ABSTRACT

AHN, MIHYE. Random Effect Selection in Linear Mixed Models. (Under the direction of Dr. Hao Helen Zhang and Dr. Wenbin Lu.)

The selection of random effects in linear mixed models is an important yet challenging problem in practice. We propose a robust and unified framework for automatically selecting random effects and estimating the covariance components in linear mixed models. A moment-based loss function is constructed for the covariance parameters of random effects and then a sandwich nonnegative garrote penalty is imposed for the selection of random effects. Compared to typical approaches, the new estimator does not require any distribution assumptions on random effects or the error terms. Large sample theories show that the resulting estimator is consistent for both random effects selection and variance components estimation. Due to the nature of variance-covariance matrix of random effects, our procedure involves two challenging computation problems: nonlinear semidefinite programming and nonlinear programming with a linear inequality constraint. Computational strategies are suggested to tackle these issues as well as the model tuning. Furthermore, we propose a natural way to incorporate the selection of fixed effects for the procedure after choosing random effects. Simulation studies reveal that a better selection of random effects can yield efficiency gain for the estimation of fixed effects. We use numerous examples to illustrate the behaviors of the proposed approach and compare it with existing ones.

In addition, we propose two alternative methods for speeding up computation

and approximating to the original method. Since the objective function in the original method has up to fourth order terms and it makes the computation tedious, we suggest using a linear approximation to the penalized variance-covariance matrix. It reduces the objective function up to second order, and then the approximate function can be solved by the quadratic programming in some statistical softwares. We show that the approximate methods also perform well and often outperform the original method. Furthermore, the computation speed is remarkably improved in the approximate methods. We finally apply these approaches to a data set from the Amsterdam Growth and Health Study.

Random Effect Selection in Linear Mixed Models

by
Mihye Ahn

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2010

APPROVED BY:

Dr. Daowen Zhang

Dr. Howard Bondell

Dr. Wenbin Lu
Co-Chair of Advisory Committee

Dr. Hao Helen Zhang
Chair of Advisory Committee

DEDICATION

To my parents SeongIn Ahn and KyungSook Park.

BIOGRAPHY

Mihye Ahn was born in Incheon, South Korea on January 25, 1980. After graduating from Parkmoon girls' high school, she began her undergraduate work at Inha University in statistics. One year later, she decided to double major in business administration. When she was a junior, Inha University started to offer the financial analytics major which is a joint degree program from statistics and business departments. Mihye also applied to the joint program and finally earned three Bachelor's degrees in statistics, business administration, and financial analytics in February 2002. During her undergraduate studies, she worked as a vice-president of student bodies in the department of statistics and college of natural science for two years. Recognized leadership and outstanding academic performance, she received the Presidential Award from Kim Dae-Jung who was ex-president of Korea and the Nobel Peace Prize laureate.

Mihye continued her graduate work at the same school. She completed her Master of Science degree in statistics in February 2004 and wrote her thesis 'Identification of special causes in dynamic process adjustment with dead time' under the direction of Dr. Jae June Lee. Since 2005, she has continued her studies in statistics towards a Ph.D. degree at North Carolina State University.

ACKNOWLEDGEMENTS

First and foremost I would like to express my deepest gratitude to my advisors, Dr. Hao Helen Zhang and Dr. Wenbin Lu, for their kind guidance and continuous support during preparing this dissertation. Whenever I encountered hard problems, they gave me prompt advice and encouraged me to continue this research. Words cannot express how lucky I am to have worked with them.

I have to say my sincere thanks to committee members: Dr. Howard Bondell and Dr. Daowen Zhang for taking their precious time to be on my committee and providing helpful suggestions. Dr. Tao Pang in mathematics department for attending my oral preliminary exam and final oral exam as a graduate school representative.

I deeply appreciate Dr. Jae June Lee, my advisor for my master's thesis at Inha University. He convinced me to pursue a doctoral degree in the united states. Without his help, I could not have continued my graduate career. I thank Drs. Hongsuk Jorn, Jean Kyung Kim, Jinsoo Hwang, Heon Jin Park, and Jinho Park for introducing the joy of statistics during my time at Inha University.

I would like to thank statistics department at NCSU for providing financial support during my graduate studies. I could not have completed this work without their support.

Thanks my friends and colleagues. I regrettably cannot acknowledge individually by name here, but I would not forget wonderful days with them.

Last but not least, I am very grateful to my parents, my sister and brother for their endless love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Background

Over the years, variable selection methods have received much attention and been applied to various fields. Using uninformative variables not only wastes money or time but also reduce estimation efficiency or prediction accuracy. Selecting an appropriate set of important variables helps to reduce the variances of parameter estimates. By deleting some noise variables, we might improve the precision of the estimates. Moreover, a simpler model can enhance model interpretation, and are therefore helpful for understanding the underlying regression relationship.

When we have p variables, the total number of possible models is 2^p. With a large number of variables, identifying the optimal model within the large model space can be computationally burdensome. Sequential procedures, such as forward selection and backward elimination, can be used for variable selection with relatively less computation. However, these selection procedures are discrete in that one variable is either added or deleted at a time, and hence provide unstable results. (Breiman, 1995)

For continuous selection process, penalized likelihood methods have been developed recently, including nonnegative garrote (Breiman, 1995), LASSO Tibshirani (1996), Fan & Li (2001), Zou & Hastie (2005), Adaptive LASSO (Zou, 2006; Zhang & Lu, 2007; Wang et al., 2007), group LASSO (Yuan & Lin, 2006), LSA (Wang & Leng, 2007), and OSCAR (Bondell & Reich, 2008). In the following Section 1.3, we will review these methods which are designed for selecting fixed effects. Many model selection criteria have been proposed, including the Akaike information criterion (AIC) (Akaike, 1973) and Bayesian information criterion (BIC) (Schwarz, 1978). In the following, we present a detailed description of several selection criteria.

In many simple settings, we assume that the observations are independent and identically distributed. However, in more complex settings where data are repeatedly measured on a subject or taken over time on the same unit, the observations within each subject might be no longer independent. To capture the variation within each subject, Laird & Ware (1982) developed a linear mixed model containing both fixed effects and random effects. In practice, this model is useful for modeling longitudinal data, spatial data, panel data, or clustered observations.

Assume that the number of subjects is m and the number of measurements on subject i is n_i. A general linear mixed model can be written as

$$y_{ij} = X_{ij}^{\mathrm{T}}\boldsymbol{\beta} + Z_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_i + \varepsilon_{ij} \qquad (i = 1, 2, \ldots, m; j = 1, 2, \ldots, n_i), \tag{1.1}$$

where y_{ij} is the response, $X_{ij} = (X_{ij1}, \ldots, X_{ijp})^{\mathrm{T}}$ are the fixed-effect covariates for the jth observation of the ith subject, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is the $p \times 1$ vector of fixed-effect coefficients, $Z_{ij} = (Z_{ij1}, \ldots, Z_{ijq})^{\mathrm{T}}$ are the random-effect covariates for the jth observation of the ith subject, $\boldsymbol{\gamma}_i = (\gamma_{i1}, \ldots, \gamma_{iq})^{\mathrm{T}}$ is the subject-specific $q \times 1$ vector of random-effect coefficients, and ε_{ij}'s are the error terms. Furthermore, we assume that $\boldsymbol{\gamma}_i$ have mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma} = [\sigma_{jk}]$ with $1 \leq j, k \leq q$, the error $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})^{\mathrm{T}}$ have mean $\mathbf{0}$ and variance-covariance matrix $\sigma_\varepsilon^2 \mathbf{I}_{n_i}$, and $\boldsymbol{\varepsilon}_i$'s are independent of the $\boldsymbol{\gamma}_i$'s.

An important problem in applying (1.1) is how to choose important fixed and random effects. Here, important fixed effects refer to those with nonzero coefficients, and important random effects refer to random effects whose coefficients truly vary among subjects. The selection of random effects plays a crucial role in model estimation and inference for (1.1). If some important random effects are missing from the model, the covariance structure would be underfitted. On the other hand, if too many unimportant random effects are included in the model, the covariance matrix of random effects could be singular and cause numerical unstability of the estimates. Furthermore, choosing an appropriate set of random effects would capture the data variability more effectively, lead to efficiency gain in the fixed effect estimation, and eventually improve prediction accuracy for future data.

In the context of mixed-effects models, variables selection methods need to be modified to identify nonzero random effects. However, the selection method for random effects is challenging due to the nature of variance-covariance matrix of random effects. Unlike fixed effects selection, there exist the only few methods for selecting random effects (See Chen & Dunson (2003), Cai & Dunson (2006), Kinney & Dunson

(2007), Jiang et al. (2008), Bondell et al. (2010), and Ibrahim et al. (2010)). These approaches assume that the random effects and the error terms follow a normal distribution. In addition, there have been efforts to modify selection criteria in order to be used properly in mixed effect models (Wolfinger (1993), Diggle et al. (1994), Vaida & Blanchard (2005), Pu & Niu (2006), and Ibrahim et al. (2008)). In Section 1.4, we will review these existing methods and selection criteria in detail.

In this chapter, we first review traditional methods and selection criteria for comparing and assessing candidate models, which were originally developed for linear models. We then describe penalized regression, such as Lasso and SCAD, which has been recently developed and is a continuous shrinkage method. In linear mixed models, we motivate the need for random effects selection. We review existing methods for random effects selection.

1.2 Classical Methods for Fixed Effects Selection

Consider a linear regression model

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$ is an $N \times 1$ vector of responses, $\mathbf{X}$ is an $N \times p$ design matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients, and $\boldsymbol{\varepsilon}$ is an $N \times 1$ vector of random errors with $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_N)$. In general, the method of least squares is used to estimate the regression coefficients from the data.

In practice, various model selection criteria have been proposed to compare candidate models and to select the best model. Different criteria have different motivations and perform better for some problems in practice. We give a brief review on those widely used criteria in this following. Hocking (1976) extensively reviewed selection criteria with some examples.

1.2.1 Selection Criteria

Residual Mean Square: MSE

The residual mean square is defined as

$$MSE = \frac{\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2}{N-p} = \frac{RSS}{N-p},$$

where p is the number of variables in the fitted model, RSS is the residual sum of squares and $\widehat{y}_i$ is the fitted value of y_i. This is widely used to evaluate how well the model is fitted to the data. We prefer the candidate model with the minimum MSE. For small data sets, the MSE might not work effectively.

Coefficient of Determination: R^2

The coefficient of determination has been widely used as a measure of the capability of the model to fit the data, and defined as

$$R^2 = \frac{\sum\limits_{i=1}^{N}(\widehat{y}_i - \bar{y})^2}{\sum\limits_{i=1}^{N}(y_i - \bar{y})^2} = \frac{SSR}{SST} = 1 - \frac{RSS}{SST},$$

where $\bar{y}$ is the overall mean of y, $SSR = \sum_{i=1}^{N}(\widehat{y}_i - \bar{y})^2$ is the regression sum of squares and $SST = \sum_{i=1}^{N}(y_i - \bar{y})^2$ is the total sum of squares. This can be viewed as the ratio the explained variance to the total variance. As the number of parameters used in the model increases, R^2 increases. Therefore, R^2 achieves the maximum when all variables enter in the model. Based on R^2, we select the candidate model having the largest R^2. As a result, the chosen model might be overfitted.

Adjusted R^2

The drawback of R^2 leads to the modification of R^2. The adjusted R^2 is defined as

$$R^2_{adj} = 1 - \frac{RSS/(N-p)}{SST/(N-1)} = 1 - \frac{(N-1)MSE}{SST} = 1 - \Big(\frac{N-1}{N-p}\Big)(1 - R^2). \qquad (1.2)$$

The R^2_{adj} penalizes bigger models. As seen in (1.2), the minimum MSE and the maximum R^2_{adj} yield the same model selection. That is, comparing models in terms of MSE is identical to that in terms of R^2_{adj}.

Mallows' C_p

The statistic C_p, proposed by Mallows (1973), is defined as

$$C_p = \frac{RSS}{\widehat{\sigma}^2} - N + 2p,$$

where $\widehat{\sigma}^2$ is the residual mean squares in the full model. The C_p was motivated as an unbiased estimate of prediction accuracy of the candidate model. If the model with p variables is proper, $E(C_p)$ is approximately equal to p. Therefore, we find points close to the $C_p = p$ line on the plot of C_p versus p. Also, it might be good to select points below the $C_p = p$ line due to random variation. As a result, we prefer choosing the candidate model with small C_p value about equal to p. Generally, many statistical software packages select the model having the smallest C_p. Mallows (1995) studied the property of a C_p plot when p is large and there exist many weak effects. Some modified versions of Mallows' C_p are described with some examples in Miller (2002).

Information Criteria

Akaike Information Criterion (AIC) is originally proposed by Akaike (1973) to consider the number of parameters as a standard comparing the candidate models. His idea is to impose a penalty for model complexity to the log likelihood. In general, the AIC is defined as

$$\text{AIC} = -2\log(likelihood) + 2p.$$

Hurvich & Tsai (1989) showed that AIC brings about overfitting in the small sample, and suggested using AIC_C, a corrected version of AIC,

$$AIC_C = AIC + \frac{2(p+1)(p+2)}{N-p-2}.$$

For several variants of AIC, see McQuarrie & Tsai (1998). Another information criterion is the Bayesian Information Criterion (BIC), proposed by Schwarz (1978),

$$\text{BIC} = -2\log(likelihood) + p\log N.$$

BIC is motivated in the Bayesian approach to model selection. Schwarz (1978) made an appropriate modification of maximum likelihood using the asymptotic behavior of Bayes estimators. We desire the model with smaller AIC or BIC. Miller (2002) stated that using AIC tends to choose a little larger models than using Mallows' C_p. Information criteria and C_p statistic consider the trade-off between σ^2 and p. We cannot say which criterion is better than the others. However, we can consider the behavior of these criteria as follows. When $N > e^2$, BIC penalizes larger models more heavily, and hence it prefers simpler models. Moreover, BIC is asymptotically consistent for model selection. That is, the probability that BIC yields the correct model approaches 1 as $N \to \infty$. Contrary to BIC, AIC tends to select more complex models as $N \to \infty$. BIC also has disadvantages; BIC often chooses too simple models for finite samples. Hurvich & Tsai (1989) showed that BIC may poorly perform in small samples. Haughton (1988) showed that BIC is consistent when the true model is fixed. If the dimensionality of the true model increases with N, AIC is also consistent (Shibata, 1981).

PRESS

Allen (1981) proposed the prediction sum of squares (PRESS) which is defined as

$$PRESS = \sum_{i=1}^{N} (y_i - \widehat{y}_{(i)})^2 = \sum_{i=1}^{N} e_{(i)}^2,$$

where $\widehat{y}_{(i)}$ denotes the predicted value of the ith response when the model is fitted without using the ith observation. PRESS provides detailed information about the stability of the candidate models. However, PRESS requires an excessively greedy computation. Breiman & Spector (1992) showed that non-resampling estimates including PRESS statistic lead to inaccurate estimates of the mean squared error of prediction. To overcome this problem, they used cross-validation and bootstrap methods.

K-fold Cross-Validation: CV

Cross-validation is a method that uses part of data to fit the model and the rest part to test the performance of the fitted model. We first divide the data into K parts randomly. For $k = 1, \cdots, K$, two sets $\mathfrak{D}_1$ and $\mathfrak{D}_2$ consist of the kth part and the remaining parts, respectively. We fit the model to $\mathfrak{D}_2$, and compute the prediction error of the fitted model with $\mathfrak{D}_1$. When K is equal to N, this method is called 'leave-one-out cross-validation (LOOCV)'. In fact, PRESS statistic uses the LOOCV method. LOOCV has low bias but can have high variance. Moreover, the computational burden can be excessive. Breiman & Spector (1992) showed that fivefold cross-validation is better than leave-one-out cross-validation in model selection.

Generalized Cross-Validation: GCV

Craven & Wahba (1979) proposed the generalized cross-validation (GCV) as a computational shortcut for LOOCV. The GCV approximation is defined by

$$GCV = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{y_i - \widehat{y}_i}{1 - \mathrm{tr}(\mathbf{S})/N} \right)^2,$$

where $\widehat{\boldsymbol{y}} = \mathbf{S}\boldsymbol{y}$ and $\mathrm{tr}(\mathbf{S})$ is the effective number of parameters. For linear models, we have

$$GCV = \frac{RSS}{N} \frac{1}{(1 - p/N)^2}. \tag{1.3}$$

By Taylor expansion,

$$GCV \approx \frac{RSS}{N} + 2\widehat{\sigma}^2 \frac{p}{N}.$$

Since $\frac{RSS}{N} \to \sigma^2$ as $N \to \infty$, GCV yields the same result as AIC and Mallow's C_p asymptotically. Various GCV-style statistics have been proposed, and they will be described in Section 2.3.2.

1.2.2 Computational Methods

If we have p variables, there exist 2^p candidate models. As the number of variables increases, the number of computations needed rapidly increases. For efficient and effective computation, many algorithms have been proposed. In this section, we will go over traditional procedures that are in common use.

All Possible Regressions

All possible regressions are in fact to compare 2^p candidate models. However, it requires considerably greedy computations. Furnival & Wilson (1974) proposed the leaps and bounds algorithm to perform all possible regression efficiently. They employed the lexicographic algorithm and also performed an exhaustive search. The algorithm is quite useful in linear models with $p < 40$. The idea of the algorithm is to use information obtained from previous steps. As a result, we can reduce the computational burden. Their algorithm offers the best m models of each size, where

m is set by the user. They provided the Fortran subroutine which is available in many statistical softwares. When we find the best subset by the leaps and bound algorithm, C_p, R^2, and R^2_{adj}, described in Section 1.2.1, are available as a criterion for comparing candidate models. The *best subset selection* is to choose the best one among all possible subsets. It tends to result in a model with too many variables, and the final model would be very unstable.

Forward Selection

For forward selection, the procedure starts with no variables in the model. First, for all variables not included in the model we check which variable has the largest partial F-statistic. If the partial F-statistic is greater than a pre-determined F value, the variable is added to the model. The pre-determined F value is often called 'F-to-enter'. The above procedure is continued until new variable cannot be added to the model any more. Roecker (1991) showed that forward selection can provide slightly smaller prediction error and less bias compared to all possible regressions.

Backward Elimination

Backward elimination is the simplest procedure for variable selection and works in the opposite direction of forward selection. At first, the procedure begins with all variables in the model. We then compute the partial F-statistic for each variable out of the model. If the smallest partial F-statistic is less than a pre-determined F value, the variable is excluded from the model. This pre-determined F value is sometimes called 'F-to-remove'. The backward elimination also stops when the partial F-statistics for variables not belonging to the model are all greater than F-to-remove. Forward selection and backward elimination are more economical than the all possible regressions. Since we start with all variables in the model, backward elimination can be performed only when $p < N$.

Stepwise Regression

The stepwise regression can be thought as a combination of forward selection and backward elimination. At each step, one variable may be either entered or removed. Therefore, the same variable can be again added to the model after excluded. Note that this procedure allows the move of only one variable at one step.

These methods described above are easy to understand and perform, but the selection results are unstable. Because the selection procedure is discrete; that is, variables are either remained or dropped from the model, even small changes in the data might lead to quite different results for variable selection. This can also result in worse prediction accuracy. In next section, we review penalized approaches which are continuous selection procedures.

1.3 Penalized Regression for Fixed Effects

The main idea of penalized approaches is to impose a certain type of penalty to the regression coefficients, such that they are shrunk towards zeros, and some small coefficients will become exactly zero, achieving the purpose of variable selection. As a continuous variable selection procedure, the penalized method gives better prediction and small variance than traditional searching methods when the model is properly tuned. In general, the penalized likelihood function is of the form

$$-2\log(likelihood) + P_\lambda(\beta),$$

where $P_\lambda(\beta)$ is the penalty and $\lambda \geq 0$. As λ increases, the penalty term also increases and and impose more shrinkage on the coefficients. We can select the proper value of λ by using the variable selection criteria described in Section 1.2.1.

1.3.1 Ridge Regression

When multicollinearity is present, the ordinary least squares estimates might behave very poor. The variances of the estimates are considerably inflated, and the estimation is quite unstable. Hoerl & Kennard (1970b,a) introduced ridge regression which produces a little biased estimator of regression coefficients. They allow a small bias to achieve a smaller variance. The ridge estimate minimizes the penalized residual sum of squares with L2 penalty term:

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} \big(y_i - \sum_{j=1}^{p} \beta_j x_{ij}\big)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}, \tag{1.4}$$

where $\lambda \geq 0$ is often referred to as a shrinkage parameter and $\mathbf{X}$ is in standardized form. We assume $\boldsymbol{y}$ is centered, so the intercept can be omitted from the model. The ridge estimator can be viewed as a slight modification of the normal equations and is expressed as $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{y}$. The ridge estimator shrinks the ordinary least squares estimator toward the origin. In addition to the ridge estimator, other biased estimators that will be described later are also called *shrinkage estimators*. Even though the ridge regression is more stable process, it is not appropriate for variable selection because it does not make any of coefficients zero. For variable selection, many approaches have considered the use of different penalties which make coefficients shrink toward zero.

1.3.2 Nonnegative Garrote

Breiman (1995) introduced the *nonnegative garrote* that is used to do subset regression. Suppose that $\mathbf{X}$ is standardized and $\boldsymbol{y}$ has mean zero. The nonnegative garrote minimizes

$$\sum_{i=1}^{N}\Big(y_i - \sum_{j=1}^{p} c_j \widetilde{\beta}_j x_{ij}\Big)^2 \quad \text{subject to} \;\; c_j \geq 0, \;\; \sum_{j=1}^{p} c_j \leq t, \tag{1.5}$$

where $t \geq 0$ is a tuning parameter and $\widetilde{\beta}_j$'s are the ordinary least squares estimates. As t decreases, the garrote becomes tighter. As a result, many c_j's are set to zero and the rest of coefficients are shrunken. Breiman (1995) stated that the nonnegative garrote is stabler than subset selection and is scale invariant, while ridge regression is not scale invariant. Similar to ridge regression, the nonnegative garrote has smaller prediction error. Moreover, the nonnegative garrote is consistent in variable selection (Zou, 2006). The nonnegative garrote's drawback is that the estimates depend on the ordinary least squares estimates. Due to this property, the nonnegative garrote may behave poorly when variables are highly correlated or the model using the ordinary least squares estimates is overfitted. In addition, the nonnegative garrote results in a more complex model than subset regression in general. Miller (2002) showed that the nonnegative garrote gives similar results to backward elimination with a real data.

1.3.3 LASSO

Tibshirani (1996) proposed a new regression method for 'least absolute shrinkage and selection operator', called the *lasso*. It also shrinks some coefficients to zeros. The Lasso estimate is defined by

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} \Big(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \Big)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^{p} |\beta_j| \leq t, \tag{1.6}$$

where $t \geq 0$ is a tuning parameter.

Leng et al. (2004) have shown that the Lasso solution is not consistent in variable selection. Zou (2006) studied a necessary condition so that the Lasso variable selection is consistent. Zhao & Yu (2006) also provided an almost necessary and sufficient condition for consistent Lasso solution. In other words, the Lasso is not always consistent. Besides, the Lasso might shrink coefficients more than expected, and this leads to asymptotically biased estimates.

Furthermore, the Lasso cannot choose more predictors than N. When there is a group of variables and their correlations are quite high, the Lasso does not identify the group. When predictors are highly correlated, the Lasso has even the worse performance than ridge regression.

1.3.4 SCAD

Even though the Lasso provides both variable selection and shrinkage, it might shrink large coefficients more than is expected. To overcome this drawback, Fan & Li (2001) proposed the 'smoothly clipped absolute deviation (SCAD)' penalty which imposes different penalties to regression coefficients depending on their magnitudes. That is, the idea is to penalize uninformative variables heavily and informative variables lightly. The SCAD estimates can be expressed as

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} \Big(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \Big)^2 + \sum_{j=1}^{p} q_\lambda(|\beta_j|) \right\},$$

where q_λ is a penalty function. For more detailed description, see Fan & Li (2001). They also introduced the concept of the 'oracle properties'. Assume that $\mathcal{A}_0$ and $\mathcal{A}_1$ which are lists of indice of zero and nonzero coefficients, respectively. Then, the oracle procedure satisfies the following properties.

1. Selection consistency: Identify $\mathcal{A}_1$ correctly; *i.e.* $\widehat{\beta}_j = 0$ for all $j \in \mathcal{A}_0$ and $\widehat{\beta}_j \neq 0$ for all $j \in \mathcal{A}_1$

2. Optimal estimation: $\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\mathcal{A}_1} - \boldsymbol{\beta}^*_{\mathcal{A}_1}) \to N(\mathbf{0}, \boldsymbol{\Sigma}^*)$ in distribution, where $\boldsymbol{\beta}^*$ is the true values of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}^*$ is the covariance matrix when assuming we know the true model in advance.

Fan and Li (2001) showed that the Lasso does not satisfy the oracle properties. In contrast, the SCAD enjoys the oracle properties. However, the SCAD is computationally difficult to implement since its penalty is not a convex function and not differentiable at zero.

1.3.5 Elastic Net

To resolve the drawbacks of the Lasso, Zou & Hastie (2005) introduced an elastic net penalty which is a mixture of L1 and L2 penalties. The elastic net estimate is

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} \Big(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \Big)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right\}.$$

Contrary to the Lasso, the elastic net performs well when $N < p$. In addition, the elastic net can identify groups of variables and often outperform the Lasso.

1.3.6 Adaptive Lasso

Zou (2006) proposed the 'adaptive lasso' which is a modified version of the Lasso. He derived a necessary condition for the inconsistency of Lasso solution. He also showed that the nonnegative garrote is consistent in variable selection. Assume that $\widetilde{\boldsymbol{\beta}}$ is a $\sqrt{N}$ consistent estimator. The adaptive Lasso estimate is defined by

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} \Big(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \Big)^2 + \lambda \sum_{j=1}^{p} \widehat{\omega}_j |\beta_j| \right\},$$

where $\lambda \geq 0$, $\widehat{\omega}_j = 1/|\widetilde{\beta}_j|^\gamma$ is a data-dependent weight, and γ is a fixed constant. This imposes different weights to each of coefficients.

The adaptive Lasso satisfies the selection consistency and enjoys the oracle properties. In addition, the adaptive Lasso is computationally easier than the SCAD because the penalty of the adaptive Lasso is a convex function. That is, the adaptive Lasso overcomes the drawbacks of the Lasso and the SCAD.

1.3.7 OSCAR

Bondell & Reich (2008) proposed the 'Octagonal Shrinkage and Clustering Algorithm for Regression'(OSCAR). The OSCAR estimate is given by

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} \Big(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \Big)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j<k} \max\big(|\beta_j|, |\beta_k|\big) \right\}.$$

When there exist some groups of correlated variables, it might be required to identify the group because of the same influence to the response variable. That is, all variables belonging to the same group are either chosen or removed at the same time. Bondell & Reich (2008) showed that the OSCAR has good performance when there are high correlations between variables.

1.4 Variable Selection in Linear Mixed Models

In Section 1.3, we reviewed a variety of penalized approaches which have been recently developed. All the methods mentioned above can be directly applied to the selection of fixed effects in model (1.1). However, the selection of random effects has received little attention in the literature. Since it is not straightforward to estimate the variance-covariance matrix due to the implicit form of random effects, the selection of random effects is much more challenging than selecting fixed effects.

1.4.1 Selection Criteria

Vaida & Blanchard (2005) showed the general definition of AIC is not appropriate in mixed effects model. They proposed the conditional AIC, cAIC. There are several versions of cAIC; using either maximum likelihood or restricted maximum likelihood estimation, and the asymptotic version and finite-sample corrected version. The general form of cAIC is given by

$$cAIC = -2\log g\big(\boldsymbol{y}|\widehat{\boldsymbol{\beta}}(\boldsymbol{y}), \boldsymbol{\gamma}(\boldsymbol{y})\big) + 2\rho,$$

where g is the approximating model for fitting the data and ρ is the effective degrees of freedom. Pu & Niu (2006) extended the generalized information criterion (GIC), proposed by Rao & Wu (1989), which is a generalization of AIC and BIC. There are two stages for selecting both fixed and random effects: first, only fixed effects are selected and then random effects are selected. Ibrahim et al. (2008) used IC_Q criterion given by

$$IC_Q(\lambda) = -2Q(\widehat{\theta}_\lambda|\widehat{\theta}_0) + c_n(\widehat{\theta}_\lambda),$$

where Q is the penalized likelihood function , $\widehat{\theta}_0$ is the unpenalized maximum likelihood estimate, $c_n(\theta)$ is a function of the data and the fitted model.

1.4.2 Penalized Approaches

Chen & Dunson (2003) proposed a Bayesian approach for selecting random effects, using a modified Cholesky decomposition to reparameterize the linear mixed effects model:

$$\text{var}(\boldsymbol{\gamma}_i) = \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Lambda},$$

where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_q)$ is a nonnegative $q \times q$ diagonal matrix and $\boldsymbol{\Gamma}$ is a $q \times q$ triangular matrix with 1's in the diagonal entries. If $\lambda_l = 0$, the lth column and row of $\boldsymbol{\Sigma}$ are all zero. This means that the lth random effect is not important. They select a prior with point mass at zero for the random effects variances. The posterior is computed by Markov chain Monte Carlo algorithm. However, they did not provide any theoretical properties of their estimator. Cai & Dunson (2006) and Kinney & Dunson (2007) extended this approach to the generalized linear mixed model and the logistic mixed effects models, respectively.

Recently, Bondell et al. (2010) proposed a likelihood-based method which jointly selects both fixed and random effects. They also adopted the modified Cholesky decomposition instead of using innate covariance matrix. After reparameterization, the penalized log-likelihood function subject to an L1 penalty with adaptive weights is minimized via the constrained EM algorithm to obtain the estimates of both the coefficients of fixed effects and the variances of random effects. They also showed the proposed estimator enjoys the Oracle properties. Ibrahim et al. (2010) proposed to minimize the penalized likelihood with the SCAD and adaptive Lasso penalties. For selecting the proper values of tuning parameters, they used IC_Q statistic proposed by Ibrahim et al. (2008). They showed that the resulting estimator has desirable properties; consistency, sparsity properties, and asymptotic normality.

All of these approaches require the normality assumption for the random effects and the error terms, and therefore the validity of their inferences heavily depends on the distribution assumption. This motivates us to develop a more robust and flexible approach, without assuming any distributions, for random effects selection in linear mixed effects model. Different from existing methods, we propose to first construct a moment-based loss function for variance components and then achieve random-effect selection by minimizing the penalized loss. Therefore, our estimator is robust against non-normality of data, and its inference does not rely on any distributional assumption on the random effects or the errors. We further prove that the new estimator has good theoretical properties such as selection consistency, root-m consistency, and asymptotic normality. The estimator is also easy to implement and enjoys computational advantages over some competitive methods. To more speed up the computation, we also propose two alternative methods which are approximate to the original method and computationally efficient. The step for selecting fixed effects can be easily added to our algorithms by applying adaptive lasso penalty for the fixed effects coefficient. To be appropriate for the moment-based method, we suggest modifying selection criteria such as GCV and BIC used in general. We examine the performance of selection criteria and the proposed methods via simulation studies.

Chapter 2

Random Effects Selection

2.1 New Methodology

Alternatively but equivalently, in matrix form, model (1.1) can be written as

$$\boldsymbol{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i \qquad (i = 1, 2, \ldots, m), \tag{2.1}$$

where $\boldsymbol{y}_i$ is the $n_i \times 1$ response vector for observations of subject i, $\mathbf{X}_i$ is the $n_i \times p$ design matrix for the fixed effects, and $\mathbf{Z}_i$ is the $n_i \times q$ design matrix for the random effects, and $\boldsymbol{\varepsilon}_i$ is the $n_i \times 1$ vector of errors for observations of subject i. Note that $\mathrm{var}(\boldsymbol{y}_i) = \sigma_\varepsilon^2\mathbf{I}_{n_i} + \mathbf{Z}_i\boldsymbol{\Sigma}\mathbf{Z}_i^{\mathrm{T}}$ for $i = 1, \ldots, m$, which naturally incorporates heterogeneity among subjects. A large body of literature to estimate the parameters in (2.1) are on maximum likelihood (ML) and restricted maximum likelihood (REML) estimations by assuming that $\boldsymbol{\gamma}_i$'s and $\boldsymbol{\varepsilon}_i$'s are all normally distributed. See Laird & Ware (1982), Jennrich & Schluchter (1986), and Lindstrom & Bates (1988). In this dissertation, we develop an alternative moment-based approach to estimating the model parameters, which does not require any specification on the distributions of the random effects and errors.

Denote the total number of observations by $N = \sum_{i=1}^{m} n_i$. To further facilitate the following presentation, we now express (2.1) in a compressed form

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y} = (\boldsymbol{y}_1^{\mathrm{T}}, \ldots, \boldsymbol{y}_m^{\mathrm{T}})^{\mathrm{T}}$ is a $N \times 1$ vector, $\mathbf{X} = (\mathbf{X}_1^{\mathrm{T}}, \ldots, \mathbf{X}_m^{\mathrm{T}})^{\mathrm{T}}$ is a $N \times p$ matrix, $\mathbf{Z} = \mathrm{diag}(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)$ is a $N \times mq$ block diagonal matrix, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^{\mathrm{T}}, \ldots, \boldsymbol{\gamma}_m^{\mathrm{T}})^{\mathrm{T}}$ is a $mq \times 1$ vector, and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^{\mathrm{T}}, \ldots, \boldsymbol{\varepsilon}_m^{\mathrm{T}})^{\mathrm{T}}$ is a $N \times 1$ vector. The 'diag' operator is defined as follows: diag(A) is a vector of diagonal elements of A if A is a matrix, a diagonal matrix with elements of A along the diagonal if A is a vector, and a block diagonal matrix whose submatrices along the diagonal are $A_1, \ldots, A_a$ if A consists of matrices $A_1, \ldots, A_a$.

Define

$$y_{ijk} = (y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})(y_{ik} - \boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\beta}), \tag{2.2}$$

where y_{ij} is the jth entry of $\boldsymbol{y}_i$ and $\boldsymbol{x}_{ij}$ is the jth row of $\mathbf{X}_i$ ($i = 1, \ldots, m; j = 1, \ldots, n_i$; $k = j, \ldots, n_i$). Its expectation is the second-order cross-moment of $\mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i$:

$$\begin{aligned} E(y_{ijk}) &= E\big[(\boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_{ij})(\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_{ik})\big] \\ &= \begin{cases} \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{z}_{ik} + \sigma_{\varepsilon}^2 & \text{if } j = k \\ \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{z}_{ik} & \text{otherwise,} \end{cases} \end{aligned} \tag{2.3}$$

where $\boldsymbol{z}_{ij}$ is the jth row of $\mathbf{Z}_i$. If $\boldsymbol{\beta}$ were known, a moment estimator for $\boldsymbol{\Sigma}$ can be obtained by minimizing the squared-error loss function

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\left(y_{ijk}-\boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{z}_{ik}\right)^2.$$

Since $\boldsymbol{\beta}$ is generally unknown, we propose to obtain an unbiased initial estimate $\widetilde{\boldsymbol{\beta}}$ first. A natural choice would be the ordinary least squares estimate from a 'naive' linear regression model $\boldsymbol{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\eta}$ by assuming working independence. Substituting $\widetilde{\boldsymbol{\beta}}$ into (2.2), we can obtain the estimate of y_{ijk} by $\widetilde{y}_{ijk}=(y_{ij}-\boldsymbol{x}_{ij}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})(y_{ik}-\boldsymbol{x}_{ik}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})$, which consequently yields an objective function to estimate $\boldsymbol{\Sigma}$ by minimizing

$$L_0(\boldsymbol{\Sigma})=\sum_{i=1}^{m}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\left(\widetilde{y}_{ijk}-\boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{z}_{ik}\right)^2. \tag{2.4}$$

Let $\widetilde{\boldsymbol{\Sigma}}=[\widetilde{\sigma}_{jk}]$ be the solution to (2.4). For large samples, $\widetilde{\boldsymbol{\Sigma}}$ satisfies positive semidefiniteness, *i.e.*, $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{a}\geq 0$ for all $\boldsymbol{a}\in\mathbb{R}^q$. However, for small samples, $\widetilde{\boldsymbol{\Sigma}}$ is not guaranteed to be positive semidefinite, thus we require a constraint $\boldsymbol{\Sigma}\succeq 0$ to (2.4). The minimization problem subject to this constraint is asymptotically equivalent to the minimization of (2.4). In practice, we hence propose to minimize

$$L_0(\boldsymbol{\Sigma})=\sum_{i=1}^{m}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\left(\widetilde{y}_{ijk}-\boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{z}_{ik}\right)^2 \quad \text{subject to} \quad \boldsymbol{\Sigma}\succeq 0. \tag{2.5}$$

By plugging $\widetilde{\boldsymbol{\Sigma}}$ into the squared-error loss function $\sum_{i=1}^{m}\sum_{j=k}\left(\widetilde{y}_{ijk}-\boldsymbol{z}_{ij}^{T}\boldsymbol{\Sigma}\boldsymbol{z}_{ik}-\sigma_{\varepsilon}^2\right)^2$ which is obtained from (2.3) and minimizing the quantity with respect to σ_{ε}^2, we have the estimate $\widetilde{\sigma}_\varepsilon^2$. Since the variance should be nonnegative, we instead take

$$\widetilde{\sigma}_\varepsilon^2 = \max\left(0, \;\; \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\widetilde{y}_{ijj} - \boldsymbol{z}_{ij}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{z}_{ij})\right).$$

Our main goal is to identify important random effects for model (1.1). Note that if the lth ($1 \le l \le q$) random effect is not important, then $\mathrm{var}(\gamma_{il}) = 0$ for all i, which is equivalent to setting all the elements in the lth column and the lth row of $\boldsymbol{\Sigma}$ to be zero. Since the moment-based estimator $\widetilde{\boldsymbol{\Sigma}}$ does not have a sparse structure, we now propose a penalized approach to produce a sparsity estimate from it. In particular, we suggest to minimize the following objective function

$$\begin{aligned} Q_R(\mathbf{D}) &= \sum_{i=1}^{m}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\left(\widetilde{y}_{ijk} - \boldsymbol{z}_{ij}^{\mathrm{T}}\mathbf{D}\widetilde{\boldsymbol{\Sigma}}\mathbf{D}\boldsymbol{z}_{ik}\right)^2 + \lambda\sum_{i=1}^{q} d_i \qquad (2.6) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\left(\widetilde{y}_{ijk} - \sum_{r=1}^{q}\sum_{s=1}^{q} z_{ijr}z_{iks}\widetilde{\sigma}_{rs}d_r d_s\right)^2 + \lambda\sum_{i=1}^{q} d_i, \end{aligned}$$

subject to all $d_i \ge 0$, where $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_q)$, z_{ijk} is the kth element of $\boldsymbol{z}_{ij}$, $\widetilde{\sigma}_{ij}$ is the (i,j)th entry of $\widetilde{\boldsymbol{\Sigma}}$, and $\lambda \ge 0$ is a tuning parameter. Let $\widehat{\mathbf{D}} = \mathrm{diag}(\widehat{d}_1, \ldots, \widehat{d}_q)$ denote the minimizer of $Q_R(\mathbf{D})$. As λ becomes larger, $\widehat{d}_i$'s tend to be shrunk towards zero more. The subscript 'R' in $Q_R(\mathbf{D})$ refers to random effect selection. The reason we impose the matrix $\mathbf{D}$ on both sides of $\widetilde{\boldsymbol{\Sigma}}$ is due to the property of a covariance matrix, which should be positive semi-definite and symmetric. The sandwich structure of $\mathbf{D}\widetilde{\boldsymbol{\Sigma}}\mathbf{D}$ yields a matrix that satisfies these properties. In addition, when $d_i = 0$, not only the ith diagonal element of $\mathbf{D}\widetilde{\boldsymbol{\Sigma}}\mathbf{D}$ but also the ith column and row become to have the value of zero. That is, the variance of the ith random effect and the corresponding covariances with other random effects are all zero. The penalty in the second term on the right-hand side of (2.6) is inspired by a nonnegative garrote penalty (Breiman, 1995). For illustration with the structure of $\mathbf{D\Sigma D}$, let us assume that there are four random effect factors, and the first two factors are important. Then, $\mathbf{D\Sigma D}$ gives

$$\begin{pmatrix} d_1^2\sigma_{11} & d_1 d_2 \sigma_{12} & 0 & 0 \\ d_1 d_2 \sigma_{12} & d_2^2 \sigma_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The third and fourth diagonal elements corresponding to unimportant random effect factors have the value of zero, and the corresponding covariances are also zero. Hence, this is a desirable penalty on $\mathbf{\Sigma}$ yielding a positive semidefinite and symmetric matrix.

