PARAMETER ESTIMATIONS IN LINEAR MIXED MODELS WITH HEAVY-TAILED AND SKEW DISTRIBUTIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

TUĞBA KAPUCU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
STATISTICS

DECEMBER 2023

Approval of the thesis:

PARAMETER ESTIMATIONS IN LINEAR MIXED MODELS WITH HEAVY-TAILED AND SKEW DISTRIBUTIONS

submitted by **TUĞBA KAPUCU** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Statistics, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, **Graduate School of Natural and Applied Sciences** ________________

Prof. Dr. Özlem İlk Dağ
Head of the Department, **Statistics, METU** ________________

Assoc. Prof. Dr. Fulya Gökalp Yavuz
Supervisor, **Statistics, METU** ________________

Examining Committee Members:

Prof. Dr. Zeynep Kalaylıoğlu
Statistics, METU ________________

Assoc. Prof. Dr. Fulya Gökalp Yavuz
Statistics, METU ________________

Prof. Dr. Olçay Arslan
Statistics, Ankara University ________________

Prof. Dr. Özlem İlk Dağ
Statistics, METU ________________

Assoc. Prof. Dr. Fatma Zehra Doğru
Statistics, Giresun University ________________

Date: 21.12.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name : Tuğba Kapucu

Signature :

ABSTRACT

PARAMETER ESTIMATIONS IN LINEAR MIXED MODELS WITH HEAVY-TAILED AND SKEW DISTRIBUTIONS

Kapucu, Tuğba
Doctor of Philosophy, Statistics
Supervisor: Assoc. Prof. Dr. Fulya Gökalp Yavuz

December 2023, 127 pages

Modern statistical modeling grapples with the complexities of continuous longitudinal data, where traditional linear mixed models (LMMs) may fall short in capturing the full spectrum of nuances. In this research, we present a novel approach that marries the intricacies of a multivariate skew Laplace distribution for random effects with the robustness of a multivariate Laplace distribution for error terms. This approach allows our model to capture both skewness and heavy tails in continuous longitudinal data. By expressing the skew Laplace distribution as a normal mean-variance mixture distribution and modeling the error term of the proposed framework as a scale mixture of normal distribution, we establish a hierarchical structure for our model. Importantly, the skew Laplace distribution involves fewer parameters compared to the skew t distribution, which is also a robust alternative of normal distribution, simplifying the estimation process. Parameter estimations for our skew Laplace linear mixed model (SL-LMM) are achieved through the Expectation Conditional Maximization (ECM) algorithm, an extended EM-type algorithm. Within the ECM algorithm, a Bayesian approach is employed to infer unobserved latent variables, with a specific case of the Markov Chain Monte Carlo (MCMC) method, known as the Metropolis algorithm, utilized for parameter estimation. To validate our model, we apply it to schizophrenia data and conduct several simulation studies, comprehensively evaluating the model's performance under varying conditions. The results show that the proposed model accurately estimates the underlying parameters and presents a better fit when compared to the competitive models.

Keywords: Linear Mixed Models, Skew Laplace Distribution, ECM Algorithm, Metropolis Algorithm, Longitudinal Data

ÖZ

LİNEER KARMA MODELLERDE KALIN KUYRUKLU VE ÇARPIK DAĞILIMLARA DAYALI PARAMETRE TAHMİNLERİ

Kapucu, Tuğba
Doktora, İstatistik
Tez Yöneticisi: Doç. Dr. Fulya Gökalp Yavuz

Aralık 2023, 127 sayfa

Modern istatistiksel modelleme, geleneksel lineer karma modellerin (LMMs) tüm nüanslarını yakalamakta yetersiz kaldığı durumlarda sürekli boylamsal verilerin karmaşıklığının üstesinden gelmeye çalışmaktadır. Bu araştırmada, rastgele etkiler için çok değişkenli çarpık Laplace dağılımının karmaşıklığını, hata terimleri için çok değişkenli Laplace dağılımının sağlamlığıyla birleştiren yeni bir yaklaşım sunuyoruz. Bu yaklaşım, modelimizin sürekli boylamsal verilerde hem çarpıklığı hem de ağır kuyrukları yakalamasına olanak tanımaktadır. Çarpık Laplace dağılımını normal ortalama-varyans karışım dağılımı olarak ifade ederek ve önerilen çerçevenin hata terimini normal dağılımın ölçek karması olarak modelleyerek, modelimiz için hiyerarşik bir yapı oluşturuyoruz. Daha da önemlisi, çarpık Laplace dağılımı, normal dağılımın sağlam bir alternatifi olan çarpık t dağılımına kıyasla daha az parametre içerir ve tahmin sürecini basitleştirir. Çarpık Laplace lineer karma modelimiz (SL-LMM) için parametre tahminleri, genişletilmiş bir EM tipi algoritma olan Beklenti Koşullu Maksimizasyon (ECM) algoritması aracılığıyla elde edilir. ECM algoritmasında, parametre tahmini için Markov zincirli Monte Carlo (MCMC) yönteminin özel bir durumu olarak bilinen Metropolis algoritmasıyla, gözlemlenmeyen gizli değişkenleri çıkarmak için bir Bayes yaklaşımı kullanılmıştır. Modelimizi doğrulamak için şizofreni verisine uyguluyoruz ve değişen koşullar altında modelimizin performansını kapsamlı bir şekilde değerlendirerek çeşitli simülasyon çalışmaları yürütüyoruz. Sonuçlar, önerilen modelin temel parametreleri doğru bir şekilde tahmin ettiğini ve alternatif modellerle karşılaştırıldığında daha iyi uyum sağladığını göstermektedir.

Anahtar Kelimeler: Lineer Karma Modeller, Çarpık Laplace Dağılımı, ECM Algoritması, Metropolis Algoritması, Boylamsal Data

To my childhood...

ACKNOWLEDGMENTS

First of all, I would like to express my deepest gratitude to my supervisor Assoc. Prof. Dr. Fulya Gökalp Yavuz for her guidance, advice, criticism, encouragements, and insight throughout the research. This thesis owes its successful completion to exceptional mentor's invaluable guidance and support. Without her persistent and expert assistance, I would not have been able to achieve this milestone. Her guidance and teachings have broadened my horizons, not just in academic pursuits, but in life as a whole. I am grateful for the new perspectives and knowledge she has bestowed upon me.

I would like to express my deepest gratitude to my thesis committee members Prof. Dr. Olçay Arslan, Prof. Dr. Özlem İlk Dağ, Prof. Dr. Zeynep Kalaylıoğlu, and Assoc. Prof. Dr. Fatma Zehra Doğru for their valuable guidance, constructive comments, and suggestions.

I would also like to express my special thanks to my dear friend Tuğba Erdem for her everlasting support and companionship in my statistics adventure.

Finally, I would like to express my gratitude to my family for their unconditional support, understanding and love that they always give me.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS	
ECM	Expectation Conditional Maximization
EM	Expectation-Maximization
EP	Exponential Power
LMM	Linear Mixed Model
L-LMM	Laplace Linear Mixed Model
MCMC	Markov Chain Monte Carlo
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
MH	Metropolis Hasting
MSL	Multivariate Skew Laplace
NMVM	Normal Mean-Variance Mixture
SL-LMM	Skew Laplace Linear Mixed Model
SMN	Scale Mixture of Normal
SMSN	Scale Mixture of Skew Normal

CHAPTER 1

INTRODUCTION

1.1 Literature Review

Linear mixed models (LMMs) were implemented in the field of animal sciences for the first time by Eisenhart [1] and Henderson [2]. In addition to LMMs that includes random effects, the literature also presents the availability of random ANOVA (analysis of variance) models. Explanation of particular ANOVA models can be found in Searle's [3] work on linear models. Although the extensive use of these models, Harville [4] indicates that just considering linear models in terms of ordinary regression and ANOVA models induces to skip many potential applications. Specially, the observations are generated by stochastic process as they all may not have been taken at the same time (Harville [4]). Based on the study of Harville [4], Laird and Ware [5] introduce a family of two-stage models for repeated measurements. Stage 1 introduces population parameters, individual effects, and within-person variation, and Stage 2 introduces between-person variation (Laird and Ware [5]). Growth models and repeated-measures models are special cases in this family. The general formula of LMM (Laird and Ware [5]) has the following form:

$$\begin{aligned} &y = X\beta + Zb + \epsilon, \\ &\epsilon \sim N(0, R), \qquad (1.1) \\ &b \sim N(0, D), \end{aligned}$$

where y is a vector of responses, β is vector of corresponding fixed effects coefficients, X and Z are design matrices for fixed and random effects, respectively. b is vector of random effect coefficients, ϵ is a vector of residuals and assumed to be independent of b. Since LMM has one extra random effect parameter in addition to the response variable, coefficients, and error terms of a linear model, often referred to as linear models with random effects.

For traditional LMMs, it is essential to meet the normal distribution assumption for both random effects and error terms to facilitate parameter estimation. However, the normality assumption becomes particularly sensitive during parameter estimation when dealing with outliers, heavy-tailed, and/or skewed data. To address this limitation of the normality assumption, the literature introduces more flexible approaches that employ a wider range of distributions within the framework of LMMs. The t-distribution is one of the most commonly used distribution among these various distributions (Verbeke and Lesaffre [6], Pinheiro et al. [7], Lin [8]). Pinheiro et al. [7] developed a LMM in which random effect and error terms have multivariate t-distribution and they implemented an EM type algorithm to estimate the parameters of interest. Although t-distribution is more suitable than normal distribution for handling heavy tailed data, it necessitates an additional step of estimating the degrees of freedom. Moreover, the t-distributed model is primarily employed for smaller degrees of freedom, as it tends to converge towards a normal distribution as the degrees of freedom increases. Manghi et al. [9] define the multilevel models with elliptical distributions to permit light and heavy tailed error distributions. A robust LMM with multivariate Laplace distribution, which is known as the heavy tailed alternative to the normal distribution, is proposed by Yavuz and Arslan [10] and the model definitions and parameter estimations are implemented under this assumption. It is concluded that the performance of the Laplace distribution is better than t-distribution for LMM in terms of providing smaller prediction errors of the parameter estimations. The advantage of Laplace distribution is to be written as a scale mixture of normal (SMN) distribution and the number of parameters is less than t-distribution that makes the estimation procedures simplier. Expextation Maximization (EM) algorithm for LMM with Laplace distributed random effects and error terms and also with only Laplace distributed random effects is used to generate the parameter of interest. As a result of the simulation and real data example, performance of Laplace distribution over normal distribution for LMM is shown in their studies (Yavuz and Arslan [10]).

For LMM, skew distributions are also used when normality assumptions are violated, especially when skewness is present in the data. In order to fit the data accurately, different skewed distributions are constructed in the literature. Arellano-Valle et al. [11] proposed a skew-normal linear mixed model (SNLMM) based on the multivariate skew-normal (SN) distribution introduced by Azzalini and Valle [12]. They assumed that both random effects and error term have SN distribution. Recently, Lin and Lee [13] develop additional tools for a simplified version of SNLMM (Arellano-Valle et al. [11]) by assuming only the random effects are assumed to follow multivariate SN distributions. Lachos et al. [14] propose skew-normal/independent linear mixed model (SNI-LMM) where random effects follow a skew-normal/independent (SNI) distribution (Lachos et al. [15]) and within-subject errors follow a normal independent distribution. Schumacher et al. [16] propose an extension of the SNI-LMM, which provides flexibility in capturing the effects of skewness and heavy tails simultaneously when continuous repeated measures are serially correlated.

Ho and Lin [17] introduce skew *t* linear mixed model (ST-LMM) which is considered as an extension of LMMs providing flexibility in capturing the effects of skewness and heavy tails simultaneously. An appealing key feature of the ST-LMM is that it can be formulated as two flexible stochastic representations, which is useful for easy setup of EM-type algorithms for estimation and other theoretical purposes. A general skew *t* mixed model that allows different degrees of freedom for random effects and error distribution is developed by Choudhary et al. [18]. As a result of their independence, their degrees of freedom differ, which overcomes the limitation of ST-LMM and extends the range of mixed models. A very recent study conducted by Mohammadi and Kazemi [19] based on the LMM with both random effect and error terms distributed skew-Huber distribution, which is mainly not only helpful in minimizing outlying data points but also helpful for asymmetric situations. However, skew *t* and skew-Huber distributions include additional parameters-degrees of freedom v and tuning parameter c, which controls the impact of the outliers, respectively. The number of parameters of the skew Laplace distribution is less than those distributions, making it our preferred choice.

This study primarily focuses on developing a model capable of handling skewed and heavy-tailed distributed data. For this aim, an alternative multivariate skew Laplace (MSL) distribution, defined by Arslan [20], is used in this study. It is an extension of multivariate symmetric Laplace distribution, which is a special case of the multivariate Kotz-type distribution. It is worthwhile to distinguish this distribution from the multivariate Laplace distribution proposed by Kozuboswki and Podgorski [21], [22]; Kotz et al. [23], [24]. Furthermore, in this study, we employ the Laplace distribution, also known as double exponential distribution, which a specific case of exponential power (EP) distribution. Broad explanations about the multivariate EP distribution and its stochastic and probabilistic characteristics, transformation properties, marginal and conditional distributions can be found in Gómez et al. [25].

Following the model definitions, it is necessary to estimate the parameters to make inferences about the model. Since LMM does not provide closed form solutions for parameters estimations and these estimations include some other unknown parameters, iterative estimation algorithms are needed in the estimation process of LMM. The EM-type or Newton-Raphson algorithms are amongst the iteration methods used in LMM.

EM algorithm is first introduced by Dempster et al. [26] and computation of the maximum likelihood estimates from incomplete data is given in this study via the EM algorithm. The EM algorithm for LMM is developed by Laird and Ware [5] and it is applied for repeated measures by Laird et al. [27]. Specification of covariates for missing observations is prevented by showing that EM algorithm can be used with multivariate normal LMMs with an arbitrary covariance structure and missing data. By utilizing the general formulation, it becomes possible to find the closed-form solutions for maximizing the complete data for a wider range of models. Along with EM algorithm, Newton-Raphson is implemented for estimating the parameters of LMM by Lindstrom and Bates [28]. Derivatives for both maximum likelihood (ML) and restricted maximum likelihood (REML) estimations are formulated. The preference for the EM algorithm can be attributed to two key factors: it primarily employs computationally tractable ML estimations based on complete data, and it demonstrates stable convergence, increasing the likelihood with each iteration (Meng and Rubin [29]). However, EM algorithm is less appealing in case of the related complete-data ML estimation itself is complex. Meng and Rubin [29] extended the EM algorithm to the Expectation Conditional Maximization (ECM) algorithm, which provides easiness of complete-data conditional ML estimation by replacing a complicated M-step of the EM with multiple computationally simpler CM-steps. With several illustrative examples, they showed that the ECM algorithm shares all of the attractive convergence properties of the EM. Expectation-Conditional Maximization Either (ECME) algorithm, presented by Liu and Rubin [30], is another generalized EM algorithm with a faster convergence rate than EM or ECM. In addition to EM algorithm, ECM and ECME algorithms are generally applied in LMM.

1.2 Purpose of the Thesis

LMM is widely applicable and significant method for the analysis of various data types such as clustered data, longitudinal data, and repeated measurements. Since the experiments are conducted on the same subjects repeatedly over time in those data types, the independency assumption of linear model is violated. This violation leads to the definition of LMMs, which encompass both fixed effects and subject-specific random effects. In fact, LMMs can be defined as the extended version of classical regression models by adding extra random effect terms. It is a model in which errors are normally distributed but the assumptions of independent observations and constant variance of the errors do not require to be satisfied. LMM is a very efficient method to explain both within and between observations variability not only for balanced data but also for unbalanced data.

As a result of the literature review, LMM is a highly preferred method to explain within and between subjects variability in longitudinal and clustered data types. However in the occurrence of heavy tailness and skewness in the data, one needs to seek a more general parametric family of distributions with additional parameters in regulating possible skewness and thickness of the tails. In this study, we extend LMM with adapting multivariate skew Laplace (MSL) distribution into the random effect as an alternative to the multivariate skew t distribution. Incorporating a skewed distribution into the model necessitates the introduction of more parameters compared to the classical approach. Despite the additional complexity, this approach offers the advantages of improved data modeling, allowing for the inclusion of outlier observations as well as accounting for skewness and heavy-tailedness in the data. Therefore, instead of the skew t distribution, a skew distribution with fewer parameters is preferred which makes the estimation procedures more manageable. Also, expressing the skew Laplace distribution as a normal mean-variance mixture (NMVM) distribution enables us to define the proposed model in a hierarchical form. The parameter estimations of the skew Laplace LMM (SL-LMM or S-LLMM) are not in a closed form, and therefore an iterative algorithm is required for the final parameter estimations. So, they are implemented with the EM-type algorithm for LMM with skew Laplace distributed random effects and multivariate Laplace distributed error terms.

1.3 Original Contribution to Literature

Multivariate Skew Laplace (MSL) distribution for LMM is proposed as an alternative to the skew normal and skew t distributions for the first time in this study. While the other mentioned distributions are written as a scale mixture of skew-normal (SMSN) distribution, the Skew Laplace distribution is expressed in a normal mean-variance mixture (NMVM) form, which provides a distinct and valuable option for representing skewness parameter as fixed but unknown. The hierarchical representation of Skew Laplace Linear Mixed Model (SL-LMM) has been written for the first time using the definition of stochastic representation of NMVM based on MSL distribution. Furthermore, the scaling variable and related mixing function for the error terms of the proposed model has been defined for the first time, enabling the model to be expressed in a hierarchical form. Since the hierarchical representation of the SL-LMM provides a significant advantage in applying the EM algorithm, the EM algorithm approach has been used to obtain maximum likelihood estimators (MLEs) of the parameters. To conclude the parameter estimation procedure, a special case of Markov Chain Monte Carlo (MCMC) method, namely the Metropolis algorithm, is applied to find moments of the latent variable, which is unobserved but imputed from the data for each subject. The Metropolis algorithm has enabled us to proceed with the EM algorithm, even when obtaining an analytical solution for the moments of the latent variables is unfeasible. Utilizing both EM and Bayesian methods in conjunction with LMMs marks a noteworthy achievement. The study involves the generation of new R codes in SL-LMM for simulations and data analysis for the proposed and other models used for comparisons.

CHAPTER 2

LINEAR MIXED MODELS

2.1 Definition and Model Formulation of Linear Mixed Model

LMM can be defined as the extended version of classical regression with additional random effect terms. Random effects are utilized to account for heterogeneity in the data, which can be caused by subject-or cluster-related effects or by spatial correlations (Wu [31]). The random effects in the models depict the influence of each individual or cluster on repeated observations that are not considered by observed covariates (Wu [31]). It is possible to make individual-specific inferences from these random effects, as they can be viewed as cluster effects or individual effects. Further the correlation between repeated measurements from the same individual or cluster can be added to the model. Laird and Ware [5] define multivariate normal LMM in two-stage. While the first stage specifies population parameters, individual effects, and within-subject variation, the second stage specifies between-subject variation (Laird and Ware [5]).

Let y_{ij} be the response value for individual i at time t_{ij}, $i = 1,2, \dots, n$, $j = 1,2, \dots, n_i$, and let y_i be $(n_i \times 1)$ dimensional dependent variable vector. In the first stage the model is defined as

$$y_i = X_i\beta + Z_i b_i + e_i, \tag{2.1}$$

where β is a $(p \times 1)$ vector of corresponding fixed effects coefficients, X_i and Z_i are $(n_i \times p)$ and $(n_i \times q)$ known design matrixes respectively, b_i is a $(q \times 1)$ vector of random effects. e_i is a vector of residuals of length n_i and defined as $e_i \sim N(0, R_i)$. R_i is $(n_i \times n_i)$ dimensional positive definite variance-covariance matrix for error terms. The random effect term (b_i) is assumed to be normally distributed with zero mean and D variance-covariance matrix $\left(b_i \sim N(0, D)\right)$ and assumed to be independent of random errors (e_i). So, the marginal distribution of the response variable y_i is given by

$$y_i \sim N(X_i\beta, Z_iDZ_i{}' + R_i). \tag{2.2}$$

As a result, a general LMM can be written as follows (Laird and Ware [5])

$$\begin{aligned} y_i &= X_i\beta + Z_ib_i + e_i, \\ e_i &\sim N(0, R_i), \\ b_i &\sim N(0, D), \end{aligned} \tag{2.3}$$

where b_i and e_i are independent. Conditional distribution of y_i given random effect term, b_i can be easily defined as below

$$y_i \mid b_i \sim N(X_i\beta + Z_ib_i,\ R_i). \tag{2.4}$$

And the joint probability density function of y_i and b_i is as follows:

$$\begin{bmatrix} y_i \\ b_i \end{bmatrix} \sim N_{n_i+q}\left(\begin{bmatrix} X_i\beta \\ 0 \end{bmatrix}, \begin{bmatrix} Z_iDZ_i{}' + R_i & Z_iD \\ DZ_i{}' & D \end{bmatrix}\right), \quad i = 1, \dots, n. \tag{2.5}$$

The generalized representation of LMM is as follows with the response vector y, X and Z are fixed and random design matrices, respectively; β is the fixed-effect vector, b is the random effect vector and the e is the error vector:

$$y = X\beta + Zb + e. \tag{2.6}$$

In the Bayesian framework, a LMM can be represented as a hierarchical model comprising three hierarchical levels. At the first level, a linear model is constructed based on the fixed and random effects. In the second level, the distributions of the fixed and random effects are determined based on the variance component parameters. Finally, at the third level, a prior distribution is assumed for the fixed effects and variance components.

$$y \mid \beta, b, \theta \sim N(X\beta + Zb, R), \tag{2.7}$$

$$b|\theta \sim N(0, D),$$

$$(\beta, \theta) \sim \pi(\beta, \theta),$$

where $R = R(\theta)$ and $D = D(\theta)$. π denotes a prior distribution. In many applications, $\pi(\beta, \theta) = \pi_1(\beta)\, \pi_2(D)$ for known β and D, where π_1 and π_2 are specified prior distributions (Jiang [32]). This type of hierarchical representation, along with similar ones, will be commonly utilized in the upcoming chapters.

2.1.1 Matrix Specification of Linear Mixed Models

The relevant matrix notations for the model given in Equation (2.3) with the assumptions $b_i \sim N(0, D)$ and $e_i \sim N(0, R_i)$ can be defined as following, let y_i represents a vector of continuous response for the i-th subject,

$$y_i = \begin{bmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{n_i i} \end{bmatrix}.$$

Note that while the number of repeats, n_i is the same for balanced data, it varies form one subject to another for unbalanced data. X_i is $(n_i \times p)$ design matrix, which denotes the known values of the p covariates $X^1, \dots, X^p$ for each n_i observations on the i-th subject,

$$X_i = \begin{bmatrix} {X_{1i}}^{(1)} & {X_{1i}}^{(2)} & \cdots & {X_{1i}}^{(p)} \\ {X_{2i}}^{(1)} & {X_{2i}}^{(2)} & \cdots & {X_{2i}}^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ {X_{n_i i}}^{(1)} & {X_{n_i i}}^{(2)} & \cdots & {X_{n_i i}}^{(p)} \end{bmatrix}.$$

If a model has an intercept term, then the value in the first column of X_i for all observations would be 1. To simplify the presentation, it is assumed that X_i consists of independent columns or rows, and none of them can be expressed as a linear combination of the others (West et al. [33]). The β is a vector of p unknown fixed-effect parameters,

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}.$$

Z_i is $(n_i \times q)$ design matrix, which denotes the known values of the q covariates $Z^1, \dots, Z^q$ for the i-th subject,

$$Z_i = \begin{bmatrix} Z_{1i}{}^{(1)} & Z_{1i}{}^{(2)} & \cdots & Z_{1i}{}^{(q)} \\ Z_{2i}{}^{(1)} & Z_{2i}{}^{(2)} & \cdots & Z_{2i}{}^{(q)} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n_i i}{}^{(1)} & Z_{n_i i}{}^{(2)} & \cdots & Z_{n_i i}{}^{(q)} \end{bmatrix}.$$

The b_i vector for the i-th subject denotes a vector of q random effects associated with the q covariates in the Z_i design matrix,

$$b_i = \begin{bmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{qi} \end{bmatrix}.$$

The q random effects in the b_i vector are assumed to be distributed multivariate normal with mean vector 0 and D variance-covariance matrix $b_i \sim N(0, D)$. Variance-covariance matrix D is positive definite and symmetric. The elements of D are as follows:

$$D = Var(b_i) = \begin{bmatrix} Var(b_{1i}) & cov(b_{1i}, b_{2i}) & \cdots & cov(b_{1i}, b_{qi}) \\ cov(b_{1i}, b_{2i}) & Var(b_{2i}) & \cdots & cov(b_{2i}, b_{qi}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(b_{1i}, b_{qi}) & cov(b_{2i}, b_{qi}) & \cdots & Var(b_{qi}) \end{bmatrix}.$$

The elements of D are defined as functions of a small set of covariance parameters, represented by θ_D. It is worth noting that the elements are subject to the constraints (structured) imposed by the vector θ_D. The unstructured (positive definite and symmetric) variance components (diagonal) are the most commonly used structures for D matrix. For a more detailed explanation about these structures, one can examine West et al. [33].

Finally, e_i is a vector of n_i residuals, in which each element represents the residual corresponding to an observed response at a particular time t for the i-th subject. The variations in the length of the e_i vectors occur because some subjects may have more data observations due to factors such as missing data or dropouts.

$$e_i = \begin{bmatrix} e_{1i} \\ e_{2i} \\ \vdots \\ e_{n_i i} \end{bmatrix}.$$

A standard linear model assumes that residuals are independent, while those in an LMM can be correlated due to taking repeated observations on the same subject (West et al. [33]). However, it is assumed that residuals with different subjects are independent of each other. Additionally, the vector of residuals and random effects are independent of each other. The n_i residuals in the e_i vector for a given subject i are assumed to follow a multivariate normal distribution with mean vector 0 and variance-covariance matrix R_i, $e_i \sim N(0, R_i)$. The general form of the R_i is given below:

$$R_i = Var(e_i) = \begin{bmatrix} Var(e_{1i}) & cov(e_{1i}, e_{2i}) & \cdots & cov(e_{1i}, e_{n_i i}) \\ cov(e_{1i}, e_{2i}) & Var(e_{2i}) & \cdots & cov(e_{2i}, e_{n_i i}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(e_{1i}, e_{n_i i}) & cov(e_{2i}, e_{n_i i}) & \cdots & Var(e_{n_i i}) \end{bmatrix}.$$

The elements of R_i are defined as functions of a small set of covariance parameters, represented by θ_R. There are many different covariance structures for the R_i. The diagonal structure, compound symmetry, first-order autoregressive structure denoted by AR(1) are among the most commonly used structures for the matrix R_i. More detailed information regarding the properties of these structures is available in West et al. [33], one can refer to this source for a better understanding.

2.2 Maximum Likelihood Estimate of Fixed-Effect Parameter in LMM

The ML estimation of fixed-effect parameter is discussed in this section. Before moving the estimation steps, let us give some implications for the model (2.6). By applying the method of generalized least square (GLS), we get

$$\hat{\beta} = (X'V^{-1}X)^{-}X'V^{-1}y \tag{2.8}$$

where $(X'V^{-1}X)^{-}$ is any generalized inverse of $X'V^{-1}X$. Note that the estimate $\hat{\beta}$ given in (2.8) is the best linear unbiased estimator (BLUE) of β (West et al. [33]) with the variance,

$$Var(\hat{\beta}) = (X'V^{-1}X)^{-}.$$

This also gives the BLUE of $X\beta$ (Puntanen and Styan [34], Searle [35])

$$BLUE(X\beta) = X(X'V^{-1}X)^{-}X'V^{-1}y. \tag{2.9}$$

The variance V of the model (2.6) is positive definite symmetric matrix and defined as

$$V = Var(y) = Var(X\beta + Zb + e) = Var(Zb + e) = ZVar(b)Z' + Var(e)$$
$$V = ZDZ' + R. \tag{2.10}$$

The variance formula of V given in (2.10) is generalized for r random variables by Hartley and Rao [36], Searle et al. [37], McCulloch and Searle [38], Searle [35] as follows:

$D = var(b) = \{_d\, {\sigma_i}^2 \mathbf{I}_{q_i}\}_{i=1}^{r}$ and b_i has a $(q_i \times 1)$ dimension. Then,

$$V = ZDZ' + R = \sum_{i=1}^{r} Z_i Z_i' \sigma_i^2 + \sigma^2 \mathbf{I}_N \tag{2.11}$$

for y with dimension $(N \times 1)$.

Hartley and Rao [36] suggest redefining D to include $\sigma^2 \mathbf{I}_N$ by giving

$\sigma_0{}^2 \equiv \sigma^2, \quad Z_0 \equiv \mathbf{I}_N \quad$ and $\quad q_0 \equiv N.$

Then,

$$D_* = \begin{bmatrix} \sigma_0{}^2 \mathbf{I}_{q_0} & 0 \\ 0 & D \end{bmatrix} = \{_d\ \sigma_i{}^2 \mathbf{I}_{q_i}\}_{i=0}^{r}.$$

So, (2.11) can be rewritten as below (McCulloch and Searle [38])

$$V = \sum_{i=0}^{r} Z_i Z_i' \sigma_i{}^2. \tag{2.12}$$

Eq. (2.12) can be written as

$V = Z_* D_* Z_*'$ for $Z_* = [Z_0 \quad Z_1 \quad Z_2 \quad \cdots \quad Z_r] = [Z_0 \quad Z]$.

The corresponding probability density function for the model in (2.2) is

$$f(y|\beta, V) = \frac{1}{(2\pi)^{\frac{-N}{2}} |V|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}(y - X\beta)' V^{-1} (y - X\beta)\right\} \tag{2.13}$$

where $V = ZDZ' + R$.

So, the log-likelihood function is defined as:

$$l(\beta, V) = -\frac{1}{2} N ln(2\pi) - \frac{1}{2} ln|V| - \frac{1}{2}(y - X\beta)' V^{-1} (y - X\beta). \tag{2.14}$$

While it is possible to simultaneously estimate β and V through optimization, many algorithms simplify by profiling β from $l(\beta, V)$ (West et al. [33]). MLE of the fixed-effects parameter will be explained under two headings: when the V is known and it is unknown.

2.2.1 ML Estimate of Fixed-Effect Parameter when V is Known

Since V is known, β is the only parameter to be estimated. The first order derivative of Eq. (2.14) with respect to β is obtained as follows

$$\frac{\partial l}{\partial \beta} = X'V^{-1}(y - X\beta). \tag{2.15}$$

When we equate Eq. (2.15) into zero and substitute β^0 instead of β, we get

$$\begin{aligned} &X'V^{-1}y - X'V^{-1}X\beta^0 = 0 \\ &\beta^0 = (X'V^{-1}X)^{-}X'V^{-1}y. \end{aligned} \tag{2.16}$$

Since β^0 changes with respect to $(X'V^{-1}X)^{-}$, we focus on $X\beta^0$ which is invariant (McCulloch and Searle [38]). Therefore, the ML estimation of $X\beta$ is

$$ML(X\beta) = X\beta^0 = X(X'V^{-1}X)^{-}X'V^{-1}y. \tag{2.17}$$

2.2.2 ML Estimate of Fixed-Effect Parameter when V is Unknown

Let start with the estimation of the V by maximizing Eq. (2.14) with respect to parameters given in V. Let φ represent each parameter in V (McCulloch and Searle [38]) and the first-order derivatives of the log-likelihood function with respect to parameters are obtained as follows (Appendix A-2):

$$\frac{\partial l}{\partial \varphi_k} = \frac{1}{2}(y - X\beta)'V^{-1}\frac{\partial V}{\partial \varphi_k}V^{-1}(y - X\beta) - \frac{1}{2}tr\left(V^{-1}\frac{\partial V}{\partial \varphi_k}\right) \tag{2.18}$$

$$V_{\partial_k} = \frac{\partial V}{\partial \varphi_k}, \quad \hat{V}_{\partial_k} = \frac{\partial V}{\partial \varphi_k}\Big| V = \hat{V}$$

$$tr\left(\hat{V}^{-1}\hat{V}_{\partial_k}\right) = \left(y_i - X_i\hat{\beta}\right)'V^{-1}\hat{V}_{\partial_k}V^{-1}\left(y_i - X_i\hat{\beta}\right) \tag{2.19}$$

$$tr\left(\hat{V}^{-1}\hat{V}_{\partial_k}\right) = y'\hat{P}\,\hat{V}_{\partial_k}\hat{P}y, \tag{2.20}$$

$$\hat{P} = \hat{V}^{-1} - \hat{V}^{-1}X\left(X'\hat{V}^{-1}X\right)^{-}X'\hat{V}^{-1}. \tag{2.21}$$

If the variance formula V is defined as in the Eq. (2.12) and then by taking the first-order derivative with respect to ${\sigma_i}^2$, we get

$$\frac{\partial V}{\partial {\sigma_i}^2} = \frac{\partial(\sum_i Z_i {Z_i}' {\sigma_i}^2)}{\partial {\sigma_i}^2} = Z_i {Z_i}'. \tag{2.22}$$

When the Eq. (2.22) is substituted into Eq. (2.18) and set equal to zero,

$$\frac{\partial l}{\partial {\sigma_i}^2} = \frac{1}{2}(y - X\beta)'V^{-1}Z_i{Z_i}'V^{-1}(y - X\beta) - \frac{1}{2}tr(V^{-1}Z_i{Z_i}'), \tag{2.23}$$

$$tr\left(\hat{V}^{-1}Z_i{Z_i}'\right) = \left(y - X\hat{\beta}\right)'\hat{V}^{-1}Z_i{Z_i}'\hat{V}^{-1}\left(y - X\hat{\beta}\right) \tag{2.24}$$

is obtained. ML estimation of $X\beta$ is required to be found for the Eq. (2.24), so

$$ML(X\beta) = X\hat{\beta} = X\left(X'\hat{V}^{-1}X\right)^{-}X'\hat{V}^{-1}y. \tag{2.25}$$

It is worth noting that whatever $\hat{V}$ is, the ML estimation of $X\beta$ equals to Eq. (2.25). Further it is not BLUE of $X\beta$ now, since the property of BLUE is valid for the case where the variance components are known. Let's go back to Eq. (2.24) and write the equality below instead of $\left(y_i - X_i\hat{\beta}\right)$ term.

$$\left(y - X\hat{\beta}\right) = y - X\left(X'\hat{V}^{-1}X\right)^{-}X'\hat{V}^{-1}y$$

$$\hat{V}^{-1}\left(y - X\hat{\beta}\right) = \hat{V}^{-1}\left[y - X\left(X'\,\hat{V}^{-1}X\right)^{-}X'\hat{V}^{-1}y\,\right] \tag{2.26}$$

$$= \left[\hat{V}^{-1} - \hat{V}^{-1}X\left(X'\,\hat{V}^{-1}X\right)^{-}X'\hat{V}^{-1}\right]y = \hat{P}y$$

Then the Eq. (2.24) is updated as

$$tr\left(\hat{V}^{-1}Z_i{Z_i}'\right) = y'\hat{P}Z_i{Z_i}'\hat{P}y. \tag{2.27}$$

As can be seen in the above equations, $\hat{V}^{-1}$ is used in the analysis. And, the above equations, which need to be calculated for each variance component are non-linear in nature. Therefore, instead of analytic solutions, iterative methods will be utilized to solve these equations (Searle et al. [37]).

