Robust multivariate mixture regression models

by

Xiongya Li

B.S., EMPORIA STATE UNIVERSITY, 2012

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Abstract

In this dissertation, we proposed a new robust estimation procedure for two multivariate mixture regression models and applied this novel method to functional mapping of dynamic traits. In the first part, a robust estimation procedure for the mixture of classical multivariate linear regression models is discussed by assuming that the error terms follow a multivariate Laplace distribution. An EM algorithm is developed based on the fact that the multivariate Laplace distribution is a scale mixture of the multivariate standard normal distribution. The performance of the proposed algorithm is thoroughly evaluated by some simulation and comparison studies. In the second part, the similar idea is extended to the mixture of linear mixed regression models by assuming that the random effect and the regression error jointly follow a multivariate Laplace distribution. Compared with the existing robust t procedure in the literature, simulation studies indicate that the finite sample performance of the proposed estimation procedure outperforms or is at least comparable to the robust t procedure. Comparing to t procedure, there is no need to determine the degrees of freedom, so the new robust estimation procedure is computationally more efficient than the robust t procedure. The ascent property for both EM algorithms are also proved. In the third part, the proposed robust method is applied to identify quantitative trait loci (QTL) underlying a functional mapping framework with dynamic traits of agricultural or biomedical interest. A robust multivariate Laplace mapping framework was proposed to replace the normality assumption. Simulation studies show the proposed method is comparable to the robust multivariate t-distribution developed in literature and outperforms the normal procedure. As an illustration, the proposed method is also applied to a real data set.

Robust multivariate mixture regression models

by

Xiongya Li

B.S., Emporia State University, 2012

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Approved by:

Major Professor
Dr. Weixing Song

Copyright

Abstract

In this dissertation, we proposed a new robust estimation procedure for two multivariate mixture regression models and applied this novel method to functional mapping of dynamic traits. In the first part, a robust estimation procedure for the mixture of classical multivariate linear regression models is discussed by assuming that the error terms follow a multivariate Laplace distribution. An EM algorithm is developed based on the fact that the multivariate Laplace distribution is a scale mixture of the multivariate standard normal distribution. The performance of the proposed algorithm is thoroughly evaluated by some simulation and comparison studies. In the second part, the similar idea is extended to the mixture of linear mixed regression models by assuming that the random effect and the regression error jointly follow a multivariate Laplace distribution. Compared with the existing robust t procedure in the literature, simulation studies indicate that the finite sample performance of the proposed estimation procedure outperforms or is at least comparable to the robust t procedure. Comparing to t procedure, there is no need to determine the degrees of freedom, so the new robust estimation procedure is computationally more efficient than the robust t procedure. The ascent property for both EM algorithms are also proved. In the third part, the proposed robust method is applied to identify quantitative trait loci (QTL) underlying a functional mapping framework with dynamic traits of agricultural or biomedical interest. A robust multivariate Laplace mapping framework was proposed to replace the normality assumption. Simulation studies show the proposed method is comparable to the robust multivariate t-distribution developed in literature and outperforms the normal procedure. As an illustration, the proposed method is also applied to a real data set.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First and foremost I would like to thank my major advisor Dr. Weixing Song. It has been an honor to be his Ph.D. student. He has taught me, both consciously and unconsciously, how good research is done. I appreciate all his contributions of time and ideas to make my Ph.D. experience productive and stimulating. I am also thankful for the excellent example he has provided as a successful statistician and professor.

Besides my advisor, I would like to thank my committee members: Dr. James Neill, Dr. Juan Du and Dr. Weiqun Wang for their time, interest, and helpful comments. A special thank goes to Dr. James Neill for offering me the opportunity to enhance my knowledge at Department of Statistics and to become a K-state family member.

Also I would like to thank Bonnie Messmer and Jo Blackburn for offering generous help on the daily routine over the past few years. There are no words coming to close to describing the gratitude to my friends. It has been a memorable time to study and discuss statistical questions with them.

Lastly, I would like to thank my family for all their love and encouragement. For my parents who raised me with a full of love and supported me in all my pursuits. And most of all for my loving, supportive, encouraging, and patient wife Yinhao Du whose faithful support during the final stages of this Ph.D. is so appreciated

Chapter 1

Introduction

Among various robust estimation procedures, using heavy-tailed distributions to achieve the robustness has been received more and more attentions from statisticians and practitioners. In this thesis, we shall propose a new robust estimation procedures in regression models using the multivariate Laplace distributions. Due to their natural connection to least absolute deviation criterion and normal distributions, the proposed estimation procedure enjoys robustness and the computational efficiency at the same time. We will first consider the robust estimation procedure in the classical multivariate regression models, then the methodology will be extended to mixture of the linear mixed models.

1.1 Robust Mixture Multivariate Regression

Finite mixture regression modeling is an efficient tool to investigate the relationship between a response variable and a set of predictors when the underlying population consists of several unknown latent homogeneous groups, and it has been already applied for more than a hundred years since **?**. More real examples on finite mixture modeling can be found in **?**, **?**, **?**, **?** and the references therein. Statistical inferences have been discussed extensively for finite mixture modeling when the normality is assumed for the regression error in each cluster. However, since the likelihood function for normal mixture regression models often exhibits complicated form, so the maximum likelihood is hard to derive, and the unknown regression parameters are usually estimated using the expectation and maximization (EM) algorithm. Due to its unweighted least squares estimation nature, the maximum likelihood estimate (MLE) of the regression parameters are not robust to the outliers and the data with heavy tails. Because of its wide application in practice, how to design robust estimation procedures in the finite mixture regression models has attracted much attention from statisticians.

Extensive research has been done for linear or mixture of linear regression models when the response variable is a scalar. For examples, **?** proposed a trimmed likelihood estimator (TLE) to robustly estimate the mixtures and the breakdown points of the TLE for the mixture component parameters is also characterized; Replacing the least square criterion in the M step of EM algorithm designed for normal mixtures, **?** achieved robustness using Tukey's bisquare and Huber's ψ-functions; A class of S-estimators were introduced in **?** which exhibit certain robustness and the parameter estimation is achieved via an expectation-conditional maximization algorithm. Inspired by **?**, **?** proposed a new robust estimation method for mixture of linear regression by assuming that the mixtures have t-distributions, the EM algorithm is made possible by the fact that t-distribution is a scale mixture of a normal distribution. Due to the selection of degrees of freedom, the procedure in **?** requires relatively heavy computation although the choice of degrees of freedom provides certain adaptivity to the data. Realizing that the Laplace distribution is also a scale mixture of normal distribution, **?** proposed an alternative robust estimation procedure by assuming the random error has a Laplace distribution, which has a natural connection with the least absolute deviation (LAD) procedure and LAD is a well-known robust estimation procedure, see **?** and **?** for more detail on LAD methodology.

Compare to the relatively extensive discussion for the univariate response cases, there is less work having been done for the multivariate linear regressions. **?** designed a robust estimation procedure using the multivariate skewed t-distribution, which offers a great deal of flexibility that accommodates asymmetry and heavy tails simultaneously, and **?** proposed a mixture of multivariate t-distribution to fit the multivariate continuous data with a large number of missing values. Similar to the cases of scalar responses, one has to decide the proper degree of freedom in order to apply these methods. Up to now, we haven't seen any work on developing robust estimation procedures for the multivariate linear regression with the multivariate Laplace distribution. We expect a multivariate version of **?** procedure should perform equally well in the multivariate linear regression. In Chapter 2, we discussed the proposed robust estimation procedure by using multivariate Laplace distribution more in detail. Following the multivariate mixture regression models, we applied to the mixture of linear mixed models.

1.2 Robust Mixture of Linear Mixed Models

Mixed-effect modeling is often applied to model the repeated measurements because of its flexibility to handle both balanced and unbalanced data with heterogeneity, thus it is widely used in the fields of biology, agriculture, and economics etc.. However, in many applications, the data may come from different clusters. In such scenarios, mixture modeling of the linear mixed effect is often considered. See **?** for an real data analysis from genetic study.

The classical linear mixed model, proposed by **?**, takes the form of $Y = X\beta + Zb + \epsilon$, where Y is an $m \times 1$ vector of responses, X is an $m \times p$ known design matrix for the fixed effects, β is an unknown $p \times 1$ vector of the fixed effects, Z is an $m \times q$ known design matrix for the random effects, b is a $q \times 1$ vector of the random effects which is often assumed to be $N_q(0, \Phi)$, the regression error ϵ is an $m \times 1$ vector of experimental errors which is assumed to be $N_m(0, \Sigma)$, and ϵ and b are assumed to be independent. It's not hard to see that Y follows multivariate normal distribution with $E(Y) = X\beta$ and $\mathrm{Cov}(Y) = Z\Phi Z^T + \Sigma$. Now suppose we have an experiment in which m subjects come from G groups or clusters, and n_i repeated measurements are collected from each subject. Let $C_{ij} = 1$ denote the i-th subject from the j-th group, $i = 1, ..., m$ and $j = 1, ..., G$. Therefore, if $C_{ij} = 1$, then the observations from the i-th subject follow the mixture of linear mixed model proposed by **?**,

$$Y_i = X_i\beta_j + Z_i b_{ij} + \epsilon_{ij}, \tag{1.2.1}$$

where $Y_i \in \mathbf{R}^{n_i}$ is the response, $X_i \in \mathbf{R}^{n_i \times p}$ and $Z_i \in \mathbf{R}^{n_i \times q}$ are the known design matrix for fixed and random effects, respectively. $\beta_j \in \mathbf{R}^p$ and $b_{ij} \in \mathbf{R}^q$ are the coefficients of the fixed and random effects, and ϵ_{ij} are the regression errors. The traditional mixed effects models often assume that

$$b_{ij} \sim N_q(0, \Phi_j) \text{ and } \epsilon_{ij} \sim N_{n_i}(0, \Sigma_{ij}), \tag{1.2.2}$$

where b_{ij}, ϵ_{ij} are independent for $i = 1, ..., m$, $j = 1, ..., G$, and Σ_{ij} is typically assumed to depend on i only through their dimensions. For example, an AR(1) covariance structure for Σ_{ij} is a usual assumption, see **?**. The covariance matrix Φ_j may be structured or unstructured. A diagonal structural of Φ_j is adopted in **?**. Alternatively, model (1.2.1) and (1.2.2) can be written as

$$\begin{pmatrix} Y_i \\ b_{ij} \end{pmatrix} \Bigg|_{C_{ij}=1} \sim N_{n_i+q}\left(\begin{pmatrix} X_i\beta_j \\ 0 \end{pmatrix}, \begin{pmatrix} Z_i\Phi_j Z_i^T + \Sigma_{ij} & Z_i\Phi_j \\ \Phi_j Z_i^T & \Phi_j \end{pmatrix} \right),$$

where $i = 1, \ldots, m$ and $j = 1, \ldots, G$. In real applications, the latent class variable C_{ij}'s are not available, thus the conditional distribution of Y_i given X_i and Z_i can be written as

$$Y_i \sim \sum_{j=1}^{G} p_j N_{n_i}(X_i\beta_j, Z_i\Phi_j Z_i^T + \Sigma_{ij}), \tag{1.2.3}$$

where $p_j = P(C_{ij} = 1)$, for $i = 1, ..., m$ and $j = 1, ..., G$.

The normal mixture linear mixed model (1.2.3) can be fitted iteratively by maximum likelihood via the expectation maximization (EM) algorithm. The general principle and implementation of the EM algorithm can be found in **?**, and an extensive introduction on this topic can be found in **?**. More examples on normal mixtures with EM algorithms can be found in **?**, **?**, **?**, **?** and **?** and the references therein. However, the normal mixture of linear mixed models exhibit non-robustness if the random effects or the regression errors have longer than normal tails. To robustify the estimation procedure in such important model, t-distribution, which has a longer tail than the normal distribution, is often used as the "working" distributions for the random effects and regression errors.

There are many existing literature on achieving the robustness in various statistical models using the t distribution. For examples, **?**, **?**, **?**, **?** and **?**, among others. Motivated by **?**, **?** proposed a robust estimation procedure using the multivariate t distribution. To be specific, given $C_{ij} = 1$, **?** assume that

$$\left(\begin{array}{c} Y_i \\ b_{ij} \end{array} \right) \Bigg| C_{ij} = 1 \sim t_{n_i+q} \left(\left(\begin{array}{c} X_i\beta_j \\ 0 \end{array} \right), \left(\begin{array}{cc} Z_i\Phi_j Z_i^T + \Sigma_{ij} & Z_i\Phi_j \\ \Phi_j Z_i^T & \Phi_j \end{array} \right), v_j \right), \tag{1.2.4}$$

where $t_n(\mu, \Lambda, \nu)$ denotes an n-dimensional multivariate t distribution with mean vector μ, covariance matrix $\nu\Lambda/(\nu - 2)$, and degrees of freedom ν. An EM algorithm is designed to fit the mixture of linear mixed models and simulation studies show that the resulting estimates possesses robustness. However, to apply **?**'s procedure, one has to estimate the degrees of freedom of the t-distribution. They proposed to apply a numerical optimization method using the profile likelihood approach, which makes the procedure computationally extensive.

In this proposal, we shall propose a new robust method by replacing the multivariate t-distribution with the multivariate Laplace distribution. Similar to the multivariate t-distribution, the multivariate Laplace distribution is also a scale mixture of the multivariate normal distribution. This enables us to construct an efficient EM algorithm to estimate the unknown parameters in the model. Through simulation studies, we shall show that the proposed estimates are robust against heavy tailed data, also the proposed estimation procedure is computationally efficient.

1.3 Multivariate Laplace distribution

In this section, the definition and some important properties of the multivariate Laplace distribution which are directly related to our estimation procedures will be introduced.

There are multiple forms of definitions of the multivariate Laplace distribution. For example, the bivariate case was introduced by **?**, and the first form in larger dimensions was discussed in **?**. Later, the multivariate Laplace was introduced as a special case of the multivariate Linnik distribution in ?, and the multivariate power exponential distribution in ? and ?. ? presented multivariate Laplace distribution as a Gaussian scale mixture. ? presented a version of the multivariate Laplace distribution formally and thoroughly discussed its properties. The multivariate Laplace distribution is an attractive alternative to the multivariate normal distribution due to its heavier tails. For its application in image and speech recognition, ocean engineering and finance, see ?.

Definition 1.3.1. *A p-dimensional random vector U is called to have a multivariate Laplace distribution $ML_p(\mu, \Sigma)$, if its density function has the form of*

$$f_U(u) = \frac{2}{(2\pi)^{p/2}|\Sigma|^{1/2}} \left[\frac{Q(u;\mu,\Sigma)}{2}\right]^{\frac{1}{2}(1-\frac{p}{2})} K_{p/2-1}\left(\sqrt{2Q(u;\mu,\Sigma)}\right), \tag{1.3.1}$$

where $Q(u;\mu,\Sigma) = (u-\mu)'\Sigma^{-1}(u-\mu)$, μ is p-dimensional location parameter, Σ is a $p \times p$ covariance matrix, and $K_m(x)$ is the modified Bessel function of the second kind with order m, which is defined as

$$K_m(x) = \frac{\Gamma(m+1/2)(2x)^m}{\sqrt{\pi}} \int_0^\infty \frac{\cos(t)}{(t^2+x^2)^{m+1/2}} dt.$$

The modified Bessel function of the second kind is the solution to the modified Bessel differential equation, and sometimes it is also called the Basset function, the modified Bessel function of the third kind, or the Macdonald function. See ?, ? for more discussion on the Bessel functions. In fact, the multivariate Laplace distribution defined in Definition 1 is also a special case of the symmetric multivariate Bessel distribution defined in ?.

The following lemma provides some important probabilistic properties about the multivariate Laplace distribution.

