Variable Selection for a Mixture of Linear Mixed Effects Models

Yian Zhang * Lei Yang† Yongzhao Shao‡

Abstract
Linear mixed effects models have been widely used in medical studies and other applications to deal with repeated-measure or clustered data. In reality, these data can be a mixture of different groups of subjects. Then, it is preferred to fit the group-specific model if the group indicator is available so that existing variable selection methods for linear mixed effects models can be used. In this study, we consider a situation that the data involve two groups of subjects but the grouping information is only partially known. We construct a finite mixture of linear mixed effects models via joint modeling of repeated measures using linear mixed effects models and missing group status using a logistic regression. We propose a penalized likelihood method with SCAD penalty function using EM algorithm for variable selection in this mixture model. BIC is applied to determine the tuning parameters and covariates. Simulation studies are used to demonstrate the performance of the proposed method.

Key Words: linear mixed effects model, variable selection, mixture model, penalty, EM

1. Introduction

Linear mixed effects model, involving fixed effects and random effects, is especially useful for handling longitudinal data (e.g., blood glucose measures of a patient at each week in a month) and clustered data (e.g., monthly incomes of family members grouped by household in a town). It has become a major research topic in statistics since C. R. Henderson first introduced the framework of the model in his 1949 and 1950 papers [1, 2]. Its estimation methods and variable selection approaches have been studied and discussed in many publications [3, 4, 5, 6, 7, 8, 9, 10]. Due to the flexibility of modeling fixed and random factors at the same time with assuming the correlation among random effects [9], it is also a highly influential statistical model which is applied in diverse disciplines such as economics [11], genetics and biology [12], physiology [13], psychology [14], environmental science [15], social and behavioral science [16, 17], and clinical trials [18, 19]. In addition, with increasing complexity of data, linear mixed effects model has been extended to incorporate with survival models via joint modeling [20] and has also been constructed as finite mixture models to account for the unobserved heterogeneity between subjects while allowing for correlations from the same subject [21, 22]. The finite mixtures of linear mixed effects models are useful in many areas and some algorithms for fitting this mixture model have been discussed in the literature such as Markov Chain Monte Carlo (MCMC) algorithm [21] and expectation-maximization (EM) algorithm [22]. However, few of articles studied the variable selection for this kind of model.

In this study, inspired by the paper of Grün in 2008 [22] and the paper of Khalili and Chen in 2007 [23], we introduce an EM algorithm for maximum likelihood (ML) estimation of a constructed finite mixtures of linear mixed effects models and propose a penalized

*Department of Environmental Medicine (Division of Biostatistics), New York University, 650 First Avenue, 5th Floor, New York, NY 10016

†Department of Environmental Medicine (Division of Biostatistics), New York University, 650 First Avenue, 5th Floor, New York, NY 10016

‡Department of Environmental Medicine (Division of Biostatistics), New York University, 650 First Avenue, 5th Floor, New York, NY 10016

variable selection procedure for it. Similar as the finite mixtures of mixed effects models in Grün's paper [22], our mixture models are also constructed by joint modeling of repeated measures using linear mixed effects models and latent group information using a concomitant model - logistic regression model. But, instead of treating the group label for each subject as complete missing for the EM framework in [22], we extend the framework to adapt for partial missing. That is, sometimes, we have already had the group information for partial subjects and have already known these groups of subjects have heterogeneity on the response variable. Comparing to Khalili and Chen [23] who proposed a likelihood variable selection procedure with a penalty such as smoothly clipped absolute deviation (SCAD) [24] in finite mixture of regression models but without modeling the random effects and without selecting the informative variables for the concomitant linkage function, we not only involve the random effects and use logistic model as a linkage in the mixture models but also propose a penalized approach which can be used for clustering the mixture types of data and select the informative variables for both fixed effects and the concomitant model simultaneously. Since the framework is based on EM algorithm, unknown parameters will be updated alternately. In each sub-step, the unknown parameters have analytical solution and the computational cost is relatively cheap. The Bayesian information criterion (BIC) [25] will be introduced to select the tuning parameter for variable selection process and the performance is demonstrated via various simulation scenarios and real data analysis.

The rest of this article is organized as follows. In section 2, we introduce the general EM framework of the ML estimation for the proposed finite mixtures of linear mixed effects models and then propose the penalized variable selection approach. Some simulation examples are introduced to demonstrate the performance of proposed method in section 3. We conclude with a brief summary in section 4.