Once we obtain $\widehat{\mathbf{D}}$, the final estimate of $\mathbf{\Sigma}$ is given by $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{D}}\widetilde{\mathbf{\Sigma}}\widehat{\mathbf{D}}$. In addition, we estimate σ_ε^2 by minimizing the squared-error loss function $\sum_{i=1}^m \sum_{j=k} \big(\widetilde{y}_{ijk} - \boldsymbol{z}_{ij}^T \widehat{\mathbf{\Sigma}} \boldsymbol{z}_{ik} - \sigma_\varepsilon^2\big)^2$. To preserve the nonnegativeness of the variance, we take

$$\widehat{\sigma}_\varepsilon^2 = \max\left(0, \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{n_i} (\widetilde{y}_{ijj} - \boldsymbol{z}_{ij}^{\mathrm{T}} \widehat{\mathbf{\Sigma}} \boldsymbol{z}_{ij})\right).$$

2.2 Asymptotic Properties

In this section, we will study the asymptotic properties of the proposed estimator $\widehat{\mathbf{\Sigma}}$. At first, we establish some lemmas to show that the initial estimates $\widetilde{\mathbf{\Sigma}}$ are root-m consistent and asymptotically normal. All proofs of Lemmas and Theorems are given in the Appendix.

Lemma 1 shows that the estimator $\widetilde{\boldsymbol{\beta}}$ is asymptotically normal. Throughout the paper, we use the script 'o' on parameters to denote the true values of corresponding parameters.

Lemma 1. *Under some regularity conditions, the ordinary least squares estimator $\widetilde{\boldsymbol{\beta}}$ satisfies that*

$$\sqrt{m}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \to N(\mathbf{0}, \mathbf{A}_1^{-1}\mathbf{B}_1\mathbf{A}_1^{-1}) \qquad \textit{in distribution},$$

where $\mathbf{A}_1 = \lim_{m\to\infty} \sum_{i=1}^m \mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i/m$ *and* $\mathbf{B}_1 = \lim_{m\to\infty} \sum_{i=1}^m \mathbf{X}_i^{\mathrm{T}}(\mathbf{Z}_i\boldsymbol{\Sigma}\mathbf{Z}_i^{\mathrm{T}} + \sigma_\varepsilon^2\mathbf{I}_{n_i})\mathbf{X}_i/m$.

To facilitate the following technical results and proofs, we reformat Σ into its vector form as $\boldsymbol{\kappa} = \mathrm{vech}(\boldsymbol{\Sigma})$, the vector consisting of $q(q+1)/2$ elements that are on and above the diagonal of $\boldsymbol{\Sigma}$. Then we can rewrite $L_0(\boldsymbol{\Sigma})$ as a function of $\boldsymbol{\kappa}$

$$L_0(\boldsymbol{\kappa}) = \sum_{i=1}^m \sum_{j=1}^{n_i-1} \sum_{k=j+1}^{n_i} \left(\widetilde{y}_{ijk} - \mathbf{Z}_{ijk}^{*\mathrm{T}}\boldsymbol{\kappa}\right)^2,$$

where $\mathbf{Z}_{ijk}^*$ is a $q(q+1)/2 \times 1$ vector consisting of $\boldsymbol{z}_{ij}$ and $\boldsymbol{z}_{ik}$, and expressed as

$$\begin{aligned}\mathbf{Z}_{ijk}^* \;=\;& (z_{ik1}z_{ij1}, z_{ik1}z_{ij2} + z_{ik1}z_{ij1}, \ldots, z_{ik1}z_{ijq} + z_{ikq}z_{ij1},\\ & z_{ik2}z_{ij2}, z_{ik2}z_{ij3} + z_{ik3}z_{ij2}, \ldots, z_{ik2}z_{ijq} + z_{ikq}z_{ij2}, \ldots, z_{ikq}z_{ijq})^{\mathrm{T}}.\end{aligned}$$

That is, the component of $\mathbf{Z}_{ijk}^*$ corresponding to σ_{ab} is $z_{ika}z_{ija}$ if $a = b$, and $z_{ika}z_{ijb} + z_{ikb}z_{ija}$ otherwise.

We define $e_{ijk} = \widetilde{y}_{ijk}(\boldsymbol{\beta}_o) - \mathbf{Z}_{ijk}^{*\mathrm{T}}\boldsymbol{\kappa}_o$ with $\widetilde{y}_{ijk}(\boldsymbol{\beta}_o) = (y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_o)(y_{ik} - x_{ik}^{\mathrm{T}}\boldsymbol{\beta}_o)$. Also, let $\boldsymbol{e}_i$ be a column vector consisting of e_{ijk}'s ($j = 1, \ldots, n_i - 1$; $k = j+1, \ldots, n_i$).

Then Lemma 2 shows the root-m consistency and asymptotic normality of $\widetilde{\boldsymbol{\kappa}}$.

LEMMA 2. *Under some regularity conditions, there exists a local minimizer* $\widetilde{\boldsymbol{\Sigma}}$ *of* (2.4) *which has the asymptotic normality:*

$$\sqrt{m}\big(\widetilde{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_o\big) \to N(\mathbf{0}, \mathbf{A}_2^{-1}\mathbf{B}_2\mathbf{A}_2^{-1}) \qquad \textit{in distribution},$$

where $\mathbf{A}_2 = \lim_{m\to\infty}\sum_{i,j,k}\mathbf{Z}_{ijk}^{*}\mathbf{Z}_{ijk}^{*\mathrm{T}}/m$, $\mathbf{B}_2 = \lim_{m\to\infty}\sum_{i=1}^{m}\mathbf{Z}_i^{*\mathrm{T}} var(\boldsymbol{e}_i)\mathbf{Z}_i^{*}/m$, *and* $\mathbf{Z}_i^{*}$ *is a matrix whose column vectors are* $\mathbf{Z}_{ijk}^{*}$*'s* ($j = 1, \ldots, n_i - 1$; $k = j+1, \ldots, n_i$).

From $\widehat{\sigma}_{ij} = \widetilde{\sigma}_{ij}\widehat{d}_i\widehat{d}_j$, we have $\widehat{d}_i = (\widehat{\sigma}_{ii}/\widetilde{\sigma}_{ii})^{1/2}$, where $\widehat{\sigma}_{ij}$ is the (i,j)th entry of $\widehat{\boldsymbol{\Sigma}}$. Hence, $\widehat{\sigma}_{ij}$ can be expressed as $\widetilde{\sigma}_{ij}(\widehat{\sigma}_{ii}\widehat{\sigma}_{jj})^{1/2}/(\widetilde{\sigma}_{ii}\widetilde{\sigma}_{jj})^{1/2}$. Accordingly, $Q_R(\mathbf{D})$ can be reparameterized as a function of $\mathrm{diag}(\boldsymbol{\Sigma}) = (\sigma_{11}, \ldots, \sigma_{qq})$; *i.e.*

$$Q_R(\sigma_{11}, \ldots, \sigma_{qq}) = \sum_{i=1}^{m}\sum_{j<k}\Big(\widetilde{y}_{ijk} - \sum_{r=1}^{q}\sum_{s=1}^{q} z_{ijr}z_{iks}\frac{\widetilde{\sigma}_{rs}(\sigma_{rr}\sigma_{ss})^{1/2}}{(\widetilde{\sigma}_{rr}\widetilde{\sigma}_{ss})^{1/2}}\Big)^2 + \lambda\sum_{i=1}^{q}\Big(\frac{\sigma_{ii}}{\widetilde{\sigma}_{ii}}\Big)^{1/2}.$$

In this representation, $\widehat{\sigma}_{ij}$ has the value of $0/0$ when $\widetilde{\sigma}_{ii}$ or $\widetilde{\sigma}_{jj} = 0$. Here we define $0/0$ as 0.

To further shorten notations, we denote $\sigma_i = \sigma_{ii}^{1/2}$ and $\widetilde{\sigma}_i = \widetilde{\sigma}_{ii}^{1/2}$ for $i = 1, \ldots, q$, and $\boldsymbol{\sigma}^* = (\sigma_1, \ldots, \sigma_q)^{\mathrm{T}}$. Then we have

$$Q_R(\boldsymbol{\sigma}^*) = \sum_{i=1}^{m}\sum_{j<k}\Big(\widetilde{y}_{ijk} - \sum_{r=1}^{q}\sum_{s=1}^{q} z_{ijr}z_{iks}\frac{\widetilde{\sigma}_{rs}}{\widetilde{\sigma}_r\widetilde{\sigma}_s}\sigma_r^*\sigma_s^*\Big)^2 + \lambda\sum_{i=1}^{q}\frac{\sigma_i^*}{\widetilde{\sigma}_i}.$$

Note that the quantity $Q_R(\boldsymbol{\sigma}^*)$ involves $\widetilde{\boldsymbol{\kappa}} \equiv \mathrm{vech}(\widetilde{\boldsymbol{\Sigma}})$, and both $\widetilde{y}_{ijk}$ and $\widetilde{\boldsymbol{\kappa}}$ depend on $\widetilde{\boldsymbol{\beta}}$. To emphasize these aspects, we write $\widetilde{\boldsymbol{\kappa}} = \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})$ and denote $Q_R(\boldsymbol{\sigma}^*)$ by $Q_R(\boldsymbol{\sigma}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}})$. Then, the objective function $Q_R(\mathbf{D})$ in (2.6) can be equivalently represented as

$$Q_R(\boldsymbol{\sigma}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}) \equiv L_R(\boldsymbol{\sigma}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}) + \lambda \sum_{i=1}^{q} \frac{\sigma_i^*}{\widetilde{\sigma}_i}.$$

In our proofs given in the Appendix, we will use the notation $Q_R(\boldsymbol{\sigma}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}})$. Let $\widehat{\boldsymbol{\sigma}}^*$ denote the minimizer of $Q_R(\boldsymbol{\sigma}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}})$. Note that $\widehat{\boldsymbol{\sigma}}^* = \mathrm{diag}(\widehat{\sigma}_{11}^{1/2}, \ldots, \widehat{\sigma}_{qq}^{1/2})$, where $\widehat{\sigma}_{ii}$'s are the diagonal elements of $\widehat{\boldsymbol{\Sigma}}$.

Using the asymptotic normalities of the initial estimates $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\Sigma}}$, we can establish the theoretical properties of the final estimator. We assume that the first b diagonal elements of $\boldsymbol{\Sigma}_o$ are nonzero and the remaining $q-b$ elements are zero; *i.e.* $\mathrm{diag}(\boldsymbol{\Sigma}_o) = (\boldsymbol{\sigma}_{10}^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}})^{\mathrm{T}}$, where $\boldsymbol{\sigma}_{10}$ is a $b \times 1$ vector with nonzero entries. Accordingly, write $\boldsymbol{\sigma}_0^* = (\boldsymbol{\sigma}_{10}^{*\mathrm{T}}, \mathbf{0}^{\mathrm{T}})^{\mathrm{T}}$ and $\widehat{\boldsymbol{\sigma}}^* = (\widehat{\boldsymbol{\sigma}}_1^{*\mathrm{T}}, \widehat{\boldsymbol{\sigma}}_2^{*\mathrm{T}})^{\mathrm{T}}$. For a matrix $\boldsymbol{A}$, let $\|\boldsymbol{A}\|$ denote the Euclidean norm of the vector $\mathrm{vech}(\boldsymbol{\Sigma})$. In the following, Theorem 1 proves the root-m consistency of $\widehat{\boldsymbol{\Sigma}}$, and Theorem 2 shows that selection consistency of random effects and the asymptotic normality of the estimated nonzero variance components.

Theorem 1 (Root-m Consistency). *If $\lambda/\sqrt{m} \to 0$ as $m \to \infty$, then the estimator $\widehat{\boldsymbol{\Sigma}}$ satisfies $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_o\| = O_p(m^{-1/2})$.*

Theorem 2 (Selection Consistency and Asymptotic Normality). *Assume that the tuning parameter satisfies $\lambda/\sqrt{m} \to 0$ and $\lambda \to \infty$ as $m \to \infty$. Then, with probability tending to 1, the root-m consistent estimator $(\widehat{\boldsymbol{\sigma}}_1^{*\mathrm{T}}, \widehat{\boldsymbol{\sigma}}_2^{*\mathrm{T}})^{\mathrm{T}}$ must satisfy:*

(a) Sparsity : $\widehat{\boldsymbol{\sigma}}_2^* = \mathbf{0}$;

(b) Asymptotic normality : $\sqrt{m}(\widehat{\boldsymbol{\sigma}}_1^* - \boldsymbol{\sigma}_{10}^*) \to N(\mathbf{0}, \mathbf{T})$ in distribution as $m \to \infty$, where $\mathbf{T}$ is defined in (A.26) *in the Appendix.*

Algorithm 1 Exact moment-based method for random effects selection

1. (obtain initial estimate of $\boldsymbol{\beta}$):

   Fit a linear regression model $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}$,

   and then obtain an initial estimate $\widetilde{\boldsymbol{\beta}}$.

2. (obtain initial estimates of $\boldsymbol{\Sigma}$ and σ_ε^2):

   Compute $\widetilde{y}_{ijk} = (y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})(y_{ik} - \boldsymbol{x}_{ik}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})$ for all i, j, k.

   Obtain $\widetilde{\boldsymbol{\Sigma}}$ by minimizing $L_0(\boldsymbol{\Sigma})$, and compute $\widetilde{\sigma}_\varepsilon^2$.

3. (obtain final estimates of $\boldsymbol{\Sigma}$ and σ_ε^2):

   Obtain $\widehat{\mathbf{D}}$ by minimizing $Q_R(\mathbf{D})$, and compute $\widehat{\boldsymbol{\Sigma}} = \widehat{\mathbf{D}}\widetilde{\boldsymbol{\Sigma}}\widehat{\mathbf{D}}$ and $\widehat{\sigma}_\varepsilon^2$.

2.3 Computation & Tuning

2.3.1 Algorithm

The estimation procedure of $(\widehat{\boldsymbol{\Sigma}}, \widehat{\sigma}_\varepsilon^2)$ can be summarized as Algorithm 1. Assume λ is fixed. In the algorithm, we need to tackle two minimization problems and both of them are challenging. In Step 2, the minimization of (2.5) is a nonlinear semidefinite programming problem. In general, the '*semidefinite programming*' implies to minimize a linear objective function subject to the semidefiniteness constraint. Though several software packages have been developed to solve linear semidefinite programming problems, there exist very few methods for solving nonlinear semidefinite programming problems. In this study, we use a free MATLAB toolbox YALMIP, which is a modeling language for rapid optimization (Löfberg, 2004). In YALMIP, all of semidefinite programming problems can be defined in a common format. That is, YALMIP is an interface to convert various problems into the common form. Due to such a flexible interface, it is convenient to implement a number of solvers and compare their results. We chose to use SeDuMi that is the most commonly used solver in YALMIP. SeDuMi was developed by Sturm (1999) for optimization over symmetric cones, and is publicly available as well.

The minimization of (2.6) in Step 3 is a nonlinear programming problem subject to a linear inequality constraint. For this problem, some statistical softwares are available; for example, ‘optim’ function in R and ‘fmincon’ function in Matlab. For faster computations, we here use a MATLAB toolbox TOMLAB for optimization (Holmström, 1999). After comparing the speed and accuracy of various solvers in TOMLAB, we decided to use the TOMLAB base module solver ‘clsSolve’ which treats a nonlinear least squares problems and has seven optimization algorithms. Among them, we select the structured MBFGS method (Wang et al., 2010) because it is known as theoretically best and expected to be best in practice. These MATLAB toolboxes provide ready-to-use functions to implement the new procedure, which can greatly save the programming effort of users. In our numerical examples, these solvers show stable and feasible performance in various settings.

2.3.2 Tuning

We need a selection criterion for comparing and assessing candidate models. To select the optimal λ in the penalty method, information criterion like BIC is widely used for choosing an appropriate model. For example, Zou (2006) suggested the use of BIC for selecting λ for the adaptive Lasso.

Since we use the moment-based approach instead of likelihood function for estimation, we need to modify the traditional criteria for model selection. To note that $\widehat{\boldsymbol{\Sigma}}$ changes over λ, we henceforth use the notation $\widehat{\boldsymbol{\Sigma}}_\lambda$. Two modified versions of GCV are given as follows:

$$GCV1_R(\lambda) = \frac{L_0(\widehat{\boldsymbol{\Sigma}}_\lambda)}{m(1 - df/m)^2}$$

and

$$GCV2_R(\lambda) = \frac{L_0(\widehat{\boldsymbol{\Sigma}}_\lambda)}{N(1 - df/N)^2},$$

where df is the number of nonzero $\widehat{d}_i$'s. The standard definition of GCV in a linear model is given in (1.3). We note that when an estimate of $\boldsymbol{\Sigma}$ is provided, L_0 can be viewed as having the form of the residual sum of squares (RSS). According, $L_0(\widehat{\boldsymbol{\Sigma}}_\lambda)$ can be seen as the RSS in the selected model with λ. Hence, we use $L_0(\widehat{\boldsymbol{\Sigma}}_\lambda)$ in place of the RSS in the selected model.

We also modify the traditional BIC as follows:

$$BIC1_R(\lambda) = \frac{L_0(\widehat{\boldsymbol{\Sigma}}_\lambda)}{L_0(\widetilde{\boldsymbol{\Sigma}})} + \frac{\log(m)}{m} \times df$$

and

$$BIC2_R(\lambda) = \frac{L_0(\widehat{\boldsymbol{\Sigma}}_\lambda)}{L_0(\widetilde{\boldsymbol{\Sigma}})} + \frac{\log(N)}{N} \times df$$

In the same way, $L_0(\widetilde{\boldsymbol{\Sigma}})$ can be seen as the RSS in the full model. In addition, σ^2 in the standard definition of BIC for normal data can be estimated by the mean squared error in the full model. Hence, we replace the mean squared error in the full model by $L_0(\widetilde{\boldsymbol{\Sigma}})$. Some simple algebra calculation will lead to BIC1$_R$ and BIC2$_R$. In Section

2.4, we will examine the performance of these criteria.

2.4 Simulation Studies

2.4.1 Setting

In this simulation, we consider two settings: one setting is simpler and the other one involves a larger number of variables and is hence more challenging. For each setting, we design five cases corresponding to different design structures of the input covariates and the error term distributions. We assume that none of a fixed intercept and a random intercept is not included in the model.

- *Setting* 1: we have five variables, $\boldsymbol{\beta} = (1, 2, 2, 0, 0)^{\mathrm{T}}$, and $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_o & \mathbf{0}_{2\times 3} \\ \mathbf{0}_{3\times 2} & \mathbf{0}_{3\times 3} \end{pmatrix}$ with $\boldsymbol{\Sigma}_o = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$
- *Setting* 2: we have ten variables, $\boldsymbol{\beta} = (1, -1, 0, 0, 0, 0, 0, 0, 0, 0)^{\mathrm{T}}$, and $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_o & \mathbf{0}_{4\times 6} \\ \mathbf{0}_{6\times 4} & \mathbf{0}_{6\times 6} \end{pmatrix}$ with $\boldsymbol{\Sigma}_o = \begin{pmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0 & 0 \\ 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 0.6 \end{pmatrix}$

For each setting, we consider five cases as follows:

- *Case* 1: $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_N)$, cov($\mathbf{X}$)=$\mathbf{I}_p$, $\mathbf{X} = \mathbf{Z}$ (Normal error, independent X)
- *Case* 2: $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_N)$, cov($\mathbf{X}$) has the compound symmetry form with constant variance 1 and constant covariance 0.5, $\mathbf{X} = \mathbf{Z}$ (Normal error, compound symmetry X)

- *Case* 3: $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_N)$, cov($\mathbf{X}$) has the AR(1) structure with $\rho = 0.5$; that is, the covariance between the ith and jth variables is $0.5^{|i-j|}$, $\mathbf{X} = \mathbf{Z}$ (Normal error, AR(1) X)

- *Case* 4: $\boldsymbol{\varepsilon}$ has t distribution with 5 degrees of freedom, cov($\mathbf{X}$)=$\mathbf{I}_p$, $\mathbf{X} = \mathbf{Z}$ (t error, independent X)

- *Case* 5: $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_N)$, cov($\mathbf{X}$)=$\mathbf{I}_p$, $\mathbf{X} \neq \mathbf{Z}$ (Normal error, independent X, $\mathbf{X} \neq \mathbf{Z}$)

For each case, we generate the data as follows:

1. Generate $\boldsymbol{\gamma}_i$ from N($\mathbf{0}$,$\boldsymbol{\Sigma}$) and $\boldsymbol{\varepsilon}_i$ from the distribution specified in each case for all i

2. Generate $\boldsymbol{x}_{ij}$ from N($\mathbf{0}$, cov($\mathbf{X}$)) for all i, j

3. For cases 1-4, set $\boldsymbol{z}_{ij} = \boldsymbol{x}_{ij}$ for all i, j

   For case 5, generate $\boldsymbol{z}_{ij}$ from N($\mathbf{0}$,$\mathbf{I}_N$) for all i, j

4. Generate $\boldsymbol{y}_i$ by summing up $\mathbf{X}_i\boldsymbol{\beta}$, $\mathbf{Z}_i\boldsymbol{\gamma}_i$, and $\boldsymbol{\varepsilon}_i$ for all i

In all the examples, we set the error variance $\sigma_\varepsilon^2 = 1$ and $n_i = 5$ for all i. With regard to the number of subjects, we consider $m = 50, 100, 200$ and 300 for both settings. In each example, 100 data sets are simulated from the model and the average performance is reported.

For random effects selection, we consider four criteria: GCV1$_R$, GCV2$_R$, BIC1$_R$, and BIC2$_R$. The "Oracle" model assumes the knowledge of the true underlying model. We report some measures for assessing the selection performance. “CZ” means the number of zero coefficients correctly estimated to be zero, “IZ” means the number of coefficients incorrectly set to zero, “C” is the frequency of selecting correctly coefficients, “U” is the frequency of under-selecting coefficients, and “O” is the frequency of over-selecting coefficients. We also report the following performance measures:

$$\begin{aligned}
\text{Error1} &= \sqrt{\frac{\sum_{i=1}^{q}\sum_{j=1}^{q}(\widehat{\sigma}_{ij}-\sigma_{ij})^2}{q^2}} \\
\text{Error2} &= \sqrt{\frac{\sum_{i=1}^{q}\sum_{j=i}^{q}(\widehat{\sigma}_{ij}-\sigma_{ij})^2}{\frac{q(q+1)}{2}}} \\
\text{Error3} &= \sum_{i=1}^{q}\sum_{j=1}^{q}|\widehat{\sigma}_{ij}-\sigma_{ij}| \\
\text{Error4} &= \max_{1\le i,j\le q}|\widehat{\sigma}_{ij}-\sigma_{ij}|
\end{aligned}$$

2.4.2 Results

Tables 2.1-2.10 show the results of random effects selection for each setting and case. Standard errors are given in parentheses. In setting 1, GCV1$_R$ seems to work well for overall cases. For small m, BIC1$_R$ tends to select sparse models, while for large m, BIC1$_R$ works the best. For most cases, GCV2$_R$ prefers larger models; that is, it is too conservative. In terms of ‘C’, BIC2$_R$ performs worse than GCV1$_R$ and BIC1$_R$. However, since GCV1$_R$ and BIC1$_R$ select sparser models, there is possibility to miss some of important random effects. From a point of view of covariance estimation, it is better to select larger models than to miss important random effects. Therefore, we see that BIC2_R is overall a good criterion for selecting and estimating random effects. Tables 2.6-2.10 present the results of setting 2, which is a more challenging example. Similar to the results from setting 1, BIC2_R shows reasonable performance in choosing random effects. In terms of selection accuracy, GCV1_R, BIC1_R, and BIC2_R have all good performance; but we recommend the use of BIC2_R because larger models are preferred than too sparse models which might lose some important information.

Table 2.1: Random effects selection and estimation results for Setting 1 and Case 1

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}^2_\varepsilon$	C	U	O
				1	2	3	4				
50	GCV1$_R$	2.94	0.11	0.139	0.163	1.274	0.511	1.257	84	11	5
				(0.006)	(0.008)	(0.073)	(0.024)	(0.036)			
	GCV2$_R$	1.22	0.00	0.154	0.170	2.329	0.454	0.888	14	0	86
				(0.005)	(0.006)	(0.113)	(0.017)	(0.041)			
	BIC1$_R$	2.98	0.29	0.155	0.183	1.382	0.589	1.376	69	29	2
				(0.008)	(0.009)	(0.076)	(0.030)	(0.045)			
	BIC2$_R$	2.55	0.04	0.141	0.162	1.535	0.483	1.131	63	4	33
				(0.006)	(0.007)	(0.097)	(0.020)	(0.039)			
	Oracle	3.00	0.00	0.125	0.145	1.095	0.457	1.001	100	0	0
				(0.005)	(0.006)	(0.052)	(0.020)	(0.036)			
100	GCV1$_R$	2.90	0.00	0.101	0.119	0.951	0.371	1.208	91	0	9
				(0.005)	(0.005)	(0.064)	(0.016)	(0.029)			
	GCV2$_R$	1.19	0.00	0.119	0.132	1.807	0.354	0.941	16	0	84
				(0.004)	(0.005)	(0.089)	(0.014)	(0.032)			
	BIC1$_R$	2.98	0.08	0.105	0.125	0.939	0.398	1.273	90	8	2
				(0.006)	(0.007)	(0.059)	(0.022)	(0.031)			
	BIC2$_R$	2.66	0.00	0.105	0.122	1.102	0.366	1.153	74	0	26
				(0.005)	(0.005)	(0.075)	(0.015)	(0.031)			
	Oracle	3.00	0.00	0.092	0.108	0.810	0.340	1.050	100	0	0
				(0.004)	(0.005)	(0.042)	(0.016)	(0.028)			
200	GCV1$_R$	2.85	0.00	0.077	0.090	0.754	0.273	1.177	88	0	12
				(0.004)	(0.004)	(0.054)	(0.013)	(0.022)			
	GCV2$_R$	0.88	0.00	0.093	0.101	1.494	0.261	0.930	3	0	97
				(0.003)	(0.003)	(0.064)	(0.011)	(0.021)			
	BIC1$_R$	2.99	0.00	0.073	0.086	0.653	0.267	1.211	99	0	1
				(0.004)	(0.004)	(0.035)	(0.013)	(0.018)			
	BIC2$_R$	2.56	0.00	0.082	0.094	0.915	0.274	1.130	73	0	27
				(0.004)	(0.004)	(0.066)	(0.012)	(0.023)			
	Oracle	3.00	0.00	0.070	0.081	0.613	0.254	1.033	100	0	0
				(0.003)	(0.004)	(0.032)	(0.011)	(0.019)			
300	GCV1$_R$	2.85	0.00	0.062	0.073	0.597	0.224	1.176	87	0	13
				(0.003)	(0.003)	(0.034)	(0.010)	(0.018)			
	GCV2$_R$	0.84	0.00	0.073	0.079	1.174	0.205	0.938	0	0	100
				(0.002)	(0.003)	(0.046)	(0.008)	(0.017)			
	BIC1$_R$	3.00	0.00	0.060	0.071	0.528	0.220	1.207	100	0	0
				(0.003)	(0.003)	(0.026)	(0.010)	(0.015)			
	BIC2$_R$	2.49	0.00	0.066	0.076	0.722	0.226	1.110	61	0	39
				(0.003)	(0.003)	(0.038)	(0.009)	(0.021)			
	Oracle	3.00	0.00	0.054	0.062	0.466	0.198	1.019	100	0	0
				(0.002)	(0.003)	(0.024)	(0.009)	(0.016)			

Table 2.2: Random effects selection and estimation results for Setting 1 and Case 2

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}_\varepsilon^2$	C	U	O
				1	2	3	4				
50	$GCV1_R$	2.84	0.15	0.168	0.198	1.624	0.604	1.229	71	15	14
				(0.008)	(0.009)	(0.103)	(0.028)	(0.047)			
	$GCV2_R$	1.31	0.01	0.207	0.228	3.275	0.585	0.848	14	1	85
				(0.009)	(0.010)	(0.203)	(0.024)	(0.048)			
	$BIC1_R$	2.99	0.42	0.189	0.227	1.675	0.723	1.448	58	42	0
				(0.009)	(0.010)	(0.083)	(0.034)	(0.050)			
	$BIC2_R$	2.48	0.05	0.178	0.204	2.021	0.594	1.091	56	5	39
				(0.008)	(0.009)	(0.145)	(0.025)	(0.052)			
	Oracle	3.00	0.00	0.147	0.172	1.268	0.553	0.978	100	0	0
				(0.007)	(0.009)	(0.065)	(0.028)	(0.044)			
100	$GCV1_R$	2.83	0.01	0.126	0.149	1.194	0.464	1.169	85	1	14
				(0.007)	(0.008)	(0.091)	(0.026)	(0.036)			
	$GCV2_R$	1.27	0.00	0.168	0.185	2.690	0.479	0.893	18	0	82
				(0.009)	(0.009)	(0.193)	(0.024)	(0.038)			
	$BIC1_R$	2.98	0.13	0.133	0.160	1.173	0.510	1.265	85	13	2
				(0.008)	(0.010)	(0.086)	(0.030)	(0.041)			
	$BIC2_R$	2.57	0.00	0.133	0.154	1.442	0.464	1.100	68	0	32
				(0.007)	(0.008)	(0.114)	(0.025)	(0.037)			
	Oracle	3.00	0.00	0.111	0.131	0.954	0.421	1.036	100	0	0
				(0.006)	(0.007)	(0.049)	(0.024)	(0.035)			
200	$GCV1_R$	2.71	0.00	0.095	0.110	0.959	0.341	1.131	76	0	24
				(0.005)	(0.005)	(0.070)	(0.016)	(0.026)			
	$GCV2_R$	0.97	0.00	0.130	0.140	2.205	0.349	0.892	7	0	93
				(0.005)	(0.006)	(0.128)	(0.014)	(0.027)			
	$BIC1_R$	2.96	0.01	0.091	0.108	0.799	0.347	1.200	95	1	4
				(0.004)	(0.005)	(0.045)	(0.018)	(0.025)			
	$BIC2_R$	2.37	0.00	0.105	0.118	1.294	0.338	1.070	61	0	39
				(0.006)	(0.006)	(0.124)	(0.016)	(0.030)			
	Oracle	3.00	0.00	0.081	0.094	0.701	0.305	1.022	100	0	0
				(0.004)	(0.004)	(0.034)	(0.014)	(0.023)			
300	$GCV1_R$	2.64	0.00	0.078	0.090	0.830	0.269	1.112	73	0	27
				(0.003)	(0.004)	(0.059)	(0.011)	(0.024)			
	$GCV2_R$	1.00	0.00	0.104	0.113	1.730	0.279	0.911	7	0	93
				(0.004)	(0.004)	(0.089)	(0.010)	(0.022)			
	$BIC1_R$	2.97	0.00	0.076	0.090	0.657	0.284	1.190	97	0	3
				(0.004)	(0.005)	(0.032)	(0.017)	(0.023)			
	$BIC2_R$	2.36	0.00	0.082	0.092	0.942	0.268	1.057	51	0	49
				(0.003)	(0.004)	(0.061)	(0.011)	(0.024)			
	Oracle	3.00	0.00	0.066	0.076	0.570	0.238	1.006	100	0	0
				(0.003)	(0.003)	(0.026)	(0.011)	(0.019)			

Table 2.3: Random effects selection and estimation results for Setting 1 and Case 3

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}_{\varepsilon}^2$	C	U	O
				1	2	3	4				
50	GCV1_R	2.89	0.16	0.165	0.195	1.555	0.604	1.234	74	16	10
				(0.007)	(0.008)	(0.084)	(0.025)	(0.046)			
	GCV2_R	1.26	0.02	0.199	0.219	3.192	0.560	0.833	16	2	82
				(0.009)	(0.009)	(0.201)	(0.023)	(0.048)			
	BIC1_R	2.98	0.38	0.185	0.221	1.655	0.705	1.372	60	38	2
				(0.008)	(0.010)	(0.080)	(0.032)	(0.050)			
	BIC2_R	2.52	0.04	0.169	0.193	1.900	0.568	1.064	61	4	35
				(0.007)	(0.008)	(0.137)	(0.022)	(0.049)			
	Oracle	3.00	0.00	0.147	0.172	1.268	0.553	0.978	100	0	0
				(0.007)	(0.009)	(0.065)	(0.028)	(0.044)			
100	GCV1_R	2.86	0.01	0.123	0.146	1.144	0.457	1.164	86	1	13
				(0.006)	(0.007)	(0.070)	(0.024)	(0.036)			
	GCV2_R	1.31	0.00	0.157	0.174	2.446	0.459	0.897	20	0	80
				(0.007)	(0.008)	(0.164)	(0.022)	(0.038)			
	BIC1_R	2.98	0.09	0.126	0.152	1.099	0.488	1.259	89	9	2
				(0.007)	(0.008)	(0.065)	(0.027)	(0.037)			
	BIC2_R	2.52	0.00	0.132	0.152	1.459	0.457	1.078	66	0	34
				(0.007)	(0.008)	(0.114)	(0.024)	(0.038)			
	Oracle	3.00	0.00	0.111	0.131	0.954	0.421	1.036	100	0	0
				(0.006)	(0.007)	(0.049)	(0.024)	(0.035)			
200	GCV1_R	2.74	0.00	0.092	0.107	0.940	0.331	1.129	82	0	18
				(0.005)	(0.005)	(0.074)	(0.015)	(0.026)			
	GCV2_R	0.95	0.00	0.125	0.135	2.125	0.339	0.889	7	0	93
				(0.005)	(0.005)	(0.120)	(0.014)	(0.027)			
	BIC1_R	2.98	0.01	0.087	0.104	0.760	0.336	1.205	97	1	2
				(0.004)	(0.005)	(0.041)	(0.017)	(0.025)			
	BIC2_R	2.48	0.00	0.098	0.112	1.123	0.330	1.078	65	0	35
				(0.005)	(0.005)	(0.092)	(0.016)	(0.026)			
	Oracle	3.00	0.00	0.081	0.094	0.701	0.305	1.022	100	0	0
				(0.004)	(0.004)	(0.034)	(0.014)	(0.023)			
300	GCV1_R	2.67	0.00	0.075	0.087	0.757	0.262	1.101	74	0	26
				(0.003)	(0.004)	(0.050)	(0.011)	(0.023)			
	GCV2_R	0.89	0.00	0.101	0.109	1.701	0.275	0.905	5	0	95
				(0.004)	(0.004)	(0.089)	(0.011)	(0.023)			
	BIC1_R	2.99	0.00	0.072	0.085	0.620	0.270	1.194	99	0	1
				(0.003)	(0.004)	(0.028)	(0.013)	(0.022)			
	BIC2_R	2.48	0.00	0.077	0.088	0.839	0.260	1.070	60	0	40
				(0.003)	(0.004)	(0.051)	(0.011)	(0.023)			
	Oracle	3.00	0.00	0.066	0.076	0.570	0.238	1.006	100	0	0
				(0.003)	(0.003)	(0.026)	(0.011)	(0.019)			

Table 2.4: Random effects selection and estimation results for Setting 1 and Case 4

m	Criterion	CZ	IZ	Error 1	Error 2	Error 3	Error 4	$\widehat{\sigma}^2_\varepsilon$	C	U	O
50	$GCV1_R$	2.97	0.34	0.172	0.205	1.555	0.664	2.032	64	34	2
				(0.008)	(0.009)	(0.082)	(0.031)	(0.055)			
	$GCV2_R$	1.50	0.01	0.170	0.190	2.420	0.524	1.526	27	1	72
				(0.005)	(0.006)	(0.113)	(0.022)	(0.050)			
	$BIC1_R$	3.00	0.59	0.201	0.239	1.792	0.784	2.203	41	59	0
				(0.008)	(0.010)	(0.076)	(0.032)	(0.059)			
	$BIC2_R$	2.72	0.12	0.159	0.185	1.597	0.572	1.833	66	12	22
				(0.007)	(0.008)	(0.086)	(0.026)	(0.051)			
	Oracle	3.00	0.00	0.132	0.154	1.157	0.486	1.618	100	0	0
				(0.007)	(0.008)	(0.063)	(0.027)	(0.043)			
100	$GCV1_R$	2.87	0.03	0.111	0.130	1.065	0.396	1.835	85	3	12
				(0.005)	(0.006)	(0.066)	(0.019)	(0.038)			
	$GCV2_R$	1.37	0.00	0.131	0.145	1.941	0.388	1.541	14	0	86
				(0.005)	(0.006)	(0.100)	(0.017)	(0.039)			
	$BIC1_R$	2.98	0.13	0.118	0.140	1.052	0.444	1.937	85	13	2
				(0.007)	(0.008)	(0.066)	(0.026)	(0.044)			
	$BIC2_R$	2.70	0.01	0.112	0.130	1.155	0.391	1.780	74	1	25
				(0.005)	(0.006)	(0.073)	(0.018)	(0.039)			
	Oracle	3.00	0.00	0.098	0.115	0.844	0.371	1.659	100	0	0
				(0.005)	(0.005)	(0.042)	(0.019)	(0.032)			
200	$GCV1_R$	2.89	0.00	0.083	0.099	0.758	0.313	1.845	89	0	11
				(0.003)	(0.004)	(0.038)	(0.013)	(0.028)			
	$GCV2_R$	1.15	0.00	0.095	0.105	1.452	0.289	1.596	9	0	91
				(0.003)	(0.004)	(0.069)	(0.012)	(0.030)			
	$BIC1_R$	3.00	0.00	0.080	0.097	0.681	0.314	1.883	100	0	0
				(0.003)	(0.004)	(0.030)	(0.013)	(0.026)			
	$BIC2_R$	2.79	0.00	0.084	0.099	0.786	0.313	1.822	79	0	21
				(0.003)	(0.004)	(0.038)	(0.013)	(0.029)			
	Oracle	3.00	0.00	0.073	0.087	0.643	0.274	1.646	100	0	0
				(0.003)	(0.004)	(0.032)	(0.013)	(0.025)			
300	$GCV1_R$	2.82	0.00	0.071	0.082	0.690	0.251	1.792	84	0	16
				(0.003)	(0.003)	(0.038)	(0.010)	(0.025)			
	$GCV2_R$	1.03	0.00	0.082	0.089	1.267	0.234	1.557	4	0	96
				(0.003)	(0.003)	(0.054)	(0.011)	(0.023)			
	$BIC1_R$	3.00	0.00	0.068	0.080	0.603	0.252	1.839	100	0	0
				(0.003)	(0.003)	(0.027)	(0.010)	(0.021)			
	$BIC2_R$	2.60	0.00	0.073	0.084	0.769	0.249	1.742	66	0	34
				(0.003)	(0.004)	(0.045)	(0.011)	(0.027)			
	Oracle	3.00	0.00	0.053	0.061	0.462	0.196	1.655	100	0	0
				(0.002)	(0.003)	(0.022)	(0.010)	(0.020)			

Table 2.5: Random effects selection and estimation results for Setting 1 and Case 5

m	Criterion	CZ	IZ	Error 1	Error 2	Error 3	Error 4	$\widehat{\sigma}^2_\varepsilon$	C	U	O
50	GCV1$_R$	2.91	0.09	0.146	0.172	1.361	0.543	1.250	85	9	6
				(0.008)	(0.010)	(0.088)	(0.030)	(0.038)			
	GCV2$_R$	1.47	0.00	0.162	0.181	2.341	0.492	0.896	22	0	78
				(0.008)	(0.009)	(0.139)	(0.028)	(0.040)			
	BIC1$_R$	2.99	0.27	0.162	0.192	1.440	0.619	1.381	72	27	1
				(0.009)	(0.011)	(0.088)	(0.035)	(0.045)			
	BIC2$_R$	2.43	0.01	0.148	0.170	1.692	0.503	1.080	60	1	39
				(0.008)	(0.009)	(0.112)	(0.028)	(0.041)			
	Oracle	3.00	0.00	0.132	0.155	1.150	0.486	1.010	100	0	0
				(0.009)	(0.010)	(0.086)	(0.032)	(0.033)			
100	GCV1$_R$	2.87	0.00	0.105	0.122	0.975	0.386	1.166	87	0	13
				(0.005)	(0.005)	(0.055)	(0.018)	(0.024)			
	GCV2$_R$	1.20	0.00	0.124	0.136	1.857	0.376	0.896	11	0	89
				(0.005)	(0.005)	(0.096)	(0.018)	(0.030)			
	BIC1$_R$	2.99	0.04	0.106	0.125	0.927	0.404	1.223	95	4	1
				(0.005)	(0.006)	(0.053)	(0.021)	(0.026)			
	BIC2$_R$	2.57	0.00	0.110	0.127	1.172	0.385	1.104	68	0	32
				(0.005)	(0.006)	(0.076)	(0.018)	(0.028)			
	Oracle	3.00	0.00	0.097	0.113	0.835	0.370	1.002	100	0	0
				(0.005)	(0.006)	(0.047)	(0.020)	(0.024)			
200	GCV1$_R$	2.94	0.00	0.073	0.086	0.650	0.277	1.170	94	0	6
				(0.003)	(0.004)	(0.032)	(0.013)	(0.019)			
	GCV2$_R$	0.89	0.00	0.088	0.096	1.412	0.256	0.893	0	0	100
				(0.003)	(0.003)	(0.061)	(0.011)	(0.021)			
	BIC1$_R$	3.00	0.00	0.072	0.085	0.615	0.275	1.185	100	0	0
				(0.003)	(0.004)	(0.029)	(0.013)	(0.017)			
	BIC2$_R$	2.59	0.00	0.076	0.088	0.793	0.268	1.112	69	0	31
				(0.003)	(0.004)	(0.047)	(0.013)	(0.021)			
	Oracle	3.00	0.00	0.067	0.078	0.589	0.247	1.007	100	0	0
				(0.003)	(0.004)	(0.030)	(0.012)	(0.017)			
300	GCV1$_R$	2.88	0.00	0.064	0.075	0.598	0.229	1.160	90	0	10
				(0.003)	(0.003)	(0.032)	(0.009)	(0.017)			
	GCV2$_R$	0.77	0.00	0.072	0.077	1.169	0.197	0.919	0	0	100
				(0.002)	(0.003)	(0.046)	(0.008)	(0.016)			
	BIC1$_R$	3.00	0.00	0.062	0.073	0.539	0.231	1.187	100	0	0
				(0.003)	(0.003)	(0.026)	(0.009)	(0.014)			
	BIC2$_R$	2.57	0.00	0.065	0.075	0.689	0.220	1.101	64	0	36
				(0.003)	(0.003)	(0.038)	(0.010)	(0.019)			
	Oracle	3.00	0.00	0.052	0.061	0.464	0.188	1.011	100	0	0
				(0.003)	(0.003)	(0.025)	(0.009)	(0.014)			

