Estimation and selection in linear mixed models with missing data under compound symmetric structure

Yi-Ching Lee and Junfeng Shang

Department of Mathematics, Southeast Missouri State University, Cape Girardeau, MO, USA; Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH, USA

ABSTRACT

It is quite appealing to extend existing theories in classical linear models to correlated responses where linear mixed-effects models are utilized and the dependency in the data is modeled by random effects. In the mixed modeling framework, missing values occur naturally due to dropouts or non-responses, which is frequently encountered when dealing with real data. Motivated by such problems, we aim to investigate the estimation and model selection performance in linear mixed models when missing data are present. Inspired by the property of the indicator function for missingness and its relation to missing rates, we propose an approach that records missingness in an indicator-based matrix and derive the likelihood-based estimators for all parameters involved in the linear mixed-effects models. Based on the proposed method for estimation, we explore the relationship between estimation and selection behavior over missing rates. Simulations and a real data application are conducted for illustrating the effectiveness of the proposed method in selecting the most appropriate model and in estimating parameters.

ARTICLE HISTORY

Received 5 June 2020
Accepted 13 August 2021

KEYWORDS

Model selection; estimation accuracy and precision; linear mixed models; compound symmetric structure; missing rate; coverage probability

1. Introduction

Fisher [4] extended classical linear models through adding random effects to model the correlated structure in data, and the resulting model is referred to as the linear mixed model (LMM).

The long-standing problem in introducing random effects to classical linear models is in estimating parameters in the covariance structure for depicting the variability by additional random effects. Analogous to classical linear models, parameter estimation in linear mixed models are also categorized to maximum likelihood-based (ML) and restricted maximum likelihood (REML) approaches. The former is obtained by maximizing the likelihood of the full data; whereas, the latter is developed by maximizing the likelihood of residuals [9].

Both ML and REML estimators are especially popular for their purposes. Miller [7,8] developed asymptotic properties of ML estimators for variance components where
consistency and asymptotic normality were shown, and Harville [5] discussed the attractive features of the maximum likelihood approach for the estimation problem in variance components. It should be noted that the above methods rely on iterative procedures to obtain final estimates. One of the most commonly used procedures is the Expectation–Maximization (EM) algorithm [3], and in fact, the \texttt{lme} function in the \texttt{lme4} library in R employs the approach by Laird and Ware [6] that incorporates the EM algorithm for estimation.

To develop an estimation method under the circumstances where missing values are present, we propose and derive the estimators that incorporate missing rates with the indicator matrix and matrix theory in the linear mixed models with compound symmetric covariance structure. We then inspect the performance of the proposed approach in selection and estimation over different missing rates. We further investigate the performance of the proposed approach with missing data from various mechanisms, such as missing completely at random (MCAR) and missing not at random (MNAR) [10], and we compare the results with the maximum likelihood estimator (MLE).

In addition to the proposed method in estimation involving missing rates, we conduct model selection in the linear mixed models, which can be fulfilled by testing the equality of means. Because of its consistency and discreteness, we utilize the commonly used selection criterion, Schwarz Information Criterion (SIC), in Schwarz [11]. With the SIC, the posterior probability for each model is computed, and the model with the maximized posterior probability out of all candidate models is selected. The performance of model selection is therefore investigated given the estimators are produced from the proposed method in estimation.

The rest of the paper is organized as follows. In Section 2, we discuss the LMMs with a compound symmetric covariance structure, propose and derive the likelihood estimator with incorporation of missing indicator matrix and study the convergence speed of the proposed estimators. In Section 3, the simulations on selection and estimation performance with various sample sizes and missing rates are conducted for demonstrating the effectiveness of the proposed method. In Section 4, a real data application is also conducted to illustrate the effectiveness of the proposed method in estimation and model selection with missing values. Section 5 concludes and provides discussion.

## 2. The proposed method of estimation in linear mixed models with missing data

### 2.1. The linear mixed model with missing data

We consider a class of linear mixed effects models with the presence of missing responses and code every missing response with 0. The model can be written as

\[ y_i = \Delta_i(X_i\beta + Z_ib_i + \epsilon_i), \]

where \( y_i \) is a \( t \)-vector of observed responses for subject \( i \), \( \Delta_i = \text{diag}\{\delta_{i1}, \ldots, \delta_{it}\} \), where \( \delta_{ij} \) denotes the missingness, with \( \delta_{ij} = 1 \) when \( y_{ij} \) is observed and 0 when \( y_{ij} \) is missing, \( X_i \) is a known \( t \times p \) covariates matrix, \( Z_i \) is a known \( t \times q \) matrix usually a subset of \( X_i \), and \( b_i \) and \( \epsilon_i \) are unobserved independent \( q \) and \( t \)-vectors of normally distributed random variables with mean zero and respective variance–covariance matrices \( D \) and \( R_i \). The
marginal distribution of \( y_i \) is \( N(\Delta_iX_i\beta, \Sigma_i) \), where \( \Sigma_i = \Delta_i(Z_iZ_i^T + R_i)\Delta_i \), which yields the log-likelihood function as

\[
\ell = -\frac{mt}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^{m} \log |\Sigma_i| - \frac{1}{2} \sum_{i=1}^{m} (y_i - \mu_i)^T \Sigma_i^{-1} (y_i - \mu_i),
\]

where \( \mu_i = \Delta_iX_i\beta \).

It should be noted that the variance–covariance component \( \Sigma_i \) is no longer positive definite, so it is not invertible. However, as shown in the Appendix, we can establish the approximations of the determinant and inverse matrix. With these approximations, we derive the likelihood and estimators of the parameters in the LMM.

### 2.2. Compound symmetric covariance structure

Let \( y_{ij} \) denote the response of subject \( i \) at time \( j \), where \( i = 1, \ldots, m \) and \( j = 1, \ldots, t \), and consider a compound symmetric covariance structure for which the correlation is equal for all time gaps between any pair of observations. We consider the presence of missing responses and code every missing response with 0. The model is written as

\[
y_{ij} = \delta_{ij}(X_i\beta + \alpha_i + \epsilon_{ij}),
\]

where \( X_i \) are \( p \)-dimensional covariates, \( \beta \) denotes fixed effects in \( \mathbb{R}^p \) where \( \beta_1, \ldots, \beta_t \) are fixed effects, respectively, at time \( 1, \ldots, t \), and the corresponding design matrix \( X_{it} \) is a \( t \)-dimensional identity matrix, \( \delta_{ij} \) denote the missingness with \( \delta_{ij} = 1 \) when \( y_{ij} \) is observed and 0 otherwise, \( \alpha_i \) is the random effect corresponding to subject \( i \), and \( \epsilon_{ij} \) denote the pure errors, for \( i = 1, \ldots, m, j = 1, \ldots, t \). With a compound symmetry structure, it is assumed that \( \alpha_i \sim N(0, \sigma^2_\alpha) \), \( \epsilon_{ij} \sim N(0, \sigma^2) \), and \( \alpha_i \) and \( \epsilon_{ij} \) are independent for all \( i = 1, \ldots, m, j = 1, \ldots, t \).

Define \( y_i^T = (y_{i1}, \ldots, y_{it}) \), \( \Delta_i = \text{diag}(\delta_{i1}, \ldots, \delta_{it}) \), and \( \epsilon_i^T = (\epsilon_{i1}, \ldots, \epsilon_{it}) \). Model (1) can be vectorized as

\[
y_i = \Delta_iX_i\beta + \alpha_i1_t + \Delta_i\epsilon_i,
\]

where \( y_i \sim N_i(\mu_i, \Sigma_i) \), \( \mu_i = \Delta_iX_i\beta \), and \( \Sigma_i = \Delta_i(\sigma^2_\alpha1_t1_t^T + \sigma^21_t)\Delta_i \), \( i = 1, \ldots, m \). We will derive the maximum likelihood estimators for all parameters in the following section.