2. Finite mixtures of linear mixed effects models

The finite mixtures of linear mixed effects models are supposed to have total K components with a linkage to predict the weights for each component, where the component is a linear mixed effects model and the linkage could be a concomitant model such as multinomial logistic model to make sure the sum of the component weights equals to 1. The superiority of this mixture mixed effects model comparing to the mixture model discussed in Khalili and Chen (2007) [23] is that random effects are also involved so that the variations of the observations from the same subject can be explained. Therefore, given that the group labels of each subject are constant, the finite mixtures of linear mixed effects models can capture the relationships between subjects in a group with the correlations between observations from a same subject by component while indicate the heterogeneity between (latent) groups on the response. The heterogeneity is represented by the differences between parameter values of each component (e.g., the coefficients of fixed effects between components). In practice, $K = 2$ is often used in many other mixture models [26, 27], for example, gender is often a good group indicator since there exists heterogeneity between males and females such as heights, hormone levels and so on. To treat the group labels as complete missing values for the EM framework of mixture models is very common. However, it is possible that the data may contain partial group labels (e.g., gender indicators of some subjects may be missing) which could be a useful prior information for getting more accurate component weights. The complete missing of group labels which discussed in Grün (2008) [22] is just a special case of the partial missing. In the rest of the section, we will set up our model based on the total number of components $K = 2$ with logistic model as the concomitant linkage by assuming partial grouping information is available. In addition, we assume that the heterogeneity between two groups is reflected by the difference of the parameter values of the fixed effects between two components only while random effects for both components cluster around a common mean. The EM framework for the ML estimation of this model will be discussed.

2.1 Model set up

Given a set of observed data $(y_{ij}, x_{ij}, z_{ij}, u_i)$, we consider the following mixture linear mixed effects model,

$$\begin{cases} y_{ij} = \beta_1^T x_{ij} + {\gamma_{i1}}^T z_{ij} + \epsilon_{ij1}, & w_i = 1, \text{(group 1)} \\ y_{ij} = \beta_2^T x_{ij} + {\gamma_{i2}}^T z_{ij} + \epsilon_{ij2}, & w_i = 0, \text{(group 2)} \\ \quad \text{logit}\,(P(w_i = 1)) = \alpha^T u_i, & \end{cases} \tag{1}$$

where $i = 1, \ldots, N$ indicates the number of subjects; $j = 1, \ldots, n_i$ is the number of observations for subject i; x_{ij} and z_{ij} are the fix and random effects covariates for subject i in jth observation respectively; $\epsilon_{ijk} \sim N(0, \sigma^2)$ is the error term for group (or component) $k = 1, 2$; β_k is fixed effects for group k and $\gamma_{ik} \sim N(0, G)$ is random effects for individual i in group k, where G is a positive-definite variance-covariance matrix. Moreover, w_i is the missing value which indicates the subject-specific group label and is related to a set of covariates u_i for subject i. In order to make sure w_i is unchanged all the time for ith subject, the covariates u_i for subject i in the logistic model should be fixed regardless of the observation j.

By assuming the partial grouping information is available, we actually have another parts of observed data $(y_{ij}^0, x_{ij}^0, z_{ij}^0, u_i^0, w_i^0)$ with given group labels, where $i = 1, \ldots, N_0$, $j = 1, \ldots, n_i^0$, and w_i^0 indicates the given group labels. Then the complete likelihood function will be derived in the following part.

2.2 Complete likelihood and EM framework

For observed data with group labels $(y_{ij}^0, x_{ij}^0, z_{ij}^0, u_i^0, w_i^0)$, denote $Y_i^0 = [y_{i1}^0, \ldots, y_{in_i^0}^0]^T$, $X_i^0 = [x_{i1}^0, \ldots, x_{in_i^0}^0]^T$ and $Z_i^0 = [z_{i1}^0, \ldots, z_{in_i^0}^0]^T$. Then it is obvious to get the log likelihood function

$$l_0 = \sum_{i=1}^{N_0} \log \left[\left(\frac{e^{\alpha^T u_i^0}}{1 + e^{\alpha^T u_i^0}} h_{i1}^0 \right)^{w_i^0} \left(\frac{1}{1 + e^{\alpha^T u_i^0}} h_{i2}^0 \right)^{1 - w_i^0} \right], \tag{2}$$

where h_{ik}^0 for $k = 1, 2$ represents the density for ith subject who is from group k with the following formula

$$\begin{aligned} h_{ik}^0 = & \left(2\pi\sigma^2\right)^{-\frac{n_i^0}{2}} \exp\left[-\frac{1}{2\sigma^2} \left(Y_i^0 - X_i^0 \beta_k - Z_i^0 \gamma_{ik}^0 \right)^T \left(Y_i^0 - X_i^0 \beta_k - Z_i^0 \gamma_{ik}^0 \right) \right] \\ & \times (2\pi |G|)^{-\frac{1}{2}} \exp\left[-\frac{1}{2} {\gamma_{ik}^0}^T G^{-1} \gamma_{ik}^0 \right]. \end{aligned} \tag{3}$$