Table 2.6: Random effects selection and estimation results for Setting 2 and Case 1

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}_{\varepsilon}^{2}$	C	U	O
				1	2	3	4				
50	GCV1$_R$	5.91	1.89	0.140	0.178	3.458	0.866	2.456	7	91	2
				(0.003)	(0.004)	(0.093)	(0.018)	(0.084)			
	GCV2$_R$	0.46	0.03	0.165	0.178	11.808	0.579	0.231	0	3	97
				(0.004)	(0.004)	(0.350)	(0.018)	(0.040)			
	BIC1$_R$	5.97	2.51	0.154	0.198	3.552	0.942	2.926	0	100	0
				(0.002)	(0.003)	(0.077)	(0.012)	(0.080)			
	BIC2$_R$	4.34	0.56	0.133	0.155	5.679	0.682	1.307	13	43	44
				(0.004)	(0.004)	(0.370)	(0.021)	(0.083)			
	Oracle	6.00	0.00	0.121	0.137	3.922	0.637	0.919	100	0	0
				(0.004)	(0.005)	(0.128)	(0.028)	(0.046)			
100	GCV1$_R$	5.73	0.78	0.102	0.126	2.875	0.640	1.812	28	59	13
				(0.002)	(0.003)	(0.091)	(0.020)	(0.061)			
	GCV2$_R$	0.44	0.01	0.127	0.137	9.028	0.468	0.377	1	1	98
				(0.003)	(0.003)	(0.246)	(0.018)	(0.038)			
	BIC1$_R$	5.96	2.03	0.130	0.167	2.943	0.843	2.636	2	97	1
				(0.003)	(0.004)	(0.084)	(0.017)	(0.071)			
	BIC2$_R$	4.75	0.22	0.099	0.116	3.999	0.533	1.314	29	21	50
				(0.003)	(0.003)	(0.263)	(0.019)	(0.060)			
	Oracle	6.00	0.00	0.089	0.101	2.883	0.476	1.013	100	0	0
				(0.003)	(0.003)	(0.082)	(0.023)	(0.041)			
200	GCV1$_R$	5.62	0.23	0.073	0.089	2.324	0.451	1.460	60	21	19
				(0.002)	(0.002)	(0.078)	(0.017)	(0.039)			
	GCV2$_R$	0.30	0.00	0.096	0.103	6.892	0.349	0.506	0	0	100
				(0.002)	(0.002)	(0.141)	(0.011)	(0.033)			
	BIC1$_R$	5.99	1.04	0.091	0.116	2.178	0.632	1.987	25	75	0
				(0.002)	(0.003)	(0.048)	(0.018)	(0.060)			
	BIC2$_R$	4.78	0.06	0.073	0.085	2.930	0.388	1.225	44	6	50
				(0.002)	(0.002)	(0.135)	(0.014)	(0.040)			
	Oracle	6.00	0.00	0.065	0.074	2.114	0.347	1.029	100	0	0
				(0.002)	(0.002)	(0.048)	(0.013)	(0.029)			
300	GCV1$_R$	5.64	0.05	0.058	0.069	1.882	0.337	1.335	64	5	31
				(0.001)	(0.002)	(0.052)	(0.013)	(0.028)			
	GCV2$_R$	0.18	0.00	0.079	0.085	5.620	0.287	0.628	0	0	100
				(0.001)	(0.002)	(0.100)	(0.009)	(0.027)			
	BIC1$_R$	5.98	0.49	0.069	0.086	1.809	0.464	1.657	59	39	2
				(0.002)	(0.003)	(0.037)	(0.020)	(0.048)			
	BIC2$_R$	5.15	0.01	0.059	0.068	2.209	0.314	1.216	51	1	48
				(0.001)	(0.002)	(0.107)	(0.011)	(0.030)			
	Oracle	6.00	0.00	0.053	0.060	1.693	0.284	1.055	100	0	0
				(0.001)	(0.002)	(0.045)	(0.010)	(0.024)			

Table 2.7: Random effects selection and estimation results for Setting 2 and Case 2

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}_\varepsilon^2$	C	U	O
				1	2	3	4				
50	GCV1$_R$	5.64	2.02	0.156	0.196	4.073	0.944	2.079	1	99	0
				(0.004)	(0.005)	(0.159)	(0.026)	(0.080)			
	GCV2$_R$	0.77	0.10	0.259	0.281	18.411	0.886	0.210	0	8	92
				(0.008)	(0.009)	(0.726)	(0.034)	(0.041)			
	BIC1$_R$	5.85	2.36	0.162	0.206	3.927	1.002	2.429	0	100	0
				(0.004)	(0.005)	(0.144)	(0.026)	(0.083)			
	BIC2$_R$	4.22	1.13	0.181	0.211	7.918	0.870	1.272	1	73	26
				(0.008)	(0.008)	(0.723)	(0.025)	(0.080)			
	Oracle	6.00	0.00	0.157	0.176	5.065	0.815	0.796	100	0	0
				(0.006)	(0.007)	(0.182)	(0.040)	(0.051)			
100	GCV1$_R$	5.44	1.28	0.129	0.159	3.731	0.776	1.662	4	88	8
				(0.003)	(0.004)	(0.195)	(0.018)	(0.064)			
	GCV2$_R$	0.35	0.06	0.214	0.230	15.921	0.698	0.248	0	6	94
				(0.006)	(0.006)	(0.528)	(0.022)	(0.040)			
	BIC1$_R$	5.90	1.83	0.134	0.172	3.187	0.850	2.141	0	100	0
				(0.003)	(0.004)	(0.109)	(0.019)	(0.066)			
	BIC2$_R$	4.45	0.81	0.138	0.162	5.682	0.709	1.243	2	63	35
				(0.004)	(0.005)	(0.429)	(0.017)	(0.076)			
	Oracle	6.00	0.00	0.123	0.137	3.986	0.638	0.927	100	0	0
				(0.004)	(0.005)	(0.137)	(0.024)	(0.049)			
200	GCV1$_R$	5.24	0.80	0.102	0.126	3.094	0.641	1.445	16	65	19
				(0.002)	(0.003)	(0.118)	(0.016)	(0.054)			
	GCV2$_R$	0.28	0.00	0.164	0.177	12.304	0.543	0.352	0	0	100
				(0.004)	(0.004)	(0.342)	(0.016)	(0.035)			
	BIC1$_R$	5.89	1.51	0.111	0.144	2.523	0.768	1.940	6	94	0
				(0.002)	(0.003)	(0.068)	(0.014)	(0.051)			
	BIC2$_R$	4.19	0.44	0.109	0.127	4.623	0.579	1.124	12	43	45
				(0.003)	(0.004)	(0.323)	(0.015)	(0.057)			
	Oracle	6.00	0.00	0.093	0.105	2.935	0.495	1.017	100	0	0
				(0.003)	(0.003)	(0.086)	(0.017)	(0.065)			
300	GCV1$_R$	5.16	0.41	0.088	0.105	2.915	0.525	1.291	22	38	40
				(0.002)	(0.003)	(0.122)	(0.018)	(0.043)			
	GCV2$_R$	0.20	0.00	0.135	0.145	10.081	0.439	0.482	0	0	100
				(0.003)	(0.003)	(0.249)	(0.014)	(0.032)			
	BIC1$_R$	5.92	1.20	0.099	0.127	2.270	0.694	1.813	16	83	1
				(0.002)	(0.003)	(0.065)	(0.017)	(0.055)			
	BIC2$_R$	4.26	0.25	0.091	0.106	3.932	0.468	1.099	11	24	65
				(0.003)	(0.003)	(0.258)	(0.017)	(0.043)			
	Oracle	6.00	0.00	0.082	0.094	2.569	0.431	1.175	100	0	0
				(0.003)	(0.004)	(0.080)	(0.018)	(0.086)			

Table 2.8: Random effects selection and estimation results for Setting 2 and Case 3

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}^2_{\varepsilon}$	C	U	O
				1	2	3	4				
50	$\mathrm{GCV1}_R$	5.78	1.92	0.155	0.195	4.037	0.924	2.186	3	95	2
				(0.003)	(0.004)	(0.148)	(0.018)	(0.085)			
	$\mathrm{GCV2}_R$	0.35	0.03	0.254	0.273	18.919	0.841	0.123	0	3	97
				(0.007)	(0.008)	(0.642)	(0.027)	(0.031)			
	$\mathrm{BIC1}_R$	5.97	2.48	0.162	0.206	3.802	1.003	2.717	2	98	0
				(0.003)	(0.004)	(0.121)	(0.019)	(0.086)			
	$\mathrm{BIC2}_R$	4.52	0.90	0.168	0.196	7.092	0.832	1.333	6	64	30
				(0.006)	(0.006)	(0.580)	(0.022)	(0.084)			
	Oracle	6.00	0.00	0.152	0.171	4.937	0.816	0.822	100	0	0
				(0.005)	(0.005)	(0.153)	(0.033)	(0.050)			
100	$\mathrm{GCV1}_R$	5.62	1.11	0.122	0.151	3.352	0.754	1.766	9	79	12
				(0.003)	(0.004)	(0.143)	(0.019)	(0.070)			
	$\mathrm{GCV2}_R$	0.40	0.02	0.195	0.210	14.651	0.633	0.266	1	2	97
				(0.005)	(0.006)	(0.467)	(0.021)	(0.041)			
	$\mathrm{BIC1}_R$	5.95	1.91	0.135	0.174	3.103	0.875	2.367	1	99	0
				(0.004)	(0.004)	(0.105)	(0.021)	(0.062)			
	$\mathrm{BIC2}_R$	4.39	0.55	0.135	0.157	6.053	0.678	1.215	12	48	40
				(0.006)	(0.006)	(0.581)	(0.019)	(0.074)			
	Oracle	6.00	0.00	0.116	0.131	3.755	0.615	0.942	100	0	0
				(0.004)	(0.004)	(0.122)	(0.024)	(0.048)			
200	$\mathrm{GCV1}_R$	5.34	0.49	0.094	0.113	2.986	0.569	1.370	27	45	28
				(0.002)	(0.003)	(0.124)	(0.018)	(0.049)			
	$\mathrm{GCV2}_R$	0.27	0.00	0.148	0.159	11.153	0.481	0.370	1	0	99
				(0.003)	(0.004)	(0.290)	(0.016)	(0.036)			
	$\mathrm{BIC1}_R$	5.94	1.39	0.106	0.137	2.455	0.733	2.031	7	93	0
				(0.002)	(0.003)	(0.068)	(0.014)	(0.052)			
	$\mathrm{BIC2}_R$	4.47	0.28	0.101	0.116	4.264	0.527	1.116	22	27	51
				(0.003)	(0.003)	(0.307)	(0.017)	(0.055)			
	Oracle	6.00	0.00	0.086	0.097	2.753	0.455	0.966	100	0	0
				(0.002)	(0.003)	(0.074)	(0.016)	(0.040)			
300	$\mathrm{GCV1}_R$	5.31	0.17	0.075	0.089	2.527	0.428	1.274	42	17	41
				(0.002)	(0.002)	(0.082)	(0.015)	(0.039)			
	$\mathrm{GCV2}_R$	0.10	0.00	0.122	0.131	9.189	0.399	0.487	0	0	100
				(0.002)	(0.002)	(0.183)	(0.011)	(0.031)			
	$\mathrm{BIC1}_R$	5.88	0.92	0.090	0.113	2.203	0.619	1.749	21	71	8
				(0.002)	(0.003)	(0.053)	(0.018)	(0.054)			
	$\mathrm{BIC2}_R$	4.78	0.12	0.079	0.091	3.118	0.412	1.146	32	12	56
				(0.002)	(0.003)	(0.184)	(0.015)	(0.040)			
	Oracle	6.00	0.00	0.073	0.083	2.322	0.382	1.060	100	0	0
				(0.002)	(0.003)	(0.062)	(0.014)	(0.049)			

Table 2.9: Random effects selection and estimation results for Setting 2 and Case 4

m	Criterion	CZ	IZ	Error 1	Error 2	Error 3	Error 4	$\widehat{\sigma}_\varepsilon^2$	C	U	O
50	GCV1$_R$	5.85	2.22	0.153	0.193	3.753	0.920	3.175	0	98	2
				(0.003)	(0.004)	(0.145)	(0.016)	(0.093)			
	GCV2$_R$	0.66	0.02	0.174	0.189	12.200	0.623	0.587	0	2	98
				(0.005)	(0.005)	(0.408)	(0.023)	(0.070)			
	BIC1$_R$	6.00	2.70	0.162	0.208	3.727	0.982	3.690	0	100	0
				(0.002)	(0.002)	(0.064)	(0.009)	(0.082)			
	BIC2$_R$	4.70	0.93	0.144	0.171	5.521	0.755	2.044	11	53	36
				(0.004)	(0.005)	(0.369)	(0.024)	(0.111)			
	Oracle	6.00	0.00	0.133	0.150	4.257	0.705	1.555	100	0	0
				(0.005)	(0.006)	(0.157)	(0.034)	(0.066)			
100	GCV1$_R$	5.88	1.09	0.110	0.137	2.877	0.692	2.674	23	72	5
				(0.003)	(0.004)	(0.083)	(0.021)	(0.066)			
	GCV2$_R$	0.50	0.00	0.134	0.146	9.600	0.487	0.811	0	0	100
				(0.003)	(0.003)	(0.260)	(0.015)	(0.057)			
	BIC1$_R$	6.00	2.36	0.141	0.181	3.130	0.892	3.553	3	97	0
				(0.003)	(0.004)	(0.083)	(0.015)	(0.069)			
	BIC2$_R$	4.78	0.40	0.107	0.126	4.180	0.577	2.009	25	35	40
				(0.003)	(0.003)	(0.283)	(0.016)	(0.078)			
	Oracle	6.00	0.00	0.093	0.106	2.991	0.498	1.706	100	0	0
				(0.002)	(0.003)	(0.085)	(0.017)	(0.046)			
200	GCV1$_R$	5.63	0.22	0.080	0.095	2.603	0.467	2.008	56	21	23
				(0.002)	(0.003)	(0.116)	(0.017)	(0.048)			
	GCV2$_R$	0.41	0.00	0.106	0.115	7.504	0.390	0.972	1	0	99
				(0.003)	(0.003)	(0.201)	(0.014)	(0.043)			
	BIC1$_R$	5.96	1.30	0.104	0.132	2.495	0.695	2.748	18	82	0
				(0.003)	(0.004)	(0.085)	(0.020)	(0.074)			
	BIC2$_R$	4.99	0.12	0.081	0.094	3.155	0.436	1.796	37	12	51
				(0.003)	(0.003)	(0.205)	(0.016)	(0.054)			
	Oracle	6.00	0.00	0.069	0.078	2.239	0.366	1.639	100	0	0
				(0.002)	(0.002)	(0.064)	(0.012)	(0.033)			
300	GCV1$_R$	5.59	0.07	0.062	0.073	2.095	0.343	1.984	66	7	27
				(0.001)	(0.002)	(0.072)	(0.013)	(0.035)			
	GCV2$_R$	0.28	0.00	0.087	0.093	6.249	0.303	1.152	0	0	100
				(0.001)	(0.002)	(0.129)	(0.010)	(0.035)			
	BIC1$_R$	5.96	0.87	0.081	0.102	2.015	0.543	2.534	39	59	2
				(0.003)	(0.004)	(0.049)	(0.024)	(0.065)			
	BIC2$_R$	5.02	0.04	0.063	0.072	2.452	0.322	1.856	49	4	47
				(0.001)	(0.002)	(0.109)	(0.012)	(0.037)			
	Oracle	6.00	0.00	0.055	0.063	1.767	0.295	1.664	100	0	0
				(0.002)	(0.002)	(0.050)	(0.011)	(0.031)			

Table 2.10: Random effects selection and estimation results for Setting 2 and Case 5

m	Criterion	CZ	IZ	Error 1	Error 2	Error 3	Error 4	$\widehat{\sigma}_{\varepsilon}^{2}$	C	U	O
50	GCV1$_R$	5.81	1.90	0.140 (0.003)	0.178 (0.004)	3.537 (0.134)	0.864 (0.016)	2.402 (0.083)	5	91	4
	GCV2$_R$	0.40	0.01	0.171 (0.005)	0.187 (0.005)	12.271 (0.405)	0.628 (0.023)	0.206 (0.035)	0	1	99
	BIC1$_R$	5.98	2.58	0.153 (0.002)	0.197 (0.003)	3.519 (0.078)	0.949 (0.012)	3.007 (0.070)	0	100	0
	BIC2$_R$	3.97	0.62	0.141 (0.005)	0.164 (0.005)	6.396 (0.469)	0.709 (0.022)	1.235 (0.087)	12	39	49
	Oracle	6.00	0.00	0.118 (0.004)	0.136 (0.005)	3.716 (0.117)	0.650 (0.028)	1.022 (0.052)	100	0	0
100	GCV1$_R$	5.71	0.79	0.102 (0.003)	0.126 (0.003)	2.890 (0.101)	0.642 (0.021)	1.736 (0.066)	27	60	13
	GCV2$_R$	0.57	0.03	0.127 (0.003)	0.139 (0.003)	8.862 (0.251)	0.472 (0.018)	0.357 (0.046)	1	3	96
	BIC1$_R$	5.94	1.99	0.128 (0.003)	0.165 (0.004)	2.965 (0.085)	0.837 (0.016)	2.556 (0.079)	5	94	1
	BIC2$_R$	4.69	0.29	0.099 (0.003)	0.117 (0.003)	3.916 (0.232)	0.527 (0.020)	1.244 (0.063)	22	28	50
	Oracle	6.00	0.00	0.088 (0.003)	0.101 (0.003)	2.805 (0.079)	0.481 (0.020)	1.004 (0.042)	100	0	0
200	GCV1$_R$	5.37	0.18	0.073 (0.002)	0.087 (0.002)	2.470 (0.100)	0.415 (0.015)	1.317 (0.044)	41	17	42
	GCV2$_R$	0.25	0.00	0.097 (0.002)	0.105 (0.002)	6.920 (0.161)	0.353 (0.013)	0.437 (0.035)	0	0	100
	BIC1$_R$	5.96	1.12	0.093 (0.002)	0.120 (0.003)	2.164 (0.053)	0.649 (0.019)	2.018 (0.060)	24	76	0
	BIC2$_R$	4.64	0.04	0.074 (0.002)	0.085 (0.002)	3.065 (0.160)	0.376 (0.013)	1.116 (0.045)	31	4	65
	Oracle	6.00	0.00	0.065 (0.002)	0.074 (0.002)	2.074 (0.063)	0.355 (0.015)	0.975 (0.034)	100	0	0
300	GCV1$_R$	5.43	0.04	0.058 (0.002)	0.068 (0.002)	2.047 (0.101)	0.317 (0.011)	1.226 (0.034)	57	4	39
	GCV2$_R$	0.13	0.00	0.082 (0.001)	0.088 (0.002)	5.929 (0.115)	0.288 (0.009)	0.505 (0.030)	0	0	100
	BIC1$_R$	5.96	0.51	0.068 (0.002)	0.085 (0.003)	1.752 (0.045)	0.464 (0.021)	1.632 (0.047)	54	45	1
	BIC2$_R$	4.83	0.01	0.060 (0.002)	0.068 (0.002)	2.436 (0.137)	0.299 (0.009)	1.119 (0.037)	44	1	55
	Oracle	6.00	0.00	0.053 (0.001)	0.060 (0.002)	1.707 (0.050)	0.287 (0.009)	0.985 (0.027)	100	0	0

Chapter 3

Both Fixed and Random Effects Selection

3.1 Methodology and Algorithms

In previous sections, we focused on the estimation of $(\widehat{\mathbf{\Sigma}}, \widehat{\sigma}_{\varepsilon}^2)$ and the selection of random effects for (1.1). Now we add one more step to select an important set of fixed effects and estimate the coefficients. With the knowledge of $\mathbf{\Sigma}$ and σ_{ε}^2, more efficient estimator of $\boldsymbol{\beta}$ rather than $\widetilde{\boldsymbol{\beta}}$ is the generalized least squares estimator, which is obtained by minimizing the weighted residual sum of squares

$$(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{Z}\mathbf{\Sigma}\mathbf{Z}^{\mathrm{T}} + \sigma_{\varepsilon}^2\mathbf{I}_N)^{-1}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}).$$

The inverse matrix of $(\mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}^{\mathrm{T}} + \sigma_\varepsilon^2\mathbf{I}_N)$ can be decomposed into $\mathbf{Q}^{\mathrm{T}}\mathbf{Q}$, where $\mathbf{Q}$ is an upper triangular matrix. Then, the weighted RSS being minimized can be written as

$$L_F(\boldsymbol{\beta} \mid \boldsymbol{\Sigma}, \sigma_\varepsilon^2) = (\boldsymbol{y}^* - \mathbf{X}^*\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{y}^* - \mathbf{X}^*\boldsymbol{\beta}), \tag{3.1}$$

where $\boldsymbol{y}^* = \mathbf{Q}\boldsymbol{y}$ and $\mathbf{X}^* = \mathbf{Q}\mathbf{X}$.

In practice, since $\boldsymbol{\Sigma}$ and σ_ε^2 are generally unknown, we propose using their estimators $(\widehat{\boldsymbol{\Sigma}}, \widehat{\sigma}_\varepsilon^2)$ in (3.1). This yields the generalized least squares estimator $\widehat{\boldsymbol{\beta}}_G$ by minimizing $L_F(\boldsymbol{\beta} \mid \widehat{\boldsymbol{\Sigma}}, \widehat{\sigma}_\varepsilon^2)$. Note that minimizing $L_F(\boldsymbol{\beta} \mid \boldsymbol{\Sigma} = 0, \sigma_\varepsilon^2 = 1)$ yields the ordinary least squares estimate $\widetilde{\boldsymbol{\beta}}$. To achieve sparsity in the estimate of $\boldsymbol{\beta}$, we propose minimizing

$$Q_F(\boldsymbol{\beta}) = L_F(\boldsymbol{\beta} \mid \widehat{\boldsymbol{\Sigma}}, \widehat{\sigma}_\varepsilon^2) + \tau \sum_{j=1}^{p} w_j |\beta_j|, \tag{3.2}$$

where $\tau \geq 0$ is a tuning parameter and w_j's are data-dependent weights used in the adaptive lasso penalty (Zou, 2006). Here the subscript 'F' refers to fixed effects selection. In case of $w_j = 1$, it gives the lasso estimate (Tibshirani, 1996). We can also use good estimators such as $\widetilde{\boldsymbol{\beta}}$ or $\widehat{\boldsymbol{\beta}}_G$ for different weights w_j's; that is, w_j can be $1/|\widetilde{\boldsymbol{\beta}}_j|$ or $1/|\widehat{\boldsymbol{\beta}}_{G,j}|$. In summary, in order to select both random and fixed effects, we add the step of solving (3.2) into our algorithm proposed in Section 2. Then, the procedure for both fixed and random effects is summarized in Algorithm 2.

Algorithm 2 Moment-based method for both effects selection

1. (obtain initial estimate of $\boldsymbol{\beta}$):

   Fit a linear regression model $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}$,

   and then obtain an initial estimate $\widetilde{\boldsymbol{\beta}}$.

2. (obtain initial estimates of $\boldsymbol{\Sigma}$ and σ_ε^2):

   Compute $\widetilde{y}_{ijk} = (y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})(y_{ik} - \boldsymbol{x}_{ik}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})$ for i, j, k.

   Obtain $\widetilde{\boldsymbol{\Sigma}}$ by minimizing $L_0(\boldsymbol{\Sigma})$, and compute $\widetilde{\sigma}_\varepsilon^2$.

3. (obtain final estimates of $\boldsymbol{\Sigma}$ and σ_ε^2):

   Obtain $\widehat{\mathbf{D}}$ by minimizing $Q_R(\mathbf{D})$, and compute $\widehat{\boldsymbol{\Sigma}} = \widehat{\mathbf{D}}\widetilde{\boldsymbol{\Sigma}}\widehat{\mathbf{D}}$ and $\widehat{\sigma}_\varepsilon^2$.

4. (obtain final estimate of $\boldsymbol{\beta}$):

   Obtain $\widehat{\boldsymbol{\beta}}$ by minimizing $Q_F(\boldsymbol{\beta})$.

3.2 Selection Criteria

To select a proper τ for (3.2), we can use selection criteria such as CV, GCV and BIC. Here we suggest modifying GCV and BIC in a similar way to criteria defined in random effects selection procedure. Specifically, we consider

$$BIC1_F(\tau) = \frac{L_F(\widehat{\boldsymbol{\beta}}_\tau|\widehat{\boldsymbol{\Sigma}}, \widehat{\sigma}_\varepsilon^2)}{L_F(\widetilde{\boldsymbol{\beta}}|\widetilde{\boldsymbol{\Sigma}}, \widetilde{\sigma}_\varepsilon^2)} + \frac{\log(N)}{N} \times df$$

and

$$BIC2_F(\tau) = \frac{L_F(\widehat{\boldsymbol{\beta}}_\tau|\widehat{\boldsymbol{\Sigma}}, \widehat{\sigma}_\varepsilon^2)}{L_F(\widehat{\boldsymbol{\beta}}_G|\widehat{\boldsymbol{\Sigma}}, \widehat{\sigma}_\varepsilon^2)} + \frac{\log(N)}{N} \times df,$$

where df is the number of nonzeros in $\widehat{\boldsymbol{\beta}}_\tau$. In the same way as BIC_R, $L_F(\widehat{\boldsymbol{\beta}}_G|\widehat{\boldsymbol{\Sigma}},\widehat{\sigma}^2_\varepsilon)$ plays a role of the RSS in the full model, while $L_F(\widehat{\boldsymbol{\beta}}_\tau|\widehat{\boldsymbol{\Sigma}},\widehat{\sigma}^2_\varepsilon)$ is that for the selected model with τ.

We can modify GCV for fixed effects selection in a similar way to BIC_F and GCV_R.

$$GCV_F(\tau) = \frac{L_F(\widehat{\boldsymbol{\beta}}_\tau|\widehat{\boldsymbol{\Sigma}},\widehat{\sigma}^2_\varepsilon)}{N(1-df/N)^2} \tag{3.3}$$

We also consider four versions of df in (3.3) as follows:

1. $df = \text{tr}\{\mathbf{X}(\mathbf{X}^\mathrm{T}\mathbf{X}+\tau\mathbf{W}_1^-)^{-1}\mathbf{X}^\mathrm{T}\}$
2. $df =$ the number of nonzero coefficients
3. $df = \text{tr}\{\mathbf{X}(\mathbf{X}^\mathrm{T}\mathbf{X}+\tau\mathbf{W}_1^-)^{-1}\mathbf{X}^\mathrm{T}\} -$ the number of zero coefficients
4. $df = \text{tr}\{\mathbf{X}(\mathbf{X}^\mathrm{T}\mathbf{X}+\tau\mathbf{W}_2^-)^{-1}\mathbf{X}^\mathrm{T}\} -$ the number of zero coefficients

where $\mathbf{W}_1 = \text{diag}(|\widehat{\boldsymbol{\beta}}_\tau|)$, $\mathbf{W}_2 = \text{diag}(2|\widehat{\boldsymbol{\beta}}_\tau|)$ and $\mathbf{W}^-$ denotes a generalized inverse matrix of $\mathbf{W}$. We will examine the performance of these criteria in Section 3.3.

3.3 Simulation Studies

In this simulation study, we employ the same settings used in Section 2.4. In Section 2.4, we found that it is good to use BIC2_R to choose the best λ for parameter tuning in random effects selection. Therefore, we only report the results using BIC2_R in this section. For selecting fixed effects, we tried three different weights: $w_j = 1, 1/|\widetilde{\beta}_j|$, and $1/|\widehat{\beta}_{G,j}|$. We denote these methods 'Lasso', 'ALasso1', and 'ALasso2', respectively.

For parameter tuning in fixed effects selection, we considered seven criteria $\mathrm{BIC1}_F$, $\mathrm{BIC2}_F$, $\mathrm{GCV1}_F$, $\mathrm{GCV2}_F$, $\mathrm{GCV3}_F$, $\mathrm{GCV4}_F$, and CV illustrated in Sections 1.2.1 and 3.2.

We evaluate the performance in four aspects: random effects selection, fixed effects selection, median of model error (MME), and computation time. The model error (ME) is computed as

$$ME = (\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta})^{\mathrm{T}} \mathrm{cov}(\mathbf{X})(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}),$$

where τ is selected by a criterion given in Section 2.3.2. We report its median over 100 simulated data sets. "Both" is a performance measure, which is the frequency of selecting both random and fixed effects correctly. In each example, 100 data sets are simulated from the model and the average performance is reported.

For each weight and criterion, the simulation results are summarized in Appendix B. For random effects selection, we used $\mathrm{BIC2}_R$ statistic. Among seven criteria for fixed effects selection, $\mathrm{BIC2}_F$ yielded the best performance. Hence, the results presented in this section are obtained from using $\mathrm{BIC2}_F$.

Tables 3.1 and 3.2 show the performance of Lasso, ALasso1, and ALasso2 for each setting. ALasso1 and ALasso2 are obviously better than Lasso in all cases. In addition, they tend to select nearly correct models as m increases. Moreover, ALasso2 has better or comparable performance because it uses $\widehat{\boldsymbol{\beta}}_G$ which is a better estimate of $\boldsymbol{\beta}$. For all the results, see Appendix B.

In Tables 3.3 and 3.4, we summarize the biases, standard errors, and MME for the estimates of important fixed effects in each case under both settings. In order to show

Table 3.1: Comparison of Lasso, ALasso1, and ALasso2 for fixed effects selection in Setting 1

Case#	m	method	CZ	IZ	MME	C	U	O	Both
Case 1	50	Lasso	1.07	0.02	0.069	30	1	69	23
		ALasso1	1.94	0.01	0.048	93	1	6	60
		ALasso2	1.91	0.01	0.045	90	1	9	57
		Oracle	2.00	0.00	0.038	100	0	0	100
	100	Lasso	1.07	0.00	0.051	31	0	69	26
		ALasso1	1.95	0.00	0.028	95	0	5	70
		ALasso2	1.97	0.00	0.027	97	0	3	71
		Oracle	2.00	0.00	0.022	100	0	0	100
	200	Lasso	1.12	0.00	0.023	35	0	65	29
		ALasso1	1.98	0.00	0.010	98	0	2	71
		ALasso2	1.98	0.00	0.010	98	0	2	71
		Oracle	2.00	0.00	0.009	100	0	0	100
	300	Lasso	1.12	0.00	0.014	35	0	65	25
		ALasso1	1.99	0.00	0.007	99	0	1	60
		ALasso2	1.99	0.00	0.006	99	0	1	60
		Oracle	2.00	0.00	0.006	100	0	0	100
Case 2	50	Lasso	1.20	0.06	0.093	33	3	64	17
		ALasso1	1.89	0.06	0.055	88	3	9	52
		ALasso2	1.91	0.06	0.059	90	3	7	52
		Oracle	2.00	0.00	0.049	100	0	0	100
	100	Lasso	1.08	0.04	0.041	30	2	68	24
		ALasso1	1.97	0.02	0.032	95	2	3	66
		ALasso2	1.95	0.03	0.031	93	2	5	65
		Oracle	2.00	0.00	0.028	100	0	0	100
	200	Lasso	1.15	0.00	0.016	36	0	64	22
		ALasso1	1.98	0.00	0.012	98	0	2	59
		ALasso2	1.98	0.00	0.010	98	0	2	59
		Oracle	2.00	0.00	0.009	100	0	0	100
	300	Lasso	1.16	0.00	0.013	38	0	62	20
		ALasso1	1.96	0.00	0.007	97	0	3	50
		ALasso2	1.99	0.00	0.006	99	0	1	50
		Oracle	2.00	0.00	0.010	100	0	0	100
Case 3	50	Lasso	1.16	0.10	0.082	30	6	64	20
		ALasso1	1.92	0.08	0.062	89	5	6	57
		ALasso2	1.91	0.08	0.061	88	5	7	56
		Oracle	2.00	0.00	0.049	100	0	0	100
	100	Lasso	1.15	0.07	0.046	33	3	64	22
		ALasso1	1.97	0.05	0.036	95	3	2	64
		ALasso2	1.97	0.05	0.033	95	3	2	64
		Oracle	2.00	0.00	0.028	100	0	0	100
	200	Lasso	1.19	0.00	0.023	39	0	61	23

Table 3.1 Continued

Case#	m	method	CZ	IZ	MME	C	U	O	Both
		ALasso1	1.98	0.00	0.013	98	0	2	63
		ALasso2	1.98	0.00	0.011	98	0	2	63
		Oracle	2.00	0.00	0.009	100	0	0	100
	300	Lasso	1.30	0.00	0.016	46	0	54	30
		ALasso1	1.97	0.00	0.007	98	0	2	59
		ALasso2	1.99	0.00	0.006	99	0	1	59
		Oracle	2.00	0.00	0.010	100	0	0	100
Case 4	50	Lasso	1.23	0.00	0.087	40	0	60	25
		ALasso1	1.96	0.00	0.054	96	0	4	63
		ALasso2	1.98	0.00	0.054	98	0	2	65
		Oracle	2.00	0.00	0.044	100	0	0	100
	100	Lasso	1.22	0.00	0.061	42	0	58	31
		ALasso1	1.93	0.00	0.032	94	0	6	69
		ALasso2	1.94	0.00	0.031	94	0	6	70
		Oracle	2.00	0.00	0.029	100	0	0	100
	200	Lasso	1.22	0.00	0.026	37	0	63	31
		ALasso1	1.98	0.00	0.018	98	0	2	77
		ALasso2	1.98	0.00	0.017	98	0	2	77
		Oracle	2.00	0.00	0.017	100	0	0	100
	300	Lasso	1.27	0.00	0.019	40	0	60	30
		ALasso1	1.98	0.00	0.008	99	0	1	66
		ALasso2	1.99	0.00	0.007	99	0	1	65
		Oracle	2.00	0.00	0.010	100	0	0	100
Case 5	50	Lasso	1.50	0.00	0.035	58	0	42	38
		ALasso1	1.88	0.00	0.018	89	0	11	53
		ALasso2	1.90	0.00	0.018	90	0	10	54
		Oracle	2.00	0.00	0.014	100	0	0	100
	100	Lasso	1.70	0.00	0.014	75	0	25	49
		ALasso1	1.93	0.00	0.007	93	0	7	64
		ALasso2	1.96	0.00	0.007	96	0	4	65
		Oracle	2.00	0.00	0.006	100	0	0	100
	200	Lasso	1.74	0.00	0.008	78	0	22	58
		ALasso1	1.97	0.00	0.005	97	0	3	68
		ALasso2	1.98	0.00	0.004	98	0	2	68
		Oracle	2.00	0.00	0.004	100	0	0	100
	300	Lasso	1.75	0.00	0.005	76	0	24	48
		ALasso1	2.00	0.00	0.002	100	0	0	64
		ALasso2	2.00	0.00	0.002	100	0	0	64
		Oracle	2.00	0.00	0.002	100	0	0	100

Table 3.2: Comparison of Lasso, ALasso1, and ALasso2 for fixed effects selection in Setting 2

Case#	m	method	CZ	IZ	MME	C	U	O	Both
Case 1	50	Lasso	6.59	0.12	0.197	36	8	56	6
		ALasso1	7.34	0.05	0.094	62	4	34	8
		ALasso2	7.44	0.09	0.081	66	6	28	10
		Oracle	8.00	0.00	0.038	100	0	0	100
	100	Lasso	6.95	0.02	0.080	45	1	54	16
		ALasso1	7.59	0.02	0.029	78	1	21	24
		ALasso2	7.75	0.02	0.027	83	1	16	23
		Oracle	8.00	0.00	0.018	100	0	0	100
	200	Lasso	7.23	0.00	0.045	49	0	51	23
		ALasso1	7.79	0.00	0.014	86	0	14	39
		ALasso2	7.91	0.00	0.015	92	0	8	41
		Oracle	8.00	0.00	0.010	100	0	0	100
	300	Lasso	7.34	0.00	0.024	54	0	46	33
		ALasso1	7.91	0.00	0.008	95	0	5	49
		ALasso2	7.99	0.00	0.007	99	0	1	51
		Oracle	8.00	0.00	0.006	100	0	0	100
Case 2	50	Lasso	6.31	0.20	0.189	24	15	61	0
		ALasso1	7.15	0.03	0.103	58	2	40	0
		ALasso2	7.27	0.09	0.103	59	6	35	0
		Oracle	8.00	0.00	0.053	100	0	0	100
	100	Lasso	6.48	0.08	0.081	27	7	66	1
		ALasso1	7.40	0.02	0.039	70	2	28	2
		ALasso2	7.51	0.06	0.034	71	5	24	2
		Oracle	8.00	0.00	0.018	100	0	0	100
	200	Lasso	6.89	0.00	0.043	43	0	57	7
		ALasso1	7.87	0.00	0.019	89	0	11	11
		ALasso2	7.87	0.00	0.018	87	0	13	12
		Oracle	8.00	0.00	0.012	100	0	0	100
	300	Lasso	7.02	0.00	0.023	41	0	59	4
		ALasso1	7.83	0.00	0.008	88	0	12	11
		ALasso2	7.91	0.00	0.007	95	0	5	11
		Oracle	8.00	0.00	0.005	100	0	0	100
Case 3	50	Lasso	6.07	0.18	0.202	13	12	75	1
		ALasso1	7.25	0.09	0.112	51	7	42	3
		ALasso2	7.27	0.10	0.102	55	7	38	4
		Oracle	8.00	0.00	0.044	100	0	0	100
	100	Lasso	6.13	0.12	0.090	20	9	71	2
		ALasso1	7.35	0.03	0.038	69	2	29	10
		ALasso2	7.45	0.09	0.031	71	7	22	11
		Oracle	8.00	0.00	0.017	100	0	0	100
	200	Lasso	6.37	0.00	0.047	20	0	80	6

Table 3.2 Continued

Case#	m	method	CZ	IZ	MME	C	U	O	Both
		ALasso1	7.81	0.00	0.022	87	0	13	21
		ALasso2	7.80	0.00	0.017	85	0	15	20
		Oracle	8.00	0.00	0.013	100	0	0	100
	300	Lasso	6.99	0.00	0.025	41	0	59	14
		ALasso1	7.86	0.00	0.008	91	0	9	29
		ALasso2	7.96	0.00	0.007	96	0	4	31
		Oracle	8.00	0.00	0.006	100	0	0	100
Case 4	50	Lasso	6.76	0.09	0.193	37	7	56	7
		ALasso1	7.36	0.03	0.098	60	3	37	7
		ALasso2	7.38	0.08	0.094	60	5	35	9
		Oracle	8.00	0.00	0.041	100	0	0	100
	100	Lasso	6.85	0.02	0.097	40	2	58	12
		ALasso1	7.77	0.01	0.033	83	1	16	23
		ALasso2	7.74	0.01	0.031	82	1	17	24
		Oracle	8.00	0.00	0.023	100	0	0	100
	200	Lasso	7.17	0.00	0.056	50	0	50	20
		ALasso1	7.84	0.00	0.020	89	0	11	34
		ALasso2	7.90	0.00	0.018	95	0	5	35
		Oracle	8.00	0.00	0.013	100	0	0	100
	300	Lasso	7.43	0.00	0.036	59	0	41	29
		ALasso1	7.92	0.00	0.012	92	0	8	47
		ALasso2	7.94	0.00	0.011	94	0	6	47
		Oracle	8.00	0.00	0.008	100	0	0	100
Case 5	50	Lasso	6.42	0.00	0.073	45	0	55	9
		ALasso1	6.72	0.00	0.029	62	0	38	11
		ALasso2	7.04	0.00	0.031	61	0	39	9
		Oracle	8.00	0.00	0.014	100	0	0	100
	100	Lasso	6.95	0.00	0.027	61	0	39	14
		ALasso1	7.40	0.00	0.011	82	0	18	21
		ALasso2	7.58	0.00	0.009	89	0	11	22
		Oracle	8.00	0.00	0.006	100	0	0	100
	200	Lasso	7.52	0.00	0.012	70	0	30	24
		ALasso1	7.68	0.00	0.005	85	0	15	30
		ALasso2	7.78	0.00	0.005	91	0	9	30
		Oracle	8.00	0.00	0.003	100	0	0	100
	300	Lasso	7.67	0.00	0.007	78	0	22	37
		ALasso1	7.79	0.00	0.003	89	0	11	43
		ALasso2	7.87	0.00	0.003	95	0	5	43
		Oracle	8.00	0.00	0.002	100	0	0	100

that the efficiency of the weighted least squares estimates can be improved by random effects selection, we compare three methods as well as the Oracle procedure. The first two methods fit the model without random effects selection: the OLS is based on the ordinary least squares estimate, and the generalized least squares estimate GLS1 is computed using the initial estimates $\widetilde{\boldsymbol{\Sigma}}$ and $\widetilde{\sigma}_{\varepsilon}^2$. The third method, GLS2, uses $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\sigma}_{\varepsilon}^2$ which are computed after selecting random effects with BIC2_R. Tables 3.3 and 3.4 show that GLS2 is more efficient than OLS and GLS1 and gives a smaller MME. Here the Oracle's MMEs serve as a baseline for comparison. These results confirm that a proper random effects selection in linear mixed models can lead to efficient gain in the estimation of fixed effects.