Lemma 1.3.1. *Suppose a random vector $U \sim ML_p(\mu, \Sigma)$, then*

(i) $EU = \mu$ and $Cov(U) = \Sigma$;

(ii) The characteristic function of U is given by

$$\phi_U(t) = \frac{\exp(it'\mu)}{1 + t'\Sigma t/2}, \ t \in \mathbb{R}^p;$$

(iii) Let V be a scalar random variable with density function $f_V(v) = e^{-v} I(v \geqslant 0)$, Z be a p-dimensional standard normal random vector, that is $N_p(0, I)$, V and Z are independent. Then

$$U = \sqrt{V} \Sigma^{1/2} Z \sim ML_p(0, \Sigma);$$

(iv) Assume V and U are defined as above. Then

$$E\left(\frac{1}{V} \middle| U = u\right) = \sqrt{\frac{2}{u'\Sigma^{-1}u}} \frac{K_{-p/2}(\sqrt{2u'\Sigma^{-1}u})}{K_{1-p/2}(\sqrt{2u'\Sigma^{-1}u})}.$$

For the sake of brevity, these results are only summarized in Lemma 1.3.1 and their proofs can be founded in relevant literatures mentioned above. From (i) and the density function of the multivariate Laplace distribution, we can see that the Multivariate Laplace distribution is uniquely determined by its mean vector and covariance matrix. From (ii) we can see that the multivariate Laplace distribution defined by (1.3.1) indeed is a natural extension of univariate Laplace distribution. Similar to the univariate Laplace distribution, (iii) indicates that the multivariate Laplace distribution is a scale mixture of multivariate normal distribution, and this property makes it feasible to develop an efficient EM algorithm to implement the proposed robust estimation procedure. The property (iv) is emphasized here since it plays a crucial role in the E step of the developed EM algorithm.

Chapter 2

Robust Mixture Multivariate Regression by Multivariate Laplace Distribution

In this chapter, we first introduce the mixture of multivariate linear regression models in Section 2.1. The EM algorithm will be developed in Section 2.2 for both multivariate linear and mixture of multivariate linear regression models. Section 2.3 includes some simulation studies to evaluate the performance of the proposed methods and comparison studies with some existing methods will be also made. The proof of the ascent property of the proposed EM algorithm is deferred to Appendix A.

2.1 Mixture of Multivariate Linear Regression

Let G be a latent class variable such that given $G = j$, $j = 1, 2, \ldots, g$, $g \geq q$, a p-dimensional response Y and a q-dimensional predictor X are in one of the following multivariate linear regression models

$$Y = \beta_j' X + \varepsilon_j, \tag{2.1.1}$$

where, for each j, β_j is a $q\times p$ unknown regression coefficient matrix, and ε_j is a p-dimensional random error. Assume ε_j's are independent of X and it is commonly assumed that the density functions f_j of ε_j's are members in a location-scale family with mean 0 and covariance Σ_j. If we further suppose $P(G=j)=\pi_j, j=1,...,g$, then conditioning on X, the density function of Y is given by

$$f(y|x,\theta)=\sum_{j=1}^{g}\pi_j f_\varepsilon(y-\beta_j'x,0,\Sigma_j), \tag{2.1.2}$$

where $\theta=\{\pi_1,\beta_1,\Sigma_1,...,\pi_g,\beta_g,\Sigma_g\}$. The model (2.1.2) is the so called mixture multivariate regression models. The unknown parameters could be estimated by the maximum likelihood estimator (MLE), which maximizes the log-likelihood function (2.1.3) based on an independent sample $(X_i,Y_i), i=1,...,n$ from (2.1.2),

$$L_n(\theta)=\sum_{i=1}^{n}\log\left[\sum_{j=1}^{g}\pi_j f_\varepsilon(Y_i,\beta_j'X_i,\Sigma_j)\right]. \tag{2.1.3}$$

If $g=1$, then the mixture linear regression model is simply a multivariate linear regression model. The proposed robust estimation procedure is applicable for both multivariate linear regression models and mixture multivariate linear regression models.

The traditional maximum likelihood estimation procedure is based on the normality assumption. However, no explicit solution is available due to the untractable expression of (2.1.3), and EM algorithm thus developed to obtain its the maximizer, which is also evidenced from the simulation results presented in section 2.3. As we mentioned in section 1.1, the MLE based on the normality assumption is sensitive to outliers or heavy-tailed error distribution, and we shall develop a robust estimation procedure by assuming that the error distributions are Laplacian.

2.2 Robust Estimation Procedure

We start with the simpler case of $g=1$. The methodology developed for this case has its own interest, and moreover, the arguments used in this case can help us understand well the logic of the methodology development in the case of $g > 1$, without entangling us with the complexity of the notations.

2.2.1 Robust Estimation in Multivariate Linear Regression Models

Assume that $\varepsilon \sim ML_p(0, \Sigma)$. and $(Y_i', X_i')', i = 1, ..., n$ is a sample of size n from model (2.1.1) with $g = 1$. Then the likelihood function of β, Σ has the form of

$$L(\beta, \Sigma; \mathbf{Y}, \mathbf{X}) = \prod_{i=1}^{n} f_\varepsilon(Y_i, \beta' X_i, \Sigma),$$

where $\mathbf{Y} = (Y_1', Y_2', \ldots, Y_n')'$ and $\mathbf{X} = (X_1', X_2', \ldots, X_n')'$, and $f_\varepsilon(\cdot)$ is defined in (1.3.1). Due to the non-differentiable nature of the Laplace density function, the likelihood function is hard to maximize w.r.t. β and Σ. However, based on (iii) of Lemma 1.3.1, if the latent observation $\mathbf{V} = (V_1, V_2, \ldots, V_n)$ is available, then the complete likelihood function becomes

$$\begin{aligned} & L(\beta, \Sigma; \mathbf{Y}, \mathbf{X}, \mathbf{V}) \\ = \; & \prod_{i=1}^{n} \frac{1}{(2\pi V_i)^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2V_i} (Y_i - \beta' X_i)' \Sigma^{-1} (Y_i - \beta' X_i) - V_i \right), \end{aligned}$$

and the log-likelihood function therefore can be written as

$$\begin{aligned} & \ell(\beta, \Sigma; \mathbf{Y}, \mathbf{X}, \mathbf{V}) = \log L(\beta, \Sigma; \mathbf{Y}, \mathbf{X}, \mathbf{V}) \\ = \; & -\frac{p}{2} \sum_{i=1}^{n} \log(2\pi V_i) - \sum_{i=1}^{n} \frac{1}{2V_i} (Y_i - \beta' X_i)' \Sigma^{-1} (Y_i - \beta' X_i) - \sum_{i=1}^{n} V_i - \frac{n}{2} \log |\Sigma|. \end{aligned}$$

Following the two steps in the EM algorithm, assuming that $\beta^{(m)}$ and $\Sigma^{(m)}$ are the values of β and Σ for the m-th iteration, and denote

$$w_i = E\left[\frac{1}{V_i} \middle| Y_i, X_i, \beta^{(m)}, \Sigma^{(m)} \right],$$

then we have to calculate the conditional expectation

$$
\begin{aligned}
&E[\ell(\beta,\Sigma;\mathbf{Y},\mathbf{X},\mathbf{V})|\mathbf{Y},\mathbf{X},\beta^{(m)},\Sigma^{(m)}]\\
&=-\frac{np}{2}\log 2\pi-\frac{n}{2}\log|\Sigma|-\frac{1}{2}\sum_{i=1}^{n}w_i(Y_i-\beta' X_i)'\Sigma^{-1}(Y_i-\beta' X_i)\\
&\quad-\sum_{i=1}^{n}E[\log V_i|Y_i,X_i,\beta^{(m)},\Sigma^{(m)}]-\sum_{i=1}^{n}E[V_i|Y_i,X_i,\beta^{(m)},\Sigma^{(m)}],
\end{aligned}
$$

and maximize the conditional expectation with respect to β and Σ. Clearly, it is sufficient to minimize

$$
\frac{n}{2}\log|\Sigma|+\frac{1}{2}\sum_{i=1}^{n}w_i(Y_i-\beta' X_i)'\Sigma^{-1}(Y_i-\beta' X_i), \tag{2.2.1}
$$

with respect to β and Σ. From (iv) in Lemma 1.3.1, and recall the notation $Q(u;\mu,\Sigma)$ defined in Definition 1, we could obtain

$$
w_i=\frac{K_{-p/2}\left(\sqrt{2Q(Y_i;\beta^{(m)\prime}X_i,\Sigma^{(m)})}\right)}{K_{1-p/2}\left(\sqrt{2Q(Y_i;\beta^{(m)\prime}X_i,\Sigma^{(m)})}\right)}\sqrt{\frac{2}{Q(Y_i;\beta^{(m)\prime}X_i,\Sigma^{(m)})}}.
$$

If we further denote $\mathbf{W}=\mathrm{diag}(w_1,w_2,...,w_n)$, then the minimizer of (2.2.1) has the forms of

$$
\begin{aligned}
&\beta^{(m+1)}=(\mathbf{X'WX})^{-\mathbf{1}}\mathbf{X'WY},\\
&\Sigma^{(m+1)}=n^{-1}\mathbf{Y}'[\mathbf{W}-\mathbf{WX}(\mathbf{X'WX})^{-\mathbf{1}}\mathbf{X'W}]\mathbf{Y}.
\end{aligned}
$$

In particular, if $p=1$, then $K_{-1/2}(x)=K_{1/2}(x)=1$, so $w_i=\sigma^{(m)}\sqrt{2}/|Y_i-\beta^{(m)}X_i|$. This reproduces the result in **?**.

2.2.2 Multivariate Mixture Regression Models

In this section, we shall consider the mixture of multivariate linear regression model. Assume that $\varepsilon_j \sim ML(0, \Sigma_j)$ and $\sum_{j=1}^{g} \pi_j = 1$, $g > 1$. Define

$$G_{ij} = \begin{cases} 1 & \text{if i-th observation } (X_i, Y_i) \text{ is from j-th component;} \\ 0 & \text{otherwise.} \end{cases}$$

Let (X_i, Y_i, G_{ij}), $i = 1, 2, ..., n$, $j = 1, 2, ..., g$ be a sample from the model (2.1.1). Once again, recall the notation $Q(u; \mu, \Sigma)$ in Definition 1, and we further denote $Q_{ij} = Q(Y_i; \beta_j' X_i, \Sigma_j)$ for the sake of convenience, the complete likelihood function $L(\theta)$ of $\theta = (\beta_1, ..., \beta_g, \Sigma_1, ..., \Sigma_g, \pi_1, ..., \pi_g)$ can be written as

$$\prod_{i=1}^{n}\prod_{j=1}^{g}\left\{\left(\frac{2\pi_j}{(2\pi)^{p/2}|\Sigma_j|^{1/2}}\right)\left[\frac{Q_{ij}}{2}\right]^{1/2-p/4} K_{p/2-1}\left(\sqrt{2Q_{ij}}\right)\right\}^{G_{ij}}.$$

Based on (iii) in Lemma 1.3.1, similar to the discussion for the case of $g = 1$, for each (X_i, Y_i), if we can further observe V_i, $i = 1, 2, .., n$, then the complete log-likelihood function of θ, the collection of all unknown parameters, will be given by

$$\begin{aligned} L(\theta) = & \sum_{i=1}^{n}\sum_{j=1}^{g} G_{ij}\log\pi_j - \frac{p}{2}\sum_{i=1}^{n}\sum_{j=1}^{g} G_{ij}\log(2\pi V_i) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{g} G_{ij}\log|\Sigma_j| \\ & - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{g}\frac{G_{ij}}{V_i}Q_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{g} G_{ij}V_i. \end{aligned}$$

With the initial values for $\theta^{(0)} = (\pi^{(0)}, \beta^{(0)}, \Sigma^{(0)})$, we have to calculate

$$\begin{aligned} E[L(\theta)|\theta^{(0)}, \mathbf{D}] = & \sum_{i=1}^{n}\sum_{j=1}^{g}\tau_{ij}\log\pi_j - \frac{p}{2}\sum_{i=1}^{n}\sum_{j=1}^{g} E[G_{ij}\log 2\pi V_i|\theta^{(0)}, \mathbf{D}] \\ & - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{g}\tau_{ij}\log|\Sigma_j| - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{g} E\left[\frac{G_{ij}}{V_i}Q_{ij}\middle|\,\theta^{(0)}, \mathbf{D}\right] - \sum_{i=1}^{n}\sum_{j=1}^{g} E(G_{ij}|\theta^{(0)}, \mathbf{D}), \end{aligned}$$

where we use $\mathbf{D}$ to denote the complete data set for the sake of brevity. The above conditional expectation can be further written as

$$\sum_{i=1}^{n}\sum_{j=1}^{g}\tau_{ij}\log\pi_j - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{g}\tau_{ij}\log|\Sigma_j| - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{g}\tau_{ij}\delta_{ij}Q_{ij} + \mathbf{R_n}, \tag{2.2.2}$$

where

$$\tau_{ij} = E[G_{ij}|\theta^{(0)}, \mathbf{D}],\ \ \delta_{ij} = E\left[\frac{1}{V_i}\middle|\theta^{(0)}, \mathbf{D}, G_{ij} = 1\right],$$

and the reminder term $\mathbf{R_n}$ does not depend on the unknown parameters. In fact, we have

$$E\left[\frac{G_{ij}}{V_i}\middle|\theta^{(0)}, \mathbf{D}\right] = E\left[\frac{1}{V_i}\middle|\theta^{(0)}, \mathbf{D}, G_{ij} = 1\right] P(G_{ij} = 1|\theta^{(0)}, \mathbf{D}) = \tau_{ij}\delta_{ij}.$$

Denote $Q_{ij}^{(0)} = Q(Y_i; X_i'\beta_j^{(0)}, \Sigma_j^{(0)})$, we know that

$$\delta_{ij} = \frac{\sqrt{2}K_{-p/2}\left(\sqrt{2Q_{ij}^{(0)}}\right)}{\sqrt{Q_{ij}^{(0)}}K_{1-p/2}\left(\sqrt{2Q_{ij}^{(0)}}\right)}.$$

One can further show that , by applying Bayesian formula,

$$\begin{aligned}\tau_{ij} &= P(G_{ij} = 1|\theta^{(0)}, \mathbf{D}) \\ &= \frac{\pi_j^{(0)}|\Sigma_j|^{-1/2}[Q_{ij}^{(0)}]^{1/2-p/4}K_{p/2-1}\left(\sqrt{2Q_{ij}^{(0)}}\right)}{\sum_{l=1}^{g}\pi_l^{(0)}|\Sigma_l|^{-1/2}[Q_{il}^{(0)}]^{1/2-p/4}K_{p/2-1}\left(\sqrt{2Q_{il}^{(0)}}\right)}.\end{aligned}$$

Base on the above discussion, the EM algorithm for estimating θ is as follows:

EM Algorithm:

(1) Choosing initial values for β, Σ, π, say $\beta^{(0)}, \Sigma^{(0)}, \pi^{(0)}$; then at the $k+1$-th iteration,

(2) E-Step: Calculate $\tau_{ij}^{(k+1)}$, $\delta_{ij}^{(k+1)}$ from above equations with (0) replaced by (k);

(3) M-Step: Update β, Σ, π with

$$
\begin{aligned}
\pi_j^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}\tau_{ij}^{(k+1)}, \\
\beta_j^{(k+1)} &= (\mathbf{X}'\mathbf{W}_j\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}_j\mathbf{Y}), \\
\Sigma_j^{(k+1)} &= \frac{\mathbf{Y}'(\mathbf{W}_j - \mathbf{W}_j\mathbf{X}(\mathbf{X}'\mathbf{W}_j\mathbf{X})^{-1}\mathbf{X}\mathbf{W}_j)\mathbf{Y}}{\sum_{i=1}^{n}\tau_{ij}^{(k+1)}}.
\end{aligned}
$$

where $\mathbf{W}_j = \text{diag}(\tau_{1j}^{(k+1)}\delta_{1j}^{(k+1)}, \tau_{2j}^{(k+1)}\delta_{2j}^{(k+1)}, \ldots, \tau_{nj}^{(k+1)}\delta_{nj}^{(k+1)})$.