Similarly, for the data without grouping information $(y_{ij}, x_{ij}, z_{ij}, u_i)$, denote $Y_i = [y_{i1}, \ldots, y_{in_i}]^T$, $X_i = [x_{i1}, \ldots, x_{in_i}]^T$ and $Z_i = [z_{i1}, \ldots, z_{in_i}]^T$. In addition, as mentioned above, we let the missing value w_i indicates the latent group labels. The complete log likelihood function shows as follows

$$l = \sum_{i=1}^{N} \log \left[\left(\frac{e^{\alpha^T u_i}}{1 + e^{\alpha^T u_i}} h_{i1} \right)^{w_i} \left(\frac{1}{1 + e^{\alpha^T u_i}} h_{i2} \right)^{1 - w_i} \right], \tag{4}$$

where

$$h_{ik} = \left(2\pi\sigma^2\right)^{-\frac{n_i}{2}} \exp\left[-\frac{1}{2\sigma^2}\left(Y_i - X_i\beta_k - Z_i\gamma_{ik}\right)^T\left(Y_i - X_i\beta_k - Z_i\gamma_{ik}\right)\right] \times \left(2\pi|G|\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\gamma_{ik}{}^T G^{-1}\gamma_{ik}\right]. \tag{5}$$

Therefore, the complete log likelihood for data with partial grouping information with the total number of groups (or components) $K = 2$ can be simply expressed as $l_c = l_0 + l$. Since partial group labels are missing, EM algorithm is a most common approach to solve the problem.

Let $-l_c$ denote the negative complete log likelihood, EM algorithm can optimize it through updating the parameters alternatively between expectation step (E-step) and maximization step (M-step) until convergence. For E-step, the missing group labels will be predicted by using available data and current parameter values via a Bayesian formula such that $p_i = \mathrm{E}(w_i) = \frac{e^{\alpha^T u_i} h_{i1}}{e^{\alpha^T u_i} h_{i1} + h_{i2}}$. Then, we can get the expected negative complete log likelihood by replacing w_i with p_i in $-l_c$. For M-step, the expected negative complete log likelihood can be optimized via two parts separately. The first part regarding the concomitant linkage function, we attempt to find $\hat{\alpha}$ which can minimize the following formula

$$\hat{\alpha} = \mathrm{argmin}_\alpha \left\{\sum_{i=1}^{N_0}\left[\log\left(1 + e^{\alpha^T u_i^0}\right) - \alpha^T u_i^0 w_i^0\right] + \sum_{i=1}^{N}\left[\log\left(1 + e^{\alpha^T u_i}\right) - \alpha^T u_i p_i\right]\right\}. \tag{6}$$

Newton-Raphson method is used to solve $\hat{\alpha}$ in formula (6) due to lack of analytical solution. The second part regarding the components of mixture, a set of parameters, including β_k, γ_{ik}, γ_{ik}^0, σ^2 and G, needs to be solved to make the following formula minimize,

$$\begin{aligned} &\frac{n_0 + n}{2}\log\sigma^2 + \frac{N_0 + N}{2}\log|G| \\ &+ \sum_{i=1}^{N_0}\sum_{k=1}^{2}\frac{I_k^0}{2}\left[\frac{1}{\sigma^2}\left(Y_i^0 - X_i^0\beta_k - Z_i^0\gamma_{ik}^0\right)^T\left(Y_i^0 - X_i^0\beta_k - Z_i^0\gamma_{ik}^0\right) + \gamma_{ik}^0{}^T G^{-1}\gamma_{ik}^0\right] \\ &+ \sum_{i=1}^{N}\sum_{k=1}^{2}\frac{I_k}{2}\left[\frac{1}{\sigma^2}\left(Y_i - X_i\beta_k - Z_i\gamma_{ik}\right)^T\left(Y_i - X_i\beta_k - Z_i\gamma_{ik}\right) + \gamma_{ik}{}^T G^{-1}\gamma_{ik}\right], \end{aligned} \tag{7}$$

where $n_0 = \sum_{i=1}^{N_0} n_i^0$ and $n = \sum_{i=1}^{N} n_i$ represent total observations of N_0 subjects with group labels and total observations of N subjects without group labels respectively; I_k^0 and I_k are the indicator of group labels for subjects given group labels and indicator of expected group labels for subjects without group labels such that

$$I_k^0 = \begin{cases} w_i^0 & k = 1 \\ 1 - w_i^0 & k = 2 \end{cases} \quad \text{and} \quad I_k = \begin{cases} p_i & k = 1 \\ 1 - p_i & k = 2 \end{cases}.$$

All parameters in formula (7) have closed form solutions by taking first derivative on one parameter given the rest of parameters are fixed and then making the equation equal to zero.