2.3 Multivariate Normal Distribution

A random vector $X = \left(X_1, X_2, \dots, X_q\right)'$ is said to have a multivariate normal distribution if its density function is

$$f(x) = \frac{1}{(2\pi)^{q/2}\,|\Sigma|^{1/2}} exp\left(-\frac{1}{2}(x-\mu)'\,\Sigma^{-1}\,(x-\mu)\right), \tag{2.28}$$

$x \in R^q, \quad i = 1,2, \dots, q.$

where $\mu \in R^q$ is the location parameter, Σ is $(q \times q)$ positive definite variance-covariance matrix.

Normal distribution is a fundamental statistical concept, and many statistical methods rely on it as a critical assumption. Therefore, it is also an essential aspect of classical LMMs. The first assumption in classical LMM (Laird and Ware [5]) is that the random effect and random error terms follow a normal distribution. Since several features related to the multivariate normal distribution will be used in inferences, some known parts of this distribution have been added to Appendix B.

2.4 Expectation-Maximization (EM) Algorithm

Finding the ML and REML estimations in LMMs by equating the derivatives of the log-likelihoods to zero leads to nonlinear equations for the estimators. In fact, solving these complex nonlinear equations is not enough, because the log-likelihood must be maximized within the parameter space (Searle et al. [37]). Therefore, iterative methods such as Expectation Maximization (EM) algorithm, which was introduced by Demster et al. [26], can be used for calculating the parameter estimators. As the name suggests, the algorithm comprises two primary steps: the computation of conditional expected values for the respective function and the maximization of simplified likelihoods. Iterations are repeated until the convergency occurs between the two steps. A notable characteristic of the EM algorithm is that since it conducts ML estimation on complete data, the parameters always remain in the parameter space. Adaptation, convergence and handling missing data are other features of the EM algorithm (Wu [39]). In addition, no distributional assumption or being used for large data sets can be counted among the reasons why it is widely used (Wu [31]). The drawback of the EM algorithm is that it only calculates the estimates but does not give variance estimates as other iterative methods, such as Newton-Raphson and methods of scoring. Additional computations must be performed to get variance estimates, as shown in Louis's study [40].

Let us define $L(\theta|Y)$ as the log-likelihood function of the unknown parameter θ:

$$L(\theta|Y) = p(Y|\theta) = \int p(Y, Z|\theta)\ dZ$$

where $Y = (y_1, \dots, y_n)$ is the observed data, $Z = (Z_1, \dots, Z_n)$ is unobserved/missing data and $W = (Y, Z)$ is the complete data. The joint distribution function $f(Y, Z)$ of these data sets belongs the exponential family distribution and θ is defined as parameter, $\theta \in \Theta$ $(\Theta, parameter\ space)$. EM algorithm is applied by maximizing the likelihoods of the observed data with respect to θ. Define $Q\left(\theta \middle| \theta^{(t)}\right)$ as the expected value of the complete data log-likelihood. An initial value is chosen for the parameter and then expected value of the complete data log-likelihood is calculated as following:

$$Q\left(\theta \middle| \theta^{(t)}\right) = E_{\theta^{(t)}}[L(\theta|W)|Y = y], \tag{2.29}$$

for $t = 0,1,2, \dots$.

EM alternates between an Expectation step and a Maximization step until convergency, as illustrated by the following expression:

Expectation Step (E-Step): Compute the expectation of $L(\theta|W)$.

$$Q(\theta \mid \theta^{(t)}) = E_{\theta^{(t)}}[L(\theta|W)|Y = y]. \tag{2.30}$$

Maximization Step (M-Step): Estimate unknown parameter θ by using the outputs from E-step.

$$\theta^{(t+1)} \epsilon \; argmax_{\theta} Q\left(\theta \middle| \theta^{(t)}\right). \tag{2.31}$$

This process alternates between E and M steps and the estimation of the next parameter from the previous estimation value is performed. Until the condition

$$L\left(\theta^{(t+1)} \middle| Y\right) - L\left(\theta^{(t)} \middle| Y\right) > \varepsilon \tag{2.32}$$

is met, the iteration process continuous for a defined value of ε, $(\varepsilon > 0)$. When the convergence occurs, the last found value of $\theta^{(t+1)}$ is taken as $\hat{\theta}$ predictive value. It is worth noted that the value of likelihood function shows increasing series as the iterations continue, that is

$$L\left(\theta^{(t+1)} \middle| Y\right) \geq L\left(\theta^{(t)} \middle| Y\right). \tag{2.33}$$

And as a result of this, EM algorithm converges to a local or global maximum of $L(\theta)$ (Demster et. al [26]). Pinheiro et al. [7] recommend running the EM algorithm many times by assigning different initial values.

Meng and Rubin [29] extended EM algorithm to Expectation Conditional Maximization (ECM) algorithm, which provides easiness of complete-data conditional maximum likelihood estimation by replacing a complicated M-step of EM with multiple computationally simpler CM-steps. With several illustrative examples, they show that the ECM algorithm shares all of the attractive convergence properties of EM. Considering the inherent simplicity and efficiency of the algorithm, ECM will be employed in the scope of this study.

2.5 EM Algorithm for Linear Mixed Models

When applying the EM algorithm to LMM, the random effect $b_i = (b_1, b_2, \dots, b_r)$ is taken as missing data (Z is above) and $y_i = (y_1, y_2, \dots, y_N)$ is as the observed data. Since the random effect b_i is unknown, it is not possible to compute the expected value of these random effects in the E-step of EM. Fortunately, the EM algorithm provides a way to compute alternative values that can be used instead. Let's remember the model (2.6) with the response vector y; while X and Z are fixed and random design matrices, respectively. β is the fixed-effect vector, b is the random effect vector and the e is the error vector. By taking reference to Searle et al. [37], the random effects in the model (2.6) can be written as follows:

$$Zb = [Z_1 \quad \cdots \quad Z_r] \begin{bmatrix} b_1 \\ \vdots \\ b_r \end{bmatrix} = \sum_{i=1}^{r} Z_i b_i,$$

$$Z_i = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1q_1} \\ Z_{21} & Z_{22} & \cdots & Z_{2q_2} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{N1} & Z_{N2} & \cdots & Z_{Nq_i} \end{bmatrix}_{N\text{x}q_i} \tag{2.34}$$

where $b_i = \left(b_{i1}, b_{i2}, \dots, b_{iq_i}\right)'$ $i = 1,2, \dots, r$ and

$$E(b_i) = 0 \qquad \forall_i,$$

$$Var(b_i) = \sigma_i^2 I_{q_i} \qquad \forall_i,$$

$$cov\left(b_i, b_j'\right) = 0 \qquad i \neq j, \tag{2.35}$$

$$Var(b) = \{_d\, \sigma_i^2 I_{q_i}\}_{i=1}^{r} = \begin{bmatrix} \sigma_1^2 I_{q_1} & 0 & \cdots & 0 \\ 0 & \sigma_2^2 I_{q_2} & \cdots & 0 \\ \cdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r^2 I_{q_r} \end{bmatrix}.$$

Please refer to Appendix A-1 for a comprehensive overview of the general notation used for matrices.

If the random effects b_i are known, the estimation of the variance covariance of the random effects in LMM can be carried out using the least square estimation: $Var(b_i) = \sigma_i^2 = \frac{b_i\, b_i'}{q_i}$ with the assumption that $b_i \sim N\left(0, \sigma_i^2 I_{q_i}\right)$ (Searle et al. [37]).

However, since the random effects are unknown, instead of the unknown values of b_i terms, the expectations of them are used in the EM algorithm. In order to proceed, it is necessary to obtain the joint distribution of variables y and b. Let's first write random effect terms as in (2.34) in the model (2.6),

$$y = X\beta + \sum_{i=1}^{r} Z_i b_i + e. \tag{2.36}$$

Rewrite Eq. (2.36) by taking:

$$b_0 = e, \qquad {\sigma_0}^2 = {\sigma_e}^2, \qquad q_0 = N, \qquad Z_0 = I_N,$$

$$y = X\beta + \sum_{i=0}^{r} Z_i b_i. \tag{2.37}$$

$$V = Var(y) = ZDZ' + R = \sum_{i=1}^{r} Z_i\, {Z_i}'{\sigma_i}^2 + {\sigma_e}^2 I_N = \sum_{i=0}^{r} Z_i\, {Z_i}'{\sigma_i}^2. \tag{2.38}$$

$$cov(y, {b_j}') = cov\left(X\beta + \sum_{i=0}^{r} Z_i\, b_i, {b_j}'\right) = Z_j cov(b_j, {b_j}') = {\sigma_j}^2 Z_j. \tag{2.39}$$

The joint distribution function of y and b is a multivariate normal distribution with μ and Σ as

$$f_{y,b_1,\dots,b_r}(y, b_1, \dots, b_r) = (2\pi)^{-\frac{1}{2}\sum_{i=0}^{r} q_i}\, |\Sigma|^{-\frac{1}{2}}\, exp\left(-\frac{1}{2}Q\right),$$

$$Q = [(y - X\beta)' \quad {b_1}' \quad \dots \quad {b_r}'\,]\, \Sigma^{-1} \begin{bmatrix} y - X\beta \\ b_1 \\ \vdots \\ b_r \end{bmatrix} \tag{2.40}$$

where,

$$\mu = \begin{bmatrix} X\beta \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{2.41}$$

$$\Sigma = \begin{bmatrix} V & \{_r\, \sigma_i^2 Z_i\}_{i=1}^r \\ \{_c\, \sigma_i^2 Z_i'\}_{i=1}^r & \{_d\, \sigma_i^2 I_{q_i}\}_{i=1}^r \end{bmatrix}, \tag{2.42}$$

To simplify (2.40) in terms of variance components, determinant of Σ is written by using the formula given below (Appendix M5. in Searle et al. [37]).

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D||A - BD^{-1}C| \tag{2.43}$$

By substituting the related values in the formula (2.43), we get

$$\begin{vmatrix} V & \{_r\, \sigma_i^2 Z_i\}_{i=1}^r \\ \{_c\, \sigma_i^2 Z_i'\}_{i=1}^r & \{_d\, \sigma_i^2 I_{q_i}\}_{i=1}^r \end{vmatrix}. \tag{2.44}$$

Thus, $|D|$ and D^{-1} are obtained as below given in Yavuz [41]:

$$|D| = \begin{vmatrix} (\sigma_1^2 I_{q_1})_{q_1 \text{x} q_1} & 0 & \cdots & 0 \\ 0 & (\sigma_2^2 I_{q_2})_{q_2 \text{x} q_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\sigma_r^2 I_{q_r})_{q_r \text{x} q_r} \end{vmatrix} = \prod_{i=1}^{r} (\sigma_i^2)^{q_i}, \tag{2.45}$$

$$D^{-1} = \begin{bmatrix} (\sigma_1^{-2} I_{q_1})_{q_1 \text{x} q_1} & 0 & \cdots & 0 \\ 0 & (\sigma_2^{-2} I_{q_2})_{q_2 \text{x} q_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\sigma_r^{-2} I_{q_r})_{q_r \text{x} q_r} \end{bmatrix}. \tag{2.46}$$

We can write the $BD^{-1}C$ term in the Eq. (2.43) as

$$[\sigma_1{}^2 Z_1 \quad \dots \quad \sigma_r{}^2 Z_r]\begin{bmatrix} (\sigma_1{}^{-2} I_{q_1})_{q_1 \mathrm{x} q_1} & 0 & \cdots & 0 \\ 0 & (\sigma_2{}^{-2} I_{q_2})_{q_2 \mathrm{x} q_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\sigma_r{}^{-2} I_{q_r})_{q_r \mathrm{x} q_r} \end{bmatrix}\begin{bmatrix} \sigma_1{}^2 Z_1{}' \\ \vdots \\ \sigma_r{}^2 Z_r{}' \end{bmatrix}$$

$$BD^{-1}C = [\sigma_1{}^2 Z_1 \dots \sigma_r{}^2 Z_r]\begin{bmatrix} Z_1{}' \\ \vdots \\ Z_r{}' \end{bmatrix} = \sum_{i=1}^{r} \sigma_i{}^2 Z_i\, Z_i{}'. \tag{2.47}$$

Therefore, the term $|A - BD^{-1}C|$ in (2.43) is obtained as below

$$|A - BD^{-1}C| = \left| V - \sum_{i=1}^{r} \sigma_i{}^2 Z_i\, Z_i{}' \right|. \tag{2.48}$$

V is defined like $\sum_{i=1}^{r} Z_i\, Z_i{}' \sigma_i{}^2 + \sigma_e{}^2 I_N$ as in (2.38) in the above equation as

$$V - \sum_{i=1}^{r} \sigma_i{}^2 Z_i\, Z_i{}' = \sum_{i=1}^{r} Z_i\, Z_i{}' \sigma_i{}^2 + \sigma_e{}^2 I_N - \sum_{i=1}^{r} \sigma_i{}^2 Z_i\, Z_i{}' = \sigma_e{}^2 I_N. \tag{2.49}$$

Then, it is obtained

$$|A - BD^{-1}C| = |\sigma_e{}^2 I_N| = (\sigma_e{}^2)^N. \tag{2.50}$$

We ended up calculation of $|\Sigma|$ by substituting the equations (2.45) and (2.50) in the formula (2.43) as follows

$$|\Sigma| = \prod_{i=1}^{r} (\sigma_i{}^2)^{q_i} (\sigma_e{}^2)^N = \prod_{i=0}^{r} (\sigma_i{}^2)^{q_i}. \tag{2.51}$$

We also need to find the inverse of Σ given in (2.42). Let's partition Σ as $[\Sigma_1\ \Sigma_2]$. To do this, property of generalized inverse of the partitioned matrix is used (Appendix A-3).

$$\Sigma' \Sigma = \begin{bmatrix} \Sigma_1{}' \\ \Sigma_2{}' \end{bmatrix} [\Sigma_1 \quad \Sigma_2] = \begin{bmatrix} \Sigma_1{}' \Sigma_1 & \Sigma_1{}' \Sigma_2 \\ \Sigma_2{}' \Sigma_1 & \Sigma_2{}' \Sigma_2 \end{bmatrix} \tag{2.52}$$

where,

$\Sigma_1{}'\Sigma_1 = (V)_{N\text{x}N}$,

$$\Sigma_1 = \begin{bmatrix} \sigma_1 Z_1{}' \\ \vdots \\ \sigma_r Z_r{}' \end{bmatrix},$$

$$\Sigma_2{}'\Sigma_2 = \begin{bmatrix} \left(\sigma_1{}^2 I_{q_1}\right)_{q_1\text{x}q_1} & 0 & \cdots & 0 \\ 0 & \left(\sigma_2{}^2 I_{q_2}\right)_{q_2\text{x}q_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\sigma_r{}^2 I_{q_r}\right)_{q_r\text{x}q_r} \end{bmatrix},$$

$$\Sigma_2 = \begin{bmatrix} \sigma_1 I_{q_1} & 0 & \cdots & 0 \\ 0 & \sigma_2 I_{q_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r I_{q_r} \end{bmatrix},$$

$$\Sigma_1{}'\Sigma_2 = \left[\ (\sigma_1{}^2 Z_1)_{N\text{x}q_1} \quad (\sigma_2{}^2 Z_2)_{N\text{x}q_2} \quad \cdots \quad (\sigma_r{}^2 Z_r)_{N\text{x}q_r}\right],$$

$$\Sigma_2{}'\Sigma_1 = \begin{bmatrix} (\sigma_1{}^2 Z_1{}')_{q_1\text{x}N} \\ (\sigma_2{}^2 Z_2{}')_{q_2\text{x}N} \\ \vdots \\ (\sigma_r{}^2 Z_r{}')_{q_r\text{x}N} \end{bmatrix}.$$

We can start taking the inverse of Σ by using the formula (2.53) together with the partitioned matrices and vectors given below:

$$\Sigma^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & (\Sigma_2{}'\Sigma_2)^{-} \end{bmatrix} + \begin{bmatrix} I \\ -(\Sigma_2{}'\Sigma_2)^{-}\Sigma_2{}'\Sigma_1 \end{bmatrix} (\Sigma_1{}'M_2\Sigma_1)^{-}\left[I \quad -\Sigma_1{}'\Sigma_2(\Sigma_2{}'\Sigma_2)^{-}\right] \tag{2.53}$$

where $M_2 = I - \Sigma_2(\Sigma_2{}'\Sigma_2)^{-}\Sigma_2{}'$,

$$(\Sigma_2{}'\Sigma_2)^{-} = \begin{bmatrix} \left(\sigma_1{}^{-2} I_{q_1}\right)_{q_1 \mathrm{x} q_1} & 0 & \cdots & 0 \\ 0 & \left(\sigma_2{}^{-2} I_{q_2}\right)_{q_2 \mathrm{x} q_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\sigma_r{}^{-2} I_{q_r}\right)_{q_r \mathrm{x} q_r} \end{bmatrix}, \tag{2.54}$$

$$= \{_d\, \sigma_i{}^{-2} I_{q_i}\}$$

$$(\Sigma_2{}'\Sigma_2)^{-}\Sigma_2{}'\Sigma_1 = \begin{bmatrix} Z_1{}' \\ \vdots \\ Z_r{}' \end{bmatrix} = \{_c\, Z_i{}'\}, \tag{2.55}$$

$$\Sigma_1{}'\Sigma_2(\Sigma_2{}'\Sigma_2)^{-} = [Z_1 \quad \cdots \quad Z_r] = \{_r\, Z_i\}, \tag{2.56}$$

$$\Sigma_1{}'M_2\Sigma_1 = \Sigma_1{}'[I - \Sigma_2(\Sigma_2{}'\Sigma_2)^{-}\Sigma_2{}']\Sigma_1 = \Sigma_1{}'\Sigma_1 - \Sigma_1{}'\Sigma_2(\Sigma_2{}'\Sigma_2)^{-}\Sigma_2{}'\Sigma_1$$

$$= V - [Z_1 \quad \dots \quad Z_r] \begin{bmatrix} (\sigma_1{}^2 Z_1{}')_{q_1 \mathrm{x} N} \\ (\sigma_2{}^2 Z_2{}')_{q_2 \mathrm{x} N} \\ \vdots \\ (\sigma_r{}^2 Z_r{}')_{q_r \mathrm{x} N} \end{bmatrix} = V - \sum_{i=1}^{r} Z_i\, Z_i{}'\sigma_i{}^2 = \sigma_0{}^2. \tag{2.57}$$

So, the inverse of the variance matrix Σ is obtained as follows

$$\Sigma^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \{_d\, \sigma_i{}^{-2} I_{q_i}\} \end{bmatrix} + \begin{bmatrix} I \\ -\{_c\, Z_i{}'\} \end{bmatrix} \sigma_0{}^{-2} I_N [I \quad -\{_r\, Z_i\}]. \tag{2.58}$$

When we substitute the determinant (2.51) and inverse of the variance matrix Σ (2.58) in the joint distribution function given in (2.40),

$$f_{y,b_1,\dots,b_r}(y, b_1, \dots, b_r; \theta)$$

$$= \frac{1}{(2\pi)^{\frac{1}{2}(\sum_{i=0}^{r} q_i)} (\prod_{i=0}^{r} (\sigma_i{}^2)^{q_i})^{\frac{1}{2}}} exp\left[-\frac{1}{2} \sum_{i=1}^{r} \frac{b_i{}' b_i}{\sigma_i{}^2} \right.$$
$$\left. -\frac{1}{2} \left(y - X\beta - \sum_{i=1}^{r} Z_i b_i \right)' \left(y - X\beta - \sum_{i=1}^{r} Z_i b_i \right) / \sigma_0{}^2 \right]. \tag{2.59}$$

Note that $\theta = (\beta, \sigma)'$, and take the logarithm of Eq. (2.59),

$$l = -\frac{1}{2}\left(\sum_{i=0}^{r} q_i\right) ln(2\pi) - \frac{1}{2}\left(\sum_{i=0}^{r} q_i \, ln\sigma_i{}^2\right) - \frac{1}{2}\sum_{i=0}^{r} \frac{b_i{}'b_i}{\sigma_i{}^2}$$

$$- \frac{\frac{1}{2}(y - X\beta - \sum_{i=1}^{r} Z_i b_i)'(y - X\beta - \sum_{i=1}^{r} Z_i b_i)}{\sigma_0{}^2} \tag{2.60}$$

$$= -\frac{1}{2}\left(\sum_{i=0}^{r} q_i\right) ln(2\pi) - \frac{1}{2}\sum_{i=0}^{r} q_i \, ln\sigma_i{}^2 - \frac{1}{2}\sum_{i=1}^{r} \frac{b_i{}'b_i}{\sigma_i{}^2}.$$

By applying some algebra, note that $(y - X\beta - \sum_{i=1}^{r} Z_i b_i) = e$ and $e = b_0$, we get the complete data log-likelihood based on the model (2.6) as in (2.60).

To estimate $\sigma_i{}^2$, the following derivative operations should be performed.

$$-\frac{\partial l}{\partial \sigma_i{}^2} = -\frac{1}{2}\frac{q_i}{\sigma_i{}^2} + \frac{1}{2}\frac{b_i{}'b_i}{(\sigma_i{}^2)^2} \qquad i = 0,1, \dots, r \tag{2.61}$$

$$\hat{\sigma}_i{}^2 = \frac{b_i{}'b_i}{q_i} \qquad i = 1, \dots, r. \tag{2.62}$$

Notice that in the estimation formula of $\sigma_i{}^2$ given in (2.62), $b_i{}'b_i$ term is not known. Instead of this term, conditional expectation of it, $E(b_i{}'b_i|y)$ will be used. To calculate this conditional expectation, we need the conditional distribution of random effect b_i given y. By using the definition of marginal and conditional distribution of a normal distributed random variable (Appendix B-2), we can write

$$b_i|y \sim N\big(\sigma_i{}^2 Z_i{}'V^{-1}(y - X\beta), \sigma_i{}^2 I_{q_i} - \sigma_i{}^4 Z_i{}'V^{-1}Z_i\big) \quad \text{with}$$

$$E(b_i|y) = \sigma_i{}^2 Z_i{}'V^{-1}(y - X\beta), \tag{2.63}$$

$$cov(b_i|y) = \sigma_i{}^2 I_{q_i} - \sigma_i{}^2 Z_i{}'V^{-1}\sigma_i{}^2 Z_i = \sigma_i{}^2 I_{q_i} - \sigma_i{}^4 Z_i{}'V^{-1}Z_i. \tag{2.64}$$

To calculate $E(b_i{}'b_i|y)$, one can make use of the expected value feature in the quadratic form as mentioned in Appendix B-3,

$$E(b_i{}'b_i|y) = tr\left(\sigma_i{}^2 I_{q_i} - \sigma_i{}^4 Z_i{}'V^{-1}Z_i\right) + \sigma_i{}^4(y - X\beta)'V^{-1}Z_iZ_i{}'V^{-1}(y - X\beta). \tag{2.65}$$

We can estimate β by excluding all random effect terms from dependent vector y and then by applying generalized LSE (Searle et al. [37]),

$$X\hat{\beta} = X(X'X)^{-}X'\left(y - \sum_{i=1}^{r} Z_i\, b_i\right). \tag{2.66}$$

We need to calculate conditional expectation given below to estimate $X\beta$:

$$\begin{aligned}\left(y - \sum_{i=1}^{r} Z_i\, b_i \mid y\right) &= X\beta - E\left(\sum_{i=1}^{r} Z_i\, b_i \mid y\right) = X\beta - \sum_{i=1}^{r} Z_i\, E(b_i \mid y) \\ &= X\beta - \sum_{i=1}^{r} Z_i\left(\sigma_i{}^2 Z_i{}'V^{-1}(y - X\beta)\right) \\ &= X\beta - \sum_{i=1}^{r} \sigma_i{}^2 Z_iZ_i{}'V^{-1}(y - X\beta)\,.\end{aligned} \tag{2.67}$$

Remember $V = \sum_{i=1}^{r} Z_i\, Z_i{}'\sigma_i{}^2 + \sigma_e{}^2 I_N$ and then write $\sum_{i=1}^{r} \sigma_i{}^2 Z_i\, Z_i{}' = V - \sigma_e{}^2 I_N$ in the Eq. (2.67),

$$\begin{aligned}E\left(y - \sum_{i=1}^{r} Z_i\, b_i \mid y\right) &= X\beta - (V - \sigma_e{}^2 I_N)\, V^{-1}\,(y - X\beta) \\ &= X\beta + \sigma_e{}^2\, V^{-1}\,(y - X\beta).\end{aligned} \tag{2.68}$$

By obtaining conditional expectation terms of random effect b_i, we can now move forward with EM algorithm steps.

E-step:

Decide the initial values of β^0 and σ^0, and take $m = 0$.

We can write conditional expectation of $b_i'b_i$ as follows by using (2.65)

$$\hat{t}_i^{\,m} = E(b_i'b_i|y)|_{\beta=\beta^{(m)}\ \sigma^2=\sigma^{2(m)}}$$

$$= tr\left(\sigma_i^{2(m)} I_{q_i} - \sigma_i^{4(m)} Z_i' \left(V^{(m)}\right)^{-1} Z_i\right) \tag{2.69}$$

$$+\sigma_i^{4(m)}\left(y - X\beta^{(m)}\right)'\left(V^{(m)}\right)^{-1} Z_i Z_i' \left(V^{(m)}\right)^{-1}\left(y - X\beta^{(m)}\right).$$

Note that V^m is calculated by writing $\sigma_i^{2(m)}$ in V instead of σ_i^2 for $i = (1,2,\dots,r)$.

The conditional expectation denoted by $\hat{s}^{(m)}$ is calculated as below by using (2.68)

$$\hat{s}^{(m)} = E\left(y - \sum_{i=1}^{r} Z_i\, b_i \mid y\right)|_{\beta=\beta^{(m)}\ \sigma^2=\sigma^{2(m)}}$$

$$= X\beta^{(m)} + \sigma_e^{2(m)}\left(V^{(m)}\right)^{-1}\left(y - X\beta^{(m)}\right). \tag{2.70}$$

M-step:

Find the parameters by maximizing the log-likelihood of the complete data.

$$\sigma_i^{2(m+1)} = \frac{\hat{t}_i^{\,m}}{q_i} \qquad i = 0,1,\dots,r \tag{2.71}$$

$$X\beta^{(m+1)} = X(X'X)^{-}X'\hat{s}^{(m)} \tag{2.72}$$

If convergence happens, take $\hat{\sigma}_i^{\,2} = \sigma_i^{2(m+1)}$ and $\hat{\beta} = \beta^{(m+1)}$. If not, return to step 1 by increasing the m by one.

CHAPTER 3

LAPLACE LINEAR MIXED MODELS

The multivariate Laplace distribution, which is used instead of the standard normal distribution for robust models, will be examined in this section. Following the explanations, the use of this distribution for LMM (Yavuz and Arslan [10], Yavuz [41]) will be detailed. The Skew Laplace Linear Mixed Model (SL-LMM) represents an extension of the Laplace Linear Mixed Model (L-LMM). Consequently, the Laplace LMM frequently finds favor in statistical inference due to its demonstrated reliability and precision.

3.1 Multivariate Exponential Power Distribution

In the context of robustness in Bayesian modeling, the exponential power (EP) family, first introduced by Box and Tiao [42], has become a prominent method. West [43] presents the normal scale and mixture properties of EP, while also exploring its connection to the stable distribution class. Infact, the multivariate EP distribution is an extension of the multivariate normal distribution, with an additional parameter K that determines the kurtosis. This parameter distinguishes the EP distribution from the normal distribution.

Let $X = (X_1, \dots, X_n)'$ be an n-dimensional random vector. Then X is distributed $EP_n\big((\mu, \Sigma), \mathrm{K}\big)$ with parameters, $\mu \in \mathbb{R}$, Σ is $(n \times n)$ positive definite symmetric matrix and $\mathrm{K} \in (0,\ \infty)$, if its density is given as below

$$f(x; \mu, \Sigma, \mathrm{K}) = d\ |\Sigma|^{-\frac{1}{2}}\, exp\left(-\frac{1}{2}\big((x-\mu)'\ \Sigma^{-1}\ (x-\mu)\big)^{\mathrm{K}}\right), \tag{3.1}$$

$$d = \frac{n\Gamma\left(\frac{n}{2}\right)}{\pi^{\frac{n}{2}}\,\Gamma\left(1+\frac{n}{2\mathrm{K}}\right)2^{1+\frac{n}{2\mathrm{K}}}},$$

where K is the kurtosis parameter [Arslan [20], Fang et al. [44], Gómez-Sánchez-Manzano et al. [45], Gómez et al. [25]). When $\mathrm{K} = 1$, it is clear that the distribution becomes normal.

One of the reasons why the EP distribution used in Laplace LMM is that it can be expressed as a scale mixture of normal (SMN). Before giving this demonstration, a brief reminder will be given about the representation of SMN distribution, which can be expressed as follows:

Definition 3.1: Scale mixture of normal distributions (Lange and Sinsheimer [46]**)**

Let X be a $(n \times 1)$ random vector, $\mu \in \mathbb{R}$ a $(n \times 1)$ location vector, Σ a $(n \times n)$ positive-definite dispersion matrix, and U be a positive random variable with cdf $H(u; v)$, $H(0) = 0$; where v is a scalar or parameter vector indexing the distribution of U. The multivariate SMN class of distributions, denoted by $\mathrm{X} \sim SMN_n(\mu, \Sigma, H)$ can be defined with the following probability density function:

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, H) = 2\int_0^\infty \phi_n(\mathbf{x}; \boldsymbol{\mu}, \kappa(u), \boldsymbol{\Sigma})\, \boldsymbol{\Phi}\left(\kappa(u)^{-1/2}\boldsymbol{\Sigma}^{-\mathbf{1/2}}(\mathbf{x}-\boldsymbol{\mu})\right) dH(u; \boldsymbol{v}), \qquad \mathbf{x} \in R^n$$

for some positive scaling/weight function $\kappa(u)$, where $\phi(.\,; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the pdf of the n-variate normal distribution with a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$, $\Sigma^{-1/2}$ is such that $\Sigma^{-1/2}\,\Sigma^{-1/2} = \Sigma^{-1}$, and $\Phi(.)$ denotes the cumulative distribution function (cdf) of the standard normal distribution.

Theorem 3.1 Representation of EP as a scale mixture of normal distribution

Gómez-Sánchez-Manzano et al. [45] show that exponential power distribution $EP_n(\mu, \Sigma, \mathrm{K})$ is a SMN distribution if and only if its kurtosis parameter K belongs to the interval $(0, 1]$; in this case they also show the mixing distribution function $\mathrm{H_K}$ for $\mathrm{K} \in (0, 1)$ (**Theorem 2.1** in their study) as,

$$h_{\mathrm{K}}(v) = \frac{2^{1+\frac{n}{2}-\frac{n}{2\mathrm{K}}}\,\Gamma\left(1+\frac{n}{2}\right)}{\Gamma\left(1+\frac{n}{2\mathrm{K}}\right)}\; v^{n-3}\, S_K\left(v^{-2}; 2^{1-\frac{1}{2\mathrm{K}}}\right), \qquad v > 0.$$

S_K is the stable distribution with having the characteristic function for $\mathrm{K} \in (0, 1)$ as (Samorodnitsky and Taqqu [47]).

$$\varphi(t) = exp\left\{-\sigma^{\mathrm{K}}\, |t|^{\mathrm{K}}\, e^{-i\frac{\pi}{2}\mathrm{K}\, sign(t)}\right\}.$$

3.1.1 Multivariate Laplace Distribution (Double Exponential Distribution)

Multivariate Laplace distribution (double exponential distribution) is a special case of multivariate EP distribution; if $\mathrm{K} = \frac{1}{2}$, and the mixture form is denoted as $X \sim SMN_n\left(\mu, \Sigma, \mathrm{K} = \frac{1}{2}\right)$ with the scaling variable v^2 and the following density of mixing function (Theorem 2.1, Gomez et al. [25]).

$$h_{1/2}(v) = \frac{\Gamma\left(\frac{n}{2}\right)}{2^{\frac{n}{2}}\,\Gamma(n)}\; v^{n-3}\, S_{\frac{1}{2}}\left(v^{-2}\,;\,\frac{1}{2}\right). \tag{3.2}$$

Samorodnitsky and Taqqu [47] show that, as stated in Gómez-Sánchez- Manzano et al. [45], for $y > 0$, $S_{1/2}(y;\sigma) = \frac{1}{2}\pi^{-\frac{1}{2}}\,\sigma^{\frac{1}{2}}\,y^{-\frac{3}{2}}\exp\left(-\frac{1}{4}\,\sigma y^{-1}\right)$, often called as Lévy density, is an inverted-gamma distribution $IG\left(\frac{1}{2},\frac{\sigma}{4}\right)$. When it is rewritten with parameters in (3.2), $S_{1/2}\left(v^{-2};\frac{1}{2}\right) = 2^{-\frac{3}{2}}\pi^{-\frac{1}{2}}\,v^{-3}\,\exp\left(-\frac{1}{8}\,v^{-2}\right)$ and substitute into (3.2), the mixing distribution becomes a generalized gamma density (Johnson et al. [48]).

$$h_{1/2}(v) = \frac{1}{2^{\frac{3n+1}{2}}\Gamma\left(\frac{n+1}{2}\right)} v^n \exp\left(-\frac{1}{8}v^2\right) I_{(0,\infty)}(v) \quad (3.3)$$

3.2 Laplace Linear Mixed Models

In both theoretical implications and practical applications related to LMM, it is assumed that the error terms follow a normal distribution with random effects. However, when dealing with data sets that have a thicker-tailed distribution or contain outlier observations, the validity of certain assumptions comes into question. Because these assumptions might have a significant impact on the accuracy of regression and variance estimates. Instead of relying on normal distributions, it is necessary to use robust distributions that can make strong and reliable predictions in such situations. The Laplace distribution is a common alternative to the normal distribution in statistical models, as stated in various literary works. Li et al. [49], discuss robust mixture multivariate linear regression by multivariate Laplace distribution. The multivariate Laplace distribution is employed for robust modeling, and a mixture form of the probabilistic partial least squares model is adopted for multimodal description by Yang et al. [50]. Shi et al. [51] also fit a robust mixture regression model by using Laplace distribution. Yavuz [41] , Yavuz and Arslan [10] use multivariate Laplace distribution instead of normality assumption to generate parameter estimators for LMM that have Laplace distributed random effects and error terms as well as only Laplace distributed random effects. The following subsection shows in detail how the multivariate Laplace distribution for both random effects and subject-specific errors is adapted to LMM (Yavuz [41], Yavuz and Arslan [10]). One can see Yavuz [41] for the LMM constructed with only Laplace random effects.