### 2.3. The maximum likelihood estimators with missing data

Denoting the approximation of \( \Sigma_i \) as \( \Sigma_i^* \) and by Remarks 3 and 5 in the Appendix, we have

\[
(S_i^*)^{-1} = \frac{1}{\sigma^2} \left( \Delta_i - \frac{\sigma^2_\alpha}{\sigma^2 + \sigma^2_\alpha} \Delta_i1_t1_t^T \Delta_i \right)
\]

and

\[
det(\Sigma_i^*) = \left(1 + \frac{\sigma^2_\alpha}{\sigma^2} \rho_i \right)^{(\sigma^2)^{\rho_i}}.
\]

From the marginal distribution for the model with one random effect and the approximations of the variance–covariance component, the log-likelihood of the model takes the
form of
\[
\ell = -\frac{mt}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^{m} \log \left( 1 + \frac{\sigma^2_t}{\sigma^2} p_i \right) - \frac{1}{2} \sum_{i=1}^{m} p_i \log \sigma^2
\]
\[
- \frac{1}{2} \sum_{i=1}^{m} (y_i - \mu_i)^T (\Sigma^*_i)^{-1} (y_i - \mu_i).
\]

Differentiating the log-likelihood with respect to \( \beta \) and setting the result equal to zero yields the maximum likelihood estimator:
\[
\hat{\beta} = \left( \sum_{i=1}^{m} X_i^T (\Sigma^*_i)^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{m} X_i^T (\Sigma^*_i)^{-1} y_i \right).
\]

To find the maximum likelihood estimators of \( \sigma^2 \) and \( \sigma^2_t \), we first obtain the partial derivatives of \((\Sigma^*_i)^{-1}\) with respect to \( \sigma^2 \) and \( \sigma^2_t \), as shown in the following equations:
\[
\frac{\partial (\Sigma^*_i)^{-1}}{\partial \sigma^2} = -\Delta_i \frac{1}{(\sigma^2)^2} \Delta_i 1_1^T \Delta_i \left( -\frac{2\sigma^2_t}{\sigma^2} - p_i \right) \frac{1}{(\sigma^2)^2} \left( \frac{\sigma^2}{\sigma^2_t} + p_i \right)^{-2},
\]
\[
\frac{\partial (\Sigma^*_i)^{-1}}{\partial \sigma^2_t} = -\Delta_i 1_1^T \Delta_i \left( \frac{1}{\sigma^2 + \sigma^2_t p_i} \right)^2.
\]

For the brevity of notation, we define
\[
C_{1i} = (y_i - \mu_i)^T \Delta_i (y_i - \mu_i) \quad \text{and} \quad C_{2i} = (y_i - \mu_i)^T \Delta_i 1_1^T \Delta_i (y_i - \mu_i).
\]

The maximum likelihood estimator of \( \sigma^2_t \) is derived as shown in the following equation:
\[
\hat{\sigma}^2_t = \left( \sum_{i=1}^{m} \frac{p_i^2}{(\sigma^2 + \sigma^2_t p_i)^2} \right)^{-1} \left( \sum_{i=1}^{m} \frac{C_{2i}}{(\sigma^2 + \sigma^2_t p_i)^2} - \frac{\sigma^2 p_i}{(\sigma^2 + \sigma^2_t p_i)^2} \right).
\]

We can recall that \( p_i = 1_1^T \Delta_i 1_i \), and let \( w_{ti} = \sigma^2 + \sigma^2_t p_i \), so the likelihood estimator of \( \sigma^2_t \) takes a form of
\[
\hat{\sigma}^2_t = \left( \sum_{i=1}^{m} \left( \frac{p_i}{w_{ti}} \right)^2 \right)^{-1} \left( \sum_{i=1}^{m} \frac{C_{2i}}{(w_{ti})^2} - \sum_{i=1}^{m} \frac{\sigma^2 p_i}{(w_{ti})^2} \right).
\]

Similarly, the maximum likelihood estimator of \( \sigma^2 \) is derived and takes the following form:
\[
\hat{\sigma}^2 = \left( \sum_{i=1}^{m} p_i \right)^{-1} \left( \sum_{i=1}^{m} C_{1i} - \sum_{i=1}^{m} \frac{\sigma^2_t C_{2i}}{w_{ti}} \right).
\]

The detailed likelihood estimators derivations are presented in Section A.2 of the Appendix. Since they are not of closed forms, an iterative procedure is required to obtain the final solutions. We utilize the fixed-point iterative approach in the following procedure:
(1) Choose initial values for \( \beta, \sigma^2, \sigma^*_T \).

(2) At iteration \( k \), we update the parameter estimates with these three equations:

\[
\beta^{(k+1)} = \left( \sum_{i=1}^{m} X_i^T \left( (\Sigma_i^*)^{(k)} \right)^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{m} X_i^T \left( (\Sigma_i^*)^{(k)} \right)^{-1} y_i \right),
\]

\[
(\sigma^2)^{(k+1)} = \left( \sum_{i=1}^{m} \left( \frac{p_i}{w_i^{(k)}} \right)^2 \right)^{-1} \left( \sum_{i=1}^{m} \frac{C_{2i}}{(w_i^{(k)})^2} - \sum_{i=1}^{m} \frac{(\sigma^2)^{(k)} p_i}{(w_i^{(k)})^2} \right),
\]

and

\[
(\sigma^2_T)^{(k+1)} = \left( \sum_{i=1}^{m} \right)^{-1} \left( \sum_{i=1}^{m} C_{1i} - \sum_{i=1}^{m} \frac{(\sigma^2)^{(k)} C_{2i}}{w_i^{(k)}} \right).
\]

(3) Repeat step (2) until convergence.

The convergence properties are evaluated by the simulations in Section 2.5.

2.4. Theoretical properties

By model (2), we have \( y_i \sim N_i(\mu_i, \Sigma_i) \), with \( \mu_i = \Delta_i X_i \beta \) and \( \Sigma_i = \Delta_i (\sigma^2 T I_T + \sigma^2_0 I_T) \Delta_i \), for \( i = 1, \ldots, m \). To consistently denote the variance–covariance components with the approximations of the determinant and the inverse of the variance–covariance components in use to obtain the proposed estimators, we rewrite the model as \( y_i \sim N_i(\mu_i, \Sigma_i^*) \), where \( \Sigma_i^* = \Delta_i (\sigma^2 T I_T + \sigma^2_0 I_T) \Delta_i \), with the corresponding determinant and inverse as in Remarks 3 and 5 in the Appendix. We show that \( \hat{\beta}, \hat{\sigma}_T^2 \), and \( \hat{\sigma}^2 \) in Equations (3), (4), and (5) are unbiased estimators with equations (A4), (A5), and (A6) in Section A.3 of the Appendix.

2.5. Simulation studies of convergence in estimation

We evaluate the successive fixed-point iteration technique in calculating the proposed estimators by simulated data, and we construct the convergence plots for the estimates of \( \beta, \sigma^2 \) and \( \sigma^2_T \), with data under the missing rates of 10\%, 50\%, and 80\%.

The samples are generated from a compound symmetric mixed-effects model for 100 individuals (subjects) with three repeated measurements. The fixed effects associated with time are chosen to be \( \beta_i \), and the coefficients for each known covariate are set to be \( \beta^T \), where \( \beta^T = (21, 11, 7) \) and \( \beta^T = (-20, 13, 27, 0, 0) \) yielding the model where data truly consists of \( x_1, x_2, x_3 \) covariates with the known covariates \( x_1, x_2, x_3, x_4, \) and \( x_5 \) generated uniformly. The variance components are chosen to be \( \sigma^2 = 1 \) and \( \sigma^2_T = 4 \). We study the convergence behaviors under two scenarios: (1) with the model being correctly specified and (2) with the model being incorrectly specified, i.e. fitting data with redundant covariate terms. All initial values for fixed effects parameters are chosen to be zero and initial values for variances are set to be one.

Fitting the data with Scenario (1), the model including the covariates \( x_1, x_2, \) and \( x_3 \), the estimates of \( \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \sigma^2 \), and \( \sigma^2_T \) are computed by the iterative procedure
The iteration processes are recorded and displayed in Figure 1. The top row shows the convergence behavior of the fixed effects $\beta$s in the model, and all of them converge within five iterations for the missing rates of 10% and 50%. For the data that have 80% of missing values, the process takes about 10 iterations to converge.

Both variance estimates $\sigma^2$ and $\sigma^2_\tau$ converge within five iterations as well with the missing rates of 10% and 50%, as shown in middle and bottom rows of Figure 1. Similarly, both estimates take about 25 iterations to converge when 80% of data are not observed.

In Scenario (2), the analogous analyzes are carried out with the specified model, including a redundant variable $x_4$ and $x_5$, and the results are shown in Figure 2. All estimates converge within the first five iterations. The process takes about two more iterations to converge when the missing rate increases to 80%.

The aforementioned simulation results on convergence demonstrate that generally the iterative process of estimating the parameters using the proposed method only takes a few iterations to converge even with a high missing rate (80%).

### 3. Simulations on selection and estimation performance

#### 3.1. Objectives and generic settings

There are four objectives in this section for the simulations to examine (1) the selection performance under a series of data generated with various missing rates based on the proposed estimators, (2) the selection performance with missing data from different missing mechanisms based on the proposed estimators, (3) the accuracy, precision, and coverage...
Figure 2. Convergence of (top to bottom) $\beta$, $\sigma^2$, and $\sigma^2_\tau$ with (left to right) 10%, 50%, and 80% missing values for Scenario (2). The beta indices 1–6 denote the estimates for $\beta_{t1}$, $\beta_{t2}$, $\beta_{t3}$, $\beta_1$, $\beta_2$, and $\beta_3$.

probability of the estimators from different missing mechanisms, and (4) the comparison of the proposed approach with MLEs in estimation.