The solutions for each parameter are listed below:

$$\begin{aligned}
\hat{\beta}_k &= \text{argmin}_{\beta_k}\Bigg\{ \sum_{i=1}^{N_0} \left[\frac{\sqrt{I_k^0}}{\sigma}\left[\left(Y_i^0 - Z_i^0\gamma_{ik}^0\right) - X_i^0\beta_k\right]\right]^T \left[\frac{\sqrt{I_k^0}}{\sigma}\left[\left(Y_i^0 - Z_i^0\gamma_{ik}^0\right) - X_i^0\beta_k\right]\right] \\
&\quad + \sum_{i=1}^{N} \left[\frac{\sqrt{I_k}}{\sigma}\left[\left(Y_i - Z_i\gamma_{ik}\right) - X_i\beta_k\right]\right]^T \left[\frac{\sqrt{I_k}}{\sigma}\left[\left(Y_i - Z_i\gamma_{ik}\right) - X_i\beta_k\right]\right]\Bigg\}, \\
\hat{\sigma}^2 &= \frac{1}{n+n_0}\Bigg\{ \sum_{i=1}^{N_0}\sum_{k=1}^{2} I_k^0 \left(Y_i^0 - X_i^0\beta_k - Z_i^0\gamma_{ik}^0\right)^T \left(Y_i^0 - X_i^0\beta_k - Z_i^0\gamma_{ik}^0\right) \\
&\quad + \sum_{i=1}^{N}\sum_{k=1}^{2} I_k \left(Y_i - X_i\beta_k - Z_i\gamma_{ik}\right)^T \left(Y_i - X_i\beta_k - Z_i\gamma_{ik}\right)\Bigg\}, \\
\hat{G} &= \frac{1}{N+N_0}\left\{\sum_{i=1}^{N_0}\sum_{k=1}^{2} \frac{I_k^0}{2}\gamma_{ik}^0{\gamma_{ik}^0}^T + \sum_{i=1}^{N}\frac{I_k}{2}\gamma_{ik}{\gamma_{ik}}^T\right\}, \\
\hat{\gamma}_{ik} &= \left[\frac{I_k}{\sigma^2}Z_i^TZ_i + I_kG^{-1}\right]^{-1} \times \frac{I_k}{\sigma^2}Z_i^T\left(Y_i - X_i\beta_k\right), \\
\hat{\gamma}_{ik}^0 &= \left[\frac{I_k^0}{\sigma^2}{Z_i^0}^TZ_i^0 + I_k^0G^{-1}\right]^{-1} \times \frac{I_k^0}{\sigma^2}{Z_i^0}^T\left(Y_i^0 - X_i^0\beta_k\right).
\end{aligned} \tag{8}$$

The equation of $\hat{\beta}_k$ in formula (8) is formulated in the form of least squares. Several packages such as *ncvreg* [28] in the statistical software R can be applied to solve $\hat{\beta}_k$ directly depending on this form. Table 1 summerizes the computing algorithm for parameters updating using the proposed EM framework.

Table 1: Computing algorithm for parameters updating

1.	Initialize $\alpha^{(0)}$, $\sigma^{2(0)}$, $G^{(0)}$, $\beta_k^{(0)}$, $\gamma_{ik}^{(0)}$ and ${\gamma_{ik}^0}^{(0)}$ to get $p_i^{(0)}$.
2.	a) Given $p_i^{(t)}$, update α, σ^2, G, β_k, γ_{ik} and γ_{ik}^0 alternately in M-step.
	b) Given $\alpha^{(t)}$, $\sigma^{2(t)}$, $G^{(t)}$, $\beta_k^{(t)}$, $\gamma_{ik}^{(t)}$ and ${\gamma_{ik}^0}^{(t)}$, update $p_i^{(t+1)}$ in E-step.
3.	Repeat step 2 until convergence.

(0) parameter value at the beginning (i.e., initializer).
(t) parameter value at current level.

2.3 Variable selection using penalized approach

Based on the EM framework described in section 2.2, variable selection via a penalized approach will be introduced. A penalized method is used to enhance the predictive power of the model via producing sparse solutions [24]. We utilize SCAD penalty in this study due to some good properties such as continuity and unbiasedness [24, 29]. Variable selection is applied on both the logistic linkage (α) and fixed effects (β_k) simultaneously. For selecting informative variables for β_k $(k = 1, 2)$, the equation of $\hat{\beta}_k$ in formula (8) is replaced by a penalized version in the M-step:

$$\begin{aligned}
\hat{\beta}_k^* =&\text{argmin}_{\beta_k}\Bigg\{ \sum_{i=1}^{N_0} \left[\frac{\sqrt{I_k^0}}{\sigma}\left[\left(Y_i^0 - Z_i^0\gamma_{ik}^0\right) - X_i^0\beta_k\right]\right]^T \left[\frac{\sqrt{I_k^0}}{\sigma}\left[\left(Y_i^0 - Z_i^0\gamma_{ik}^0\right) - X_i^0\beta_k\right]\right] \\
&+ \sum_{i=1}^{N} \left[\frac{\sqrt{I_k}}{\sigma}\left[\left(Y_i - Z_i\gamma_{ik}\right) - X_i\beta_k\right]\right]^T \left[\frac{\sqrt{I_k}}{\sigma}\left[\left(Y_i - Z_i\gamma_{ik}\right) - X_i\beta_k\right]\right] + p_{\lambda_k}(\beta_k)\Bigg\},
\end{aligned} \tag{9}$$

where $p_{\lambda_k}(\beta_k)$ is the penalty with tuning parameter λ_k (i.e., λ_1 is for β_1 and λ_2 is for β_2). Thus, we can get the SCAD-penalized least squares estimators. However, for selecting informative α, formula (6) cannot be transformed into least squares directly due to lack of analytic solution. In order to apply the penalized approach, the second order Taylor expansion was studied to be a good solution in this case. Let the first and second order derivative of likelihood formula (6) be:

$$l'(\alpha) = \sum_{i=1}^{N_0}\left[\left(\frac{e^{\alpha^T u_i^0}}{1+e^{\alpha^T u_i^0}} - w_i^0\right)u_i^0\right] + \sum_{i=1}^{N}\left[\left(\frac{e^{\alpha^T u_i}}{1+e^{\alpha^T u_i}} - p_i\right)u_i\right], \tag{10}$$

and

$$l''(\alpha) = \sum_{i=1}^{N_0}\left[\frac{e^{\alpha^T u_i^0}}{(1+e^{\alpha^T u_i^0})^2}u_i^0 {u_i^0}^T\right] + \sum_{i=1}^{N}\left[\frac{e^{\alpha^T u_i}}{(1+e^{\alpha^T u_i})^2}u_i {u_i}^T\right]. \tag{11}$$

Then, given estimation $\alpha^{(s)}$, using the second order Taylor expansion formula, we have penalized likelihood as:

$$l(\alpha) \approx l(\alpha^{(s)}) + l'(\alpha^{(s)})(\alpha - \alpha^{(s)}) + \frac{1}{2}l''(\alpha^{(s)})(\alpha - \alpha^{(s)}) + p_{\lambda_3}(\alpha), \tag{12}$$

where $p_{\lambda_3}(\alpha)$ is the penalty for α. Thus, to minimize $l(\alpha)$ is equivalent to find

$$\hat{\alpha}^* = \text{argmin}_\alpha \left\{\left[l'(\alpha^{(s)})\alpha - \alpha^{(s)T} l''(\alpha^{(s)})\alpha\right] + \frac{1}{2}\alpha^T l''(\alpha^{(s)})\alpha + p_{\lambda_3}(\alpha)\right\}. \tag{13}$$

By taking formula (10) and (11) into formula (13), we can rewrite it into the least squares form and supersede formula (6) in M-step so that α can be iteratively updated by the regular penalized procedure as β_k.

The updating procedure for the rest of parameters is the same as the procedure in the EM framework proposed in the section 2.2. Total three penalties ($p_{\lambda_k}(\beta_k)_{k=1,2}$ and $p_{\lambda_3}(\alpha)$) are supposed to involve in the penalized framework. In this study, we set $\lambda_1 = \lambda_2 = \lambda_3$ to save the computational cost and BIC criteria will be introduced to find the optimal tuning parameter. The algorithm is summarized in Table 2.

Table 2: Computing algorithm for parameters updating (penalized)

1.	Assign a λ and initialize $\alpha^{(0)}$, $\sigma^{2(0)}$, $G^{(0)}$, $\beta_k^{(0)}$, $\gamma_{ik}^{(0)}$ and ${\gamma_{ik}^0}^{(0)}$ to get $p_i^{(0)}$.
2.	a) Given $p_i^{(t)}$, update α, σ^2, G, β_k, γ_{ik} and γ_{ik}^0 alternately in M-step.
	b) Given $\alpha^{(t)}$, $\sigma^{2(t)}$, $G^{(t)}$, $\beta_k^{(t)}$, $\gamma_{ik}^{(t)}$ and ${\gamma_{ik}^0}^{(t)}$, update $p_i^{(t+1)}$ in E-step.
3.	Repeat step 2 until convergence.
4.	Compute BIC.
5.	Repeat step 1 to 4 to find a λ that gives minimum BIC.

$^{(0)}$ parameter value at the beginning (i.e., initializer).
$^{(t)}$ parameter value at current level.