Now we compare our proposed method to the MLE-based method of Bondell et al. (2010) (denoted by BKG in the tables). When implementing the method of Bondell et al. (2010), we set the options 't.fracs' to the default value and 'eps' to 0.001, since the computation was extremely slow if a default value is used for 'eps'. Tables 3.5 and 3.6 compare the performance of mixed-effects selection in our proposed method and BKG method.

In Setting 1, we observe that the proposed method gives a smaller MME in most cases and a comparable MME in the rest of cases. With regard to fixed effects selection, in all the cases our method consistently performs better than BKG by showing a higher frequency of identifying the correct model structure for the fixed effects. In terms of random effects selection, the proposed method shows much better performance than BKG in cases 1, 4, and 5, where the variables are uncorrelated or the error term is non-normal distributed data. The results for case 4 is not surprising,

Table 3.3: Biases and standard errors of the estimated fixed-effect coefficients in Setting 1

Case#	m	method	Mean of biases			Standard error			MME
			β_1	β_2	β_3	β_1	β_2	β_3	
Case1	50	OLS	0.030	-0.010	0.004	0.199	0.174	0.110	0.089
		GLS1	0.035	-0.014	0.014	0.285	0.168	0.079	0.063
		GLS2	0.016	-0.008	0.011	0.177	0.152	0.072	0.055
		Oracle	0.018	-0.008	0.014	0.169	0.155	0.068	0.038
	100	OLS	0.015	-0.010	0.001	0.126	0.133	0.074	0.039
		GLS1	0.008	-0.004	0.007	0.128	0.123	0.060	0.032
		GLS2	0.008	-0.005	0.008	0.123	0.123	0.056	0.030
		Oracle	0.008	-0.006	0.010	0.121	0.122	0.055	0.022
	200	OLS	0.011	0.003	0.000	0.093	0.090	0.050	0.020
		GLS1	0.008	0.007	0.004	0.079	0.084	0.038	0.013
		GLS2	0.007	0.007	0.004	0.080	0.084	0.037	0.012
		Oracle	0.007	0.006	0.005	0.081	0.085	0.036	0.009
	300	OLS	0.011	0.000	0.001	0.070	0.078	0.043	0.015
		GLS1	0.009	0.002	0.002	0.062	0.068	0.034	0.008
		GLS2	0.010	0.002	0.002	0.062	0.069	0.033	0.008
		Oracle	0.010	0.001	0.003	0.063	0.068	0.033	0.006
Case2	50	OLS	0.025	-0.015	0.007	0.228	0.203	0.143	0.097
		GLS1	0.009	-0.025	0.019	0.202	0.183	0.106	0.072
		GLS2	0.005	-0.018	0.013	0.198	0.171	0.093	0.064
		Oracle	0.018	-0.013	0.008	0.181	0.166	0.081	0.049
	100	OLS	0.021	-0.008	0.003	0.146	0.160	0.094	0.043
		GLS1	0.006	-0.010	0.011	0.138	0.142	0.080	0.043
		GLS2	0.004	-0.012	0.009	0.130	0.136	0.070	0.036
		Oracle	0.008	-0.011	0.010	0.124	0.131	0.066	0.028
	200	OLS	0.010	0.006	0.005	0.110	0.108	0.061	0.025
		GLS1	0.004	0.002	0.004	0.084	0.089	0.048	0.014
		GLS2	0.002	0.001	0.005	0.085	0.088	0.046	0.013
		Oracle	0.005	0.002	0.007	0.081	0.088	0.043	0.009
	300	OLS	0.011	0.000	0.005	0.093	0.087	0.056	0.016
		GLS1	0.006	-0.001	0.003	0.069	0.072	0.041	0.008
		GLS2	0.007	-0.002	0.003	0.069	0.072	0.040	0.008
		Oracle	0.009	-0.001	0.004	0.067	0.071	0.040	0.006
Case3	50	OLS	0.030	-0.014	0.004	0.227	0.194	0.149	0.097
		GLS1	0.017	-0.026	0.015	0.196	0.188	0.110	0.072
		GLS2	0.012	-0.013	0.010	0.192	0.171	0.094	0.060
		Oracle	0.021	-0.014	0.008	0.179	0.170	0.076	0.049
	100	OLS	0.019	-0.010	0.002	0.138	0.153	0.096	0.043
		GLS1	0.012	-0.010	0.009	0.131	0.145	0.080	0.043
		GLS2	0.012	-0.011	0.011	0.126	0.137	0.070	0.036
		Oracle	0.011	-0.012	0.010	0.122	0.133	0.062	0.028

Table 3.3 Continued

Case#	m	method	Mean of biases			Standard error			MME
			β_1	β_2	β_3	β_1	β_2	β_3	
	200	OLS	0.008	0.003	0.003	0.101	0.101	0.063	0.025
		GLS1	0.007	0.003	0.002	0.080	0.090	0.050	0.014
		GLS2	0.006	0.003	0.003	0.081	0.089	0.047	0.012
		Oracle	0.007	0.001	0.006	0.081	0.089	0.040	0.009
	300	OLS	0.010	-0.002	0.003	0.080	0.083	0.056	0.016
		GLS1	0.009	-0.001	0.001	0.064	0.071	0.041	0.008
		GLS2	0.009	-0.002	0.001	0.064	0.072	0.040	0.008
		Oracle	0.010	-0.002	0.003	0.065	0.072	0.037	0.006
Case4	50	OLS	-0.022	0.009	0.022	0.184	0.182	0.114	0.082
		GLS1	-0.019	0.004	0.008	0.173	0.169	0.092	0.070
		GLS2	-0.018	0.005	0.013	0.171	0.167	0.099	0.062
		Oracle	-0.016	0.008	0.013	0.168	0.169	0.095	0.044
	100	OLS	-0.009	-0.013	-0.005	0.170	0.152	0.087	0.059
		GLS1	-0.013	-0.015	-0.003	0.146	0.135	0.073	0.038
		GLS2	-0.013	-0.015	-0.003	0.144	0.137	0.071	0.038
		Oracle	-0.014	-0.015	-0.002	0.140	0.137	0.070	0.029
	200	OLS	-0.014	0.011	-0.011	0.111	0.106	0.065	0.028
		GLS1	-0.005	0.010	-0.011	0.098	0.092	0.052	0.021
		GLS2	-0.006	0.012	-0.011	0.097	0.092	0.052	0.022
		Oracle	-0.006	0.011	-0.011	0.098	0.092	0.052	0.017
	300	OLS	0.001	-0.002	0.005	0.080	0.078	0.051	0.018
		GLS1	0.001	-0.003	0.003	0.070	0.067	0.039	0.011
		GLS2	0.000	-0.002	0.003	0.070	0.068	0.038	0.011
		Oracle	0.000	-0.002	0.002	0.069	0.069	0.038	0.008
Case5	50	OLS	-0.025	0.007	0.003	0.104	0.110	0.113	0.059
		GLS1	-0.004	0.014	0.007	0.083	0.081	0.084	0.029
		GLS2	-0.008	0.009	0.004	0.077	0.078	0.078	0.030
		Oracle	-0.012	0.009	-0.002	0.073	0.075	0.078	0.014
	100	OLS	-0.024	-0.001	0.003	0.079	0.073	0.080	0.025
		GLS1	-0.008	-0.005	0.002	0.056	0.051	0.051	0.013
		GLS2	-0.007	-0.003	0.004	0.053	0.048	0.049	0.012
		Oracle	-0.006	-0.001	0.003	0.053	0.044	0.050	0.006
	200	OLS	-0.014	0.004	0.000	0.061	0.054	0.059	0.013
		GLS1	-0.005	0.002	0.005	0.044	0.037	0.042	0.007
		GLS2	-0.005	0.002	0.004	0.042	0.036	0.041	0.007
		Oracle	-0.004	0.002	0.004	0.042	0.034	0.040	0.004
	300	OLS	-0.011	0.005	-0.002	0.049	0.046	0.044	0.009
		GLS1	-0.004	0.003	0.001	0.036	0.029	0.032	0.004
		GLS2	-0.004	0.003	0.002	0.035	0.028	0.031	0.004
		Oracle	-0.004	0.003	0.002	0.034	0.028	0.031	0.002

Table 3.4: Biases and standard errors of the estimated fixed-effect coefficients in Setting 2

Case#	m	method	Mean of biases		Standard error		MME
			β_1	β_2	β_1	β_2	
Case1	50	OLS	0.000	-0.017	0.202	0.198	0.249
		GLS1	-0.004	-0.016	0.357	0.304	0.539
		GLS2	-0.014	-0.011	0.232	0.209	0.172
		Oracle	-0.004	-0.014	0.155	0.170	0.038
	100	OLS	0.009	-0.005	0.131	0.141	0.117
		GLS1	-0.021	-0.022	0.217	0.228	0.103
		GLS2	0.004	-0.002	0.111	0.129	0.076
		Oracle	0.003	0.007	0.103	0.123	0.018
	200	OLS	-0.002	-0.009	0.093	0.108	0.061
		GLS1	-0.010	-0.010	0.092	0.102	0.044
		GLS2	-0.007	-0.004	0.084	0.092	0.040
		Oracle	-0.008	-0.002	0.081	0.088	0.010
	300	OLS	-0.001	-0.011	0.085	0.083	0.040
		GLS1	-0.002	-0.011	0.072	0.065	0.023
		GLS2	-0.003	-0.010	0.070	0.066	0.022
		Oracle	-0.003	-0.009	0.069	0.063	0.006
Case2	50	OLS	-0.014	-0.022	0.235	0.259	0.292
		GLS1	0.096	-0.079	0.751	0.859	0.824
		GLS2	0.017	-0.044	0.319	0.276	0.176
		Oracle	0.032	-0.031	0.167	0.182	0.053
	100	OLS	0.012	-0.002	0.162	0.186	0.136
		GLS1	-0.037	-0.019	0.490	0.340	0.195
		GLS2	0.009	-0.017	0.166	0.166	0.085
		Oracle	0.000	0.023	0.117	0.132	0.018
	200	OLS	0.003	-0.015	0.125	0.135	0.068
		GLS1	0.025	-0.030	0.193	0.223	0.053
		GLS2	-0.007	-0.007	0.102	0.114	0.036
		Oracle	-0.007	0.002	0.092	0.096	0.012
	300	OLS	0.000	-0.016	0.106	0.107	0.039
		GLS1	0.008	-0.005	0.118	0.091	0.028
		GLS2	-0.003	-0.010	0.095	0.072	0.021
		Oracle	-0.002	-0.008	0.077	0.067	0.005
Case3	50	OLS	0.007	-0.015	0.220	0.247	0.289
		GLS1	0.032	-0.051	0.378	0.641	0.721
		GLS2	0.001	-0.021	0.262	0.230	0.183
		Oracle	-0.002	-0.005	0.162	0.185	0.044
	100	OLS	0.018	0.003	0.144	0.176	0.133
		GLS1	-0.007	0.027	0.288	0.274	0.169
		GLS2	0.027	0.001	0.193	0.195	0.079
		Oracle	-0.005	0.020	0.112	0.134	0.017

Table 3.4 Continued

Case#	m	method	Mean of biases β_1	Mean of biases β_2	Standard error β_1	Standard error β_2	MME
	200	OLS	0.007	-0.013	0.113	0.128	0.067
		GLS1	0.011	-0.030	0.139	0.139	0.053
		GLS2	0.000	-0.008	0.102	0.119	0.038
		Oracle	-0.009	0.002	0.089	0.097	0.013
	300	OLS	0.004	-0.015	0.096	0.103	0.041
		GLS1	0.000	-0.017	0.102	0.096	0.028
		GLS2	-0.005	-0.008	0.077	0.069	0.023
		Oracle	-0.005	-0.006	0.075	0.067	0.006
Case4	50	OLS	-0.012	-0.027	0.235	0.209	0.238
		GLS1	0.031	-0.023	0.344	0.307	0.354
		GLS2	-0.007	-0.027	0.220	0.208	0.194
		Oracle	-0.011	-0.031	0.192	0.157	0.041
	100	OLS	-0.001	0.009	0.128	0.136	0.115
		GLS1	-0.008	0.013	0.150	0.159	0.110
		GLS2	-0.004	0.010	0.112	0.137	0.096
		Oracle	0.001	0.012	0.108	0.128	0.023
	200	OLS	0.007	0.028	0.116	0.097	0.064
		GLS1	0.002	0.022	0.103	0.092	0.051
		GLS2	0.003	0.021	0.097	0.088	0.048
		Oracle	0.002	0.019	0.098	0.088	0.013
	300	OLS	0.010	0.002	0.092	0.074	0.044
		GLS1	0.000	0.007	0.084	0.070	0.034
		GLS2	0.002	0.006	0.084	0.068	0.031
		Oracle	0.002	0.005	0.082	0.069	0.008
Case5	50	OLS	-0.012	-0.010	0.118	0.147	0.178
		GLS1	-0.010	-0.049	0.215	0.249	0.325
		GLS2	-0.010	-0.032	0.157	0.155	0.130
		Oracle	0.000	-0.016	0.084	0.097	0.014
	100	OLS	-0.010	-0.003	0.080	0.092	0.082
		GLS1	-0.010	-0.011	0.119	0.151	0.086
		GLS2	-0.006	-0.002	0.100	0.087	0.045
		Oracle	-0.001	0.003	0.062	0.068	0.006
	200	OLS	-0.005	-0.003	0.058	0.073	0.039
		GLS1	0.009	-0.006	0.088	0.091	0.029
		GLS2	0.007	0.003	0.049	0.059	0.021
		Oracle	0.005	0.002	0.045	0.050	0.003
	300	OLS	-0.008	-0.004	0.051	0.053	0.026
		GLS1	0.009	-0.001	0.078	0.051	0.016
		GLS2	0.001	-0.004	0.041	0.041	0.013
		Oracle	0.002	-0.002	0.038	0.038	0.002

as the MLE method depends on the normal assumption for the error term. In cases 2 and 3 where input variables are moderately correlated, our method produces slightly better selection for $m = 50$ and 100 while BKG method performs slightly better when $m = 200$ and 300.

In Setting 2, the proposed method shows a smaller or comparable MME in all the cases. In terms of random effects selection, our method shows better performance in cases 1, 4, and 5 and worse performance in cases 2 and 3. In terms of fixed effects selection, the new method gives an overall significantly better performance than BKG.

The above results imply that when variables are correlated and the error term is normally distributed, BKG could be a better choice in terms of random effects selection. However, BKG does not work well for fixed effects selection even though BKG was developed for selecting both effects simultaneously. In practice, when the variables are somewhat uncorrelated with each other or the data are not guaranteed to follow the normal distribution, the proposed method can be a reasonable alternative for selecting both random and fixed effects and model prediction.

Table 3.7 compares the computation times for implementing our proposed method and BKG method for each setting and case. We note that the computation time of the new method is substantially shorter than that of BKG in setting 1. For setting 2, the difference is even more significant, especially when m is large. This is a big advantage of the new method.

Table 3.5: Selection and estimation results of both effects for Setting 1

Case#	m	method	Random effect CZ	Random effect IZ	Random effect C	Random effect O	Fixed effect CZ	Fixed effect IZ	Fixed effect C	Fixed effect O	MME	Both
Case1	50	Our	2.55	0.04	63	33	1.91	0.01	90	9	0.0450	57
		BKG	2.30	0.00	40	60	1.89	0.00	89	11	0.0562	37
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0382	100
	100	Our	2.66	0.00	74	26	1.97	0.00	97	3	0.0269	71
		BKG	2.40	0.00	52	48	1.86	0.00	86	14	0.0328	49
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0225	100
	200	Our	2.56	0.00	73	27	1.98	0.00	98	2	0.0100	71
		BKG	2.59	0.00	69	31	1.75	0.00	75	25	0.0245	54
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0088	100
	300	Our	2.49	0.00	61	39	1.99	0.00	99	1	0.0063	60
		BKG	2.78	0.00	82	18	1.79	0.00	79	21	0.0242	63
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0057	100
Case2	50	Our	2.48	0.05	56	39	1.91	0.06	90	7	0.0594	52
		BKG	2.44	0.00	51	49	1.87	0.00	89	11	0.0583	49
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0490	100
	100	Our	2.57	0.00	68	32	1.95	0.03	93	5	0.0311	65
		BKG	2.55	0.00	59	41	1.89	0.00	89	11	0.0357	56
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0275	100
	200	Our	2.37	0.00	61	39	1.98	0.00	98	2	0.0100	59
		BKG	2.64	0.00	70	30	1.87	0.00	87	13	0.0304	63
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0090	100
	300	Our	2.36	0.00	51	49	1.99	0.00	99	1	0.0062	50
		BKG	2.79	0.00	80	20	1.80	0.00	80	20	0.0193	60
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0058	100
Case3	50	Our	2.52	0.04	61	35	1.91	0.08	88	7	0.0606	56
		BKG	2.42	0.00	51	49	1.87	0.00	88	12	0.0619	47
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0490	100
	100	Our	2.52	0.00	66	34	1.97	0.05	95	2	0.0333	64
		BKG	2.50	0.00	57	43	1.89	0.00	90	10	0.0321	54
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0275	100
	200	Our	2.48	0.00	65	35	1.98	0.00	98	2	0.0106	63
		BKG	2.73	0.00	78	22	1.81	0.00	81	19	0.0279	63
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0090	100
	300	Our	2.48	0.00	60	40	1.99	0.00	99	1	0.0063	59
		BKG	2.82	0.00	84	16	1.82	0.00	82	18	0.0245	67
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0058	100
Case4	50	Our	2.72	0.12	66	22	1.98	0.00	98	2	0.0545	65
		BKG	2.18	0.02	35	63	1.90	0.00	90	10	0.0633	33
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0442	100
	100	Our	2.70	0.01	74	25	1.94	0.00	94	6	0.0314	70
		BKG	2.47	0.00	53	47	1.88	0.00	89	11	0.0363	49
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0292	100

Table 3.5 Continued

Case#	m	method	Random effect CZ	Random effect IZ	Random effect C	Random effect O	Fixed effect CZ	Fixed effect IZ	Fixed effect C	Fixed effect O	MME	Both
	200	Our	2.79	0.00	79	21	1.98	0.00	98	2	0.0174	77
		BKG	2.46	0.00	55	45	1.81	0.00	81	19	0.0231	47
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0172	100
	300	Our	2.60	0.00	66	34	1.99	0.00	99	1	0.0073	65
		BKG	2.75	0.00	80	20	1.75	0.00	75	25	0.0212	59
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0083	100
Case5	50	Our	2.43	0.01	60	39	1.90	0.00	90	10	0.0175	54
		BKG	2.38	0.00	41	59	1.71	0.00	74	26	0.0210	31
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0143	100
	100	Our	2.57	0.00	68	32	1.96	0.00	96	4	0.0067	65
		BKG	2.51	0.00	56	44	1.72	0.00	72	28	0.0081	36
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0060	100
	200	Our	2.59	0.00	69	31	1.98	0.00	98	2	0.0042	68
		BKG	2.83	0.00	86	14	1.83	0.00	83	17	0.0063	71
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0037	100
	300	Our	2.57	0.00	64	36	2.00	0.00	100	0	0.0023	64
		BKG	2.49	0.00	51	49	1.80	0.00	80	20	0.0036	33
		Oracle	3.00	0.00	100	0	2.00	0.00	100	0	0.0022	100

Table 3.6: Selection and estimation results of both effects for Setting 2

Case#	m	method	Random effect CZ	Random effect IZ	Random effect C	Random effect O	Fixed effect CZ	Fixed effect IZ	Fixed effect C	Fixed effect O	MME	Both
Case1	50	Our	4.34	0.56	13	44	7.44	0.09	66	28	0.0806	10
		BKG	4.88	0.04	22	74	7.27	0.01	48	51	0.0676	13
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0377	100
	100	Our	4.75	0.22	29	50	7.75	0.02	83	16	0.0270	23
		BKG	4.65	0.02	16	82	7.61	0.00	67	33	0.0268	15
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0181	100
	200	Our	4.78	0.06	44	50	7.91	0.00	92	8	0.0152	41
		BKG	4.63	0.00	16	84	7.42	0.00	51	49	0.0164	7
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0096	100
	300	Our	5.15	0.01	51	48	7.99	0.00	99	1	0.0067	51
		BKG	5.11	0.00	43	57	7.36	0.00	46	54	0.0106	19
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0056	100
Case2	50	Our	4.22	1.13	1	26	7.27	0.09	59	35	0.1027	0
		BKG	4.83	0.07	20	74	7.42	0.03	53	44	0.0907	15
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0529	100
	100	Our	4.45	0.81	2	35	7.51	0.06	71	24	0.0336	2
		BKG	4.79	0.04	25	71	7.56	0.00	65	35	0.0306	20
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0181	100
	200	Our	4.19	0.44	12	45	7.87	0.00	87	13	0.0180	12
		BKG	4.87	0.00	28	72	7.65	0.00	68	32	0.0200	22
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0117	100
	300	Our	4.26	0.25	11	65	7.91	0.00	95	5	0.0066	11
		BKG	5.30	0.00	51	49	7.45	0.00	54	46	0.0099	22
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0055	100
Case3	50	Our	4.52	0.90	6	30	7.27	0.10	55	38	0.1019	4
		BKG	4.79	0.01	20	79	7.50	0.04	60	37	0.0734	12
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0445	100
	100	Our	4.39	0.55	12	40	7.45	0.09	71	22	0.0306	11
		BKG	4.70	0.01	16	83	7.60	0.00	69	31	0.0297	13
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0170	100
	200	Our	4.47	0.28	22	51	7.80	0.00	85	15	0.0171	20
		BKG	4.77	0.00	24	76	7.53	0.00	61	39	0.0175	16
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0127	100
	300	Our	4.78	0.12	32	56	7.96	0.00	96	4	0.0072	31
		BKG	5.27	0.00	52	48	7.41	0.00	53	47	0.0115	30
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0060	100
Case4	50	Our	4.70	0.93	11	36	7.38	0.08	60	35	0.0939	9
		BKG	4.97	0.34	15	66	7.63	0.04	68	28	0.1280	11
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0411	100
	100	Our	4.78	0.40	25	40	7.74	0.01	82	17	0.0310	24
		BKG	4.54	0.19	9	79	7.56	0.00	65	35	0.0406	4
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0229	100

Table 3.6 Continued

Case#	m	method	Random effect CZ	IZ	C	O	Fixed effect CZ	IZ	C	O	MME	Both
	200	Our	4.99	0.12	37	51	7.90	0.00	95	5	0.0177	35
		BKG	4.59	0.06	12	83	7.50	0.00	60	40	0.0209	9
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0131	100
	300	Our	5.02	0.04	49	47	7.94	0.00	94	6	0.0106	47
		BKG	4.67	0.01	23	76	7.24	0.00	45	55	0.0110	10
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0077	100
Case5	50	Our	3.97	0.62	12	49	7.04	0.00	61	39	0.0314	9
		BKG	4.67	0.09	11	82	7.42	0.00	60	40	0.0307	6
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0135	100
	100	Our	4.69	0.29	22	50	7.58	0.00	89	11	0.0085	22
		BKG	4.53	0.03	12	85	7.60	0.00	70	30	0.0104	8
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0064	100
	200	Our	4.64	0.04	31	65	7.78	0.00	91	9	0.0046	30
		BKG	4.71	0.00	24	76	7.42	0.00	53	47	0.0049	12
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0034	100
	300	Our	4.83	0.01	44	55	7.87	0.00	95	5	0.0026	43
		BKG	5.14	0.00	46	54	7.38	0.00	57	43	0.0031	22
		Oracle	6.00	0.00	100	0	8.00	0.00	100	0	0.0016	100

Table 3.7: Comparison of computation times (min/run) for each setting and case

Case#	method	Setting 1				Setting 2			
		m=50	m=100	m=200	m=300	m=50	m=100	m=200	m=300
Case1	Our	0.4	0.8	4.7	8.1	0.8	5.3	12.0	19.6
	BKG	2.9	9.0	29.4	56.4	7.4	22.9	139.1	560.3
Case2	Our	0.4	2.2	5.0	8.0	0.9	5.5	13.9	23.5
	BKG	4.3	11.9	27.9	84.6	9.7	28.1	189.0	506.2
Case3	Our	0.4	2.8	5.8	8.2	0.8	5.4	16.4	19.0
	BKG	3.7	10.0	27.6	82.4	12.0	26.6	193.1	499.6
Case4	Our	0.3	0.7	1.4	7.7	2.5	6.2	11.9	19.6
	BKG	3.4	5.6	26.7	77.2	11.3	17.6	146.8	460.3
Case5	Our	0.3	0.7	4.8	8.9	0.8	1.6	13.0	17.9
	BKG	1.4	3.3	10.4	29.7	6.5	23.9	109.4	286.5

3.4 Real Example

3.4.1 Description

For illustration, we consider a real data example: the Amsterdam Growth and Health Study (Kemper, 1995). The goal of this study is to investigate the relationship between lifestyle and health in adolescence and young adulthood. This data consists of a response variable y, which is the total serum cholesterol measured over six time points, and five explanatory variables x_1 to x_5. In particular, x_1 is fitness level at baseline measured as maximal oxygen uptake on a treadmill, x_2 is body fatness estimated by the sum of the thickness of four skinfolds, x_3 is smoking behavior (0=no,1=yes), x_4 is gender (0=female, 1=male), and x_5 is the measurement time coded as $(1, 2, \ldots, 6)$.

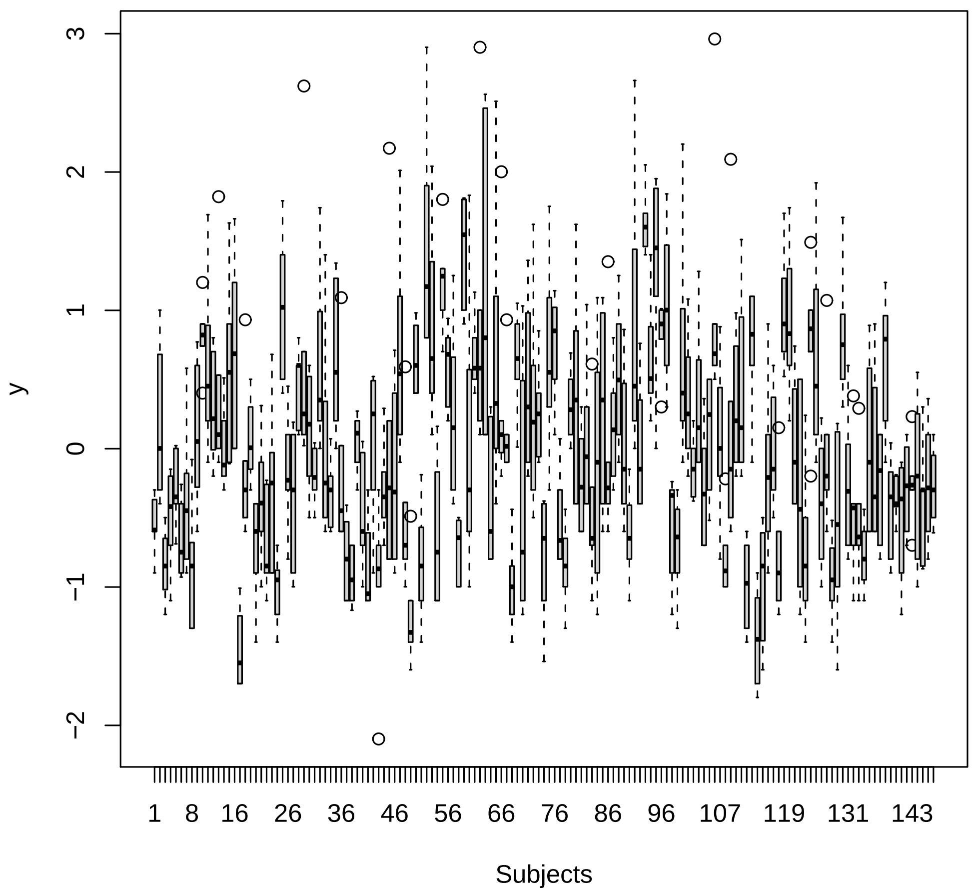


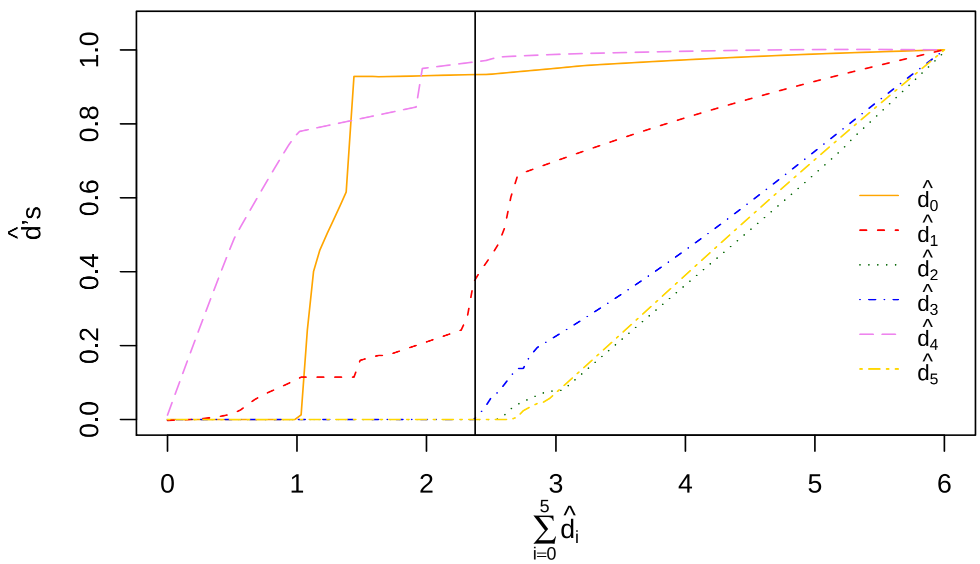
Figure 3.1: Boxplot of a response variable over subjects

The number of subjects is 147, and the total number of observations is 882. Figure 3.1 is the boxplot of the response over 147 subjects. It shows heterogeneity among subjects.

Twisk (2003) analyzed this data using various techniques for longitudinal data, and the dataset is available from his website. Azari et al. (2006) studied this data for selecting a set of fixed effects in a linear regression model, and they included some quadratic and interaction terms as fixed effects. Here we consider the linear mixed effects model for the data with all the five covariates for both fixed and random effects. We center the response variable y and standardize x_1, x_2, and x_5 before fitting the new procedure. Thus, we exclude an intercept from fixed effects of the model, while a random intercept is allowed. The $\mathrm{BIC2}_R$ and $\mathrm{BIC2}_F$ were used to choose the tuning parameters for random and fixed effects, respectively. In Step 4, we used the generalized least squares estimates $\widehat{\boldsymbol{\beta}}_G$ for constructing the weights in (3.2). For comparison, we also fitted the full model by including all explanatory variables as fixed and random effects. We used the `lmer` function from `lme4` package in R, which used the restricted maximum likelihood (REML) estimation based on normality assumption by default.

3.5 Results

In Figure 3.2, we plotted the diagonal elements of $\widehat{\mathbf{D}}$ and $\widehat{\boldsymbol{\Sigma}}$ as functions of $s = \sum_1^q d_i$. The optimal value of s, 2.38, was chosen by using $\mathrm{BIC2}_R$. When $s = q$, all of $\widehat{d_i}$'s are 1 which means all effects are important. The resulting estimates for random effects are presented in Table 3.8. We observe that the new methods selected x_1, x_4 plus a



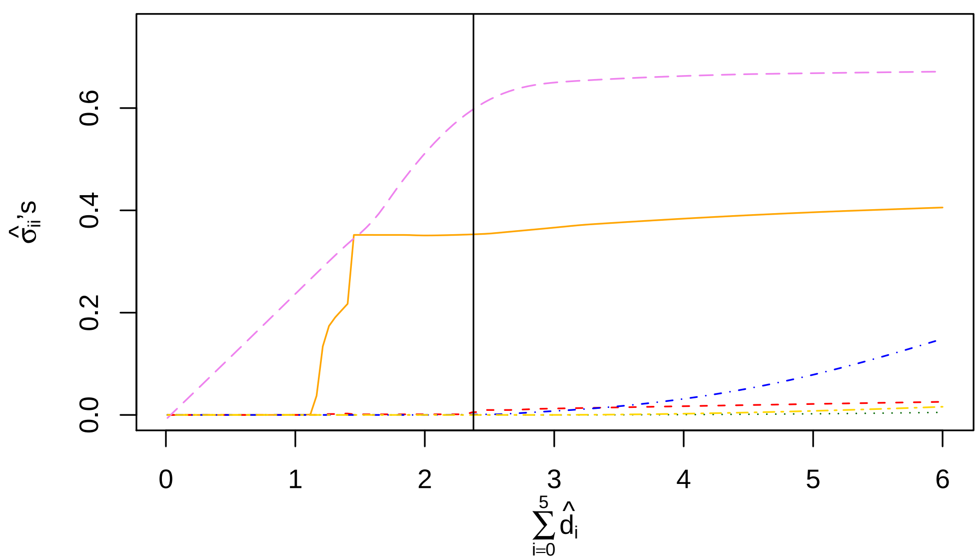


Figure 3.2: Profiles of variance matrix estimates for random effects as $s = \sum_1^q d_i$ is varied. A vertical line is drawn at $s = 2.38$, the optimal value chosen by BIC2_R

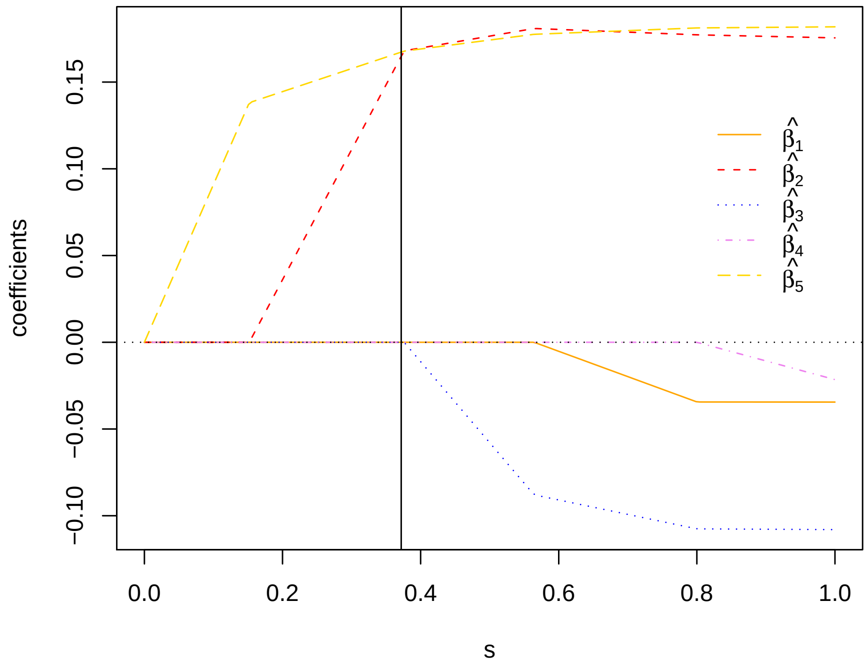


Figure 3.3: Profiles of coefficients for fixed effects for the Amsterdam growth and health study data. A vertical line is drawn at $s = 0.37$, the optimal value chosen by $\mathrm{BIC2}_F$

random intercept as important random effects. Due to a smaller random effects in the final model, the remaining unexplained variance is absorbed into the error variance, and therefore the error variance given by the new method is a little bigger than that in the REML estimation.

Figure 3.3 shows the profiles of coefficient for fixed effects as $s = \sum_1^p |\widehat{\beta}_j|/|\widehat{\beta}_{G,j}|$ ranges from zero to one. Table 3.8 also displays the estimates of fixed effects obtained from our proposed method. The proposed method selected x_2 and x_5, which is con-

Table 3.8: Mixed effects selection results for the Amsterdam Growth and Health Study data: Comparison of estimates from REML and the proposed methods. Standard errors are given in parentheses

Random effect		
Variable	REML estimates	Our estimates
var(γ_{0i})	0.522	0.347
var(γ_{1i})	0.014	0.006
var(γ_{2i})	0.042	0
var(γ_{3i})	0.073	0
var(γ_{4i})	0.961	0.624
var(γ_{5i})	0.037	0
var(ε_i)=σ_ε^2	0.197	0.253
Fixed effect		
Variable	REML estimates	Our estimates
x_1	-0.027 (0.050)	0
x_2	0.190 (0.035)	0.165
x_3	-0.106 (0.060)	0
x_4	0.126 (0.072)	0
x_5	0.169 (0.023)	0.167

sistent with the results from the REML estimation. Specifically, the REML estimates of x_2 and x_5 have t-statistics of 5.50 and 7.26, respectively, which are the only two highly significant fixed effects.

Chapter 4

Computationally Efficient Methods for Random Effects Selection in Linear Mixed Models

4.1 Motivation

In Section 2, we proposed a moment-based method which is used for selecting random effects in linear mixed models. The main step in random effects selection is to minimize a fourth order objective function (2.6). This is a nonlinear programming problem subject to a linear inequality constraint. Some algorithms have been recently developed to solve such an optimization problem. However, most of them are not publicly available, and highly depend on initial values. That is, if the initial value is too far from the true one, it might fail to converge. Moreover, it requires computationally intensive tasks and is hence time-consuming as shown in Table 3.5 and 3.6. In this section, we propose two alternative methods which are computationally efficient and even perform better.

In order to obtain the initial estimates of $\mathbf{\Sigma}$ and σ_ε^2, we minimize L_0 in (2.5), which is a nonlinear semi-definite programming problem. As in Section 2, SeDuMi (Sturm, 1999) can be used for dealing with the semi-definiteness constraint. Given the initial estimates, $Q_R(\mathbf{D})$ is minimized to produce sparse solutions of $\mathbf{D}$. However, the minimization includes the fourth order function in d_i's, which makes the computation slow. Instead of the complicated function, we propose using approximate methods that are quadratic programming problems. The quadratic programming can be easily implemented in some statistical softwares such as Matlab and R. For fixed effects selection, we use the same procedure which proposed in Section 3 and is computationally easy to be implemented. For choosing tuning parameters we recommend BIC2_R and BIC2_F in fixed and random effects selection, respectively.

4.2 Alternative Efficient Methods

4.2.1 Iterative Approximation Method

The basic idea to speed up the exact method is to replace $\mathbf{D\Sigma D}$ in (2.6) with a linear function of d_i's. Let $\mathbf{D}^{(0)}$ be the initial guess for $\mathbf{D}$. In a matrix Taylor expansion, the first order approximation for $\mathbf{D\Sigma D}$ at $\mathbf{D} = \mathbf{D}^{(0)}$ is given by

$$\mathbf{D\Sigma D} \approx \mathbf{D\Sigma D}^{(0)} + \mathbf{D}^{(0)}\mathbf{\Sigma D} - \mathbf{D}^{(0)}\mathbf{\Sigma D}^{(0)}.$$

Algorithm 3 Iterative approximation method

1. (obtain initial estimate of $\boldsymbol{\beta}$):

   Fit a linear regression model $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}$,

   and then obtain an initial estimate $\widetilde{\boldsymbol{\beta}}$.

2. (obtain initial estimates of $\boldsymbol{\Sigma}$ and σ_ε^2):

   Compute $\widetilde{y}_{ijk} = (y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})(y_{ik} - \boldsymbol{x}_{ik}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})$ for i, j, k.

   Obtain $\widetilde{\boldsymbol{\Sigma}}$ by minimizing $L_0(\boldsymbol{\Sigma})$, and compute $\widetilde{\sigma}_\varepsilon^2$.

3. (obtain final estimates of $\boldsymbol{\Sigma}$ and σ_ε^2):

   3.1 Start with $\mathbf{D}^{(0)} = \mathbf{I}_q$

   3.2. For $s = 1, 2, \cdots$, obtain $\mathbf{D}^{(s)}$ by minimizing

   $$\sum_{i=1}^{m}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\left(\widetilde{y}_{ijk} - \boldsymbol{z}_{ij}^T\big(\mathbf{D}\widetilde{\boldsymbol{\Sigma}}\mathbf{D}^{(s-1)} + \mathbf{D}^{(s-1)}\widetilde{\boldsymbol{\Sigma}}\mathbf{D} - \mathbf{D}^{(s-1)}\widetilde{\boldsymbol{\Sigma}}\mathbf{D}^{(s-1)}\big)\boldsymbol{z}_{ik}\right)^2$$
   $$+\lambda\sum_{i=1}^{q} d_i, \qquad \text{subject to all } d_i \geq 0$$

   3.3. Repeat step 3.2 until convergence, and obtain $\widehat{\mathbf{D}}$ in the last iteration

   3.4. Compute the final estimate $\widehat{\boldsymbol{\Sigma}} = \widehat{\mathbf{D}}\widetilde{\boldsymbol{\Sigma}}\widehat{\mathbf{D}}$ and $\widehat{\sigma}_\varepsilon^2$

4. (obtain final estimate of $\boldsymbol{\beta}$):

   Obtain $\widehat{\boldsymbol{\beta}}$ by minimizing $Q_F(\boldsymbol{\beta})$.