(4) Repeat (2) and (3) until certain convergence criterion is met.

The ascent property is a very important characteristic possessed by the EM algorithm in parametric models. It implies that after each iteration, the likelihood at the newly updated estimate is no less than the likelihoods at the previous estimates. In the following, we will present a theorem which states that the proposed EM algorithm in the current context also has this desired property.

Theorem 1. *Let $\theta^{(k)}$ denote the estimate of θ in the k-th iteration of the EM algorithm, then for any n,*

$$L_n(\theta^{(k+1)}) \geq L_n(\theta^{(k)}),$$

where $L_n(\theta)$ is defined in (2.1.3).

The main proof of Theorem 1 is similar to the proof of Theorem 3 in **?**. However, a nontrivial modification is needed to accommodate both latent variables G and V. If we further assume that all Σ_j's are equal, then the common covariance matrix can be estimated by

$$\Sigma^{(k+1)} = n^{-1}\sum_{j=1}^{g}\mathbf{Y}'(\mathbf{W}_j - \mathbf{W}_j\mathbf{X}(\mathbf{X}'\mathbf{W}_j\mathbf{X})^{-1}\mathbf{X}\mathbf{W}_j)\mathbf{Y}.$$

One can easily check that if Y is one-dimensional, then all the formulae listed in the M-step of the above EM algorithm are exactly the same as in **?**.

As we mentioned in section 1.1, the robustness of the EM procedure developed in **?** is well aligned with the close connection between the MLE of the regression coefficients when the error term has a Laplace distribution and the LAD regression, this is also true for the EM procedure we developed above for the multivariate case. Also, from the M-step, we can see that the estimate of the regression coefficients β_j's indeed is a weighted least squares estimate, and the factor $\delta_{ij}^{(k+1)}$ from the weights $w_{ij}^{(k+1)}$ depends on the $Q_{ij}^{(k)} = (Y_i - (\beta_j^{(k)})'X_i)'(\Sigma_j^{(k)})^{-1}(Y_i - (\beta_j^{(k)})'X_i)$ in a rather complicated way. However, for each i, j, $\delta_{ij}^{(k+1)}$ is indeed a decreasing function of $Q_{ij}^{(k)}$, which indicates that similar to the scalar response case discussed in **?**, less weights will be received for those observations with larger residuals in the estimation procedure, which guarantees the robustness of the proposed EM algorithm.

Note that when $Q_{ij} = 0$, δ_{ij} will be infinite. This creates some difficulties when we program since very big value of δ_{ij} would make the computation very unstable. For $g = 1$ and scalar response case, Phillips (2002) noticed that this problem rarely arises, but this does occur often in our case. Similar to **?**, in our simulation study, we adopt a hard threshold rule to control the effect of extremely small Q_{ij} values in each iteration step. Under this rule, $\delta_{ij}^{(k+1)}$ will be assigned a value of 10^6 if the corresponding Q_{ij} equals 0. To see the effects of different choices of the threshold values, we also tried other threshold values, such as $10^8, 10^{10}$, and all these choices produce similar results. Therefore, only the results for 10^6 are reported. Note that numerical instability could also occur if the weights are very small, to deal with this, we use the another hard threshold rule on the value of τ_{ij}, if $\tau_{ij}^{k+1} > 10^{-6}$, then τ_{ij}^{k+1} itself will be used for the next iteration; otherwise, 10^{-6} will be used as the weight for the next iteration. Same technique is used in **?** and **?**. For more deep discussion on this issue in the case of $g = 1$, see **?**.

To conclude this section, we would like to point out that the proposed EM algorithm based on the multivariate Laplace distribution is robust against outliers along the y-direction, but not in the x-direction. Therefore, certain modification is needed to equip the proposed method with some robustness against the outliers in x-direction. Here we recommend a pre-screening method. That is, exclude the observations which is deemed to be an outlier in x-direction before applying the proposed EM algorithm. For this purpose, we first calculate the leverage value for each observation using the formula $h_{jj} = n^{-1} + (n-1)^{-1}MD_j$, where $MD_j = (X_j - \bar{X})'S^{-1}(X_j - \bar{X})$, $\bar{X}$, S are the sample mean and sample covariance matrix of X_j's, respectively. The j-th observation will be identified as a high leverage point if $h_{jj} > 2p/n$, where p is the dimension of X. Some robust estimation of the population mean and covariance matrix of X can be used instead of the sample mean and sample covariance. For example, the minimum covariance determinant (MCD) estimators developed in **?**, and the Stahel-Donoho (SD) estimator from **?** and **?**. More discussion on this matter can be found in **?** and **?**.

2.3 Simulation Studies

To evaluate the performance of the proposed robust estimation procedure, we conduct some simulation studies in this section. In the first simulation, a comparison study is made between the prosed method and the MCD-based robust multivariate regression procedure discussed in **?**. Note that this study is done only for the non-mixture case, due to the MCD-based robust estimation procedure does not have a clearly workable extension to the mixture cases. In the second simulation, a case of $g > 1$ will be considered. We shall compare the proposed method with other two methods, the traditional MLE assuming the error has a multivariate normal density and the robust mixture regression model based on the multivariate t distribution.

2.3.1 Simulation 1: $g = 1$ case

Among many robust multivariate regression procedures, the one based on the MCD has been enjoyed great popularity since its introduction by **?**. To be specific, let $Z_i = (Y_i, X_i)'$, and $\mathbf{Z}_n = (Z_1, ..., Z_n)$. The MCD regression first looks for the subset $(Z_{i_1}, ..., Z_{i_h})$ of size h of $\mathbf{Z}_n$ whose covariance matrix has the smallest determinant, then the usual least squared estimation procedure (mcdLSE) is applied to the selected subset to obtain the estiamtes for the regression coefficients. Common choices of h are $h \approx n/2$ or $3n/4$. To increase the efficiency, **?** proposed three reweighted least squared estimation procedures by reweighting the location (mcdLoc), the regression (mcdReg), both the location and regression (mcdLR).

In the simulation, the data are generated from the multivariate regression models $Y = \beta' X + \epsilon$, where β is a $q \times p = 4 \times 10$ matrix with entries randomly generated from a uniform distribution on $[0, 10]$, X follows a $4-$dimensional multivariate normal distribution with mean 0 and identity covariance matrix. The regression errors ϵ are chosen from 6 different distributions: (a) the multivariate standard normal; (b) the multivariate Laplace distribution with identity covariance matrix; (c) the multivariate t distribution with degrees of freedom 1; (d) the multivariate t distribution with degrees of freedom 3; (e) the normal mixture $0.95N(0, I) + 0.05N(0, 50I)$, and (f) a multivariate normal with 5% $x-$direction high leverage outliers, all $x-$values being 10 and all $y-$values 2.

Case (a) is often used to evaluate the efficiency of different estimation methods compared to the traditional MLE when error is exactly multivariate normally distributed and there are no outliers. Under case (b), the proposed estimation procedure will provide the MLE of unknown parameters, which, as in the first case, would serve as a baseline to evaluate the performance of other estimation procedures. Both case (c) and (d) are heavy tailed distributions and are often used in the literature to mimic the outlier situations. Case (e) would produce 5% low leverage outliers, and in case (f) 5% of the observations are replicated serving as the high leverage outliers, which will be used to check the robustness of estimation procedures against the high leverage outliers.

The sample size of $n = 200$ is used in the simulation study. For each case, the simulation is repeated 200 times. the average L_2-norm of the differences of the estimated β values from their true values are used as the criterion to evaluate the performance of the proposed (L-EM) and MCD based estimation procedures. Table 2.1 is a summary of the simulation results.

Clearly, one can see the proposed estimation procedure performs better than the MCD-based estimation procedures for all chosen scenarios (a)-(e), and the MLE based on the

Error	Normal	mcdLSE	mcdLoc	mcdReg	mcdLR	L-EM
(a)	0.2060	0.2990	0.2424	0.5472	0.3102	0.2517
(b)	0.4097	0.4487	0.3920	0.6791	0.4457	0.3358
(c)	40868.81	0.8786	0.8227	1.0766	0.8419	0.8012
(d)	40869.43	1.2239	1.1637	1.4826	1.2001	1.1172
(e)	40870.17	1.5287	1.4184	2.0247	1.5314	1.4045
(f)	41888.27	1.8193	1.6604	2.5675	1.8476	123.5556

Table 2.1: Simulation 1 results for $g = 1$.

normal distribution is not resistant to the outliers at all. However, the worse performance in (f) indicates that the proposed estimate is not robust to the outliers in $x-$directions.

2.3.2 Simulation 2: $g > 1$ case

For convenience, we will denote N-EM the EM algorithm based on Normal distribution, t-EM the EM algorithm based on t distribution and L-EM the EM algorithm based on Laplace distribution. To implement the $t-$EM procedure, the profile likelihood method discussed in **?** is adopted to determine the proper degrees of freedom. The data are generated from the mixture of multivariate linear regression models with $g = 2$: $Y = \beta_1'X + \epsilon_1$ if $G = 1$ and $Y = \beta_2' + \epsilon_2$ if $G = 2$, where G is a component indicator of Y with $P(G = 1) = 0.25$. The true regression coefficients are chosen to be

$$\beta_1 = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 3 \end{pmatrix}, \ \beta_2 = -\beta_1.$$

The covariance $X \in \mathcal{R}^2$ are generated from $N_2(0, I_{2\times2})$, the random errors ϵ_1 and ϵ_2 have the same distribution as ϵ. We will consider the following six error distributions: (a) $\epsilon \sim N(0, I_{3\times3})$; (b) $\epsilon \sim$ 3$-$dimensional Laplace distribution with mean 0 and identity covariance matrix; (c) $\epsilon \sim t_1$, the 3$-$dimensional t distribution with 1 degrees of freedom, denoted as $MT_3(1)$; (d) $\epsilon \sim t_3$, the 3$-$dimensional t distribution with 3 degrees of freedom, denoted as $MT_3(3)$; (e) $\epsilon \sim 0.95N(0, I_{3\times3}) + 0.05N(0, 50I_{3\times3})$; (f) $\epsilon \sim N(0, I_{3\times3})$ with 5% high leverage outliers in both $x-$ and $y-$ directions ($X_1 = X_2 = 5, Y = 100$); and (g) $\epsilon \sim N(0, I_{3\times3})$ with

5% high leverage outliers only in $x-$direction ($X_1 = X_2 = 5, Y = 0$).

The sample size of $n = 100$ is used in the simulation study. For each case, the simulation is repeated 200 times. Same criterion as in simulation 1 are used as the criterion to evaluate the performance of various estimation procedures, except that this is done separately for π, $(\beta_{11}, \beta_{12}, \beta_{13})$ and $(\beta_{21}, \beta_{22}, \beta_{23})$. The simulation results are summarized in Table 2.2.

From the simulation results, we can see that if the true distribution of ϵ is normal, the MSEs of traditional MLE procedure are slightly smaller than two robust estimation procedures, which indicates the proposed estimation procedure and the procedure based on the multivariate t distribution are as efficient as the traditional MLE. For other cases when the distribution of ϵ has heavier tail or there are high leverage outliers in the data set, traditional MLE fails to provide reasonable estimates. The robust estimation via multivariate t distribution performs well, except when high leverage outliers are present in the data set. The computation of robust multivariate t distribution is intensive due to the estimation of degrees of freedom parameters. The simulation results clearly show that the proposed method in the paper outperforms or is at least comparable to any other methods except for some scenarios, for example, when ϵ has a lighter tail, the MSEs of proposed method are slightly larger than the traditional MLE method. However, when the ϵ has a heavier tail, the MSEs of proposed method are comparable to robust multivariate t distribution, and when the high leverage outliers are present in both directions, the proposed method outperforms any other methods. It is also clear that the proposed method and the t-procedures is not very robust when the outliers appear in the x-direction.

In summary, the simulation results indicate that the performance of the proposed robust estimation procedure is, in most of cases, comparable to the $t-$procedure. However, the extra step for finding a proper degrees of freedom makes the $t-$procedure is more computationally extensive than the proposed estimation procedure. Also, when $p = 1$, the natural connection between the LAD (least absolute deviation) estimate and the MLE based on Laplace distributions appears more attractive. That said, we do not intend to say the proposed robust estimation procedure is better than the $t-$procedure in all aspects. In fact, the extra degrees of freedom might provide $t-$procedure an extra adaptivity to the data. Except for propos-

Error	N-EM	L-EM	t-EM
(a)	(0.002, 0.027, 0.015)	(0.003, 0.056, 0.022)	(0.003, 0.042, 0.022)
(b)	(0.002, 0.082, 0.030)	(0.002, 0.020, 0.022)	(0.002, 0.032, 0.014)
(c)	(0.034, 5.054, 2.604)	(0.004, 0.190, 0.039)	(0.004, 0.062, 0.026)
(d)	(0.004, 0.162, 0.077)	(0.003, 0.030, 0.042)	(0.003, 0.033, 0.019)
(e)	(0.003, 0.336, 0.176)	(0.002, 0.043, 0.045)	(0.002, 0.032, 0.029)
(f)	(0.031, 49.362, 3.645)	(0.004, 0.073, 0.196)	(0.017, 11.409, 0.260)
(g)	(0.009, 12.870, 0.075)	(0.006, 2.478, 0.992)	(0.004, 7.409, 0.016)

Table 2.2: Simulation 2 results for $g > 1$.

ing a computationally efficient robust estimation procedure for the mixtures of multivariate linear regression, the significance of this paper is to provide another alternative to robustly estimate the regression parameters in such models. In real application, collectively using all the available robust estimation methods might provide us more accurate information on the data structures.

Chapter 3

Robust Mixture of Linear Mixed Models by Multivariate Laplace Distribution

In last chapter, we have shown that the estimation procedure based on multivariate Laplace distribution indeed possesses certain robustness in multivariate mixture regression models. In this chapter, we will extend the similar methodology to the mixture of linear mixed models. Section 3.1 will introduce the mixture of linear mixed models with multivariate Laplace distribution. An EM algorithm will be constructed in Section 3.2, together with a theoretical result on the ascent property of the proposed EM algorithm. The proof of the ascent property of the proposed EM algorithm is postponed to Appendix B. Finite sample performance of the proposed robust models will be evaluated through simulation studies , as well as a sensitivity study using the proposed method on a real data set in Section 3.3.

3.1 Mixture Linear Mixed Models with Multivariate Laplace Distribution

Instead of assuming the random components to have multivariate t-distribution, we assume that given $C_{ij} = 1$, the joint distribution of (Y_i, b_{ij}) is

$$\left.\begin{pmatrix} Y_i \\ b_{ij} \end{pmatrix}\right|_{C_{ij}=1} \sim ML_{n_i+q}\left(\begin{pmatrix} X_i\beta_j \\ 0 \end{pmatrix}, \begin{pmatrix} Z_i\Phi_j Z_i^T + \Sigma_{ij} & Z_i\Phi_j \\ \Phi_j Z_i^T & \Phi_j \end{pmatrix}\right).$$

For simplicity, we assume the error covariance has a diagonal form $\Sigma_{ij} = \sigma_j^2 I_i$, for $i = 1, ..., m$, $j = 1, ..., G$, and I_i is identity matrix throughout the paper. The marginal distribution of Y_i thus can be written as

$$Y_i \sim \sum_{j=1}^{G} p_j ML_{n_i}(X_i\beta_j, Z_i\Phi_j Z_i^T + \Sigma_{ij}) \tag{3.1.1}$$

and the log-likelihood function for given observed data is

$$L_m(\theta) = \sum_{i=1}^{m} \ln\left\{\sum_{j=1}^{G} p_j ML_{n_i}(Y_i - X_i\beta_j, Z_i\Phi_j Z_i^T + \Sigma_{ij})\right\}, \tag{3.1.2}$$

where θ denotes the collection of all unknown parameters.