All data are generated from a compound symmetric normal mixed-effects model with three repeated measures in each subject. The number of subjects in the data varies and is specified in each of the simulations in this section. The mean parameter associated with time $j$ is denoted as $\beta_{tj}$, and the mean parameter associated with the covariate $i$ is $\beta_i$. The between-subject and within-subject variances are denoted, the same as before, by $\sigma^2_\tau$ and $\sigma^2$, respectively.

3.2. Selection performance with missing rates

It is well known that the SIC, in Schwarz [11], is consistent in model selection and suitable for discrete models. Some variants of the SIC or alternate methods have been developed for improving the selection performance. The SIC$_f$ in Shang [12] is known as a variant criterion of the SIC in linear mixed models. In the case of considering different prior model probabilities, the SIC$_f$ is favored compared to the SIC. Delattre et al. [2] derived an alternative to SIC by substituting the sample size in the penalty term with two components: one referring to the number of estimated fixed effects and the other associating with the number of estimated variance components in random effects. Similar to the method in Shang and Cavanaugh [13], the posterior probability is a measure for selecting the most appropriate model.

Our main goal is to examine the effectiveness of the proposed method in estimation in Section 2 with the presence of missing data. For succinctness, we utilize the most
Table 1. Variables included in 10 candidate models.

| Model 1 | Variables | Model 2 | Variables | Model 3 | Variables | Model 4 | Variables | Model 5 | Variables | Model 6 | Variables | Model 7 | Variables | Model 8 | Variables | Model 9 | Variables | Model 10 | Variables |
|---------|-----------|---------|-----------|---------|-----------|---------|-----------|---------|-----------|---------|-----------|---------|-----------|---------|-----------|---------|-----------|---------|-----------|
| x_1     |           | x_1, x_2 |           |         | x_1, x_2 | x_3, x_4 |           | x_1, x_2 | x_3, x_4 |          | x_2, x_3 | x_4, x_5 |           | x_1, x_2 | x_3, x_4 | x_5     | x_1, x_2 | x_3, x_5 | x_1, x_2 | x_3, x_5 | x_1, x_2 |

commonly used selection criterion, SIC, to conduct the model selection in the mixed models.

Since the fixed-point iterative method yields stable estimates, we move forward to illustrate the model selection performance with the implementation of the proposed method. We evaluate the performance of the proposed method with a series of data generated under a sequence of missing rates ranging from 0% to 80%, and we use the SIC as the assessment of the selection.

Data are generated from a compound symmetric normal mixed effects model with three repeated measures for each subject, and subjects of 30, of 50, and of 100 are considered. The mean parameters are chosen to be $\beta^T_{ij} = (21, 11, 7)$ and $\beta^T = (-20, 13, 27, 0, 0)$ yielding a true model consisting of $x_1, x_2$, and $x_3$ covariates. We carry out studies with (1) small variances $\sigma^2 = 1, \sigma^2_\tau = 4$ and (2) large variances $\sigma^2 = 16, \sigma^2_\tau = 100$. These parameters are randomly specified, and different $\beta$s and both small and large variance component parameters are considered. This is the approach to selecting the true parameters through all simulations in this section. Among all $2^5$ possible models, we consider the following 10 models for illustrative purposes (Table 1):

We compare the selection performance based on the SIC values in a sequence of missing proportions from 0% to 80%. In each iteration, the SIC will be computed using the proposed estimators and the model that yields the lowest SIC is the most appropriate. The accuracy rate is calculated as the ratio of the correct model yielding the lowest SIC to the total number of replicates. For each setting, the process is repeated 200 times, resulting in the model selection rate of the true model as the percentage of model 3 yielding the lowest SIC.

Figure 3 summarizes the correctly selected rates of the true model against the missing rates of 0%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, and 80% on the x-axis with sample sizes of 30, 50, and 100. As the sample size increases, the selection rates of the true model increase slightly regardless of the percent of missing values presented in the data. With a sample size of 100, it is observable that the selected rates of the true model are higher than 90% until the missing proportion rises to 40%. It is expected that the selected rates of the true model drop drastically when the majority of data are not observed. With a 60% missing rate in a sample of 100, only 40 observations are considered, and the SIC favors the model with more covariates such as model 4, model 5 or model 9.

Figure 4 demonstrates the selection performance with larger variances, $\sigma^2 = 16$ and $\sigma^2_\tau = 100$. The effectiveness of the proposed estimators on data with larger variability for the selected rates of the true model is ideal until the missing rate increases to 40%, just as in the previous scenario. From these two cases, we summarize that the proposed method can be utilized to return a reliable model selection result in choosing a set of variables with a moderate amount of missing data.
In addition to choosing variables, the SIC can also serve as an assessment for hypothesis type questions, such as testing $\beta_{t1} = \beta_{t2} = \beta_{t3}$ against $\beta_{t1} = \beta_{t2}, \beta_{t3}$. Here, we conduct a study to evaluate the performance in choosing these types of models by the SIC. We consider the same model specification for data with the mean parameters of $\beta_{t1}^T = (5, 5, 5)$ and $\beta^T = (-20, 13, 27)$ and the variances $\sigma^2 = 1$ and $\sigma^2_r = 4$. For illustrative purposes, we consider the three candidate models shown in Table 2. With the setup of the parameters, model 3 is expected to yield the lowest SIC.
Figure 5. Selection performance with three candidate models.

Figure 5 shows the selection result with a missing proportion less than 20% is optimal. A further investigation finds that the difference in the SIC between the three models is only within 0.01 to 1 when the missing proportion reaches 40%, though not shown here. One resolution for observing the results is to calculate the posterior probabilities for each candidate model by the SIC values, and this implementation is advised in real data analysis not in the simulations since it would be redundant to output all 200 sets of posterior probabilities.

3.3. Selection performance under MNAR

With an indicator function considered as a random variable, the core of the parameter estimation and model inferences rely on the assumption of the missing mechanism. In the derivation of the proposed estimators, we view the missing indicator as a part of the design matrix, so the assumption of the missing mechanism has no influence on the proposed estimators. Nevertheless, this paper is dedicated to evaluating the selection performance under as many scenarios as possible. In this subsection, we investigate the selection performance with the missing responses generated under missing not at random (MNAR).

The samples are generated in the same fashion as in the previous section with the mean parameters as \( \beta^T_{tj} = (5, 5, 5) \) and \( \beta^T = (-20, 13, 27) \) and two sets of variances (1) \( \sigma^2 = 1 \) and \( \sigma^2_\tau = 9 \) and (2) \( \sigma^2 = 16 \) and \( \sigma^2_\tau = 100 \). As mentioned previously in Section 3.2, these parameters are randomly specified, and different \( \beta \)s and both small and large variance component parameters are considered. Again, this principle is the approach to selecting true parameters through all simulations in this section. Missingness is generated from a Bernoulli distribution with the specified missing proportion as the success \( p \), and the probability of missing remains the same if the proceeding observation is observed and shifts to 0.25 otherwise. The resulting missingness follows the MNAR mechanism, where the missingness depends on unobserved data.
Figure 6 summarizes the selected rates of the true model against the missing rates of 0%, 5%, 10%, 20%, 50%, 60%, 70%, and 80% on 200 replicates, and the data are generated with small variances. Each trajectory represents the selected rates with the specified sample size. As the sample size increases, the selection rate increases regardless of the percent of missing values present in the data. Compared with Figure 3 where data are generated under MCAR, the selection rates seem better, and it is in fact due to the missingness generating process where some observations have only 25% chance of missing.

Figure 7 shows the competitive performance under large variances scenario versus the missing rates compared with the results shown in Figure 6 with small variances, indicating that the proposed method behaves well in selecting the true model with large variances.
Table 3. Parameter estimates (standard errors) for parameters on data with small variances.

| Missing rate = 0% Estimates (SE) | Missing rate = 20% Estimates (SE) | Missing rate = 50% Estimates (SE) | Missing rate = 80% Estimates (SE) |
|----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| MCAR, $\beta = (15, 16, 17, -10, 13, 27)^T$, $\sigma^2 = 1, \sigma^2_\tau = 9, m = 100, t = 3$ | MNAR, $\beta = (15, 16, 17, -10, 13, 27)^T$, $\sigma^2 = 1, \sigma^2_\tau = 9, m = 100, t = 3$ |
| $\beta_1$ | 14.9991 (0.9417) | 15.0381 (0.9281) | 15.0411 (1.1148) | 14.9909 (2.0783) |
| $\beta_2$ | 15.9984 (0.9453) | 16.0394 (0.9253) | 16.0476 (1.1091) | 15.9863 (2.0590) |
| $\beta_3$ | 16.9939 (0.9407) | 17.0466 (0.9170) | 17.0483 (1.1069) | 16.9906 (2.0656) |
| $\beta_4$ | -10.0004 (0.0177) | -10.0003 (0.0180) | -10.0000 (0.0216) | -10.0001 (0.0388) |
| $\beta_5$ | 13.0003 (0.0180) | 13.0001 (0.0176) | 13.0009 (0.0212) | 12.9985 (0.0409) |
| $\beta_6$ | 26.9991 (0.0175) | 26.9999 (0.0171) | 27.0007 (0.0208) | 27.0003 (0.0419) |
| $\sigma^2$ | 0.9946 (0.0984) | 0.9968 (0.1223) | 0.9936 (0.1931) | 0.9735 (0.5678) |
| $\sigma^2_\tau$ | 8.9931 (1.3103) | 8.8768 (1.3093) | 8.9620 (1.4142) | 8.6646 (2.2040) |

3.4. Estimation under MCAR and MNAR

Following the same data generating process to obtain samples with 100 observations, we investigate the accuracy and precision of the estimates through 1000 replicates under the scenarios of $\sigma^2 = 1, \sigma^2_\tau = 9$ and $\sigma^2 = 4, \sigma^2_\tau = 100$, each with missingness from MCAR and MNAR.