The implementation of our proposed approach in R is not difficult with current available R packages. The initializers listed in Table 1 and Table 2 ($\alpha^{(0)}$, $\sigma^{2(0)}$, $G^{(0)}$, $\beta_k^{(0)}$, $\gamma_{ik}^{(0)}$ and ${\gamma_{ik}^0}^{(0)}$) are achieved by using the data from the subjects with group indicator.

3. Simulations

In this section, we do some simulation studies based on our proposed method. We generated x_{ij}, z_{ij} and u_i independently from $U(-3, 3)$, the random effects γ_{ik}^0 and γ_{ik} from

$N(0, G)$, error terms ϵ_{ijk} from $N(0, 1)$ for $k = 1, 2$, and set $n_i = n_i^0 = D$, where D indicates that all the subjects have the same duplicates number.

Example 1: In the first example, we set

$$G = \begin{bmatrix} 0.15 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.45 \end{bmatrix},$$

which is a diagonal matrix. Then we set $\beta_1 = (1, 1, 1, 0, 0)$, $\beta_2 = (-1, -1, -1, 0, 0)$ and $\beta_1 = (0.3, 0.3, 0.3, 0, 0)$, $\beta_2 = (-0.3, -0.3, -0.3, 0, 0)$ separately with $\alpha = (1, -1, 0)$ or $(0.5, -0.5, 0)$ in addition to $D = 4$, the number of subjects with labels $N_0 = 100$ or 200, and the number of subjects without labels $N = 200$. Table 3 summarized parameter estimation results without penalization. It shows that the parameter estimation results based on the proposed EM framework are approximate to the true values. In addition, the estimation accuracy is affected by both the number of N_0 and the true effect size of the parameters. The lower number of N_0 with lower parameter effects can cause some estimation bias especially for α. However, it can be remitted by increase the number of N_0. For proposed penalized variable selection method, each scenario is duplicated for 50 times and the performance is summarized in Table 4.

Example 2: In the second example, we set

$$G = \begin{bmatrix} 0.15 & 0.05 & 0.05 \\ 0.05 & 0.3 & 0.05 \\ 0.05 & 0.05 & 0.45 \end{bmatrix},$$

which is a general covariance matrix. Other settings are the same as example 1 and the penalized variable selection performance is showed in Table 5.

In Table 4 and 5, ”AP” is the average number of predictors to be selected in the model. ”AIP” is the average number of informative predictors to be selected. ”ANP” is the average number of non-informative predictors to be selected. ”#C” is the number times of correctly selected. ”#O” is the number times of overestimate and ”#U” is the number times of underestimate. According to the results, the average number of informative variables should be selected is varied from 7.2 to 8.0 with average number of non-informative variables being selected is from 0.3 to 2.5 which implies that our method has a good power to identify the informative variables. Meanwhile, the computational time for each scenario is pretty reasonable. From both examples, the effect size of the variables and the number of subjects with labels (N_0) are factors which can impact the variable selection accuracy. Usually, the higher N_0 and higher effect size can provide more accurate results. In addition, the variance-covariance matrix of random effects, G, has some effects on detect the informative variables.

4. Conclusion

In this study, we constructed a mixture linear mixed effects model to deal with a situation that the group information is partially missing. The proposed variable selection method for this model is penalization likelihood based with SCAD penalty function. EM framework and BIC are efficiently incorporated in our method. Our proposed approach can do variable selection for both concomitant linkage and fixed effects in the mixture model simultaneously which is a novelty comparing to the published literature. Based on our simulation results, the introduced method can identify informative predictors efficiently. With larger size of subjects with group indicator, the model even has more power to detect the true informative predictors and sift out non-informative variables. Meanwhile, from the parameter estimation results, the estimation accuracy is also related to the size of the subjects with grouping information which cause estimation accuracy varies among applications. We also find that the assumption of random effects can also affect the selection results. In the literature, some studies such as Bondell and others [4] have studied the variable selection in both fixed and random effects in linear mixed effects model. In the future study, we may refer to their methods to incorporate variable selection for random effects in the mixture linear mixed effects model.

References

[1] HENDERSON, C.R. (1949) Estimation of changes in herd environment. *Journal of Dairy Science*, **32**, 706-715.

[2] HENDERSON, C.R. (1950) Estimation of genetic parameters. *Annals of Mathematical Statistics*, **21**, 309-310 (abstract).

[3] LAIRD, N. M. AND WARE, J. H. (1982) Random-effects models for longitudinal data. *Biometrics*, **38(4)**, 963-974.

[4] BONDELL, HOWARD D., KRISHNA, ARUN AND GHOSH, SUJIT K. (2012) Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics*, **66(4)**, 1069-1077.

[5] FAN, YINGYING AND LI, RUNZE. (2012) Variable selection in linear mixed effects models. *Annals of Statistics*, **40(4)**, 2043-2068.