Directly maximizing the above log-likelihood function is infeasible due to its untractable form. In the following, an EM algorithm is pursued to obtain the MLE, which is made possible by the fact that the multivariate Laplace distribution is a scale mixture of multivariate normal distribution. In fact, from the discussion in section 1.3, for each pair (i, j), $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, G$ the following hierarchical model leads to a marginal multivariate Laplace distribution of Y_i as in (3.1.1).

$$Y_i|b_{ij}, V_{ij} \sim \sum_{j=1}^{G} p_j N(X_i\beta_j + Z_i b_{ij}, V_{ij}\Sigma_{ij}),$$
$$b_{ij}|V_{ij} \sim N(0, V_{ij}\Phi_j), \quad V_{ij} \sim g(v) = e^{-v} I(v > 0). \tag{3.1.3}$$

Imposing multivariate Laplace distributions on random effects and regression errors simultaneously, the robustness can be achieved at both levels of the within subjects errors and between subjects errors. Similar to the t procedure in **?**, we assume that $b_{ij} \sim ML_q(0, \Phi_j)$ and $\epsilon_{ij} \sim ML_{n_i}(0, \Sigma_{ij})$. In the meanwhile, we also assume that given V_{ij}, b_{ij} is independent of ϵ_{ij}, then the above hierarchical structure can be expressed as the conventional mixed effects model $Y_i = X_i\beta_j + Z_i b_{ij} + \epsilon_{ij}$ for $i = 1, ..., m$, $j = 1, \ldots, G$. On the other hand, we also can obtain a two level hierarchical structure

$$Y_i|V_{ij} \sim \sum_{j=1}^{G} p_j N(X_i\beta_j, V_{ij}(Z_i\Phi_j Z_i^T + \Sigma_{ij})), \quad V_{ij} \sim g(v) = e^{-v} I(v > 0).$$

3.2 EM Algorithm for Robust Mixture Linear Mixed Models

The complexity of the log-likelihood function (3.1.2) makes it hard to maximize directly. In this section, we shall develop an efficient EM algorithm to obtain the MLE. The EM algorithm is made possible by utilizing the information from three missing components, the latent class variable C_{ij}, the missing scale variable V_{ij} and the random effects b_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, G$, as well as the hierarchical structure (3.1.3).

For convenience, denote $\mathbf{D}$ as the complete data set including all X_i, Y_i, Z_i, b_{ij}, C_{ij}, V_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, G$. Here we treat $\{b_{ij}, V_{ij}, C_{ij}\}$ as missing data. From the hierarchical model (3.1.3), the complete likelihood function of $\Theta = \big\{(p_j, \beta_j, \sigma_j^2, \Phi_j), j = 1, ..., G\big\}$ based on $\mathbf{D}$ can be written as

$$\prod_{i=1}^{m}\prod_{j=1}^{G}\left\{p_j f(Y_i; X_i\beta_j + Z_i b_{ij}, V_{ij}\Sigma_{ij}) f(b_{ij}; 0, V_{ij}\Phi_j) g(V_{ij})\right\}^{C_{ij}}.$$

Therefore the complete log-likelihood function is

$$\begin{aligned}
L(\Theta|\mathbf{D}) =& \sum_{i=1}^{m}\sum_{j=1}^{G} C_{ij}\ln(p_j) + \sum_{i=1}^{m}\sum_{j=1}^{G} C_{ij}\left\{-\frac{1}{2}\ln|V_{ij}\sigma_j^2 I_i| - \frac{1}{2}U_{ij}^T(V_{ij}\sigma_j^2 I_i)^{-1}U_{ij}\right\}\\
&+ \sum_{i=1}^{m}\sum_{j=1}^{G} C_{ij}\left\{-\frac{1}{2}\ln|V_{ij}\Phi_j| - \frac{1}{2}b_{ij}^T(V_{ij}\Phi_j)^{-1}b_{ij}\right\} + R_n\\
=& L_1(p|\mathbf{D}) + L_2(\beta,\sigma^2|\mathbf{D}) + L_3(\Phi|\mathbf{D}) + R_n,
\end{aligned} \tag{3.2.1}$$

where $U_{ij} = Y_i - X_i\beta_j - Z_i b_{ij}$, R_n is the collection of all terms which do not involve unknown parameters and hence plays no role in the subsequent EM procedure, and a little bit of abuse of notations, p, β, σ^2, Φ are the collections of all corresponding parameters, respectively. After some simple algebra, we can rewrite $L_2(\beta,\sigma^2|\mathbf{D})$ and $L_3(\Phi|\mathbf{D})$ as

$$\begin{aligned}
L_2(\beta,\sigma^2|\mathbf{D}) =& -\sum_{i=1}^{m}\sum_{j=1}^{G}\frac{n_i C_{ij}}{2}\ln(\sigma_j^2)\\
&-\sum_{i=1}^{m}\sum_{j=1}^{G}\frac{C_{ij}}{2V_{ij}\sigma_j^2}\mathrm{tr}\left\{(Y_i - Z_i b_{ij})(Y_i - Z_i b_{ij})^T\right\}\\
&+\sum_{i=1}^{m}\sum_{j=1}^{G}\frac{C_{ij}}{V_{ij}\sigma_j^2}\beta_j^T X_i^T(Y_i - Z_i b_{ij}) - \sum_{i=1}^{m}\sum_{j=1}^{G}\frac{C_{ij}}{2V_{ij}\sigma_j^2}\beta_j^T X_i^T X_i\beta_j,\\
L_3(\Phi|\mathbf{D}) =& -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{G} C_{ij}\ln|\Phi_j| - \frac{1}{2}\mathrm{tr}\left\{\Phi_j^{-1}\sum_{i=1}^{m}\sum_{j=1}^{G} C_{ij}V_{ij}^{-1}b_{ij}b_{ij}^T\right\}.
\end{aligned}$$

By the EM algorithm protocol, we have to derive the conditional distribution of (3.2.1) given the observed data set on Y, X, Z and initial estimates for all unknown parameters. For this purpose, denote $\Theta^{(0)}$ as the initial estimate of Θ, and we have to find the following expectations

$$p_{ij}^{(1)} = E(C_{ij} = 1 | \Theta = \Theta^{(0)}, Y), \tag{3.2.2}$$

$$v_{ij}^{(1)} = E\left(\frac{1}{V_{ij}} \middle| \Theta = \Theta^{(0)}, Y, C_{ij} = 1 \right), \tag{3.2.3}$$

$$b_{ij}^{(1)} = E(b_{ij} | \Theta = \Theta^{(0)}, Y, C_{ij}, V_{ij}), \tag{3.2.4}$$

$$\Omega_{ij}^{(1)} = \frac{1}{V_{ij}} \mathrm{Cov}(b_{ij} | \Theta = \Theta^{(0)}, Y, C_{ij}, V_{ij}). \tag{3.2.5}$$

Denote

$$Q_{ij}^{(0)} = (Y_i - X_i \beta_j^{(0)})^T (Z_i \Phi_j^{(0)} Z_i^T + \Sigma_{ij}^{(0)})^{-1} (Y_i - X_i \beta_j^{(0)}).$$

(3.2.2) can be calculated using the Bayesian formula,

$$p_{ij}^{(1)} = \frac{p_j^{(0)} |Z_i \Phi_j^{(0)} Z_i^T + \Sigma_{ij}^{(0)}|^{-\frac{1}{2}} (Q_{ij}^{(0)})^{\frac{1}{2} - \frac{n_i}{4}} K_{(n_i/2-1)} \left(\sqrt{2 Q_{ij}^{(0)}} \right)}{\sum_{j=1}^G p_j^{(0)} |Z_i \Phi_j^{(0)} Z_i^T + \Sigma_{ij}^{(0)}|^{-\frac{1}{2}} (Q_{ij}^{(0)})^{\frac{1}{2} - \frac{n_i}{4}} K_{(n_i/2-1)} \left(\sqrt{2 Q_{ij}^{(0)}} \right)}. \tag{3.2.6}$$

The calculation of (3.2.3) follows the similar argument as in **?**, and

$$v_{ij}^{(1)} = \sqrt{2} K_{-n_i/2} \left(\sqrt{2 Q_{ij}^{(0)}} \right) \Big/ \sqrt{Q_{ij}^{(0)}} K_{1-n_i/2} \left(\sqrt{2 Q_{ij}^{(0)}} \right).$$

Based on the hierarchical structure (3.1.1), given $C_{ij} = 1$ and V_{ij}, (b_{ij}, Y_i) are jointly normal, so given $Y_i, C_{ij} = 1, V_{ij}$, b_{ij} follows a q-dimensional normal distribution with mean $\Gamma_{ij}(Y_i - X_i \beta_j)$, and covariance matrix $V_{ij}(\Phi_j - \Gamma_{ij} Z_i \Phi_j))$, where $\Gamma_{ij} = \Phi_j Z_i^T (Z_i \Phi_j Z_i^T + \sigma_j^2 I_i)^{-1}$, therefore,

$$\begin{aligned} b_{ij}^{(1)} &= \Phi_j^{(0)} Z_i^T (Z_i \Phi_j^{(0)} Z_i^T + \sigma_j^2 I_i)^{-1} (Y_i - X_i \beta_j^{(0)}) \\ &= (\Phi_j^{-1(0)} + \sigma_j^{-2(0)} Z_i^T I_i^{-1} Z_i)^{-1} \sigma_j^{-2(0)} Z_i^T I_i^{-1} (Y_i - X_i \beta_j^{(0)}), \end{aligned} \tag{3.2.7}$$

$$
\begin{aligned}
\Omega_{ij}^{(1)} &= \Phi_j^{(0)} - \Phi_j^{(0)} Z_i^T (Z_i \Phi_j^{(0)} Z_i^T + \sigma_j^{2(0)} I_i)^{-1} Z_i \Phi_j^{(0)} \\
&= \left(\Phi_j^{-1(0)} + \sigma_j^{-2(0)} Z_i^T I_i^{-1} Z_i \right)^{-1}. \qquad (3.2.8)
\end{aligned}
$$

Also, we can show that

$$
E\left[C_{ij}/V_{ij} | \, \Theta = \Theta^{(0)}, Y\right] = E\left[I(C_{ij} = 1)/V_{ij} | \, \Theta = \Theta^{(0)}, Y\right]
$$

$$
= E\left[1/V_{ij} | \, \Theta = \Theta^{(0)}, Y, C_{ij} = 1\right] P(C_{ij} = 1 | \Theta = \Theta^{(0)}, Y) = v_{ij}^{(1)} p_{ij}^{(1)}.
$$

Based on all the results above, given the observed data Y and the initial estimate $\Theta^{(0)}$, the conditional expectations of $L_1(p|\mathbf{D}), L_2(\beta, \sigma^2|\mathbf{D})$ and $L_3(\Phi|\mathbf{D})$ can be calculated as follows

$$
E(L_1(p|\mathbf{D})|Y, \Theta^{(0)}) = \sum_{i=1}^{m} \sum_{j=1}^{G} p_{ij}^{(1)} \ln(p_j), \qquad (3.2.9)
$$

$$
\begin{aligned}
E(L_2(\beta, \sigma^2|\mathbf{D})|Y, \Theta^{(0)}) &= -\sum_{i=1}^{m} \sum_{j=1}^{G} p_{ij}^{(1)} \frac{n_i}{2} \ln(\sigma_j^2) \\
&- \sum_{i=1}^{m} \sum_{j=1}^{G} p_{ij}^{(1)} \frac{1}{2\sigma_j^2} \mathrm{tr} \left[\left\{ Z_i \Omega_{ij}^{(1)} Z_i^T + v_{ij}^{(1)} (Y_i - Z_i b_{ij}^{(1)})(Y_i - Z_i b_{ij}^{(1)})^T \right\} \right] \\
&+ \sum_{i=1}^{m} \sum_{j=1}^{G} p_{ij}^{(1)} \frac{1}{\sigma_j^2} v_{ij}^{(1)} \beta_j^T X_i^T (Y_i - Z_i b_{ij}^{(1)}) - \sum_{i=1}^{m} \sum_{j=1}^{G} p_{ij}^{(1)} \frac{1}{2\sigma_j^2} v_{ij}^{(1)} \beta_j^T X_i^T X_i \beta_j, \qquad (3.2.10)
\end{aligned}
$$

and

$$
\begin{aligned}
&E(L_3(\Phi|\mathbf{D})|Y, \Theta^{(0)}) \\
&= -\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{G} p_{ij}^{(1)} \ln |\Phi_j| - \frac{1}{2} \mathrm{tr} \left\{ \Phi_j^{-1} \sum_{i=1}^{m} \sum_{j=1}^{G} p_{ij}^{(1)} \left(v_{ij}^{(1)} b_{ij}^{(1)} b_{ij}^{(1)T} + \Omega_{ij}^{(1)} \right) \right\}. \qquad (3.2.11)
\end{aligned}
$$

Since $L_1(p|\mathbf{D}), L_2(\beta, \sigma^2|\mathbf{D})$ and $L_3(\Phi|\mathbf{D})$ involve separate sets of unknown parameters, thus to maximize the conditional expectation of the complete log-likelihood function is amount to maximize the three L-functions with respect to their own unknown parameters. Thus, we have the following EM algorithm.

EM Algorithm

(I) Start with an initial value for $\Theta^{(0)} = (p_j^{(0)}, \beta_j^{(0)}, \Phi_j^{(0)}, \sigma_j^{2(0)})$, for $j = 1, ..., G$.

(II) *E-Step:* In the $(k+1)$-th iteration, calculate $p_{ij}^{(k+1)}$, $v_{ij}^{(k+1)}$, $b_{ij}^{(k+1)}$ and $\Omega_{ij}^{(k+1)}$ by using (3.2.2), (3.2.3), (3.2.4) and (3.2.5), for $i = 1, ..., m$ and $j = 1, ..., G$, with (0) replaced by (k), and (1) replaced by $(k+1)$. Subsequently, the constructions of conditional expected complete log-likelihood functions are based on (3.2.9), (3.2.10) and (3.2.11).

(III) *M-Step:* Maximize the expected complete log-likelihood functions, or $E(L_1(p|\mathbf{D})|Y, \Theta^{(k)})$, $E(L_2(\beta, \sigma^2|\mathbf{D})|Y, \Theta^{(k)})$ and $E(L_3(\Phi|\mathbf{D})|Y, \Theta^{(k)})$ with respect to their own unknown parameters, resulting the following formulas:

(i) Maximizing $E(L_1(p|\mathbf{D})|Y, \Theta^{(k)})$ with respect to p leads to $p_j^{(k+1)}$

$$p_j^{(k+1)} = \frac{1}{m}\sum_{i=1}^{m} p_{ij}^{(k)}, \quad j = 1, \ldots, G;$$

(ii) Maximizing $E(L_2(\beta, \sigma^2|\mathbf{D})|Y, \Theta^{(k)})$ with respect to β, σ^2 gives

$$\beta_j^{(k+1)} = \left(\sum_{i=1}^{m} p_{ij}^{(k)} v_{ij}^{(k)} X_i^T X_i\right)^{-1} \sum_{i=1}^{m} p_{ij}^{(k)} v_{ij}^{(k)} X_i^T (Y_i - Z_i b_{ij}^{(k)})$$

and letting $U_{ij}^{(k)} = Y_i - X_i\beta_j^{(k+1)} - Z_i b_{ij}^{(k)}$, σ^2 is updated by

$$\sigma_j^{2(k+1)} = \frac{\sum_{i=1}^{m} p_{ij}^{(k)} \left[v_{ij}^{(k)} U_{ij}^{(k)T} U_{ij}^{(k)} + \mathrm{tr}\left(\Omega_{ij}^{(k)} Z_i^T Z_i\right)\right]}{\sum_{i=1}^{m} n_i p_{ij}^{(k)}};$$

(iii) Maximizing $E(L_3(\Phi|\mathbf{D})|Y, \Theta^{(k)})$ with respect to Φ provides

$$\Phi_j^{(k+1)} = \frac{\sum_{i=1}^m p_{ij}^{(k)} \left[v_{ij}^{(k)} b_{ij}^{(k)} (b_{ij}^{(k)})^T + \Omega_{ij}^{(k)} \right]}{\sum_{i=1}^m p_{ij}^{(k)}};$$

(IV) Repeat steps (II) and (III) until convergence.