Using this approximation, we can replace (2.6) by

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\Big(\widetilde{y}_{ijk} - \boldsymbol{z}_{ij}^T\big(\mathbf{D}\widetilde{\boldsymbol{\Sigma}}\mathbf{D}^{(0)} + \mathbf{D}^{(0)}\widetilde{\boldsymbol{\Sigma}}\mathbf{D} - \mathbf{D}^{(0)}\widetilde{\boldsymbol{\Sigma}}\mathbf{D}^{(0)}\big)\boldsymbol{z}_{ik}\Big)^2 + \lambda\sum_{i=1}^{q} d_i.$$

This is a quadratic function in d_i's. Hence, minimizing can be simply performed by a quadratic programming. Iteratively, we can update $\mathbf{D}$ by using the current estimate in place of $\mathbf{D}^{(0)}$. In other words, at the sth iteration we have

$$\mathbf{D}^{(s)}\boldsymbol{\Sigma}\mathbf{D}^{(s)} \approx \mathbf{D}^{(s)}\boldsymbol{\Sigma}\mathbf{D}^{(s-1)} + \mathbf{D}^{(s-1)}\boldsymbol{\Sigma}\mathbf{D}^{(s)} - \mathbf{D}^{(s-1)}\boldsymbol{\Sigma}\mathbf{D}^{(s-1)}.$$

However, it is not guaranteed that the resulting matrix from the right-hand side is positive definite. Hence, we instead report the matrix on the left-hand side as the resulting expression for the final estimate $\widehat{\boldsymbol{\Sigma}}$. The complete procedure for replacing the minimization of (2.6) is described in Algorithm 3, called the '*iterative approximation method*'.

4.2.2 One-step Approximation Method

In this section, we propose more simplified algorithm which is approximate to the exact method. We set the initial guess for $\mathbf{D}$ as $\mathbf{I}_q$. From the linear approximation for $\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}$ at $\mathbf{D} = \mathbf{I}_q$, we have

$$\mathbf{D}\boldsymbol{\Sigma}\mathbf{D} \approx \mathbf{D}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{D} - \boldsymbol{\Sigma}.$$

Algorithm 4 One-step approximation method

1. (obtain initial estimate of $\boldsymbol{\beta}$):

   Fit a linear regression model $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}$,

   and then obtain an initial estimate $\widetilde{\boldsymbol{\beta}}$.

2. (obtain initial estimates of $\boldsymbol{\Sigma}$ and σ_ε^2):

   Compute $\widetilde{y}_{ijk} = (y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})(y_{ik} - \boldsymbol{x}_{ik}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})$ for i, j, k.

   Obtain $\widetilde{\boldsymbol{\Sigma}}$ by minimizing $L_0(\boldsymbol{\Sigma})$, and compute $\widetilde{\sigma}_\varepsilon^2$.

3. (obtain final estimates of $\boldsymbol{\Sigma}$ and σ_ε^2):

   3.1 Obtain $\widehat{\mathbf{D}}$ by minimizing

   $$\sum_{i,j,k}\Big(\widetilde{y}_{ijk} - \boldsymbol{z}_{ij}^T\big(\mathbf{D}\widetilde{\boldsymbol{\Sigma}} + \widetilde{\boldsymbol{\Sigma}}\mathbf{D} - \widetilde{\boldsymbol{\Sigma}}\big)\boldsymbol{z}_{ik}\Big)^2 + \lambda\sum_{i=1}^{q} d_i \text{ subject to all } d_i \geq 0$$

   3.3. Compute the final estimate $\widehat{\boldsymbol{\Sigma}} = \widehat{\mathbf{D}}\widetilde{\boldsymbol{\Sigma}}\widehat{\mathbf{D}}$ and $\widehat{\sigma}_\varepsilon^2$

4. (obtain final estimate of $\boldsymbol{\beta}$):

   Obtain $\widehat{\boldsymbol{\beta}}$ by minimizing $Q_F(\boldsymbol{\beta})$.

Then, the objective function in (2.6) is changed to

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\Big(\widetilde{y}_{ijk}-\boldsymbol{z}_{ij}^T\big(\mathbf{D}\widetilde{\boldsymbol{\Sigma}}+\widetilde{\boldsymbol{\Sigma}}\mathbf{D}-\widetilde{\boldsymbol{\Sigma}}\big)\boldsymbol{z}_{ik}\Big)^2+\lambda\sum_{i=1}^{q}d_i,$$

with the same constraints $d_i \geq 0$ for all i. We call this procedure the '*one-step approximation method*' as described in Algorithm 4. For the same reason as in the iterative approximation method, we report $\widehat{\mathbf{D}}\widetilde{\boldsymbol{\Sigma}}\widehat{\mathbf{D}}$ as the final estimate of $\boldsymbol{\Sigma}$.

4.3 Simulation Studies

In this section, we compare the performance of the proposed methods and the exact method. We use the same simulation settings as in Section 2. Two settings are considered: One is a simple example and the other is a harder one with a large number of variables. For each setting, we consider five different assumptions about the error term distribution and covariance structure of $\mathbf{X}$. In tables, we denote the proposed method in Section 2 by 'Exact', Iterative approximation method by 'Iterative', and One-step approximation method by 'One-step'. The process times (in seconds) required to implement each method are averaged over 100 data sets.

Tables 4.1-4.5 show the results of random effects selection in Setting 1. In Cases 1, 4, and 5 with independent $\mathbf{X}$, all of three methods have quite similar performance. In Cases 2 and 3 where variables are correlated, One-step approximation method shows even better or similar performance to other methods. Tables 4.6-4.10 give the comparison of Exact method and approximation methods. Similar to Setting 1, One-step approximation method is comparable or slightly better than the others. In Cases

2 and 3, One-step approximation method shows much better performance than the Exact method.

Table 4.1: Random effects selection and estimation results of the exact method and two approximate methods for Setting 1 and Case 1

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}_{\varepsilon}^{2}$	C	U	O
				1	2	3	4				
50	Exact	2.55	0.04	0.141 (0.006)	0.162 (0.007)	1.535 (0.097)	0.483 (0.020)	1.131 (0.039)	63	4	33
	Iterative	2.57	0.02	0.139 (0.006)	0.159 (0.007)	1.554 (0.108)	0.458 (0.019)	1.037 (0.039)	66	2	32
	One-step	2.54	0.01	0.140 (0.006)	0.161 (0.007)	1.539 (0.101)	0.476 (0.018)	1.097 (0.042)	70	1	29
	Oracle	3.00	0.00	0.125 (0.005)	0.145 (0.006)	1.095 (0.052)	0.457 (0.020)	1.001 (0.036)	100	0	0
100	Exact	2.66	0.00	0.105 (0.005)	0.122 (0.005)	1.102 (0.075)	0.366 (0.015)	1.153 (0.031)	74	0	26
	Iterative	2.65	0.00	0.105 (0.005)	0.120 (0.005)	1.129 (0.078)	0.353 (0.015)	1.067 (0.030)	70	0	30
	One-step	2.65	0.00	0.105 (0.005)	0.122 (0.005)	1.110 (0.076)	0.367 (0.015)	1.151 (0.031)	74	0	26
	Oracle	3.00	0.00	0.092 (0.004)	0.108 (0.005)	0.810 (0.042)	0.340 (0.016)	1.050 (0.028)	100	0	0
200	Exact	2.56	0.00	0.082 (0.004)	0.094 (0.004)	0.915 (0.066)	0.274 (0.012)	1.130 (0.023)	73	0	27
	Iterative	2.54	0.00	0.082 (0.004)	0.093 (0.004)	0.937 (0.064)	0.267 (0.012)	1.052 (0.021)	66	0	34
	One-step	2.59	0.00	0.083 (0.004)	0.095 (0.004)	0.914 (0.065)	0.276 (0.012)	1.132 (0.023)	75	0	25
	Oracle	3.00	0.00	0.070 (0.003)	0.081 (0.004)	0.613 (0.032)	0.254 (0.011)	1.033 (0.019)	100	0	0
300	Exact	2.49	0.00	0.066 (0.003)	0.076 (0.003)	0.722 (0.038)	0.226 (0.009)	1.110 (0.021)	61	0	39
	Iterative	2.66	0.00	0.064 (0.003)	0.072 (0.003)	0.668 (0.039)	0.218 (0.009)	1.060 (0.019)	70	0	30
	One-step	2.54	0.00	0.067 (0.003)	0.076 (0.003)	0.716 (0.039)	0.227 (0.010)	1.109 (0.022)	65	0	35
	Oracle	3.00	0.00	0.054 (0.002)	0.062 (0.003)	0.466 (0.024)	0.198 (0.009)	1.019 (0.016)	100	0	0

Table 4.2: Random effects selection and estimation results of the exact method and two approximate methods for Setting 1 and Case 2

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}_{\varepsilon}^{2}$	C	U	O
				1	2	3	4				
50	Exact	2.48	0.05	0.178	0.204	2.021	0.594	1.091	56	5	39
				(0.008)	(0.009)	(0.145)	(0.025)	(0.052)			
	Iterative	2.44	0.06	0.188	0.213	2.227	0.611	0.958	55	6	39
				(0.009)	(0.010)	(0.176)	(0.025)	(0.050)			
	One-step	2.49	0.03	0.177	0.203	2.011	0.597	1.041	61	3	36
				(0.008)	(0.009)	(0.154)	(0.024)	(0.048)			
	Oracle	3.00	0.00	0.147	0.172	1.268	0.553	0.978	100	0	0
				(0.007)	(0.009)	(0.065)	(0.028)	(0.044)			
100	Exact	2.57	0.00	0.133	0.154	1.442	0.464	1.100	68	0	32
				(0.007)	(0.008)	(0.114)	(0.025)	(0.037)			
	Iterative	2.56	0.00	0.141	0.162	1.582	0.474	0.989	70	0	30
				(0.008)	(0.009)	(0.131)	(0.025)	(0.039)			
	One-step	2.62	0.00	0.131	0.152	1.396	0.459	1.082	73	0	27
				(0.007)	(0.008)	(0.112)	(0.025)	(0.038)			
	Oracle	3.00	0.00	0.111	0.131	0.954	0.421	1.036	100	0	0
				(0.006)	(0.007)	(0.049)	(0.024)	(0.035)			
200	Exact	2.37	0.00	0.105	0.118	1.294	0.338	1.070	61	0	39
				(0.006)	(0.006)	(0.124)	(0.016)	(0.030)			
	Iterative	2.52	0.00	0.105	0.119	1.239	0.346	1.000	68	0	32
				(0.006)	(0.006)	(0.118)	(0.015)	(0.028)			
	One-step	2.50	0.00	0.101	0.115	1.206	0.333	1.050	72	0	28
				(0.006)	(0.006)	(0.121)	(0.015)	(0.028)			
	Oracle	3.00	0.00	0.081	0.094	0.701	0.305	1.022	100	0	0
				(0.004)	(0.004)	(0.034)	(0.014)	(0.023)			
300	Exact	2.36	0.00	0.082	0.092	0.942	0.268	1.057	51	0	49
				(0.003)	(0.004)	(0.061)	(0.011)	(0.024)			
	Iterative	2.64	0.00	0.082	0.093	0.891	0.273	1.015	73	0	27
				(0.004)	(0.004)	(0.067)	(0.011)	(0.022)			
	One-step	2.67	0.00	0.079	0.091	0.848	0.269	1.056	78	0	22
				(0.004)	(0.004)	(0.066)	(0.012)	(0.023)			
	Oracle	3.00	0.00	0.066	0.076	0.570	0.238	1.006	100	0	0
				(0.003)	(0.003)	(0.026)	(0.011)	(0.019)			

Table 4.3: Random effects selection and estimation results of the exact method and two approximate methods for Setting 1 and Case 3

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}_{\varepsilon}^2$	C	U	O
				1	2	3	4				
50	Exact	2.52	0.04	0.169	0.193	1.900	0.568	1.064	61	4	35
				(0.007)	(0.008)	(0.137)	(0.022)	(0.049)			
	Iterative	2.57	0.03	0.170	0.193	1.942	0.561	0.986	65	3	32
				(0.008)	(0.009)	(0.154)	(0.023)	(0.047)			
	One-step	2.47	0.03	0.170	0.195	1.953	0.570	1.047	60	3	37
				(0.007)	(0.008)	(0.139)	(0.022)	(0.049)			
	Oracle	3.00	0.00	0.147	0.172	1.268	0.553	0.978	100	0	0
				(0.007)	(0.009)	(0.065)	(0.028)	(0.044)			
100	Exact	2.52	0.00	0.132	0.152	1.459	0.457	1.078	66	0	34
				(0.007)	(0.008)	(0.114)	(0.024)	(0.038)			
	Iterative	2.67	0.00	0.127	0.146	1.345	0.440	1.036	73	0	27
				(0.007)	(0.008)	(0.102)	(0.023)	(0.036)			
	One-step	2.61	0.00	0.129	0.150	1.385	0.452	1.104	75	0	25
				(0.006)	(0.007)	(0.107)	(0.024)	(0.039)			
	Oracle	3.00	0.00	0.111	0.131	0.954	0.421	1.036	100	0	0
				(0.006)	(0.007)	(0.049)	(0.024)	(0.035)			
200	Exact	2.48	0.00	0.098	0.112	1.123	0.330	1.078	65	0	35
				(0.005)	(0.005)	(0.092)	(0.016)	(0.026)			
	Iterative	2.47	0.00	0.101	0.113	1.201	0.325	1.009	65	0	35
				(0.005)	(0.006)	(0.099)	(0.016)	(0.026)			
	One-step	2.47	0.00	0.099	0.113	1.162	0.331	1.078	68	0	32
				(0.005)	(0.005)	(0.098)	(0.015)	(0.027)			
	Oracle	3.00	0.00	0.081	0.094	0.701	0.305	1.022	100	0	0
				(0.004)	(0.004)	(0.034)	(0.014)	(0.023)			
300	Exact	2.48	0.00	0.077	0.088	0.839	0.260	1.070	60	0	40
				(0.003)	(0.004)	(0.051)	(0.011)	(0.023)			
	Iterative	2.65	0.00	0.079	0.090	0.838	0.265	0.992	72	0	28
				(0.004)	(0.004)	(0.057)	(0.012)	(0.024)			
	One-step	2.67	0.00	0.078	0.090	0.825	0.267	1.087	78	0	22
				(0.003)	(0.004)	(0.062)	(0.012)	(0.024)			
	Oracle	3.00	0.00	0.066	0.076	0.570	0.238	1.006	100	0	0
				(0.003)	(0.003)	(0.026)	(0.011)	(0.019)			

Table 4.4: Random effects selection and estimation results of the exact method and two approximate methods for Setting 1 and Case 4

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}_{\varepsilon}^{2}$	C	U	O
				1	2	3	4				
50	Exact	2.72	0.12	0.159	0.185	1.597	0.572	1.833	66	12	22
				(0.007)	(0.008)	(0.086)	(0.026)	(0.051)			
	Iterative	2.74	0.13	0.157	0.181	1.595	0.551	1.722	65	13	22
				(0.007)	(0.008)	(0.092)	(0.026)	(0.052)			
	One-step	2.57	0.03	0.163	0.192	1.779	0.570	1.782	70	3	27
				(0.009)	(0.010)	(0.145)	(0.029)	(0.056)			
	Oracle	3.00	0.00	0.132	0.154	1.157	0.486	1.618	100	0	0
				(0.007)	(0.008)	(0.063)	(0.027)	(0.043)			
100	Exact	2.70	0.01	0.112	0.130	1.155	0.391	1.780	74	1	25
				(0.005)	(0.006)	(0.073)	(0.018)	(0.039)			
	Iterative	2.68	0.01	0.114	0.131	1.218	0.386	1.684	72	1	27
				(0.006)	(0.006)	(0.081)	(0.018)	(0.040)			
	One-step	2.80	0.00	0.111	0.131	1.046	0.419	1.789	84	0	16
				(0.005)	(0.006)	(0.062)	(0.019)	(0.037)			
	Oracle	3.00	0.00	0.098	0.115	0.844	0.371	1.659	100	0	0
				(0.005)	(0.005)	(0.042)	(0.019)	(0.032)			
200	Exact	2.79	0.00	0.084	0.099	0.786	0.313	1.822	79	0	21
				(0.003)	(0.004)	(0.038)	(0.013)	(0.029)			
	Iterative	2.78	0.00	0.081	0.095	0.799	0.291	1.744	79	0	21
				(0.003)	(0.004)	(0.045)	(0.012)	(0.030)			
	One-step	2.75	0.00	0.078	0.091	0.782	0.282	1.749	82	0	18
				(0.004)	(0.004)	(0.054)	(0.013)	(0.027)			
	Oracle	3.00	0.00	0.073	0.087	0.643	0.274	1.646	100	0	0
				(0.003)	(0.004)	(0.032)	(0.013)	(0.025)			
300	Exact	2.60	0.00	0.073	0.084	0.769	0.249	1.742	66	0	34
				(0.003)	(0.004)	(0.045)	(0.011)	(0.027)			
	Iterative	2.71	0.00	0.071	0.080	0.738	0.237	1.686	73	0	27
				(0.003)	(0.004)	(0.044)	(0.011)	(0.025)			
	One-step	2.73	0.00	0.063	0.073	0.624	0.225	1.743	76	0	24
				(0.003)	(0.003)	(0.038)	(0.010)	(0.025)			
	Oracle	3.00	0.00	0.053	0.061	0.462	0.196	1.655	100	0	0
				(0.002)	(0.003)	(0.022)	(0.010)	(0.020)			

Table 4.5: Random effects selection and estimation results of the exact method and two approximate methods for Setting 1 and Case 5

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}_{\varepsilon}^{2}$	C	U	O
				1	2	3	4				
50	Exact	2.43	0.01	0.148	0.170	1.692	0.503	1.080	60	1	39
				(0.008)	(0.009)	(0.112)	(0.028)	(0.041)			
	Iterative	2.54	0.01	0.147	0.168	1.642	0.489	1.003	63	1	36
				(0.008)	(0.009)	(0.105)	(0.028)	(0.039)			
	One-step	2.49	0.01	0.147	0.169	1.640	0.503	1.080	63	1	36
				(0.008)	(0.009)	(0.108)	(0.029)	(0.041)			
	Oracle	3.00	0.00	0.132	0.155	1.150	0.486	1.010	100	0	0
				(0.009)	(0.010)	(0.086)	(0.032)	(0.033)			
100	Exact	2.57	0.00	0.110	0.127	1.172	0.385	1.104	68	0	32
				(0.005)	(0.006)	(0.076)	(0.018)	(0.028)			
	Iterative	2.57	0.00	0.111	0.125	1.203	0.374	1.020	63	0	37
				(0.005)	(0.005)	(0.075)	(0.017)	(0.027)			
	One-step	2.53	0.00	0.110	0.126	1.190	0.384	1.083	66	0	34
				(0.005)	(0.006)	(0.081)	(0.017)	(0.028)			
	Oracle	3.00	0.00	0.097	0.113	0.835	0.370	1.002	100	0	0
				(0.005)	(0.006)	(0.047)	(0.020)	(0.024)			
200	Exact	2.59	0.00	0.076	0.088	0.793	0.268	1.112	69	0	31
				(0.003)	(0.004)	(0.047)	(0.013)	(0.021)			
	Iterative	2.67	0.00	0.075	0.085	0.772	0.260	1.041	69	0	31
				(0.003)	(0.004)	(0.044)	(0.012)	(0.020)			
	One-step	2.59	0.00	0.076	0.087	0.783	0.270	1.104	69	0	31
				(0.003)	(0.004)	(0.047)	(0.013)	(0.022)			
	Oracle	3.00	0.00	0.067	0.078	0.589	0.247	1.007	100	0	0
				(0.003)	(0.004)	(0.030)	(0.012)	(0.017)			
300	Exact	2.57	0.00	0.065	0.075	0.689	0.220	1.101	64	0	36
				(0.003)	(0.003)	(0.038)	(0.010)	(0.019)			
	Iterative	2.71	0.00	0.062	0.071	0.644	0.209	1.056	74	0	26
				(0.003)	(0.003)	(0.039)	(0.009)	(0.018)			
	One-step	2.61	0.00	0.064	0.074	0.677	0.221	1.110	67	0	33
				(0.003)	(0.003)	(0.040)	(0.009)	(0.020)			
	Oracle	3.00	0.00	0.052	0.061	0.464	0.188	1.011	100	0	0
				(0.003)	(0.003)	(0.025)	(0.009)	(0.014)			

Table 4.6: Random effects selection and estimation results of the exact method and two approximate methods for Setting 2 and Case 1

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}_{\varepsilon}^{2}$	C	U	O
				1	2	3	4				
50	Exact	4.34	0.56	0.133	0.155	5.679	0.682	1.307	13	43	44
				(0.004)	(0.004)	(0.370)	(0.021)	(0.083)			
	Iterative	4.73	0.43	0.136	0.154	5.806	0.640	0.885	18	35	47
				(0.004)	(0.004)	(0.344)	(0.020)	(0.065)			
	One-step	4.32	0.32	0.132	0.153	5.905	0.657	1.279	23	24	53
				(0.004)	(0.004)	(0.426)	(0.020)	(0.082)			
	Oracle	6.00	0.00	0.121	0.137	3.922	0.637	0.919	100	0	0
				(0.004)	(0.005)	(0.128)	(0.028)	(0.046)			
100	Exact	4.75	0.22	0.099	0.116	3.999	0.533	1.314	29	21	50
				(0.003)	(0.003)	(0.263)	(0.019)	(0.060)			
	Iterative	5.11	0.17	0.100	0.113	4.033	0.498	0.987	37	16	47
				(0.003)	(0.003)	(0.230)	(0.020)	(0.051)			
	One-step	4.79	0.10	0.098	0.115	3.972	0.520	1.315	41	10	49
				(0.003)	(0.003)	(0.253)	(0.018)	(0.063)			
	Oracle	6.00	0.00	0.089	0.101	2.883	0.476	1.013	100	0	0
				(0.003)	(0.003)	(0.082)	(0.023)	(0.041)			
200	Exact	4.78	0.06	0.073	0.085	2.930	0.388	1.225	44	6	50
				(0.002)	(0.002)	(0.135)	(0.014)	(0.040)			
	Iterative	5.06	0.02	0.075	0.085	3.122	0.363	0.973	47	2	51
				(0.002)	(0.002)	(0.148)	(0.012)	(0.040)			
	One-step	5.02	0.02	0.072	0.085	2.772	0.392	1.269	51	2	47
				(0.001)	(0.002)	(0.129)	(0.012)	(0.039)			
	Oracle	6.00	0.00	0.065	0.074	2.114	0.347	1.029	100	0	0
				(0.002)	(0.002)	(0.048)	(0.013)	(0.029)			
300	Exact	5.15	0.01	0.059	0.068	2.209	0.314	1.216	51	1	48
				(0.001)	(0.002)	(0.107)	(0.011)	(0.030)			
	Iterative	5.28	0.00	0.060	0.068	2.344	0.292	1.042	55	0	45
				(0.002)	(0.002)	(0.104)	(0.010)	(0.031)			
	One-step	5.30	0.00	0.059	0.069	2.173	0.317	1.239	66	0	34
				(0.001)	(0.002)	(0.103)	(0.011)	(0.029)			
	Oracle	6.00	0.00	0.053	0.060	1.693	0.284	1.055	100	0	0
				(0.001)	(0.002)	(0.045)	(0.010)	(0.024)			

Table 4.7: Random effects selection and estimation results of the exact method and two approximate methods for Setting 2 and Case 2

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}_{\varepsilon}^2$	C	U	O
				1	2	3	4				
50	Exact	4.22	1.13	0.181	0.211	7.918	0.870	1.272	1	73	26
				(0.008)	(0.008)	(0.723)	(0.025)	(0.080)			
	Iterative	4.54	1.13	0.196	0.228	8.445	0.934	0.908	0	69	31
				(0.008)	(0.009)	(0.739)	(0.029)	(0.080)			
	One-step	3.87	0.65	0.188	0.217	9.360	0.847	1.034	4	47	49
				(0.008)	(0.008)	(0.807)	(0.027)	(0.087)			
	Oracle	6.00	0.00	0.157	0.176	5.065	0.815	0.796	100	0	0
				(0.006)	(0.007)	(0.182)	(0.040)	(0.051)			
100	Exact	4.45	0.81	0.138	0.162	5.682	0.709	1.243	2	63	35
				(0.004)	(0.005)	(0.429)	(0.017)	(0.076)			
	Iterative	4.68	0.70	0.158	0.183	6.924	0.761	0.882	7	56	37
				(0.007)	(0.007)	(0.595)	(0.022)	(0.070)			
	One-step	4.26	0.43	0.146	0.169	6.947	0.695	1.082	20	35	45
				(0.006)	(0.006)	(0.643)	(0.019)	(0.077)			
	Oracle	6.00	0.00	0.123	0.137	3.986	0.638	0.927	100	0	0
				(0.004)	(0.005)	(0.137)	(0.024)	(0.049)			
200	Exact	4.19	0.44	0.109	0.127	4.623	0.579	1.124	12	43	45
				(0.003)	(0.004)	(0.323)	(0.015)	(0.057)			
	Iterative	4.77	0.42	0.119	0.137	5.007	0.598	0.844	19	37	44
				(0.004)	(0.004)	(0.384)	(0.019)	(0.056)			
	One-step	4.91	0.25	0.104	0.122	4.098	0.565	1.155	36	23	41
				(0.003)	(0.004)	(0.322)	(0.016)	(0.055)			
	Oracle	6.00	0.00	0.093	0.105	2.935	0.495	1.017	100	0	0
				(0.003)	(0.003)	(0.086)	(0.017)	(0.065)			
300	Exact	4.26	0.25	0.091	0.106	3.932	0.468	1.099	11	24	65
				(0.003)	(0.003)	(0.258)	(0.017)	(0.043)			
	Iterative	4.69	0.18	0.099	0.113	4.450	0.467	0.818	37	17	46
				(0.004)	(0.004)	(0.319)	(0.016)	(0.046)			
	One-step	4.92	0.14	0.087	0.102	3.509	0.462	1.109	42	13	45
				(0.003)	(0.003)	(0.262)	(0.016)	(0.043)			
	Oracle	6.00	0.00	0.082	0.094	2.569	0.431	1.175	100	0	0
				(0.003)	(0.004)	(0.080)	(0.018)	(0.086)			

Table 4.8: Random effects selection and estimation results of the exact method and two approximate methods for Setting 2 and Case 3

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}_{\varepsilon}^2$	C	U	O
				1	2	3	4				
50	Exact	4.52	0.90	0.168	0.196	7.092	0.832	1.333	6	64	30
				(0.006)	(0.006)	(0.580)	(0.022)	(0.084)			
	Iterative	4.84	0.69	0.178	0.203	7.603	0.826	0.927	11	53	36
				(0.007)	(0.007)	(0.611)	(0.024)	(0.076)			
	One-step	4.27	0.53	0.170	0.195	7.910	0.799	1.203	8	39	53
				(0.007)	(0.007)	(0.645)	(0.023)	(0.089)			
	Oracle	6.00	0.00	0.152	0.171	4.937	0.816	0.822	100	0	0
				(0.005)	(0.005)	(0.153)	(0.033)	(0.050)			
100	Exact	4.39	0.55	0.135	0.157	6.053	0.678	1.215	12	48	40
				(0.006)	(0.006)	(0.581)	(0.019)	(0.074)			
	Iterative	4.90	0.38	0.139	0.157	6.045	0.662	0.871	24	37	39
				(0.006)	(0.006)	(0.534)	(0.019)	(0.062)			
	One-step	4.18	0.23	0.140	0.160	6.995	0.649	1.094	29	21	50
				(0.006)	(0.006)	(0.642)	(0.020)	(0.077)			
	Oracle	6.00	0.00	0.116	0.131	3.755	0.615	0.942	100	0	0
				(0.004)	(0.004)	(0.122)	(0.024)	(0.048)			
200	Exact	4.47	0.28	0.101	0.116	4.264	0.527	1.116	22	27	51
				(0.003)	(0.003)	(0.307)	(0.017)	(0.055)			
	Iterative	5.05	0.14	0.101	0.114	4.212	0.489	0.878	41	14	45
				(0.003)	(0.004)	(0.268)	(0.017)	(0.050)			
	One-step	4.50	0.12	0.100	0.115	4.530	0.503	1.109	39	11	50
				(0.003)	(0.004)	(0.351)	(0.016)	(0.058)			
	Oracle	6.00	0.00	0.086	0.097	2.753	0.455	0.966	100	0	0
				(0.002)	(0.003)	(0.074)	(0.016)	(0.040)			
300	Exact	4.78	0.12	0.079	0.091	3.118	0.412	(1.146)	32	12	56
				(0.002)	(0.003)	(0.184)	(0.015)	(0.040)			
	Iterative	5.39	0.09	0.080	0.090	3.034	0.392	0.971	53	9	38
				(0.002)	(0.003)	(0.164)	(0.014)	(0.037)			
	One-step	5.07	0.04	0.078	0.090	3.043	0.408	1.174	52	4	44
				(0.002)	(0.003)	(0.207)	(0.013)	(0.040)			
	Oracle	6.00	0.00	0.073	0.083	2.322	0.382	1.060	100	0	0
				(0.002)	(0.003)	(0.062)	(0.014)	(0.049)			

Table 4.9: Random effects selection and estimation results of the exact method and two approximate methods for Setting 2 and Case 4

m	Criterion	CZ	IZ	Error 1	Error 2	Error 3	Error 4	$\widehat{\sigma}_{\varepsilon}^2$	C	U	O
50	Exact	4.70	0.93	0.144	0.171	5.521	0.755	2.044	11	53	36
				(0.004)	(0.005)	(0.369)	(0.024)	(0.111)			
	Iterative	4.80	0.78	0.149	0.171	5.995	0.725	1.545	8	49	43
				(0.004)	(0.005)	(0.366)	(0.024)	(0.108)			
	One-step	3.85	0.46	0.150	0.173	6.987	0.713	1.669	9	33	58
				(0.005)	(0.006)	(0.468)	(0.025)	(0.117)			
	Oracle	6.00	0.00	0.133	0.150	4.257	0.705	1.555	100	0	0
				(0.005)	(0.006)	(0.157)	(0.034)	(0.066)			
100	Exact	4.78	0.40	0.107	0.126	4.180	0.577	2.009	25	35	40
				(0.003)	(0.003)	(0.283)	(0.016)	(0.078)			
	Iterative	5.23	0.31	0.105	0.121	4.001	0.538	1.674	35	27	38
				(0.003)	(0.003)	(0.229)	(0.017)	(0.069)			
	One-step	4.62	0.14	0.107	0.126	4.363	0.571	1.826	36	12	52
				(0.003)	(0.004)	(0.245)	(0.025)	(0.077)			
	Oracle	6.00	0.00	0.093	0.106	2.991	0.498	1.706	100	0	0
				(0.002)	(0.003)	(0.085)	(0.017)	(0.046)			
200	Exact	4.99	0.12	0.081	0.094	3.155	0.436	1.796	37	12	51
				(0.003)	(0.003)	(0.205)	(0.016)	(0.054)			
	Iterative	5.17	0.07	0.084	0.094	3.349	0.413	1.522	44	7	49
				(0.003)	(0.003)	(0.195)	(0.015)	(0.050)			
	One-step	4.89	0.03	0.077	0.090	3.036	0.415	1.870	39	3	58
				(0.002)	(0.002)	(0.169)	(0.014)	(0.047)			
	Oracle	6.00	0.00	0.069	0.078	2.239	0.366	1.639	100	0	0
				(0.002)	(0.002)	(0.064)	(0.012)	(0.033)			
300	Exact	5.02	0.04	0.063	0.072	2.452	0.322	1.856	49	4	47
				(0.001)	(0.002)	(0.109)	(0.012)	(0.037)			
	Iterative	5.11	0.01	0.067	0.074	2.719	0.306	1.616	46	1	53
				(0.002)	(0.002)	(0.119)	(0.010)	(0.037)			
	One-step	5.21	0.00	0.064	0.075	2.407	0.346	1.881	59	0	41
				(0.002)	(0.002)	(0.149)	(0.010)	(0.044)			
	Oracle	6.00	0.00	0.055	0.063	1.767	0.295	1.664	100	0	0
				(0.002)	(0.002)	(0.050)	(0.011)	(0.031)			

Table 4.10: Random effects selection and estimation results of the exact method and two approximate methods for Setting 2 and Case 5

m	Criterion	CZ	IZ	Error				$\widehat{\sigma}_{\varepsilon}^2$	C	U	O
				1	2	3	4				
50	Exact	3.97	0.62	0.141 (0.005)	0.164 (0.005)	6.396 (0.469)	0.709 (0.022)	1.235 (0.087)	12	39	49
	Iterative	4.46	0.50	0.143 (0.005)	0.165 (0.005)	6.319 (0.404)	0.689 (0.024)	0.902 (0.080)	14	36	50
	One-step	3.78	0.44	0.144 (0.005)	0.168 (0.005)	6.887 (0.508)	0.701 (0.022)	1.214 (0.102)	9	30	61
	Oracle	6.00	0.00	0.118 (0.004)	0.136 (0.005)	3.716 (0.117)	0.650 (0.028)	1.022 (0.052)	100	0	0
100	Exact	4.69	0.29	0.099 (0.003)	0.117 (0.003)	3.916 (0.232)	0.527 (0.020)	1.244 (0.063)	22	28	50
	Iterative	4.98	0.18	0.102 (0.003)	0.116 (0.003)	4.080 (0.199)	0.492 (0.019)	0.890 (0.059)	31	17	52
	One-step	4.44	0.08	0.099 (0.003)	0.116 (0.003)	4.211 (0.238)	0.509 (0.018)	1.155 (0.066)	32	7	61
	Oracle	6.00	0.00	0.088 (0.003)	0.101 (0.003)	2.805 (0.079)	0.481 (0.020)	1.004 (0.042)	100	0	0
200	Exact	4.64	0.04	0.074 (0.002)	0.085 (0.002)	3.065 (0.160)	0.376 (0.013)	1.116 (0.045)	31	4	65
	Iterative	4.97	0.04	0.077 (0.002)	0.087 (0.002)	3.181 (0.154)	0.363 (0.014)	0.898 (0.042)	35	4	61
	One-step	4.58	0.02	0.076 (0.002)	0.087 (0.002)	3.214 (0.186)	0.382 (0.013)	1.112 (0.048)	46	2	52
	Oracle	6.00	0.00	0.065 (0.002)	0.074 (0.002)	2.074 (0.063)	0.355 (0.015)	0.975 (0.034)	100	0	0
300	Exact	4.83	0.01	0.060 (0.002)	0.068 (0.002)	2.436 (0.137)	0.299 (0.009)	1.119 (0.037)	44	1	55
	Iterative	5.09	0.01	0.062 (0.002)	0.069 (0.002)	2.549 (0.123)	0.290 (0.009)	0.949 (0.036)	43	1	56
	One-step	4.83	0.00	0.060 (0.002)	0.069 (0.002)	2.489 (0.155)	0.305 (0.010)	1.112 (0.040)	50	0	50
	Oracle	6.00	0.00	0.053 (0.001)	0.060 (0.002)	1.707 (0.050)	0.287 (0.009)	0.985 (0.027)	100	0	0

Tables 4.11 and 4.12 provide the results of fixed effects selection. As recommended in Section 3, we used the adaptive lasso penalty with the weight $w = 1/|\widehat{\beta}_G|$. In most cases, the three methods have almost the same results in both settings. In setting 2, One-step approximation method has much better performance in Cases 2 and 3, and it is due to better random-effects selection.

In conclusion, the one-step approximation method has a good alternative to the exact method which has better or similar performance. Though the iterative approximation method repeated until $\widehat{\mathbf{D}}$ converges, it does not give better selection and estimation for random effects. In terms of computation speed, the exact method can be implemented faster than BKG method as seen in Section 3. However, the two approximate methods are executed even faster than the exact method because the objective function is of second order. Even when variables are somewhat correlated, the approximate methods work well.

