boldsymbol{\beta}}^M|\boldsymbol{b}_i)|_{\boldsymbol{b}_i = \tilde{\boldsymbol{b}}_i}$, Then we have

$$\boldsymbol{b}_{0i} - \hat{\boldsymbol{b}}_i = \left(\dot{g}_{i,\boldsymbol{b}_i}^C\left(\hat{\boldsymbol{\beta}}^M|\tilde{\boldsymbol{b}}_i\right)\right)^{-1} g_i^C\left(\hat{\boldsymbol{\beta}}^M|\boldsymbol{b}_0\right).$$

By Theorem 2, $\hat{\boldsymbol{\beta}}^M \xrightarrow{p} \boldsymbol{\beta}_0$, we have

$$\boldsymbol{b}_{0i} - \hat{\boldsymbol{b}}_i \to \left(\dot{g}_{i,\boldsymbol{b}_i}^C\left(\boldsymbol{\beta}_0|\tilde{\boldsymbol{b}}_i\right)\right)^{-1} g_i^C(\boldsymbol{\beta}_0|\boldsymbol{b}_0).$$

Since $\dot{\boldsymbol{\mu}}_{ij,\boldsymbol{b}_i}(\boldsymbol{\beta}_0|\tilde{\boldsymbol{b}}_i)$ is bounded in probability, we have $(\dot{g}_{i,\boldsymbol{b}_i}^C(\boldsymbol{\beta}_0|\tilde{\boldsymbol{b}}_i))^{-1} = O_p(r_n^{-1})$. Then there exists a constant K_1 such that

$$g_i^C(\boldsymbol{\beta}_0|\boldsymbol{b}_0) = \left|\sum_{k=1}^{n}\sum_{j=1}^{n} c_{ijk}\dot{\boldsymbol{\mu}}_{ik,b_i}e_{ij}\right| \le K_1 r_n \left|\frac{n}{r_n}\sum_{k=1}^{n} c_{ijk}\frac{1}{n}\sum_{j=1}^{n} e_{ij}\right| = O_p(r_n\bar{e}_i),$$

where $\frac{n}{r_n}\sum_{k=1}^{n} c_{ijk} = O_p(1)$, and $\frac{1}{n}\sum_{j=1}^{n} e_{ij} = \bar{e}_i$. Since $(\dot{g}_{i,\boldsymbol{b}_i}^C(\boldsymbol{\beta}_0|\tilde{\boldsymbol{b}}_i))^{-1} = O_p(r_n^{-1})$ and $g_i^C(\boldsymbol{\beta}_0|\boldsymbol{b}_0) = O_p(r_n\bar{e}_i)$, then $\boldsymbol{b}_{0i} - \hat{\boldsymbol{b}}_i = O_p(\bar{e}_i)$. Hence, we need to prove that $\bar{e}_i = O_p(n^{-1/2})$, that is $E(|\bar{e}_i|^2) = O_p(n^{-1})$.

$$\begin{aligned} E\left(|\bar{e}_i|^2\right) &\le \frac{2}{n^2}\sum_{t=1}^{n}\sum_{j=1}^{t} E|e_{ij}e_{it}| \le \frac{2}{n^2}\sum_{t=1}^{n}\sum_{j=1}^{t} E\left(|e_{ij}||E\left(e_{it}|e_{ij}\right)|\right) \\ &\le \frac{2}{n^2}\sum_{t=1}^{n}\sum_{j=1}^{t} c_jc_t\varphi_{t-j} = \frac{2}{n^2}\sum_{k=1}^{n}\varphi_k\sum_{j=1}^{n-k} c_jc_{j+k}, \end{aligned}$$

Under the condition that the sequence of random variables e_{ij} satisfies the L_2 mixingale condition and $\sum_{k=1}^{\infty}\varphi_k < \infty$, this implies $E(|\bar{e}_i|^2) = O_p(n^{-1})$. So this proof is complete! □

References

Breslow, N. E., and D. G. Clayton. 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88:9–25. doi:10.2307/2290687.