In the case of small variances, the results are presented in Table 3. All fixed effect estimates are close to their actual values, and the standard error of estimates increases as the missing rate increases, a direct consequence from the loss of a large portion of data. Under MCAR, the estimates of both $\sigma^2$ and $\sigma^2_\tau$ have a downward bias, but the magnitude of the bias is very small. With 80% of missingness, the estimates of $\sigma^2$ and $\sigma^2_\tau$ are off about 2% to 3% from their respective true values.

Under MNAR, the resulting estimates do not exhibit a systematic downward bias as the missing proportion increases. Overall, all estimates are very accurate with decreasing precision as the missing rate escalates.

The estimates on data with more substantial variances, $\sigma^2 = 4$ and $\sigma^2_\tau = 100$, are shown in Table 4, the results have a similar pattern as those in the small variances.

All simulations results in Tables 3 and 4 demonstrate that the proposed estimators generate satisfactory results in estimation regardless of the missing mechanism with the presence of moderate missingness in responses.

3.5. Comparison with MLE

In this subsection, we compare the proposed estimators with the maximum likelihood estimates on data with 0%, 20%, 50%, and 80% missing rates. Missing data are generated under missing completely at random (MCAR) and missing not at random (MNAR). Each scenario is carried out with (1): $\sigma^2 = 1$ and $\sigma^2_\tau = 9$, and (2): $\sigma^2 = 16$ and $\sigma^2_\tau = 100$. Both cases are iterated 1000 times to obtain the final estimates and standard errors. The
Table 4. Parameter estimates (standard errors) for parameters on data with large variances.

| Missing rate = 0% | Missing rate = 20% | Missing rate = 50% | Missing rate = 80% |
|-------------------|-------------------|-------------------|-------------------|
| Estimates (SE)    | Estimates (SE)    | Estimates (SE)    | Estimates (SE)    |
| \( \beta_1 \)     | 14.8808 (3.0378)  | 15.0493 (3.5739)  | 14.9918 (4.9442)  | 15.0819 (12.1032) |
| \( \beta_2 \)     | 15.8835 (3.0616)  | 16.0358 (3.5627)  | 16.0752 (4.9249)  | 16.1145 (12.1415) |
| \( \beta_3 \)     | 16.8898 (3.0338)  | 17.0529 (3.5542)  | 17.0683 (4.9400)  | 17.0349 (12.1465) |
| \( \sigma^2 \)    | 15.9156 (3.0294)  | 15.9565 (3.5508)  | 15.9443 (4.9225)  | 15.9015 (12.0873) |
| \( \tau \)        | 99.9108 (14.8410) | 98.5492 (14.8615) | 99.3245 (16.2892) | 95.4243 (26.4056) |

Maximum likelihood estimates are obtained from \( \text{lmer} \) function in \( \text{lme4} \) package in R (Version 3.6.2) where the missing values are omitted in the calculations. It should be noted that the differences between the proposed estimators and MLE are (1) the iterative algorithm imposed to obtain the final solutions, and (2) the former utilizes matrix theory to obtain the approximations of the determinant and the inversion of the variance–covariance components.

To find the coverage probabilities for the confidence interval (95%) estimation, the estimation process is iterated 500 times where each iteration contains 1000 data sets generated from each simulation setting: (1-a) MCAR with small variances, (1-b) MNAR with small variances, (2-a) MCAR with large variances, and (2-b) MNAR with large variances. The coverage probability is calculated over the 500 empirical 95% confidence intervals in each simulation setting.

3.5.1. Scenario (1-a): MCAR with \( \sigma^2 = 1 \) and \( \sigma^2 = 9 \)

Table 5 exhibits the mean estimate and the standard error on 1000 replicates. All fixed effects, \( \beta_{11}, \beta_{12}, \beta_{13}, \beta_1, \beta_2, \) and \( \beta_3 \) are almost identical in both the proposed estimators and MLE. A more noticeable increase in the standard errors of the estimates is present in all fixed effects as the missing proportion increases. The proposed method yields a more accurate \( \sigma^2 \) and \( \sigma^2 \) estimates under all considered missing rates, and the differences become more noticeable with moderate and excessive missingness.

To better investigate the biases, we calculate the following reference value called relative bias defined as

\[
\frac{1}{N} \sum_{k=1}^{N} \frac{\theta_k - \theta}{\theta},
\]

where \( \theta_k \) denotes the estimate of the \( k \)th sample, \( \theta \) is the true value, and \( N \) is the number of replications in the study. In our simulation settings, \( N = 1000 \). Table 6 features the relative
Table 5. Estimates (standard errors) of parameters for Scenario (1-a).

| Missing rate | Proposed est. (SE) | MLE (SE) | Proposed est. (SE) | MLE (SE) |
|--------------|-------------------|----------|-------------------|----------|
|              | 0%                | 20%      | 50%               | 80%      |
| $\beta_1$   | 14.9688 (0.7796)  | 14.9688 (0.7796) | 15.0134 (0.9118)  | 15.0135 (0.9117) |
| $\beta_2$   | 15.9695 (0.7857)  | 15.9695 (0.7857) | 16.0099 (0.9072)  | 16.0101 (0.9084) |
| $\beta_3$   | -10.0004 (0.0141) | -10.0004 (0.0141) | -9.9997 (0.0169)  | -9.9999 (0.0169) |
| $\beta_4$   | 13.0004 (0.0147)  | 13.0004 (0.0147) | 12.9994 (0.0176)  | 12.9994 (0.0176) |
| $\sigma^2$  | 0.9946 (0.0984)   | 0.9768 (0.1000)  | 0.9968 (0.1223)   | 0.9628 (0.1148)  |
| $\sigma^2_{\tau}$  | 8.9931 (1.3103)   | 8.9206 (1.2950)  | 8.9768 (1.2950)   | 8.8555 (1.2972)  |

Table 6. Relative biases of parameters for Scenario (1-a).

| Missing rate | Prop. est. | MLE | Prop. est. | MLE | Prop. est. | MLE | Prop. est. | MLE |
|--------------|-----------|-----|-----------|-----|-----------|-----|-----------|-----|
|              | 0%        |     | 20%       |     | 50%       |     | 80%       |     |
| $\beta_1$   | -0.00208  | -0.00208 | 0.00089  | 0.00090 | -0.00020  | -0.00020 | 0.00185  | 0.00124 |
| $\beta_2$   | -0.00191  | -0.00191 | 0.00062  | 0.00063 | 0.00116  | 0.00117 | 0.00224  | 0.00140 |
| $\beta_3$   | -0.00170  | -0.00170 | 0.00083  | 0.00084 | 0.00101  | 0.00101 | 0.00068  | -0.00017 |
| $\beta_4$   | 0.00004   | 0.00004 | -0.00003 | -0.00003 | 0.00006  | 0.00006 | 0.00015  | 0.00016 |
| $\sigma^2$  | 0.00003   | 0.00003 | -0.00005 | -0.00004 | 0.00014  | 0.00013 | 0.00007  | 0.00008 |
| $\sigma^2_{\tau}$  | <0.00001 | <0.00001 | -0.00002 | -0.00002 | 0.00002  | 0.00002 | 0.00003  | 0.00002 |
| $\sigma^2$  | -0.00543  | -0.02323 | -0.0323  | -0.03718 | -0.00642 | -0.07371 | -0.02646 | -0.37818 |
| $\sigma^2_{\tau}$  | -0.00077 | -0.00882 | -0.01369 | -0.01606 | -0.00962 | -0.00423 | -0.03727 | 0.00473 |

Biases, and the values in the table reassure our findings that the proposed method yields notably better estimates of $\sigma^2$ and $\sigma^2_{\tau}$.