Table 4.11: Fixed effects selection and estimation results of the exact method and two approximate methods for Setting 1

Case#	m	method	CZ	IZ	MME	C	U	O	Both	Time (sec/run)
Case 1	50	Exact	1.91	0.01	0.045	90	1	9	57	21.3
		Iterative	1.91	0.02	0.048	89	2	9	61	6.5
		One-step	1.91	0.01	0.046	90	1	9	64	2.0
		Oracle	2.00	0.00	0.038	100	0	0	100	-
	100	Exact	1.97	0.00	0.027	97	0	3	71	45.7
		Iterative	1.97	0.00	0.025	97	0	3	67	10.9
		One-step	1.98	0.00	0.027	98	0	2	73	2.0
		Oracle	2.00	0.00	0.022	100	0	0	100	-
	200	Exact	1.98	0.00	0.010	98	0	2	71	281.7
		Iterative	1.98	0.00	0.010	98	0	2	65	20.3
		One-step	1.98	0.00	0.010	98	0	2	73	2.4
		Oracle	2.00	0.00	0.009	100	0	0	100	-
	300	Exact	1.99	0.00	0.006	99	0	1	60	486.9
		Iterative	1.99	0.00	0.006	99	0	1	69	30.0
		One-step	1.99	0.00	0.006	99	0	1	64	3.0
		Oracle	2.00	0.00	0.006	100	0	0	100	-
Case 2	50	Exact	1.91	0.06	0.059	90	3	7	52	22.8
		Iterative	1.89	0.10	0.066	84	6	10	49	20.8
		One-step	1.91	0.06	0.060	90	3	7	56	1.4
		Oracle	2.00	0.00	0.049	100	0	0	100	-
	100	Exact	1.95	0.03	0.031	93	2	5	65	134.9
		Iterative	1.97	0.07	0.029	93	4	3	68	12.0
		One-step	1.95	0.03	0.032	93	2	5	70	2.1
		Oracle	2.00	0.00	0.028	100	0	0	100	-
	200	Exact	1.98	0.00	0.010	98	0	2	59	301.5
		Iterative	1.98	0.00	0.010	98	0	2	66	21.4
		One-step	1.98	0.00	0.010	98	0	2	70	2.4
		Oracle	2.00	0.00	0.009	100	0	0	100	-
	300	Exact	1.99	0.00	0.006	99	0	1	50	479.7
		Iterative	1.99	0.00	0.006	99	0	1	72	30.3
		One-step	1.99	0.00	0.006	99	0	1	77	3.2
		Oracle	2.00	0.00	0.010	100	0	0	100	-
Case 3	50	Exact	1.91	0.08	0.061	88	5	7	56	21.3
		Iterative	1.91	0.10	0.057	87	6	7	61	6.1
		One-step	1.91	0.09	0.060	88	5	7	55	1.4
		Oracle	2.00	0.00	0.049	100	0	0	100	-
	100	Exact	1.97	0.05	0.033	95	3	2	64	166..7
		Iterative	1.95	0.05	0.030	93	3	4	70	11.3
		One-step	1.96	0.03	0.032	95	2	3	72	1.8
		Oracle	2.00	0.00	0.028	100	0	0	100	-
	200	Exact	1.98	0.00	0.011	98	0	2	63	345.6

Table 4.11 Continued

Case#	m	method	CZ	IZ	MME	C	U	O	Both	Time (sec/run)
		Iterative	1.97	0.00	0.011	97	0	3	63	20.9
		One-step	1.98	0.00	0.011	98	0	2	67	2.4
		Oracle	2.00	0.00	0.009	100	0	0	100	-
	300	Exact	1.99	0.00	0.006	99	0	1	59	491.3
		Iterative	1.99	0.00	0.006	99	0	1	71	88.7
		One-step	1.99	0.00	0.006	99	0	1	77	2.6
		Oracle	2.00	0.00	0.010	100	0	0	100	-
Case 4	50	Exact	1.98	0.00	0.054	98	0	2	65	20.1
		Iterative	1.98	0.00	0.052	98	0	2	64	6.3
		One-step	1.95	0.02	0.067	95	1	4	69	1.9
		Oracle	2.00	0.00	0.044	100	0	0	100	-
	100	Exact	1.94	0.00	0.031	94	0	6	70	42.2
		Iterative	1.93	0.00	0.032	93	0	7	68	10.7
		One-step	1.96	0.00	0.028	96	0	4	80	2.0
		Oracle	2.00	0.00	0.029	100	0	0	100	-
	200	Exact	1.98	0.00	0.017	98	0	2	77	85.9
		Iterative	1.98	0.00	0.018	98	0	2	77	19.4
		One-step	1.95	0.00	0.017	95	0	5	77	2.6
		Oracle	2.00	0.00	0.017	100	0	0	100	-
	300	Exact	1.99	0.00	0.007	99	0	1	65	459.1
		Iterative	1.99	0.00	0.007	99	0	1	72	29.0
		One-step	2.00	0.00	0.010	100	0	0	76	3.2
		Oracle	2.00	0.00	0.010	100	0	0	100	-
Case 5	50	Exact	1.90	0.00	0.018	90	0	10	54	20.8
		Iterative	1.90	0.00	0.018	90	0	10	57	6.3
		One-step	1.90	0.00	0.017	90	0	10	57	1.8
		Oracle	2.00	0.00	0.014	100	0	0	100	-
	100	Exact	1.96	0.00	0.007	96	0	4	65	41.5
		Iterative	1.96	0.00	0.007	96	0	4	60	10.4
		One-step	1.96	0.00	0.007	96	0	4	63	2.0
		Oracle	2.00	0.00	0.006	100	0	0	100	-
	200	Exact	1.98	0.00	0.004	98	0	2	68	290.9
		Iterative	1.97	0.00	0.004	97	0	3	67	19.0
		One-step	1.98	0.00	0.004	98	0	2	68	2.4
		Oracle	2.00	0.00	0.004	100	0	0	100	-
	300	Exact	2.00	0.00	0.002	100	0	0	64	531.7
		Iterative	2.00	0.00	0.002	100	0	0	74	30.0
		One-step	2.00	0.00	0.002	100	0	0	67	3.1
		Oracle	2.00	0.00	0.002	100	0	0	100	-

Table 4.12: Fixed effects selection and estimation results of the exact method and two approximate methods for Setting 2

Case#	m	method	CZ	IZ	MME	C	U	O	Both	Time (sec/run)
Case 1	50	Exact	7.44	0.09	0.081	66	6	28	10	49.0
		Iterative	7.23	0.13	0.103	62	8	30	15	10.9
		One-step	7.40	0.07	0.085	65	6	29	19	2.2
		Oracle	8.00	0.00	0.038	100	0	0	100	-
	100	Exact	7.75	0.02	0.027	83	1	16	23	317.7
		Iterative	7.71	0.01	0.027	82	1	17	29	17.6
		One-step	7.71	0.02	0.025	83	1	16	35	2.5
		Oracle	8.00	0.00	0.018	100	0	0	100	-
	200	Exact	7.91	0.00	0.015	92	0	8	41	719.5
		Iterative	7.83	0.00	0.014	88	0	12	42	32.1
		One-step	7.92	0.00	0.015	93	0	7	51	3.1
		Oracle	8.00	0.00	0.010	100	0	0	100	-
	300	Exact	7.99	0.00	0.007	99	0	1	51	1174.2
		Iterative	7.98	0.00	0.007	98	0	2	55	47.1
		One-step	7.99	0.00	0.007	99	0	1	66	3.9
		Oracle	8.00	0.00	0.006	100	0	0	100	-
Case 2	50	Exact	7.27	0.09	0.103	59	6	35	0	53.1
		Iterative	7.02	0.27	0.151	48	20	32	0	11.7
		One-step	7.05	0.19	0.131	51	14	35	3	2.2
		Oracle	8.00	0.00	0.053	100	0	0	100	-
	100	Exact	7.51	0.06	0.034	71	5	24	2	331.9
		Iterative	7.12	0.17	0.052	61	12	27	6	20.0
		One-step	7.35	0.08	0.046	68	6	26	18	2.5
		Oracle	8.00	0.00	0.018	100	0	0	100	-
	200	Exact	7.87	0.00	0.018	87	0	13	12	836.6
		Iterative	7.56	0.08	0.020	80	6	14	19	34.8
		One-step	7.81	0.02	0.018	88	2	10	32	3.2
		Oracle	8.00	0.00	0.012	100	0	0	100	-
	300	Exact	7.91	0.00	0.007	95	0	5	11	1412.6
		Iterative	7.84	0.01	0.007	91	1	8	37	50.4
		One-step	7.82	0.00	0.006	92	0	8	41	4.0
		Oracle	8.00	0.00	0.005	100	0	0	100	-
Case 3	50	Exact	7.27	0.10	0.102	55	7	38	4	48.6
		Iterative	7.02	0.19	0.115	53	12	35	9	11.2
		One-step	7.20	0.13	0.117	57	9	34	8	2.3
		Oracle	8.00	0.00	0.044	100	0	0	100	-
	100	Exact	7.45	0.09	0.031	71	7	22	11	323.0
		Iterative	7.28	0.09	0.035	64	6	30	20	18.9
		One-step	7.36	0.10	0.039	73	7	20	29	2.6
		Oracle	8.00	0.00	0.017	100	0	0	100	-
	200	Exact	7.80	0.00	0.017	85	0	15	20	984.7

Table 4.12 Continued

Case#	m	method	CZ	IZ	MME	C	U	O	Both	Time (sec/run)
		Iterative	7.66	0.04	0.019	84	2	14	38	95.8
		One-step	7.71	0.02	0.019	83	2	15	38	3.4
		Oracle	8.00	0.00	0.013	100	0	0	100	-
	300	Exact	7.96	0.00	0.007	96	0	4	31	1141.8
		Iterative	7.92	0.00	0.007	96	0	4	53	50.8
		One-step	7.92	0.00	0.007	95	0	5	50	3.5
		Oracle	8.00	0.00	0.006	100	0	0	100	-
Case 4	50	Exact	7.38	0.08	0.094	60	5	35	9	150.4
		Iterative	7.39	0.10	0.082	62	7	31	7	10.9
		One-step	7.40	0.03	0.095	62	2	36	7	2.2
		Oracle	8.00	0.00	0.041	100	0	0	100	-
	100	Exact	7.74	0.01	0.031	82	1	17	24	374.1
		Iterative	7.74	0.00	0.026	83	0	17	33	18.0
		One-step	7.81	0.00	0.045	82	0	18	32	2.5
		Oracle	8.00	0.00	0.023	100	0	0	100	-
	200	Exact	7.90	0.00	0.018	95	0	5	35	715.2
		Iterative	7.94	0.01	0.018	94	1	5	42	33.2
		One-step	7.93	0.00	0.015	94	0	6	37	3.2
		Oracle	8.00	0.00	0.013	100	0	0	100	-
	300	Exact	7.94	0.00	0.011	94	0	6	47	1175.8
		Iterative	7.93	0.00	0.012	93	0	7	44	46.9
		One-step	7.87	0.00	0.012	91	0	9	54	3.7
		Oracle	8.00	0.00	0.008	100	0	0	100	-
Case 5	50	Exact	7.04	0.00	0.031	61	0	39	9	47.6
		Iterative	6.70	0.01	0.034	60	1	39	12	10.9
		One-step	6.78	0.00	0.032	61	0	39	8	2.3
		Oracle	8.00	0.00	0.014	100	0	0	100	-
	100	Exact	7.58	0.00	0.009	89	0	11	22	96.5
		Iterative	7.40	0.00	0.008	81	0	19	30	57.7
		One-step	7.65	0.00	0.009	85	0	15	30	2.8
		Oracle	8.00	0.00	0.006	100	0	0	100	-
	200	Exact	7.78	0.00	0.005	91	0	9	30	777.8
		Iterative	7.74	0.00	0.005	89	0	11	35	32.0
		One-step	7.73	0.00	0.005	87	0	13	44	3.2
		Oracle	8.00	0.00	0.003	100	0	0	100	-
	300	Exact	7.87	0.00	0.003	95	0	5	43	1076.7
		Iterative	7.90	0.00	0.003	95	0	5	41	45.4
		One-step	7.84	0.00	0.003	96	0	4	49	3.5
		Oracle	8.00	0.00	0.002	100	0	0	100	-

4.4 Real Example

In this section, we apply our proposed methods to the Amsterdam Growth and Health Study data and compare the results from the three methods. Figures 4.1 and 4.2 show the profiles of the diagonal elements of $\widehat{\mathbf{D}}$ and $\widehat{\mathbf{\Sigma}}$ as functions of $s = \sum_1^q d_i$, using the iterative approximation method and the one-step approximation method, respectively. The optimal values of s were chosen by BIC2_R. Figure 4.3 presents the coefficients for fixed effects using two approximation methods. As for well-known lasso estimates, the profiles are piece-wise linear. Table 4.13 presents the resulting estimates for random effects obtained from the proposed methods. The three proposed methods selected x_1, x_4 plus a random intercept as important random effects. Table 4.13 also shows the results of fixed effects selection from the Exact method and approximation methods. All of the three methods selected x_2 and x_5, and this result is consistent with REML estimation in Section 3.5. In conclusion, two approximate methods provided very similar estimation and selection when compared to the exact method.

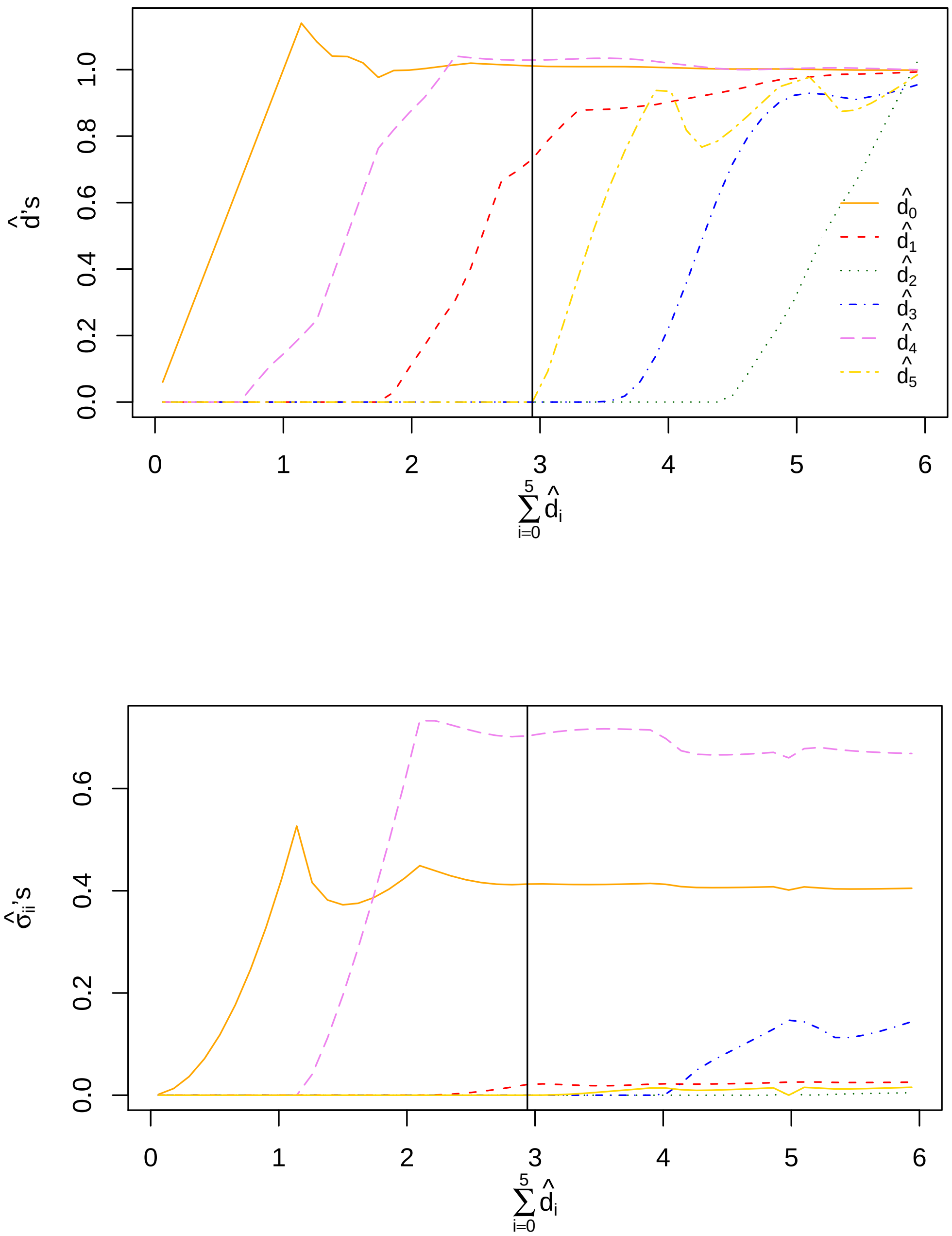


Figure 4.1: Profiles of variance matrix estimates for random effects as $s = \sum_1^q d_i$ is varied, using the iterative approximation method. A vertical line is drawn at $s = 2.94$, the optimal value chosen by BIC2_R

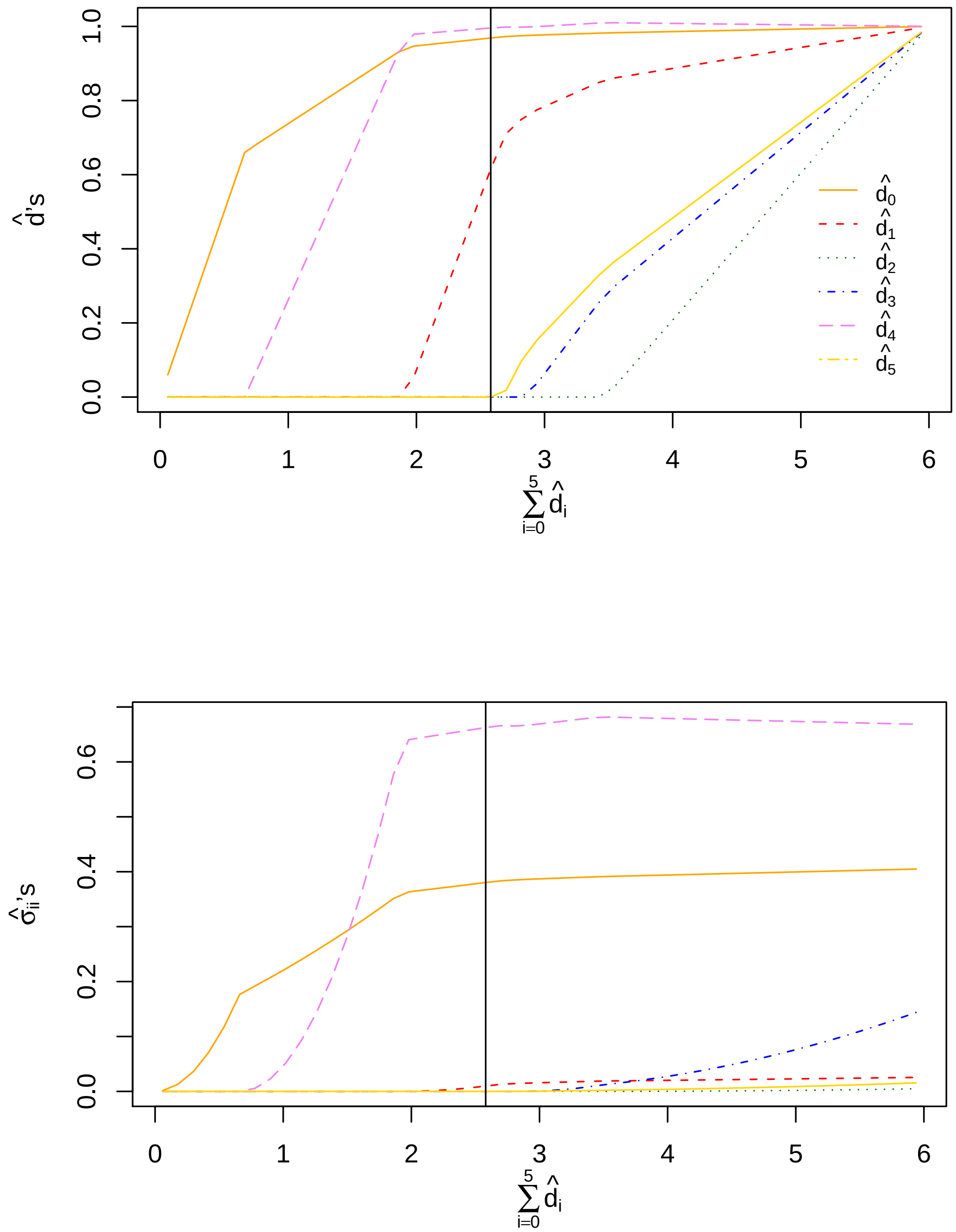


Figure 4.2: Profiles of variance matrix estimates for random effects as $s = \sum_1^q d_i$ is varied, using the one-step approximation method. A vertical line is drawn at $s = 2.58$, the optimal value chosen by BIC2$_R$

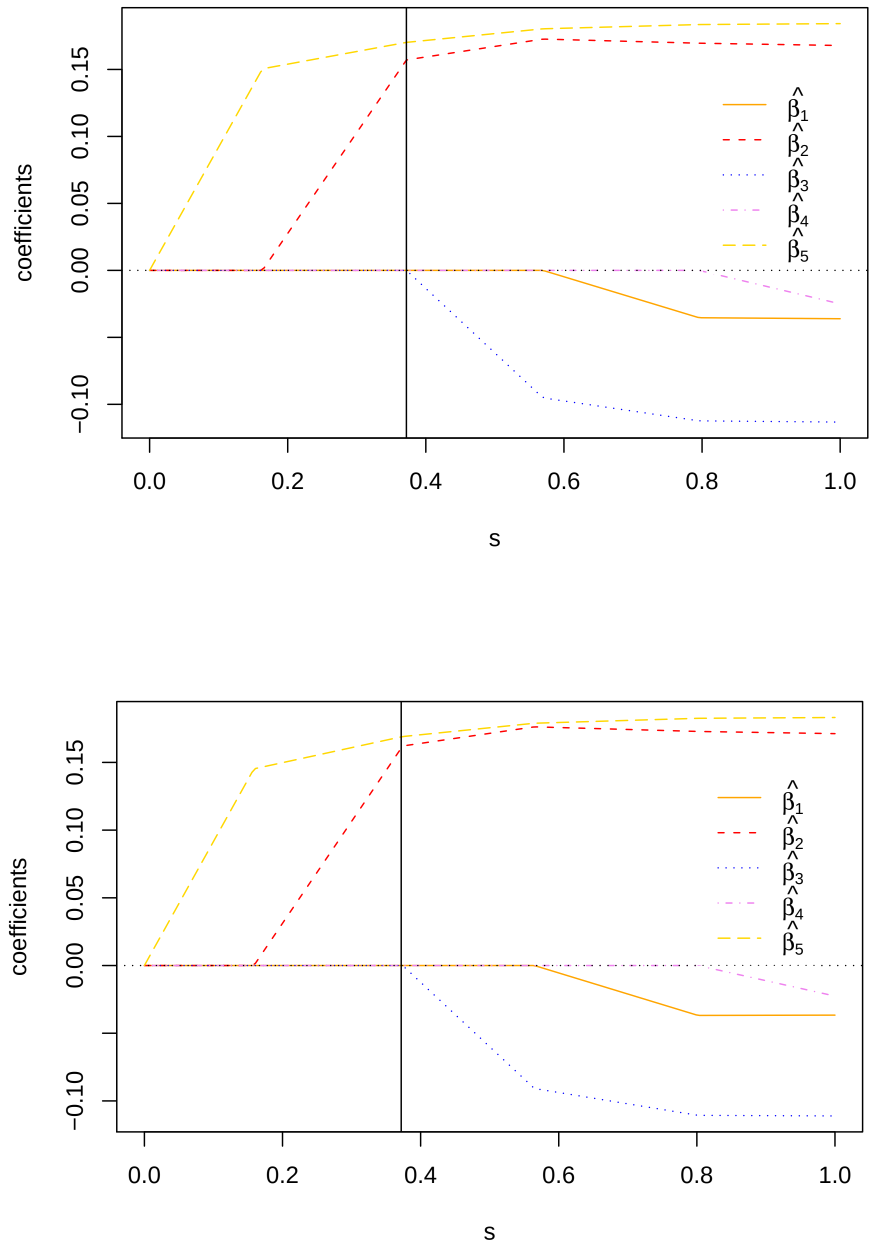


Figure 4.3: Profiles of coefficients for fixed effects using the iterative approximation method and one-step approximation method. A vertical line is drawn at $s = 0.37$, the optimal value chosen by BIC2_F

Table 4.13: Mixed effects selection results for the Amsterdam Growth and Health Study data: Comparison of estimates from REML and the proposed methods. Standard errors are given in parentheses

Random effect				Fixed effect			
Variable	Exact method	Iterative method	One-step method	Variable	Exact method	Iterative method	One-step method
var(γ_{0i})	0.347	0.413	0.380				
var(γ_{1i})	0.006	0.021	0.010	x_1	0	0	0
var(γ_{2i})	0	0	0	x_2	0.165	0.157	0.161
var(γ_{3i})	0	0	0	x_3	0	0	0
var(γ_{4i})	0.624	0.703	0.663	x_4	0	0	0
var(γ_{5i})	0	0	0	x_5	0.167	0.170	0.169
var(ε_i)=σ_ε^2	0.253	0.214	0.232				

Chapter 5

Conclusion and Future Work

We propose a robust method for random and fixed effects selection in linear mixed models. Our theoretical and numerical results suggest that the proposed method is a promising tool for the analysis of correlated data in practice. The proposed method depends on two challenging optimization problems. Though we have identified some good strategies and solvers for computing the solution, there is much room for a further improvement of computation efficiency.

One possibility could be linear approximation to the original nonlinear objective function used in random-effect selection. Then, the objective function can be solved by quadratic programming. We have considered two approximation methods; One is an iterative method and the other is simpler one-step approximation method. Through simulation studies, we found that the approximation can greatly speed up the computation for our procedure. Moreover, the approximate methods perform as well as or better than the exact moment-based method. In particular, the approximate methods overcome the limitation of the exact moment-based method in that they show good performance even when variables are moderately correlated.

We might extend our approach to the generalized linear mixed model in order to be used for binary or count data. We remain this extension for further study. In addition, we might make some modifications to better identify random effect factors even in case of correlated variables. We leave this topic for future research as well.

REFERENCES

Akaike, H. (1973). Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika* **60**, 255–265.

Allen, D. (1981). The prediction sum of squares as a criterion for selecting prediction variables. Tech. Rep. 23, Department of Statistics, University of Kentucky.

Azari, R., Li, L. & Tsai, C.-L. (2006). Longitudinal data model selection. *Computational Statistics and Data Analysis* **50**, 3053–3066.

Bondell, H., Krishna, A. & Ghosh, S. (2010). Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics* , In Press.

Bondell, H. D. & Reich, B. J. (2008). Simultaneous regression shrinkage, variable selection and clustering of predictors with oscar. *Biometrics* **64**, 115–123.

Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics* **37**, 373–384.

Breiman, L. & Spector, P. (1992). Submodel selection and evaluation in regression. the x-random case. *International Statistical Review* **60**, 291–319.

Cai, B. & Dunson, D. (2006). Bayesian covariance selection in generalized linear mixed models. *Biometrics* **62**, 446–457.

Chen, Z. & Dunson, D. (2003). Random effects selection in linear mixed models. *Biometrics* **59**, 762–769.

Craven, P. & Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik* **31**, 377–403.

Diggle, P., Liang, K. & Zeger, S. (1994). *Analysis of Longitudinal Data*. New York: Oxford University Press.

Fan, J. & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360.

Furnival, G. & Wilson, R. (1974). Regression by leaps and bounds. *Technometrics* **16**, 499–511.

HAUGHTON, D. (1988). On the choice of a model to fit data from an exponential family. *The annals of statistics* **16**, 342–355.

HOCKING, R. (1976). The analysis and selection of variables in linear regression. *Biometrics* **32**, 1–49.

HOERL, A. & KENNARD, R. (1970a). Ridge regression: application to nonorthogonal problems. *Technometrics* **12**, 69–82.

HOERL, A. & KENNARD, R. (1970b). Ridge regression: based estimation for nonorthogonal problems. *Technometrics* **12**, 55–67.

HOLMSTRÖM, K. (1999). The tomlab optimization environment in matlab. *Advanced Modeling and Optimization* **1**, 47–69.

HURVICH, C. & TSAI, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297–307.

IBRAHIM, J., ZHU, H., GARCIA, R. & GUO, R. (2010). Fixed and random effects selection in mixed effects models. *Biometrics* , In press.

IBRAHIM, J., ZHU, H. & TANG, N. (2008). Model selection criteria for missing-data problems using the em algorithm. *Journal of the American Statistical Association* **103**, 1648–1658.

JENNRICH, R. & SCHLUCHTER, M. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**, 805–820.

JIANG, J., RAO, J., GU, Z. & NGUYEN, T. (2008). Fence methods for mixed models selection. *Annals of Statistics* **36**, 1669–1692.

KEMPER, H., ed. (1995). *The Amsterdam Growth Study: A Longitudinal Analysis of Health, Fitness and Lifestyle*, vol. 6 of *HK Sport Science Monograph Series*. Champaign IL: Human Kinetics Publishers.

KINNEY, S. & DUNSON, D. (2007). Fixed and random effects selection in linear and logistic models. *Biometrics* **63**, 690–698.

LAIRD, N. M. & WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

Leng, C., Lin, Y. & Wahba, G. (2004). A note on the lasso and related procedures in model selection. Tech. rep., Department of Statistics, University of Wisconsin at Madison.

Lindstrom, M. & Bates, D. (1988). Newton-raphson and em algorithms for linear mixed-effects models for repeated measures data. *Journal of the American Statistical Association* **83**, 1014–1022.

Löfberg, J. (2004). Yalmip : A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*. Taipei, Taiwan.

Mallows, C. (1973). Some comments on c_p. *Technometrics* **15**, 661–675.

Mallows, C. (1995). More comments on c_p. *Technometrics* **37**, 362–372.

McQuarrie, A. & Tsai, C.-L. (1998). *Regression and time series model selection*. World Scientific.

Miller, A. (2002). *Subset selection in Regression*. New York: Chapman and Hill, 2nd ed.

Pu, P. & Niu, F. (2006). Selecting mixed-effects models based on generalized information criterion. *Journal of Multivariate Analysis* **97**, 733–758.

Rao, C. & Wu, Y. (1989). A strongly consistent procedure for model selection in a regression problem. *Biometrika* **76**, 369–374.

Roecker, E. (1991). Prediction error and its estimation for subset-selected models. *Technometrics* **20**, 459–468.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464.

Shibata, R. (1981). An optimal selection of regression variables. *Biometrika* **68**, 45–54.

Sturm, J. F. (1999). Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software* **11–12**, 625–653.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B.* **58**, 267–288.

Twisk, J. W. (2003). *Applied Longitudinal Data Analysis for Epidemilogy-Practical Guide*. New York: Cambridge University press.

Vaida, F. & Blanchard, S. (2005). Conditional akaike information for mixed-effects models. *Biometrika* **92**, 351–370.

Wang, F., Li, D.-H. & Qi, L. (2010). Global Convergence of Gauss-Newton-MBFGS Method for Solving the Nonlinear Least Squares Problem. *Advanced Modeling and Optimization* **12**, 1–20.

Wang, H. & Leng, C. (2007). Unified lasso estimation by least squares approximation. *Journal of the American Statistical Association* **102(479)**, 1039–1048.

Wang, H., Li, G. & Jiang, G. (2007). Robust regression shrinkage and consistent variable selection via the lad-lasso. *Journal of Business and Economics Statistics* **20**, 347–355.

Wolfinger, R. (1993). Covariance structure selection in general mixed models. *Communications in Statistics, Simulation and Computation* **22**, 1079–1106.

Yuan, M. & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B* **68**, 49–67.

Zhang, H. & Lu, W. (2007). Adaptive lasso for cox's proportional hazards model. *Biometrika* **94**, 1–13.

Zhao, P. & Yu, B. (2006). On model selection consistency of lasso. *Journal of machine learning research* **7**, 2541–2563.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* **101**, 1418–1429.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B.* **67**, 301–320.

APPENDICES

Appendix A

Asymptotic Properties

A.1 Proof of Lemma 1

The ordinary least square estimator $\widetilde{\boldsymbol{\beta}}$ minimizes $\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})^2$, and hence satisfies

$$\frac{\partial}{\partial\boldsymbol{\beta}}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})^2\Big|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\boldsymbol{x}_{ij}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})\boldsymbol{x}_{ij}=\mathbf{0}.$$

Thus, $\widetilde{\boldsymbol{\beta}}$ is an M-estimator which satisfies $\sum_{i=1}^{m}\psi_{1i}(\widetilde{\boldsymbol{\beta}})=\mathbf{0}$ with $\psi_{1i}(\boldsymbol{\beta})=\sum_{j=1}^{n_i}(y_{ij}-\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\boldsymbol{x}_{ij}=\mathbf{0}$. Then, we have

$$\mathbf{A}_{1m}=\frac{1}{m}\sum_{i=1}^{m}\left\{-\frac{\partial}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\psi_{1i}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}\right\}=\frac{1}{m}\sum_{i=1}^{m}\left\{\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{T}}\right\}=\frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i.$$

By the law of large numbers,

$$\mathbf{A}_{1m}\xrightarrow{p}\mathbf{A}_1=\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i\quad\text{as}\quad m\to\infty.$$

We also have

$$\mathbf{B}_{1m} = \frac{1}{m}\sum_{i=1}^{m}\psi_{1i}(\widetilde{\boldsymbol{\beta}})\psi_{1i}^{\mathrm{T}}(\widetilde{\boldsymbol{\beta}}) = \frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{X}_i,$$

where $\boldsymbol{\eta}_i = \mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i$ as defined in Step 1 of our Algorithm. By the law of large numbers,

$$\begin{aligned}\mathbf{B}_{1m} \xrightarrow{p} \mathbf{B}_1 &= \frac{1}{m}\sum_{i=1}^{m}E\big(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{X}_i\big) \qquad \text{as} \;\; m \to \infty \\ &= \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_i^{\mathrm{T}}(\mathbf{Z}_i\boldsymbol{\Sigma}\mathbf{Z}_i^{\mathrm{T}} + \sigma_\varepsilon^2\mathbf{I}_{n_i})\mathbf{X}_i,\end{aligned}$$

since $\mathrm{var}(\boldsymbol{\eta}_i) = \mathrm{var}(\mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i) = \mathbf{Z}_i\boldsymbol{\Sigma}\mathbf{Z}_i^{\mathrm{T}} + \sigma_\varepsilon^2\mathbf{I}_{n_i}$.

We assume that $\mathbf{A}_1$ is nonsingular and $\mathbf{B}_1$ is finite. Using $\widetilde{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{y} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\eta}$, we have

$$\sqrt{m}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) = \sqrt{m}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\eta} = \left(\frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\eta},$$

where $\boldsymbol{\eta} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$. The Lindeberg's central limit theorem yields

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\eta}_i \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{B}_1) \quad \text{as} \;\; m \to \infty.$$

Putting this all together, by Slutsky's theorem, we have

$$\sqrt{m}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{A}_1^{-1}\mathbf{B}_1\mathbf{A}_1^{-1}) \quad \text{as} \;\; m \to \infty. \tag{A.1}$$

A.2 Proof of Lemma 2

Because $\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})$ minimizes $L_0\big(\boldsymbol{\kappa}(\widetilde{\boldsymbol{\beta}})\big)$, we have

$$\frac{\partial}{\partial \boldsymbol{\kappa}} L_0\big(\boldsymbol{\kappa}(\widetilde{\boldsymbol{\beta}})\big)\bigg|_{\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})} = -2\sum_{i,j,k}\big(\widetilde{y}_{ijk} - \mathbf{Z}_{ijk}^{*\mathrm{T}}\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})\big)\mathbf{Z}_{ijk}^{*} = \mathbf{0}. \tag{A.2}$$

Define

$$G_m\big(\boldsymbol{\kappa}(\boldsymbol{\beta});\boldsymbol{\beta}\big) = -\frac{1}{2m}\Big(\frac{\partial}{\partial \boldsymbol{\kappa}} L_0\big(\boldsymbol{\kappa}(\boldsymbol{\beta})\big)\Big) = \frac{1}{m}\sum_{i,j,k}\big(y_{ijk} - \mathbf{Z}_{ijk}^{*\mathrm{T}}\boldsymbol{\kappa}(\boldsymbol{\beta})\big)\mathbf{Z}_{ijk}^{*}.$$

Hence, $G_m\big(\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}});\widetilde{\boldsymbol{\beta}}\big) = \mathbf{0}$. In M-estimation approach, we can also define $\psi_{2i}\big(\boldsymbol{\kappa}(\boldsymbol{\beta})\big) = \sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\big(y_{ijk} - \mathbf{Z}_{ijk}^{*\mathrm{T}}\boldsymbol{\kappa}(\boldsymbol{\beta})\big)\mathbf{Z}_{ijk}^{*}$. That is, $G_m\big(\boldsymbol{\kappa}(\boldsymbol{\beta});\boldsymbol{\beta}\big) = \frac{1}{m}\sum_{i=1}^{m}\psi_{2i}\big(\boldsymbol{\kappa}(\boldsymbol{\beta})\big)$. Then, we have

$$\begin{aligned}
\mathbf{A}_{2m} &= \frac{1}{m}\sum_{i=1}^{m}\bigg\{-\frac{\partial}{\partial \boldsymbol{\kappa}^{\mathrm{T}}}\psi_{2i}\big(\boldsymbol{\kappa}(\boldsymbol{\beta})\big)\bigg|_{\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}}\bigg\} \\
&= -\frac{\partial}{\partial \boldsymbol{\kappa}^{\mathrm{T}}} G_m\big(\boldsymbol{\kappa}(\boldsymbol{\beta});\boldsymbol{\beta}\big)\bigg|_{\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}} \\
&= -\frac{1}{m}\sum_{i=1}^{m}\bigg\{\frac{\partial}{\partial \boldsymbol{\kappa}^{\mathrm{T}}}\sum_{j,k}\big(y_{ijk} - \mathbf{Z}_{ijk}^{*\mathrm{T}}\boldsymbol{\kappa}(\boldsymbol{\beta})\big)\mathbf{Z}_{ijk}^{*}\bigg|_{\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}}\bigg\} \\
&= \frac{1}{m}\sum_{i,j,k}\mathbf{Z}_{ijk}^{*}\mathbf{Z}_{ijk}^{*\mathrm{T}} \\
&\xrightarrow{p} \lim_{m\to\infty}\frac{1}{m}\sum_{i,j,k}\mathbf{Z}_{ijk}^{*}\mathbf{Z}_{ijk}^{*\mathrm{T}} = \mathbf{A}_2
\end{aligned} \tag{A.3}$$

and

$$
\begin{aligned}
\mathbf{B}_{2m} &= \frac{1}{m}\sum_{i=1}^{m}\psi_{2i}\big(\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})\big)\psi_{2i}^{\mathrm{T}}\big(\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})\big) \\
&\xrightarrow{p} \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}E\big(\psi_{2i}(\boldsymbol{\kappa}_o)\psi_{2i}^{\mathrm{T}}(\boldsymbol{\kappa}_o)\big) \\
&\xrightarrow{p} \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\mathbf{Z}_i^{*\mathrm{T}}\mathrm{var}(e_i)\mathbf{Z}_i^{*} = \mathbf{B}_2, \qquad \text{(A.4)}
\end{aligned}
$$

since $\psi_{2i}(\boldsymbol{\kappa}_o) = \sum_{j,k}\big(\widetilde{y}_{ijk}(\boldsymbol{\beta}_o) - \mathbf{Z}_{ijk}^{*\mathrm{T}}\boldsymbol{\kappa}_o\big)\mathbf{Z}_{ijk}^{*} = \sum_{j,k} e_{ijk}\mathbf{Z}_{ijk}^{*} = \mathbf{Z}_i^{*\mathrm{T}}\boldsymbol{e}_i$. We assume that $\mathbf{A}_2$ is nonsingular and $\mathbf{B}_2$ is finite.

By the Lindeberg's central limit theorem, we have

$$
\begin{aligned}
\sqrt{m}G_m(\boldsymbol{\kappa}_o;\boldsymbol{\beta}_o) &= \frac{1}{\sqrt{m}}\sum_{ijk}\big(\widetilde{y}_{ijk}(\boldsymbol{\beta}_o) - \mathbf{Z}_{ijk}^{*\mathrm{T}}\boldsymbol{\kappa}_o\big)\mathbf{Z}_{ijk}^{*} \\
&= \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\mathbf{Z}_i^{*\mathrm{T}}\boldsymbol{\eta}_i \\
&\xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{B}_2). \qquad \text{(A.5)}
\end{aligned}
$$

From the Taylor expansion around $\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}) = \boldsymbol{\kappa}_o$ and $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_o$, we have

$$
\begin{aligned}
\mathbf{0} &= G_m\big(\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}});\widetilde{\boldsymbol{\beta}}\big) \\
&= G_m\big(\boldsymbol{\kappa}_o;\boldsymbol{\beta}_o\big) + \Big(\frac{\partial}{\partial\widetilde{\boldsymbol{\kappa}}^{\mathrm{T}}}G_m\big(\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}});\widetilde{\boldsymbol{\beta}}\big)\Big|_{\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o}\Big)\big(\widetilde{\boldsymbol{\kappa}}(\boldsymbol{\beta}_o) - \boldsymbol{\kappa}_o\big) \\
&+ \Big(\frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}}G_m\big(\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}});\widetilde{\boldsymbol{\beta}}\big)\Big|_{\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o}\Big)\big(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_o\big) \\
&+ \Big(\frac{\partial}{\partial\widetilde{\boldsymbol{\kappa}}^{\mathrm{T}}}G_m\big(\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}});\widetilde{\boldsymbol{\beta}}\big)\Big|_{\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o}\Big)\Big(\frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}}\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\beta}_o}\Big)\big(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_o\big) + o_p(m^{-1/2}). \qquad \text{(A.6)}
\end{aligned}
$$

In (A.2), $\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})$ satisfies $\sum_{i,j,k}\big(\widetilde{y}_{ijk} - \mathbf{Z}^{*\mathrm{T}}_{ijk}\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})\big)\mathbf{Z}^{*}_{ijk} = \mathbf{0}$. Taking derivative with respect to $\widetilde{\boldsymbol{\beta}}$, it follows that

$$\begin{aligned}
\mathbf{0} &= \frac{\partial}{\partial \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}} \sum_{i,j,k} \big(\widetilde{y}_{ijk} - \mathbf{Z}^{*\mathrm{T}}_{ijk}\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})\big)\mathbf{Z}^{*}_{ijk} \\
&= \sum_{i,j,k} \Big(\mathbf{Z}^{*}_{ijk} \frac{\partial}{\partial \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}} \widetilde{y}_{ijk} - \mathbf{Z}^{*}_{ijk}\mathbf{Z}^{*\mathrm{T}}_{ijk} \frac{\partial}{\partial \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}} \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})\Big).
\end{aligned}$$

Then, we have

$$\frac{\partial}{\partial \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}} \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}) = \Bigg(\sum_{i,j,k} \mathbf{Z}^{*}_{ijk}\mathbf{Z}^{*\mathrm{T}}_{ijk}\Bigg)^{-1} \Bigg(\sum_{i,j,k} \mathbf{Z}^{*}_{ijk} \frac{\partial}{\partial \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}} \widetilde{y}_{ijk}\Bigg). \tag{A.7}$$

In the right-hand side of (A.7), we have

$$\begin{aligned}
\frac{1}{m}\sum_{i,j,k} \mathbf{Z}^{*}_{ijk} \frac{\partial}{\partial \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}} \widetilde{y}_{ijk}\bigg|_{\boldsymbol{\beta}_o} &= \frac{1}{m}\sum_{i,j,k} \mathbf{Z}^{*}_{ijk}\Big(- \boldsymbol{x}_{ij}\big(y_{ik} - \boldsymbol{x}^{\mathrm{T}}_{ik}\widetilde{\boldsymbol{\beta}}\big) - \boldsymbol{x}_{ik}\big(y_{ij} - \boldsymbol{x}^{\mathrm{T}}_{ij}\widetilde{\boldsymbol{\beta}}\big)\Big)\bigg|_{\boldsymbol{\beta}_o} \\
&= \frac{1}{m}\sum_{i,j,k} \mathbf{Z}^{*}_{ijk}\big(- \boldsymbol{x}_{ij}\eta_{ik} - \boldsymbol{x}_{ik}\eta_{ij}\big) \\
&\xrightarrow{p} \mathbf{0},
\end{aligned} \tag{A.8}$$

since by the law of large numbers

$$\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n_i} \eta_{ij} \xrightarrow{p} 0. \tag{A.9}$$

By (A.3) and (A.8), (A.7) becomes

$$\frac{\partial}{\partial \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}} \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})\bigg|_{\boldsymbol{\beta}_o} \xrightarrow{p} \mathbf{0}. \tag{A.10}$$

Using (A.3), (A.10) and the fact that $\frac{1}{m}\sum_{i,j} a_{ij}\eta_{ij} \xrightarrow{p} 0$, it follows that

$$\begin{aligned}
\frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}} G_m\big(\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}});\widetilde{\boldsymbol{\beta}}\big)\Big|_{\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o} &= \frac{1}{m}\sum_{i,j,k}\mathbf{Z}^*_{ijk}\bigg(\frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}}\widetilde{y}_{ijk} - \mathbf{Z}^{*\mathrm{T}}_{ijk}\frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}}\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})\bigg)\Big|_{\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o} \\
&= \frac{1}{m}\sum_{i,j,k}\mathbf{Z}^*_{ijk}(-\boldsymbol{x}_{ij}\eta_{ik} - \boldsymbol{x}_{ik}\eta_{ij}) \\
&\xrightarrow{p} \mathbf{0}. \qquad \text{(A.11)}
\end{aligned}$$

Rearranging (A.6) with (A.3), (A.5), (A.10), (A.11), and Lemma 1, we have that

$$\sqrt{m}\big(\widetilde{\boldsymbol{\kappa}}(\boldsymbol{\beta}_o) - \boldsymbol{\kappa}_o\big) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{A}_2^{-1}\mathbf{B}_2\mathbf{A}_2^{-1}). \qquad \text{(A.12)}$$

Putting (A.10) and Lemma 1, the Taylor expansion around $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_o$ yields

$$\begin{aligned}
\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}) &= \widetilde{\boldsymbol{\kappa}}(\boldsymbol{\beta}_o) + \Big(\frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}}\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\beta}_o}\Big)(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) + o_p\Big(\frac{1}{\sqrt{m}}\Big) \\
&= \widetilde{\boldsymbol{\kappa}}(\boldsymbol{\beta}_o) + o_p\Big(\frac{1}{\sqrt{m}}\Big),
\end{aligned}$$

and thus $\sqrt{m}\big(\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}) - \boldsymbol{\kappa}_o\big) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{A}_2^{-1}\mathbf{B}_2\mathbf{A}_2^{-1})$.