One of the desired properties should be possessed for any EM algorithms is the ascent property, that is, the values of the likelihood function are nondecreasing after each iteration. This result is summarized in the following theorem.

Theorem 2. *Let $\theta^{(k)}$ denote the estimate of θ in the k^{th} iteration of the EM algorithm, then for any m,*

$$L_m(\theta^{(k+1)}) \geqslant L_m(\theta^{(k)}),$$

where $L_m(\theta)$ is the log-likelihood function defined in (3.1.2).

As we mentioned earlier, the major difference of the proposed EM algorithm from **?** EM algorithm is that we don't have to estimate the degrees of freedom as in multivariate t-distribution. This has significant effect on the computational aspect, since it is very time-consuming to search for the right degrees of freedom over a grid of candidate values. Of course, we must acknowledge that with the degrees of the freedom, the EM algorithm based on the multivariate t-distribution does have certain flexibility to fit the data in an adaptive way.

It is well known that the choice of initial values is crucial in the EM algorithms. If the initial value is not properly chosen, then the updated estimates might converge to the boundary points where variance is small and log-likelihood function is large. Some recommendations on choosing the initial values can be found in the literature. For example, the profile likelihood method proposed in **?**, the trimmed likelihood estimates suggested in **?**, and the K-means method used in **?**. In this paper, we simply adopted some initial values around true values in the simulation studies for the sake of simplicity.

3.3 Numerical Studies

To assess the relative performance of the proposed estimation procedure to the exiting procedure based on the multivariate t-distribution, we conducted some simulation studies in this section, as well as a sensitivity study to the lung function growth study on girls in Topeka, KS.

3.3.1 Simulation Studies

The simulated data are generated from the following two-components mixture linear mixed model

$$Y_i = \begin{cases} X_i\beta_1 + Z_i b_{i1} + \epsilon_{i1}, & \text{if } C_{i1} = 1, \\ X_i\beta_2 + Z_i b_{i2} + \epsilon_{i2}, & \text{if } C_{i2} = 1. \end{cases}$$

where $i = 1, ..., m$, $\beta_1 = (1, 1, 0, 0)^T$, $\beta_2 = (0, 0, 1, 1)^T$, and $p_1 = P(C_i = 1) = 0.4$. The rows of the covariates $X_i \in \mathbb{R}^{n_i \times 4}$ are independently generated from $N_4(0, I)$. The rows of $Z_i \in \mathbb{R}^{n_i \times 2}$ are independently generated from $N_2(0, I)$. The following 4 different scenarios on the distributions of the random effects and the regression errors are considered. In all cases, Σ_{ij} is chosen to be identity matrix and Φ_j be a diagonal matrix with diagonal elements 1 and off-diagonal elements 0.5.

(1) Normal distribution: $\epsilon_{ij} \sim N_{n_i}(0, \Sigma_{ij})$, $b_{ij} \sim N_2(0, \Phi_j)$.

(2) t distribution: $\epsilon_{ij} \sim t_{n_i}(0, \Sigma_{ij}, \nu)$, $b_{ij} \sim t_2(0, \Phi_j, \nu)$ with degrees of freedom being 1, 3 and 5.

(3) Laplace distribution: $\epsilon_{ij} \sim ML_{n_i}(0, \Sigma_{ij})$, $b_{ij} \sim ML_2(0, \Phi_j)$.

(4) Contaminated normal distribution:

$$\epsilon_{ij} \sim 0.95N_{n_i}(0, I) + 0.05N_{n_i}(0, 25I), \quad b_{ij} \sim 0.95N_2(0, I) + 0.05N_2(0, 25I).$$

The random errors ϵ_{i1} and ϵ_{i2} are chosen to be independent, so are the random effects b_{i1}

and b_{i2} for $i = 1, ..., m$. Each simulation is replicated 200 times for four different cases: (1) $n_i = 8$, $m = 100$; (2) $n_i = 8$, $m = 200$; (3)$n_i = 6$, $m = 200$; (4)$n_i = 6$, $m = 400$. The initial values are chosen to be $\pi_1 = 0.5$, $\beta_1 = (0.5, 0.5, 0, 0)^T$, and $\beta_2 = (0, 0, 0.5, 0.5)^T$, and similar to **?**, the median squared errors (MedSE) and the relative efficiencies of the proposed Laplace mixture method, and the t mixture method to the conventional normal mixture mixed model are reported. The simulation results are presented in Table 3.1-3.4. The values reported in these tables are the median of standard errors based on 200 simulations, and the values in the parentheses are the relative efficiency which is defined as the ratio of the median standard errors from the normal procedure (in the numerator) and the proposed robust estimation procedure (in the denominator).

Clearly the performance of the proposed method is at least comparable to the robust t method proposed by **?**, and when the random effects and random errors have heavy tail distributions, both Laplace and t procedures are superior to the normal mixture model. Interestingly, for the case of the random terms being from t distribution with degrees of freedom one, the proposed method works even better than the t mixture method. When the random effects and random errors are normally distributed, the MedSEs from both robust procedures are slightly larger than those of normal mixtures in general, as expected. However, the superiority is not significant. When the random effects and random errors are from the contaminated normal, the t mixture procedures seems better than the Laplace mixture method, in particular, when the total sample size consisting of the number of subjects and the repeated measurements from each subjects, get bigger.

We also report the mean square errors (MSEs) for all cases in Table 3.5-3.8, and in these tables the values in the parentheses are the relative efficiencies which are defined as the ratio of the MSEs from the normal procedure (in the numerator) and the proposed robust estimation procedure (in the denominator). Similar patterns are exhibited as in Table 3.1-3.4.

3.3.2 Sensitivity Study

To evaluate how the relevant factors influence the lung function growth, a longitudinal study of air pollution and health was conducted in six cities across the USA. See **?** for a detailed description. Similar to **?**, we only focus on the data collected from 300 female participants living in Topeka, Kansas. To apply our proposed method, subjects with only one record are omitted from the model fitting, resulting a data set with 252 subjects aged between 2 and 12.

The response variable Y is chosen to be the logarithmic forced expiratory volume per second, which is critically important in the diagnosis of obstructive and restrictive diseases and is a commonly used measure of lung function from the pulmonary function tests. To model the lung growth pattern over time, the variable age, X, is treated as both the fixed effect and the random effect. Similar to **?**, a three-component mixture of linear mixed model is chosen to fit the data. In particular, for the i-th subject in the j-th component, the response variable and the covariate age are fitted with $Y_{ij} = \beta_{0j} + b_{ij0} + (\beta_{1j} + b_{ij1})X_{ij} + \varepsilon_{ij}$. For comparison purpose, in addition to the estimation results using the proposed Laplace procedure, the results from the normal mixture and t-mixture proposed in **?** are also reported, as seen in the Table 3.9. Besides the estimate of the parameters, the values in the parentheses are the bootstrap standard deviation of the corresponding estimates based on 100 bootstrap samples.

The first part of Table 3.9 contains the fitting result using the original data set from the longitudinal study. The degrees of freedom chosen by the t procedure is 28, which implies that the data are very close to normal and all three methods perform equally well. To check the robustness of the proposed procedure, we contaminated the data by adding 10 to all the response values in the first subject, and adding 10 to the first two subjects, respectively, then refit the data using the three procedures. By comparing the estimates under different cases, we can see that the estimates from both t and Laplace procedures do not vary too much, indicating that the proposed Laplace estimation procedure and the t procedure possess certain robustness. However, the normal procedure provides different estimates when outliers present. It is noticeable that in most cases the bootstrap standard deviations from the proposed procedure are smaller than those from t and normal procedures, which implies the Laplace procedure is much more stable. We also note that the robust t procedure could adjust the degrees of freedom to achieve robustness, but it takes much time to find out the right degrees of freedom to fit the data. So, comparing to the t procedure, the proposed Laplace procedure is more efficient.

Estimate	Method	t_1	t_3	t_5	Laplace	Normal	Contaminated
$\hat{\pi}_1$	N-MM	0.0057	0.0031	0.0020	0.0016	0.0011	0.0017
	t-MM	0.0045	0.0026	0.0014	0.0018	0.0011	0.0013
		(1.27)	(1.19)	(1.43)	(0.89)	(1.00)	(1.31)
	L-MM	0.0047	0.0019	0.0011	0.0017	0.0011	0.0014
		(1.21)	(1.63)	(1.82)	(0.94)	(1.00)	(1.21)
$\hat{\beta}_{11}$	N-MM	0.4067	0.1249	0.1290	0.0331	0.0018	0.0899
	t-MM	0.0089	0.0039	0.0048	0.0007	0.0014	0.0104
		(45.67)	(32.03)	(26.87)	(47.29)	(1.29)	(8.64)
	L-MM	0.0057	0.0019	0.0024	0.0003	0.0012	0.0019
		(71.35)	(65.74)	(53.75)	(110.33)	(1.50)	(47.32)
$\hat{\beta}_{21}$	N-MM	0.4354	0.1406	0.1182	0.0299	0.0033	0.0916
	t-MM	0.0088	0.0055	0.0032	0.0088	0.0030	0.0093
		(49.48)	(25.56)	(36.94)	(3.40)	(1.10)	(9.84)
	L-MM	0.0066	0.0054	0.0036	0.0005	0.0032	0.0025
		(65.97)	(26.04)	(32.83)	(59.98)	(1.03)	(36.64)
$\hat{\beta}_{31}$	N-MM	0.3614	0.0782	0.0665	0.0216	0.0026	0.0440
	t-MM	0.0050	0.0034	0.0034	0.0007	0.0037	0.0016
		(72.28)	(23.00)	(19.56)	(30.86)	(0.70)	(27.5)
	L-MM	0.0037	0.0035	0.0022	0.0005	0.0037	0.0026
		(97.68)	(22.34)	(30.23)	(43.2)	(0.70)	(16.92)
$\hat{\beta}_{41}$	N-MM	0.3068	0.0826	0.0627	0.0132	0.0017	0.0459
	t-MM	0.0037	0.0054	0.0019	0.0006	0.0016	0.0014
		(8.29)	(15.30)	(33.00)	(22.00)	(1.06)	(32.79)
	L-MM	0.0045	0.0025	0.0024	0.0005	0.0019	0.0025
		(68.18)	(33.04)	(26.13)	(26.40)	(0.89)	(18.36)
$\hat{\beta}_{12}$	N-MM	0.2036	0.0265	0.0162	0.0047	0.0015	0.0170
	t-MM	0.0028	0.0023	0.0011	0.0004	0.0018	0.0014
		(72.71)	(11.52)	(14.73)	(11.75)	(0.83)	(12.14)
	L-MM	0.0025	0.0021	0.0009	0.0004	0.0018	0.0019
		(81.44)	(12.61)	(18.00)	(11.75)	(0.83)	(8.95)
$\hat{\beta}_{22}$	N-MM	0.1762	0.0245	0.0172	0.0049	0.0012	0.0136
	t-MM	0.0035	0.0022	0.0014	0.0005	0.0012	0.0013
		(50.34)	(11.14)	(12.29)	(9.80)	(1.00)	(10.46)
	L-MM	0.0027	0.0014	0.0011	0.0003	0.0014	0.0017
		(65.26)	(17.50)	(15.64)	(16.33)	(0.86)	(8.00)
$\hat{\beta}_{32}$	N-MM	0.2350	0.0727	0.0600	0.0115	0.0015	0.0437
	t-MM	0.0035	0.0033	0.0020	0.0006	0.0020	0.0060
		(67.14)	(22.03)	(30.00)	(19.17)	(0.75)	(7.28)
	L-MM	0.0029	0.0019	0.0011	0.0004	0.0015	0.0028
		(81.03)	(38.26)	(54.55)	(28.75)	(1.00)	(15.61)
$\hat{\beta}_{42}$	N-MM	0.1655	0.0641	0.0556	0.0171	0.0015	0.0415
	t-MM	0.0030	0.0023	0.0013	0.0007	0.0015	0.0046
		(55.17)	(27.87)	(42.77)	(24.43)	(1.00)	(9.02)
	L-MM	0.0018	0.0022	0.0015	0.0005	0.0019	0.0011
		(91.94)	(29.14)	(37.07)	(34.2)	(0.79)	(37.73)

Table 3.1: MedSE: $n_i = 8$, $m = 100$

Estimate	Method	t_1	t_3	t_5	Laplace	Normal	Contaminated
$\hat{\pi}_1$	N-MM	0.0021	0.0029	0.0019	0.0007	0.0008	0.0013
	t-MM	0.0006	0.0011	0.0010	0.0006	0.0009	0.0005
		(3.50)	(2.64)	(1.90)	(1.17)	(0.89)	(2.60)
	L-MM	0.0006	0.0009	0.0008	0.0005	0.0009	0.0008
		(3.50)	(3.22)	(2.38)	(1.40)	(0.89)	(1.63)
$\hat{\beta}_{11}$	N-MM	0.1345	0.1448	0.1112	0.0426	0.0008	0.0029
	t-MM	0.0005	0.0028	0.0023	0.0003	0.0016	0.0015
		(269.00)	(51.71)	(48.35)	(142.00)	(0.50)	(1.93)
	L-MM	0.0002	0.0013	0.0014	0.0001	0.0015	0.0019
		(672.50)	(111.38)	(79.43)	(426.00)	(0.53)	(1.53)
$\hat{\beta}_{21}$	N-MM	0.1373	0.1369	0.1187	0.0530	0.0008	0.0042
	t-MM	0.0007	0.0036	0.0022	0.0003	0.0014	0.0010
		(196.14)	(38.03)	(53.95)	(176.67)	(0.57)	(4.20)
	L-MM	0.0004	0.0017	0.0013	0.0002	0.0012	0.0015
		(343.25)	(80.53)	(91.31)	(265.00)	(0.67)	(2.80)
$\hat{\beta}_{31}$	N-MM	0.0730	0.0757	0.0686	0.0186	0.0012	0.0021
	t-MM	0.0010	0.0022	0.0013	0.0004	0.0011	0.0012
		(73.00)	(34.41)	(52.77)	(46.50)	(1.09)	(1.75)
	L-MM	0.0004	0.0016	0.0014	0.0002	0.0012	0.0014
		(182.5)	(47.31)	(49.00)	(93.00)	(1.00)	(1.50)
$\hat{\beta}_{41}$	N-MM	0.0760	0.0775	0.0627	0.0211	0.0011	0.0033
	t-MM	0.0005	0.0027	0.0019	0.0003	0.0013	0.0008
		(152.00)	(28.70)	(33.00)	(70.33)	(0.85)	(4.13)
	L-MM	0.0003	0.0015	0.0018	0.0001	0.0013	0.0014
		(253.33)	(51.67)	(34.83)	(211.00)	(0.85)	(2.36)
$\hat{\beta}_{12}$	N-MM	0.0166	0.0220	0.0173	0.0055	0.0005	0.0016
	t-MM	0.0002	0.0009	0.0014	0.0001	0.0006	0.0009
		(83.00)	(24.44)	(12.36)	(55.00)	(0.83)	(1.78)
	L-MM	0.0002	0.0012	0.0010	0.0001	0.0008	0.0011
		(83.00)	(18.33)	(17.30)	(55.00)	(0.63)	(1.45)
$\hat{\beta}_{22}$	N-MM	0.0189	0.0240	0.0147	0.0045	0.0003	0.0014
	t-MM	0.0005	0.0012	0.0007	0.0002	0.0004	0.0010
		(37.80)	(20.00)	(21.00)	(22.50)	(0.75)	(1.40)
	L-MM	0.0002	0.0008	0.0008	0.0001	0.0008	0.0012
		(94.50)	(30.00)	(18.38)	(45.00)	(0.38)	(1.17)
$\hat{\beta}_{32}$	N-MM	0.0587	0.0680	0.0561	0.0242	0.0009	0.0016
	t-MM	0.0004	0.0019	0.0011	0.0002	0.0013	0.0006
		(146.75)	(35.79)	(51.00)	(121.00)	(0.69)	(2.67)
	L-MM	0.0002	0.0011	0.0007	0.0001	0.0011	0.0007
		(293.50)	(61.82)	(80.14)	(242.00)	(0.82)	(2.29)
$\hat{\beta}_{42}$	N-MM	0.0562	0.0644	0.0549	0.0237	0.0006	0.0019
	t-MM	0.0006	0.0017	0.0014	0.0003	0.0007	0.0010
		(93,67)	(37.88)	(39.21)	(79.00)	(0.86)	(1.90)
	L-MM	0.0003	0.0010	0.0007	0.0002	0.0008	0.0012
		(187.33)	(64.40)	(78.43)	(118.50)	(0.75)	(1.58)