3.2.1 Model Formulation of Laplace Linear Mixed Model

Laplace Linear Mixed Model (L-LMM) with the assumption of multivariate Laplace distribution for both random effects and errors is proposed as follows:

$$
\begin{aligned}
&Y_i = X_i\beta + Z_i b_i + e_i, \qquad\qquad i = 1, \dots, n,\\
&b_i \sim ML_q(0, D),\\
&e_i \sim ML_{n_i}(0, \mathrm{R}_i).
\end{aligned} \tag{3.4}
$$

The density function of b_i's is given as follows:

$$f_{ML}(b_i; 0, D) = c\, |D|^{-\frac{1}{2}}\, exp\left(-\frac{1}{2}(b_i{}'D^{-1}b_i)^{\frac{1}{2}}\right),$$

$$c = \frac{q\Gamma\left(\frac{q}{2}\right)}{\pi^{\frac{q}{2}}\, \Gamma(1+q)\, 2^{q+1}},$$

where y_i denotes $(n_i \times 1)$ vector of continuous responses for the i-th subject, β denotes $(p \times 1)$ vector of unknown fixed effects describing the population mean, X_i denotes $(n_i \times p)$ dimensional full-rank design matrix, b_i denotes $(q \times 1)$ vector of unobservable random effects, Z_i denotes $(n_i \times q)$ dimensional full-rank design matrix, e_i denotes $(n_i \times 1)$ vector of residual errors assumed to be independent of b_i.

Moreover, R_i is a $(n_i \times n_i)$ variance-covariance matrix of within-individual measurements. $\boldsymbol{D}$ is the variance-covariance matrix of the random effects. The joint distribution of $\boldsymbol{y_i}$ and $\boldsymbol{b_i}$ is defined by Yavuz [41] as

$$\begin{bmatrix} y_i \\ b_i \end{bmatrix} \sim Laplace_{n_i+q}\left(\begin{bmatrix} X_i\beta \\ 0 \end{bmatrix}, \begin{bmatrix} Z_iDZ_i{}' + R_i & Z_iD \\ DZ_i{}' & D \end{bmatrix}\right), \qquad i = 1, \dots, n. \tag{3.5}$$

Then, by integrating out $\boldsymbol{b_i}$ terms from (3.5), the marginal distribution of $\boldsymbol{y_i}$ is obtained as

$$\boldsymbol{y_i} \sim Laplace_{n_i}(\boldsymbol{X_i\beta}, \boldsymbol{Z_iDZ_i}{}' + \boldsymbol{R_i}),$$

($\boldsymbol{V_i} = \boldsymbol{Z_i}\boldsymbol{D}\boldsymbol{Z_i}' + \boldsymbol{R_i}$) as stated in Yavuz [41], Arslan [20] and Gómez et al. [25].

The corresponding conditional distributions of $[\boldsymbol{y_i}', \boldsymbol{b_i}']'$ is written as follows by using the scale mixture property of Laplace distribution and Theorem 3.1 (Gómez-Sánchez-Manzano et al. [45]).

$$\begin{bmatrix} y_i \\ b_i \end{bmatrix} | \tau_i \sim N_{n_i+q}\left(\begin{bmatrix} X_i\beta \\ 0 \end{bmatrix}, {\tau_i}^2 \begin{bmatrix} Z_iDZ_i' + R_i & Z_iD \\ DZ_i' & D \end{bmatrix}\right),$$

$$\tau_i \sim h_{1/2}(\tau_i). \tag{3.6}$$

Based on (3.6), L-LMM model given in (3.4) can be hierarchically represented as

$$\boldsymbol{y_i}|\boldsymbol{b_i}, \tau_i \sim N_{n_i}(\boldsymbol{X_i\beta} + \boldsymbol{Z_i b_i}, {\tau_i}^2\boldsymbol{R_i}),$$

$$\boldsymbol{b_i}|\tau_i \sim N_q(0, {\tau_i}^2\boldsymbol{D}),$$

$$\tau_i \sim h_{1/2}(\tau_i), \tag{3.7}$$

$$h_{1/2}(\tau_i) = \frac{1}{2^{\frac{3n+1}{2}}\Gamma\left(\frac{n+1}{2}\right)} {\tau_i}^n \exp\left(-\frac{1}{8}{\tau_i}^2\right) I_{(0,\infty)}(\tau_i).$$

3.2.2 EM Algorithm for Laplace LMM

This section focuses on the application of EM algorithm for estimating parameters in LMM with Laplace distributed random effects and error terms (Yavuz [41]). EM algorithm is implemented for the cases where b_i and τ_i are considered to be missing observations (i) and where b_i is taken out of the density function by integration (ii).

3.2.2.1 EM Algorithm for the case $\boldsymbol{b_i}$ and $\boldsymbol{\tau_i}$ are considered as missing

The hierarchical representation of the L-LMM with multivariate Laplace distributed random effects and error terms, including the conditional distributions of the y_i and b_i as well as the distribution of the mixing variable τ_i is defined in the previous section (3.7). Besides the joint distribution of y_i and b_i is given in the Eq. (3.6). Based on this hierarchical model, EM algorithm steps will be implemented as follow.

$\theta = [\beta', D', R_i{}']'$ is full parameter space, $y = [y_1{}', \dots, y_n{}']'$ is observed data, $b = [b_1{}', \dots, b_n{}']'$ and $\tau = [\tau_1{}', \dots, \tau_n{}']'$ are treated as "missing data". Therefore, the complete data include y, b and τ. Due to the conditional structure of model (3.7), the joint density of complete data can be factored into the product of the conditional densities of y, b, and τ. As a result, the density function can be written as $f(\mathbf{y_i}|\, b_i, \tau_i)\, f(\mathbf{b_i}|\, \tau_i)\, f(\tau_i)$. The complete data log-likelihood based on the model (3.7) is given by

$$\begin{aligned} lnL &= \sum_{i=1}^{n} lnf(y_i; b_i, \tau_i) + \sum_{i=1}^{n} lnf(b_i; \tau_i) + \sum_{i=1}^{n} lnf(\tau_i) \\ &= L_1(\boldsymbol{\beta}, \boldsymbol{R_i}|y_i, b_i, \tau_i) + L_2(\boldsymbol{D}|b_i, \tau_i) + L_3(\tau_i) \end{aligned} \tag{3.8}$$

$$\begin{aligned} lnL = &-\frac{1}{2}\sum_{i=1}^{n}\{(y_i - X_i\beta - Z_i b_i)'\,(\tau_i{}^2 R_i)^{-1}(y_i - X_i\beta - Z_i b_i) + ln|R_i|\} \\ &-\frac{1}{2}\left\{nln|\mathrm{D}| + \sum_{i=1}^{n} \tau_i{}^{-2} b_i{}' D^{-1} b_i\right\} + constant. \end{aligned} \tag{3.9}$$

'$Constant$' means the terms without any parameters to be estimated. To make it easier to display, Eq. (3.8) is expressed as $L_1(\boldsymbol{\beta}, \boldsymbol{R_i}|y_i\,, b_i, \tau_i) = L_1$ and $L_2(\boldsymbol{D}|b_i, \tau_i) = L_2$. By using the equation provided below (3.10), the first term in Eq. (3.8) is once again rewritten and then placed into L_1.

$$(y_i - X_i\beta - Z_i b_i)' R_i{}^{-1}(y_i - X_i\beta - Z_i b_i) \tag{3.10}$$

$$= (y_i - Z_i b_i)' R_i^{-1} (y_i - Z_i b_i) + (X_i\beta)' R_i^{-1} (X_i\beta)$$

$$-2(X_i\beta)' R_i^{-1} (y_i - Z_i b_i)$$

$$= trace\left[R_i^{-1} (y_i - Z_i b_i)(y_i - Z_i b_i)'\right] + \beta' X_i' R_i^{-1} X_i \beta$$

$$-2\beta' X_i' R_i^{-1} (y_i - Z_i b_i).$$

$$L_1 = -\frac{1}{2}\sum_{i=1}^{n}\Big[trace\big(\tau_i^{-2} R_i^{-1} (y_i - Z_i b_i)(y_i - Z_i b_i)'\big) + \tau_i^{-2}\beta' X_i' R_i^{-1} X_i \beta - 2\tau_i^{-2}\beta' X_i' R_i^{-1} (y_i - Z_i b_i) + ln|R_i|\Big] \tag{3.11}$$

$$L_2 = -\frac{1}{2}\left\{nln|\mathrm{D}| + \sum_{i=1}^{n} \tau_i^{-2}\, b_i' D^{-1} b_i\right\} = -\frac{1}{2}\left\{nln|D| + trace\left(D^{-1}\sum_{i=1}^{n} \tau_i^{-2} b_i b_i'\right)\right\}. \tag{3.12}$$

In the log-likelihood functions defined in (3.11) and (3.12), the aim is to estimate β, D and R_i. To perform the E-step of the ECM algorithm, it is necessary to compute the conditional expectations of the log-likelihood function given.

$$E\left[lnL(\boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{R_i} | y, b, \tau) | y, \hat{\theta}\right] =$$

$$E\left[(\boldsymbol{L_1}) | y, \hat{\theta}\right] + E\left[(\boldsymbol{L_2}) | y, \hat{\theta}\right] + constant.$$

$$E\left[(\boldsymbol{L_1}) | y, \hat{\theta}\right] = -\frac{1}{2}\sum_{i=1}^{n}\Big[trace\left(\hat{\tau}_i^{-2} R_i^{-1} \boldsymbol{E}\big((\boldsymbol{y_i} - \boldsymbol{Z_i b_i})(\boldsymbol{y_i} - \boldsymbol{Z_i b_i})'\big)\right) + \hat{\tau}_i^{-2}\beta' X_i' R_i^{-1} X_i \beta - 2\hat{\tau}_i^{-2}\beta' X_i' R_i^{-1}\big(y_i - Z_i \boldsymbol{\hat{b}_i}\big) + ln|R_i|\Big] \tag{3.13}$$

$$E\left[(\boldsymbol{L_2})|y,\hat{\theta}\right] = -\frac{1}{2}\left\{nln|D| + trace\left(D^{-1}\sum_{i=1}^{n}\boldsymbol{E}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}{}^{-2}\boldsymbol{b_i}\boldsymbol{b_i}'\right)\right)\right\} \tag{3.14}$$

The bold written expectation term in Eq. (3.13) can be edited as follows:

$$\boldsymbol{E}[(\boldsymbol{y_i} - \boldsymbol{Z_i b_i})(\boldsymbol{y_i} - \boldsymbol{Z_i b_i})'] = y_i y_i' - y_i \hat{b}_i{}' Z_i' - Z_i \hat{b}_i y_i' + Z_i \boldsymbol{E}(\boldsymbol{b_i b_i}')Z_i'$$

$$cov(b_i, b_i') = \boldsymbol{E}(\boldsymbol{b_i b_i}') - E(b_i)\,E(b_i') \tag{3.15}$$

$$\boldsymbol{E}(\boldsymbol{b_i b_i}') = cov(b_i, b_i') + E(b_i)\,E(b_i') = \hat{\Omega}_i + \hat{b}_i \hat{b}_i{}'.$$

By substituting (3.15) in (3.13) and (3.14) the final form of $E\left[(L_1)|y,\hat{\theta}\right]$ and $E\left[(L_2)|y,\hat{\theta}\right]$ are obtained as below:

$$E\left[(L_1)|y,\hat{\theta}\right] = -\frac{1}{2}\sum_{i=1}^{n}\left[trace\left(\hat{\tau}_i{}^{-2}R_i{}^{-1}\left(\left(y_i - Z_i\hat{b}_i\right)\left(y_i - Z_i\hat{b}_i\right)' + Z_i\hat{\Omega}_i Z_i'\right)\right)\right.$$
$$\left. + \hat{\tau}_i{}^{-2}\beta' X_i' R_i{}^{-1} X_i \beta - 2\hat{\tau}_i{}^{-2}\beta' X_i' R_i{}^{-1}\left(y_i - Z_i\hat{b}_i\right) + ln|R_i|\right], \tag{3.16}$$

$$E\left[(L_2)|y,\hat{\theta}\right] = -\frac{\text{n}}{2}ln|D| - \frac{1}{2}trace\left(D^{-1}\sum_{i=1}^{n}\hat{\tau}_i{}^{-2}\left(\hat{\Omega}_i\hat{b}_i\hat{b}_i{}'\right)\right). \tag{3.17}$$

In the E-step, instead of $\boldsymbol{b_i}$, $\boldsymbol{\tau_i}$ and $\boldsymbol{b_i b_i}'$ in Eq. (3.16) and (3.17) regarded as missing data, we can calculate their conditional expected values;

$$\hat{b}_i = E\left(b_i \mid \theta = \hat{\theta}, y\right), \tag{3.18}$$

$$\hat{\tau}_i{}^{-2} = E\left(\tau_i{}^{-2} \mid \theta = \hat{\theta}, y\right), \tag{3.19}$$

$$\hat{\Omega}_i = cov\left(b_i \mid \theta = \hat{\theta}, y\right). \tag{3.20}$$

From the joint distribution defined in (3.6), conditional expectation (3.18) and covariance (3.20) of the random effect variable are obtained by using the definition of marginal and conditional distribution of a normal distributed random variable (Appendix B-2).

$$\hat{b}_i = \hat{D}Z_i'\hat{V}^{-1}(y_i - X_i\hat{\beta}) = \hat{D}Z_i'(Z_i\hat{D}Z_i' + \hat{R}_i)^{-1}(y_i - X_i\hat{\beta})$$
$$= \left(\hat{D}^{-1} + Z_i'\hat{R}_i^{-1}Z_i\right)^{-1} Z_i'\hat{R}_i^{-1}(y_i - X_i\hat{\beta}) \qquad (3.21)$$

$$\hat{\Omega}_i = \hat{\tau}_i^2\hat{D} - \hat{\tau}_i^2\hat{D}Z_i'(\hat{\tau}_i^2\hat{V}_i)^{-1}Z_i\hat{D}\,\hat{\tau}_i^2 = \hat{\tau}_i^2(\hat{D} - \hat{D}Z_i'\hat{V}^{-1}Z_i\hat{D})$$
$$= \hat{\tau}_i^2\left(\hat{D} - \hat{D}Z_i'(Z_i\hat{D}Z_i' + \hat{R}_i)^{-1}Z_i\hat{D}\right) = \hat{\tau}_i^2\left(\hat{D}^{-1} + Z_i'\hat{R}_i^{-1}Z_i\right)^{-1} \qquad (3.22)$$

The result of the Schur complement given in Appendix A-4 is used to find (3.22). As noticed, the covariance of random effect terms $(\hat{\Omega}_i)$ depends on τ_i with $\hat{\tau}_i^2$ term. In order to find both $\hat{\tau}_i^{-2}$ (3.19) and also $\hat{\tau}_i^2$, conditional distribution of τ_i should be found first.

So, related calculations are carried out as follows:

$$\boldsymbol{f(\tau_i|y_i)} = \frac{\boldsymbol{f(y_i|\tau_i)\,f(\tau_i)}}{\boldsymbol{f(y_i)}} = \frac{N\big((X_i\beta),\ \tau_i^2(R_i + Z_iDZ_i')\big)\,h_{1/2}(\tau_i)}{Laplace\ (X_i\beta,\ Z_iDZ_i' + R_i)} \qquad (3.23)$$

Let write $V_i = Z_iDZ_i' + R_i$ in the Eq. (3.23).

$$f(\tau_i|y_i) = \frac{\dfrac{1}{(2\pi)^{\frac{n}{2}}\,|\tau_i^2V_i|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}(y_i - X_i\beta)'\,(\tau_i^2V_i)^{-1}(y_i - X_i\beta)\right\} \text{ x } h_{1/2}(\tau_i)}{\dfrac{n\Gamma\left(\frac{n}{2}\right)}{\pi^{\frac{n}{2}}\,\Gamma(n+1)\,2^{n+1}|V_i|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}\left((y_i - X_i\beta)'\,(V_i)^{-1}(y_i - X_i\beta)\right)^{\frac{1}{2}}\right\}}$$

$$= \frac{\Gamma(n+1)\ 2^{\frac{n}{2}+1}}{n\Gamma\left(\frac{n}{2}\right)} \tau_i^{-1}\ exp\left\{-\frac{1}{2}(y_i - X_i\beta)'\ (\tau_i^2 V_i)^{-1}(y_i - X_i\beta)\right.$$

$$\left. + \frac{1}{2}\left((y_i - X_i\beta)'\ (V_i)^{-1}(y_i - X_i\beta)\right)^{\frac{1}{2}}\right\} \text{x}\ h_{1/2}(\tau_i). \tag{3.24}$$

Let say $C_i = (y_i - X_i\beta)'\ (V_i)^{-1}(y_i - X_i\beta)$, and rewrite (3.24) by using also $\Gamma(n+1) = n\Gamma(n)$ as below:

$$f(\tau_i|y_i) = \frac{\Gamma(n)}{\Gamma\left(\frac{n}{2}\right)}\ 2^{\frac{n}{2}+1}\tau_i^{-1}\ exp\left\{-\frac{1}{2}\tau_i^{-2}C_i + \frac{1}{2}C_i^{\frac{1}{2}}\right\} \text{x}\ h_{1/2}(\tau_i). \tag{3.25}$$

Substitute $h_{1/2}(\tau_i)$ given before in Eq. (3.3) into the equation above (3.25).

$$f(\tau_i|y_i) = \frac{\Gamma(n)}{\Gamma\left(\frac{n}{2}\right)}\ 2^{\frac{n}{2}+1}\tau_i^{-1}\ exp\left\{-\frac{1}{2}\tau_i^{-2}C_i + \frac{1}{2}C_i^{\frac{1}{2}}\right\}$$
$$\text{x}\frac{1}{2^{\frac{3n+1}{2}}\Gamma\left(\frac{n+1}{2}\right)}\ \tau_i^{n}\ \exp\left(-\frac{1}{8}\tau_i^2\right) I_{(0,\infty)}(\tau_i). \tag{3.26}$$

Rearranging the conditional probability distribution function (3.26) yields:

$$f(\tau_i|y_i) = \frac{\Gamma(n)}{\Gamma\left(\frac{n}{2}\right)\left(\frac{n+1}{2}\right)}\ 2^{\frac{1}{2}-n} exp\left(\frac{1}{2}C_i^{\frac{1}{2}}\right)\tau_i^{n-1} exp\left\{-\frac{1}{2}\left(\tau_i^{-2}C_i\right.\right.$$
$$\left.\left. + \frac{1}{4}\tau_i^2\right)\right\} I_{(0,\infty)}(\tau_i). \tag{3.27}$$

The part of (3.27) depending on τ_i is as follows,

$$f(\tau_i|y_i) \propto \tau_i^{n-1}\ exp\left[-\frac{1}{2}\left(\tau_i^{-2}C_i + \frac{1}{4}\tau_i^2\right)\right]. \tag{3.28}$$

$\zeta_i = \tau_i^2$

$$\tau_i = {\zeta_i}^{1/2}$$

$$\frac{\partial \tau_i}{\partial \zeta} = \frac{1}{2{\zeta_i}^{1/2}}$$

When variable substitution is applied with the above transformations, Generalized Inverse Gaussian (GIG) distributions is achieved. Before proceeding with the calculation of ${\hat{\tau}_i}^2$ and ${\hat{\tau}_i}^{-2}$, it should be noted that the distribution and moment of the GIG are defined as follow (Barndorff-Nielsen [52]; Barndorff-Nielsen and Halgreen [53]).

$$h(x;\lambda,\chi,\psi) = \frac{(\psi/\chi)^{\lambda/2}}{2M_\lambda\left(\sqrt{\chi\psi}\right)} x^{\lambda-1} exp\left(-\frac{1}{2}(\chi x^{-1} + \psi x)\right), \qquad x > 0$$

with the parameters satisfying the following constraints:

$$if\ \lambda < 0, \qquad \chi > 0, \psi \geq 0$$

$$if\ \lambda = 0, \qquad \chi > 0, \psi > 0$$

$$if\ \lambda > 0, \qquad \chi \geq 0, \psi > 0$$

$M_\lambda(.)$ is defined as the modified Bessel function of the third kind. The moment of this distribution is given below

$$E(X^\alpha) = \left(\frac{\chi}{\psi}\right)^{\alpha/2} \frac{M_{\lambda+\alpha}\left(\sqrt{\chi\psi}\right)}{M_\lambda\left(\sqrt{\chi\psi}\right)}. \tag{3.29}$$

To calculate ${\hat{\tau}_i}^2$ and ${\hat{\tau}_i}^{-2}$, which are crucial for this study, we use the above-mentioned moment formula (3.29). Before we proceed with the calculations, it is important to make a note that the following transformations have been applied.

$$E(\zeta_i) = E({\tau_i}^2)$$

$$E\left({\zeta_i}^{-1}\right) = E({\tau_i}^{-2})$$

$$\chi = C_i$$

$\psi = 1/4$

$\lambda = \frac{n}{2}$

The moments are as follows for $\alpha = 1$ and $\alpha = -1$, respectively.

$$E({\tau_i}^2) = E(\zeta_i) = \left(\frac{C_i}{1/4}\right)^{1/2} \frac{M_{\frac{n}{2}+1}\left(\sqrt{C_i \frac{1}{4}}\right)}{M_{\frac{n}{2}}\left(\sqrt{C_i \frac{1}{4}}\right)} \tag{3.30}$$

$$E({\tau_i}^{-2}) = E({\zeta_i}^{-1}) = \left(\frac{C_i}{1/4}\right)^{-1/2} \frac{M_{\frac{n}{2}-1}\left(\sqrt{C_i \frac{1}{4}}\right)}{M_{\frac{n}{2}}\left(\sqrt{C_i \frac{1}{4}}\right)} \tag{3.31}$$

By obtaining distribution and moment of τ_i, we can confidently move forward with the crucial parameter estimation procedure. Parameter estimates are obtained by maximizing the conditional expected values of log-likelihood functions. To estimate variance-covariance matrix related to errors R_i, the following derivative operations should be performed to Eq. (3.16) (Graybill [54], Chap. 10):

$$\begin{aligned} \frac{E\left[(L_1)|y,\hat{\theta}\right]}{\partial {R_i}^{-1}} = -\frac{1}{2}\sum_{i=1}^{n}\Bigg[& trace\left(\frac{{\hat{\tau}_i}^{-2}{R_i}^{-1}\left((y_i - Z_i\hat{b}_i)(y_i - Z_i\hat{b}_i)'\right)}{\partial {R_i}^{-1}}\right) \\ & + \frac{trace\left({\hat{\tau}_i}^{-2}{R_i}^{-1}Z_i\hat{\Omega}_i{Z_i}'\right)}{\partial {R_i}^{-1}} + \frac{{\hat{\tau}_i}^{-2}\beta'{X_i}'{R_i}^{-1}X_i\beta}{\partial {R_i}^{-1}} \\ & - \frac{2{\hat{\tau}_i}^{-2}\beta'{X_i}'{R_i}^{-1}(y_i - Z_i\hat{b}_i)}{\partial {R_i}^{-1}} + \frac{ln|R_i|}{\partial {R_i}^{-1}}\Bigg] \end{aligned} \tag{3.32}$$

The first order derivatives of all terms in (3.32) are given below:

$$\frac{\partial trace\left({\hat{\tau}_i}^{-2}{R_i}^{-1}(y_i - Z_i\hat{b}_i)(y_i - Z_i\hat{b}_i)'\right)}{\partial {R_i}^{-1}} = \tag{3.33}$$

$$\hat{\tau}_i^{-2}\left(2\left(y_i - Z_i\hat{b}_i\right)\left(y_i - Z_i\hat{b}_i\right)' - D_{aa'}\right),$$

$a = \left(y_i - Z_i\hat{b}_i\right).$

$$\frac{\partial trace\left(\hat{\tau}_i^{-2}R_i^{-1}Z_i\hat{\Omega}_iZ_i'\right)}{\partial R_i^{-1}} = 2\hat{\tau}_i^{-2}Z_i\hat{\Omega}_iZ_i' \tag{3.34}$$

$$\frac{\partial\left(\hat{\tau}_i^{-2}\beta'X_i'R_i^{-1}X_i\beta\right)}{\partial R_i^{-1}} = \hat{\tau}_i^{-2}(2X_i\beta\beta'X_i' - D_{bb'}), \tag{3.35}$$

$b = X_i\beta.$

$$\frac{\partial\left(2\hat{\tau}_i^{-2}\beta'X_i'R_i^{-1}\left(y_i - Z_i\hat{b}_i\right)\right)}{\partial R_i^{-1}} = 2\hat{\tau}_i^{-2}X_i\beta\left(y_i - Z_i\hat{b}_i\right)' \tag{3.36}$$

$ln|R_i| = -ln\left|R_i^{-1}\right|,$

$$\frac{\partial ln\left|R_i^{-1}\right|}{\partial R_i^{-1}} = (2R_i - D_R). \tag{3.37}$$

When the results from (3.33) and (3.37) are substituted in the equation (3.32) and set to zero, we get the estimation of R_i as show in (3.38).

$$\begin{aligned}\frac{E\left[(L_1)|y,\hat{\theta}\right]}{\partial R_i^{-1}} = -\frac{1}{2}\sum_{i=1}^{n}\Big[&\hat{\tau}_i^{-2}\left(2\left(y_i - Z_i\hat{b}_i\right)\left(y_i - Z_i\hat{b}_i\right)' - D_{aa'}\right) + 2\hat{\tau}_i^{-2}Z_i\hat{\Omega}_iZ_i' \\ &+ \hat{\tau}_i^{-2}(2X_i\beta\beta'X_i' - D_{bb'}) - 2\hat{\tau}_i^{-2}X_i\beta\left(y_i - Z_i\hat{b}_i\right)' - (2R_i - D_R)\Big] \\ &= 0\end{aligned}$$

$$\hat{R}_i = \frac{1}{n}\sum_{i=1}^{n}\hat{\tau}_i^{-2}\left[\left(y_i - Z_i\hat{b}_i\right)\left(y_i - Z_i\hat{b}_i\right)' + Z_i\hat{\Omega}_iZ_i' + X_i\hat{\beta}\hat{\beta}'X_i' - X_i\hat{\beta}\left(y_i - Z_i\hat{b}_i\right)'\right] \tag{3.38}$$

To estimate the β parameters, we first take derivative of $E\left[(L_1)|y,\hat{\theta}\right]$ with respect to β and then set it equal to zero. This allows us to obtain estimation of the parameter, as shown below

$$\hat{\beta} = \left(\sum_{i=1}^{n} \hat{\tau}_i^{-2} X_i' \hat{R}_i^{-1} X_i\right)^{-1} \sum_{i=1}^{n} \hat{\tau}_i^{-2} X_i' \hat{R}_i^{-1} \left(y_i - Z_i \hat{b}_i\right). \tag{3.39}$$

After applying the same logic to take the derivative of Eq. (3.17), $E\left[(L_2)|y,\hat{\theta}\right]$, with respect to D, we get

$$\hat{D} = \frac{1}{n}\sum_{i=1}^{n} \hat{\tau}_i^{-2} \left(\hat{\Omega}_{\mathrm{i}} + \hat{b}_i \hat{b}_i'\right). \tag{3.40}$$

If we proceed to the ECM algorithm steps:

E-step: For a given $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$**,** calculate $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$, $\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}^{-2}$, $\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}^{2}$and $\hat{\boldsymbol{\Omega}}_{\boldsymbol{i}}$ for $i = 1, \dots, n$.

$$\hat{b}_i^{(m)} = \left(\hat{D}^{-1(m)} + Z_i' \hat{R}_i^{-1(m)} Z_i\right)^{-1} Z_i' \hat{R}_i^{-1(m)} \left(y_i - X_i \hat{\beta}^{(m)}\right), \tag{3.41}$$

$$\hat{\Omega}_i^{(m)} = \hat{\tau}_i^{2(m)} \left(\hat{D}^{-1(m)} + Z_i' \hat{R}_i^{-1(m)} Z_i\right)^{-1}, \tag{3.42}$$

$$\hat{\tau}_i^{2(m)} = \left(\frac{\hat{C}}{1/4}\right)^{1/2} \frac{M_{\frac{n}{2}+1}\left(\sqrt{\hat{C}\frac{1}{4}}\right)}{M_{\frac{n}{2}}\left(\sqrt{\hat{C}\frac{1}{4}}\right)}, \tag{3.43}$$

$$\hat{\tau}_i^{-2(m)} = \left(\frac{\hat{C}}{1/4}\right)^{-1/2} \frac{M_{\frac{n}{2}-1}\left(\sqrt{\hat{C}\frac{1}{4}}\right)}{M_{\frac{n}{2}}\left(\sqrt{\hat{C}\frac{1}{4}}\right)}. \tag{3.44}$$

CM-step 1: Fix $\boldsymbol{R}_{\boldsymbol{i}} = \hat{\boldsymbol{R}}_{\boldsymbol{i}}$ and update $\hat{\boldsymbol{\beta}}$ for $i = 1, \dots, n$ by maximizing $E\left[L_1\left(\boldsymbol{\beta}, \hat{\boldsymbol{R}}|\boldsymbol{y}, \boldsymbol{b}, \boldsymbol{\tau}\right)|\boldsymbol{y}, \hat{\boldsymbol{\theta}}\right]$ over $\boldsymbol{\beta}$.

$$\hat{\beta}^{(m+1)} = \left(\sum_{i=1}^{n} \hat{\tau}_i^{-2(m)} X_i' \hat{R}_i^{-1(m)} X_i\right)^{-1} \sum_{i=1}^{n} \hat{\tau}_i^{-2(m)} X_i' \hat{R}_i^{-1(m)} \left(y_i - Z_i \hat{b}_i^{(m)}\right). \tag{3.45}$$

CM-step 2: Fix $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ and update $\widehat{\boldsymbol{R}}_{\boldsymbol{i}}$ for $i = 1, \dots, n$ by maximizing $E\left[L_1\left(\boldsymbol{R}, \widehat{\boldsymbol{\beta}} | \boldsymbol{y}, \boldsymbol{b}, \boldsymbol{\tau}\right) | y, \hat{\theta}\right]$ over $\boldsymbol{R}_{\boldsymbol{i}}$.

$$\begin{aligned} \hat{R}_i^{(m+1)} &= \frac{1}{n} \sum_{i=1}^{n} \hat{\tau}_i^{-2(m)} \left[\left(y_i - Z_i \hat{b}_i^{(m)}\right)\left(y_i - Z_i \hat{b}_i^{(m)}\right)' + Z_i \widehat{\Omega}_i^{(m)} Z_i' \right. \\ &\quad \left. + X_i \hat{\beta}^{(m)} \hat{\beta}'^{(m)} X_i' - X_i \hat{\beta}^{(m)} \left(y_i - Z_i \hat{b}_i^{(m)}\right)'\right]. \end{aligned} \tag{3.46}$$

CM-step 3: Update $\widehat{\boldsymbol{D}}$ by maximizing $E\left[L_2(\boldsymbol{D}|\boldsymbol{b}, \boldsymbol{\tau}) | y, \hat{\theta}\right]$ over $\boldsymbol{D}$.

$$\hat{D}^{(m+1)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\tau}_i^{-2(m)} \left(\widehat{\Omega}_i^{(m)} + \hat{b}_i^{(m)} \hat{b}_i'^{(m)}\right). \tag{3.47}$$

3.2.2.2 EM Algorithm for the case just $\boldsymbol{\tau_i}$ is considered as missing

Yavuz [41] and Pinheiro [7] apply EM algorithm by integrating out related random effect term b_i from the joint distribution of y_i and b_i, and using just τ_i as missing data. The hierarchical representation of the Laplace distributed LMM defined in (3.5) is given by (3.7). The below model has been attained by integrating out b_i from the joint distribution of y_i and b_i.

$$\begin{aligned} & y_i \mid \tau_i \sim N_{n_i}(X_i \beta, \tau_i^2 V_i), \\ & \tau_i \sim h_{1/2}(\tau_i). \\ & (V_i = Z_i D Z_i' + R_i) \end{aligned} \tag{3.48}$$

The joint probability density function of y_i and τ_i, representing the complete data is as follows

$$f(y_i|\tau_i)\,f(\tau_i) = \frac{1}{|\tau_i{}^2 V_i|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}(y_i - X_i\beta)'\,(\tau_i{}^2 V_i)^{-1}(y_i - X_i\beta)\right\}$$
$$\text{x}\ \frac{1}{2^{\frac{3n+1}{2}}\Gamma\left(\frac{n+1}{2}\right)}\ v^n \exp\left(-\frac{1}{8}v^2\right) I_{(0,\infty)}(v). \tag{3.49}$$

The complete data log-likelihood based on the model (3.48) is given by

$$lnL = \sum_{i=1}^{n} lnf(y_i\,;\tau_i) + \sum_{i=1}^{n} lnf(\tau_i) = L(\boldsymbol{\beta}, \boldsymbol{V_i}|y,\tau) + constant. \tag{3.50}$$

'Constant' denotes the terms that do not require any parameter estimation. Letting $L = L(\beta, V_i|y,\tau)$, we get

$$lnL = -\frac{1}{2}\sum_{i=1}^{n}\{n_i log|\tau_i{}^2 V_i| + (y_i - X_i\beta)'\,(\tau_i{}^2 V_i)^{-1}(y_i - X_i\beta)\} + constant. \tag{3.51}$$

$$E\left(L\middle|y,\hat{\theta}\right) = -\frac{1}{2}\sum_{i=1}^{n}\left\{n_i log|V_i| + \hat{\tau}_i{}^{-2}(y_i - X_i\beta)'\,V_i{}^{-1}(y_i - X_i\beta)\right\} \tag{3.52}$$

$$\hat{\tau}_i{}^{-2} = E\left(\tau_i{}^{-2}|\,\theta = \hat{\theta}, y\right) \tag{3.53}$$

Parameter estimates will be carried out by maximizing (3.52) with respect to parameters. The fixed effect parameter is given as below

$$\hat{\beta} = \left(\sum_{i=1}^{n}\hat{\tau}_i{}^{-2}X_i{}'\hat{V}_i{}^{-1}X_i\right)^{-1}\sum_{i=1}^{n}\hat{\tau}_i{}^{-2}X_i{}'\hat{V}_i{}^{-1}y_i. \tag{3.54}$$

Defining $V = \sum_{i=0}^{q} Z_i Z_i{}'\,\sigma_i{}^2$ as in the previous chapter (2.38) and φ representing each parameter in V, so the first-order derivatives of the parameters as obtained as follows:

$$\frac{\partial E(L)}{\partial \varphi_k} = -\frac{1}{2}\sum_{i=1}^{n}\left\{n_i trace\left(V_i^{-1}\frac{\partial V}{\partial \varphi_k}\right)\right\} - \frac{1}{2}\sum_{i=1}^{n}\left\{\hat{\tau}_i^{-2}(y_i - X_i\beta)'\, V_i^{-1}\frac{\partial V}{\partial \varphi_k}V_i^{-1}\,(y_i - X_i\beta)\right\} \tag{3.55}$$

$$\frac{\partial V}{\partial \varphi_k} = \frac{\partial V}{\partial \sigma_i^2} = \frac{\partial\left(\sum_j Z_j Z_j'\sigma_i^2\right)}{\partial \sigma_i^2} = Z_i Z_i' \tag{3.56}$$

$$\frac{\partial E(L)}{\partial \sigma_i^2} = -\frac{1}{2}\sum_{i=1}^{n}\left\{n_i trace(V^{-1}Z_iZ_i') + \hat{\tau}_i^{-2}(y_i - X_i\beta)'\, V^{-1}Z_iZ_i'V^{-1}(y_i - X_i\beta)\right\} \tag{3.57}$$

$$\sum_{i=1}^{n} n_i trace\left(\hat{V}^{-1}Z_iZ_i'\right) = \sum_{i=1}^{n}\hat{\tau}_i^{-2}\,(y_i - X_i\beta)'\,\hat{V}^{-1}Z_iZ_i'\hat{V}^{-1}(y_i - X_i\beta) \tag{3.58}$$

The inferences are the same as the classical LMM inferences, except containing $\hat{\tau}_i^{-2}$ terms. The rest of the inferences are given in the LMM section, so it is not explained here again. During the application of the EM algorithm, τ_i is treated as missing data. Therefore, in the E-step, its conditional density (3.59) needs to be determined to calculate $E\left(\tau_i^{-2} |\, \theta = \hat{\theta}, y\right)$, and then substitute the observed results in the parameter estimates.

$$f(\tau_i \mid y_i) = \frac{f(y_i|\tau_i)\, f(\tau_i)}{f(y_i)} = \frac{N\left((X_i\beta),\ \tau_i^2(R_i + Z_iDZ_i')\right) h_{1/2}(\tau_i)}{N\,(X_i\beta + Z_ib_i,\ \tau_i^2R_i)} \tag{3.59}$$

CHAPTER 4

SKEW LAPLACE LINEAR MIXED MODELS

In this section, we introduce the Skew Laplace Linear Mixed Model (SL-LMM) employing the multivariate skew Laplace (MSL) distribution, a well-established statistical distribution renowned for its robustness in statistical inference. This study presents novel contributions by delineating the model specifications, hierarchical structure, likelihood function formulations, parameter estimations, and the ECM algorithm specifically designed for SL-LMM, marking the first conprehensive exposition of these aspects within the context of this model. We initiate this section with a concise reintroduction of the MSL distribution, elucidating its key properties and characteristics. Subsequently, we seamlessly incorporate this distribution within the framework of a LMM, thereby extending its application to this particular statistical context.