A.3 Proof of Theorem 1

To prove theorem 1, it is sufficient to show that for any given $\epsilon > 0$, there exists a large constant C such that

$$\text{pr}\Big[\inf_{\boldsymbol{\sigma}^*\in B_m(C)} Q_R\big(\boldsymbol{\sigma}^*;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}\big) > Q_R\big(\boldsymbol{\sigma}^*_o;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}\big)\Big] \geq 1-\epsilon, \tag{A.13}$$

where the C-ball $B_m(C) = \{\boldsymbol{\sigma}^* : \boldsymbol{\sigma}^* = \boldsymbol{\sigma}^*_o + \boldsymbol{u}/\sqrt{m}, \|\boldsymbol{u}\| \leq C\}$. The derivatives of Q_R and L_R with respect to $\boldsymbol{\sigma}^*$ or $\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})$ are the right derivatives because $\boldsymbol{\sigma}$ is defined in the set of nonnegative real q-vectors, denoted by $\mathbb{R}^{q+}$.

From the Taylor expansion of L_R around $\boldsymbol{\sigma}^* = \boldsymbol{\sigma}^*_o$, we have

$$\begin{aligned}
& Q_R(\boldsymbol{\sigma}^*_o + \frac{\boldsymbol{u}}{\sqrt{m}};\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}) - Q_R(\boldsymbol{\sigma}^*_o;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}) \\
= \;& L_R(\boldsymbol{\sigma}^*_o + \frac{\boldsymbol{u}}{\sqrt{m}};\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}) - L_R(\boldsymbol{\sigma}^*_o;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}) + \lambda\sum_{i=1}^{q}\Big(\frac{\sigma^*_{io}+u_i/\sqrt{m}}{\widetilde{\sigma}_i} - \frac{\sigma^*_{io}}{\widetilde{\sigma}_i}\Big) \\
= \;& \Big(\frac{\boldsymbol{u}}{\sqrt{m}}\Big)^{\mathrm{T}} S_R(\boldsymbol{\sigma}^*_o;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}) + \frac{1}{2}\Big(\frac{\boldsymbol{u}}{\sqrt{m}}\Big)^{\mathrm{T}}\nabla S_R(\boldsymbol{\sigma}^*_o;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}})\Big(\frac{\boldsymbol{u}}{\sqrt{m}}\Big) + \lambda\sum_{i=1}^{q}\frac{u_i/\sqrt{m}}{\widetilde{\sigma}_i},
\end{aligned} \tag{A.14}$$

where $S_R(\boldsymbol{\sigma}^*_o;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}) = \frac{\partial}{\partial\boldsymbol{\sigma}^*}L_R(\boldsymbol{\sigma}^*;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}})\big|_{\boldsymbol{\sigma}^*_o}$ and $\nabla S_R(\boldsymbol{\sigma}^*_o;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}})$
$= \frac{\partial^2}{\partial\boldsymbol{\sigma}^*\partial\boldsymbol{\sigma}^{*\mathrm{T}}}L_R(\boldsymbol{\sigma}^*;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}})\big|_{\boldsymbol{\sigma}^*_o}$.
Note that by the law of large numbers

$$\frac{1}{m}\sum_{i,j,k} e_{ijk} \xrightarrow{p} 0. \tag{A.15}$$

Therefore, we can obtain

$$\frac{1}{m}\Big(\frac{\partial}{\partial\widetilde{\boldsymbol{\kappa}}^{\mathrm{T}}}S_R(\boldsymbol{\sigma}_o^*;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o}\Big) \xrightarrow{p} \mathbf{E} \tag{A.16}$$

and

$$\frac{1}{m}\Big(\frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}}S_R(\boldsymbol{\sigma}_o^*;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o}\Big) \xrightarrow{p} \mathbf{0}, \tag{A.17}$$

where $\mathbf{E}$ is a $q \times q(q+1)/2$ matrix.

Using (A.10), (A.16), (A.17), Lemmas 1 and 2, the first-order Taylor expansion around $\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}) = \boldsymbol{\kappa}_o$ and $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_o$ yields

$$\begin{aligned}
S_R(\boldsymbol{\sigma}_o^*;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}) &= S_R(\boldsymbol{\sigma}_o^*;\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o) + \Big(\frac{\partial}{\partial\widetilde{\boldsymbol{\kappa}}^{\mathrm{T}}}S_R(\boldsymbol{\sigma}_o^*;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o}\Big)(\widetilde{\boldsymbol{\kappa}}(\boldsymbol{\beta}_o) - \boldsymbol{\kappa}_o) \\
&\quad + \Big(\frac{\partial}{\partial\widetilde{\boldsymbol{\kappa}}^{\mathrm{T}}}S_R(\boldsymbol{\sigma}_o^*;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o}\Big)\Big(\frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}}\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\beta}_o}\Big)(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \\
&\quad + \Big(\frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}}S_R(\boldsymbol{\sigma}_o^*;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o}\Big)(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \\
&\quad + o_p(\|\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}) - \boldsymbol{\kappa}_o\|) + o_p(\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_o\|) \\
&= S_R(\boldsymbol{\sigma}_o^*;\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o) + o_p(m)O_p(1/\sqrt{m}) + o_p(m)o_p(1)O_p(1/\sqrt{m}) \\
&\quad + o_p(m)O_p(1/\sqrt{m}) + o_p(1/\sqrt{m}) + o_p(1/\sqrt{m}) \\
&= S_R(\boldsymbol{\sigma}_o^*;\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o) + o_p(\sqrt{m})
\end{aligned} \tag{A.18}$$

and

$$\nabla S_R(\boldsymbol{\sigma}_o^*;\widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}) = \nabla S_R(\boldsymbol{\sigma}_o^*;\boldsymbol{\kappa}_o,\boldsymbol{\beta}_o) + o_p(\sqrt{m}). \tag{A.19}$$

The tth component of $S_R(\boldsymbol{\sigma}_o^*; \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o)$ is

$$-2\sum_{i,j,k} e_{ijk} \sum_{l=1}^{q} (z_{ijt}z_{ikl} + z_{ijl}z_{ikt})\frac{\sigma_{tl,o}}{\sigma_t},$$

where $\sigma_{tl,o}$ is the (t,l)th element of Σ_o. Hence, $S_R(\boldsymbol{\sigma}_o^*; \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o)$ can be expressed in a form of $\sum_{i=1}^m \mathbf{W_i}^{\mathrm{T}}\boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is a column vector consisting of e_{ijk}'s. By the central limit theorem, we have

$$\frac{1}{\sqrt{m}} S_R(\boldsymbol{\sigma}_o^*; \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{F}) \quad \text{as} \quad n \to \infty, \tag{A.20}$$

where $\mathbf{F} = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^m E\big(\mathbf{W_i}^{\mathrm{T}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}\mathbf{W_i}\big) = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^m \mathbf{W_i}^{\mathrm{T}}\mathrm{var}(\boldsymbol{e}_i)\mathbf{W_i}$ and a $q \times q$ positive semidefinite matrix. This implies $S_R(\boldsymbol{\sigma}_o^*; \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o)/\sqrt{m} = O_p(1)$.

In addition, by the law of large numbers, it follows that

$$\frac{1}{m}\nabla S_R(\boldsymbol{\sigma}_o^*; \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o) = \mathbf{H} + o_p(1), \tag{A.21}$$

where $\mathbf{H}$ is a $q \times q$ positive semidefinite matrix.

Then, using (A.18) and (A.19), the first and second terms of (A.14) become

$$\begin{aligned}
&\Big(\frac{\boldsymbol{u}}{\sqrt{m}}\Big)^{\mathrm{T}}\Big(S_R(\boldsymbol{\sigma}_o^*; \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o) + o_p(\sqrt{m})\Big) + \frac{1}{2}\Big(\frac{\boldsymbol{u}}{\sqrt{m}}\Big)^{\mathrm{T}}\Big(\nabla S_R(\boldsymbol{\sigma}_o^*; \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o) + o_p(\sqrt{m})\Big)\Big(\frac{\boldsymbol{u}}{\sqrt{m}}\Big)\\
&= \boldsymbol{u}^{\mathrm{T}}\big(O_p(1) + o_p(1)\big) + \frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\big(H + o_p(1) + o_p(m^{-1/2})\big)\boldsymbol{u}\\
&= \boldsymbol{u}^{\mathrm{T}}O_p(1) + \frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\big(H + o_p(1)\big)\boldsymbol{u}.
\end{aligned} \tag{A.22}$$

We assume that the first b diagonal elements of $\mathbf{\Sigma}_o$ are nonzero. For $i = 1, 2, \ldots, b$, we have

$$\frac{1}{\widetilde{\sigma}_i} = \frac{1}{\sigma_{i,o}} - \frac{1}{\sigma_{i,o}^2}(\widetilde{\sigma}_i - \sigma_{i,o}) + o_p(|\widetilde{\sigma}_i - \sigma_{i,o}|) = \frac{1}{\sigma_{i,o}} + O_p\left(\frac{1}{\sqrt{m}}\right).$$

Hence, if $\lambda/\sqrt{m} = O_p(1)$, then

$$\frac{\lambda}{\sqrt{m}} \sum_{i=1}^{q} \frac{u_i}{\widetilde{\sigma}_i} \geq \frac{\lambda}{\sqrt{m}} \sum_{i=1}^{b} \frac{u_i}{\widetilde{\sigma}_i} = \frac{\lambda}{\sqrt{m}} \sum_{i=1}^{b} u_i \Big(\frac{1}{\sigma_{i,o}} + O_p\big(\frac{1}{\sqrt{m}}\big)\Big) = \sum_{i=1}^{b} u_i O_p(1). \quad \text{(A.23)}$$

Combining (A.22) and (A.23), it follows that

$$Q_R(\boldsymbol{\sigma}_o^* + \frac{\boldsymbol{u}}{\sqrt{m}}; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}) - Q_R(\boldsymbol{\sigma}_o^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}) \geq \boldsymbol{u}^{\mathrm{T}} O_p(1) + \frac{1}{2} \boldsymbol{u}^{\mathrm{T}} \big(H + o_p(1)\big) \boldsymbol{u} + \sum_{i=1}^{b} u_i O_p(1).$$

If we choose a sufficiently large constant C, the second term dominates the other terms. Therefore, (A.13) holds. This means that $\sqrt{m}(\widehat{\sigma}_i^* - \sigma_{i,o}^*) = O_p(1)$. By the Delta method, it is easy to show that $\sqrt{m}(\widehat{\sigma}_{ii} - \sigma_{ii,o}) = O_p(1)$. Since $\widehat{\sigma}_{ij} = \widetilde{\sigma}_{ij} \sqrt{\widehat{\sigma}_{ii} \widehat{\sigma}_{jj}} / \sqrt{\widetilde{\sigma}_{ii} \widetilde{\sigma}_{jj}}$, root-$m$ consistency also holds for the off-diagonal elements of $\widehat{\mathbf{\Sigma}}$. Hence, $\sqrt{m}(\widehat{\sigma}_{ij} - \sigma_{ij,o}) = O_p(1)$ for $i, j = 1, \ldots, q$.

A.4 Proof of Theorem 2

We first prove that $\widehat{\boldsymbol{\sigma}}_2^* = 0$ with probability tending to 1. It is enough to show that for any sequence $\boldsymbol{\sigma}_1^*$ satisfying $\|\boldsymbol{\sigma}_1^* - \boldsymbol{\sigma}_{10}^*\| = O_p(m^{-1/2})$ and for any constant C,

$$Q_R\big((\boldsymbol{\sigma}_1^*, \mathbf{0}); \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}\big) = \min_{\|\boldsymbol{\sigma}_2^*\| \le Cm^{-1/2}} Q_R\big((\boldsymbol{\sigma}_1^*, \boldsymbol{\sigma}_2^*); \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}\big).$$

From (A.18), (A.19), (A.21), and the second-order Taylor expansion around $\boldsymbol{\sigma}^* = \boldsymbol{\sigma}_o^*$, we obtain

$$\begin{aligned} L_R(\boldsymbol{\sigma}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}) \;\; = \;\; & L_R(\boldsymbol{\sigma}_o^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}) + (\boldsymbol{\sigma}^* - \boldsymbol{\sigma}_o^*)^{\mathrm{T}} S_R(\boldsymbol{\sigma}_o^*; \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o) \\ & + \frac{1}{2}(\boldsymbol{\sigma}^* - \boldsymbol{\sigma}_o^*)^{\mathrm{T}} \nabla S_R(\boldsymbol{\sigma}_o^*; \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o)(\boldsymbol{\sigma}^* - \boldsymbol{\sigma}_o^*) + o_p(1). \end{aligned}$$

For $t = b+1, \ldots, q$,

$$\begin{aligned} \frac{\partial}{\partial \sigma_t^*} Q_R(\boldsymbol{\sigma}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}) \;\; &= \;\; \frac{\partial}{\partial \sigma_t^*} L_R(\boldsymbol{\sigma}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}) + \lambda \frac{1}{\widetilde{\sigma}_t} \\ &= \;\; [S_R(\boldsymbol{\sigma}_o^*; \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o)]_t + [\nabla S_R(\boldsymbol{\sigma}_o^*; \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o)]_t \sigma_t^* + \frac{\sqrt{m}\lambda}{|O_p(1)|} \\ &= \;\; \sqrt{m}\Big(O_p(1) + \frac{\lambda}{|O_p(1)|}\Big), \end{aligned}$$

where $[A]_t$ denotes the tth element if A is a vector, and the tth row vector if A is a matrix. Since $\lambda \to \infty$, $\frac{\partial}{\partial \sigma_t^*} Q_R(\boldsymbol{\sigma}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}) > 0$. Hence, $Q_R((\boldsymbol{\sigma}_1^*, \mathbf{0}); \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}) < Q_R((\boldsymbol{\sigma}_1^*, \boldsymbol{\sigma}_2^*); \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}})$. This completes the proof.

Next we show part (b) of Theorem 2, which is asymptotic normality of $\widehat{\boldsymbol{\sigma}}_1^*$. Define $S_{1R}(\boldsymbol{\sigma}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}})$ be the first b elements of $S_R(\boldsymbol{\sigma}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}})$. From part (a), we can write $\widehat{\boldsymbol{\sigma}}^* = (\widehat{\boldsymbol{\sigma}}_1^{*\mathrm{T}}, \mathbf{0}^{\mathrm{T}})^{\mathrm{T}}$. Hence, we have

$$\begin{aligned}\mathbf{0} &= \frac{\partial}{\partial \boldsymbol{\sigma}_1^*} Q_R(\boldsymbol{\sigma}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\sigma}^* = (\widehat{\boldsymbol{\sigma}}_1^{*\mathrm{T}}, \mathbf{0}^{\mathrm{T}})^{\mathrm{T}}} \\ &= S_{1R}(\widehat{\boldsymbol{\sigma}}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}) + \lambda\Big(\frac{1}{\widetilde{\boldsymbol{\sigma}}_1}, \ldots, \frac{1}{\widetilde{\boldsymbol{\sigma}}_b}\Big)^{\mathrm{T}}.\end{aligned}$$

From the first-order Taylor expansion around $(\widehat{\boldsymbol{\sigma}}^*, \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}) = (\boldsymbol{\sigma}_o^*, \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o)$, we have

$$\begin{aligned}&S_{1R}(\widehat{\boldsymbol{\sigma}}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}) = S_{1R}(\boldsymbol{\sigma}_o^*; \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o) + \Big(\frac{\partial}{\partial \widehat{\boldsymbol{\sigma}}_1^{*\mathrm{T}}} S_{1R}(\widehat{\boldsymbol{\sigma}}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\sigma}_o^*, \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o}\Big)(\widehat{\boldsymbol{\sigma}}_1^* - \boldsymbol{\sigma}_{10}^*) \\ &+\Big(\frac{\partial}{\partial \widetilde{\boldsymbol{\kappa}}^{\mathrm{T}}} S_{1R}(\widehat{\boldsymbol{\sigma}}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\sigma}_o^*, \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o}\Big)\big(\widetilde{\boldsymbol{\kappa}}(\boldsymbol{\beta}_o) - \boldsymbol{\kappa}_o\big) \\ &+\Big(\frac{\partial}{\partial \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}} S_{1R}(\widehat{\boldsymbol{\sigma}}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\sigma}_o^*, \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o}\Big)(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \\ &+\Big(\frac{\partial}{\partial \widetilde{\boldsymbol{\kappa}}^{\mathrm{T}}} S_{1R}(\widehat{\boldsymbol{\sigma}}^*; \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\sigma}_o^*, \boldsymbol{\kappa}_o, \boldsymbol{\beta}_o}\Big)\Big(\frac{\partial}{\partial \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}} \widetilde{\boldsymbol{\kappa}}(\widetilde{\boldsymbol{\beta}})\Big|_{\boldsymbol{\beta}_o}\Big)(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) + o_p(\sqrt{m}).\end{aligned}$$

Since $\lambda/\sqrt{m} \to 0$ and $\widetilde{\sigma}_i \xrightarrow{p} \sigma_{i,o} \neq 0$ for $i = 1, \ldots, b$, we have

$$\frac{1}{\sqrt{m}}\lambda\Big(\frac{1}{\widetilde{\sigma}_1}, \ldots, \frac{1}{\widetilde{\sigma}_b}\Big)^{\mathrm{T}} \xrightarrow{p} \mathbf{0}. \tag{A.24}$$

Putting (A.10), (A.16), (A.17), (A.20), (A.21), (A.24), Lemmas 1 and 2 into (A.24), we have, by Slutsky's theorem,

$$\sqrt{m}(\widehat{\boldsymbol{\sigma}}_1^* - \boldsymbol{\sigma}_{10}^*) \xrightarrow{\mathcal{D}} N\big(\mathbf{0}, \mathbf{T}\big), \tag{A.25}$$

where $\mathbf{T} = \mathbf{H}_1^{-1}(\mathbf{F}_1 + \mathbf{E}_1\mathbf{A}_2^{-1}\mathbf{B}_2\mathbf{A}_2^{-1}\mathbf{E}_1^{\mathrm{T}})\mathbf{H}_1^{-1}$ with $\mathbf{E}_1$, $\mathbf{F}_1$, and $\mathbf{H}_1$ being the first $b \times q(q+1)/2$, $b \times b$, and $b \times b$ submatrices of $\mathbf{E}$, $\mathbf{F}$, and $\mathbf{H}$, respectively.

By the Delta theorem, we have

$$\sqrt{m}(\widehat{\boldsymbol{\sigma}}_1 - \boldsymbol{\sigma}_{10}) \xrightarrow{\mathcal{D}} N\big(\mathbf{0}, \mathbf{T}_2\big), \tag{A.26}$$

where $\mathbf{T}_2 = 4\text{diag}(\sqrt{\boldsymbol{\sigma}_{10}})\mathbf{T}\text{diag}(\sqrt{\boldsymbol{\sigma}_{10}})$.

Appendix B

Simulation Results

B.1 Fixed effect selection results: Comparison of selection criteria

Table B.1: Fixed effect selection and estimation results for Setting 1 and Case 1 when applying the Lasso

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.22	0.12	0.076	33	5	62	25
	BIC2	1.07	0.02	0.069	30	1	69	23
	GCV1	0.47	0.02	0.059	1	1	98	1
	GCV2	0.69	0.02	0.060	14	1	85	11
	GCV3	1.02	0.02	0.063	27	1	72	20
	GCV4	1.02	0.02	0.063	27	1	72	20
	CV	0.39	0.02	0.060	6	1	93	6
	Oracle	2.00	0.00	0.038	100	0	0	100
100	BIC1	1.17	0.00	0.052	36	0	64	31
	BIC2	1.07	0.00	0.051	31	0	69	26
	GCV1	0.40	0.00	0.034	1	0	99	0
	GCV2	0.67	0.00	0.034	12	0	88	10
	GCV3	0.83	0.00	0.039	16	0	84	13
	GCV4	0.83	0.00	0.039	16	0	84	13
	CV	0.31	0.00	0.033	3	0	97	1
	Oracle	2.00	0.00	0.022	100	0	0	100
200	BIC1	1.19	0.00	0.024	38	0	62	32
	BIC2	1.12	0.00	0.023	35	0	65	29
	GCV1	0.37	0.00	0.013	3	0	97	2
	GCV2	0.64	0.00	0.015	11	0	89	8
	GCV3	0.86	0.00	0.017	16	0	84	13
	GCV4	0.88	0.00	0.018	18	0	82	15
	CV	0.33	0.00	0.013	3	0	97	2
	Oracle	2.00	0.00	0.009	100	0	0	100
300	BIC1	1.18	0.00	0.015	39	0	61	28
	BIC2	1.12	0.00	0.014	35	0	65	25
	GCV1	0.34	0.00	0.009	3	0	97	3
	GCV2	0.62	0.00	0.010	15	0	85	9
	GCV3	0.86	0.00	0.012	19	0	81	13
	GCV4	0.89	0.00	0.011	20	0	80	14
	CV	0.28	0.00	0.008	5	0	95	5
	Oracle	2.00	0.00	0.006	100	0	0	100

Table B.2: Fixed effect selection and estimation results for Setting 1 and Case 1 when applying the ALasso1

m	Criterion$_F$	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.95	0.11	0.051	90	5	5	58
	BIC2	1.94	0.01	0.048	93	1	6	60
	GCV1	1.27	0.01	0.050	39	1	60	24
	GCV2	1.60	0.01	0.048	64	1	35	39
	GCV3	1.83	0.01	0.044	83	1	16	52
	GCV4	1.87	0.01	0.049	86	1	13	55
	CV	0.84	0.00	0.057	22	0	78	13
	Oracle	2.00	0.00	0.038	100	0	0	100
100	BIC1	1.97	0.01	0.028	96	1	3	72
	BIC2	1.95	0.00	0.028	95	0	5	70
	GCV1	1.24	0.00	0.029	35	0	65	25
	GCV2	1.68	0.00	0.030	70	0	30	55
	GCV3	1.88	0.00	0.028	88	0	12	65
	GCV4	1.88	0.00	0.028	88	0	12	65
	CV	0.81	0.02	0.031	14	2	84	12
	Oracle	2.00	0.00	0.022	100	0	0	100
200	BIC1	1.98	0.00	0.010	98	0	2	71
	BIC2	1.98	0.00	0.010	98	0	2	71
	GCV1	1.11	0.00	0.010	28	0	72	18
	GCV2	1.64	0.00	0.011	71	0	29	52
	GCV3	1.79	0.00	0.011	80	0	20	57
	GCV4	1.86	0.00	0.011	86	0	14	62
	CV	0.78	0.00	0.012	16	0	84	10
	Oracle	2.00	0.00	0.009	100	0	0	100
300	BIC1	2.00	0.00	0.007	100	0	0	61
	BIC2	1.99	0.00	0.007	99	0	1	60
	GCV1	1.09	0.00	0.006	29	0	71	16
	GCV2	1.65	0.00	0.007	72	0	28	44
	GCV3	1.89	0.00	0.007	89	0	11	50
	GCV4	1.89	0.00	0.007	89	0	11	50
	CV	0.62	0.00	0.008	13	0	87	7
	Oracle	2.00	0.00	0.006	100	0	0	100

Table B.3: Fixed effect selection and estimation results for Setting 1 and Case 1 when applying the ALasso2

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.97	0.11	0.048	92	5	3	59
	BIC2	1.91	0.01	0.045	90	1	9	57
	GCV1	1.26	0.01	0.048	39	1	60	27
	GCV2	1.65	0.01	0.050	68	1	31	42
	GCV3	1.86	0.01	0.045	85	1	14	54
	GCV4	1.86	0.01	0.046	85	1	14	54
	CV	0.86	0.06	0.070	13	4	83	9
	Oracle	2.00	0.00	0.038	100	0	0	100
100	BIC1	1.99	0.01	0.027	98	1	1	73
	BIC2	1.97	0.00	0.027	97	0	3	71
	GCV1	1.25	0.00	0.028	33	0	67	24
	GCV2	1.70	0.00	0.030	72	0	28	55
	GCV3	1.90	0.00	0.028	90	0	10	66
	GCV4	1.91	0.00	0.027	91	0	9	67
	CV	0.93	0.05	0.031	19	4	77	13
	Oracle	2.00	0.00	0.022	100	0	0	100
200	BIC1	1.98	0.00	0.010	98	0	2	71
	BIC2	1.98	0.00	0.010	98	0	2	71
	GCV1	1.16	0.00	0.010	32	0	68	21
	GCV2	1.66	0.00	0.012	72	0	28	53
	GCV3	1.83	0.00	0.011	83	0	17	60
	GCV4	1.89	0.00	0.011	89	0	11	64
	CV	0.92	0.08	0.014	17	7	76	12
	Oracle	2.00	0.00	0.009	100	0	0	100
300	BIC1	2.00	0.00	0.006	100	0	0	61
	BIC2	1.99	0.00	0.006	99	0	1	60
	GCV1	1.19	0.00	0.006	38	0	62	23
	GCV2	1.64	0.00	0.007	72	0	28	44
	GCV3	1.87	0.00	0.006	87	0	13	49
	GCV4	1.89	0.00	0.006	89	0	11	50
	CV	0.83	0.01	0.009	18	1	81	9
	Oracle	2.00	0.00	0.006	100	0	0	100

Table B.4: Fixed effect selection and estimation results for Setting 1 and Case 2 when applying the Lasso

m	Criterion$_F$	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.27	0.23	0.104	31	10	59	16
	BIC2	1.20	0.06	0.093	33	3	64	17
	GCV1	0.58	0.06	0.069	6	3	91	3
	GCV2	0.84	0.06	0.063	16	3	81	10
	GCV3	1.09	0.06	0.082	26	3	71	14
	GCV4	1.09	0.06	0.073	26	3	71	14
	CV	0.53	0.06	0.069	6	3	91	3
	Oracle	2.00	0.00	0.049	100	0	0	100
100	BIC1	1.12	0.13	0.041	30	5	65	24
	BIC2	1.08	0.04	0.041	30	2	68	24
	GCV1	0.48	0.04	0.035	6	2	92	5
	GCV2	0.69	0.04	0.035	15	2	83	11
	GCV3	0.93	0.04	0.039	23	2	75	17
	GCV4	1.00	0.04	0.041	27	2	71	21
	CV	0.44	0.04	0.034	7	2	91	4
	Oracle	2.00	0.00	0.028	100	0	0	100
200	BIC1	1.20	0.00	0.019	39	0	61	23
	BIC2	1.15	0.00	0.016	36	0	64	22
	GCV1	0.55	0.00	0.014	7	0	93	3
	GCV2	0.79	0.00	0.014	19	0	81	12
	GCV3	1.00	0.00	0.016	27	0	73	14
	GCV4	1.00	0.00	0.015	27	0	73	14
	CV	0.48	0.00	0.016	11	0	89	7
	Oracle	2.00	0.00	0.009	100	0	0	100
300	BIC1	1.17	0.00	0.014	38	0	62	20
	BIC2	1.16	0.00	0.013	38	0	62	20
	GCV1	0.43	0.00	0.009	3	0	97	3
	GCV2	0.73	0.00	0.011	18	0	82	9
	GCV3	0.95	0.00	0.012	25	0	75	14
	GCV4	0.95	0.00	0.012	25	0	75	14
	CV	0.38	0.00	0.009	5	0	95	3
	Oracle	2.00	0.00	0.010	100	0	0	100

Table B.5: Fixed effect selection and estimation results for Setting 1 and Case 2 when applying the ALasso1

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.93	0.23	0.067	85	10	5	51
	BIC2	1.89	0.06	0.055	88	3	9	52
	GCV1	1.24	0.05	0.061	38	3	59	20
	GCV2	1.62	0.06	0.056	67	3	30	38
	GCV3	1.85	0.06	0.060	84	3	13	48
	GCV4	1.86	0.06	0.057	85	3	12	49
	CV	0.87	0.06	0.072	23	3	74	11
	Oracle	2.00	0.00	0.049	100	0	0	100
100	BIC1	1.98	0.11	0.032	93	5	2	65
	BIC2	1.97	0.02	0.032	95	2	3	66
	GCV1	1.23	0.02	0.032	34	2	64	25
	GCV2	1.72	0.02	0.032	74	2	24	51
	GCV3	1.87	0.02	0.032	86	2	12	60
	GCV4	1.88	0.02	0.032	87	2	11	61
	CV	0.77	0.01	0.038	20	1	79	16
	Oracle	2.00	0.00	0.028	100	0	0	100
200	BIC1	1.98	0.00	0.012	98	0	2	59
	BIC2	1.98	0.00	0.012	98	0	2	59
	GCV1	1.20	0.00	0.012	34	0	66	19
	GCV2	1.65	0.00	0.013	73	0	27	41
	GCV3	1.86	0.00	0.012	86	0	14	50
	GCV4	1.89	0.00	0.012	89	0	11	53
	CV	0.75	0.00	0.016	18	0	82	11
	Oracle	2.00	0.00	0.009	100	0	0	100
300	BIC1	1.96	0.00	0.007	97	0	3	50
	BIC2	1.96	0.00	0.007	97	0	3	50
	GCV1	1.15	0.00	0.006	32	0	68	18
	GCV2	1.64	0.00	0.008	71	0	29	35
	GCV3	1.84	0.00	0.007	85	0	15	40
	GCV4	1.86	0.00	0.007	87	0	13	41
	CV	0.58	0.00	0.008	15	0	85	8
	Oracle	2.00	0.00	0.010	100	0	0	100

Table B.6: Fixed effect selection and estimation results for Setting 1 and Case 2 when applying the ALasso2

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.95	0.23	0.060	87	10	3	51
	BIC2	1.91	0.06	0.059	90	3	7	52
	GCV1	1.21	0.05	0.056	38	3	59	21
	GCV2	1.64	0.06	0.060	68	3	29	40
	GCV3	1.86	0.06	0.055	85	3	12	48
	GCV4	1.86	0.06	0.057	85	3	12	48
	CV	0.99	0.11	0.091	26	7	67	14
	Oracle	2.00	0.00	0.049	100	0	0	100
100	BIC1	1.97	0.12	0.031	92	5	3	65
	BIC2	1.95	0.03	0.031	93	2	5	65
	GCV1	1.25	0.02	0.032	34	2	64	25
	GCV2	1.73	0.02	0.033	73	2	25	50
	GCV3	1.89	0.03	0.030	87	2	11	60
	GCV4	1.89	0.03	0.030	87	2	11	60
	CV	0.79	0.06	0.041	17	5	78	12
	Oracle	2.00	0.00	0.028	100	0	0	100
200	BIC1	1.98	0.00	0.010	98	0	2	59
	BIC2	1.98	0.00	0.010	98	0	2	59
	GCV1	1.22	0.00	0.011	37	0	63	20
	GCV2	1.65	0.00	0.013	73	0	27	41
	GCV3	1.88	0.00	0.011	88	0	12	52
	GCV4	1.90	0.00	0.010	90	0	10	54
	CV	0.90	0.06	0.017	19	5	76	11
	Oracle	2.00	0.00	0.009	100	0	0	100
300	BIC1	1.99	0.00	0.006	99	0	1	50
	BIC2	1.99	0.00	0.006	99	0	1	50
	GCV1	1.15	0.00	0.006	31	0	69	15
	GCV2	1.68	0.00	0.007	75	0	25	37
	GCV3	1.88	0.00	0.006	88	0	12	41
	GCV4	1.90	0.00	0.006	90	0	10	42
	CV	0.82	0.04	0.008	13	4	83	7
	Oracle	2.00	0.00	0.010	100	0	0	100

Table B.7: Fixed effect selection and estimation results for Setting 1 and Case 3 when applying the Lasso

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.25	0.23	0.095	33	10	57	21
	BIC2	1.16	0.10	0.082	30	6	64	20
	GCV1	0.58	0.06	0.073	4	3	93	2
	GCV2	0.86	0.06	0.067	20	3	77	14
	GCV3	1.11	0.07	0.080	27	4	69	17
	GCV4	1.12	0.07	0.076	27	4	69	17
	CV	0.46	0.06	0.069	7	3	90	4
	Oracle	2.00	0.00	0.049	100	0	0	100
100	BIC1	1.24	0.13	0.046	37	5	58	26
	BIC2	1.15	0.07	0.046	33	3	64	22
	GCV1	0.45	0.06	0.034	2	3	95	1
	GCV2	0.69	0.07	0.038	14	3	83	10
	GCV3	0.98	0.07	0.045	24	3	73	17
	GCV4	0.98	0.07	0.045	24	3	73	17
	CV	0.43	0.07	0.037	6	3	91	3
	Oracle	2.00	0.00	0.028	100	0	0	100
200	BIC1	1.25	0.00	0.026	42	0	58	26
	BIC2	1.19	0.00	0.023	39	0	61	23
	GCV1	0.43	0.00	0.014	2	0	98	1
	GCV2	0.71	0.00	0.013	15	0	85	11
	GCV3	1.01	0.00	0.020	27	0	73	18
	GCV4	1.03	0.00	0.019	28	0	72	18
	CV	0.38	0.00	0.017	6	0	94	5
	Oracle	2.00	0.00	0.009	100	0	0	100
300	BIC1	1.33	0.00	0.016	48	0	52	31
	BIC2	1.30	0.00	0.016	46	0	54	30
	GCV1	0.47	0.00	0.009	1	0	99	1
	GCV2	0.74	0.00	0.009	16	0	84	11
	GCV3	0.94	0.00	0.011	21	0	79	13
	GCV4	0.96	0.00	0.011	23	0	77	14
	CV	0.42	0.00	0.009	6	0	94	4
	Oracle	2.00	0.00	0.010	100	0	0	100

Table B.8: Fixed effect selection and estimation results for Setting 1 and Case 3 when applying the ALasso1

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.92	0.23	0.065	84	10	6	55
	BIC2	1.92	0.08	0.062	89	5	6	57
	GCV1	1.28	0.06	0.058	41	3	56	24
	GCV2	1.69	0.06	0.058	72	3	25	41
	GCV3	1.84	0.06	0.058	83	3	14	50
	GCV4	1.85	0.07	0.059	83	4	13	50
	CV	0.90	0.06	0.070	29	3	68	15
	Oracle	2.00	0.00	0.049	100	0	0	100
100	BIC1	1.98	0.11	0.036	93	5	2	63
	BIC2	1.97	0.05	0.036	95	3	2	64
	GCV1	1.22	0.03	0.032	32	3	65	23
	GCV2	1.71	0.05	0.034	72	3	25	46
	GCV3	1.88	0.05	0.034	87	3	10	58
	GCV4	1.92	0.05	0.035	90	3	7	60
	CV	0.83	0.02	0.042	25	2	73	19
	Oracle	2.00	0.00	0.028	100	0	0	100
200	BIC1	1.98	0.00	0.013	98	0	2	63
	BIC2	1.98	0.00	0.013	98	0	2	63
	GCV1	1.14	0.00	0.012	29	0	71	19
	GCV2	1.60	0.00	0.013	70	0	30	41
	GCV3	1.88	0.00	0.012	88	0	12	55
	GCV4	1.89	0.00	0.011	89	0	11	56
	CV	0.66	0.00	0.016	17	0	83	11
	Oracle	2.00	0.00	0.009	100	0	0	100
300	BIC1	1.97	0.00	0.007	98	0	2	59
	BIC2	1.97	0.00	0.007	98	0	2	59
	GCV1	1.16	0.00	0.006	31	0	69	18
	GCV2	1.60	0.00	0.008	67	0	33	36
	GCV3	1.83	0.00	0.007	84	0	16	49
	GCV4	1.87	0.00	0.007	88	0	12	51
	CV	0.61	0.00	0.009	16	0	84	8
	Oracle	2.00	0.00	0.010	100	0	0	100

Table B.9: Fixed effect selection and estimation results for Setting 1 and Case 3 when applying the ALasso2

m	Criterion$_F$	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.93	0.23	0.063	84	10	6	55
	BIC2	1.91	0.08	0.061	88	5	7	56
	GCV1	1.26	0.05	0.056	41	3	56	24
	GCV2	1.70	0.06	0.059	73	3	24	43
	GCV3	1.85	0.06	0.057	84	3	13	51
	GCV4	1.87	0.06	0.057	86	3	11	53
	CV	0.97	0.14	0.076	25	8	67	18
	Oracle	2.00	0.00	0.049	100	0	0	100
100	BIC1	1.98	0.11	0.033	93	5	2	63
	BIC2	1.97	0.05	0.033	95	3	2	64
	GCV1	1.29	0.03	0.032	36	3	61	23
	GCV2	1.74	0.05	0.034	72	3	25	47
	GCV3	1.89	0.05	0.030	87	3	10	58
	GCV4	1.90	0.05	0.031	88	3	9	59
	CV	0.91	0.09	0.049	21	7	72	14
	Oracle	2.00	0.00	0.028	100	0	0	100
200	BIC1	1.98	0.00	0.011	98	0	2	63
	BIC2	1.98	0.00	0.011	98	0	2	63
	GCV1	1.17	0.00	0.010	36	0	64	24
	GCV2	1.65	0.00	0.012	73	0	27	45
	GCV3	1.88	0.00	0.010	88	0	12	55
	GCV4	1.89	0.00	0.010	89	0	11	55
	CV	0.79	0.10	0.016	17	8	75	12
	Oracle	2.00	0.00	0.009	100	0	0	100
300	BIC1	1.99	0.00	0.006	99	0	1	59
	BIC2	1.99	0.00	0.006	99	0	1	59
	GCV1	1.18	0.00	0.006	33	0	67	21
	GCV2	1.67	0.00	0.007	72	0	28	40
	GCV3	1.86	0.00	0.006	86	0	14	50
	GCV4	1.89	0.00	0.006	89	0	11	52
	CV	0.76	0.05	0.008	13	4	83	9
	Oracle	2.00	0.00	0.010	100	0	0	100

Table B.10: Fixed effect selection and estimation results for Setting 1 and Case 4 when applying the Lasso

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.31	0.01	0.098	45	1	54	30
	BIC2	1.23	0.00	0.087	40	0	60	25
	GCV1	0.44	0.00	0.071	3	0	97	1
	GCV2	0.79	0.00	0.071	15	0	85	10
	GCV3	1.11	0.00	0.085	29	0	71	18
	GCV4	1.12	0.00	0.080	30	0	70	18
	CV	0.44	0.00	0.070	8	0	92	5
	Oracle	2.00	0.00	0.044	100	0	0	100
100	BIC1	1.25	0.00	0.067	45	0	55	33
	BIC2	1.22	0.00	0.061	42	0	58	31
	GCV1	0.37	0.00	0.048	2	0	98	1
	GCV2	0.70	0.00	0.044	16	0	84	10
	GCV3	0.97	0.00	0.055	22	0	78	16
	GCV4	0.99	0.00	0.053	23	0	77	17
	CV	0.33	0.00	0.042	8	0	92	6
	Oracle	2.00	0.00	0.029	100	0	0	100
200	BIC1	1.23	0.00	0.027	38	0	62	32
	BIC2	1.22	0.00	0.026	37	0	63	31
	GCV1	0.38	0.00	0.023	3	0	97	3
	GCV2	0.72	0.00	0.024	17	0	83	16
	GCV3	1.00	0.00	0.025	23	0	77	21
	GCV4	1.02	0.00	0.025	25	0	75	22
	CV	0.38	0.00	0.026	6	0	94	5
	Oracle	2.00	0.00	0.017	100	0	0	100
300	BIC1	1.29	0.00	0.019	40	0	60	30
	BIC2	1.27	0.00	0.019	40	0	60	30
	GCV1	0.39	0.00	0.012	1	0	99	1
	GCV2	0.72	0.00	0.013	12	0	88	10
	GCV3	0.96	0.00	0.013	20	0	80	17
	GCV4	1.01	0.00	0.013	24	0	76	18
	CV	0.38	0.00	0.013	6	0	94	5
	Oracle	2.00	0.00	0.010	100	0	0	100

Table B.11: Fixed effect selection and estimation results for Setting 1 and Case 4 when applying the ALasso1