Table 3.2: MedSE: $n_i = 8$, $m = 200$

Estimate	Method	t_1	t_3	t_5	Laplace	Normal	Contaminated
$\hat{\pi}_1$	N-MM	0.0041	0.0021	0.0010	0.0011	0.0007	0.0029
	t-MM	0.0010	0.0009	0.0006	0.0007	0.0007	0.0008
		(4.10)	(2.33)	(1.67)	(1.57)	(1.00)	(3.625)
	L-MM	0.0013	0.0012	0.0006	0.0007	0.0007	0.0011
		(3.15)	(1.75)	(1.67)	(1.57)	(1.00)	(2.64)
$\hat{\beta}_{11}$	N-MM	0.2751	0.0862	0.0053	0.0015	0.0018	0.0081
	t-MM	0.0056	0.0040	0.0018	0.0003	0.0021	0.0017
		(49.125)	(21.55)	(2.94)	(5.00)	(0.86)	(4.76)
	L-MM	0.0032	0.0025	0.0021	0.0002	0.0024	0.0020
		(85.97)	(34.48)	(2.52)	(7.50)	(0.75)	(4.05)
$\hat{\beta}_{21}$	N-MM	0.3368	0.0853	0.0026	0.0020	0.0017	0.0089
	t-MM	0.0060	0.0036	0.0023	0.0004	0.0018	0.0018
		(56.13)	(23.69)	(1.13)	(5.00)	(0.94)	(4.94)
	L-MM	0.0028	0.0025	0.0023	0.0003	0.0022	0.0024
		(120.29)	(34.12)	(1.13)	(6.67)	(0.77)	(3.71)
$\hat{\beta}_{31}$	N-MM	0.3317	0.0242	0.0060	0.0014	0.0015	0.0050
	t-MM	0.0028	0.0018	0.0022	0.0006	0.0018	0.0018
		(118.46)	(13.44)	(2.73)	(2.33)	(0.83)	(2.78)
	L-MM	0.0024	0.0021	0.0021	0.0002	0.0023	0.0023
		(138.21)	(11.52)	(2.86)	(7.00)	(0.65)	(2.17)
$\hat{\beta}_{41}$	N-MM	0.2323	0.0258	0.0044	0.0015	0.0013	0.0075
	t-MM	0.0024	0.0025	0.0014	0.0005	0.0013	0.0022
		(96.79)	(10.32)	(3.14)	(3.00)	(1.00)	(3.41)
	L-MM	0.0021	0.0027	0.0022	0.0003	0.0018	0.0024
		(110.62)	(9.56)	(2.00)	(3.00)	(0.72)	(3.13)
$\hat{\beta}_{12}$	N-MM	0.1509	0.0054	0.0023	0.0009	0.0009	0.0043
	t-MM	0.0013	0.0015	0.0011	0.0003	0.0012	0.0016
		(116.08)	(3.60)	(2.09)	(3.00)	(0.75)	(2.69)
	L-MM	0.0020	0.0015	0.0012	0.0002	0.0015	0.0017
		(75.45)	(3.60)	(1.92)	(4.50)	(0.60)	(2.53)
$\hat{\beta}_{22}$	N-MM	0.1189	0.0046	0.0017	0.0011	0.0009	0.0074
	t-MM	0.0011	0.0012	0.0008	0.0004	0.0010	0.0015
		(108.09)	(3.83)	(2.13)	(2.75)	(0.90)	(4.93)
	L-MM	0.0014	0.0016	0.0013	0.0002	0.0017	0.0010
		(84.93)	(2.88)	(1.31)	(5.50)	(0.53)	(7.40)
$\hat{\beta}_{32}$	N-MM	0.1785	0.0425	0.0020	0.0015	0.0010	0.0043
	t-MM	0.0032	0.0018	0.0014	0.0004	0.0010	0.0012
		(55.78)	(23.61)	(1.43)	(1.07)	(1.00)	(3.58)
	L-MM	0.0027	0.0013	0.0019	0.0003	0.0012	0.0016
		(66.11)	(32.69)	(1.05)	(5.00)	(0.83)	(2.69)
$\hat{\beta}_{42}$	N-MM	0.1998	0.0378	0.0022	0.0014	0.008	0.0044
	t-MM	0.0021	0.0026	0.0013	0.0005	0.0010	0.0013
		(95.14)	(14.54)	(1.69)	(2.80)	(0.80)	(3.38)
	L-MM	0.0015	0.0017	0.0010	0.0002	0.0013	0.0015
		(133.2)	(22.24)	(2.20)	(7.00)	(0.62)	(2.93)

Table 3.3: MedSE: $n_i = 6$, $m = 200$

Estimate	Method	t_1	t_3	t_5	Laplace	Normal	Contaminated
$\hat{\pi}_1$	N-MM	0.0041	0.0017	0.0006	0.0003	0.0003	0.0026
	t-MM	0.0005 (8.20)	0.0005 (3.40)	0.0005 (1.20)	0.0004 (0.75)	0.0003 (1.00)	0.0003 (8.67)
	L-MM	0.0006 (6.83)	0.0007 (2.43)	0.0005 (1.20)	0.0004 (0.75)	0.0004 (0.75)	0.0005 (5.20)
$\hat{\beta}_{11}$	N-MM	0.3392	0.0893	0.0023	0.0009	0.0009	0.0071
	t-MM	0.0050 (67.84)	0.0012 (74.42)	0.0008 (2.88)	0.0002 (4.50)	0.0001 (9.00)	0.0010 (7.10)
	L-MM	0.0014 (242.29)	0.0014 (63.79)	0.0011 (2.09)	0.0001 (9.00)	0.0009 (1.00)	0.0006 (11.83)
$\hat{\beta}_{21}$	N-MM	0.4403	0.0815	0.0020	0.0015	0.0007	0.0066
	t-MM	0.0058 (75.91)	0.0005 (163.00)	0.0007 (2.86)	0.0002 (7.50)	0.0009 (0.78)	0.0014 (4.71)
	L-MM	0.0017 (259.00)	0.0012 (67.92)	0.0010 (2.00)	0.0001 (15.00)	0.0010 (0.70)	0.0011 (6.00)
$\hat{\beta}_{31}$	N-MM	0.2340	0.0265	0.0034	0.0013	0.0008	0.0082
	t-MM	0.0018 (130.00)	0.0011 (24.09)	0.0015 (2.27)	0.0002 (6.50)	0.0007 (1.14)	0.0009 (9.11)
	L-MM	0.0013 (180.00)	0.0012 (22.08)	0.0015 (2.27)	0.0001 (13.00)	0.0009 (0.89)	0.0012 (6.83)
$\hat{\beta}_{41}$	N-MM	0.2751	0.0266	0.0026	0.0012	0.0007	0.0047
	t-MM	0.0019 (144.79)	0.0008 (33.25)	0.0015 (1.73)	0.0001 (12.00)	0.0008 (0.88)	0.0009 (5.22)
	L-MM	0.0012 (229.25)	0.0007 (38.00)	0.0011 (2.36)	0.0002 (6.00)	0.0009 (0.78)	0.0013 (3.62)
$\hat{\beta}_{12}$	N-MM	0.1295	0.0052	0.0014	0.0003	0.0006	0.0042
	t-MM	0.0011 (117.72)	0.0029 (1.79)	0.0005 (2.80)	0.0002 (1.50)	0.0006 (1.00)	0.0005 (8.40)
	L-MM	0.0010 (129.50)	0.0004 (13.00)	0.0006 (2.33)	0.0001 (3.00)	0.0007 (0.86)	0.0008 (5.25)
$\hat{\beta}_{22}$	N-MM	0.1059	0.0059	0.0011	0.0007	0.0005	0.0030
	t-MM	0.0007 (151.29)	0.0005 (11.80)	0.0006 (1.83)	0.0001 (7.00)	0.0004 (1.25)	0.0006 (5.00)
	L-MM	0.0008 (132.38)	0.0005 (11.80)	0.0007 (1.57)	0.0001 (7.00)	0.0004 (1.25)	0.0008 (3.75)
$\hat{\beta}_{32}$	N-MM	0.1600	0.0453	0.0011	0.0005	0.0003	0.0037
	t-MM	0.0026 (61.53)	0.0005 (90.60)	0.0007 (1.57)	0.0002 (2.50)	0.0004 (0.75)	0.0009 (4.11)
	L-MM	0.0009 (177.78)	0.0008 (56.63)	0.0006 (1.83)	0.0001 (5.00)	0.0004 (0.75)	0.0008 (4.63)
$\hat{\beta}_{42}$	N-MM	0.1712	0.0456	0.0013	0.0007	0.0004	0.0030
	t-MM	0.0025 (68.48)	0.0008 (57.00)	0.0006 (2.17)	0.0002 (3.50)	0.0003 (1.33)	0.0006 (5.00)
	L-MM	0.0011 (155.64)	0.0007 (65.14)	0.0007 (1.86)	0.0001 (7.00)	0.0006 (0.67)	0.0004 (7.50)

Table 3.4: MedSE: $n_i = 6$, $m = 400$

Estimate	Method	t_1	t_3	t_5	Laplace	Normal	Contaminated
$\hat{\pi}_1$	N-MM	0.0061	0.0043	0.0030	0.0035	0.0020	0.0028
	t-MM	0.0076 (0.80)	0.0044 (0.98)	0.0033 (0.91)	0.0031 (1.13)	0.0020 (1.00)	0.0025 (1.12)
	L-MM	0.0085 (0.72)	0.0035 (1.23)	0.0024 (1.25)	0.0028 (1.25)	0.0020 (1.00)	0.0032 (0.88)
$\hat{\beta}_{11}$	N-MM	11.5754	0.1379	0.1321	0.0435	0.0037	0.0917
	t-MM	0.0218 (530.98)	0.0094 (14.67)	0.0113 (11.69)	0.0015 (33.46)	0.0040 0.93()	0.0179 (5.12)
	L-MM	0.0150 (771.69)	0.0063 (21.89)	0.0083 (15.92)	0.0011 (39.55)	0.0040 (0.93)	0.0071 (12.92)
$\hat{\beta}_{21}$	N-MM	24.7536	0.1485	0.1218	0.0450	0.0062	0.0945
	t-MM	0.0161 (1537.49)	0.0106 (14.01)	0.0064 (19.03)	0.0017 (26.47)	0.0070 (0.88)	0.0181 (5.22)
	L-MM	0.0135 (1833.60)	0.0081 (16.88)	0.0066 (18.45)	0.0015 (30.00)	0.0074 (0.84)	0.0064 (14.77)
$\hat{\beta}_{31}$	N-MM	24.1803	0.0856	0.0693	0.0256	0.0050	0.0566
	t-MM	0.0138 (1752.20)	0.0092 (9.30)	0.0066 (10.50)	0.0018 (14.22)	0.0051 (0.98)	0.0046 (12.30)
	L-MM	0.0112 (2158.96)	0.0079 (10.84)	0.0055 (12.60)	0.0015 (17.07)	0.0056 (0.89)	0.0068 (8.32)
$\hat{\beta}_{41}$	N-MM	29.3003	0.0857	0.0717	0.0204	0.0036	0.0541
	t-MM	0.0173 (1693.66)	0.0108 (7.94)	0.0052 (13.79)	0.0018 (11.33)	0.0038 (0.95)	0.0034 (15.91)
	L-MM	0.0144 (2034.74)	0.0070 (12.24)	0.0040 (17.93)	0.0015 (13.60)	0.0043 (0.84)	0.0052 (10.40)
$\hat{\beta}_{12}$	N-MM	11.3499	0.0299	0.0195	0.0082	0.0028	0.0150
	t-MM	0.0073 (1554.78)	0.0043 (6.95)	0.0033 (5.91)	0.0011 (7.45)	0.0029 (0.97)	0.0030 (5.00)
	L-MM	0.0068 (1669.10)	0.0036 (8.31)	0.0028 (6.96)	0.0010 (8.20)	0.0033 (0.85)	0.0040 (3.75)
$\hat{\beta}_{22}$	N-MM	22.1387	0.0295	0.0212	0.0077	0.0026	0.0162
	t-MM	0.0059 (3752.32)	0.0047 (6.28)	0.0033 (6.42)	0.0011 (7.00)	0.0027 (0.96)	0.0030 (5.40)
	L-MM	0.0055 (4025.22)	0.0047 (6.28)	0.0028 (7.57)	0.0008 (9.63)	0.0030 (0.87)	0.0040 (4.05)
$\hat{\beta}_{32}$	N-MM	21.0212	0.0760	0.0621	0.0229	0.0031	0.0468
	t-MM	0.0070 (3003.03)	0.0051 (14.90)	0.0041 (15.53)	0.0010 (22.90)	0.0035 (0.89)	0.0117 (4.00)
	L-MM	0.0063 (3336.70)	0.0040 (19.00)	0.0031 (20.03)	0.0007 (32.71)	0.0033 (0.94)	0.0040 (11.70)
$\hat{\beta}_{42}$	N-MM	21.6504	0.0752	0.0564	0.0237	0.0033	0.0471
	t-MM	0.0070 (3092.91)	0.0059 (12.75)	0.0031 (18.19)	0.0014 (16.93)	0.0031 (1.06)	0.0110 (4.28)
	L-MM	0.0063 (3436.57)	0.0045 (16.71)	0.0031 (18.19)	0.0011 (21.55)	0.0035 (0.94)	0.0030 (15.70)