4.1 Multivariate Skew Laplace Distribution

The multivariate Laplace distribution is a family of probability distributions that includes various extensions derived from the univariate Laplace distributions. (e.g., see Anderson [55], Kotz et al. [24]). A special case of the multivariate Kotz-type distribution has been studied recently (Plungpongpun [56] and Naik et al. [57]) to improve statistical methods that enable the execution of various statistical inferences for multivariate data. As mentioned in the previous chapter, the multivariate Laplace distribution, which is a special case of multivariate EP distribution, is presented [Arslan [20], Fang et al. [44], Gómez et al. [25], Gómez-Sánchez-Manzano et al. [45], Guney et al. [58]].

The multivariate generalization of the Laplace distribution is utilized as a heavy-tailed alternative to the normal distribution when robustness becomes a concern. However, the symmetric nature of the multivariate Laplace distribution may not adequately capture the characteristics of the data when skewness is present. In order to simultaneously address heavy-tailedness and skewness, it is required to introduce a skew extension of the multivariate Laplace distribution to construct a meaningful model. There are different skew extensions of the multivariate Laplace distributions in the literature. Generalized asymmetric Laplace (GAL) distribution which is a recent example of the normal mean-variance mixture (NMVM) distributions is introduced by Kotz et al. [23], and Kotz et al. [24]. Asymmetric Laplace (AL) distribution, skew extension of the multivariate Laplace distribution proposed by Anderson [55], has been introduced and thoroughly examined by Kozuboswki and Podgorski ([59], [60], [61]), Kotz et al. [24], and Kollo and Srivastava [62].

In this paper, we use an alternative MSL distribution, defined by Arslan [20]. This skew extension can effectively set up a significant model in order to handle both heavy tails and skewness simultaneously. It possesses a straightforward structure and mathematically more tractable density function in comparison to alternative multivariate skew distributions. Furthermore, Arslan [20] uses NMVM approach to define MSL distribution and defines the hierarchical representation of MSL distribution.

4.1.1 Review of Multivariate skew Laplace Distribution

In probability theory, multivariate skew Laplace distribution is a versatile multivariate distribution introduced by Arslan [20]. If a random vector X in $R^q,\ q \geq 1$ is $X \sim MSL_q(\mu, \Sigma, \gamma)$, then its density function is given as the following:

$$f_{MSL}(x; \mu, \Sigma, \gamma) = \frac{|\Sigma|^{-1/2}}{2^q \pi^{(q-1)/2}\, \alpha\, \Gamma\left(\frac{q+1}{2}\right)} e^{-\alpha\sqrt{(x-\mu)'\Sigma^{-1}(x-\mu)} + (x-\mu)'\Sigma^{-1}\gamma}, \tag{4.1}$$

where $x \in R^q$, $\mu \in R^q$ is the location parameter, Σ is the positive definite scatter matrix parameter, $\gamma \in R^q$ is the skewness or drift parameter and $\alpha = \sqrt{1 + \gamma'\Sigma^{-1}\gamma}$.

In case of $\gamma = 0$, the density function of X reduces to multivariate Laplace distribution given in Eq. (3.1). When $q = 1$, the density of the MSL distribution overlaps with the univariate density function of AL distribution as stated in Arslan [20].

Versatility of the distribution is given by Arslan [20] for the univariate skew Laplace density by setting $\mu = 0$ and $\sigma = 1$ for simplicity, which is also depicted in Figure 4.1. In this figure, this distribution is characterized by peakedness, heavy-tailedness, and skewness. The details and other properties of MSL distribution can be found in the paper of Arslan [20].

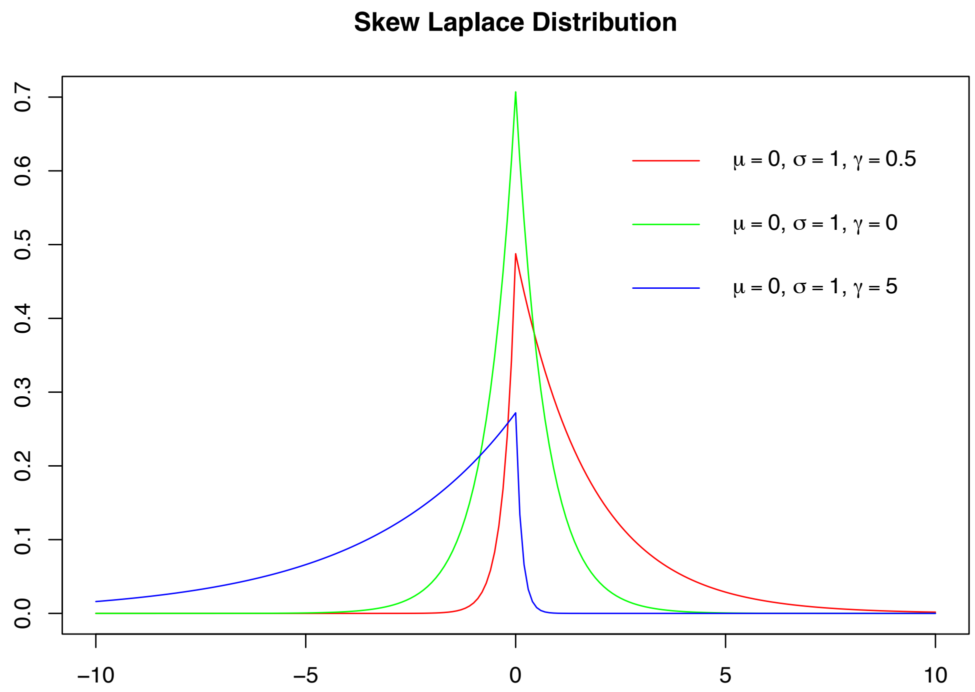

Figure 4.1. Univariate skew (asymmetric) Laplace density plots with different skewness γ parameter

Other important properties are given by the following.

i. The characteristic function of $X \sim MSL_q\ (\mu, \Sigma, \gamma)$ is

$$\Phi_X(t) = e^{it'\mu}[1 + t'\Sigma t - 2it'\gamma]^{-(q+1)/2} \quad , \ t \in R^q. \tag{4.2}$$

ii. The moments of $X \sim MSL_q\ (\mu, \Sigma, \gamma)$ is

$$E(V^{-k}) = \frac{2^k\,\Gamma\left(\frac{q+1}{2} + k\right)}{\Gamma\left(\frac{q+1}{2}\right)} = 2^k \prod_i^k \left(\frac{q+1}{2} + k - i\right). \tag{4.3}$$

Specifically, the expectation and variance of X are given by

$$E(X) = \mu + (q+1)\gamma, \tag{4.4}$$

$$Var(X) = (q+1)(\Sigma + 2\gamma\gamma'). \tag{4.5}$$

iii. If $X \sim MSL_q(\mu, \Sigma, \gamma)$ and $Y = BX + b$, where $B \in R^{p\times q}$ and $b \in R^p$, then $Y \sim MSL_p(B\mu + b,\ B\Sigma B', B\gamma)$. The proof is given in Arslan [20].

As a consequence, it is clear that if $X \sim MSL_q(\mu, \Sigma, \gamma)$, then $a'X \sim MSL_1(a'\mu, a'\Sigma a, a'\gamma)$, for any $a \in R^q$.

iv. The marginal and conditional distribution of $X \sim MSL_q\ (\mu, \Sigma, \gamma)$ if $X,\ \mu,\ \gamma,$ and Σ are partitioned as

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}, \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where $X_1,\ \mu_1$, and γ_1 are $(k \times 1)$ vectors, Σ_{11} is $(k \times k)$ matrix.

- $X_1 \sim MSL_k(\mu_1, \Sigma_{11}, \gamma_1)$,

- The conditional distribution of X_2 given X_1 is a normal variance-mean mixture distribution with the parameters $\mu_{2.1} = \mu_2 + \Sigma_{21}{\Sigma_{11}}^{-1}(x_1 - \mu_1)$, $\gamma_{2.1} = \gamma_2 - \Sigma_{21}{\Sigma_{11}}^{-1}\gamma_1$ and $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}{\Sigma_{11}}^{-1}\Sigma_{12}$.

It is detected that the conditional distribution of X_2 given X_1 belongs to the class of NMVM distributions; however, it is not a MSL distribution.

4.1.2 Representation of MSL Distribution as a Normal Mean-Variance Mixture Distribution

Arslan [20] defines MSL distribution as a normal mean-variance mixture (NMVM) form, which is one of the reasons why this distribution is being used in this study. Before proceeding with this demonstration, a review of the definition of NMVM is given in short as follows.

Review of NMVM Definition: (Definition 2.1 at Barndorff-Nielsen et al. [63])

Suppose $X = (X_1, \dots, X_q)'$ is a random vector which, for a given $v \geq 0$, follows a q-dimensional normal distribution with covariance matrix $v\Sigma$ and mean vector $\mu + v\gamma$, where Σ is a symmetric, positive-definite $(q \times q)$ matrix with determinant 1, and μ and γ are vectors of dimension q. Suppose furthermore that v follows a probability distribution F on $[0, \infty)$. Note that we call distribution function F by the name of mixing/scaling distribution function. Then, it is said that the distribution of X is a normal variance-mean mixture with position μ, drift or skewness γ, structure matrix Σ and mixing distribution F. In case of $\gamma = 0$, we speak of normal variance mixture or well-known as the scale mixture of normal distribution.

In order to avoid an unidentified scale factor, condition $|\Sigma| = 1$ is imposed. As an alternative model underlying a normal variance-mean mixture with $\Sigma = \mathrm{I}$ is the position of a q-dimensional Brownian motion with skewness γ, starting at μ and detected at a random time v. One can see Barndorff-Nielsen and Darroch [64] and the related application about therein. The probability density function of X is given as;

$$P(x) = exp\{(x-\mu)\Sigma^{-1}\gamma'\} \int (2\pi v)^{-q/2} \, exp\left\{-\frac{1}{2}(x-\mu)\,(v\Sigma)^{-1}\,(x-\mu)' - \frac{1}{2} v\gamma\Sigma^{-1}\gamma'\right\} F(dv),$$

and characteristic function

$$g(\vartheta) = e^{i\vartheta\mu'} \varphi\left(i\vartheta\gamma' - \tfrac{1}{2}\vartheta\Sigma\vartheta'\right),$$

where φ is the moment generating function of F, that is $\varphi(t) = E\{exp(tv)\}$.

The logic of NMVM is that randomness is being introduced into the covariance matrix and the mean vector of multivariate normal distribution via the mixing variable v (Arslan [20]). The member of NMVM class of distributions can be obtained by choosing appropriate mixing random variable v. The generalized hyperbolic (GH) distribution introduced by Barndorff-Nielsen [53], [52] as stated in Arslan [20] is a well-known subclass of the NMVM distribution. Pourmousa et al. [65] present multivariate NMVM distribution based on Birnbaum-Saunders distribution.

Arslan [20] defines multivariate skew Laplace distribution given in (4.1) by using NMVM approach as follows:

$$X = \mu + V^{-1}\gamma + \sqrt{V^{-1}}\,\Sigma^{1/2} Z \tag{4.6}$$

where $Z \sim N(0, I_q)$ and V has an inverse gamma distribution with the density function

$$g(v) = \frac{1}{\Gamma\left(\frac{q+1}{2}\right) 2^{\frac{q+1}{2}}} \, v^{-\left(\frac{q+1}{2}+1\right)} e^{-\frac{1}{2v}}, \qquad v > 0, \tag{4.7}$$

Z and V are independent of each other. Then, the random variable X given in Eq. (4.6) in R^q has a MSL distribution with the density function given in Eq. (4.1). Arslan [20] (**Proposition 1** in the paper) gives the proof.

Notice that X has a MSL distribution with the following the representation in (4.6), where $V = v \sim IG\left(\frac{q+1}{2}, \frac{1}{2}\right)$ the pdf is given in (4.7). It is said that the random variable X has a NMVM distribution with mixing variable $V = v$, mixing function $V^{-1} = v^{-1}$. Arslan [20] indicates that the conditional distribution of X given $V = v$ is $N_q \sim (\mu + v^{-1}\gamma,\ v^{-1}\Sigma)$. Therefore, the hierarchical representation of NMVM base on MSL random variable is given as follows:

$$X|V = v \sim N_q(\mu + v^{-1}\gamma, v^{-1}\Sigma),$$
$$V \sim IG\left(\frac{q+1}{2}, \frac{1}{2}\right). \tag{4.8}$$

4.2 Skew Laplace Linear Mixed Models

In an endeavor to enhance the versatility and robustness of LMMs, we introduce a novel approach by incorporating the MSL distribution into the random effect as an alternative to the multivariate skew t-distribution. This innovative adaptation introduces additional parameters, which, while potentially complicating the model, endows it with superior capabilities in terms of modelling, thereby accommodating skewness, heavy tailness, and outlier observations in the data. Therefore, instead of the skew t -distribution, a skew distribution with fewer parameters is preferred which makes the estimation procedures tractable. Also, being able to write skew Laplace distribution as a NMVM distribution allows us to define the proposed model as hierarchical form of the proposed model. It is imperative to note that MSL distribution marks its debut within the context of LMM in this study. This represents a notable departure from conventional practises employed in previous studies within the LMM domain.

To obtain parameter estimations for the skew Laplace LMM (SL-LMM), it is essential to acknowledge that these estimates cannot be derived through straightforward closed-form solutions. Instead, an iterative algorithm is imperative for arriving at the final parameter estimations. In this context, we employ an EM-type algorithm designed specifically for LMMs featuring skew Laplace distributed random effects and multivariate Laplace distributed error terms. This method ensures that the parameter estimation process is both effective and robust, making it a valuable tool for modeling complex data distributions.

4.2.1 Model Formulation of SL-LMM

Skew Laplace Linear Mixed Model (SL-LMM) with the assumption of multivariate skew Laplace distribution for random effects and multivariate Laplace distribution for errors is proposed as follows:

$$\begin{aligned} &Y_i = X_i\beta + Z_i b_i + e_i, \qquad i = 1, \dots, n,\\ &b_i \sim MSL_q(0, D, \gamma),\\ &e_i \sim ML_{n_i}(0, \Lambda_i). \end{aligned} \tag{4.9}$$

The density function of b_i's can be defined as follows:

$$f_{MSL}(b_i; 0, \mathrm{D}, \gamma) = \frac{|\mathrm{D}|^{-1/2}}{2^q \pi^{(q-1)/2}\, \alpha\, \Gamma\left(\frac{q+1}{2}\right)} e^{-\alpha\sqrt{b_i{}'\mathrm{D}^{-1}b_i} + b_i{}'\mathrm{D}^{-1}\gamma},$$

where $\boldsymbol{y}_i$ denotes $(n_i \times 1)$ vector of continuous responses for the i-th subject, $\boldsymbol{\beta}$ denotes $(p \times 1)$ vector of unknown fixed effects describing the population mean, $\boldsymbol{X}_i$ denotes $(n_i \times p)$ dimensional full-rank design matrix, $\boldsymbol{b}_i$ denotes $(q \times 1)$ vector of unobservable random effects, $\boldsymbol{Z}_i$ denotes $(n_i \times q)$ dimensional full-rank design matrix, $\boldsymbol{e}_i$ denotes $(n_i \times 1)$ vector of residual errors assumed to be independent of $\boldsymbol{b}_i$.

Moreover, without loss of generality, let us define $\boldsymbol{\Lambda}_i = \sigma^2 \boldsymbol{R}_i$ which is a $(n_i \times n_i)$ variance-covariance matrix of within-individual measurements. $\boldsymbol{R}_i$ are known matrices, generally equal to the identity. $\boldsymbol{D}$ is the variance-covariance matrix of the random effects and $\boldsymbol{\gamma}$ is a $(q \times 1)$ vector of skewness parameters for the random effects.

We present in the previous chapter that EP distribution $EP_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathrm{K})$ is a SMN distribution if and only if its kurtosis parameter K belongs to the interval $(0, 1]$; in this case the mixing distribution function H_{K} for $\mathrm{K} \in (0, 1)$ is given in Theorem 3.1 (Gomez-Sanchez-Manzano et al. [45]). Yavuz and Arslan [10] also proposes LMM based on this special case of EP distribution, specifically multivariate Laplace distribution when the kurtosis parameter $\mathrm{K} = \frac{1}{2}$, and the mixture form is denoted as $SMN_n\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathrm{K} = \frac{1}{2}\right)$ with generalized gamma distributed mixing function Eq. (3.3) and the scaling variable v^2 . Remember Eq. (3.3),

$$h_{1/2}(v) = \frac{1}{2^{\frac{3n+1}{2}} \Gamma\left(\frac{n+1}{2}\right)} v^n \exp\left(-\frac{1}{8} v^2\right) I_{(0,\infty)}(v) .$$

If a random variable $V = v$ has density $h_{\frac{1}{2}}$, then the variable V^2 has the gamma distribution $G\left(\frac{n+1}{2}, \frac{1}{8}\right)$. We can show this result with a basic transformation as below.

$Y = V^2$ then $V = \sqrt{Y}$

$$\frac{dV}{dY} = \frac{1}{2\sqrt{y}}$$

$$f_Y(y) = f_V\left(v = \sqrt{y}\right) . \left|\frac{dV}{dY}\right|$$

$$f_Y(y) = \underbrace{\frac{1}{8^{\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right)} y^{\frac{n-1}{2}} exp\left(-\frac{1}{8} y\right)}_{\sim Gamma\left(\frac{n+1}{2},\ \frac{1}{8}\right)} \tag{4.10}$$

The density function given in Eq. (4.10) and inverse gamma distribution with the density function given in Eq. (4.7) show similarity in distribution, the first one is $Gamma\left(\frac{n+1}{2},\frac{1}{8}\right)$ and the second one is $IG\left(\frac{n+1}{2},\frac{1}{2}\right)$. Therefore, for our model SL-LMM, it can be written that residual errors $e_i \sim ML_{n_i}(\mathbf{0},\boldsymbol{\Lambda}_i)$, is a SMN distribution with the scaling variable $\frac{1}{v^2}$.

Then SL-LMM model given in (4.9) can be hierarchically represented as

$$\boldsymbol{y}_i|\boldsymbol{b}_i,\tau_i \sim N_{n_i}\left(\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{Z}_i\boldsymbol{b}_i\,,\frac{1}{{\tau_i}^2}\boldsymbol{\Lambda}_i\right),$$

$$\boldsymbol{b}_i|\tau_i \sim N_q\left(\frac{1}{\tau_i}\boldsymbol{\gamma},\ \frac{1}{\tau_i}\boldsymbol{D}\right),$$

$$\tau_i \sim g\left(\frac{q+1}{2},\frac{1}{2}\right), \tag{4.11}$$

$$g(\tau_i)=\frac{1}{\Gamma\left(\frac{q+1}{2}\right)2^{\frac{q+1}{2}}}\,{\tau_i}^{-\left(\frac{q+1}{2}+1\right)}e^{-\frac{1}{2\tau_i}}, \qquad \tau_i>0.$$

Using Eq. (4.6) it is written that the random effect variable b_i has a NMVM distribution with mixing variable τ, mixing function τ^{-1} and with density function $g(\tau_i)$ given above.

$$b_i={\tau_i}^{-1}\gamma+\sqrt{{\tau_i}^{-1}}D^{1/2}Z \ \text{ with } Z\sim N\left(0,I_q\right). \tag{4.12}$$

Notice that $Y_i=X_i\beta+A({b_i}'{e_i}')'$ with $A=\left(Z_iI_{n_i}\right)$ (Schumacher et al. [66]). Using the third property of MSL distribution given in the subsection 4.1.1, we have

$$Y\sim MSL_{n_i}(X_i\beta,\ Z_iD{Z_i}'+\Lambda_i,Z_i\gamma). \tag{4.13}$$

By using the NMVM property of the skew Laplace distribution, we can write the joint distribution of y_i and b_i as:

$$\begin{bmatrix} y_i \\ b_i \end{bmatrix} | \tau_i \sim N_{n_i+q}\left(\begin{bmatrix} X_i\beta + \tau_i{}^{-1}Z_i\gamma \\ \tau_i{}^{-1}\gamma \end{bmatrix}, \begin{bmatrix} \tau_i{}^{-1}Z_iDZ_i{}' + \tau_i{}^{-2}\Lambda_i & \tau_i{}^{-1}Z_iD \\ \tau_i{}^{-1}DZ_i{}' & \tau_i{}^{-1}D \end{bmatrix}\right), \tag{4.14}$$

$$\begin{aligned} cov(y, b') = cov\big((X\beta + Zb + e), b'\big) &= cov(Zb, b') + cov(e, b') \\ &= Zcov(b, b') = Z\tau^{-1}D, \\ cov(b, y') = cov(b, (X\beta + Zb + e)') &= cov(b, b'Z') + cov(b, e') \\ &= cov(b, b')Z' = \tau^{-1}DZ'. \end{aligned}$$

4.2.2 EM Algorithm for Skew Laplace LMM

This section includes the MLEs of unknown parameters of SL-LLM (4.9) with EM-type algorithm. For the MLEs of the parameters in the proposed model, we utilize the EM algorithm by using hierarchical model with both b_i and τ_i are treated as missing.

Remember the condition for variance covariance matrix of MSL distributed random effect terms; that is positive-definite matrix with $|D| = 1$. Therefore, ML estimation is executed for two distinct cases. In the first case, we maximize the likelihood without taking into account the constraint $|D| = 1$ imposed on D. And then we regard the constraint and take it into account during the maximization procedure. We investigate both cases to comprehensively explore the subject matter starting with the non-constraint case first.

$\theta = [\beta', \sigma^2, D', \gamma']'$ is full parameter space, $y = [y_1{}', \dots, y_n{}']'$ is observed data, $b = [b_1{}', \dots, b_n{}']'$ and $\tau = [\tau_1{}', \dots, \tau_n{}']'$ are treated as "missing data". Therefore, the complete data include y, b and τ. Due to the conditional structure of model (4.11), the joint density of complete data can be factored into the product of the conditional densities of y, b, and τ. As a result, the density function can be written as $f(\boldsymbol{y_i}|\, b_i, \tau_i)\, f(\boldsymbol{b_i}|\, \tau_i)\, f(\tau_i)$.

The complete data log-likelihood based on the model (4.11) is given by

$$lnL = \sum_{i=1}^{n} lnf(y_i; b_i, \tau_i) + \sum_{i=1}^{n} lnf(\, b_i; \tau_i) + \sum_{i=1}^{n} lnf(\, \tau_i) \qquad (4.15)$$

$$= L_1(\boldsymbol{\beta}, \boldsymbol{\sigma^2}|y_i, b_i, \tau_i) + L_2(\mathbf{D}, \boldsymbol{\gamma}|b_i, \tau_i) + L_3(\tau_i),$$

$$lnL = -\frac{n_i n}{2} ln\sigma^2 - \frac{1}{2}\sum_{i=1}^{n} \frac{{\tau_i}^2}{\sigma^2} (y_i - X_i\beta - Z_i b_i)' \, {R_i}^{-1}(y_i - X_i\beta - Z_i b_i)$$

$$-\frac{1}{2}\left\{nln|\mathrm{D}| + \sum_{i=1}^{n} \tau_i (b_i - {\tau_i}^{-1}\gamma)' D^{-1} (b_i - {\tau_i}^{-1}\gamma)\right\} + constant. \qquad (4.16)$$

For clarity and organization, Eq. (4.15) is expressed as $L_1(\boldsymbol{\beta}, \boldsymbol{\sigma^2}|y_i \,, b_i, \tau_i) = L_1$ and $L_2(\boldsymbol{D}, \boldsymbol{\gamma}|b_i, \tau_i) = L_2$. In equation (4.16), '$constant$' means the terms without any parameters to be estimated. The second and third terms in Eq. (4.16) can be rewritten respectively in L_1 and L_2 by using the equation given below.

$$\begin{aligned} &(y_i - X_i\beta - Z_i b_i)' {R_i}^{-1} (y_i - X_i\beta - Z_i b_i) \\ &\quad = (y_i - Z_i b_i)' {R_i}^{-1} (y_i - Z_i b_i) + (X_i\beta)' {R_i}^{-1} (X_i\beta) \\ &\quad -2(X_i\beta)' {R_i}^{-1} (y_i - Z_i b_i) \\ &\quad = trace\left[{R_i}^{-1} (y_i - Z_i b_i)(y_i - Z_i b_i)'\right] + \beta' {X_i}' {R_i}^{-1} X_i \beta \\ &\quad -2\beta' {X_i}' {R_i}^{-1} (y_i - Z_i b_i). \end{aligned} \qquad (4.17)$$

After some simple algebraic operations, L_1 and L_2 are given as below.

$$L_1 = -\frac{n_i n}{2} ln\sigma^2 - \sum_{i=1}^{n} \frac{{\tau_i}^2}{2\sigma^2} trace\left[{R_i}^{-1}(y_i - Z_i b_i)(y_i - Z_i b_i)'\right]$$

$$+\sum_{i=1}^{n} \frac{{\tau_i}^2}{\sigma^2} \beta' {X_i}' {R_i}^{-1} (y_i - Z_i b_i) - \sum_{i=1}^{n} \frac{{\tau_i}^2}{2\sigma^2} \beta' {X_i}' {R_i}^{-1} X_i \beta. \qquad (4.18)$$

$$L_2 = -\frac{1}{2}\left\{nln|D| + \sum_{i=1}^{n} \tau_i (b_i - {\tau_i}^{-1}\gamma)' D^{-1} (b_i - {\tau_i}^{-1}\gamma)\right\}$$

$$= -\frac{1}{2}\left\{nln|D| + trace\left(D^{-1}\sum_{i=1}^{n}\tau_i b_i b_i{}'\right) - \sum_{i=1}^{n}2\gamma' D^{-1} b_i + \sum_{i=1}^{n}\tau_i{}^{-1}\gamma' D^{-1}\gamma\right\} \tag{4.19}$$

We want to estimate $\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2$ and $\boldsymbol{\gamma}$ parameters in log-likelihood functions defined in (4.18) and (4.19). For this aim, we use ECM algorithm. E-step of ECM algorithm requires the conditional expectations of log-likelihood function written as below:

$$E\left[lnL(\boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{\sigma^2}, \boldsymbol{\gamma}|y, b, \tau)|y, \hat{\theta}\right] =$$

$$E\left[(\boldsymbol{L_1})|y, \hat{\theta}\right] + E\left[(\boldsymbol{L_2})|y, \hat{\theta}\right] + constant.$$

$$E\left[(\boldsymbol{L_1})|y, \hat{\theta}\right] = -\frac{n_i n}{2} ln\sigma^2 - \sum_{i=1}^{n}\frac{1}{2\sigma^2} trace\left[\hat{\tau}_i{}^2 R_i{}^{-1}\boldsymbol{E}\left((\boldsymbol{y_i} - \boldsymbol{Z_i b_i})(\boldsymbol{y_i} - \boldsymbol{Z_i b_i})'\right)\right] + \sum_{i=1}^{n}\frac{\hat{\tau}_i{}^2}{\sigma^2}\beta' X_i{}' R_i{}^{-1}\left(y_i - Z_i\widehat{\boldsymbol{b}}_{\boldsymbol{i}}\right) - \sum_{i=1}^{n}\frac{\hat{\tau}_i{}^2}{2\sigma^2}\beta' X_i{}' R_i{}^{-1} X_i\beta. \tag{4.20}$$

$$E\left[(\boldsymbol{L_2})|y, \hat{\theta}\right] = -\frac{1}{2}\left\{\mathrm{n}ln|\mathrm{D}| + trace\left(D^{-1}\sum_{i=1}^{n}\hat{\tau}_i\boldsymbol{E}(\boldsymbol{b_i b_i}')\right) - \sum_{i=1}^{n}2\gamma' D^{-1}\hat{b}_i + \sum_{i=1}^{n}\hat{\tau}_i{}^{-1}\gamma' D^{-1}\gamma\right\}. \tag{4.21}$$

The bold written expectation term in Eq. (4.20) can be edited as follows:

$$\boldsymbol{E}[(\boldsymbol{y_i} - \boldsymbol{Z_i b_i})(\boldsymbol{y_i} - \boldsymbol{Z_i b_i})'] = y_i y_i{}' - y_i\hat{b}_i{}' Z_i{}' - Z_i\hat{b}_i y_i{}' + Z_i\boldsymbol{E}(\boldsymbol{b_i b_i}')Z_i{}'$$

$$cov(b_i, b_i{}') = \boldsymbol{E}(\boldsymbol{b_i b_i}') - E(b_i)\,E(b_i{}'), \tag{4.22}$$

$$\boldsymbol{E}(\boldsymbol{b_i b_i}') = cov(b_i, b_i{}') + E(b_i)\,E(b_i{}') = \widehat{\Omega}_i + \hat{b}_i\hat{b}_i{}'.$$

By substituting (4.22) in (4.20) and (4.21), the final form of $E\left[(\boldsymbol{L_1})|y,\hat{\theta}\right]$ and $E\left[(\boldsymbol{L_2})|y,\hat{\theta}\right]$ are obtained as below:

$$
\begin{aligned}
E\left[(\boldsymbol{L_1})|y,\hat{\theta}\right] = & -\frac{n_i n}{2} ln\sigma^2 \\
& -\sum_{i=1}^{n} \frac{1}{2\sigma^2} trace\left[\hat{\tau}_i{}^2 R_i{}^{-1}\left((y_i - Z_i\widehat{\boldsymbol{b}}_{\boldsymbol{i}})(y_i - Z_i\widehat{\boldsymbol{b}}_{\boldsymbol{i}})'\right.\right. \\
& \left.\left. + Z_i\widehat{\Omega}_i Z_i{}'\right)\right] + \sum_{i=1}^{n} \frac{\hat{\tau}_i{}^2}{\sigma^2} \beta' X_i{}' R_i{}^{-1}(y_i - Z_i\widehat{\boldsymbol{b}}_{\boldsymbol{i}}) \\
& -\sum_{i=1}^{n} \frac{\hat{\tau}_i{}^2}{2\sigma^2} \beta' X_i{}' R_i{}^{-1} X_i \beta
\end{aligned}
\tag{4.23}
$$

$$
\begin{aligned}
E\left[(\boldsymbol{L_2})|y,\hat{\theta}\right] = & -\frac{1}{2}\left\{\mathrm{n}ln|\mathrm{D}| + trace\left(D^{-1}\sum_{i=1}^{n}\hat{\tau}_i\left(\widehat{\Omega}_i + \hat{b}_i\hat{b}_i{}'\right)\right)\right. \\
& \left. -\sum_{i=1}^{n} 2\gamma' D^{-1}\hat{b}_i + \sum_{i=1}^{n}\hat{\tau}_i{}^{-1}\gamma' D^{-1}\gamma\right\}.
\end{aligned}
\tag{4.24}
$$

We can write the conditional expectations of the terms $\boldsymbol{b_i}$, $\boldsymbol{\tau_i}$, and $\boldsymbol{b_i}\boldsymbol{b_i}'$ in E-step as below:

$$\hat{b}_i = E(b_i|\theta = \hat{\theta}, y), \tag{4.25}$$

$$\widehat{\Omega}_i = cov(b_i|\theta = \hat{\theta}, y), \tag{4.26}$$

$$\hat{\tau}_i = E(\tau_i|\theta = \hat{\theta}, y), \tag{4.27}$$

$$\hat{\tau}_i{}^2 = E(\tau_i{}^2|\theta = \hat{\theta}, y), \tag{4.28}$$

$$\hat{\tau}_i{}^{-1} = E(\tau_i{}^{-1}|\theta = \hat{\theta}, y). \tag{4.29}$$

From the joint distribution defined in (4.14), conditional expectation (4.25) and covariance (4.26) of the random effect variable are obtained by using the definition of marginal and conditional distribution of a normal distributed random variable (Appendix B-2) as follows:

$$\hat{b}_i = \hat{\tau}_i^{-1}\hat{\gamma} + \hat{\tau}_i^{-1}\hat{D}Z_i'\left(Z_i\hat{\tau}_i^{-1}\hat{D}Z_i' + \hat{\tau}_i^{-2}\hat{\Lambda}_i\right)^{-1}\left(y_i - X_i\hat{\beta} - \hat{\tau}_i^{-1}Z_i\hat{\gamma}\right)$$

$$= \hat{\tau}_i^{-1}\hat{\gamma} + \left(\hat{\tau}_i\hat{D}^{-1} + \frac{\hat{\tau}_i^{\,2}}{\hat{\sigma}^2}Z_i'R_i^{-1}Z_i\right)^{-1}\frac{\hat{\tau}_i^{\,2}}{\hat{\sigma}^2}Z_i'R_i^{-1} \left(y_i - X_i\hat{\beta} - Z_i\hat{\tau}_i^{-1}\hat{\gamma}\right), \tag{4.30}$$

$$\hat{\Omega}_i = \hat{\tau}_i^{-1}\hat{D} - \hat{\tau}_i^{-1}\hat{D}Z_i'\left(Z_i\hat{\tau}_i^{-1}\hat{D}Z_i' + \hat{\tau}_i^{-2}\hat{\Lambda}_i\right)^{-1}Z_i\hat{\tau}_i^{-1}\hat{D}$$

$$= \left(\hat{\tau}_i\hat{D}^{-1} + \frac{\hat{\tau}_i^{\,2}}{\hat{\sigma}^2}Z_i'R_i^{-1}Z_i\right)^{-1}. \tag{4.31}$$

The result of the Schur complement (Appendix A-4) is utilized to find (4.31). SL-LMM model (4.11) requires τ_i, that is regarded as missing but imputed from the data for each subject. Additionally, we need to find $\hat{\tau}_i$, $\hat{\tau}_i^{\,2}$ and $\hat{\tau}_i^{-1}$ to achieve estimators of the parameter. For this aim, the conditional distribution of τ_i is required. We will find it with a Bayesian approach where the conditional distribution of it can be defined as $\boldsymbol{f(\tau_i|y_i)} = \frac{f(y_i|\tau_i)\,f(\tau_i)}{f(y_i)}$ and then calculate the expectation of τ_i as below.

$$\hat{\tau}_i = E\left(\tau_i|\theta = \hat{\theta}, y\right) = \int_0^\infty \tau_i\ f(\tau_i|y_i)\ d\tau_i\,,$$

$$\hat{\tau}_i^{\,2} = E\left(\tau_i^2|\theta = \hat{\theta}, y\right) = \int_0^\infty \tau_i^2\ f(\tau_i|y_i)\ d\tau_i\,,$$

$$\hat{\tau}_i^{-1} = E\left(\tau_i^{-1}|\theta = \hat{\theta}, y\right) = \int_0^\infty \tau_i^{-1}\ f(\tau_i|y_i)\ d\tau_i.$$

The conditional distribution function of τ_i is found as follows:

$$\boldsymbol{f(\tau_i|y_i)} = \frac{\boldsymbol{f(y_i|\tau_i)f(\tau_i)}}{\boldsymbol{f(y_i)}} = \frac{N\left((X_i\beta + \tau_i^{-1}Z_i\gamma), \tau_i^{-2}(\tau_iZ_iDZ_i' + \Lambda_i)\right)\,g(\tau_i)}{MSL\ (X_i\beta,\ Z_iDZ_i' + \Lambda_i,\ Z_i\gamma)} \tag{4.32}$$

$$= \frac{2^n \pi^{(n-1)/2} \, \alpha \, \Gamma\left(\frac{n+1}{2}\right) \, |V_i|^{-1/2}}{(2\pi)^{n/2} \, |\tau_i^{-2} \mathrm{A}_i|^{1/2}} exp\left\{-\frac{1}{2}\omega_i^T \, (\tau_i^{-2} \, \mathrm{A}_i)^{-1}\omega_i\right\} \mathrm{x}$$

$$exp\left\{-\alpha\left((y_i - X_i\beta)'V_i^{-1}(y_i - X_i\beta)\right)^{\frac{1}{2}} + (y_i - X_i\beta)'V_i^{-1} \, Z_i \, \gamma\right\}^{-1} \tag{4.33}$$

$$\boldsymbol{f(\tau_i|y_i)} = \frac{\alpha \, |V_i|^{1/2}}{(\sqrt{2\pi}) \, exp\left\{-\alpha\left((y_i - X_i\beta)'V_i^{-1}(y_i - X_i\beta)\right)^{\frac{1}{2}} + (y_i - X_i\beta)'V_i^{-1} \, Z_i \, \gamma\right\}} \mathrm{x}$$

$$\frac{\tau_i^{-\left(\frac{n+1}{2}\right)} \, e^{-\frac{1}{2\tau_i}} \, exp\left\{-\frac{1}{2}\omega_i' \, (\tau_i^{-2} \, \mathrm{A}_i)^{-1}\omega_i\right\}}{|\mathrm{A}_i|^{1/2}} \tag{4.34}$$

where $(y_i - X_i\beta - \tau_i^{-1}Z_i \, \gamma) = \omega_i$, $V_i = (Z_iDZ_i' + \Lambda_i)$, $\mathrm{A}_i = \tau_i \, Z_iDZ_i' + \Lambda_i$, and $\alpha = \sqrt{1 + (Z_i\gamma)'(Z_iDZ_i' + \Lambda_i)^{-1}Z_i\gamma}$.