Cho, H., P. Wang, and A. Qu. 2017. Personalize treatment for longitudinal data using unspecified random-effects model. *Statistica Sinica* 27:187–205. doi:10.5705/ss.202015.0120.

Diaz, F. J., T. E. Rivera, R. C. Josiassen, and J. Leon. 2007. Individualizing drug dosage by using a random intercept linear model. *Statistics in Medicine* 26 (9):2052–73. doi:10.1002/sim.2636.

Diaz, F. J., H. W. Yeh, and J. Leon. 2012. Role of statistics random-effects linear models in personalized medicine. *Current Pharmacogenomics and Personalized Medicine* 10:22–32. doi:10.2174/1875692111201010022.

Fang, S., H. Zhang, and L. Sun. 2016. Joint analysis of longitudinal data with additive mixed effect model for informative observation times. *Journal of Statistical Planning and Inference* 169:43–55. doi:10.1016/j.jspi.2015.08.001.

Jiang, B. J. 1999. Conditional inference about generalized linear mixed models. *The Annals of Statistics* 27 (6):1974–2007. doi:10.1214/009053605000000543.

Jiang, J., and W. Zhang. 2001. Robust estimation in generalized linear mixed models. *Biometrika* 88 (3):753–65. doi:10.1093/biomet/88.3.753.

Laird, N. M., and J. H. Ware. 1982. Random-effects models for longitudinal data. *Biometrics* 38 (4):963–74.

McCulloch, C. E. 1997. Maximum likelihood algorithm for generalized linear mixed models. *Journal of the American Statistical Association* 92 (437):162–70. doi:10.2307/2291460.

McCulloch, C. E., S. R. Searle, and J. M. Neuhaus. 2008. *Generalized, linear, and mixed models*. 2nd ed. New York, NY: John Wiley.

Ortega, J. M., and W. C. Rheinboldt. 1973. *Iterative solutions of nonlinear equations in several variables*. Cambridge, MA: Academic Press.

Vonesh, E. F., H. Wang, L. Nie, and D. Majumdar. 2002. Conditional second-order generalized estimating equations for generalized linear and nonlinear mixed-effects models. *Journal of the American Statistical Association* 97 (457):271–83. doi:10.1198/016214502753479400.

Wang, N. 2003. Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika* 90 (1):43–52. doi:10.1093/biomet/90.1.43.

Wang, Y. G., and V. Carey. 2003. Working correlation structure misspecification, estimation and covariate design: Implications for generalized estimating equations performance. *Biometrika* 90 (1):29–41. doi:10.1093/biomet/90.1.29.

Wang, P., G. F. Tsai, and A. Qu. 2012. Conditional inference functions for mixed-effects models with unspecified random-effects distribution. *Journal of the American Statistical Association* 107 (498):725–36. doi:10.1080/01621459.2012.665199.

Wedderburn, R. W. 1974. Quasi-likelihood functions, generalized linear models and the gauss-newton method. *Biometrika* 61:439–47. doi:10.2307/2334725.

Westgate, PM. 2012. A bias-corrected covariance estimate for improved inference with quadratic inference functions. *Statistics in Medicine* 31 (29):4003–22. doi:10.1002/sim.5479.

Westgate, PM. 2013. A bias correction for covariance estimators to improve inference with generalized estimating equations that use an unstructured correlation matrix. *Statistics in Medicine* 32 (16):2850–8. doi:10.1002/sim.5709.

Xing, Y. C., L. L. Xu, W. Q. Ma, and Z. C. Zhu. 2018. Conditional mix-GEE models for longitudinal data with unspecified random-effects distributions. *Communications in Statistics - Theory and Methods* 47 (4):862–76. doi:10.1080/03610926.2016.1267763.

Xu, L. L., N. Lin, B. X. Zhang, and N. Z. Shi. 2012. A finite mixture model for working correlation matrices in generalized estimating equations. *Statistica Sinica* 22 (2):755–76. doi:10.5705/ss.2010.090.