Table 7 features the coverage probabilities in Scenario (1-a) for 95% confidence intervals. For the parameters of fixed effects, the values of coverage probabilities show that as the missing rate increases, the proposed method performs better than the MLE method, closer to the true confidence level, indicating that the proposed method is a good choice when the data have missing values in the LMM. Here, we comment on the coverage probabilities for the variance components $\sigma^2$ and $\sigma^2_{\tau}$. The coverage probabilities for $\sigma^2$ in the proposed method are significantly better than those for MLE due to the bias of MLE estimator of $\sigma^2$. As observed in the estimates of $\sigma^2$ in Table 5, the MLE estimate is notably off the true value. For $\sigma^2_{\tau}$, the coverage probabilities show that the proposed method outperforms the MLE method when the missing rate is small or medium, but not when the missing rate is as high as 80%.
Table 7. 95% Coverage probabilities of parameters for Scenario (1-a).

| Parameter | MCAR, m = 100, t = 3, β = (15, 16, 17, −10, 13, 27)^T, σ^2 = 1, σ^2 τ = 9 |
|-----------|---------------------------------------------------------------------|
|           | Missing rate = 0% | Missing rate = 20% | Missing rate = 50% | Missing rate = 80% |
| β_{t1}    | Prop. est. MLE     | Prop. est. MLE     | Prop. est. MLE     | Prop. est. MLE     |
| β_{t2}    | 0.944 0.958        | 0.944 0.948        | 0.952 0.952        | 0.972 0.954        |
| β_{t3}    | 0.936 0.956        | 0.956 0.954        | 0.954 0.958        | 0.968 0.952        |
| β_{t1}    | 0.908 0.924        | 0.952 0.970        | 0.966 0.966        | 0.948 0.946        |
| β_{t2}    | 0.924 0.940        | 0.944 0.966        | 0.968 0.979        | 0.960 0.950        |
| β_{t3}    | 0.956 0.954        | 0.968 0.952        | 0.960 0.958        | 0.952 0.954        |
| σ^2       | 0.964 0.304        | 0.936 0.148        | 0.920 0.012        | 0.928 0.001        |
| σ^2 τ     | 0.920 0.892        | 0.928 0.888        | 0.940 0.908        | 0.688 0.956        |

3.5.2. Scenario (1-b): MNAR with σ^2 = 1 and σ^2 τ = 9

We conduct a similar study with the scenario that the missing values are generated under MNAR where the succeeding response has 25% chance of missing if the previous observation is missing. Therefore, the missingness depends on the response itself and thus follows the MNAR mechanism. All parameters and sample sizes remain the same as in the previous scenario.

Table 8 displays the mean estimates and standard errors of 1000 replicates for all parameters, and the estimates of the proposed method in the table demonstrate that they are very close to and closer to the true parameters compared with those under MLE. Table 9 shows the corresponding relative biases, and these relative biases are very close to zero. Note that the unbiased property of the proposed estimators is proved in Section A.3 of the Appendix. The observed values in Tables 8 and 9 confirm the two aspects of the proposed estimators: (1) the fixed effects associated with the covariates β_{t1}, β_{t2}, and β_{t3} estimates obtained from the proposed estimators are unbiased regardless of the missing proportions, and (2) the proposed estimates are less biased compared with the MLE method.

Again, the coverage probabilities for the interval estimation are calculated using 500 replicates where each replicate contains 1000 generated data. Table 10 shows that for the fixed effect parameters, the proposed method produces the coverage probabilities closer to the true confidence level as the missing rate increases, which implies the proposed method performs well with various amount of missing values. As mentioned previously, the coverage probabilities for σ^2 in the proposed method are significantly better than those for MLE due to the bias of MLE estimator of σ^2 and the bias is shown in the point estimates of σ^2 in Table 8. For σ^2 τ, the coverage probabilities show that the proposed method outperforms the MLE method when the missing rate is small or medium, but not when the missing rate is very high.

3.5.3. Scenario (2-a): MCAR with σ^2 = 16 and σ^2 τ = 100

The mean estimates, standard errors, and relative biases for the MCAR case are shown in Tables 11 and 12. The performance from both approaches in estimating the fixed effects, β_{t1}, β_{t2}, β_{t3}, β_{t1}, β_{t2}, and β_{t3}, are almost identical. Again, a noticeable increase in the standard errors of the estimates is observed in all fixed effects as the missing rate increases. For the variance components, the proposed method produces much more accurate estimate values under all considered missing rates than the MLE method.
Table 8. Estimates (standard errors) of parameters for Scenario (1-b).

|                | Missing rate = 0% | Missing rate = 20% | Missing rate = 50% | Missing rate = 80% |
|----------------|-------------------|--------------------|--------------------|--------------------|
|                | Proposed est. (SE) | MLE (SE)           | Proposed est. (SE) | MLE (SE)           |
| $\beta_1$     | 14.9991 (0.9417)   | 14.9990 (0.9417)   | 15.0381 (0.9281)   | 15.0382 (0.9282)   |
| $\beta_2$     | 15.9984 (0.9453)   | 15.9984 (0.9453)   | 16.0394 (0.9253)   | 16.0395 (0.9295)   |
| $\beta_3$     | 16.9939 (0.9407)   | 16.9938 (0.9407)   | 17.0446 (0.9170)   | 17.0446 (0.9171)   |
| $\beta_4$     | -10.0004 (0.0177)  | -10.0004 (0.0178)  | -10.0003 (0.0180)  | -10.0003 (0.0180)  |
| $\beta_5$     | 13.0003 (0.0180)   | 13.0003 (0.0181)   | 13.0001 (0.0176)   | 13.0001 (0.0177)   |
| $\beta_6$     | 26.9991 (0.0175)   | 26.9991 (0.0176)   | 26.9999 (0.0171)   | 26.9999 (0.0171)   |
| $\sigma^2_1$  | 1.004 (0.1257)     | 0.9677 (0.1258)    | 0.9931 (0.1274)    | 0.9681 (0.1206)    |
| $\sigma^2_2$  | 8.943 (1.3619)     | 8.8372 (1.3399)    | 8.9785 (1.3677)    | 8.8769 (1.3520)    |

Table 9. Relative biases of parameters for Scenario (1-b).

|                | Missing rate = 0% | Missing rate = 20% | Missing rate = 50% | Missing rate = 80% |
|----------------|-------------------|--------------------|--------------------|--------------------|
|                | Prop. est.        | MLE                | Prop. est.        | MLE                |
| $\beta_1$     | -0.00005          | -0.00007           | 0.00254           | 0.00255           |
| $\beta_2$     | -0.00009          | -0.00010           | 0.00247           | 0.00247           |
| $\beta_3$     | -0.00036          | -0.00036           | 0.00262           | 0.00263           |
| $\beta_4$     | 0.00004           | 0.00004            | 0.00003           | 0.00003           |
| $\beta_5$     | 0.00002           | 0.00002            | 0.00001           | 0.00001           |
| $\beta_6$     | 0.00003           | -0.00003           | 0.00001           | 0.00001           |
| $\sigma^2_1$  | 0.00400           | -0.03234           | -0.0692           | -0.03191          |
| $\sigma^2_2$  | -0.00633          | -0.01809           | -0.00239          | -0.01367          |

Table 10. 95% Coverage probabilities of parameters for Scenario (1-b).

|                | Missing rate = 0% | Missing rate = 20% | Missing rate = 50% | Missing rate = 80% |
|----------------|-------------------|--------------------|--------------------|--------------------|
|                | Prop. est.        | MLE                | Prop. est.        | MLE                |
| $\beta_1$     | 0.948             | 0.944              | 0.938             | 0.938              |
| $\beta_2$     | 0.944             | 0.942              | 0.946             | 0.946              |
| $\beta_3$     | 0.938             | 0.948              | 0.946             | 0.946              |
| $\beta_4$     | 0.968             | 0.966              | 0.964             | 0.968              |
| $\beta_5$     | 0.952             | 0.960              | 0.952             | 0.952              |
| $\beta_6$     | 0.940             | 0.946              | 0.960             | 0.960              |
| $\sigma^2_1$  | 0.940             | 0.132              | 0.952             | 0.140              |
| $\sigma^2_2$  | 0.960             | 0.904              | 0.912             | 0.872              |

Table 13 features the coverage probabilities for Scenario (2-a) MCAR with large variances, and the values show that both the proposed method and MLE produce competitive results in estimating the parameters of fixed effects. Again, as in Scenarios (1-a) and (1-b),
Table 11. Estimates (standard errors) of parameters for Scenario (2-a).