When the conditional distribution of τ_i, shown as above in Eq. (4.34) is examined, the analytical solution of expected values of τ_i, τ_i^2 and τ_i^{-1} cannot be found. In the light of the absence of an analytical solution, The Metropolis Algorithm, which is a type of Markov Chain Monte Carlo (MCMC) methods is employed as a scientifically robust method to address the problem at hand. Before explaining MCMC steps, let us finalize with the parameter estimation procedures.

Parameter estimations will be obtained by maximizing the conditional expected values of the log-likelihood functions. Fixed effect parameter β and variance covariance of error terms σ^2 will be estimated by maximizing the conditional expectations of log-likelihood function (4.23). Variance covariance matrix of random effect terms D and skewness parameter γ will be estimated by maximizing the conditional expectations of log-likelihood function (4.24).

In order to estimate the variance covariance matrix for errors σ^2, the following differentiation operations should be performed.

$$\frac{\partial E\left[(L_1)|y,\hat{\theta}\right]}{\partial\sigma^2} = -\frac{n_i n}{2} ln\sigma^2 \tag{4.35}$$

$$-\sum_{i=1}^{n}\frac{1}{2\sigma^2} trace\left[\hat{\tau}_i{}^2 R_i{}^{-1}\left(\left(y_i - Z_i\widehat{\boldsymbol{b}}_{\boldsymbol{i}}\right)\left(y_i - Z_i\widehat{\boldsymbol{b}}_{\boldsymbol{i}}\right)' + Z_i\widehat{\Omega}_i Z_i{}'\right)\right]$$

$$+\sum_{i=1}^{n}\frac{\hat{\tau}_i{}^2}{\sigma^2}\beta' X_i{}' R_i{}^{-1}\left(y_i - Z_i\widehat{\boldsymbol{b}}_{\boldsymbol{i}}\right) - \sum_{i=1}^{n}\frac{\hat{\tau}_i{}^2}{2\sigma^2}\beta' X_i{}' R_i{}^{-1} X_i\beta$$

$$= -\frac{n_i n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n} trace\left[\hat{\tau}_i{}^2 R_i{}^{-1}\left(\left(y_i - Z_i\widehat{\boldsymbol{b}}_{\boldsymbol{i}}\right)\left(y_i - Z_i\widehat{\boldsymbol{b}}_{\boldsymbol{i}}\right)' + Z_i\widehat{\Omega}_i Z_i{}'\right)\right]$$

$$-2\hat{\tau}_i{}^2\beta' X_i{}' R_i{}^{-1}\left(y_i - Z_i\widehat{\boldsymbol{b}}_{\boldsymbol{i}}\right) + \hat{\tau}_i{}^2\beta' X_i{}' R_i{}^{-1} X_i\beta. \tag{4.36}$$

Remember

$$(y_i - X_i\beta - Z_i b_i)'\, R_i{}^{-1}(y_i - X_i\beta - Z_i b_i)$$

$$= (y_i - Z_i b_i)' R_i{}^{-1}(y_i - Z_i b_i) - 2(X_i\beta)' R_i{}^{-1}(y_i - Z_i b_i) + (X_i\beta)' R_i{}^{-1}(X_i\beta)$$

$$= trace\left[R_i{}^{-1}(y_i - Z_i b_i)(y_i - Z_i b_i)'\right] - 2\beta' X_i{}' R_i{}^{-1}(y_i - Z_i b_i) + \beta' X_i{}' R_i{}^{-1} X_i\beta.$$

Therefore, Eq. (4.36) can be expressed as Eq. (4.37) as below and then set equal to zero.

$$\frac{\partial E\left[(L_1)|y,\hat{\theta}\right]}{\partial\sigma^2} = -\frac{n_i n}{2\sigma^2}$$

$$+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}\hat{\tau}_i{}^2\left[\left(y_i - X_i\hat{\beta} - Z_i\hat{b}_i\right)' R_i{}^{-1}\left(y_i - X_i\hat{\beta}\right.\right. \tag{4.37}$$

$$\left.\left. - Z_i\hat{b}_i\right) + trace\left(\widehat{\Omega}_i Z_i{}' R_i{}^{-1} Z_i\right)\right] = 0$$

As a result, parameter estimation of σ^2 is as below:

$$\hat{\sigma}^2 = \frac{1}{n_i n}\sum_{i=1}^{n} \hat{\tau}_i{}^2 \left[\left(y_i - X_i\hat{\beta} - Z_i\hat{b}_i\right)' R_i{}^{-1}\left(y_i - X_i\hat{\beta} - Z_i\hat{b}_i\right) + trace\left(\hat{\Omega}_i Z_i{}' R_i{}^{-1} Z_i\right)\right]. \tag{4.38}$$

In order to estimate the fixed effects parameter β, we take derivatives with respect to β in the conditional expectations of log-likelihood function (4.23) and then set equal to zero. So, we get

$$\hat{\beta} = \left(\sum_{i=1}^{n} \frac{\hat{\tau}_i{}^2}{\hat{\sigma}^2} X_i{}' R_i{}^{-1} X_i\right)^{-1} \sum_{i=1}^{n} \frac{\hat{\tau}_i{}^2}{\hat{\sigma}^2} X_i{}' R_i{}^{-1}\left(y_i - Z_i\hat{b}_i\right). \tag{4.39}$$

Covariance matrix of random effect terms D and the skewness parameter γ are estimated by maximizing the conditional expectations of log-likelihood function (4.24). To estimate D, the following differentiation operations should be performed (Graybill [54], Chap. 10):

$$\frac{\partial E\left[(L_2)|y,\hat{\theta}\right]}{\partial D^{-1}} = -\frac{1}{2}\left\{\mathrm{n}ln|\mathrm{D}| + trace\left(D^{-1}\sum_{i=1}^{n}\hat{\tau}_i\left(\hat{\Omega}_i + \hat{b}_i\hat{b}_i{}'\right)\right) - \sum_{i=1}^{n} 2\gamma' D^{-1}\hat{b}_i + \sum_{i=1}^{n}\hat{\tau}_i{}^{-1}\gamma' D^{-1}\gamma\right\} \tag{4.40}$$

$$\frac{\partial trace\left(D^{-1}\sum_{i=1}^{n}\hat{\tau}_i\left(\hat{\Omega}_i + \hat{b}_i\hat{b}_i{}'\right)\right)}{\partial D^{-1}} = \sum_{i=1}^{n}\hat{\tau}_i\left(\hat{\Omega}_i + \hat{b}_i\hat{b}_i{}'\right) \tag{4.41}$$

$$\frac{\partial \sum_{i=1}^{n} 2\gamma' D^{-1}\hat{b}_i}{\partial D^{-1}} = \sum_{i=1}^{n} 2\,\gamma\hat{b}_i{}' \tag{4.42}$$

$$\frac{\partial \sum_{i=1}^{n}\hat{\tau}_i{}^{-1}\gamma' D^{-1}\gamma}{\partial D^{-1}} = \sum_{i=1}^{n}\hat{\tau}_i{}^{-1}\left(2\gamma\gamma' - D_{\gamma\gamma'}\right) \tag{4.43}$$

$$ln|D| = -ln|D^{-1}|,$$

$$\frac{\partial ln|D^{-1}|}{\partial D^{-1}} = D. \tag{4.44}$$

When the results from (4.41) and (4.44) are substituted in the equation (4.40) and set to zero, we get the estimation of D as show in (4.45).

$$\frac{\partial E\left[(L_2)|y,\hat{\theta}\right]}{\partial D^{-1}} = \frac{n}{2}\mathrm{D} - \frac{1}{2}\left\{\sum_{i=1}^{n} \hat{\tau}_i \left(\widehat{\Omega}_i + \hat{b}_i\hat{b}_i{'}\right) - \sum_{i=1}^{n} 2\gamma\hat{b}_i{'} + \sum_{i=1}^{n} \hat{\tau}_i{}^{-1}\left(2\gamma\gamma' - D_{\gamma\gamma'}\right)\right\} = 0$$

$$\widehat{D} = \frac{1}{n}\sum_{i=1}^{n} \hat{\tau}_i \left(\widehat{\Omega}_i + \hat{b}_i\hat{b}_i{'}\right) - 2\,\gamma\hat{b}_i{'} + 2\hat{\tau}_i{}^{-1}\gamma\gamma'. \tag{4.45}$$

With the intention of estimating skewness parameter γ, we derive the conditional expectations of the log-likelihood function (4.24) with respect to γ and subsequently equate it to zero. As a result, the parameter estimation of the skewness parameter is obtained as follows:

$$\hat{\gamma} = \left(\sum_{i=1}^{n} \hat{\tau}_i{}^{-1}\right)^{-1} \sum_{i=1}^{n} \hat{b}_i. \tag{4.46}$$

So far, we get the MLEs of the parameters $\beta, \sigma^2, D, \gamma$ without considering the constraint on D. Following that we will shift our attention to maximize the likelihood with taking into account the constraint $|D| = 1$ imposed on D for our proposed model. The optimization problem under this constraint is as follows

$$\text{Maximize } l(\beta, \sigma^2, D, \gamma)$$

with the condition $|D| = 1$.

In order to find the MLEs for $\beta, \sigma^2, D, \gamma$ parameters of the proposed model SL-LLM (4.9) subject to the constraint $|D| = 1$, we again apply the EM algorithm approach.

The E-step of the EM algorithm will be the same with the first case since we do not use the constraint at the E-step of the EM algorithm. However, we have to maximize $E\left[lnL(\beta, D, \sigma^2, \gamma | y, b, \tau) | y, \hat{\theta}\right]$ subject to the restriction $|D| = 1$. To solve the constrained optimization problem, we employ the method of Lagrange multiplier as Arslan [20] and derive the subsequent estimators for β, σ^2, D, and γ.

$$\hat{\sigma}^2 = \frac{1}{n_i n}\sum_{i=1}^{n} \hat{\tau}_i^{\,2}\left[\left(y_i - X_i\hat{\beta} - Z_i\hat{b}_i\right)' R_i^{-1}\left(y_i - X_i\hat{\beta} - Z_i\hat{b}_i\right) + trace\left(\hat{\Omega}_i Z_i' R_i^{-1} Z_i\right)\right], \tag{4.47}$$

$$\hat{\beta} = \left(\sum_{i=1}^{n} \frac{\hat{\tau}_i^{\,2}}{\hat{\sigma}^2} X_i' R_i^{-1} X_i\right)^{-1} \sum_{i=1}^{n} \frac{\hat{\tau}_i^{\,2}}{\hat{\sigma}^2} X_i' R_i^{-1}\left(y_i - Z_i\hat{b}_i\right), \tag{4.48}$$

$$\hat{\gamma} = \left(\sum_{i=1}^{n} \hat{\tau}_i^{\,-1}\right)^{-1} \sum_{i=1}^{n} \hat{b}_i, \tag{4.49}$$

$$\hat{D} = \frac{\sum_{i=1}^{n} \hat{\tau}_i\left(\hat{\Omega}_i + \hat{b}_i\,\hat{b}_i'\right) - 2\hat{\gamma}\,\hat{b}_i' + 2\hat{\tau}_i^{\,-1}\hat{\gamma}\,\hat{\gamma}'}{\left|\sum_{i=1}^{n} \hat{\tau}_i\left(\hat{\Omega}_i + \hat{b}_i\,\hat{b}_i'\right) - 2\hat{\gamma}\,\hat{b}_i' + 2\hat{\tau}_i^{\,-1}\hat{\gamma}\,\hat{\gamma}'\right|^{1/q}}. \tag{4.50}$$

It is worth noting that the estimates for the error covariance, fixed effects parameter, and skewness parameter remain unchanged in the unconstrained case. However, variance covariance estimator of random effect terms exhibit difference up to a scale factor.

We then move the following ECM algorithm:

E-Step: For a given $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, calculate, $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$, $\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}^{\boldsymbol{-1}}$, $\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}$, $\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}^{\boldsymbol{2}}$ and $\hat{\boldsymbol{\Omega}}_{\boldsymbol{i}}$ for $i = 1, \dots, n$.

$$\hat{b}_i^{(m)} = \hat{\tau}_i^{-(m)}\hat{\gamma}^{(m)} + \left(\hat{\tau}_i^{(m)}\hat{D}^{-1(m)} + \frac{\hat{\tau}_i^{2(m)}}{\hat{\sigma}^{2(m)}} Z_i' R_i^{-1} Z_i\right)^{-1} \frac{\hat{\tau}_i^{2(m)}}{\hat{\sigma}^{2(m)}} Z_i' R_i^{-1}$$

$$\left(y_i - X_i\hat{\beta}^{(m)} - Z_i\hat{\tau}_i^{-(m)}\hat{\gamma}^{(m)}\right), \tag{4.51}$$

$$\widehat{\Omega}_i^{(m)} = \left(\hat{\tau}_i^{(m)}\widehat{D}^{-1(m)} + \frac{\hat{\tau}_i^{2(m)}}{\widehat{\sigma}^{2(m)}} Z_i' R_i^{-1} Z_i\right)^{-1}. \tag{4.52}$$

$\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}^{-\mathbf{1}}, \hat{\boldsymbol{\tau}}_{\boldsymbol{i}}, \hat{\boldsymbol{\tau}}_{\boldsymbol{i}}^{\mathbf{2}}$ will be added after applying Bayesian inference which will be explained in the next section.

CM-Step 1: Fix $\boldsymbol{\sigma}^2 = \widehat{\boldsymbol{\sigma}}^2$ and update $\widehat{\boldsymbol{\beta}}$ for $i = 1, \dots, n$ by maximizing $E\left[L_1(\boldsymbol{\beta}, \widehat{\boldsymbol{\sigma}}^2 | \boldsymbol{y}, \boldsymbol{b}, \boldsymbol{\tau}) | \boldsymbol{y}, \widehat{\boldsymbol{\theta}}\right]$ over $\boldsymbol{\beta}$, leading to

$$\hat{\beta}^{(m+1)} = \left(\sum_{i=1}^{n} \hat{\tau}_i^{2(m)} X_i' R_i^{-1} X_i\right)^{-1} \sum_{i=1}^{n} \hat{\tau}_i^{2(m)} X_i' R_i^{-1} \left(y_i - Z_i \hat{b}_i^{(m)}\right). \tag{4.53}$$

CM-step 2: Fix $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ and update $\widehat{\boldsymbol{\sigma}}^2$ for $i = 1, \dots, n$ by maximizing $E\left[L_1\left(\boldsymbol{\sigma}^2, \widehat{\boldsymbol{\beta}} | \boldsymbol{y}, \boldsymbol{b}, \boldsymbol{\tau}\right) | \boldsymbol{y}, \widehat{\boldsymbol{\theta}}\right]$ over $\boldsymbol{\sigma}^2$.

$$\begin{aligned} \widehat{\sigma}^{2(m+1)} = \frac{1}{n_i n} \sum_{i=1}^{n} \hat{\tau}_i^{2(m)} \Big[& \left(y_i - X_i\hat{\beta}^{(m)} - Z_i\hat{b}_i^{(m)}\right)' R_i^{-1} \left(y_i - X_i\hat{\beta}^{(m)} \right. \\ & \left. - Z_i\hat{b}_i^{(m)}\right) + trace\left(\widehat{\Omega}_i^{(m)} Z_i' R_i^{-1} Z_i\right) \Big]. \end{aligned} \tag{4.54}$$

CM-Step 3: Update $\widehat{\boldsymbol{D}}$ by maximizing $E\left[L_2(\boldsymbol{D}, \boldsymbol{\gamma} |\ \boldsymbol{b}, \boldsymbol{\tau}) | \boldsymbol{y}, \widehat{\boldsymbol{\theta}}\right]$ over $\boldsymbol{D}$, that is

$$\widehat{D}^{(m+1)} = \frac{\sum_{i=1}^{n} \hat{\tau}_i^{(m)} \left(\widehat{\Omega}_i^{(m)} + \hat{b}_i^{(m)} \hat{b}_i'^{(m)}\right) - 2\gamma \hat{b}_i'^{(m)} + 2\hat{\tau}_i^{-(m)}\, \hat{\gamma}^{(m)} \hat{\gamma}'^{(m)}}{\left|\sum_{i=1}^{n} \hat{\tau}_i^{(m)} \left(\widehat{\Omega}_i^{(m)} + \hat{b}_i^{(m)} \hat{b}_i'^{(m)}\right) - 2\gamma \hat{b}_i'^{(m)} + 2\hat{\tau}_i^{-(m)}\, \hat{\gamma}^{(m)} \hat{\gamma}'^{(m)}\right|^{1/q}} \tag{4.55}$$

or update the equation,

$$\begin{aligned} \widehat{D}^{(m+1)} = \sum_{i=1}^{n} \hat{\tau}_i^{(m)} \left(\widehat{\Omega}_i^{(m)} + \hat{b}_i^{(m)} \hat{b}_i'^{(m)}\right) - 2\gamma \hat{b}_i'^{(m)} \\ + 2\hat{\tau}_i^{-(m)}\, \hat{\gamma}^{(m)} \hat{\gamma}'^{(m)}. \end{aligned} \tag{4.56}$$

CM-Step 4: Update $\hat{\boldsymbol{\gamma}}$ by maximizing $E\left[L_2(\boldsymbol{D}, \boldsymbol{\gamma} |\ \boldsymbol{b}, \boldsymbol{\tau})|\boldsymbol{y}, \widehat{\boldsymbol{\theta}}\right]$ over $\boldsymbol{\gamma}$, that is

$$\hat{\gamma}^{(m+1)} = \left(\sum_{i=1}^{n} \hat{\tau}_i^{\,-(m)}\right)^{-1} \sum_{i=1}^{n} \hat{b}_i^{\,(m)}. \quad (4.57)$$

A sequence of related parameters is created by repeating the cycles.

4.3 Bayesian Approach

From a Bayesian perspective, probability distributions are treated as if the parameters of the likelihood are random variables (Givens and Hoeting [67]). Assume that the distribution of Y is defined by the parameter θ. $f(\theta)$ is called as prior distribution, representing the density assigned to θ before observing the data. $L(Y|\theta)$ is denoted as likelihood or sampling density. Then, Bayes' theorem follows:

$$f(\boldsymbol{\theta}|y) = c\, f(\boldsymbol{\theta}) f(\boldsymbol{y}|\theta) = c\, f(\boldsymbol{\theta})\, L(\boldsymbol{\theta}|y), \quad (4.58)$$

where $f(\boldsymbol{\theta}|y)$ is the posterior density of $\boldsymbol{\theta}$, and c is the normalizing constant, often difficult to compute directly.

$$c = \int f(\boldsymbol{\theta})\, L(\boldsymbol{\theta}|y)\, d(\boldsymbol{\theta}).$$

In Bayesian inference, it is common to have to calculate expectations using probability distributions that are only known up to a normalizing constant. In certain cases, it may not be feasible to obtain samples from these distributions. In such situations, the Bayesian approach can provide miraculous solutions.

Monte Carlo simulation is a valuable technique for estimating integrals in various situations. For example, when conducting Bayesian analyses, it is common to express moments of posterior distributions in the form of an integral, but it is usually not possible to evaluate them analytically (Givens and Hoeting [67]). Even though simple Monte Carlo techniques such as rejection and importance sampling are generally very versatile, having a strong proposal distribution is essential to ensure their efficiency (Robert and Casella [68]). This might be challenging, especially in problems with a high number of dimensions. Therefore, Markov Chain Monte Carlo (MCMC) methods have been developed to compute expectations for high-dimensional distributions that are often tough to sample.

4.4 Markov Chain Monte Carlo (MCMC)

Applying Bayesian techniques and the corresponding requirement to compute complex and multi-dimensional integrals have resulted in the use of interest and advancement in MCMC methods in 1980s (Johansen [69]). At this point, MCMC methods gained popularity, especially for approximating expectations of complex probability distributions.

Simulation and Monte Carlo integration can be applied to obtain an approximate or exact sample in case of a target density f can be valued but not conveniently sampled. Such a sample is primarily used to estimate the expectation of a function of $X \sim f(x)$.

MCMC methods can also be used to generate sample from the distribution that approximates f. However, their primary purpose is to provide a reliable method for estimating expectations of functions of X. MCMC methods are different than the simulation techniques by their iterative process and the convenience of adaptability to a wide variety of difficult problems.

Let the sequence $\{X^t\} = \left(X_1{}^{(t)}, \dots, X_p{}^{(t)}\right)$ state a Markov chain for $t = 0, 1, 2, \dots$ When the chain is both irresolvable and not periodic, the distribution of X^t converges to the limiting stationary distribution of the chain. So, MCMC sampling construct such kinds of Markov chain for which the stationary distribution is equivalent to the target distribution f (Givens and Hoeting [67]). An extensive variety of algorithms has been asserted for the construction of a proper chain in the literature. It is also important to note that X^t are serially dependent and may differ substantially from f when t is too small (Givens and Hoeting [67]).

4.4.1 Metropolis Hasting Algorithm

The Metropolis-Hasting (MH) algorithm, which was first defined by Metropolis et al. [70] and then extended by Hastings [71], is a commonly used MCMC methods. The main purpose of its usage is to simulate observations that are derived from complex distributions (Hitchcock [72]). The MH algorithm allows to sample from a probability distribution which we will call our target distribution, even if we do not know the normalizing constant (Givens and Hoeting [67]). It consists of picking an arbitrary starting value and the iteratively accepting or rejecting candidate samples drawn another distribution that is easy to sample (Robert and Casella [68]).

In the MH algorithm, there are two distributions that are taken into the consideration: the target distribution P and the proposal distribution $q(\mathrm{x}^*|\mathrm{x})$. A candidate sample x^* for the new Markov chain state is drawn from the proposal distribution. According to MH algorithm, the chain moves from its current state x to a new state x^* with the probability

$$\alpha(\mathrm{x},\mathrm{x}^*) = min\left[1, \frac{P(\mathrm{x}^*)\, q(\mathrm{x}|\mathrm{x}^*)}{P(\mathrm{x})\, q(\mathrm{x}^*|\mathrm{x})}\right], \tag{4.59}$$

or remains at x.

We can present the algorithmic form of the MH algorithm as in the paper of Chib and Greenberg [73]:

1. Initialize
   i. Pick an arbitrary starting value x^0.
   ii. Set $t = 0$.
2. For each iteration:
   i. Generate a random candidate value x^* from $q(\mathrm{x}^*|\mathrm{x}^t)$.
   ii. Calculate the probability of moving,
   $$\alpha(\mathrm{x}^t,\mathrm{x}^*) = \min\left[1, \frac{P(\mathrm{x}^*)\, q(\mathrm{x}^t|\mathrm{x}^*)}{P(\mathrm{x}^t)\, q(\mathrm{x}^*|\mathrm{x}^t)}\right]$$
   iii. Accept or reject

- Generate a random number $u \sim Unif\ (0, 1)$,
- If $u \leq \alpha$, accept the candidate. Set $\mathrm{x}^{t+1} = \mathrm{x}^*$,
- If $> \alpha$, reject the candidate. Set $\mathrm{x}^{t+1} = \mathrm{x}$.

iv. Increment t, set $t = t + 1$.

In order to prove that the MH algorithm converges to $P(\mathrm{x})$, it is essential to demonstrate that the chain produced through the algorithm shown in (4.59) is an ergodic chain that is aperiodic and irreducible. The chain's lack of periodicity can be attributed to the fact that the proposed move can always be rejected. To ensure irreducibility, we must ensure that the support of $q(\mathrm{x})$ includes the support of $P(\mathrm{x})$ (Debski [74]). Based on Tierney [75] findings, in this case $P(\mathrm{x})$ is the invariant distribution of the chain (Debski [74]). Although MH algorithm is a straightforward and highly effective approach, it requires the fulfillment of specific conditions (Mosegaard and Tarantola [76], Bosch et al. [77]). For example, when dealing with complicated, multidimensional target distributions, selecting an appropriate proposal distribution $q(\mathrm{x})$ can be quite challenging and require consideration. There are different categories of MH algorithms depending on the choice of the proposal distribution $q(\mathrm{x})$ as stated in Debski [74]. Since we work exclusively with a symmetric proposal distribution, we will just provide an explanation of Metropolis algorithm.

4.4.1.1 Metropolis Algorithm

In Metropolis algorithm, the proposal distribution $q(\mathrm{x}^*|\mathrm{x}^t)$ is symmetric:

$q(\mathrm{x}^*|\mathrm{x}^t) = q(\mathrm{x}^t|\mathrm{x}^*)$.

Then acceptance/rejection ratio given in Eq. (4.59) can be rewritten as

$$\alpha(\mathrm{x}, \mathrm{x}^*) = min\left[1, \frac{P(\mathrm{x}^*)}{P(\mathrm{x})}\right] \tag{4.60}$$

Peskun [78] demonstrate that the Metropolis algorithm satisfies the optimal performance condition among various MH algorithms. It is proven that as the number of iterations goes to infinity, the invariant distribution is guaranteed to be reached by MH algorithm. So, it is possible that the chain has not yet reached the invariant stage even after a certain number of iterations (Debski [74]). When $q(\mathrm{x}^*|\mathrm{x}^t)$ is too narrow, during a finite run of the chain, it can result in excessively sampling a specific area of the space. On the other hand, if $q(\mathrm{x}^*|\mathrm{x}^t)$ is too wide, it is possible that the rate of rejection is elevated because of an effort to gather samples from different areas of the space. The ideal scenario is when the sequence explores all local maxima of $P(\mathrm{x})$, which generally results in an acceptance rate of approximately 50% (Peskun [78], Gilks et al. [79], Jackman [80]). However, in the context of high-dimensional problems, where the complexity and dimensionality of the data pose significant challenges, an acceptance ratio of around 23% can indeed be considered acceptable (Robert and Casella [68]).

4.4.1.2 A Study with Metropolis Algorithm

To conclude the parameter estimation procedure started in the section 4.2.2, for SL-LMM with the EM algorithm, we need to calculate the corresponding moments of τ_i: $\hat{\tau}_i$, ${\hat{\tau}_i}^2$ and ${\hat{\tau}_i}^{-1}$. Although we can express moments of τ_i in the form of an integral, we cannot evaluate them analytically. Therefore, we use Metropolis algorithm to sample from the conditional distribution function of τ_i, $f(\tau_i|y_i)$ that is complicated, high-dimensional, and tough to sample. And then we get the expected values of τ_i, ${\tau_i}^2$ and ${\tau_i}^{-1}$ respectively.

In the conditional distribution function of τ_i in Eq. (4.32) the denominator $f(y_i)$ might be thought as a normalizing constant, since it does not contain any terms related τ_i. Therefore, instead of the exact form of $f(\tau_i|y_i)$, we can use a function that is proportional to it in the Metropolis algorithm. Let us define the posterior probability of τ that is proportional to the sampling density for the data multiplied by the prior probability for τ,

$f(\tau_i|y_i) \propto f(y_i|\tau_i)\, f(\tau_i).$

$f(y_i|\tau_i)\, f(\tau_i) \propto N\big((X_i\beta + \tau_i{}^{-1} Z_i\gamma), \tau_i{}^{-2}(\tau_i Z_i D Z_i{}' + \Lambda_i)\big) \text{ x } g(\tau_i).$

$$f(\tau_i|y_i) \propto \frac{1}{(2\pi)^{n/2}\, |\tau_i{}^{-2}\mathrm{A}_i|^{1/2}} exp\left\{-\frac{1}{2}\omega_i{}'(\tau_i{}^{-2}\mathrm{A}_i)^{-1}\omega_i\right\} \text{ x } \tag{4.61}$$

$$\frac{1}{\Gamma\left(\frac{n+1}{2}\right) 2^{\frac{n+1}{2}}}\, \tau_i{}^{-\left(\frac{n+1}{2}+1\right)} e^{-\frac{1}{2\tau_i}}$$

where $(y_i - X_i\beta - \tau_i{}^{-1} Z_i\, \gamma) = \omega_i$ and $\mathrm{A}_i = \tau_i\, Z_i D Z_i{}' + \Lambda_i$. Excluding all constant terms that are independent of τ, the target function is obtained in the following form:

$$f(\tau_i|y_i) \propto \frac{exp\left\{-\frac{1}{2}\omega_i{}'\, (\tau_i{}^{-2}\mathrm{A}_i)^{-1}\omega_i\right\}}{|\mathrm{A}_i|^{1/2}} \tau_i{}^{-\left(\frac{n+1}{2}\right)}\, e^{-\frac{1}{2\tau_i}} \tag{4.62}$$

where $(y_i - X_i\beta - \tau_i{}^{-1} Z_i\gamma) = \omega_i$ and $\mathrm{A}_i = \tau_i Z_i D Z_i{}' + \Lambda_i$. For proposal distribution in Metropolis algorithm, we first choose a uniform distribution centered at τ with a width of 2c.

$$\tau_{proposed} \sim Unif(\tau - c, \tau + c)$$

We start with an arbitrary point $\tau_0 = 0.5$ to be the first sample in the Markov chain. The c is deliberately chosen very small, such as 0.001, because when c increases, τ tends to get negative values. However, τ has a constrained parameter space; that its interval is between 0 and infinity. When τ takes on negative values, it poses a challenge in the implementation of the Metropolis algorithm in R. Specifically, the svd() function encounters difficulties when attempting to calculate the inverse due to the presence of negative values. Based on the sample presented in Figure 4.2 indicates that the sample produced is not representative of the target distribution, and the mixing needs to be improved. A significantly larger number of iterations might be required for convergence. However, even we increase the iteration number, τ gets negative values; this is easily predictable due to value of c.

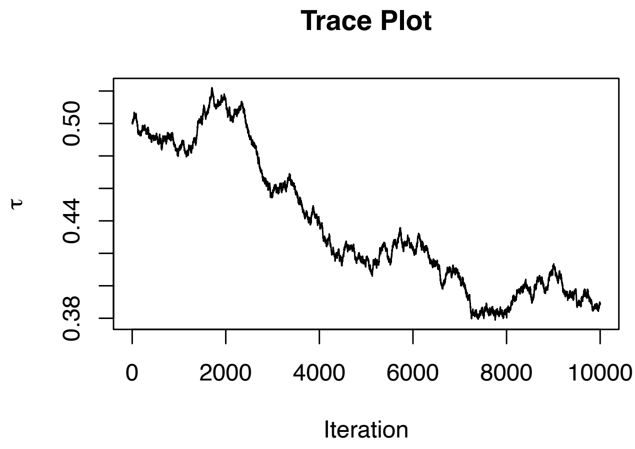

Figure 4.2. Markov Chain Monte Carlo (MCMC) visual convergence diagnostics for the Metropolis algorithm under uniform proposal distribution

To account for the constraint that τ can only take values between zero and infinity, it is essential to incorporate this restriction into the algorithm implemented in R. By incorporating the appropriate constraints, such as defining the range of τ within the code, we ensure that the generated values of τ adhere to the required bounds. This consideration is crucial in maintaining the validity and accuracy of the implementation of Metropolis algorithm. It is possible that we may end up with biased results if this approach is used.

To address this issue, an alternative approach is to work with the distribution of $ln(\tau)$ instead of τ within the Metropolis algorithm. Since natural logarithms have defined only for positive real numbers, this relieves us from this restriction. The range of $ln(\tau)$ for $\tau > 0$ is $(-\infty, \infty)$, which means that its range is all real numbers, so we keep on working Metropolis algorithm.

To get the distribution of $ln(\tau)$, the following transformation is applied:

$m = ln(\tau) \Rightarrow \qquad \tau = e^m \qquad$ and $\qquad \frac{\partial \tau}{\partial m} = e^m$,

$h_M(m) = g_\tau(\tau = e^m\ ) * |\ Jacobian\ |$,

$$h_M(m_i) = \frac{1}{\Gamma\left(\frac{n+1}{2}\right) 2^{\frac{n+1}{2}}} \, (e^m)^{-\left(\frac{n+1}{2}+1\right)} \, e^{-\frac{1}{2e^m}} \, e^m \quad \text{where} \quad -\infty < m < \infty.$$

So, the distribution of $ln(\tau)$ is obtained as

$$h_M(m_i) = \frac{1}{\Gamma\left(\frac{n+1}{2}\right) \, 2^{\frac{n+1}{2}}} \, e^{-\frac{1}{2}(nm_i + m_i + e^{-m_i})} \quad \text{where} \quad -\infty < m < \infty. \tag{4.63}$$

Since the range of the $\tau = e^m$ is all positive real numbers $(0, \infty)$ even if the Metropolis algorithm proposes any real number for m, the output of the target distribution will be a positive real number. In this way, the problem of being unable to get the inverse and calculate the determinant due to the negativity has been solved.