[6] WU, PING, LUO, XINCHAO, XU, PEIRONG AND ZHU, LIXING. (2016) New variable selection for linear mixed-effects models. *Annals of the Institute of Statistical Mathematics*, 1-20.

[7] HENDERSON, C.R. AND OTHERS. (1959) The Estimation of environmental and genetic trends from records subject to culling. *Biometrics*, **15**, 192-218.

[8] HENDERSON, C.R. (1963) Selection index and expected genetic advance. *Statistical Genetics and Plant Breeding*, 141-163. NAS-NRC 982, Washington, DC.

[9] MCLEAN, R.A. AND OTHERS. (1991) A Unified Approach to Mixed Linear Models. *The American Statistician*, **45(1)**, 54-64.

[10] LINDSTROM, M.L. AND BATES, D.M. (1988) Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, **83 (404)**, 1014-1021.

[11] CHENG, HSIAO. (2014) Analysis of panel data. *Cambridge university press*.

[12] WARRINGTON N.M. AND OTHERS. (2014) Robustness of the linear mixed effects model to error distribution assumptions and the consequences for genome-wide association studies. *Statistical Applications in Genetics and Molecular Biology*, **13(5)**, 567-587.

[13] KOERNER, T.K. AND ZHANG, Y. (2017) Application of Linear Mixed-Effects Models in Human Neuroscience Research: A Comparison with Pearson Correlation in Two Auditory Electrophysiology Studies. *Brain Science*, **7(3)**, 26.

[14] MAGEZI, D.A. (2015) Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology*, **6(2)**.

[15] MA, ZONGWEI AND OTHERS. (2016) Satellite-derived high resolution PM2.5 concentrations in Yangtze River Delta Region of China using improved linear mixed effects model. *Atmospheric Environment*, **133**, 156-164.

[16] PAN, AY-WOAN AND OTHERS. (2012) A longitudinal study of the predictors of quality of life in patients with major depressive disorder utilizing a linear mixed effect model. *Psychiatry Research*, **198**, 412-419.

[17] AZUERO, ANDRES AND OTHERS. (2010) An Application of Longitudinal Analysis with Skewed Outcomes. *Nursing Research*, **59(4)**, 301-307.

[18] GIBBONS, ROBERT D. AND HEDEKER, DONALD. (2000) Applications of mixed-effects models in biostatistics. *The Indian Journal of Statistics, Series B*, **62(1)**, 70-103.

[19] DESMET L. AND OTHERS. (2014) Linear mixed-effects models for central statistical monitoring of multicenter clinical trials. *Statistics in Medicine*, **33**, 5265-5279.

[20] WULFSOHN, M. S., AND TSIATIS, A. A. (1997) A Joint Model for Survival and Longitudinal Data Measured with Error. *Biometrics*, **53**, 330-339.

[21] LENK, P. J., AND DESARBO, W. S. (2000) Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika*, **65(1)**, 93-119.

[22] GRÜN, B. (2008) Fitting finite mixtures of linear mixed models with the EM algorithm. *P. Brito, (red.), Compstat 2008-proceedings in computational statistics*, 165173.

[23] KHALILI, ABBAS AND CHEN, JIAHUA. (2007) Variables selection in finite mixture of regression models. *Journal of the American Statistical Association*, **102(479)**, 1025-1038.

[24] FAN, JIANQING AND LI, RUNZE. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, **96**, 1348-1360.

[25] SCHWARZ, GIDEON. (1978) Estimating the dimension of a model. *Annals of Statistics*, **6(2)**, 461-464.

[26] SKRONDAL, A. AND RABE-HESKETH, S. (2004) Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models. *CRC Press*.

[27] MCNICHOLAS PAUL, D. (2016) Mixture Model-Based Classification. *CRC Press*.

[28] BREHENY, P. AND HUANG, J. (2011) Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Annals of Applied Statistics*, **5**, 232-253.

[29] XIA, XIAOCHAO AND YANG, HU. (2016) Variable selection for partially time-varying coefficient error-in-variables models. *Statistics*, **50(2)**, 278-297.

[30] BATES, DOUGLAS AND OTHERS. (2016) Linear Mixed-Effects Models using 'Eigen' and S4. *CRAN*, Version 1.1-1.2.

Table 3: The parameter estimation results for the proposed mixture model for example 1 ($N = 200$ and $D = 4$).