| Missing rate | MCAR, \( m = 100, t = 3, \beta = (15, 16, 17, -10, 13, 27)^T, \sigma^2 = 16, \sigma^2_\tau = 100 \) |
|--------------|--------------------------------------------------------------------------------------------------|
|              | Proposed est. (SE) | MLE (SE) | Proposed est. (SE) | MLE (SE) |
| \( \beta_1 \) | 14.8808 (3.0378)   | 14.8809 (3.0377) | 15.0493 (3.5739) | 15.0499 (3.5735) |
| \( \beta_2 \) | 15.8835 (3.0616)   | 15.8836 (3.0615) | 16.0358 (3.5627) | 16.0364 (3.5623) |
| \( \beta_3 \) | 16.8898 (3.0338)   | 16.8898 (3.0337) | 17.0529 (3.5542) | 17.0543 (3.5539) |
| \( \beta_4 \) | -10.0017 (0.0562)  | -10.0017 (0.0562) | -9.9998 (0.0672) | -9.9989 (0.0672) |
| \( \beta_5 \) | 13.0016 (0.0584)   | 13.0016 (0.0584) | 12.9976 (0.0698) | 12.9976 (0.0698) |
| \( \beta_6 \) | 26.9996 (0.0562)   | 26.9996 (0.0562) | 26.9979 (0.0690) | 26.9979 (0.0690) |
| \( \sigma^2 \) | 15.9156 (1.5745)   | 15.6301 (1.6000) | 15.9565 (1.9590) | 15.4098 (1.8378) |
| \( \sigma^2_\tau \) | 99.9108 (14.8410)  | 99.1046 (14.6124) | 98.5492 (14.8615) | 98.4140 (14.7673) |

Table 12. Relative biases of parameters for Scenario (2-a).

| Missing rate | MCAR, \( m = 100, t = 3, \beta = (15, 16, 17, -10, 13, 27)^T, \sigma^2 = 16, \sigma^2_\tau = 100 \) |
|--------------|--------------------------------------------------------------------------------------------------|
|              | Proposed est. | MLE | Proposed est. | MLE | Proposed est. | MLE | Proposed est. | MLE |
| \( \beta_1 \) | -0.00795 | -0.00794 | 0.00329 | 0.00333 | -0.00055 | -0.00059 | 0.00546 | 0.00393 |
| \( \beta_2 \) | -0.00728 | -0.00728 | 0.00224 | 0.00227 | -0.00470 | 0.00470 | 0.00716 | -0.00486 |
| \( \beta_3 \) | -0.00648 | -0.00648 | 0.00311 | 0.00314 | -0.00402 | 0.00402 | 0.00205 | -0.00078 |
| \( \beta_4 \) | 0.00017 | 0.00017 | -0.00012 | -0.00011 | 0.00024 | 0.00024 | 0.00064 | 0.00064 |
| \( \beta_5 \) | 0.00012 | 0.00012 | -0.00018 | -0.00018 | 0.00055 | 0.00055 | 0.00026 | -0.00030 |
| \( \beta_6 \) | -0.00002 | -0.00002 | 0.00008 | 0.00008 | 0.00009 | 0.00009 | 0.00016 | 0.00009 |
| \( \sigma^2 \) | -0.00527 | -0.02312 | -0.00272 | -0.03689 | -0.00348 | -0.07170 | -0.00616 | -0.34548 |
| \( \sigma^2_\tau \) | -0.00089 | -0.00895 | -0.01451 | -0.01586 | -0.00676 | -0.00562 | -0.04576 | 0.00894 |

the coverage probabilities for \( \sigma^2 \) in the proposed method are apparently better than those for the MLE method due to the bias of MLE estimator of \( \sigma^2 \). For \( \sigma^2_\tau \), when the missing rate is 80%, the coverage probability of proposed is not optimal. Generally, the coverage rates for \( \sigma^2 \) and \( \sigma^2_\tau \) demonstrate that the proposed method outperforms the MLE in estimation. Therefore, we can conclude that the proposed method generally behaves efficaciously with the missing values present, but may not perform well with a high missing rate.

3.5.4. Scenario (2-b): MNAR with \( \sigma^2 = 16 \) and \( \sigma^2_\tau = 100 \)

Following the same setups for generating samples, the evaluations of data with high noise and MNAR are in Tables 14 and 15. The performance from both approaches in estimating fixed effects, \( \beta_{t1}, \beta_{t2}, \beta_{t3}, \beta_1, \beta_2, \) and \( \beta_3 \), are again almost identical. It can be observed that as the missing proportion increases, the standard errors of the estimates in all fixed effects increase. For the variance components, the proposed method overall produces much more accurate values than MLE. For the \( \sigma^2_\tau \) estimates, the proposed method has less precision.
Table 13. 95% Coverage probabilities of parameters for Scenario (2-a).

| Missing rate | MCAR, m = 100, t = 3, β = (15, 16, 17, −10, 13, 27)', σ2 = 16, σ2 = 100 |
|--------------|--------------------------------------------------------------------------|
|             | Prop. est. | MLE | Prop. est. | MLE | Prop. est. | MLE | Prop. est. | MLE |
| β1          | 0.956      | 0.956 | 0.940      | 0.952 | 0.950      | 0.948 | 0.964      | 0.950 |
| β2          | 0.956      | 0.956 | 0.952      | 0.952 | 0.966      | 0.960 | 0.968      | 0.948 |
| β3          | 0.956      | 0.958 | 0.942      | 0.954 | 0.952      | 0.952 | 0.960      | 0.940 |
| β4          | 0.926      | 0.926 | 0.952      | 0.970 | 0.956      | 0.964 | 0.956      | 0.948 |
| β5          | 0.938      | 0.938 | 0.944      | 0.966 | 0.968      | 0.968 | 0.960      | 0.966 |
| β6          | 0.956      | 0.954 | 0.964      | 0.954 | 0.956      | 0.956 | 0.956      | 0.956 |
| σ2          | 0.964      | 0.312 | 0.924      | 0.152 | 0.936      | 0.020 | 0.948      | 0.001 |
| σ2          | 0.920      | 0.884 | 0.912      | 0.884 | 0.916      | 0.908 | 0.576      | 0.948 |

Table 14. Estimates (standard errors) of parameters for Scenario (2-b).

| Missing rate | MNAR, m = 100, t = 3, β = (15, 16, 17, −10, 13, 27)', σ2 = 16, σ2 = 100 |
|--------------|--------------------------------------------------------------------------|
|             | Proposed est. (SE) | MLE (SE) | Proposed est. (SE) | MLE (SE) | Proposed est. (SE) | MLE (SE) | Proposed est. (SE) | MLE (SE) |
| β1          | 14.9962 (3.6809) | 14.9958 (3.6806) | 15.1530 (3.6239) | 15.1533 (3.6241) |
| β2          | 15.9934 (3.6943) | 15.9930 (3.6941) | 16.1585 (3.6145) | 16.1587 (3.6146) |
| β3          | 16.9756 (3.6764) | 16.9751 (3.6761) | 17.1791 (3.5815) | 17.1793 (3.5816) |
| β4          | −10.0017 (0.0706) | −10.0017 (0.0706) | d10.0012 (0.0716) | −10.0012 (0.0716) |
| β5          | 13.0012 (0.0718) | 13.0012 (0.0718) | 13.0002 (0.0700) | 13.0002 (0.0700) |
| β6          | 26.9965 (0.0697) | 26.9965 (0.0697) | 26.9997 (0.0679) | 26.9997 (0.0679) |
| σ2          | 16.0740 (2.0136) | 15.4903 (2.0129) | 15.8995 (2.0400) | 15.4964 (1.9314) |
| σ2          | 99.2602 (16.4614) | 98.1705 (15.2867) | 99.7023 (15.4489) | 98.6044 (15.4470) |

compared with the MLE method with a high missing rate of 80%. For the σ2 estimates, the MLE method owns notably less accuracy because of the bias of MLE estimator of σ2. Table 16 features the coverage probabilities of confidence interval estimation for Scenario (2-b) MNAR with large variances. The values in the table show that the proposed method produces competitive coverage probabilities under almost all four missing rates for the parameters of fixed effects. As in the last three scenarios, based on the empirical confidence intervals, the coverage probabilities for σ2 in the proposed method are significantly better than those for the MLE method due to the bias of MLE estimator of σ2 and the bias is also shown in Table 14. This bias also implies the reason that the REML is more effective than the MLE for σ2 in the LMM. For σ2, the coverage probabilities show that the proposed method outperforms the MLE method when the missing rate is small or medium, but not when the missing rate is very high. We can conclude that the proposed method in general performs effectively when involving missing values even though it may not be very optimal with a high missing rate.
Table 15. Relative biases of parameters for Scenario (2-b).