So, instead of the posterior distribution of τ, we use the posterior distribution of $ln(\tau) = m$, given in the Eq. (4.63)

$$f(m_i|y_i) \propto \frac{e^{m_i}}{(2\pi)^{n/2} \, |\mathrm{A}_i|^{1/2}} exp\left\{-\frac{1}{2}\omega_i{}' \, (e^{-2m_i}\mathrm{A}_i)^{-1}\omega_i\right\} \text{ x} \\ \frac{1}{\Gamma\left(\frac{n+1}{2}\right) \, 2^{\frac{n+1}{2}}} \, e^{-\frac{1}{2}(nm_i + m_i + e^{-m_i})} \tag{4.64}$$

where $(y_i - X_i\beta - e^{-m_i}Z_i \, \gamma) = \omega_i$ and $\mathrm{A}_i = e^{m_i} \, Z_i D Z_i{}' + \Lambda_i$.

Now, our target density that is proportional to the desired probability density function is $f(m_i|y_i)$. We will generate a sequence of sample values iteratively. In our study, we have introduced a modification to the proposal distribution by incorporating a Gaussian distribution $q(\mathrm{m}^*|\mathrm{m}^t)$ proposes a candidate for the next sample value based on the previous sample value. By utilizing the normal distribution as our proposal distribution in the context of MH algorithm simulations, we have observed distinct patterns in the behavior of the Markov Chain for various starting values and variances. When considering different starting values, we have noticed that the choice of initial point can significantly impact the convergence and exploration properties of the Markov Chain. Depending on the distance between the starting value and the target distribution, we may observe varying rates of convergence. If the starting value is located in a region of high probability density within the target distribution, the Markov Chain tends to converge rapidly towards the desired distribution. Conversely, if the initial value is situated in a low-probability region, the Markov Chain may exhibit slower convergence as it explores the parameter space to reach areas of higher probability (Robert and Casella [68]).

Initialization: We start with an arbitrary point $m_0 = 0.5$ to be the first observation in the Markov chain.

$$m_{proposed} \sim Normal\ (m, 1)$$

Consequently, a candidate value for the next sample can be easily sampled from the proposal distribution $N(0.5, 1)$. For each iteration, Metropolis algorithm proceeds to suggest a candidate for the next sample value based on the current sample value. Following the previous step, with a certain probability, the candidate is either accepted; leading to the utilization of the candidate value in the subsequent iteration, or it is rejected, resulting in the discarding of the candidate value and the reuse of the current value in the next iteration. Since proposal distribution is Gaussian and exhibits symmetry, we employ Metropolis algorithm instead of MH algorithm.

We can summarize Metropolis algorithm experiment as

1. Initialize
   i. Pick an arbitrary starting value $\mathrm{m}_0 = 0.5$.
   ii. Set $t = 0$.
2. For each iteration:
   i. Generate a random candidate value m^* from $N(\mathrm{m}^*|\mathrm{m}^t)$.
   ii. Calculate the probability of moving,

$$\alpha(\mathrm{m}^t, \mathrm{m}^*) = \min\left[1, \frac{P(\mathrm{m}^*)}{P(\mathrm{m}^t)}\right]$$

   iii. Accept or reject
   - Generate a random number $u \sim Unif\ (0, 1)$,

- If $u \leq \alpha$, accept the candidate. Set $\mathrm{m}^{t+1} = \mathrm{m}^*$,
- If $> \alpha$, reject the candidate. Set $\mathrm{m}^{t+1} = \mathrm{m}$.

iv. Increment t.

The goal we wish to achieve with the Metropolis algorithm is to simulate the posterior distribution of interest. After creating a Markov chain with the desired probability distribution, we generate a sample by simulation. In order to assess if the target distribution of the solution space is matched by the posterior distribution of the sample, we must evaluate the convergence diagnostic of MCMC (Du et al. [81], Roy [82]). The assessment of sample quality through convergence diagnostics can be based on three main criteria ([83]). The first is stationarity, which determines whether the sample chain successfully represents the target distribution. The second is mixing, it determines whether the samples are effectively covering the entire posterior probability distribution. The last one is the sampling efficiency, which refers to enough iterations been returned to adequately describe the posterior distribution. Assessing the quality of sample chains can be done using various convergence diagnostics. Nevertheless, there is no definitive diagnostic that stands out as the best, as each one estimates convergence differently (Mengersen et al. [84]) and may not deliver good performance in all situations (Du et al. [81]). Therefore, it is recommended to use multiple diagnostics in conjunction with each other for optimal results (Du et al. [81], [85], Hogg et al. [85]). The MCMC convergence diagnostics for this study are given in Appendix C.

After examining and confirming the statistical MCMC convergence diagnostics of the sample, we utilize the simulated values to investigate the characteristics of the probability distribution and calculate the corresponding moments of τ. Simulations of the Metropolis algorithm are performed using varying numbers of iterations. However, the final result is determined based on a total of 20,000 iterations. By averaging accepted candidates from the stationary distribution of a Metropolis chain, we roughly estimate the expectation of τ_i for each $i = 1, \dots, n$. Similarly, we take the inverse and square of the accepted candidates, and then calculate the expectation of

${\tau_i}^{-1}$ and ${\tau_i}^{2}$, respectively. This study focuses on completing of the parameter estimation process after obtaining expected value inferences for τ. By obtaining the moments of τ terms and subsequently substituting them into the parameter estimations, we are able to conclude the parameter estimation process.

CHAPTER 5

APPLICATION

In this chapter, we first present a simulation study to evaluate our proposed model, SL-LMM, in terms of modeling and the predictive accuracy. Then, we revisit the schizophrenia data intending to detect the performance of the proposed model in real data applications. Although the theoretical aspect of the study has been formulated by assuming distinct n_i for each subject, as per the established literature, it is observed that equal n_i's can lead to a simplification of estimation and computation. Therefore, we take into account the models that are balanced in the application process to ensure the accuracy of our simulations and real data.

5.1 Simulation Study

Longitudinal data prediction is an indispensable tool for numerous practical applications and its significance cannot be overstated. An alternative way to measure the fitness of data is examining the accuracy of predictions for future observations, as pointed out by Rao [86] and Lee [87]. This study delves deeply into the subject and offers valuable insights, fostering a greater comprehension. We present a similar simulation study involved in the study of Ho and Lin [17] to assess the performance of the proposed SL-LMM approach. Furthermore, this simulation study involves a comparison of the predictive capabilities of different LMMs. The objective is to evaluate and assess the relative performance and effectiveness of these models in terms of their predictive accuracy.

In this study, 500 Monte Carlo (MC) data sets are generated from the model:

$$Y_{ij} = \beta_0 + t_{ij}\beta_1 + w_i\beta_2 + b_{0i} + t_{ij}b_{1i} + e_{ij}, \qquad (5.1)$$

$i = 1, \ldots, 100, \qquad j = 1, \ldots, 5,$

where $t_{ij} = j - 3$, $\beta_0 = 3$, $\beta_1 = 2$, $\beta_2 = 1$ and $b_i \sim MSL(0,\ 1,\ 5)$ and $e_{ij} \sim ML(0,\ 0.25)$ similar to Ho and Lin [17]. The binary indicator w_i is $w_i = 1$ if $i \leq 50$ and $w_i = 0$ if $i > 50$. As an illustrative example, the matrix representation for the 1st subject is as follows:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \end{bmatrix} = \underbrace{\begin{bmatrix} 3 \\ 3 \\ 3 \\ 3 \\ 3 \end{bmatrix}}_{\beta_0} + \underbrace{\begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}}_{t_{ij}} \underbrace{[2]}_{\beta_1} + \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{w_i} \underbrace{[1]}_{\beta_2} + b_{0i} + \underbrace{\begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}}_{t_{ij}} b_{1i} + e_{ij}.$$

R Core Team [88] is used to execute identical process across all 100 observations. Random effect terms b_i and error terms e_i are generated from multivariate skew Laplace (MSL) distribution and multivariate Laplace distribution respectively by using *LaplacesDemon* package. As evidenced by the graphical representation Figure 5.1, the generated subject-specific intercepts and subject-specific slopes exhibit a pronounced skewness.

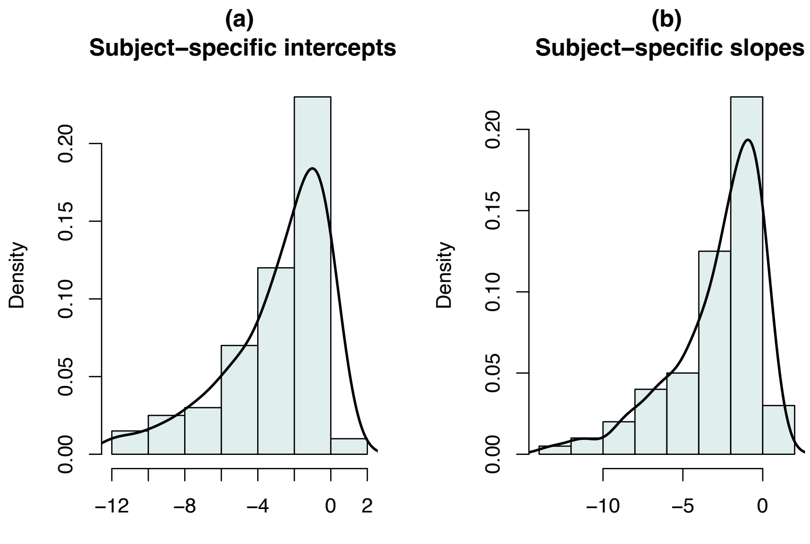

Figure 5.1. Histogram with kernel density plots of (a) subject-specific intercepts and (b) subject-specific slopes

Data generated from the model (5.1) display highly skewed and heavy-tailed random slopes based on this specification. The fundamental goal is to examine possible bias that can result from simulating true random effects from the MSL distribution and true error terms from multivariate Laplace distribution. In addition to this goal, we also check whether the fitted candidate models are adequate.

In order to compare different models, simulations are performed with a total of 500 MCMC replications under three estimation methods LMM (Model 1), Laplace (L)-LMM (Model 2), and Skew Laplace (SL)-LMM (Model 3). Also, for each simulation, models are fitted three times under the assumptions with the density of b_i represented by multivariate skew Laplace (Scenario-1), multivariate Laplace (Scenario-2), and multivariate normal (Scenario-3), respectively. The numerical results, containing the median and standard deviation of the iterated ML estimates of all unknown parameters for each scenario are given in Table 5.1, Table 5.2, and Table 5.3 respectively. To evaluate the precision of the parameter estimates $\hat{\theta}_k$ where k is the iteration number; $k \in \{1, \dots, 500\}$, mean absolute bias $\left(MAB = ave\left|\hat{\theta}_k - \theta_k\right|\right)$ and root mean square errors $\left(RMSE = \sqrt{ave\left(\hat{\theta}_k - \theta_k\right)^2}\right)$ are calculated. In addition, lower panel of Table 5.1 provides the count of preferred fits based on the AIC and BIC criteria, facilitating an objective evaluation of the model selection process.

According to Table 5.1, the proposed model SL-LMM gives better estimates in terms of exhibiting a higher degree of accuracy in capturing the actual parameter values when the skew Laplace distribution is adopted for the random effects. Also, we observe that SL-LMM gives much smaller MAB and RMSE values for β parameters and variance covariance of error terms than LMM and L-LMM, which increase modelling adequacy for the proposed model. Figure 5.2 and Figure 5.3 present the estimated beta parameters and variance covariance matrix for errors (σ^2) for Scenario-1, obtained through the application of three distinct statistical models, namely LMM (Model 1), L-LMM (Model 2) and SL-LMM (Model 3) respectively.

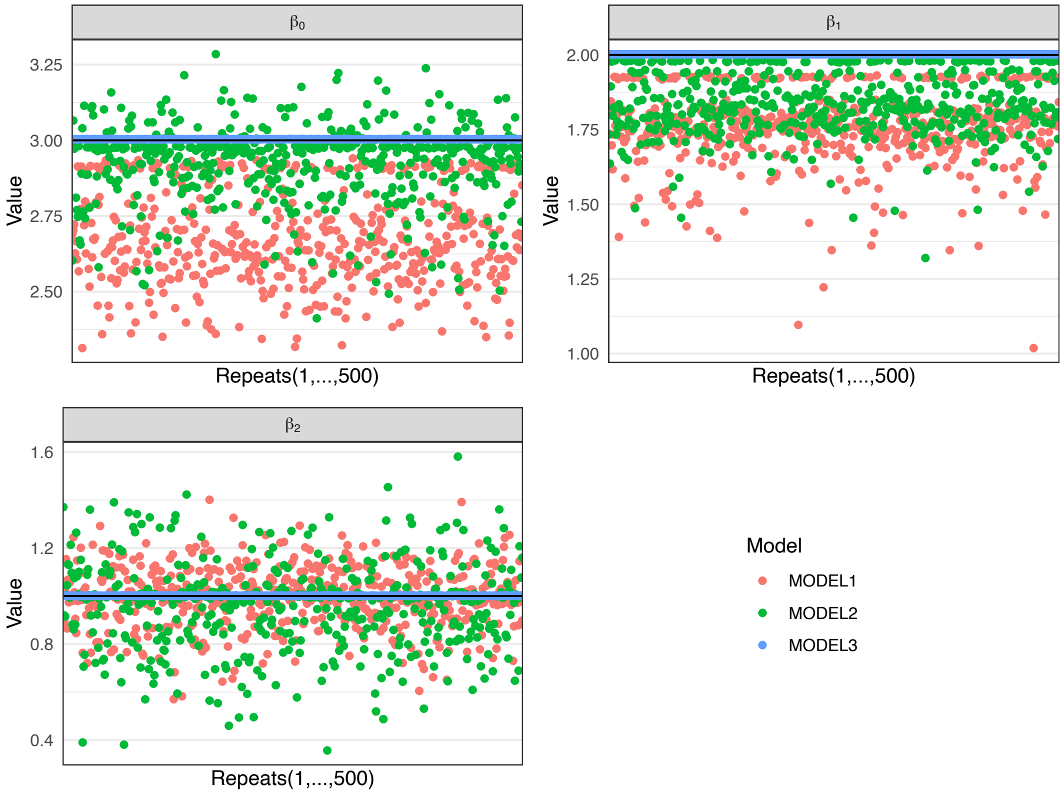

Figure 5.2. β Estimations $(n = 500)$ for Scenario-1 under LMM, L-LMM, and SL-LMM respectively

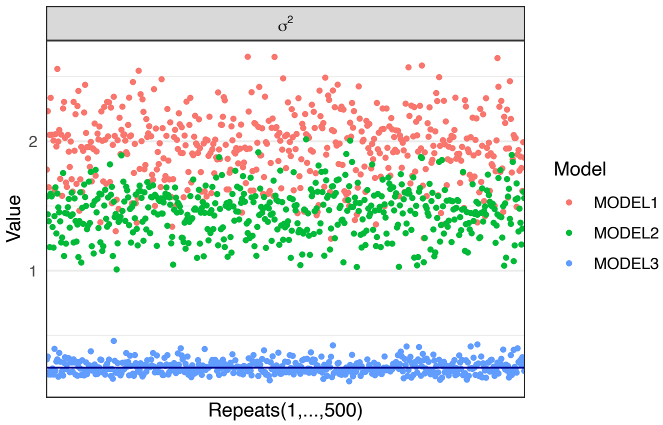

Figure 5.3. Estimation of σ^2 $(n = 500)$ for Scenario-1 under LMM, L-LMM, and SL-LMM respectively

As observed in both figures, SL-LMM demonstrates superior estimation capabilities for the respective parameters, since the parameter estimations for the proposed model is closer to the true values of the parameters. However, we cannot say the same thing for the estimation of variance covariance of random effect terms, D. As shown in the Table 5.1, the MAB and RMSE values of D_{11} and D_{22} for the model L-LMM is smaller than SL-LMM. Figure 5.4 depicts the estimation of D for Scenario-1 under LMM, L-LMM, and SL-LMM respectively.

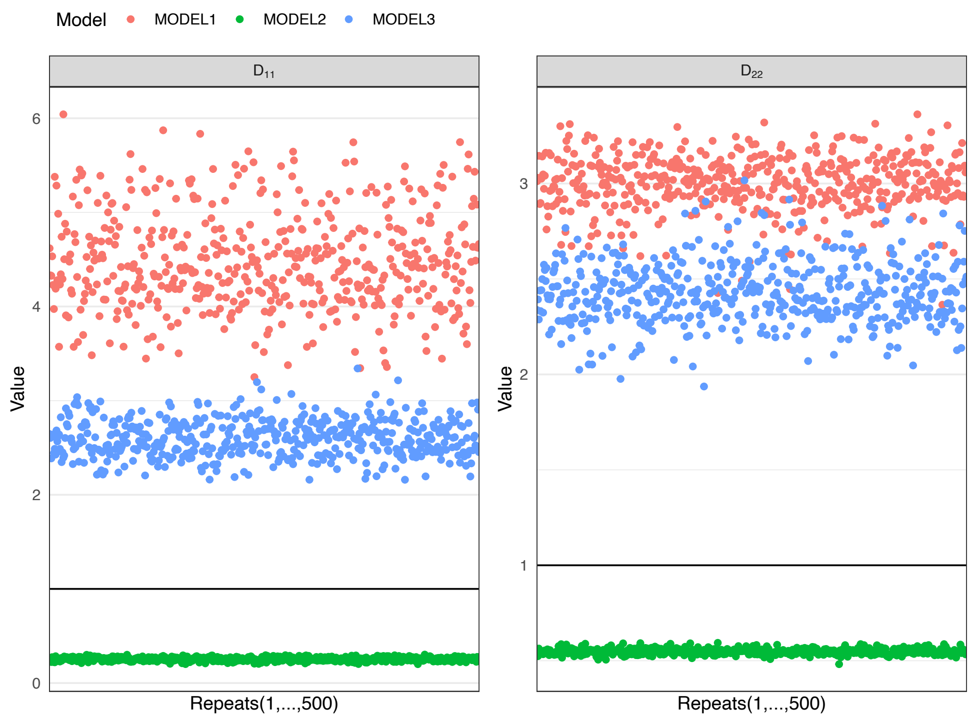

Figure 5.4. Estimation of D $(n = 500)$ for Scenario-1 under LMM, L-LMM, and SL-LMM respectively

It is worthy of mention that the fluctuation of skewness parameters which γ_1 and γ_2, are dominate comparing other parameters (Figure 5.5). At this point, please notice that LMM and L-LMM do not estimate the skewness parameters. The MAB values associated with the skewness parameter γ are smaller compared to the RMSE values. The result that 98% of BIC values align with the SL-LMM specification is expected, as the random effects were generated from MSL distribution.

Table 5.1 Parameter estimations and their MAB and RMSE values for LMM, L-LMM, and SL-LMM of the data generated data from Multivariate Skew Laplace Distribution — Scenario 1[a]

		LMM				L-LMM				SL-LMM			
	True	Est.	Sd.	MAB	RMSE	Est.	Sd.	MAB	RMSE	Est.	Sd.	MAB	RMSE
$\hat{\beta}_0$	3	2.6575	0.1549	0.3243	0.3593	2.9583	0.1304	0.1074	0.1514	3.0035	0.0005	0.0035	0.0035
$\hat{\beta}_1$	2	1.7611	0.1277	0.2468	0.2778	1.8293	0.1039	0.1642	0.1942	2.0020	0.0002	0.0035	0.0020
$\hat{\beta}_2$	1	0.9892	0.1293	0.1003	0.1292	0.9939	0.1767	0.1271	0.1812	1.0014	0.0007	0.0035	0.0016
$\hat{\sigma}^2$	0.25	1.9623	0.2646	0.7003	0.7207	1.4670	0.1841	1.2029	1.2169	0.2488	0.0522	0.0411	0.0525
$\hat{D}_{11}$	1	4.4193	0.5120	3.4790	3.5164	0.2531	0.0203	0.7461	0.7464	2.6033	0.1981	1.6032	1.6153
$\hat{D}_{22}$	1	3.0604	0.7171	3.0117	3.0310	0.5475	0.0154	0.4519	0.4522	2.4225	0.1604	1.4268	1.4358
$\hat{\gamma}_1$	5	–	–	–	–	–	–	–	–	4.9317	5.8660	2.7537	6.0299
$\hat{\gamma}_2$	5	–	–	–	–	–	–	–	–	4.7587	5.6233	2.6499	5.7318
Criterion											Frequency		
AIC		0				0				500			
BIC		0				10				490			

a Frequency represents the number of choices preferred by AIC and BIC.

Table 5.2 Parameter estimations and their MAB and RMSE values for LMM, L-LMM, and SL-LMM of the data generated data from Multivariate Laplace Distribution — Scenario 2

		LMM				L-LMM				SL-LMM			
	True	Est.	Sd.	MAB	RMSE	Est.	Sd.	MAB	RMSE	Est.	Sd.	MAB	RMSE
$\hat{\beta}_0$	3	2.9923	0.0264	0.0094	0.0279	2.9952	0.0032	0.0051	0.0059	2.9992	0.0003	0.0008	0.0009
$\hat{\beta}_1$	2	1.9990	0.0033	0.0014	0.0035	2.0014	0.0007	0.0014	0.0016	1.9995	0.0001	1.0005	0.0005
$\hat{\beta}_2$	1	1.0143	0.0384	0.0168	0.0418	1.0026	0.0047	0.0043	0.0054	0.9999	0.0004	2.0001	0.0004
$\hat{\sigma}^2$	0.25	1.2861	0.1386	1.0499	1.0590	1.7641	0.1484	1.5180	0.5252	0.2493	0.0517	0.0400	0.0522
$\widehat{D}_{11}$	1	1.3625	0.1410	0.3581	0.3843	0.1497	0.0100	0.8504	0.8504	1.0480	0.0892	0.0820	0.1009
$\widehat{D}_{22}$	1	1.2405	0.0952	0.2460	0.2637	0.1122	0.0048	0.8879	0.8879	1.0323	0.0786	0.0736	0.0932
$\hat{\gamma}_1$		—	—	—	—	—	—	—	—	0.0231	0.1485	4.9710	4.9733
$\hat{\gamma}_2$		—	—	—	—	—	—	—	—	-.0757	0.1707	5.0957	5.0985

Table 5.3 Parameter estimations and their MAB and RMSE values for LMM, L-LMM, and SL-LMM of the data generated data from Multivariate Normal Distribution — Scenario 3

		LMM				L-LMM				SL-LMM			
	True	Est.	Sd.	MAB	RMSE	Est.	Sd.	MAB	RMSE	Est.	Sd.	MAB	RMSE
$\hat{\beta}_0$	3	2.9984	0.0046	0.0038	0.0048	2.9967	0.0033	0.0038	0.0047	2.9995	0.0008	0.0006	0.0010
$\hat{\beta}_1$	2	2.0019	0.0011	0.0020	0.0022	2.0003	0.0007	0.0006	0.0008	1.9995	0.0002	1.0005	0.0005
$\hat{\beta}_2$	1	1.0010	0.0060	0.0049	0.0061	0.9945	0.0048	0.0061	0.0074	0.9999	0.0008	2.0001	0.0008
$\hat{\sigma}^2$	0.25	1.3068	0.1453	1.0624	1.0723	1.5261	0.1371	1.2856	1.2929	0.2544	0.0513	0.0403	0.0519
$\widehat{D}_{11}$	1	1.2819	0.1268	0.2895	0.3152	0.1734	0.0109	0.8264	0.8265	1.0493	0.0719	0.0697	0.0889
$\widehat{D}_{22}$	1	1.2153	0.0975	0.2182	0.2387	0.1273	0.0049	0.8726	0.8726	0.9971	0.0687	0.0549	0.0687
$\hat{\gamma}_1$		–	–	–	–	–	–	–	–	0.3065	0.3006	4.6102	4.6200
$\hat{\gamma}_2$		–	–	–	–	–	–	–	–	-.0441	0.1075	5.0555	5.0567

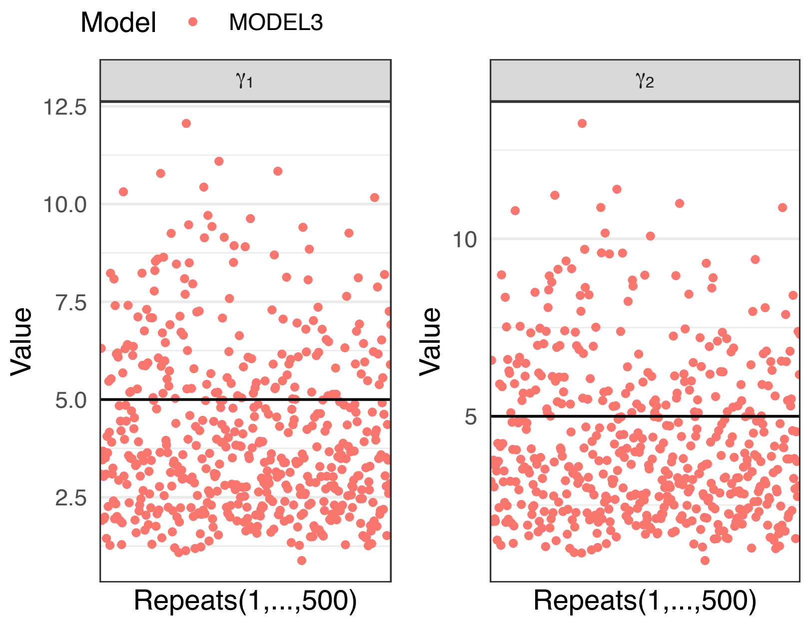

Figure 5.5. Estimation of γ_1 and γ_2 for Scenario-1 under SL-LMM

We observed from Table 5.2 and Table 5.3 that proposed model SL-LMM provides good estimates for all parameters, but it exhibits suboptimal estimates for the skewness parameter when the data is generated under Scenario 2 and 3. This inadequacy in skewness parameter estimation is an aforementioned result due to the utilization of a symmetric distribution for random slopes. The estimation of γ_1 and γ_2 from various multivariate distributions, namely skew Laplace (Scenario-1), Laplace (Scenario-2), and normal (Scenario-3), is depicted in the Figure 5.6. In this plot, the estimation of the skewness parameter around zero in models generated for a data set with a non-skewed distribution emphasizes the accuracy of the model.

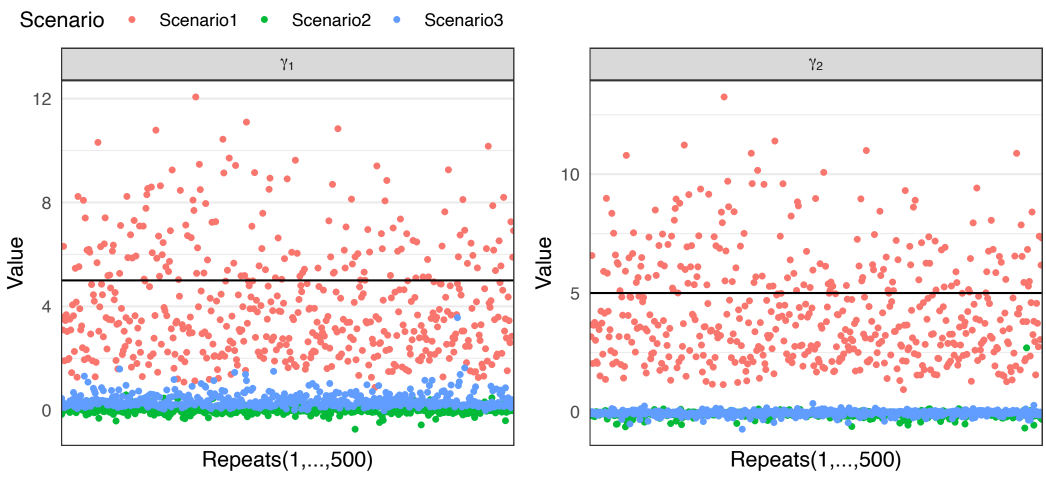

Figure 5.6. Estimation of γ_1 and γ_2 from different multivariate distributions under SL-LMM

5.2 Application: The Schizophrenia Data

5.2.1 Background and Data Summary

Schizophrenia is a severe and complex psychiatric disorder that exhibits the presence of delusions, hallucinations, suffering delusions, and occasionally disorganized behaviour and speech. To provide a motivating example for our model, a double-blinded clinical trial presented by Lapierre et al. [89], comparing the effects of three doses (low, medium, high) of a new therapy (NT) against a standard therapy (ST) for 245 patients with acute schizophrenia is proposed. The schizophrenia level of the patients is evaluated using the Brief Psychiatric Rating Scale (BPRS) at baseline (week 0) and after one, two, three, four, and six weeks of treatment. BPRS scores range from 0 to 108 and indicate the severity of a patient's symptoms. Many patients withdraw from the trial for various reasons, including those directly related to drug efficacy. The preliminary studies demonstrate that the experimental drug is effective as ST regarding antipsychotic activity and has fewer side effects (Cnaan et al. [90]). The schizophrenia data is reported by Ho and Lin [17] and loaded as supplementary material attached to their article. One can read Lapierre et al. [89], Cnaan et al. [90] and Hogan and Laird [91] for a more detailed explanation of the data.

Specifically, we highlight the comparison between 35 patients on high-dose NT and 38 patients on ST, with all having BPRS scores each week. The main objective of the study is to compare the change from baseline to week 6 in new therapy with high dose and standard therapy. The general structure of the data, and time versus BPRS score for each subject are shown in Table 5.4. Figure 5.7 represents the trajectories of individual BPRS scores changed over six visits, allied with the mean profiles for two treatment arms. It appears that several of these 73 patients have sudden jumps and drops over time in BPRS scores. For simplicity of illustration, the plot of time versus BPRS score for randomly selected 28 subjects is provided in (Figure 5.8). We mainly observe variation within individuals and between individuals. To gain deeper insights into these variations, we fit a series of individual-specific linear regression models with time variable versus BPRS scores.

Table 5.4 Schizophrenia data

		Week					
Treatment	**Subject**	**0**	**1**	**2**	**3**	**4**	**6**
NT (high dose)	Subject 1	26	22	24	21	19	17
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	Subject 35	38	44	34	34	35	38
ST	Subject 36	54	28	14	12	9	13
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	Subject 73	24	26	16	18	17	15

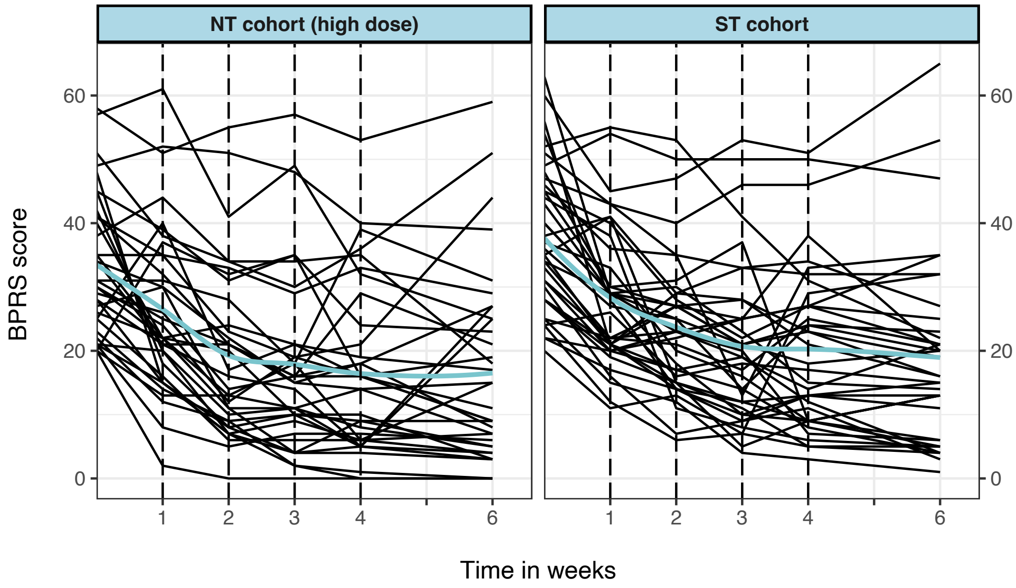

Figure 5.7. Trajectories of schizophrenia levels for 73 participants. The thicker solid line indicates the mean profile in the study

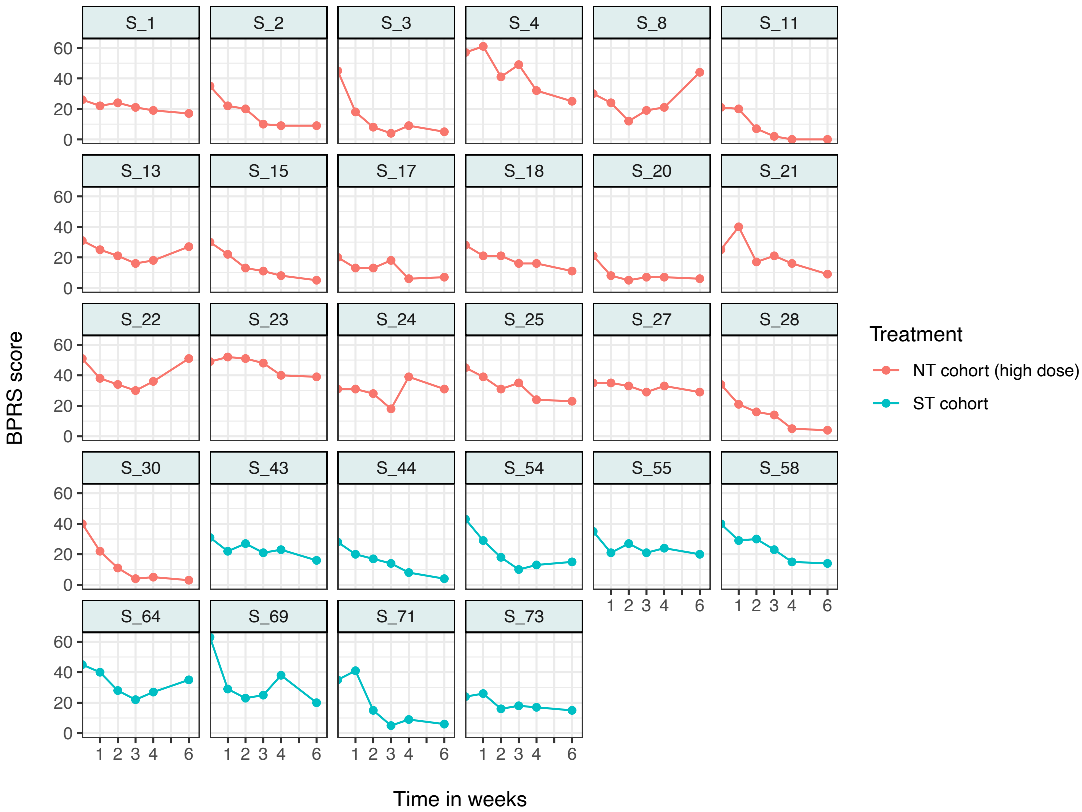

Figure 5.8. Time versus BPRS score for randomly selected 28 observations

It is obvious that BPRS score shows a non-linear decline with time for each subject taking ST and NT, but this decrease is not identical for each subject. For instance, the regression line for Subject 22 is flatter (Figure 5.9). Besides, baseline measurement (week 0) values of BPRS score differs significantly for each subject. For example, while subjects Subject 4 and Subject 23 started with the largest value of BPRS score in the baseline, Subject 11 and Subject 13 started with a quite low score compared to others. All these inferences provide an indication to fit the model on behalf of the observation with random intercept and slope. Undoubtedly, fitting mixed effects model including both fixed and random effects to the data renders the model more accurate and consistent.