N_0	α	β_1	β_2	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$
100	(-1,1,0)	(1,1,1,0,0)	(-1,-1,-1,0,0)	(0.81,-0.89,0.04)	(0.97,0.99,0.97,0.06,-0.13)	(-0.97,-1.01,-1.05,0.03,-0.06)
200	(-1,1,0)	(1,1,1,0,0)	(-1,-1,-1,0,0)	(1.11,-1.02,0.01)	(1.00,1.05,0.99,0.04,-0.01)	(-1.00,-0.98,-1.00,-0.06,-0.04)
100	(-1,1,0)	(0.3,0.3,0.3,0,0)	(-0.3,-0.3,-0.3,0,0)	(0.81,-0.60,0.13)	(0.31,0.26,0.36,0.01,-0.06)	(-0.31,-0.39,-0.42,-0.03,-0.01)
200	(-1,1,0)	(0.3,0.3,0.3,0,0)	(-0.3,-0.3,-0.3,0,0)	(0.86,-0.60,0.22)	(0.2,0.28,0.27,0.00,-0.08)	(-0.24,-0.37,-0.28,-0.01,0.00)
100	(-0.5,0.5,0)	(1,1,1,0,0)	(-1,-1,-1,0,0)	(0.59,-0.60,0.02)	(1.01,1.00,1.02,0.02,-0.04)	(-1.00,-0.97,-1.07,-0.01,0.06)
200	(-0.5,0.5,0)	(1,1,1,0,0)	(-1,-1,-1,0,0)	(0.49,-0.46,-0.03)	(0.99,0.98,1.05,-0.02,0.01)	(-1.06,-1.01,-1.03,-0.06,0.00)
100	(-0.5,0.5,0)	(0.3,0.3,0.3,0,0)	(-0.3,-0.3,-0.3,0,0)	(0.05,-0.91,-0.43)	(0.24,0.25,0.34,-0.06,0.03)	(-0.21,-0.36,-0.34,-0.01,-0.13)
200	(-0.5,0.5,0)	(0.3,0.3,0.3,0,0)	(-0.3,-0.3,-0.3,0,0)	(0.65,-0.33,0.06)	(0.26,0.31,0.32,-0.06,-0.02)	(-0.32,-0.41,-0.34,0.02,-0.02)

Table 4: The average performance of variable selection for the proposed mixture model for example 1 ($N = 200$ and $D = 4$).

N_0	α	β_1	β_2	AP	AIP	ANP	#C	#O	#U
100	(-1,1,0)	(1,1,1,0,0)	(-1,-1,-1,0,0)	9.10	8.00	1.10	30	20	0
200	(-1,1,0)	(1,1,1,0,0)	(-1,-1,-1,0,0)	8.68	8.00	0.68	41	9	0
100	(-1,1,0)	(0.3,0.3,0.3,0,0)	(-0.3,-0.3,-0.3,0,0)	9.00	7.88	1.12	17	27	6
200	(-1,1,0)	(0.3,0.3,0.3,0,0)	(-0.3,-0.3,-0.3,0,0)	9.22	8.00	1.22	27	23	0
100	(-0.5,0.5,0)	(1,1,1,0,0)	(-1,-1,-1,0,0)	10.50	8.00	2.50	9	41	0
200	(-0.5,0.5,0)	(1,1,1,0,0)	(-1,-1,-1,0,0)	9.70	8.00	1.70	23	27	0
100	(-0.5,0.5,0)	(0.3,0.3,0.3,0,0)	(-0.3,-0.3,-0.3,0,0)	8.40	7.20	1.20	7	25	18
200	(-0.5,0.5,0)	(0.3,0.3,0.3,0,0)	(-0.3,-0.3,-0.3,0,0)	9.10	7.70	1.40	26	19	5

Table 5: The average performance of variable selection for the proposed mixture model for example 2 ($N = 200$ and $D = 4$).

N_0	α	β_1	β_2	AP	AIP	ANP	#C	#O	#U
100	(-1,1,0)	(1,1,1,0,0)	(-1,-1,-1,0,0)	9.78	8.00	1.78	22	28	0
200	(-1,1,0)	(1,1,1,0,0)	(-1,-1,-1,0,0)	9.14	8.00	1.14	29	21	0
100	(-1,1,0)	(0.3,0.3,0.3,0,0)	(-0.3,-0.3,-0.3,0,0)	9.86	8.00	1.86	14	36	0
200	(-1,1,0)	(0.3,0.3,0.3,0,0)	(-0.3,-0.3,-0.3,0,0)	9.28	8.00	1.28	24	26	0
100	(-0.5,0.5,0)	(1,1,1,0,0)	(-1,-1,-1,0,0)	9.42	8.00	1.42	19	31	0
200	(-0.5,0.5,0)	(1,1,1,0,0)	(-1,-1,-1,0,0)	9.10	8.00	1.10	25	25	0
100	(-0.5,0.5,0)	(0.3,0.3,0.3,0,0)	(-0.3,-0.3,-0.3,0,0)	8.50	7.20	1.30	10	25	15
200	(-0.5,0.5,0)	(0.3,0.3,0.3,0,0)	(-0.3,-0.3,-0.3,0,0)	9.40	7.80	1.60	15	30	5