| Missing rate = 0%         | Missing rate = 20%        | Missing rate = 50%        | Missing rate = 80%        |
|---------------------------|---------------------------|---------------------------|---------------------------|
| Prop. est.                | MLE                       | Prop. est.                | MLE                       |
| $\beta_{t1}$              | $-0.00025$                | $0.01020$                 | $0.01070$                 |
| $\beta_{t2}$              | $-0.00041$                | $0.00991$                 | $0.01162$                 |
| $\beta_{t3}$              | $-0.00143$                | $0.01054$                 | $0.01120$                 |
| $\beta_1$                 | $0.00017$                 | $0.00012$                 | $0.00001$                 |
| $\beta_2$                 | $-0.00009$                | $0.00002$                 | $0.00001$                 |
| $\beta_3$                 | $-0.00013$                | $-0.00001$                | $0.00009$                 |
| $\sigma^2$                | $0.00463$                 | $-0.00628$                | $-0.00896$                |
| $\sigma^2_{\tau}$        | $-0.00740$                | $-0.01829$                | $-0.01433$                |

Table 16. 95% Coverage probabilities of parameters for Scenario (2-b).

| Missing rate = 0%         | Missing rate = 20%        | Missing rate = 50%        | Missing rate = 80%        |
|---------------------------|---------------------------|---------------------------|---------------------------|
| Prop. est.                | MLE                       | Prop. est.                | MLE                       |
| $\beta_{t1}$              | $0.940$                   | $0.938$                   | $0.952$                   |
| $\beta_{t2}$              | $0.942$                   | $0.944$                   | $0.952$                   |
| $\beta_{t3}$              | $0.946$                   | $0.948$                   | $0.952$                   |
| $\beta_1$                 | $0.968$                   | $0.960$                   | $0.964$                   |
| $\beta_2$                 | $0.962$                   | $0.950$                   | $0.932$                   |
| $\beta_3$                 | $0.946$                   | $0.948$                   | $0.948$                   |
| $\sigma^2$                | $0.944$                   | $0.948$                   | $0.932$                   |
| $\sigma^2_{\tau}$        | $0.940$                   | $0.900$                   | $0.904$                   |

In the above four scenarios, we note that for the MLE method, the between-subject variability ($\sigma^2_{\tau}$) is mostly caught in the estimation, but not for the within-subject variability ($\sigma^2$), which implies the MLE method provides a good estimate of $\sigma^2_{\tau}$ and a poor estimate of $\sigma^2$. Practically, a high missing rate reduces the between-subject variability, so the estimate of $\sigma^2_{\tau}$ is off the true value, leading to the lower values of the coverage probability under 80% missing rate compared with small and medium missing rates. Therefore, when the missing rate is very high, the coverage rate for $\sigma^2_{\tau}$ is not well-behaved for the proposed method.

3.6. Summary of selection and estimation performance

The simulations above are conducted and the results show that: (1) the performance of the proposed method in model selection under various settings are of satisfaction in the sense that the selected rates of the correct model are above 90% with a moderate amount of missingness (missing rate of 40% or less), (2) the estimates by the proposed method are close to the true parameters in both the small and large variability data, (3) the comparison of the proposed method with MLE in point estimation shows that the proposed method produces notably better estimates of both $\sigma^2$ and $\sigma^2_{\tau}$, and (4) the comparison of the proposed method with MLE in the coverage probability of confidence interval estimation shows that the proposed method performs effectively and outperforms MLE with missing values in the LMM.
Table 17. Incomplete blood lead data.

| Subject | Group | Week 0 | Week 1 | Week 4 | Week 6 |
|---------|-------|--------|--------|--------|--------|
| 2       | T     | 26.5   | .      | 19.5   | –      |
| 3       | T     | 25.8   | 23     | –      | –      |
| 5       | T     | 20.4   | 2.8    | –      | –      |
| 6       | T     | 20.4   | 5.4    | –      | 11.9   |
| 12      | T     | 24.8   | 23.1   | –      | 24.6   |
| 14      | T     | 27.9   | 6.3    | –      | 16.3   |

4. An application

In this section, we apply the proposed approach to incomplete blood lead concentrations data. The data recorded the results from the treatment of lead-exposed children, a placebo-controlled and randomized study on Succimer, a treatment that enhances urinary excretion of lead, in children with blood levels of 20–44 μg/dl. The utilized data were from TLC Research Group [14], and the collection of the data was also described in it. As mentioned in TLC Research Group [14], the data were collected following the rules of the Centers for Disease Control and Prevention (CDC) as described in CDC [1]. We analyze the blood lead levels from 100 children who were randomly assigned to either the lead Chelation treatment with Succimer or to the placebo group. Every participant was observed four times; each at baseline (μ₁), week 1 (μ₂), week 4 (μ₃), and week 6 (μ₄). The incompleteness comes from unobserved blood lead levels. A subset of data is presented in Table 17.

In this case, the ability to model the correlated blood lead levels from one individual would be essential in examining the changes in blood lead levels over the four-time courses. For this reason, we accommodate the linear mixed-effects model to measure the magnitude of the variations from subjects to subjects along with the variations from within-subject.

One straightforward model we can employ is to assign a varying intercept for each subject, where the varying intercept is characterized by the subject variability, σ². The primary objective in the analysis is to examine the changes in blood lead levels, so statistically, we test whether the blood lead levels at time t, μᵣ, for t = 1, . . . , 4, are significantly different.

Figure 8 shows the blood lead levels at each of the four weeks from 50 subjects. Notice that each individual appears to have their linear trajectory with some fluctuations over time. The variation from one subject to the other subject (from one linear trajectory to the other) is often referred to as the between-subject variation. As the name suggests, the variation is subject-specific and treating the subject as a random effect in modeling considers the magnitude of the variation. The variation from the individual observations to the individual’s trajectory is referred to as the within-subject variation. It represents the deviations among the observations within the specific individual.

The triangular-shaped points represent the average blood lead levels at each week based on the available data. For the children receiving the active treatment, the mean blood lead level at baseline is about 26 μg/dl, and the mean concentration at week 1, one week after the TLC treatment was received, dropped to 13 μg/dl. The mean blood lead levels for the children in the placebo group stayed around 26 and 24 μg/dl from week 0 to week 6. The preliminary results based on the graphical inspection of the data show that (1) the linear trajectories for group 0 should be different from group 1 and (2) the between-subject variations in both groups may vary since the individual lines are more vertically spread out in the treatment group than in the placebo group.
To statistically formulate the problem, we denote the blood lead level from subject $i$ at time $t$ as $y_{it}$ and denote the mean blood lead level at time $t$ as $\mu_t$. The model addresses the subject-specific variation, denoted as $b_i$, referring as a subject-specific random effect in the linear mixed model scheme. The effect of groups is also considered and denoted as $\beta$ with an indicator covariate $X_i$ having a value 1 if the subject was in the treatment group and a value of 0 otherwise. The explicit model specification is as follows:

$$y_{it} = \mu_t + \beta \times X_i + b_i + \epsilon_{it}, \quad \text{for } i = 1, \ldots, 100, \text{ and } t = 1, \ldots, 4,$$

with $b_i \sim N(0, \sigma^2_I)$ and $\epsilon_{it} \sim \text{iid } N(0, \sigma^2)$. Note that $\sigma^2_I$ models the subject variability, and $\epsilon_{it}$ models the within-subject variability. Since the data consist of some missing values, we employ the approach in Section 2.2, which yields likelihood estimators. Notice that the current model considers the within-subject variations for both individuals in the placebo group and individuals in the TLC treatment group are of the same magnitude.

We carry out the model selection by the implementation of the SIC as the model selection assessment to choose the most appropriate model. The model posterior probabilities are then calculated for each candidate model based on the SIC value. The model that yields the highest posterior probability will be the most appropriate model for the data. To establish the ultimate goal of the analysis, that is, to determine whether the mean blood lead levels have significant changes over time with consideration of both within and between-subject variations, the candidate models in consideration are decided by the equality between the effects on blood lead level at time $t$. Based on the graphical inspection, we consider 15 models shown in Table 18.
Table 18. Posterior model probability for each candidate model.

| Model          | Probability |
|----------------|-------------|
| Model 1        | $\mu_1 = \mu_2 = \mu_3 = \mu_4 < 0.00001$ |
| Model 2        | $\mu_1 = \mu_2 = \mu_3, \mu_4 < 0.00001$ |
| Model 3        | $\mu_1 = \mu_3 = \mu_4, \mu_2 < 0.00001$ |
| Model 4        | $\mu_1 = \mu_2 = \mu_4, \mu_3 < 0.00001$ |
| Model 5        | $\mu_1, \mu_2 = \mu_3 = \mu_4, 0.13174$ |
| Model 6        | $\mu_1 = \mu_2, \mu_3, \mu_4 < 0.00001$ |
| Model 7        | $\mu_1 = \mu_2, \mu_3, \mu_4, 0.72448$ |
| Model 8        | $\mu_1, \mu_2, \mu_3, \mu_4 < 0.00001$ |
| Model 9        | $\mu_1, \mu_2, \mu_3 = \mu_4, 0.08325$ |
| Model 10       | $\mu_1, \mu_2, \mu_3 = \mu_4, 0.00824$ |
| Model 11       | $\mu_1, \mu_2, \mu_3, \mu_4 < 0.00001$ |
| Model 12       | $\mu_1 = \mu_2, \mu_3, \mu_4 < 0.00001$ |
| Model 13       | $\mu_1 = \mu_2, \mu_3 = \mu_4 < 0.00001$ |
| Model 14       | $\mu_1 = \mu_2, \mu_3 = \mu_4 = 0.00001$ |
| Model 15       | $\mu_1, \mu_2, \mu_3, \mu_4 = 0.05228$ |

For instance, model ‘$\mu_1 = \mu_2 = \mu_3 = \mu_4$’ indicates there are no significant changes in the blood lead levels from week 0 to week 6; whereas, model ‘$\mu_1, \mu_2, \mu_3, \mu_4$’ represents that there are significant differences in the blood lead levels from week 0 to week 6. Model ‘$\mu_1 = \mu_2, \mu_3, \mu_4$’ represents the blood lead levels at week 0 and week 1 are not different but very different from the blood lead levels both at week 4 and at week 6, and differences in blood lead levels at week 4 and week 6 are also significant.