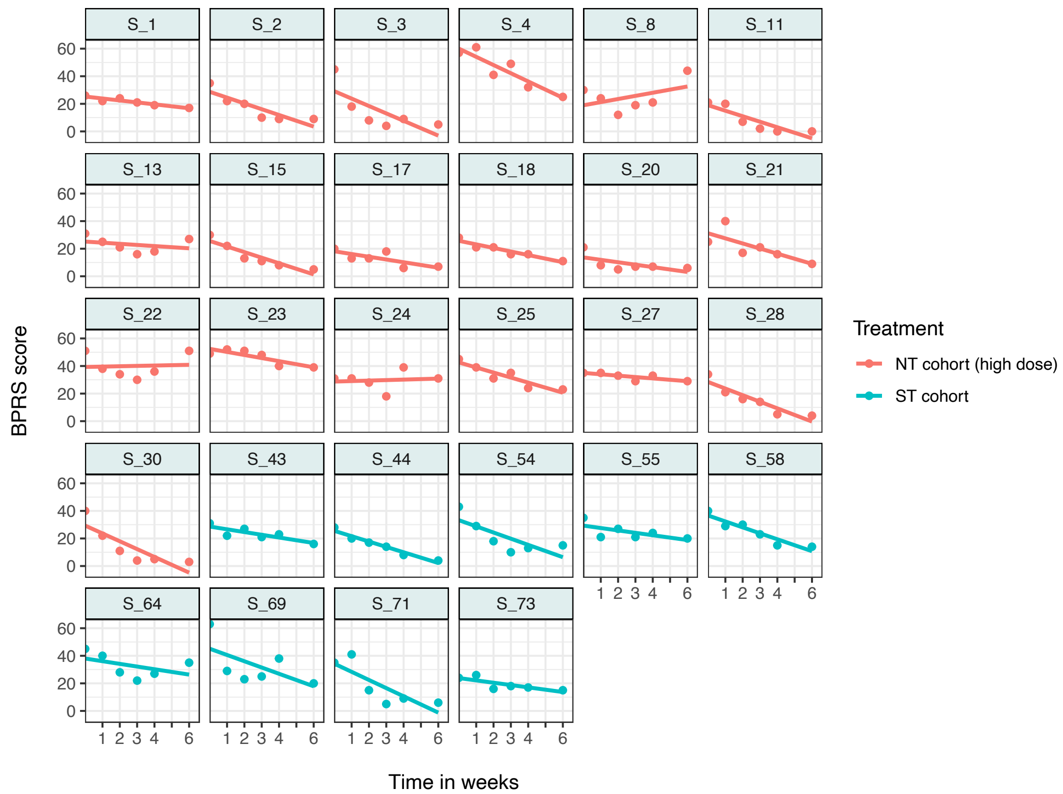

Figure 5.9. Fitted regression lines for each subject with time versus BPRS score

As a result of conducting a linear regression for each subject, we acquire as number of intercept and slope for each subject, exactly 73 via *lm.list* function in R. By performing a linear regression to obtain the fitted parameters for the entire data set, this would be equivalent to the mean of those parameters. The graph (Figure 5.10) regarding the confidence intervals of the fitted parameters emphasizes the requirement of including random effects in the model. In addition, standardized residuals versus fitted values for each treatment group (Figure 5.11) shows that there are discernible outliers in the model and the prediction is not good. Finally, boxplots of the errors for each observation (Figure 5.12) obtained from these series of linear regression underlines the differences between the subjects.

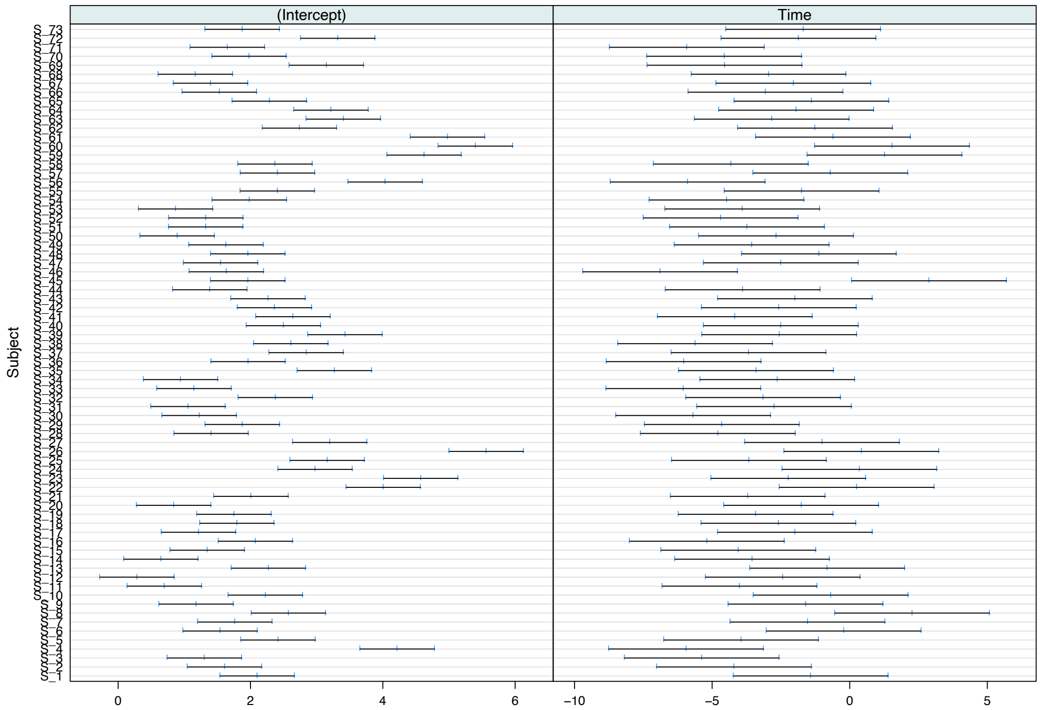

Figure 5.10. Confidence intervals of $\hat{\beta}_0$ and $\hat{\beta}_1$ for each subject

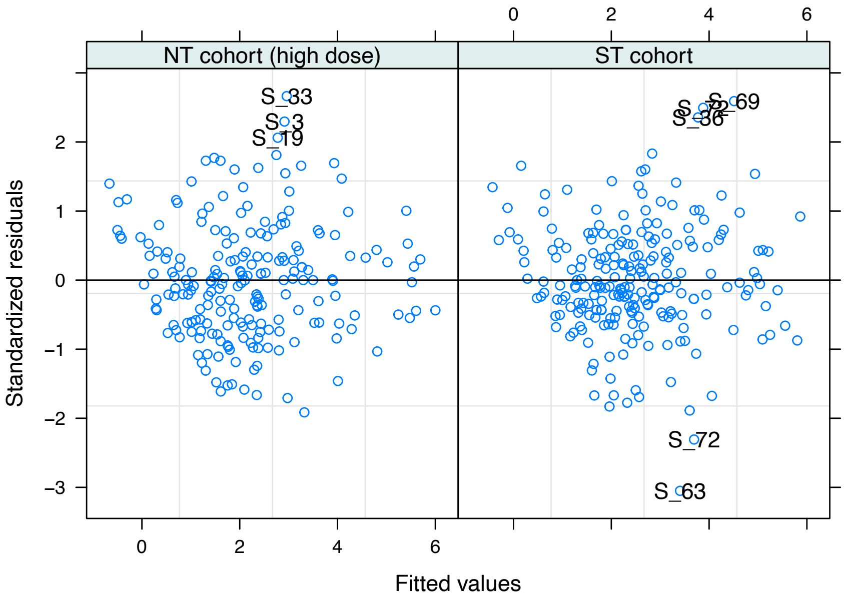

Figure 5.11. Standardized residuals versus predictions for each treatment group

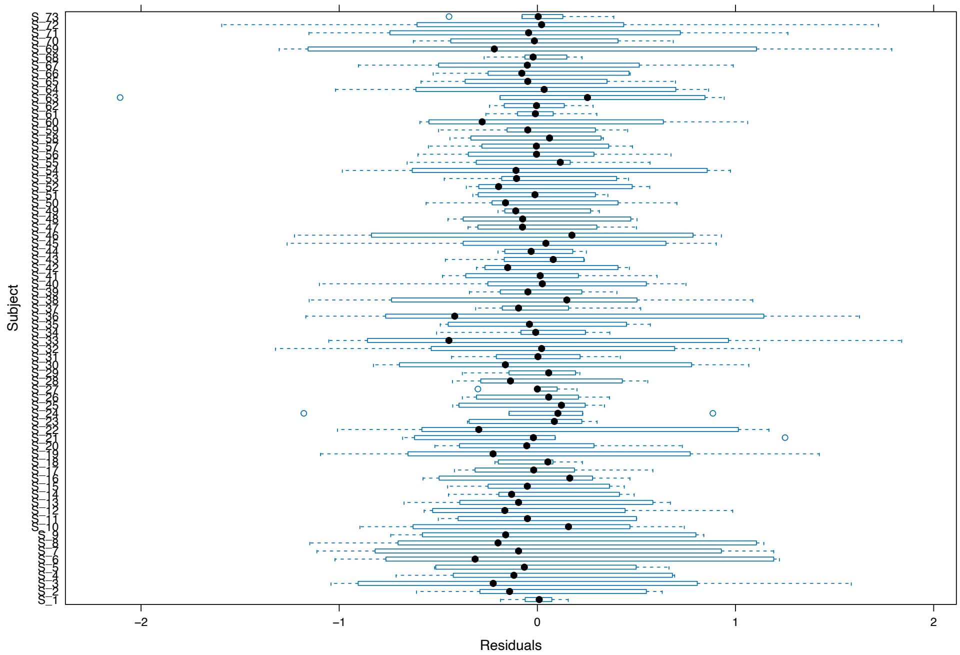

Figure 5.12. Box plot of the errors obtained from series of linear regression for each observation

Meanwhile, in addition to some within and between variations, we mainly observe a non-linear decline in the patterns of population mean profiles. In view of the large initial decreases in the BPRS scores and then a levelling-off, the quadratic trend model would provide an appropriate way to approximate the mean trajectory for the BPRS scores.

We adopt the same linear mixed model based on a curvilinear trend for the population average as in the study of Ho and Lin [17]:

$$y_{ij} = \beta_0 + \beta_1 t_j + \beta_2 {t_j}^2 + \beta_3 NT_i + b_{0i} + b_{1i} t_j + e_{ij},$$
$$i = 1, \dots, 73, \qquad j = 1, \dots, 6, \tag{5.2}$$

where y_{ij}, for consistent numeric outcomes, indicates the BPRS score divided by 10 at the j-th time point for the i-th subject, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$ is the fixed effects of explanatory variables; $\boldsymbol{NT_i}$ is an indicator variable of NT for subject i, i.e. individual level-covariate; t_j is a covariate represents the time being measured in a week from the baseline and taken as $\frac{(\mathrm{time}-3)}{10}$; $\boldsymbol{b_i} = (b_{0i}, b_{1i})'$ is the subject-specific intercepts and slopes for the random effects for the i-th subject; and e_{ij} is the within-subject error. β_3, in the model (5.2), denotes the average of improvement in the weekly BPRS scores of NT patients.

Based on this information about the model, we first fit classical LMM (Laird and Ware [5]) to the schizophrenia data. The Figure 5.13 shows the deviation of the random intercept and slope from the overall model estimate, which implies the variation among random effects. This graph once again underlines the necessity of using mixed effects model for the schizophrenia data.

In order to verify whether there is a tendency for accommodating skewness in the random effects, we examine estimates of subject-specific random effects derived from fitting LMM to the schizophrenia data. Figure 5.14 depicts histograms and corresponding normal quantile plots of the estimates of the random effect term, $\boldsymbol{b_i}$.

The plots display that there are apparent non-normal patterns for both subject-specific intercept and slope.

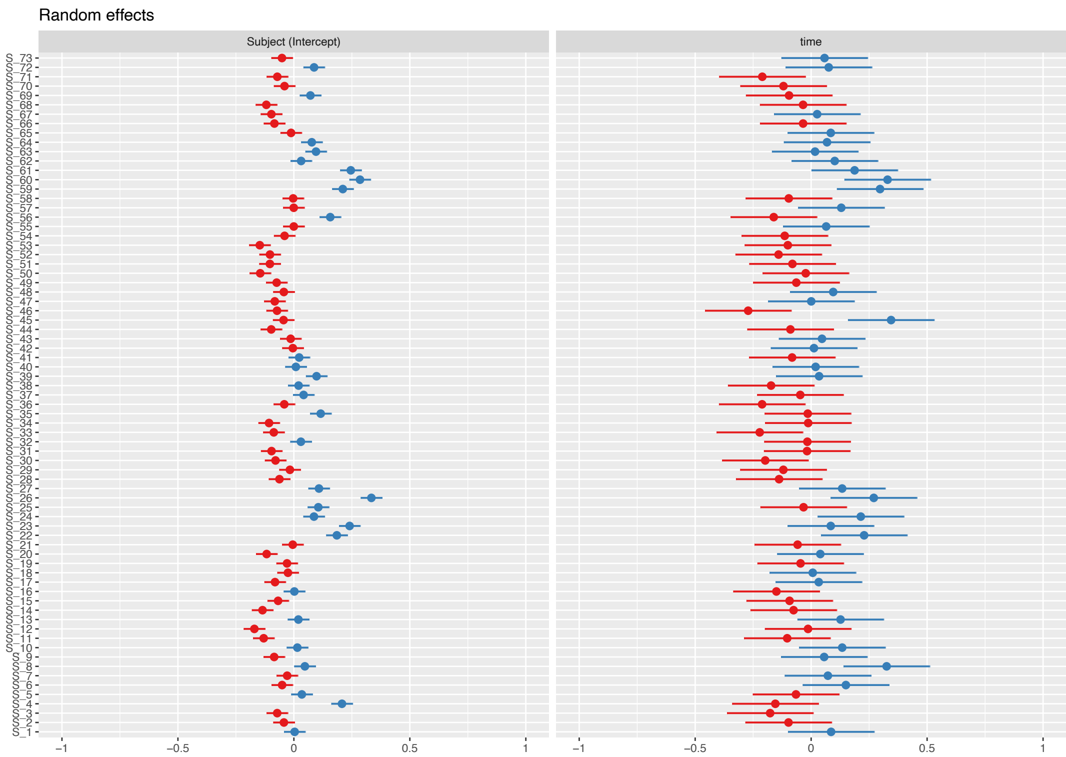

Figure 5.13. Deviation of the random intercept and slope from the LMM model estimate (red lines show the negative deviations from the fixed parameters and blue lines show the positive deviations from the fixed parameters)

Specifically, subject-specific intercept yields a highly positively skewed distribution as suggested by the left panel of Figure 5.14. In support of this, Ho and Lin [16], [17] and Schumacher et al. [16] reveal that both subject-specific intercepts and slopes are positively skewed and highlight necessity of a robust model to accommodate the skewness of random effect. Besides, Figure 5.14 also provides evidence that there are potential outliers at the level of the random effects, namely b-outliers. In order to determine whether e-outliers, i.e., outliers at the level of the within-subjects exist, we refer to Figure 5.15. The residual plots shown in the upper panels of Fig. 5.15 demonstrate that e-outliers exist.

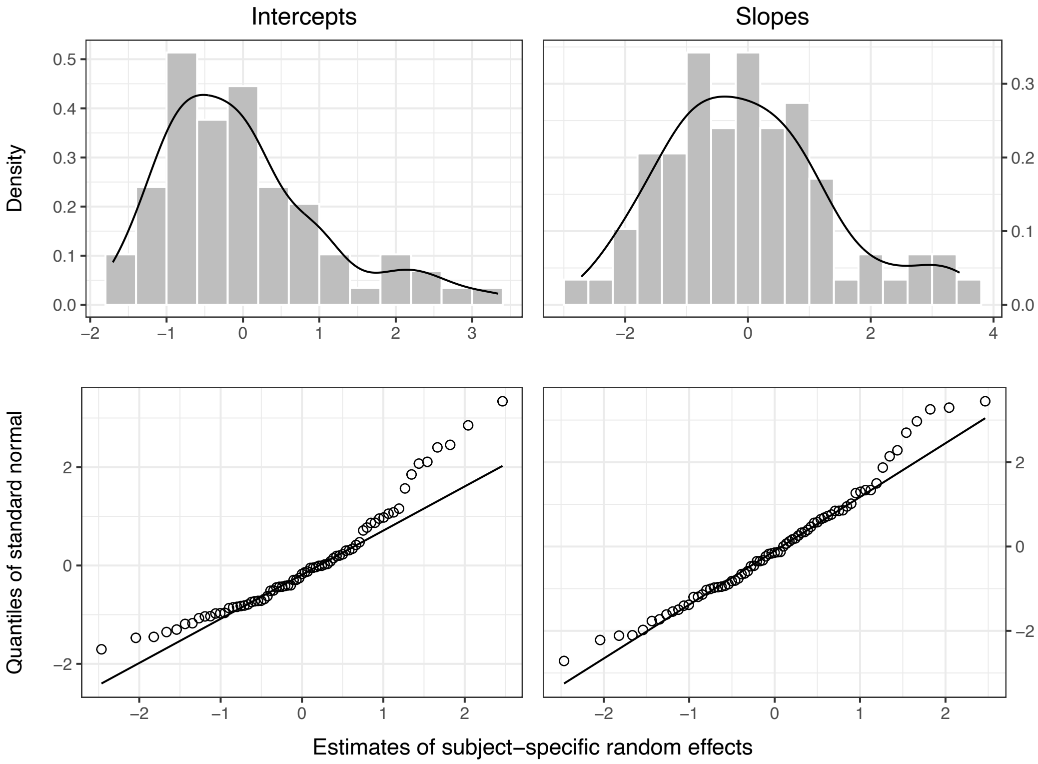

Figure 5.14. Histograms and corresponding normal quantile plots of the estimates of subject-specific random effects derived from fitting LMM to the schizophrenia data

According to the normal quantile–quantile plots shown in the lower panels, the residuals display heavy tails but appear to be quite symmetric. In the light of these observations, we intend to construct SL-LMM model with multivariate skew Laplace distributed random effects and multivariate Laplace distributed within-subject errors. In the following section, we demonstrate the results of fitting SL-LMM model to schizophrenia data with comparative results of other models.

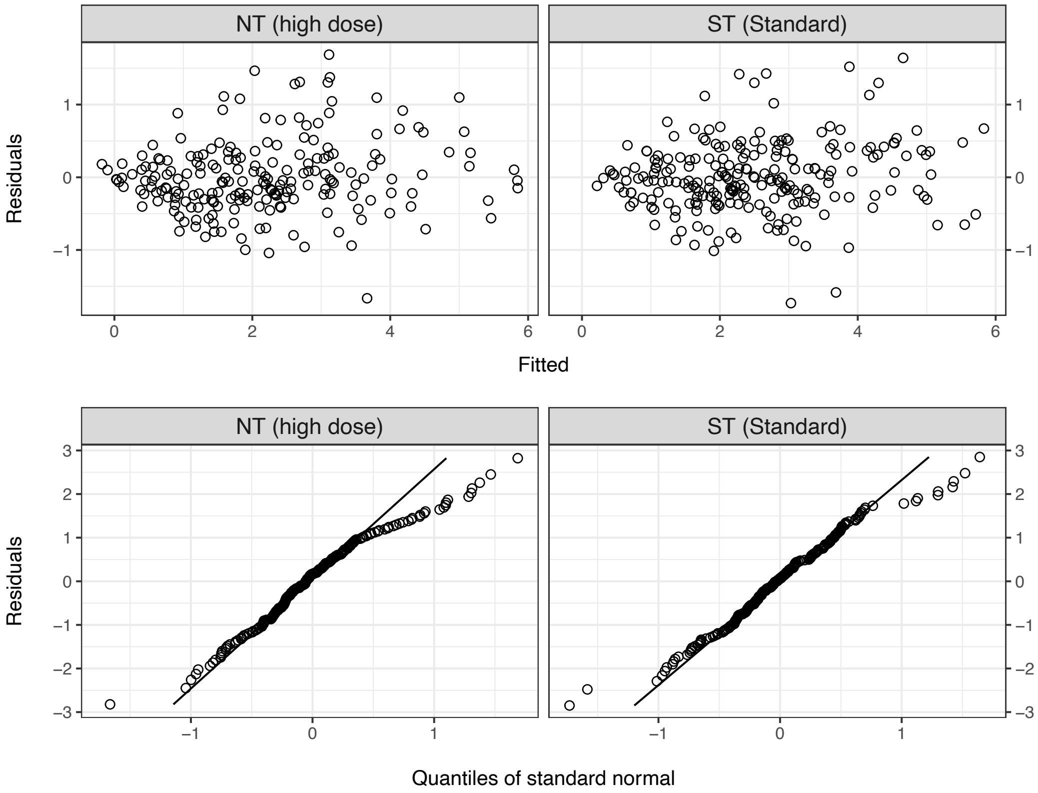

Figure 5.15. Residuals versus fitted values (upper panel) and the normal quantile plots corresponding to residuals (lower panel)

5.3 Real Data Application: The schizophrenia data

Derived from the preliminary analysis in the previous section, we are inspired to advocate the implementation of SL-LMM as a proposed model to analyse the schizophrenia data set. We start by modifying the model (5.2) within the context of SL-LMM, that is, the random effects $\boldsymbol{b_i} = (b_{0i}, b_{1i})'$ and error terms $\boldsymbol{e_i} = (e_{i1}, \dots, e_{in_i})'$ jointly distributed as,

$$\begin{bmatrix} \boldsymbol{b_i} \\ \boldsymbol{e_i} \end{bmatrix} \sim SL_{2+n_i}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} D & 0 \\ 0 & \sigma^2 \end{bmatrix}, \begin{bmatrix} \gamma \\ 0 \end{bmatrix}\right), \quad i = 1, \dots, 73.$$

We then compare the ML results under the SL-LMM with those acquired under the LMM, and L-LMM. Table 5.5 summarizes the results from ML estimations of the

SL-LMM, L-LMM, and LMM. In addition, maximized log-likelihood values $\ell(\widehat{\boldsymbol{\theta}})$, and two penalized likelihood information criteria, $\mathrm{AIC} = -2\ell(\widehat{\boldsymbol{\theta}}) + 2m$ and $\mathrm{BIC} = -2\ell(\widehat{\boldsymbol{\theta}}) + mlog(N)$, where m indicates the number of parameters and $N = \sum_{i=1}^{n} n_i$ denotes the total number of observations, are also presented in Table 5.5. While fitting the models by using the ML method, we employ the ECM algorithm discussed in Section 4 under different starting values. We observe a decline in BPRS scores over time for all fitted models due to the significant negative time effect. In order to compare the models, we computed the AIC and BIC for each model. The model with the smallest AIC or BIC value is taken as the best fit to the data. Based on these two criteria, SL-LMM with the lowest value for both criteria are selected as the best fit, followed by L-LMM and LMM.

Figure 5.16 delineates a scatter plot of estimated random effects overlapped on contour lines acquired from the marginalization of the fitted SL_2 density, accompanied with corresponding histograms of the marginal densities. Based on the plot, asymmetry and thick tails among the estimated random effects are remarkable. Furthermore, the contour level curves reflect the adequacy of using a bivariate skew Laplace distribution for the random effects since they are well-adapted to the shape of the scattering pattern.

Since simulations studies indicates that both MAP and RMSE of the skewness parameter is high, we employ a likelihood ratio test for skewness parameter as in the study of Schumacher et al. [16]. Likelihood ratio test is performed in order to test whether a restricted model represents the data well enough or not. Particularly, we test the hypothesis whether a skew model is unnecessary, which could be expressed $\mathrm{H}_0{:}\,\gamma = 0$ versus necessary expressed with $\mathrm{H}_1{:}\,\gamma \neq 0$. Let $\mathrm{H}_0{:}\,\gamma = 0$, $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_r)'$ is the hypothesis of interest, Θ is the k-dimensional parameter space of the unrestricted model, and Θ_0 is the parameter space under H_0, for $1 \leq r < k$. In order to understand the influence of H_0 in the maximum of the likelihood function, calculate the statistic $\Delta_n = 2\left(l(\hat{\theta}) - l(\hat{\theta}_0)\right)$, where $\hat{\theta}_0$ is the ML estimate of θ under the restriction in H_0. Afterwards, under H_0, Δ_n is asymptotic distributed as a chi-square random variable with k degrees of freedom $\left({\chi^2}_r \right)$ (Mood [92]).

Table 5.5 ML results from fitting SL-LMM and competitive models to the schizophrenia data

	Estimates		
Parameters	LMM	L-LMM	SL-LMM
$\hat{\beta}_0$	2.0626	1.7159	2.1925
$\hat{\beta}_1$	-2.7816	-2.8795	-2.7881
$\hat{\beta}_2$	8.3039	8.1291	7.3448
$\hat{\beta}_3$	-0.3139	0.1634	-0.5042
$\hat{\sigma}^2$	1.3832	1.3700	0.2258
$\hat{D}_{11}$	1.2753	0.2877	0.6049
$\hat{D}_{12}$	0.8536	0.1793	0.3054
$\hat{D}_{22}$	1.2785	1.4080	1.8073
$\hat{\gamma}_1$	−	−	0.1984
$\hat{\gamma}_2$	−	−	1.2110
m	8	8	10
$l(\hat{\theta})$	-727.978	-688.239	-610.998
AIC	1471.956	1392.478	1241.997
BIC	1504.614	1425.136	1264.901

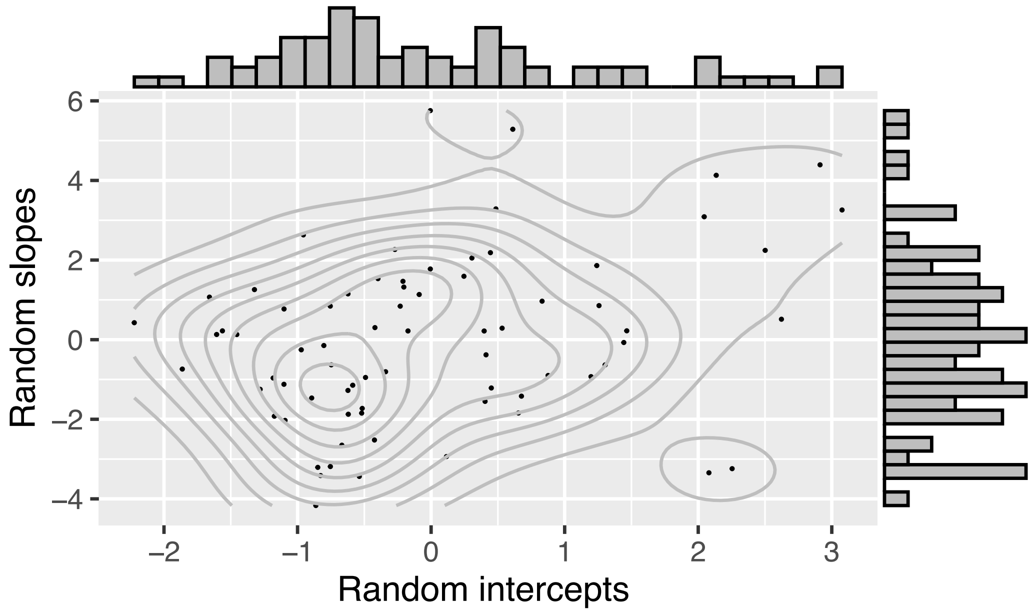

Figure 5.16. Scatter plot of estimated random effects overlapped on contour lines acquired from the marginalization of the fitted SL_2 density, accompanied with corresponding histograms of the marginal densities

As described above, the asymptotic distribution of Δ_n is χ^2 due to two restrictions under H_0. We calculate $\Delta_n = 28.761$, so the p-value of the likelihood ratio test is $P(\chi^2 > 28.761) \approx 0$. As per the previous studies, the schizophrenia data requires the skew model for modeling. Figure 5.17 depicts the fitted BPRS scores obtained from proposed model (SL-LMM) versus real valued BPRS score for each subject. As can be seen, the fitted values obtained from the proposed model demonstrate a remarkable level of concordance with the actual observed values. The figure visually shows that the fitted line resembles a diagonal line, indicating a good fit between the proposed model and the observed data. The model's success in capturing the observed variability is further supported by low AIC and BIC values.

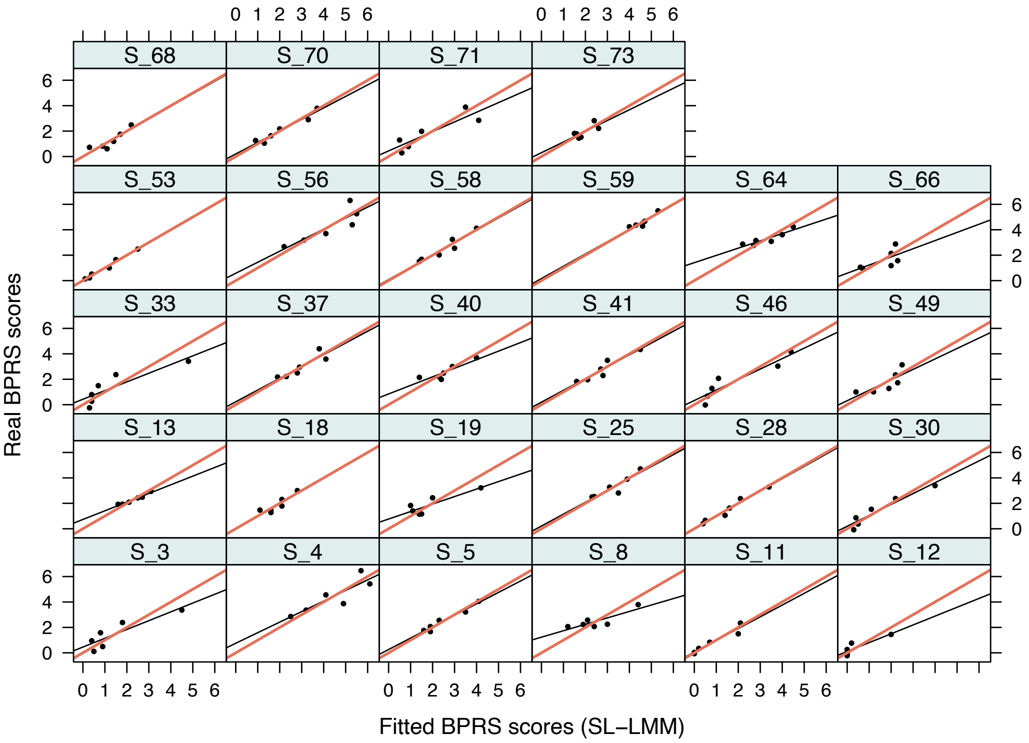

Figure 5.17. Fitted BPRS scores obtained from proposed method versus real valued BPRS scores

The SL-LMM application for practitioners highlights that the proposed model is particularly advantageous when traditional LMM approaches are not appropriate due to the presence of outliers and non-normality in the data. Its ability to handle heavy-tailed and skewed data makes it a valuable tool for researchers and practitioners seeking to analyse complex datasets in various fields. This model uses the Metropolis algorithm within the ECM algorithm to estimate the underlying parameters. Once the necessary data has been uploaded into the R environment, the next step is to execute the R codes associated with the proposed model. If you require the codes, you can request them from the author. Due to the Metropolis algorithm, the computational time required to analyse the data using this model takes approximately 6 minutes on an Apple M2 chip and 8 GB memory.

CHAPTER 6

CONCLUSION AND DISCUSSION

In this study, we extend LMM with a skew and robust distribution to effectively account for the inherent within-subject correlation that may arise in longitudinal data obtained by repeated measurements of the same subject over time. Classical LMMs are based on the fundamental assumption that both the random effects and errors are normally distributed. However, the normality assumption may not be always validated when dealing with outliers, heavy-tailed, and/or skewed data. To overcome the limitations of the normality assumption, alternative distributions with greater flexibility have been utilized for parameter estimation in the relevant literature of LMM. A robust LMM with multivariate Laplace distribution is proposed by Yavuz and Arslan [10] and the model definitions and parameter estimations are implemented under this assumption. We propose a skew generalization of Laplace LMM, called SL-LMM, as a powerful tool to handle longitudinal data with asymmetry in this study. This is the first attempt in literature using skew Laplace distribution for random effects and Laplace distribution for error terms in LMMs.

Skew Laplace distribution is expressed in a normal mean-variance mixture form, which provides a distinct and valuable option for representing skewness parameter as fixed but unknown. Expressing the skew Laplace distribution as a NMVM distribution and expressing the error term of the proposed model as scale mixture of normal distribution enable us to define the proposed model in a hierarchical form. The parameter estimations of the SL-LMM are implemented with the EM-type algorithm. To conclude the parameter estimation procedure, a special case of MCMC method, namely the Metropolis algorithm, is applied to find moments of the latent variable. As a result, both EM and Bayesian methods are utilized in the parameter estimation procedure.

Simulation studies are implemented with three different scenarios under three estimation methods LMM, Laplace (L)-LMM, and Skew Laplace (SL)-LMM. For each scenario, models are fitted three times under the assumptions that random effect terms are generated from MSL, multivariate Laplace, and multivariate normal, respectively. The proposed model, SL-LMM, yields improved estimations for accurately capturing the actual parameter values when data is generated from a multivariate skew Laplace distribution. The simulation study demonstrates that appropriate specification of the model is important for different scenarios. Numerical results shown in Section 5 indicate that the SL-LMM model for the schizophrenia data is evidently more adequate than the LMM and L-LMM. We generate new R codes in SL-LMM for simulations and data analysis for the proposed and other models used for comparisons.

There are several potential avenues for further research in this area, including the following possibilities. A worthwhile task is to develop SL-LMM with serially correlated errors due to the repeated measures of each subject collected over time (Schumacher et al. [16]). Another promising avenue for future research is variable selection in SL-LMM. Finally, since BIC for LMMs is questionable in the literature (Shen and González [93], Lai and Gao [94]) a modification or adaptation to BIC might be studied for the model selection among skew LMMs.

REFERENCES

[1] C. Eisenhart, "The Assumptions Underlying the Analysis of Variance," *Biometrics*, vol. 3, pp. 1–21, 1947, doi: 10.2307/3001534.

[2] C. R. Henderson, "Estimation of Genetic Parameters," *Ann. Math. Stat.*, vol. 21, pp. 309–310, 1950.

[3] S. R. Searle, *Linear Models*. New York: John Wiley & Sons, 1971.