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.97	0.00	0.057	97	0	3	63
	BIC2	1.96	0.00	0.054	96	0	4	63
	GCV1	1.28	0.00	0.055	41	0	59	28
	GCV2	1.73	0.00	0.056	76	0	24	50
	GCV3	1.88	0.00	0.057	88	0	12	58
	GCV4	1.89	0.00	0.054	89	0	11	58
	CV	0.89	0.01	0.058	23	1	76	17
	Oracle	2.00	0.00	0.044	100	0	0	100
100	BIC1	1.95	0.00	0.032	96	0	4	71
	BIC2	1.93	0.00	0.032	94	0	6	69
	GCV1	1.20	0.00	0.033	37	0	63	25
	GCV2	1.55	0.00	0.035	66	0	34	48
	GCV3	1.79	0.00	0.034	82	0	18	61
	GCV4	1.80	0.00	0.031	83	0	17	62
	CV	0.73	0.01	0.039	17	1	82	14
	Oracle	2.00	0.00	0.029	100	0	0	100
200	BIC1	1.98	0.00	0.018	98	0	2	77
	BIC2	1.98	0.00	0.018	98	0	2	77
	GCV1	1.14	0.00	0.018	29	0	71	24
	GCV2	1.64	0.00	0.019	69	0	31	54
	GCV3	1.89	0.00	0.018	89	0	11	71
	GCV4	1.90	0.00	0.018	90	0	10	71
	CV	0.72	0.00	0.025	21	0	79	16
	Oracle	2.00	0.00	0.017	100	0	0	100
300	BIC1	1.98	0.00	0.008	99	0	1	66
	BIC2	1.98	0.00	0.008	99	0	1	66
	GCV1	1.18	0.00	0.010	31	0	69	23
	GCV2	1.67	0.00	0.010	70	0	30	47
	GCV3	1.83	0.00	0.010	85	0	15	57
	GCV4	1.84	0.00	0.010	86	0	14	58
	CV	0.68	0.02	0.011	17	2	81	14
	Oracle	2.00	0.00	0.010	100	0	0	100

Table B.12: Fixed effect selection and estimation results for Setting 1 and Case 4 when applying the ALasso2

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.99	0.00	0.055	99	0	1	65
	BIC2	1.98	0.00	0.054	98	0	2	65
	GCV1	1.29	0.00	0.054	40	0	60	27
	GCV2	1.74	0.00	0.055	76	0	24	50
	GCV3	1.89	0.00	0.059	89	0	11	59
	GCV4	1.91	0.00	0.055	91	0	9	60
	CV	1.13	0.11	0.072	30	9	61	22
	Oracle	2.00	0.00	0.044	100	0	0	100
100	BIC1	1.96	0.00	0.033	96	0	4	71
	BIC2	1.94	0.00	0.031	94	0	6	70
	GCV1	1.16	0.00	0.031	34	0	66	24
	GCV2	1.59	0.00	0.034	67	0	33	48
	GCV3	1.81	0.00	0.032	83	0	17	62
	GCV4	1.82	0.00	0.033	83	0	17	62
	CV	0.86	0.04	0.047	15	4	81	12
	Oracle	2.00	0.00	0.029	100	0	0	100
200	BIC1	1.98	0.00	0.017	98	0	2	77
	BIC2	1.98	0.00	0.017	98	0	2	77
	GCV1	1.19	0.00	0.018	30	0	70	24
	GCV2	1.63	0.00	0.020	69	0	31	54
	GCV3	1.87	0.00	0.018	88	0	12	70
	GCV4	1.88	0.00	0.017	89	0	11	70
	CV	0.86	0.06	0.027	17	5	78	14
	Oracle	2.00	0.00	0.017	100	0	0	100
300	BIC1	2.00	0.00	0.007	100	0	0	66
	BIC2	1.99	0.00	0.007	99	0	1	65
	GCV1	1.24	0.00	0.010	34	0	66	24
	GCV2	1.66	0.00	0.010	69	0	31	44
	GCV3	1.84	0.00	0.008	85	0	15	56
	GCV4	1.87	0.00	0.008	88	0	12	57
	CV	0.82	0.03	0.013	18	3	79	15
	Oracle	2.00	0.00	0.010	100	0	0	100

Table B.13: Fixed effect selection and estimation results for Setting 1 and Case 5 when applying the Lasso

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.59	0.06	0.036	63	2	35	42
	BIC2	1.50	0.00	0.035	58	0	42	38
	GCV1	0.53	0.00	0.025	9	0	91	5
	GCV2	0.91	0.00	0.029	26	0	74	17
	GCV3	1.26	0.00	0.026	37	0	63	24
	GCV4	1.27	0.00	0.027	38	0	62	25
	CV	0.55	0.00	0.030	15	0	85	10
	Oracle	2.00	0.00	0.014	100	0	0	100
100	BIC1	1.76	0.03	0.014	77	1	22	51
	BIC2	1.70	0.00	0.014	75	0	25	49
	GCV1	0.53	0.00	0.010	8	0	92	4
	GCV2	1.06	0.00	0.012	38	0	62	23
	GCV3	1.35	0.00	0.011	50	0	50	31
	GCV4	1.39	0.00	0.011	51	0	49	32
	CV	0.65	0.00	0.011	19	0	81	13
	Oracle	2.00	0.00	0.006	100	0	0	100
200	BIC1	1.76	0.00	0.008	79	0	21	59
	BIC2	1.74	0.00	0.008	78	0	22	58
	GCV1	0.55	0.00	0.006	4	0	96	2
	GCV2	1.02	0.00	0.007	29	0	71	20
	GCV3	1.38	0.00	0.007	48	0	52	37
	GCV4	1.41	0.00	0.006	51	0	49	39
	CV	0.65	0.00	0.006	16	0	84	10
	Oracle	2.00	0.00	0.004	100	0	0	100
300	BIC1	1.77	0.00	0.005	78	0	22	50
	BIC2	1.75	0.00	0.005	76	0	24	48
	GCV1	0.63	0.00	0.003	10	0	90	8
	GCV2	0.95	0.00	0.004	29	0	71	19
	GCV3	1.29	0.00	0.004	43	0	57	26
	GCV4	1.34	0.00	0.004	47	0	53	29
	CV	0.62	0.00	0.004	14	0	86	10
	Oracle	2.00	0.00	0.002	100	0	0	100

Table B.14: Fixed effect selection and estimation results for Setting 1 and Case 5 when applying the ALasso1

m	Criterion$_F$	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.91	0.06	0.015	91	2	7	55
	BIC2	1.88	0.00	0.018	89	0	11	53
	GCV1	1.03	0.00	0.022	21	0	79	13
	GCV2	1.57	0.00	0.024	66	0	34	41
	GCV3	1.75	0.00	0.021	80	0	20	49
	GCV4	1.77	0.00	0.019	81	0	19	50
	CV	0.65	0.00	0.027	13	0	87	9
	Oracle	2.00	0.00	0.014	100	0	0	100
100	BIC1	1.97	0.03	0.007	96	1	3	66
	BIC2	1.93	0.00	0.007	93	0	7	64
	GCV1	1.08	0.00	0.008	24	0	76	16
	GCV2	1.68	0.00	0.009	72	0	28	48
	GCV3	1.86	0.00	0.007	86	0	14	60
	GCV4	1.87	0.00	0.008	87	0	13	61
	CV	0.54	0.00	0.009	12	0	88	10
	Oracle	2.00	0.00	0.006	100	0	0	100
200	BIC1	1.97	0.00	0.005	97	0	3	68
	BIC2	1.97	0.00	0.005	97	0	3	68
	GCV1	0.95	0.00	0.005	16	0	84	11
	GCV2	1.73	0.00	0.005	78	0	22	58
	GCV3	1.84	0.00	0.005	86	0	14	61
	GCV4	1.87	0.00	0.005	88	0	12	63
	CV	0.55	0.01	0.006	8	1	91	5
	Oracle	2.00	0.00	0.004	100	0	0	100
300	BIC1	2.00	0.00	0.002	100	0	0	64
	BIC2	2.00	0.00	0.002	100	0	0	64
	GCV1	1.02	0.00	0.003	20	0	80	14
	GCV2	1.69	0.00	0.003	73	0	27	47
	GCV3	1.89	0.00	0.003	89	0	11	57
	GCV4	1.90	0.00	0.003	90	0	10	58
	CV	0.67	0.00	0.004	12	0	88	10
	Oracle	2.00	0.00	0.002	100	0	0	100

Table B.15: Fixed effect selection and estimation results for Setting 1 and Case 5 when applying the ALasso2

m	Criterion$_F$	CZ	IZ	MME	C	U	O	Both
50	BIC1	1.94	0.06	0.016	92	2	6	57
	BIC2	1.90	0.00	0.018	90	0	10	54
	GCV1	1.08	0.00	0.022	25	0	75	17
	GCV2	1.59	0.00	0.023	66	0	34	40
	GCV3	1.76	0.00	0.020	79	0	21	47
	GCV4	1.81	0.00	0.019	83	0	17	50
	CV	0.76	0.07	0.027	14	6	80	8
	Oracle	2.00	0.00	0.014	100	0	0	100
100	BIC1	1.99	0.03	0.007	98	1	1	67
	BIC2	1.96	0.00	0.007	96	0	4	65
	GCV1	1.14	0.00	0.008	33	0	67	22
	GCV2	1.73	0.00	0.010	77	0	23	52
	GCV3	1.88	0.00	0.007	89	0	11	61
	GCV4	1.91	0.00	0.008	91	0	9	62
	CV	0.70	0.02	0.010	17	1	82	14
	Oracle	2.00	0.00	0.006	100	0	0	100
200	BIC1	1.99	0.00	0.004	99	0	1	68
	BIC2	1.98	0.00	0.004	98	0	2	68
	GCV1	1.02	0.00	0.005	16	0	84	11
	GCV2	1.78	0.00	0.005	82	0	18	60
	GCV3	1.90	0.00	0.004	91	0	9	63
	GCV4	1.93	0.00	0.004	93	0	7	64
	CV	0.70	0.04	0.007	12	4	84	9
	Oracle	2.00	0.00	0.004	100	0	0	100
300	BIC1	2.00	0.00	0.002	100	0	0	64
	BIC2	2.00	0.00	0.002	100	0	0	64
	GCV1	0.82	0.00	0.003	0	0	100	0
	GCV2	1.73	0.00	0.003	75	0	25	49
	GCV3	1.90	0.00	0.003	90	0	10	57
	GCV4	1.92	0.00	0.002	92	0	8	59
	CV	0.82	0.04	0.005	20	4	76	16
	Oracle	2.00	0.00	0.002	100	0	0	100

Table B.16: Fixed effect selection and estimation results for Setting 2 and Case 1 when applying the Lasso

m	Criterion$_F$	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.34	0.75	0.414	36	42	22	7
	BIC2	6.59	0.12	0.197	36	8	56	6
	GCV1	3.25	0.06	0.125	0	4	96	0
	GCV2	4.35	0.07	0.132	2	5	93	0
	GCV3	5.53	0.08	0.146	5	5	90	0
	GCV4	5.67	0.09	0.146	8	6	86	1
	CV	4.79	0.16	0.152	2	9	89	0
	Oracle	8.00	0.00	0.038	100	0	0	100
100	BIC1	7.44	0.14	0.112	64	8	28	20
	BIC2	6.95	0.02	0.080	45	1	54	16
	GCV1	3.11	0.02	0.050	0	1	99	0
	GCV2	4.63	0.02	0.057	8	1	91	5
	GCV3	5.82	0.02	0.062	14	1	85	6
	GCV4	5.97	0.02	0.063	14	1	85	6
	CV	4.17	0.06	0.057	6	3	91	3
	Oracle	8.00	0.00	0.018	100	0	0	100
200	BIC1	7.63	0.00	0.053	70	0	30	36
	BIC2	7.23	0.00	0.045	49	0	51	23
	GCV1	3.17	0.00	0.025	0	0	100	0
	GCV2	4.77	0.00	0.030	7	0	93	2
	GCV3	5.95	0.00	0.031	13	0	87	3
	GCV4	6.15	0.00	0.032	16	0	84	5
	CV	4.13	0.00	0.032	3	0	97	1
	Oracle	8.00	0.00	0.010	100	0	0	100
300	BIC1	7.54	0.00	0.027	66	0	34	36
	BIC2	7.34	0.00	0.024	54	0	46	33
	GCV1	3.27	0.00	0.015	0	0	100	0
	GCV2	4.87	0.00	0.018	8	0	92	6
	GCV3	6.03	0.00	0.018	9	0	91	7
	GCV4	6.14	0.00	0.018	12	0	88	8
	CV	4.19	0.00	0.016	5	0	95	4
	Oracle	8.00	0.00	0.006	100	0	0	100

Table B.17: Fixed effect selection and estimation results for Setting 2 and Case 1 when applying the ALasso1

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.58	0.64	0.201	48	36	16	6
	BIC2	7.34	0.05	0.094	62	4	34	8
	GCV1	4.51	0.03	0.107	3	2	95	0
	GCV2	6.04	0.03	0.107	21	2	77	2
	GCV3	6.63	0.04	0.099	34	3	63	4
	GCV4	6.74	0.04	0.099	38	3	59	4
	CV	6.11	0.06	0.113	29	4	67	1
	Oracle	8.00	0.00	0.038	100	0	0	100
100	BIC1	7.71	0.13	0.032	82	8	10	24
	BIC2	7.59	0.02	0.029	78	1	21	24
	GCV1	4.00	0.02	0.046	3	1	96	1
	GCV2	6.56	0.02	0.038	32	1	67	10
	GCV3	7.08	0.02	0.033	52	1	47	15
	GCV4	7.16	0.02	0.031	52	1	47	15
	CV	5.70	0.02	0.046	24	1	75	5
	Oracle	8.00	0.00	0.018	100	0	0	100
200	BIC1	7.83	0.00	0.014	90	0	10	41
	BIC2	7.79	0.00	0.014	86	0	14	39
	GCV1	3.84	0.00	0.024	1	0	99	0
	GCV2	6.49	0.00	0.019	33	0	67	12
	GCV3	7.13	0.00	0.018	55	0	45	27
	GCV4	7.23	0.00	0.017	55	0	45	27
	CV	5.61	0.00	0.027	20	0	80	10
	Oracle	8.00	0.00	0.010	100	0	0	100
300	BIC1	7.92	0.00	0.008	96	0	4	50
	BIC2	7.91	0.00	0.008	95	0	5	49
	GCV1	4.17	0.00	0.013	3	0	97	3
	GCV2	6.64	0.00	0.011	36	0	64	23
	GCV3	7.26	0.00	0.010	60	0	40	33
	GCV4	7.29	0.00	0.009	58	0	42	34
	CV	5.84	0.00	0.012	26	0	74	16
	Oracle	8.00	0.00	0.006	100	0	0	100

Table B.18: Fixed effect selection and estimation results for Setting 2 and Case 1 when applying the ALasso2

m	Criterion$_F$	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.63	0.66	0.197	50	38	12	7
	BIC2	7.44	0.09	0.081	66	6	28	10
	GCV1	4.30	0.06	0.109	2	4	94	0
	GCV2	6.09	0.08	0.107	27	5	68	4
	GCV3	6.76	0.09	0.098	41	6	53	5
	GCV4	6.85	0.09	0.092	42	6	52	5
	CV	6.96	0.12	0.191	40	9	51	3
	Oracle	8.00	0.00	0.038	100	0	0	100
100	BIC1	7.77	0.12	0.026	85	7	8	27
	BIC2	7.75	0.02	0.027	83	1	16	23
	GCV1	4.08	0.02	0.047	1	1	98	1
	GCV2	6.71	0.02	0.036	33	1	66	9
	GCV3	7.20	0.02	0.032	53	1	46	16
	GCV4	7.21	0.02	0.032	53	1	46	16
	CV	6.91	0.04	0.060	45	3	52	14
	Oracle	8.00	0.00	0.018	100	0	0	100
200	BIC1	7.96	0.00	0.014	97	0	3	42
	BIC2	7.91	0.00	0.015	92	0	8	41
	GCV1	4.01	0.00	0.025	2	0	98	0
	GCV2	6.56	0.00	0.021	33	0	67	13
	GCV3	7.27	0.00	0.018	58	0	42	27
	GCV4	7.32	0.00	0.016	57	0	43	28
	CV	6.93	0.00	0.035	51	0	49	25
	Oracle	8.00	0.00	0.010	100	0	0	100
300	BIC1	7.99	0.00	0.007	99	0	1	51
	BIC2	7.99	0.00	0.007	99	0	1	51
	GCV1	4.12	0.00	0.014	2	0	98	2
	GCV2	6.73	0.00	0.011	40	0	60	24
	GCV3	7.30	0.00	0.008	60	0	40	36
	GCV4	7.36	0.00	0.008	60	0	40	37
	CV	6.88	0.00	0.026	52	0	48	25
	Oracle	8.00	0.00	0.006	100	0	0	100

Table B.19: Fixed effect selection and estimation results for Setting 2 and Case 2 when applying the Lasso

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.49	1.41	1.000	17	75	8	0
	BIC2	6.31	0.20	0.189	24	15	61	0
	GCV1	3.12	0.07	0.125	0	5	95	0
	GCV2	4.20	0.07	0.137	5	4	91	0
	GCV3	5.33	0.10	0.149	8	7	85	0
	GCV4	5.61	0.11	0.150	10	8	82	0
	CV	4.44	0.18	0.149	5	11	84	0
	Oracle	8.00	0.00	0.053	100	0	0	100
100	BIC1	7.14	0.55	0.186	40	31	29	2
	BIC2	6.48	0.08	0.081	27	7	66	1
	GCV1	3.00	0.02	0.058	0	1	99	0
	GCV2	4.34	0.03	0.061	3	2	95	0
	GCV3	5.40	0.05	0.064	9	4	87	0
	GCV4	5.56	0.05	0.067	12	4	84	0
	CV	4.32	0.15	0.065	5	8	87	0
	Oracle	8.00	0.00	0.018	100	0	0	100
200	BIC1	7.39	0.14	0.054	59	7	34	10
	BIC2	6.89	0.00	0.043	43	0	57	7
	GCV1	3.24	0.00	0.029	0	0	100	0
	GCV2	4.64	0.00	0.032	5	0	95	0
	GCV3	5.68	0.00	0.030	9	0	91	2
	GCV4	5.89	0.00	0.033	13	0	87	3
	CV	4.33	0.01	0.035	2	1	97	0
	Oracle	8.00	0.00	0.012	100	0	0	100
300	BIC1	7.44	0.02	0.029	64	1	35	7
	BIC2	7.02	0.00	0.023	41	0	59	4
	GCV1	3.11	0.00	0.013	0	0	100	0
	GCV2	4.69	0.00	0.015	3	0	97	1
	GCV3	5.80	0.00	0.018	14	0	86	1
	GCV4	5.94	0.00	0.018	15	0	85	1
	CV	4.34	0.00	0.018	5	0	95	1
	Oracle	8.00	0.00	0.005	100	0	0	100

Table B.20: Fixed effect selection and estimation results for Setting 2 and Case 2 when applying the ALasso1

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.66	1.24	1.000	24	67	9	0
	BIC2	7.15	0.03	0.103	58	2	40	0
	GCV1	4.32	0.04	0.108	6	3	91	0
	GCV2	5.97	0.04	0.113	25	3	72	0
	GCV3	6.60	0.04	0.100	37	3	60	0
	GCV4	6.71	0.04	0.101	39	3	58	0
	CV	6.44	0.10	0.144	33	8	59	0
	Oracle	8.00	0.00	0.053	100	0	0	100
100	BIC1	7.56	0.47	0.079	59	26	15	2
	BIC2	7.40	0.02	0.039	70	2	28	2
	GCV1	4.00	0.01	0.051	2	1	97	0
	GCV2	5.99	0.01	0.051	23	1	76	0
	GCV3	6.56	0.01	0.042	37	1	62	1
	GCV4	6.69	0.01	0.039	38	1	61	2
	CV	6.16	0.10	0.066	28	7	65	0
	Oracle	8.00	0.00	0.018	100	0	0	100
200	BIC1	7.92	0.14	0.019	87	7	6	12
	BIC2	7.87	0.00	0.019	89	0	11	11
	GCV1	4.12	0.00	0.022	3	0	97	0
	GCV2	6.32	0.00	0.023	32	0	68	4
	GCV3	7.13	0.00	0.021	55	0	45	6
	GCV4	7.22	0.00	0.020	58	0	42	6
	CV	5.48	0.00	0.023	16	0	84	3
	Oracle	8.00	0.00	0.012	100	0	0	100
300	BIC1	7.88	0.02	0.008	92	1	7	11
	BIC2	7.83	0.00	0.008	88	0	12	11
	GCV1	4.03	0.00	0.013	2	0	98	0
	GCV2	6.60	0.00	0.012	36	0	64	1
	GCV3	7.14	0.00	0.011	54	0	46	5
	GCV4	7.20	0.00	0.010	55	0	45	5
	CV	5.27	0.00	0.012	14	0	86	0
	Oracle	8.00	0.00	0.005	100	0	0	100

Table B.21: Fixed effect selection and estimation results for Setting 2 and Case 2 when applying the ALasso2

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.65	1.27	1.000	22	71	7	0
	BIC2	7.27	0.09	0.103	59	6	35	0
	GCV1	4.30	0.05	0.114	3	3	94	0
	GCV2	6.11	0.05	0.115	29	3	68	0
	GCV3	6.77	0.05	0.105	44	3	53	0
	GCV4	6.83	0.05	0.103	45	3	52	0
	CV	6.92	0.22	0.193	40	17	43	0
	Oracle	8.00	0.00	0.053	100	0	0	100
100	BIC1	7.62	0.47	0.075	61	26	13	2
	BIC2	7.51	0.06	0.034	71	5	24	2
	GCV1	4.16	0.01	0.050	2	1	97	0
	GCV2	6.11	0.01	0.050	21	1	78	0
	GCV3	6.85	0.01	0.036	35	1	64	2
	GCV4	6.89	0.01	0.035	36	1	63	2
	CV	6.92	0.16	0.085	41	11	48	0
	Oracle	8.00	0.00	0.018	100	0	0	100
200	BIC1	7.89	0.14	0.020	85	7	8	12
	BIC2	7.87	0.00	0.018	87	0	13	12
	GCV1	4.20	0.00	0.023	2	0	98	0
	GCV2	6.39	0.00	0.022	31	0	69	4
	GCV3	6.93	0.00	0.020	44	0	56	6
	GCV4	7.02	0.00	0.020	45	0	55	6
	CV	6.94	0.02	0.049	51	2	47	7
	Oracle	8.00	0.00	0.012	100	0	0	100
300	BIC1	7.94	0.02	0.007	96	1	3	11
	BIC2	7.91	0.00	0.007	95	0	5	11
	GCV1	4.03	0.00	0.013	2	0	98	0
	GCV2	6.52	0.00	0.012	34	0	66	3
	GCV3	7.21	0.00	0.011	53	0	47	4
	GCV4	7.28	0.00	0.009	55	0	45	4
	CV	6.81	0.00	0.031	54	0	46	4
	Oracle	8.00	0.00	0.005	100	0	0	100

Table B.22: Fixed effect selection and estimation results for Setting 2 and Case 3 when applying the Lasso

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.46	1.46	1.000	11	78	11	1
	BIC2	6.07	0.18	0.202	13	12	75	1
	GCV1	3.01	0.04	0.137	0	3	97	0
	GCV2	3.91	0.05	0.157	3	4	93	1
	GCV3	5.19	0.12	0.167	3	9	88	1
	GCV4	5.29	0.12	0.175	5	9	86	1
	CV	4.27	0.19	0.153	2	11	87	0
	Oracle	8.00	0.00	0.044	100	0	0	100
100	BIC1	6.98	0.49	0.191	34	27	39	7
	BIC2	6.13	0.12	0.090	20	9	71	2
	GCV1	3.16	0.06	0.057	0	4	96	0
	GCV2	4.05	0.07	0.063	3	4	93	0
	GCV3	5.02	0.12	0.061	3	9	88	0
	GCV4	5.19	0.12	0.064	5	9	86	0
	CV	4.45	0.25	0.060	2	14	84	0
	Oracle	8.00	0.00	0.017	100	0	0	100
200	BIC1	7.01	0.14	0.058	38	8	54	8
	BIC2	6.37	0.00	0.047	20	0	80	6
	GCV1	3.11	0.00	0.030	0	0	100	0
	GCV2	4.10	0.00	0.034	3	0	97	1
	GCV3	5.18	0.00	0.035	5	0	95	2
	GCV4	5.41	0.00	0.036	5	0	95	2
	CV	4.01	0.03	0.034	0	2	98	0
	Oracle	8.00	0.00	0.013	100	0	0	100
300	BIC1	7.31	0.06	0.028	51	3	46	20
	BIC2	6.99	0.00	0.025	41	0	59	14
	GCV1	3.33	0.00	0.013	0	0	100	0
	GCV2	4.47	0.00	0.017	4	0	96	1
	GCV3	5.61	0.00	0.018	5	0	95	1
	GCV4	5.84	0.00	0.019	7	0	93	2
	CV	4.21	0.02	0.018	2	1	97	0
	Oracle	8.00	0.00	0.006	100	0	0	100

Table B.23: Fixed effect selection and estimation results for Setting 2 and Case 3 when applying the ALasso1

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.60	1.32	1.000	18	70	12	1
	BIC2	7.25	0.09	0.112	51	7	42	3
	GCV1	4.05	0.02	0.124	1	1	98	1
	GCV2	5.92	0.02	0.121	20	1	79	1
	GCV3	6.65	0.02	0.107	32	1	67	1
	GCV4	6.70	0.02	0.111	33	1	66	1
	CV	6.33	0.09	0.145	29	6	65	1
	Oracle	8.00	0.00	0.044	100	0	0	100
100	BIC1	7.53	0.32	0.068	67	17	16	11
	BIC2	7.35	0.03	0.038	69	2	29	10
	GCV1	4.07	0.01	0.051	1	1	98	0
	GCV2	6.05	0.03	0.044	25	3	72	1
	GCV3	6.64	0.03	0.036	38	2	60	2
	GCV4	6.73	0.03	0.035	41	2	57	2
	CV	6.05	0.19	0.044	15	11	74	1
	Oracle	8.00	0.00	0.017	100	0	0	100
200	BIC1	7.87	0.12	0.022	84	6	10	19
	BIC2	7.81	0.00	0.022	87	0	13	21
	GCV1	4.28	0.00	0.025	3	0	97	0
	GCV2	6.24	0.00	0.023	30	0	70	9
	GCV3	6.96	0.00	0.022	48	0	52	10
	GCV4	7.02	0.00	0.022	50	0	50	11
	CV	5.63	0.03	0.023	19	2	79	3
	Oracle	8.00	0.00	0.013	100	0	0	100
300	BIC1	7.90	0.04	0.008	92	2	6	29
	BIC2	7.86	0.00	0.008	91	0	9	29
	GCV1	4.27	0.00	0.013	5	0	95	1
	GCV2	6.82	0.00	0.011	40	0	60	14
	GCV3	7.39	0.00	0.010	65	0	35	22
	GCV4	7.38	0.00	0.009	62	0	38	22
	CV	5.61	0.00	0.013	16	0	84	8
	Oracle	8.00	0.00	0.006	100	0	0	100

Table B.24: Fixed effect selection and estimation results for Setting 2 and Case 3 when applying the ALasso2

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.65	1.31	1.000	21	70	9	1
	BIC2	7.27	0.10	0.102	55	7	38	4
	GCV1	4.37	0.03	0.126	3	2	95	1
	GCV2	5.94	0.03	0.140	24	2	74	1
	GCV3	6.70	0.04	0.110	37	3	60	2
	GCV4	6.75	0.06	0.105	37	4	59	2
	CV	6.96	0.18	0.191	35	12	53	2
	Oracle	8.00	0.00	0.044	100	0	0	100
100	BIC1	7.55	0.38	0.056	66	22	12	11
	BIC2	7.45	0.09	0.031	71	7	22	11
	GCV1	4.36	0.04	0.051	2	3	95	0
	GCV2	6.14	0.05	0.045	21	3	76	1
	GCV3	6.76	0.08	0.037	38	6	56	3
	GCV4	6.84	0.08	0.036	40	6	54	3
	CV	6.97	0.21	0.078	40	12	48	5
	Oracle	8.00	0.00	0.017	100	0	0	100
200	BIC1	7.89	0.12	0.017	86	6	8	19
	BIC2	7.80	0.00	0.017	85	0	15	20
	GCV1	4.20	0.00	0.025	2	0	98	0
	GCV2	6.42	0.00	0.022	29	0	71	9
	GCV3	7.06	0.00	0.023	48	0	52	11
	GCV4	7.10	0.00	0.021	48	0	52	11
	CV	6.67	0.04	0.036	39	3	58	5
	Oracle	8.00	0.00	0.013	100	0	0	100
300	BIC1	7.98	0.04	0.007	96	2	2	32
	BIC2	7.96	0.00	0.007	96	0	4	31
	GCV1	4.21	0.00	0.014	3	0	97	0
	GCV2	6.67	0.00	0.011	39	0	61	16
	GCV3	7.31	0.00	0.011	60	0	40	21
	GCV4	7.35	0.00	0.009	61	0	39	21
	CV	6.88	0.01	0.024	53	1	46	20
	Oracle	8.00	0.00	0.006	100	0	0	100

Table B.25: Fixed effect selection and estimation results for Setting 2 and Case 4 when applying the Lasso

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.53	0.54	0.299	45	32	23	5
	BIC2	6.76	0.09	0.193	37	7	56	7
	GCV1	3.01	0.02	0.133	0	1	99	0
	GCV2	4.33	0.02	0.160	4	1	95	2
	GCV3	5.44	0.06	0.158	9	4	87	2
	GCV4	5.63	0.06	0.158	11	4	85	2
	CV	4.15	0.07	0.161	6	4	90	1
	Oracle	8.00	0.00	0.041	100	0	0	100
100	BIC1	7.33	0.08	0.106	55	5	40	18
	BIC2	6.85	0.02	0.097	40	2	58	12
	GCV1	3.27	0.00	0.060	0	0	100	0
	GCV2	4.78	0.00	0.066	6	0	94	2
	GCV3	5.72	0.01	0.065	11	1	88	4
	GCV4	5.86	0.01	0.067	12	1	87	4
	CV	4.46	0.03	0.072	9	2	89	3
	Oracle	8.00	0.00	0.023	100	0	0	100
200	BIC1	7.40	0.00	0.062	62	0	38	23
	BIC2	7.17	0.00	0.056	50	0	50	20
	GCV1	3.18	0.00	0.030	0	0	100	0
	GCV2	4.92	0.00	0.037	9	0	91	5
	GCV3	5.76	0.00	0.041	12	0	88	5
	GCV4	5.93	0.00	0.041	13	0	87	6
	CV	4.38	0.00	0.040	7	0	93	2
	Oracle	8.00	0.00	0.013	100	0	0	100
300	BIC1	7.54	0.00	0.037	66	0	34	33
	BIC2	7.43	0.00	0.036	59	0	41	29
	GCV1	3.27	0.00	0.021	0	0	100	0
	GCV2	5.05	0.00	0.024	11	0	89	5
	GCV3	5.92	0.00	0.022	14	0	86	5
	GCV4	6.06	0.00	0.022	16	0	84	5
	CV	4.47	0.00	0.025	7	0	93	5
	Oracle	8.00	0.00	0.008	100	0	0	100

Table B.26: Fixed effect selection and estimation results for Setting 2 and Case 4 when applying the ALasso1

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.72	0.37	0.129	64	22	14	7
	BIC2	7.36	0.03	0.098	60	3	37	7
	GCV1	3.89	0.00	0.128	3	0	97	0
	GCV2	5.84	0.00	0.119	18	0	82	4
	GCV3	6.62	0.00	0.101	35	0	65	5
	GCV4	6.72	0.00	0.096	37	0	63	5
	CV	6.00	0.06	0.135	27	5	68	3
	Oracle	8.00	0.00	0.041	100	0	0	100
100	BIC1	7.93	0.07	0.033	90	4	6	25
	BIC2	7.77	0.01	0.033	83	1	16	23
	GCV1	4.26	0.00	0.057	3	0	97	1
	GCV2	6.24	0.00	0.047	31	0	69	11
	GCV3	6.90	0.00	0.037	46	0	54	16
	GCV4	6.91	0.00	0.037	45	0	55	15
	CV	6.23	0.02	0.047	31	2	67	10
	Oracle	8.00	0.00	0.023	100	0	0	100
200	BIC1	7.90	0.01	0.018	94	1	5	36
	BIC2	7.84	0.00	0.020	89	0	11	34
	GCV1	4.09	0.00	0.031	4	0	96	2
	GCV2	6.33	0.00	0.023	27	0	73	9
	GCV3	7.20	0.00	0.020	56	0	44	16
	GCV4	7.19	0.00	0.020	55	0	45	16
	CV	5.84	0.01	0.023	20	1	79	6
	Oracle	8.00	0.00	0.013	100	0	0	100
300	BIC1	7.93	0.00	0.012	93	0	7	47
	BIC2	7.92	0.00	0.012	92	0	8	47
	GCV1	4.01	0.00	0.020	1	0	99	1
	GCV2	6.58	0.00	0.018	34	0	66	20
	GCV3	7.16	0.00	0.014	55	0	45	30
	GCV4	7.19	0.00	0.014	54	0	46	29
	CV	5.71	0.00	0.017	21	0	79	10
	Oracle	8.00	0.00	0.008	100	0	0	100

Table B.27: Fixed effect selection and estimation results for Setting 2 and Case 4 when applying the ALasso2

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.86	0.42	0.100	70	25	5	8
	BIC2	7.38	0.08	0.094	60	5	35	9
	GCV1	4.14	0.02	0.128	2	1	97	0
	GCV2	5.88	0.03	0.117	16	2	82	5
	GCV3	6.80	0.03	0.094	36	2	62	8
	GCV4	6.83	0.03	0.094	36	2	62	8
	CV	6.89	0.13	0.184	42	10	48	6
	Oracle	8.00	0.00	0.041	100	0	0	100
100	BIC1	7.86	0.06	0.030	88	3	9	25
	BIC2	7.74	0.01	0.031	82	1	17	24
	GCV1	4.18	0.00	0.056	2	0	98	1
	GCV2	6.26	0.00	0.050	25	0	75	8
	GCV3	6.94	0.00	0.038	47	0	53	17
	GCV4	7.04	0.00	0.036	51	0	49	18
	CV	6.96	0.04	0.067	50	3	47	10
	Oracle	8.00	0.00	0.023	100	0	0	100
200	BIC1	7.93	0.00	0.017	98	0	2	37
	BIC2	7.90	0.00	0.018	95	0	5	35
	GCV1	3.91	0.00	0.028	1	0	99	0
	GCV2	6.37	0.00	0.022	33	0	67	12
	GCV3	7.17	0.00	0.022	53	0	47	18
	GCV4	7.26	0.00	0.019	53	0	47	18
	CV	6.85	0.01	0.032	49	1	50	16
	Oracle	8.00	0.00	0.013	100	0	0	100
300	BIC1	7.95	0.00	0.011	95	0	5	47
	BIC2	7.94	0.00	0.011	94	0	6	47
	GCV1	4.19	0.00	0.020	1	0	99	1
	GCV2	6.61	0.00	0.018	38	0	62	21
	GCV3	7.09	0.00	0.016	54	0	46	30
	GCV4	7.22	0.00	0.015	55	0	45	30
	CV	6.66	0.05	0.025	39	5	56	19
	Oracle	8.00	0.00	0.008	100	0	0	100

Table B.28: Fixed effect selection and estimation results for Setting 2 and Case 5 when applying the Lasso

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.04	0.28	0.109	59	14	27	8
	BIC2	6.42	0.00	0.073	45	0	55	9
	GCV1	2.79	0.00	0.072	0	0	100	0
	GCV2	4.56	0.00	0.072	12	0	88	1
	GCV3	5.54	0.00	0.066	23	0	77	3
	GCV4	5.61	0.00	0.066	23	0	77	3
	CV	4.72	0.00	0.070	12	0	88	0
	Oracle	8.00	0.00	0.014	100	0	0	100
100	BIC1	7.49	0.02	0.040	86	1	13	20
	BIC2	6.95	0.00	0.027	61	0	39	14
	GCV1	3.07	0.00	0.025	0	0	100	0
	GCV2	5.38	0.00	0.028	24	0	76	4
	GCV3	6.02	0.00	0.025	29	0	71	7
	GCV4	6.20	0.00	0.025	32	0	68	8
	CV	4.94	0.01	0.030	16	1	83	2
	Oracle	8.00	0.00	0.006	100	0	0	100
200	BIC1	7.67	0.02	0.013	84	1	15	28
	BIC2	7.52	0.00	0.012	70	0	30	24
	GCV1	3.40	0.00	0.010	0	0	100	0
	GCV2	5.56	0.00	0.011	18	0	82	5
	GCV3	6.38	0.00	0.010	31	0	69	11
	GCV4	6.48	0.00	0.010	33	0	67	11
	CV	5.33	0.00	0.012	19	0	81	4
	Oracle	8.00	0.00	0.003	100	0	0	100
300	BIC1	7.75	0.02	0.007	83	1	16	39
	BIC2	7.67	0.00	0.007	78	0	22	37
	GCV1	3.38	0.00	0.006	0	0	100	0
	GCV2	5.53	0.00	0.007	14	0	86	7
	GCV3	6.32	0.00	0.006	24	0	76	9
	GCV4	6.50	0.00	0.006	27	0	73	11
	CV	5.52	0.00	0.007	18	0	82	5
	Oracle	8.00	0.00	0.002	100	0	0	100

Table B.29: Fixed effect selection and estimation results for Setting 2 and Case 5 when applying the ALasso1

m	Criterion_F	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.22	0.28	0.030	67	14	19	10
	BIC2	6.72	0.00	0.029	62	0	38	11
	GCV1	3.51	0.00	0.076	5	0	95	0
	GCV2	5.35	0.00	0.060	26	0	74	7
	GCV3	6.12	0.00	0.046	42	0	58	7
	GCV4	6.16	0.00	0.043	42	0	58	8
	CV	5.60	0.00	0.054	19	0	81	2
	Oracle	8.00	0.00	0.014	100	0	0	100
100	BIC1	7.53	0.02	0.012	87	1	12	20
	BIC2	7.40	0.00	0.011	82	0	18	21
	GCV1	3.35	0.00	0.027	1	0	99	0
	GCV2	6.03	0.00	0.016	33	0	67	8
	GCV3	6.60	0.00	0.015	50	0	50	13
	GCV4	6.69	0.00	0.013	51	0	49	13
	CV	5.49	0.02	0.019	17	1	82	3
	Oracle	8.00	0.00	0.006	100	0	0	100
200	BIC1	7.78	0.02	0.004	94	1	5	31
	BIC2	7.68	0.00	0.005	85	0	15	30
	GCV1	3.85	0.00	0.012	2	0	98	1
	GCV2	6.68	0.00	0.008	38	0	62	11
	GCV3	6.95	0.00	0.007	50	0	50	14
	GCV4	6.99	0.00	0.007	50	0	50	15
	CV	5.60	0.00	0.009	18	0	82	2
	Oracle	8.00	0.00	0.003	100	0	0	100
300	BIC1	7.85	0.02	0.003	94	1	5	44
	BIC2	7.79	0.00	0.003	89	0	11	43
	GCV1	3.67	0.00	0.007	0	0	100	0
	GCV2	6.57	0.00	0.005	32	0	68	13
	GCV3	7.08	0.00	0.004	55	0	45	25
	GCV4	7.17	0.00	0.004	55	0	45	25
	CV	5.78	0.00	0.005	22	0	78	11
	Oracle	8.00	0.00	0.002	100	0	0	100

Table B.30: Fixed effect selection and estimation results for Setting 2 and Case 5 when applying the ALasso2

m	Criterion$_F$	CZ	IZ	MME	C	U	O	Both
50	BIC1	7.31	0.26	0.030	68	13	19	10
	BIC2	7.04	0.00	0.031	61	0	39	9
	GCV1	3.68	0.00	0.077	2	0	98	0
	GCV2	5.64	0.00	0.057	26	0	74	5
	GCV3	6.32	0.00	0.049	39	0	61	9
	GCV4	6.43	0.00	0.046	40	0	60	9
	CV	6.86	0.02	0.091	49	2	49	5
	Oracle	8.00	0.00	0.014	100	0	0	100
100	BIC1	7.63	0.02	0.009	90	1	9	21
	BIC2	7.58	0.00	0.009	89	0	11	22
	GCV1	3.80	0.00	0.026	1	0	99	0
	GCV2	6.26	0.00	0.018	40	0	60	9
	GCV3	6.77	0.00	0.014	51	0	49	12
	GCV4	6.78	0.00	0.014	51	0	49	11
	CV	6.55	0.04	0.041	43	3	54	11
	Oracle	8.00	0.00	0.006	100	0	0	100
200	BIC1	7.79	0.02	0.004	94	1	5	31
	BIC2	7.78	0.00	0.005	91	0	9	30
	GCV1	4.10	0.00	0.012	0	0	100	0
	GCV2	6.64	0.00	0.008	40	0	60	14
	GCV3	7.05	0.00	0.007	54	0	46	17
	GCV4	7.16	0.00	0.007	56	0	44	18
	CV	6.83	0.00	0.015	49	0	51	16
	Oracle	8.00	0.00	0.003	100	0	0	100
300	BIC1	7.91	0.02	0.003	96	1	3	43
	BIC2	7.87	0.00	0.003	95	0	5	43
	GCV1	4.03	0.00	0.007	0	0	100	0
	GCV2	6.58	0.00	0.006	37	0	63	18
	GCV3	7.10	0.00	0.004	54	0	46	25
	GCV4	7.13	0.00	0.004	54	0	46	25
	CV	7.07	0.00	0.024	55	0	45	24
	Oracle	8.00	0.00	0.002	100	0	0	100