Table 18 features the 15 model posterior probabilities, suggesting that the differences in mean blood levels from week 1 to week 4 are not significant but the differences both from week 0 to week 1 and from week 4 to week 6 are statistically significant. With model 3 being the most appropriate model yielding the highest model posterior probability, the parameter estimates of mean blood lead levels are $(\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4)^T = (26.74, 20.02, 20.02, 22.27)$, and the standard deviation estimates are $\hat{\sigma}_\tau = 5.44$ and $\hat{\sigma} = 4.43$. The estimated group factor is -0.2173 with a standard error of 0.9527, which does not exhibit a significant group effect. A follow-up investigation for fitting the model without the group factor is executed, and the SIC infers the model without a group factor in consideration yields a better fit. These results may also be a consequence of the negligence of the varying scales in between-subject variations between the treatment and placebo groups. The updated model yields the estimated mean blood lead levels as $(26.41, 19.70, 19.70, 21.95)$ and the estimated standard deviations as $\hat{\sigma}_\tau = 5.44$ and $\hat{\sigma} = 4.43$.

One resolution for considering a non-homogeneity within-subject variation between the treatment and placebo groups is to fit a separate mixed effects model for each group. Following the same notations for the blood lead levels, the subject-specific random effect, and the within-subject variation, the model in consideration has a form of

$$y_{it} = \mu_t + b_i + \epsilon_{it}, \text{ for } i = 1, \ldots, 50, \text{ and } t = 1, \ldots, 4,$$

with $b_i \sim N(0, \sigma^2)$ and $\epsilon_{it} \iid N(0, \sigma^2)$.

We employ the same estimators in Equations (3), (4), and (5) to resolve the incompleteness and re-evaluate the 15 candidate models. Then the SIC values are calculated, and based on these values, the posterior probability for each model is computed. Table 19 features the posterior probabilities. With the suggested model as ‘$\mu_1, \mu_2 = \mu_3, \mu_4$’, the estimated
Table 19. Posterior model probabilities for each group.

| Model | Group 1 | Group 0 |
|-------|---------|---------|
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.00000 | 0.00001 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.00000 | 0.00008 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.00000 | <0.00001 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.01119 | 0.58262 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.00000 | 0.00110 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.82374 | 0.12234 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.03516 | 0.21833 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.00042 | <0.00001 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.00000 | 0.00001 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.00087 | 0.04370 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.00000 | 0.01276 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.00000 | 0.00040 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.00030 | <0.00001 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 0.12562 | 0.01898 |

mean blood lead levels are (26.54, 14.26, 14.26, 20.22) and the estimated standard deviations are $\hat{\sigma}_T = 4.38$ and $\hat{\sigma} = 5.41$. The aforementioned modeling process is repeated on the observations from the placebo group. Table 19 suggests that Model $'\mu_1, \mu_2 = \mu_3 = \mu_4'$ is the most appropriate among all 15 models. The estimates of mean blood lead levels are $(\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4)^T = (26.27, 24.13, 24.13, 24.13)$ and the standard deviation estimates are $\hat{\sigma}_T = 4.93$ and $\hat{\sigma} = 2.07$.

With consideration of the individual variability, the overall blood lead level for participants in the treatment group dropped drastically and are statistically significant within the first week after the active treatment was administrated, and the blood level stayed at about the same level within the month, and went up at week 6. For the placebo group, the overall blood lead level at the baseline was significantly different from those at the rest of the periods. The original study [14] mentioned circumstances, such as the exposure of lead paint in the participants’ housing, which posed potential challenges for the experiments. Nevertheless, with the implementation of linear mixed-effects modeling and the proposed approach to address data incompleteness, the final results illustrate the significance of changes in a specific time courses for both groups.

5. Concluding remarks and discussion

We propose an approach to incorporate incomplete data for estimation problems in the linear mixed-effects model with compound symmetric covariance structures. We derive likelihood-based estimators with the implementation of the indicator matrix for the accommodation of missing data. Due to the non-closed-form expressions in estimators, as in all other estimation approaches in linear mixed models, we apply the fixed-point iterative approach to obtain the final estimates. The iterative algorithm is explicitly listed, and simulations are conducted to demonstrate the successful iterative process. The simulation results show that all processes converge within the first few iterations on data with a small, moderate, and excessive amount of missingness.

To evaluate the proposed approach on incomplete data, numerous simulations are conducted to examine (1) the model selection performance under various amounts of missingness, (2) the selection performance under different missing mechanisms, (3) the
estimation performance under various missing rates, and (4) the comparison in point estimation and coverage probabilities of confidence interval estimation between the proposed method and MLE.

The simulation results demonstrate the selected rates of the true model, accessed by the SIC with the implementation of the proposed approach, are satisfactory with a moderate amount of missingness, regardless of the missing mechanisms (MCAR or MNAR).

In the studies of the comparison between the proposed approach and MLE, the proposed method provides a noticeably better estimate for the variance components $\sigma^2$ and $\tau^2$, and yields competitive estimates for fixed effects in the compound symmetric models.

We apply the proposed approach to the blood lead concentration data. The applied results illustrate the effectiveness of the proposed approach on estimation and on model selection with missing values in the linear mixed modeling framework with the compound symmetric covariance structure.

The proposed method focuses on estimation and selection in the linear mixed models under compound symmetric structures. We will extend it to other correlation structures in future research. We are currently exploring a similar method under AR(1) covariance structure in a follow-up project.

**Acknowledgments**

The authors would like to express their appreciation to an associate editor and the referees for providing thoughtful and insightful comments, which helped to improve the original version of this manuscript.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**References**

[1] CDC, *Preventing Lead Poisoning in Young Children*, US Department of Health and Human Services, Centers for Disease Control, Atlanta, 1991.

[2] M. Delattre, M. Lavielle, and M.-A. Poursat, *A note on BIC in mixed-effects models*, Electron. J. Stat. 8 (2014), pp. 456–475.

[3] A.P. Dempster, N.M. Laird, and D.B. Rubin, *Maximum likelihood from incomplete data via the em algorithm*, J. R. Stat. Soc. Ser. B (Methodol.) 39 (1977), pp. 1–38.

[4] R.A. Fisher, *The correlation between relatives on the supposition of mendelian inheritance*, Trans. R. Soc. Edinb. 52 (1918), pp. 399–433.

[5] D.A. Harville, *Maximum likelihood approaches to variance component estimation and to related problems*, J. Am. Stat. Assoc. 72 (1977), pp. 320–338.

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

[7] J.J. Miller, *Asymptotic properties and computation of maximum likelihood estimates in the mixed model of the analysis of variance*, PhD thesis, Department of Statistics, Stanford University, Stanford, California, 1973.

[8] J.J. Miller, *Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance*, Ann. Stat. 5 (1977), pp. 746–762.

[9] H.D. Patterson and R. Thompson, *Recovery of inter-block information when block sizes are unequal*, Biometrika 58 (1971), pp. 545–554.

[10] D.B. Rubin, *Inference and missing data*, Biometrika 63 (1976), pp. 581–592.
[11] G. Schwarz, *Estimating the dimension of a model*, Ann. Stat. 6 (1978), pp. 461–464.

[12] J. Shang, *A multiple comparison procedure based on a variant of the Schwarz information criterion in a mixed model*, Commun. Stat. Theory Methods 39 (2010), pp. 1095–1109.

[13] J. Shang and J.E. Cavanaugh, *A Bayesian multiple comparison procedure for order-restricted mixed models*, Int. Stat. Rev. 72 (2008), pp. 268–284.

[14] TLC Research Group, *The treatment of lead-exposed children (TLC) trial: Design and recruitment for a study of the effect of oral chelation on growth and development in toddlers*, Paediatr. Perinat. Epidemiol. 12 (1998), pp. 313–333.