[4] D. A. Harville, "Maximum likelihood approaches to variance component estimation and to related problems," *J Am Stat Assoc*, vol. 72, no. 358, pp. 320–338, 1977, doi: 10.1080/01621459.1977.10480998.

[5] N. M. Laird and J. H. Ware, "Random-Effects Models for Longitudinal Data," *Biometrics*, vol. 38, pp. 963–974, 1982, [Online]. Available: https://www.jstor.org/stable/2529876

[6] G. Verbeke and E. Lesaffre, "A Linear Mixed-Effects Model With Heterogeneity in the Random-Effects Population," *J Am Stat Assoc*, vol. 91, no. 433, pp. 217–221, 1996, doi: 10.2307/2291398.

[7] J. C. Pinheiro, C. Liu, and Y. Nianwu, "Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution," *Journal of Computational and Graphical Statistics*, vol. 10, no. 2, pp. 249–276, 2001, doi: 10.1198/10618600152628059.

[8] T.-I. Lin, "Longitudinal Data Analysis Using t Linear Mixed Models with Autoregressive Dependence Structures," *Journal of Data Science*, vol. 6, pp. 333–355, 2008, [Online]. Available: http://cran.r-project.org/

[9] R. F. Manghi, G. A. Paula, and F. J. A. Cysneiros, "On elliptical multilevel models," *J Appl Stat*, vol. 43, no. 12, pp. 2150–2171, Sep. 2016, doi: 10.1080/02664763.2015.1134445.

[10] F. G. Yavuz and O. Arslan, "Linear mixed model with Laplace distribution (LLMM)," *Statistical Papers*, vol. 59, no. 1, pp. 271–289, 2018, doi: 10.1007/s00362-016-0763-x.

[11] R. B. Arellano-Valle, H. Bolfarine, and V. H. Lachos, "Skew-normal Linear Mixed Models," *Journal of Data Science*, vol. 3, pp. 415–438, 2005.

[12] A. Azzalini and A. Dalla Valle, "The multivariate skew-normal distribution," *Biometrika*, vol. 83, no. 4, pp. 715–726, 1996, [Online]. Available: https://academic.oup.com/biomet/article-abstract/83/4/715/253295

[13] T. I. Lin and J. C. Lee, "Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data," *Stat Med*, vol. 27, no. 9, pp. 1490–1507, 2008, doi: 10.1002/sim.3026.

[14] V. H. Lachos, P. Ghosh, and R. B. Arellano-Valle, "Likelihood Based Inference for Skew-normal Independent Linear Mixed Models," *Stat Sin*, vol. 20, no. 1, pp. 303–322, 2010.

[15] V. H. Lachos, H. Bolfarine, R. B. Arellano-Valle, and L. C. Montenegro, "Likelihood-Based Inference for Multivariate Skew-Normal Regression Models," *Commun Stat Theory Methods*, vol. 36, no. 9, pp. 1769–1786, Jul. 2007, doi: 10.1080/03610920601126241.

[16] F. L. Schumacher, V. H. Lachos, and L. A. Matos, "Scale mixture of skew-normal linear mixed models with within-subject serial dependence," *Stat Med*, vol. 40, no. 7, pp. 1790–1810, Mar. 2021, doi: https://doi.org/10.1002/sim.8870.

[17] H. J. Ho and T. I. Lin, "Robust linear mixed models using the skew t distribution with application to schizophrenia data," *Biometrical Journal*, vol. 52, no. 4, pp. 449–469, 2010, doi: 10.1002/bimj.200900184.

[18] P. K. Choudhary, D. Sengupta, and P. Cassey, "A general skew-t mixed model that allows different degrees of freedom for random effects and error

distributions," *J Stat Plan Inference*, vol. 147, pp. 235–247, 2014, doi: https://doi.org/10.1016/j.jspi.2013.11.015.

[19] R. Mohammadi and I. Kazemi, "A robust linear mixed-effects model for longitudinal data using an innovative multivariate skew-Huber distribution," *J Multivar Anal*, vol. 187, Jan. 2022, doi: 10.1016/j.jmva.2021.104856.

[20] O. Arslan, "An alternative multivariate skew Laplace distribution: Properties and estimation," *Statistical Papers*, vol. 51, no. 4, pp. 865–887, 2010, doi: 10.1007/s00362-008-0183-7.

[21] T. J. Kozubowski and K. Podgórski, "A multivariate and asymmetric generalization of Laplace distribution," *Comput Stat*, vol. 15, no. 4, pp. 531–540, 2000, doi: 10.1007/pl00022717.

[22] T. Kozubowski and K. Podgorski, "Asymmetric Laplace Distributions," *The Mathematical Scientist*, vol. 25, pp. 37–46, Jan. 2000, doi: 10.1007/978-1-4612-0173-1_3.

[23] Kotz S, Kozubowski TJ, and Podgórski K, *The Laplace distribution, and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Boston: Birkhäuser, 2001.

[24] Kotz S, Kozubowski TJ, and Podgórski K, "An asymmetric multivariate Laplace Distribution," 2003.

[25] E. Gómez, M. A. Gomez-Viilegas, and J. M. Marín, "A multivariate generalization of the power exponential family of distributions," *Commun Stat Theory Methods*, vol. 27, no. 3, pp. 589–600, Jan. 1998, doi: 10.1080/03610929808832115.

[26] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977, [Online]. Available: http://www.jstor.org/stable/2984875

[27] N. Laird, N. Lange, and D. Stram, "Maximum likelihood computations with repeated measures: Application of the EM algorithm," *J Am Stat Assoc*, vol. 82, no. 397, pp. 97–105, 1987, doi: 10.1080/01621459.1987.10478395.

[28] M. J. Lindstrom and D. M. Bates, "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data," *J Am Stat Assoc*, vol. 83, no. 404, pp. 1014–1022, 1988, doi: 10.2307/2290128.

[29] X.-L. Meng and D. B. Rubin, "Maximum Likelihood Estimation via the ECM Algorithm: A General Framework," *Biometrika*, vol. 80, no. 2, pp. 267–278, 1993, doi: 10.2307/2337198.

[30] C. Liu and D. B. Rubin, "The ECME Algorithm: A Simple Extension of EM and ECM with Faster Monotone Convergence," *Biometrika*, vol. 81, no. 4, pp. 633–648, 1994, doi: 10.2307/2337067.

[31] Wu L., *Mixed Effects Models for Complex Data*. Chapman and Hall/CRC., 2010.

[32] J. Jiang, *Linear and Generalized Linear Mixed Models and Their Applications*. Springer Science+Business Media, NY., 2007. doi: 10.1007/978-0-387-47946-0.

[33] B. T. West, K. B. Welch, and A. T. Galecki, *Linear Mixed Models: A Practical Guide Using Statistical Software*. Boca Raton, Florida: CRC Press LLC, 2007.

[34] S. Puntanen and G. P. H. Styan, "The Equality of the Ordinary Least Squares Estimator and the Best Linear Unbiased Estimator," *Am Stat*, vol. 43, no. 3, pp. 153–161, 1989.

[35] S. R. Searle, "The matrix Handling of BLUE and BLUP in the mixed linear model," *Linear Algebra Appl*, vol. 264, pp. 291–311, 1997, doi: https://doi.org/10.1016/S0024-3795(96)00400-4.

[36] H. O. Hartley and J. N. K. Rao, “Maximum-Likelihood Estimation for the Mixed Analysis of Variance Model,” *Biometrika*, vol. 54, no. 1/2, pp. 93–108, 1967, doi: 10.2307/2333854.

[37] S. R. Searle, G. Casella, and C. E. McCulloch, *Variance Components*. USA: John Wiley & Sons.

[38] C. E. McCulloch and S. R. Searle, *Generalized, Linear, and Mixed Models*. USA: John Wiley & Sons Inc, 2001.

[39] C. F. J. Wu, “On the Convergence Properties of the EM Algorithm,” *The Annals of Statistics*, vol. 11, no. 1, pp. 95–103, 1983, [Online]. Available: http://www.jstor.org/stable/2240463

[40] T. A. Louis, “Finding the Observed Information Matrix when Using the EM Algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 44, no. 2, pp. 226–233, 1982, [Online]. Available: http://www.jstor.org/stable/2345828

[41] F. Yavuz, “Parameter Estimations in Linear Mixed Models With Heavy Tailed Distributions,” PhD Thesis, Yıldız Technical University , İstanbul, 2015. Accessed: Aug. 13, 2023. [Online]. Available: https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp

[42] G. E. P. Box and G. Tiao, *Bayesian inference in statistical analysis*. New York: Reading Addison-Wesley, 1973.

[43] M. West, “On scale mixtures of normal distributions,” *Biometrika*, vol. 74, no. 3, pp. 646–648, Sep. 1987, doi: 10.1093/biomet/74.3.646.

[44] K. T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*. London: Chapman and Hall, 1990.

[45] E. Gómez-Sánchez-Manzano, M. A. Gómez-Villegas, and J. M. Marín, “Multivariate Exponential Power Distributions as Mixtures of Normal

Distributions with Bayesian Applications," *Commun Stat Theory Methods*, vol. 37, no. 6, pp. 972–985, Feb. 2008, doi: 10.1080/03610920701762754.

[46] K. Lange and J. S. Sinsheimer, "Normal/Independent Distributions and Their Applications in Robust Regression," *Journal of Computational and Graphical Statistics*, vol. 2, no. 2, pp. 175–198, Jun. 1993, doi: 10.1080/10618600.1993.10474606.

[47] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes*. Boca Raton: Chapman and Hall, 2000.

[48] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 1st ed. New York: Wiley., 1994.

[49] X. Li, X. Bai, and W. Song, "Robust mixture multivariate linear regression by multivariate Laplace distribution," *Stat Probab Lett*, vol. 130, pp. 32–39, 2017, doi: https://doi.org/10.1016/j.spl.2017.06.028.

[50] X. Yang, X. Liu, and C. Xu, "Robust Mixture Probabilistic Partial Least Squares Model for Soft Sensing With Multivariate Laplace Distribution," *IEEE Trans Instrum Meas*, vol. 70, pp. 1–9, 2021, doi: 10.1109/TIM.2020.3009354.

[51] J. Shi, K. Chen, and W. Song, "Robust errors-in-variables linear regression via Laplace distribution," *Stat Probab Lett*, vol. 84, pp. 113–120, 2014, doi: https://doi.org/10.1016/j.spl.2013.09.036.

[52] O. Barndorff-Nielsen, "Hyperbolic Distributions on Hyperbolae," *Scandinavian Journal of Statistics*, vol. 5, no. 3, pp. 151–157, 1978, [Online]. Available: http://www.jstor.org/stable/4615702

[53] O. Barndorff-Nielsen and C. Halgreen, "Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 38, no. 4, pp. 309–311, 1977, doi: 10.1007/BF00533162.

[54] F. Graybill, *Matrices with Applications in Statistics*, 2nd ed. Wadsworth Group, 1983.

[55] D. N. Anderson, “A multivariate Linnik distribution,” *Stat Probab Lett*, vol. 14, no. 4, pp. 333–336, 1992, doi: https://doi.org/10.1016/0167-7152(92)90067-F.

[56] K. Plungpongpun, “Analysis of Multivariate Data Using Kotz Type Distribution,” PhD Dissertation, Old Dominion University, USA, 2003. doi: 10.25777/6rcj.

[57] D. N. Naik and K. Plungpongpun, “A Kotz-Type Distribution for Multivariate Statistical Inference,” in *Advances in Distribution Theory, Order Statistics, and Inference*, N. Balakrishnan, J. M. Sarabia, and E. Castillo, Eds., Boston, MA: Birkhäuser Boston, 2006, pp. 111–124. doi: 10.1007/0-8176-4487-3_7.

[58] Y. Guney, O. Arslan, and F. G. Yavuz, “Robust estimation in multivariate heteroscedastic regression models with autoregressive covariance structures using EM algorithm,” *J Multivar Anal*, vol. 191, p. 105026, 2022, doi: https://doi.org/10.1016/j.jmva.2022.105026.

[59] T. Kozubowski and K. Podgorski, “A Class Of Asymmetric Distributions,” *ARCH*, vol. 1, pp. 113–134, Jul. 1999.

[60] T. J. Kozubowski and K. Podgórski, “A Multivariate and Asymmetric Generalization of Laplace Distribution,” *Comput Stat*, vol. 15, no. 4, pp. 531–540, 2000, doi: 10.1007/PL00022717.

[61] T. Kozubowski and K. Podgorski, “Asymmetric Laplace Distributions,” *The Mathematical Scientist*, vol. 25, pp. 37–46, Jan. 2000, doi: 10.1007/978-1-4612-0173-1_3.

[62] T. Kollo and M. Srivastava, “Estimation and Testing of Parameters in Multivariate Laplace Distribution,” *Communications in Statistics Theory and*

Methods Vol. 33, vol. No. 10, pp. 2363–2387, Oct. 2004, doi: 10.1081/STA-200031408.

[63] O. Barndorff-Nielsen, J. Kent, and M. Sorensen, “Normal Variance-Mean Mixtures and z Distributions,” *Int Stat Rev*, vol. 50, no. 2, pp. 145–159, 1982, doi: 10.2307/1402598.

[64] O. Barndorff-Nielsen and J. N. Darroch, “A Stochastic Model for Sand Sorting in a Wind-Tunnel,” *Adv Appl Probab*, vol. 13, no. 2, pp. 282–297, 1981, doi: 10.2307/1426684.

[65] R. Pourmousa, A. Jamalizadeh, and M. Rezapour, “Multivariate normal mean–variance mixture distribution based on Birnbaum–Saunders distribution,” *J Stat Comput Simul*, vol. 85, no. 13, pp. 2736–2749, Sep. 2015, doi: 10.1080/00949655.2014.937435.

[66] F. L. Schumacher, L. A. Matos, and C. R. B. Cabral, “Canonical fundamental skew-t linear mixed models,” Sep. 2021, [Online]. Available: http://arxiv.org/abs/2109.12152

[67] G. H. Givens and J. A. Hoeting, *Computational Statistics*, 2nd ed. USA: John Wiley & Sons, 2012.

[68] C. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer, 1999.

[69] A. M. Johansen, “Markov Chain Monte Carlo,” in *International Encyclopedia of Education (Third Edition)*, P. Peterson, E. Baker, and B. McGaw, Eds., Oxford: Elsevier, 2010, pp. 245–252. doi: https://doi.org/10.1016/B978-0-08-044894-7.01347-6.

[70] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of State Calculations by Fast Computing Machines,” *J Chem Phys*, vol. 21, no. 6, pp. 1087–1092, Dec. 2004, doi: 10.1063/1.1699114.

[71] W. K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970, doi: 10.2307/2334940.

[72] D. B. Hitchcock, "A History of the Metropolis–Hastings Algorithm," *Am Stat*, vol. 57, no. 4, pp. 254–257, Nov. 2003, doi: 10.1198/0003130032413.

[73] S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings Algorithm," *Am Stat*, vol. 49, no. 4, pp. 327–335, Nov. 1995, doi: 10.1080/00031305.1995.10476177.

[74] W. Debski, "Chapter 1 - Probabilistic Inverse Theory," in *Advances in Geophysics*, vol. 52, R. Dmowska, Ed., Elsevier, 2010, pp. 1–102. doi: https://doi.org/10.1016/S0065-2687(10)52001-6.

[75] L. Tierney, "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, vol. 22, no. 4, pp. 1701–1728, Dec. 1994, doi: 10.1214/aos/1176325750.

[76] K. Mosegaard and A. Tarantola, "Monte Carlo sampling of solutions to inverse problems," *J Geophys Res*, vol. 1001, pp. 12431–12448, Jul. 1995, doi: 10.1029/94JB03097.

[77] M. Bosch, C. Barnes, and K. Mosegaard, "Multi-step samplers for improving efficiency in probabilistic geophysical inference," in *Methods and Applications of Inversion*, P. C. Hansen, B. H. Jacobsen, and K. Mosegaard, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 50–67. doi: 10.1007/BFb0010283.

[78] P. H. Peskun, "Optimum Monte-Carlo Sampling Using Markov Chains," *Biometrika*, vol. 60, no. 3, pp. 607–612, 1973, doi: 10.2307/2335011.

[79] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. Boca, Raton: Chapman & Hall/CRC Press, 1995.

[80] S. Jackman, “Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo,” *Am J Pol Sci*, vol. 44, no. 2, pp. 375–404, 2000, doi: 10.2307/2669318.

[81] H. Du, Z. Ke, G. Jiang, and S. Huang, “The Performances of Gelman-Rubin and Geweke’s Convergence Diagnostics of Monte Carlo Markov Chains in Bayesian Analysis,” *Journal of Behavioral Data Science*, vol. 2, pp. 1–24, Nov. 2022, doi: 10.35566/jbds/v2n2/p3.

[82] V. Roy, “Convergence Diagnostics for Markov Chain Monte Carlo,” *Annu Rev Stat Appl*, vol. 7, no. 1, pp. 387–412, Mar. 2020, doi: 10.1146/annurev-statistics-031219-041300.

[83] G. Gerber, B. Brooker, and U. M. Scharler, “Automated workflow for incorporation and evaluation of data uncertainty in ecological networks with autoLIMR,” *Ecol Inform*, p. 102375, 2023, doi: https://doi.org/10.1016/j.ecoinf.2023.102375.

[84] K. L. Mengersen and S. Knight, “MCMC : how do we know when to stop ?,” 1999. [Online]. Available: https://api.semanticscholar.org/CorpusID:18533865

[85] D. W. Hogg and D. Foreman-Mackey, “Data Analysis Recipes: Using Markov Chain Monte Carlo*,” *Astrophys J Suppl Ser*, vol. 236, no. 1, p. 11, 2018, doi: 10.3847/1538-4365/aab76e.

[86] C. R. Rao, “Prediction of Future Observations in Growth Curve Models,” *Statistical Science*, vol. 2, no. 4, pp. 434–447, 1987, [Online]. Available: http://www.jstor.org/stable/2245538

[87] J. C. Lee, “Prediction and Estimation of Growth Curves With Special Covariance Structures,” *J Am Stat Assoc*, vol. 83, no. 402, pp. 432–440, 1988, doi: 10.2307/2288859.

[88] R Core Team, "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2023.

[89] Y. D. Lapierre *et al.*, "A controlled dose-ranging study of remoxipride and haloperidol in schizophrenia - a Canadian multicentre trial," *Acta Psychiatr Scand*, vol. 82, no. S358, pp. 72–77, Apr. 1990, doi: https://doi.org/10.1111/j.1600-0447.1990.tb05293.x.

[90] A. Cnaan, N. M. Laird, and P. Slasor, "Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data," *Stat Med*, vol. 16, no. 20, pp. 2349–2380, Oct. 1997.

[91] J. W. Hogan and N. A. N. M. Laird, "Mixture Models for the Joint Distribution of Repeated Measures and Event TIimes," *Stat Med*, vol. 16, no. 3, pp. 239–257, Feb. 1997.

[92] A. M. Mood, *Introduction to the theory of statistics.* New York, NY, US: McGraw-Hill, 1950.

[93] N. Shen and B. González, "Bayesian information criterion for linear mixed-effects models," *arXiv preprint arXiv:2104.14725*, 2021.

[94] H. Lai and X. Gao, "Modified BIC Criterion for Model Selection in Linear Mixed Models," *Mathematics*, vol. 11, no. 9, p. 2130, 2023.

[95] S. R. Searle, *Matrix Algebra Useful for Statistics*. New York: John Wiley & Sons, 1982.

[96] G. Matsaglia and G. P. H. Styan, "Equalities and Inequalities for Ranks of Matrices †," *Linear and Multilinear Algebra*, vol. 2, no. 3, pp. 269–292, Jan. 1974, doi: 10.1080/03081087408817070.

[97] G. Marsaglia and G. P. H. Styan, "Rank Conditions for Generalized Inverses of Partitioned Matrices," *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, vol. 36, no. 4, pp. 437–442, 1974, [Online]. Available: http://www.jstor.org/stable/25049954

[98] B. J. Smith, "boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference," *J Stat Softw*, vol. 21, no. 11, pp. 1–37, Nov. 2007, doi: 10.18637/jss.v021.i11.

[99] A. E. Raftery and S. Lewis, "How many iterations in the Gibbs sampler?" *Bayesian statistics*, vol. 4, no. 2, pp. 763–773, 1992.

[100] C. M. Carlo, "Markov chain monte carlo and gibbs sampling," *Lecture notes for EEB*, vol. 581, no. 540, p. 3, 2004.

[101] J. Geweke, "Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments," *Bayesian statistics*, vol. 4, pp. 641–649, 1992.

[102] P. Heidelberger and P. D. Welch, "Simulation Run Length Control in the Presence of an Initial Transient," *Oper Res*, vol. 31, no. 6, pp. 1109–1144, Dec. 1983, doi: 10.1287/opre.31.6.1109.

APPENDICES

A. Some Results in Matrix Algebra

A-1 General Matrix Notation

General matrix notation for a p x q dimensional matrix A,

$$A = \{a_{ij}\} \qquad i = 1, \dots, p \qquad j = 1, \dots, q$$

is given below (Searle [37]):

$$A = \{_m\, a_{ij}\}_{i=1}^{p}\,{}_{j=1}^{q} = \{_m\, a_{ij}\}$$

The subscript notation r, c, d are also used denote the row vector, column vector and diagonal matrix respectively. For example,

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_t \end{bmatrix} = \{_c\, b_i\}_{i=1}^{t} = \{_c\, b_i\}, \quad b' = \{_r\, b_i\}_{i=1}^{t} = \{_r\, b_i\},$$

$$\begin{bmatrix} a_1 & 0 & 0 & \cdots & 0 \\ 0 & a_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & & a_k \end{bmatrix} = \{_d\, a_i\}_{i=1}^{k} = \{_d\, a_i\},$$

A-2 Differentiating The Inverse of A Matrix

Let A be r x c dimensional matrix with elements functions of a scalar θ. The differentiation of matrix A with respect to θ is,

$$\frac{\partial \boldsymbol{A}}{\partial \theta} = \left\{\frac{\partial a_{ij}}{\partial \theta}\right\} \quad \text{for} \quad i = 1, \dots, r \quad \text{and} \quad j = 1, \dots, c.$$

When this is applied to the product $AA^{-1} = I$ for nonsingular A, it yields the following results (Searle [95]):

$$\frac{\partial \boldsymbol{A^{-1}}}{\partial \theta} = -\boldsymbol{A}^{-1} \frac{\partial \boldsymbol{A}}{\partial \theta} \boldsymbol{A}^{-1}.$$

A-3 Differentiating The Logarithm of Determinant of A Matrix

Let A be r x c dimensional matrix with elements functions of a scalar θ. For a nonsingular A (Searle[95]),

$$\frac{\partial}{\partial \theta} ln|\boldsymbol{A}| = tr\left(\boldsymbol{A}^{-1} \frac{\partial \boldsymbol{A}}{\partial \theta}\right).$$

A-4 Generalized Inverse of A Matrix

$$A = [A_1 \quad A_2]$$

$$A'A = \begin{bmatrix} A_1' \\ A_2' \end{bmatrix} [A_1 \quad A_2] = \begin{bmatrix} A_1'A_1 & A_1'A_2 \\ A_2'A_1 & A_2'A_2 \end{bmatrix}$$

The generalized inverse of $A'A$ is given below (Searle [95])

$$A^{-1} = \begin{bmatrix} (A_1'A_1)^- & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -(A_1'A_1)^- A_1'A_2 \\ I \end{bmatrix} (A_2'M_1A_2)^- [-A_2'A_1(A_1'A_1)^- \quad I]$$

where $M_1 = I - A_1(A_1'A_1)^- A_1'$ and

$$A^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & (A_2'A_2)^- \end{bmatrix} + \begin{bmatrix} I \\ -(A_2'A_2)^- A_2'A_1 \end{bmatrix} (A_1'M_2A_1)^- [I \quad -A_1'A_2(A_2'A_2)^-]$$

where $M_2 = I - A_2(A_2'A_2)^- A_2'$.

A-5 The Schur Complement

The inverse of a nonsingular partitioned matrix is given below.

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -A^{-1}B \\ I \end{bmatrix} (D - CA^{-1}B)^{-1} [-CA^{-1} \quad I]$$

$D - CA^{-1}B$ in the above formula is the Schur complement of A (Searle et al. [37]). Some important results related to the Schur complement is defined by Marsagila and Styan [96], [97]. The two used in this study are given below.

$$(D - CA^{-1}B)^{-1} = D^{-1} + D^{-1}C(A - BD^{-1}C)BD^{-1}$$

$$(D + CA^{-1}B)^{-1} = D^{-1} - D^{-1}C(A + BD^{-1}C)BD^{-1}$$

B. Some Known Features of Multivariate Normal Distribution

B-1 Distribution of Linear Function of a Random Variable with Normal Distribution

Let X be distributed q-dimensional multivariate normal distribution; $X \sim N_q(\mu, \Sigma)$ and $Y = BX + a$. Then $Y \sim N_q(B\mu + a,\ B\Sigma B')$ for a non-singular scalar matrix B and $a \in R^q$.

B-2 Marginal and Conditional Distribution of a Normal Distributed Random Variable

A q-dimensional random vector X is said to have a multivariate normal distribution with q x 1 location vector $\mu \in \mathbb{R}$, a q x q positive-definite dispersion matrix Σ, we say $X \sim N_q(\mu, \Sigma)$.

Let X be partitioned as

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

where $X_1 : (q_1 \text{x } 1)$, $\Sigma_{11} : (q_1 \text{ x } q_1)$.

Then X_1 has q_1 dimensional normal distribution with location vector μ_1 and Σ_{11} variance-covariance matrix, $X_1 \sim N_{q_1}(\mu_1, \Sigma_{11})$ and the pdf is given:

$$f(x_1) = \frac{1}{(2\pi)^{\frac{q_1}{2}} |\Sigma_{11}|^{\frac{1}{2}}} \, exp\left(-\frac{1}{2}(x_1 - \mu_1)'(\Sigma_{11})^{-1}(x_1 - \mu_1)\right)$$

$$-\infty < x_i < \infty \text{ and } i = 1,2, \dots, q_1.$$

Conditional distribution of X_1 given X_2 is as below:

$$X_1 \mid X_2 \sim N\left(\mu_1 + \Sigma_{12}\Sigma_{22}{}^{-1}(X_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{21}\right)$$

B-3 Expectation of Quadratic form of a Normal Distributed Random Variable

Let a random vector X be distributed multivariate normal, expectation of its quadratic form, $X'AX$ is written as below:

$$E(x'A\,x) = tr(A\Sigma) + \mu' A\mu$$

C. MCMC Convergence Diagnostics for Metropolis Algorithm

We use a combination of visual and statistical convergence diagnostics to evaluate the convergence of the Metropolis algorithm. To conduct this analysis, we use the **boa** R package (Smith [98]), which offers a range of useful visual and statistical tools for assessing MCMC convergence.

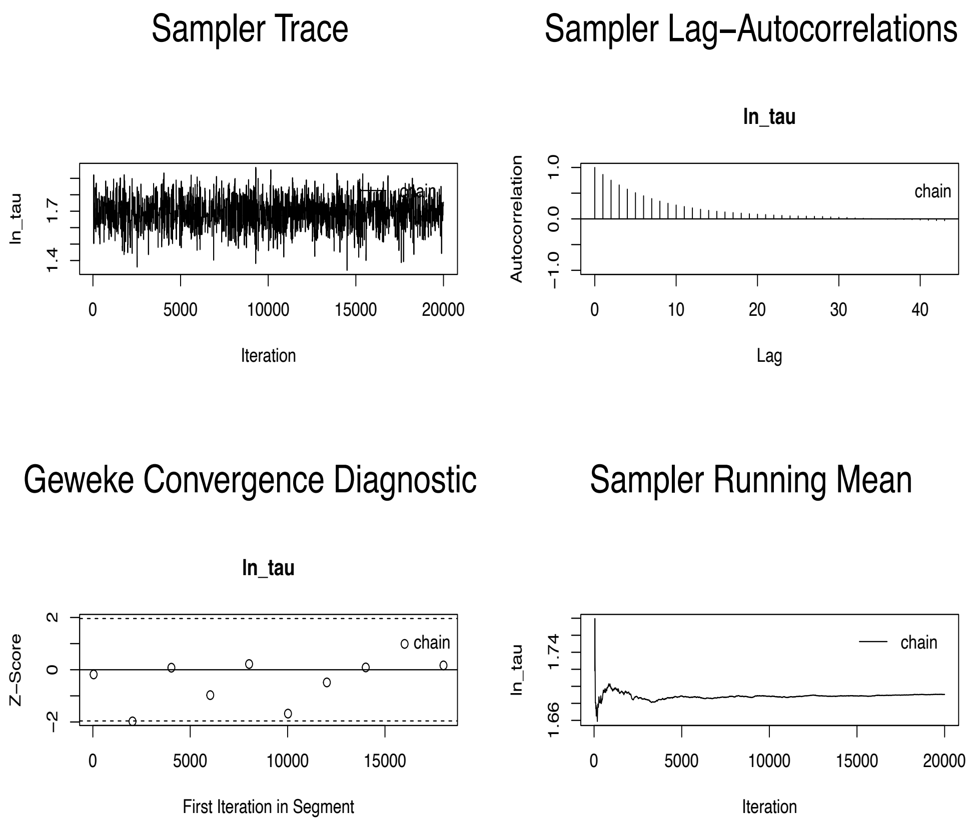

Figure C. 1. Markov Chain Monte Carlo (MCMC) visual convergence diagnostics for the Metropolis algorithm with number of iterations = 20000

Based on the "Sampler Lag-Autocorrelations" plot shown in Figure C. 1, there appears to be an issue of autocorrelation resulting in mixing problems. Although visual inspection of the plot may offer preliminary insights, it is crucial to supplement this subjective evaluation with robust statistical analysis to ensure the accuracy and validity of the findings. The Raftery-Lewis dependence factor (I) is a statistical diagnostic to assess the degree of autocorrelation between successive observations in a sample. When the value of I surpasses the threshold of 5, it signifies a high level of autocorrelation between the observations (Raftery and Lewis [99]). In the current scenario, the high autocorrelation issue has been conclusively established as the calculated dependence factor has exceeded the threshold of 5. To reduce autocorrelation, a common approach is thinning the output by storing only every *m*th point after the burn-in period (Carlo [100]). We subsampled every fifth iteration after the burn-in period and performed the convergence diagnostic assessment. Visual convergence diagnostic for the thinned sample is presented in Figure C. 2.

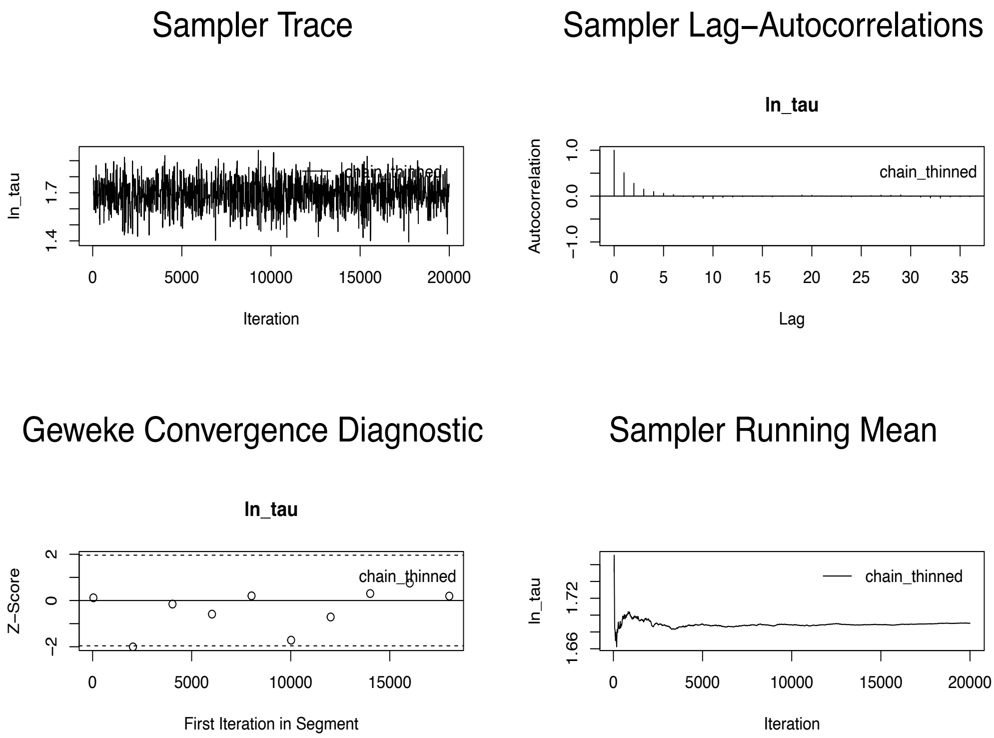

Figure C. 2. Markov Chain Monte Carlo (MCMC) visual convergence diagnostics for the Metropolis algorithm after thinning

The stationarity condition can be determined by checking the Geweke Z-score mean, considered stationary when it falls between -1.96 and 1.96, corresponding to a probability of 0.05 (Du et al. [81]). Since the different segments fall within the range, Geweke Z-scores thereof infer stationarity (Figure C. 2). To statistically assess stationarity, it is recommended to check Geweke Z-scores (Geweke [101]) and the Heidelberger-Welch diagnostic (Heidelberger and Welch [102]). The Heidelberger-Welch diagnostic comprises two parts. The first part examines whether the sampled chain is stationary, while the second part, known as the halfwidth test, determines if there are adequate iterations to accurately estimate the mean value with an acceptable degree of precision. So, this study indicates stationarity by passing both the Heidelberger-Welch stationarity and halfwidth test and the Geweke Z-score (0.12).

Mixing criteria are statistically proven through lag-k autocorrelation decreasing quickly and low Raftery-Lewis dependence factors (I$\approx \mathbf{2,4} < \mathbf{5}$) as stated by Raftery and Lewis [99]. Upon observation of the trace plot, the "hairy caterpillar" (Roy [82]) displays a random pattern that suggests a good mixing process. As observed in Figure C. 2, lag-k autocorrelation values drop to zero almost as quickly as k increases, which indicates fast mixing of MCMC chains (Roy [82]).

The effectiveness of sampling can be visually assessed by analyzing running mean plots, which exhibit the Monte Carlo time average estimates against the number of iterations (Roy [82]. When the estimated value of a sample is stabilized at a fixed number, the chain has collected a sufficient number of samples. The mean plot shows stabilization of the mean as the number of iterations increases (Figure C. 2). The Raftery-Lewis required number of iterations (N=3746) supports this outcome statistically, suggesting that the specified number of iterations (3961 after thinning) is enough.

CURRICULUM VITAE

Surname, Name: Kapucu, Tuğba

EDUCATION

Degree	Institution	Year of Graduation
MS	METU Statistics	2016
BS	METU Statistics-Minor Program	2013
BS	METU Elementary Math. Education	2013
High School	Fatma Saygın Anadolu High School, İzmir	2008

FOREIGN LANGUAGES

Advanced English

PUBLICATIONS

1. Kapucu, T., İlk Dağ, Ö., & Batmaz, İ., (2018). A Categorical Principal Component Regression on Computer-Assisted Instruction in Probability Domain. Trends and Perspectives in Linear Statistical Inference (pp.143-164), Zürich: Springer, London/Berlin.