Table 3.5: MSE: $n_i = 8$, $m = 100$

Estimate	Method	t_1	t_3	t_5	Laplace	Normal	Contaminated
$\hat{\pi}_1$	N-MM	0.0055	0.0033	0.0028	0.0015	0.0014	0.0180
	t-MM	0.0033 (1.67)	0.0019 (1.74)	0.0019 (1.47)	0.0012 (1.25)	0.0015 (0.93)	0.0015 (12.00)
	L-MM	0.0038 (1.45)	0.0020 (1.65)	0.0014 (2.00)	0.0012 (1.25)	0.0015 (0.93)	0.0019 (9.47)
$\hat{\beta}_{11}$	N-MM	17.0682	0.1508	0.1140	0.0461	0.0026	0.0372
	t-MM	0.0109 (1565.89)	0.0060 (25.13)	0.0045 (25.33)	0.0007 (65.86)	0.0058 (0.45)	0.0032 (11.63)
	L-MM	0.0071 (2403.97)	0.0028 (53.86)	0.0030 (38.00)	0.0004 (115.25)	0.0034 (0.76)	0.0034 (10.94)
$\hat{\beta}_{21}$	N-MM	18.0848	0.1406	0.1203	0.0487	0.0017	0.0728
	t-MM	0.0129 (1401.92)	0.0058 (24.24)	0.0044 (27.34)	0.0009 (54.11)	0.0042 (0.40)	0.0022 (33.09)
	L-MM	0.0078 (2318.56)	0.0039 (36.05)	0.0029 (41.48)	0.0005 (97.40)	0.0023 (0.74)	0.0027 (26.96)
$\hat{\beta}_{31}$	N-MM	76.7101	0.0813	0.0713	0.0227	0.0023	0.0626
	t-MM	0.0099 (7748.49)	0.0063 (12.90)	0.0046 (15.50)	0.0008 (28.38)	0.0025 (0.92)	0.0026 (24.08)
	L-MM	0.0067 (11449.27)	0.0034 (23.91)	0.0029 (24.59)	0.0005 (45.40)	0.0031 (0.74)	0.0030 (20.87)
$\hat{\beta}_{41}$	N-MM	11.7916	0.0815	0.0670	0.0228	0.0022	0.0278
	t-MM	0.0099 (1191.07)	0.0056 (14.55)	0.0047 (14.26)	0.0007 (32.57)	0.0023 (0.96)	0.0024 (11.58)
	L-MM	0.0062 (1901.87)	0.0039 (20.90)	0.0034 (19.71)	0.0006 (38.00)	0.0025 (0.88)	0.0028 (9.93)
$\hat{\beta}_{12}$	N-MM	15.0488	0.0237	0.0203	0.0065	0.0014	0.0333
	t-MM	0.0045 (3344.18)	0.0027 (8.78)	0.0029 (7.00)	0.0003 (21.67)	0.0015 (0.93)	0.0013 (25.62)
	L-MM	0.0040 (3762.20)	0.0021 (11.29)	0.0021 (9.67)	0.0002 (32.50)	0.0018 (0.78)	0.0015 (22.20)
$\hat{\beta}_{22}$	N-MM	14.8460	0.0266	0.0181	0.0060	0.0011	0.0230
	t-MM	0.0038 (3906.84)	0.0028 (9.50)	0.0021 (8.62)	0.0004 (15.00)	0.0012 (0.92)	0.0017 (13.53)
	L-MM	0.0038 (3906.84)	0.0020 (13.30)	0.0016 (11.31)	0.0003 (20.00)	0.0016 (0.69)	0.0019 (12.11)
$\hat{\beta}_{32}$	N-MM	67.5758	0.0703	0.0569	0.0231	0.0017	0.0204
	t-MM	0.0034 (19875.24)	0.0033 (21.30)	0.0023 (24.74)	0.0004 (57.75)	0.0042 (0.40)	0.0017 (12.00)
	L-MM	0.0031 (21798.65)	0.0020 (35.15)	0.0018 (31.61)	0.0002 (115.50)	0.0020 (0.85)	0.0020 (10.20)
$\hat{\beta}_{42}$	N-MM	8.8716	0.0682	0.0567	0.0241	0.0011	0.0149
	t-MM	0.0040 (2217.90)	0.0033 (20.67)	0.0026 (21.81)	0.0006 (40.17)	0.0036 (0.31)	0.0022 (6.77)
	L-MM	0.0033 (2688.36)	0.0022 (31.00)	0.0021 (27.00)	0.0004 (60.25)	0.0013 (0.85)	0.0025 (5.96)

Table 3.6: MSE: $n_i = 8$, $m = 200$

Estimate	Method	t_1	t_3	t_5	Laplace	Normal	Contaminated
$\hat{\pi}_1$	N-MM	0.0045	0.0027	0.0017	0.0017	0.0013	0.0409
	t-MM	0.0022 (2.05)	0.0020 (1.35)	0.0013 (1.31)	0.0014 (1.21)	0.0013 (1.00)	0.0018 (22.72)
	L-MM	0.0026 (1.73)	0.0028 (0.96)	0.0014 (1.21)	0.0013 (1.31)	0.0014 (0.93)	0.0026 (15.73)
$\hat{\beta}_{11}$	N-MM	10.9291	0.0901	0.0160	0.0032	0.0034	0.0605
	t-MM	0.0130 (840.70)	0.0149 (6.05)	0.0043 (3.72)	0.0009 (3.56)	0.0040 (0.85)	0.0038 (15.92)
	L-MM	0.0083 (1316.76)	0.0049 (18.39)	0.0043 (3.72)	0.0007 (4.57)	0.0045 (0.76)	0.0040 (15.12)
$\hat{\beta}_{21}$	N-MM	15.2047	0.0895	0.0128	0.0043	0.0039	0.1025
	t-MM	0.0122 (1246.99)	0.0136 (6.58)	0.0042 (3.05)	0.0010 (4.30)	0.0046 (0.85)	0.0043 (23.84)
	L-MM	0.0063 (2413.44)	0.0053 (16.89)	0.0047 (2.72)	0.0008 (5.38)	0.0055 (0.71)	0.0058 (17.67)
$\hat{\beta}_{31}$	N-MM	62.8445	0.0310	0.0104	0.0038	0.0036	0.1495
	t-MM	0.0058 (10835.26)	0.0032 (9.69)	0.0045 (2.31)	0.0013 (2.92)	0.0039 (0.92)	0.0043 (34.77)
	L-MM	0.0052 (12085.48)	0.0043 (7.21)	0.0054 (1.93)	0.0008 (4.75)	0.0049 (0.73)	0.0052 (28.75)
$\hat{\beta}_{41}$	N-MM	16.9928	0.0303	0.0082	0.0027	0.0031	0.1346
	t-MM	0.0064 (2655.12)	0.0047 (6.45)	0.0040 (2.05)	0.0012 (2.25)	0.0031 (1.00)	0.0047 (28.64)
	L-MM	0.0066 (2574.67)	0.0058 (5.22)	0.0049 (1.67)	0.0009 (3.00)	0.0036 (0.86)	0.0060 (22.43)
$\hat{\beta}_{12}$	N-MM	8.5129	0.0084	0.0043	0.0019	0.0020	0.0594
	t-MM	0.0029 (2935.48)	0.0028 (3.00)	0.0025 (1.72)	0.0006 (3.17)	0.0022 (0.91)	0.0029 (20.48)
	L-MM	0.0037 (2300.78)	0.0036 (2.33)	0.0030 (1.43)	0.0005 (3.80)	0.0027 (0.74)	0.0036 (16.50)
$\hat{\beta}_{22}$	N-MM	11.3715	0.0078	0.0035	0.0022	0.0027	0.0610
	t-MM	0.0032 (3553.59)	0.0030 (2.60)	0.0019 (1.84)	0.0008 (2.75)	0.0027 (1.00)	0.0030 (20.33)
	L-MM	0.0039 (2915.77)	0.0036 (2.17)	0.0023 (1.52)	0.0006 (3.67)	0.0032 (0.84)	0.0033 (18.48)
$\hat{\beta}_{32}$	N-MM	43.6165	0.0447	0.0083	0.0027	0.0022	0.0869
	t-MM	0.0076 (5739.01)	0.0086 (5.20)	0.0032 (2.59)	0.0009 (3.00)	0.0023 (0.96)	0.0024 (36.21)
	L-MM	0.0055 (7930.27)	0.0031 (14.42)	0.0038 (2.18)	0.0006 (4.50)	0.0026 (0.85)	0.0029 (29.97)
$\hat{\beta}_{42}$	N-MM	12.3743	0.0454	0.0077	0.0025	0.0021	0.0473
	t-MM	0.0057 (2170.93)	0.0095 (4.78)	0.0028 (2.75)	0.0009 (2.78)	0.0025 (0.84)	0.0024 (19.71)
	L-MM	0.0040 (3093.57)	0.0041 (11.07)	0.0028 (2.75)	0.0005 (5.00)	0.0028 (0.75)	0.0028 (16.89)

Table 3.7: MSE: $n_i = 6$, $m = 200$

Estimate	Method	t_1	t_3	t_5	Laplace	Normal	Contaminated
$\hat{\pi}_1$	N-MM	0.0045	0.0020	0.0009	0.0009	0.0009	0.0401
	t-MM	0.0015	0.0010	0.0008	0.0008	0.0009	0.0008
		(3.00)	(2.00)	(1.12)	(1.12)	(1.00)	(50.12)
	L-MM	0.0016	0.0012	0.0008	0.0008	0.0010	0.0012
		(2.81)	(1.67)	(1.12)	(1.12)	(0.90)	(33.42)
$\hat{\beta}_{11}$	N-MM	20.8041	0.0909	0.0058	0.0025	0.0016	0.1288
	t-MM	0.0080	0.0030	0.0018	0.0006	0.0017	0.0058
		(2600.51)	(30.30)	(3.22)	(4.17)	(0.94)	(22.21)
	L-MM	0.0033	0.0028	0.0022	0.0004	0.0019	0.0020
		(6304.27)	(32.46)	(2.64)	(6.25)	(0.84)	(64.40)
$\hat{\beta}_{21}$	N-MM	29.8189	0.0853	0.0056	0.0030	0.0015	0.0871
	t-MM	0.0087	0.0018	0.0017	0.0005	0.0019	0.0062
		(3427.46)	(47.39)	(3.29)	(6.00)	(0.79)	(14.05)
	L-MM	0.0036	0.0023	0.0020	0.0003	0.0021	0.0029
		(8283.03)	(37.09)	(2.80)	(10.00)	(0.71)	(30.03)
$\hat{\beta}_{31}$	N-MM	21.8550	0.0291	0.0057	0.0025	0.0018	0.0936
	t-MM	0.0035	0.0019	0.0025	0.0005	0.0019	0.0020
		(6244.29)	(15.32)	(2.28)	(5.00)	(0.95)	(46.80)
	L-MM	0.0030	0.0027	0.0028	0.0004	0.0023	0.0030
		(7285.00)	(10.78)	(2.04)	(6.25)	(0.78)	(31.20)
$\hat{\beta}_{41}$	N-MM	23.7552	0.0291	0.0054	0.0023	0.0018	0.1139
	t-MM	0.0038	0.0023	0.0023	0.0005	0.0020	0.0020
		(6251.37)	(12.65)	(2.35)	(4.60)	(0.90)	(56.95)
	L-MM	0.0029	0.0023	0.0024	0.0004	0.0022	0.0025
		(8191.45)	(12.65)	(2.25)	(5.75)	(0.82)	(45.56)
$\hat{\beta}_{12}$	N-MM	18.5801	0.0071	0.0020	0.0011	0.0011	0.0383
	t-MM	0.0019	0.0012	0.0012	0.0003	0.0012	0.0012
		(9779.00)	(5.92)	(1.67)	(3.67)	(0.92)	(31.92)
	L-MM	0.0021	0.0012	0.0013	0.0002	0.0015	0.0015
		(8847.67)	(5.92)	(1.54)	(5.50)	(0.73)	(25.53)
$\hat{\beta}_{22}$	N-MM	27.4140	0.0070	0.0020	0.0016	0.0012	0.0405
	t-MM	0.0017	0.0013	0.0015	0.0004	0.0013	0.0014
		(16125.88)	(5.38)	(1.33)	(4.00)	(0.92)	(28.93)
	L-MM	0.0019	0.0016	0.0016	0.0003	0.0015	0.0015
		(14428.42)	(4.38)	(1.25)	(5.33)	(0.80)	(27.00)
$\hat{\beta}_{32}$	N-MM	20.2915	0.0458	0.0025	0.0015	0.0008	0.0404
	t-MM	0.0039	0.0013	0.0016	0.0004	0.0010	0.0036
		(5202.95)	(35.23)	(1.56)	(3.75)	(0.80)	(11.22)
	L-MM	0.0020	0.0016	0.0016	0.0003	0.0010	0.0015
		(10145.75)	(28.62)	(1.56)	(5.00)	(0.80)	(26.93)
$\hat{\beta}_{42}$	N-MM	19.1017	0.0460	0.0032	0.0016	0.0009	0.0407
	t-MM	0.0040	0.0015	0.0015	0.0003	0.0009	0.0033
		(4775.43)	(30.67)	(2.13)	(5.33)	(1.00)	(12.33)
	L-MM	0.0021	0.0015	0.0016	0.0003	0.0012	0.0011
		(9096.05)	(30.67)	(2.00)	(5.33)	(0.75)	(37.00)

Table 3.8: MSE: $n_i = 6$, $m = 400$

	Original Data		
	Laplace	t_{28}	Normal
$\hat{\pi}_1$	0.279(0.120)	0.248(0.081)	0.235(0.045)
$\hat{\pi}_2$	0.707(0.120)	0.688(0.077)	0.704(0.040)
$\hat{\beta}_{01}$	-0.038(0.081)	-0.010(0.328)	-0.010(0.129)
$\hat{\beta}_{11}$	0.075(0.006)	0.074(0.028)	0.074(0.010)
$\hat{\beta}_{02}$	-0.362(0.043)	-0.350(0.090)	-0.341(0.080)
$\hat{\beta}_{12}$	0.091(0.003)	0.092(0.010)	0.091(0.008)
$\hat{\beta}_{03}$	-0.624(0.137)	-0.307(0.093)	-0.296(0.074)
$\hat{\beta}_{13}$	0.085(0.011)	0.074(0.010)	0.073(0.007)
	One outlier		
	Laplace	t_9	Normal
$\hat{\pi}_1$	0.338(0.166)	0.281(0.015)	0.569(0.178)
$\hat{\pi}_2$	0.649(0.163)	0.652(0.195)	0.402(0.136)
$\hat{\beta}_{01}$	-0.054(0.081)	-0.041(0.673)	-0.126(0.566)
$\hat{\beta}_{11}$	0.074(0.005)	0.074(0.062)	0.083(0.030)
$\hat{\beta}_{02}$	-0.376(0.051)	-0.361(0.369)	-0.336(0.228)
$\hat{\beta}_{12}$	0.093(0.004)	0.093(0.042)	0.090(0.028)
$\hat{\beta}_{03}$	-0.537(0.151)	-0.293(0.477)	-0.418(0.229)
$\hat{\beta}_{13}$	0.076(0.012)	0.074(0.050)	0.090(0.025)
	Two outliers		
	Laplace	t_6	Normal
$\hat{\pi}_1$	0.311(0.182)	0.297(0.178)	0.196(0.148)
$\hat{\pi}_2$	0.684(0.177)	0.630(0.249)	0.765(0.194)
$\hat{\beta}_{01}$	-0.059(0.109)	-0.056(0.553)	-0.175(0.305)
$\hat{\beta}_{11}$	0.073(0.008)	0.075(0.048)	0.083(0.011)
$\hat{\beta}_{02}$	-0.368(0.057)	-0.368(0.389)	-0.274(0.288)
$\hat{\beta}_{12}$	0.093(0.005)	0.093(0.044)	0.087(0.032)
$\hat{\beta}_{03}$	-0.095(0.136)	-0.279(0.540)	-0.335(0.362)
$\hat{\beta}_{13}$	0.066(0.011)	0.075(0.055)	0.088(0.045)

Table 3.9: Lung growth data analysis

Chapter 4

Application: Functional Mapping of Dynamic Traits with Multivariate Laplace Distribution

In previous chapters, we have shown that multivariate Laplace distribution can achieve certain robustness in multivariate mixture regression models and mixture of linear mixed models . In this chapter, we apply the proposed methodology to the functional mapping for detecting quantitative trait loci (QTL).

4.1 Introduction

Traits are complex in biological, biomedical and agricultural studies. Identifying quantitative traits has been one of the most important topics in the biology history. There are several quaitntatitative genetic models serving as instruments for predicting developmental events, see **?**. With genetic markers information and statistics, **?** proposed a quantitative trait loci (QTL) to analyze a various quantitative traits of interest. Since then, there are numbers of literatures discussing on the development of statical methods for mapping complex traits. For example, **?**, **?**, **?**, **?** and **?**. However, traditional statistical methods of QTL mapping neglect the developmental characteristic of trait. For example, body height and weight change with time. Genetic control should be represented as a function of time. **?** proposed a functional mapping approach to model developmental mean function of a dynamic trait. Identification of QTL can be resolved by testing mean differences of different QTL genotype categories.

For conventional functional mapping of development traits, normality assumption is assumed for the random error distribution. However, in many applied researches such as agricultural studies, the data distribution may have longer tails than normal distribution. The presence of extreme observations can make a huge impact on statistical inferences. A robust functional mapping becomes aspiration for identifying QTLs underlying dynamics traits.

In this chapter, we extend the proposed robust method by assuming Laplace distribution for random errors. Motivated by **?**, the proposed robust method should obtain certain robustness as t-distribution does to the QTL setup, but computationally more efficient than the t-procedure.

Since the purpose of this study is to compare the robustness of estimation procedure based on normal, Laplace, and t-distributions, we simply adopt the same mean and covariance functions as in **?**. The real data set used in **?** is reanalysed for identifying genes underlying the variation of rice tiller numbers.

In section 4.2, we shall discuss how to apply the robust procedure to a two component mixture model. An efficient EM algorithm is developed with non-parametric B-spline mean function and first order structured antedependence (SAD(1)) covariance function. The performance of the proposed method is evaluated through some simulation studies in section 4.3. A real data analysis is presented in section 4.4.

4.2 Statistical Methods

4.2.1 The Mixture Model with Multivariate Laplace Distribution

For a backcross population with n observations, each one is measured over n time points. The phenotypic vector $\mathbf{y} = [y(t_1), ..., y(t_n)]^T$ follows a multivariate distribution with a location-scale density function $f(\mathbf{y}; \alpha, \beta)$, where α and β denote location and scale parameters, respectively.

Similar to **?** and **?** functional mapping framework, we apply a mixture multivariate Laplace model to estimate unknown parameters of interest. The missing data problem can be overcome by modeling the observed phenotypic data with a two-component mixture model

$$\mathbf{y}_i \sim p(\mathbf{y}_i; \alpha, \beta) = \sum_{j=0}^{1} \pi_{i|j} f_j(\mathbf{y}_i; \alpha_j, \beta)$$

where $f_j(\mathbf{y}_i; \alpha_0, \beta)$ is the probability density function with the location parameters α_j corresponding to QTL genotype j (=1 for QQ and =0 for Qq); β contains the scale parameters common to all components; and $\pi_{i|j}$ is the mixture proportion of individual i given the QTL genotype j.

Now we assume that $\mathbf{y}$ follows a multivariate Laplace distribution. To be specific, the multivariate Laplace density function for individual i given genotype j is given by

$$f_j(\mathbf{y}_i; \Omega_j) = \frac{2}{(2\pi)^{n/2}|\Sigma_j|^{1/2}} \left[\frac{(\mathbf{y}_i-\mu_j)'\Sigma_j^{-1}(\mathbf{y}_i-\mu_j)}{2}\right]^{\frac{1}{2}(1-\frac{n}{2})} K_{n/2-1}\left(\sqrt{2(\mathbf{y}_i-\mu_j)'\Sigma_j^{-1}(\mathbf{y}_i-\mu_j)}\right)$$

where for genotype j (=0, 1), $\mu_j = [\mu_j(t_1), ..., \mu_j(t_n)]$ denotes the mean vector, Σ_j is a positive definite covariance matrix, $K_m(x)$ is the modified Bessel Function of the second kind with order m, and $\Omega_j = (\mu_j, \Sigma_j)$ contains all the parameter of interest from the genotype j.

At a specific time point t, the relationship between the observation and the mean functional mapping can be expressed by a linear model

$$y_i(t) = z_i\mu_1(t) + (1 - z_i)\mu_0(t) + e_i(t)$$

where z_i =0 or 1 if the QTL genotype is Qq or QQ; and $e_i(t)$ is the error term following a Laplace distribution with mean zero and variance $\sigma^2(t)$. The errors at two different time points t_{n_1} and t_{n_2} are correlated with correlation coefficient $\rho(t_{n_1}, t_{n_2})$. Similar to **?**, we apply a more flexible developmental mean function by a non-parametric B-spline technique. An antedependence covariance model is adopted for the non-stationary covariance structure.

Assuming independence among individuals, the joint likelihood function can be expressed as

$$L(\Omega) = \prod_{i=1}^{N} [\pi_{i|0} f_0(\mathbf{y}_i|\Omega_0) + \pi_{i|1} f_1(\mathbf{y}_i|\Omega_1)]$$

where $\pi_{i|j} = P(z_{i|j} = 1)$, and $\pi_{i|0} + \pi_{i|1} = 1$. According to **?**, the mixture proportions can be solved by the conditional probabilities of QTL genotypes given the marker information in a backcross design. We define Ω as the collection of all unknown parameters. In the meanwhile, we use Ω_l to denote the locations of the QTL with respect to markers, and denote $\Omega_g = (\Omega_m, \Omega_c)$ the multivariate Laplace distribution mean vectors and covariance matrices. Since the complexity of the log-likelihood function makes it hard to maximize directly, we develop an efficient EM algorithm in the next section.

4.2.2 EM Algorithm

Define $z_i = 0$ or 1 if the QTL genotype is Qq or QQ, respectively. The n-dimensional random observations $\mathbf{y}_i, (i = 1, ..., N)$ are generated independently from a two-component mixture of multivariate Laplace distribution with proportions $\pi_{i|0}$ and $\pi_{i|1}$

$$f(\mathbf{y}_i; \Omega) = \pi_{i|0} f_0(\mathbf{y}_i; \Omega_0) + \pi_{i|1} f_1(\mathbf{y}_i; \Omega_1)$$

where $\pi_{i|j} = P(z_{i|j} = 1)$, $\Omega = (\Omega_0, \Omega_1)$ and $\Omega_j = (\mu_j, \Sigma_j)$. By the property of the multivariate Laplace distribution, we have

$$\mathbf{y}_i|v_i, z_{i|j} = 1 \sim N_n(\mu_j, v_i\Sigma_j), \text{ for } i = 1, ..., N, \ j = 0, 1$$
$$v_i|z_{i|j} = 1 \sim f_{v_i}(v) = e^{-v}\mathcal{I}(v \geqslant 0)$$

The complete data log-likelihood function can be expressed as $\ell(\Omega) = \ell_0(\pi) + \ell_1(\mu, \Sigma|y, v)$, where

$$\ell_0(\pi) = \sum_{i=1}^{N}\sum_{j=0}^{1} z_{i|j} log(\pi_{i|j})$$
$$\ell_1(\mu, \Sigma|y, v) = \sum_{i=1}^{N}\sum_{j=0}^{1} z_{i|j}\left\{-\frac{n}{2}log(2\pi) - \frac{1}{2}log|\Sigma_j| - \frac{1}{2v_i}(y_i - \mu_j)'\Sigma_j^{-1}(y_i - \mu_j) - v_i\right\}$$

Assume the two multivariate Laplace components to have the same covariance structure, i.e., $\Sigma_1 = \Sigma_2 = \Sigma$. Since we adopt same frame modeling framework as **?**, a uniform quadratic B-spline with degree 5 is adopted to model the time-dependent mean function. Additionally, the normalized basis matrix B is following

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.3906 & 0.5391 & 0.0703 & 0 & 0 \\ 0.0625 & 0.6563 & 0.2813 & 0 & 0 \\ 0 & 0.3828 & 0.6094 & 0.0078 & 0 \\ 0 & 0.1250 & 0.7500 & 0.1250 & 0 \\ 0 & 0.0078 & 0.6094 & 0.3828 & 0 \\ 0 & 0 & 0.2813 & 0.6563 & 0.0625 \\ 0 & 0 & 0.0703 & 0.5391 & 0.3906 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and η_j as the base genotypic vector for genotype j

$$\eta_j = [\eta_{1j}\ \eta_{2j}\ \eta_{3j}\ \eta_{4j}\ \eta_{5j}]'$$

whose entries are the mean parameters to be estimated with the mean vector given by $\mu_j = \mathbf{B}\eta_j$. For the SAD(1) covariance structure with constant innovation variance $\sigma_t^2 = \sigma^2$, one can easily show that $\Sigma^{-1} = \frac{1}{\sigma^2} L^T L$, $|\Sigma| = (\sigma^2)^n$, and

$$\begin{aligned}(y_i-\mu_j)^T\Sigma^{-1}(y_i-\mu_j) &= \frac{1}{\sigma^2}(y_i-\mu_j)^T\Gamma(\psi)(y_i-\mu_j)\\ &= \frac{1}{\sigma^2}\Bigg\{-2\psi\sum_{m=2}^{n-1}\left[y_i(t_m)-\mu_j(t_m)\right]\left[y_i(t_{m+1})-\mu_j(t_{m+1})\right]\\ &\qquad + (\psi^2+1)\sum_{m=1}^{n-1}\left[y_i(t_m)-\mu_j(t_m)\right]^2 - \left[y_i(t_n)-\mu_j(t_n)\right]^2\Bigg\}\end{aligned}$$

where L is a lower triangular matrix with 1's on the diagonal and with the negative of the antedependence coefficient ψ as the off-diagonal entries,

$$L = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ -\psi & 1 & 0 & 0 & \cdots & 0 \\ 0 & -\psi & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & -\psi & 1 & 0 \\ 0 & \cdots & \cdots & 0 & -\psi & 1 \end{bmatrix}, \Gamma(\psi) = \begin{bmatrix} 1+\psi^2 & -\psi & 0 & \cdots & \cdots & 0 \\ -\psi & 1+\psi^2 & -\psi & 0 & \cdots & 0 \\ 0 & -\psi & 1+\psi^2 & -\psi & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & -\psi & 1+\psi^2 & -\psi \\ 0 & \cdots & \cdots & 0 & -\psi & 1 \end{bmatrix}$$

The maximum likelihood estimator of the unknown parameters can be obtained by using the following EM algorithm. At k^{th} iteration in the **E**-step, the posterior probability of the observed trait vector $\mathbf{y}_i$ belonging to the genotype j can be expressed as

$$\hat{z}_{i|j}^{(k)} = E(z_{i|j} = 1|\mathbf{y}_i, \hat{\Omega}^{(k)}) = \frac{\pi_{i|j} f_j(\mathbf{y}_i; \hat{\Omega}_j^{(k)})}{\sum_{j=0}^{1} \pi_{i|j} f_j(\mathbf{y}_i; \hat{\Omega}_j^{(k)})}$$

where $\hat{\Omega}_j = (\hat{\eta}_j, \hat{\sigma}^2, \hat{\psi})$, , and $j = 0, 1$. The conditional expectation of $1/v_i$ given $z_{i|j} = 1$ is calculated by lemma 1.3.1 as follows,

$$\hat{v}_{ij}^{(k)} = E\left(\frac{1}{v_i}\middle|\mathbf{y}_i, z_{i|j} = 1, \hat{\Omega}^{(k)}\right) = \frac{K_{-n/2}\left(\sqrt{2\hat{Q}_{ij}(\mathbf{y}_i; \hat{\Omega}_j^{(k)})}\right)}{K_{1-n/2}\left(\sqrt{2\hat{Q}_{ij}(\mathbf{y}_i; \hat{\Omega}_j^{(k)})}\right)}\sqrt{\frac{2}{\hat{Q}_{ij}(\mathbf{y}_i; \hat{\Omega}_j^{(k)})}}$$

where $\hat{Q}_{ij}(\mathbf{y}_i; \hat{\Omega}_j^{(k)}) = \frac{1}{\hat{\sigma}^2}(y_i - \mathbf{B}\hat{\eta}_j)^T L^T L(y_i - \mathbf{B}\hat{\eta}_j)$.

In the **M**-step, the updates for η_j, σ^2 and ψ are obtained as:

$$\hat{\eta}_j^{(k+1)} = \left\{\sum_{i=1}^{N}\sum_{j=0}^{1} \hat{z}_{i|j}^{(k)} \hat{v}_{ij}^{(k)} B^T L^T L B\right\}^{-1} \left\{\sum_{i=1}^{N}\sum_{j=0}^{1} \hat{z}_{i|j}^{(k)} \hat{v}_{ij}^{(k)} B^T L^T L \mathbf{y}_i\right\}$$

$$\hat{\sigma}^{2(k+1)} = \frac{\sum_{i=1}^{N}\sum_{j=0}^{1} \hat{z}_{i|j}^{(k)} \hat{v}_{ij}^{(k)} (\mathbf{y}_i - B\hat{\eta}_j^{(k+1)})^T L^T L(\mathbf{y}_i - B\hat{\eta}_j^{(k+1)})}{Nn}$$

$$\hat{\psi}^{(k+1)} = \frac{\sum_{i=1}^{N}\sum_{j=0}^{1} \hat{z}_{i|j}^{(k)} \hat{v}_{ij}^{(k)} \sum_{m=2}^{n-1}\left[y_i(t_m) - B_m^T\hat{\eta}_j^{(k+1)}\right]\left[y_i(t_{m+1}) - B_{m+1}^T\hat{\eta}_j^{(k+1)}\right]}{\sum_{i=1}^{N}\sum_{j=0}^{1} \hat{z}_{i|j}^{(k)} \hat{v}_{ij}^{(k)} \sum_{m=1}^{n-1}\left[y_i(t_m) - B_m^T\hat{\eta}_j^{(k+1)}\right]^2}$$

where B_m and B_{m+1} are the m^{th} and $(m+1)^{\text{th}}$ row of the design matrix B. The above procedures are iterated until certain convergence criterion is achieved. The converged values are the MLEs of the parameters.

In the QTL mapping studies, it's complicated and inefficient to estimate location parameter directly. Instead, people commonly use a grid search method to estimate the QTL location by searching for a putative QTL at every $2cM$ on an interval by two flanking markers, see **?**. Then a log-likelihood ratio test is performed to test QTL location for each marker interval. The hypothesis test can be designed in the following way

$$\begin{cases} H_0 : \Omega_{m0} = \Omega_{m1} \\ H_1 : \Omega_{m0} \neq \Omega_{m1} \end{cases} \tag{4.2.1}$$

where H_0 is the reduced model, which can be fit by a single model, and H_1 is the full model, which there exists different models to fit the data. The test statistics can be calculated as log-likelihood ratio (LR) of the reduced to the full model. The critical threshold is based on permutation tests proposed by **?**. We summarize the log-likelihood ratio test statistics for each testing position of QTL in a graph, which is called LR profile plot. The location of QTL can be determined by the peak of the profile plot which is the MLE.

4.3 Simulation Studies

In this section, we conduct a preliminary simulation study to evaluate the performance of the proposed robust estimation procedure. The design of simulation studies is similar to **?**. Data is generated from a backcross population with $100cM$ long linkage group, composed of 6 equidistant markers, under the assumption that QTL governs the whole developmental process. We assume that a putative QTL which affects a development process to be located at $48cM$ away from the first marker on the linkage group, where it's between the $3rd$ and $4th$ markers.

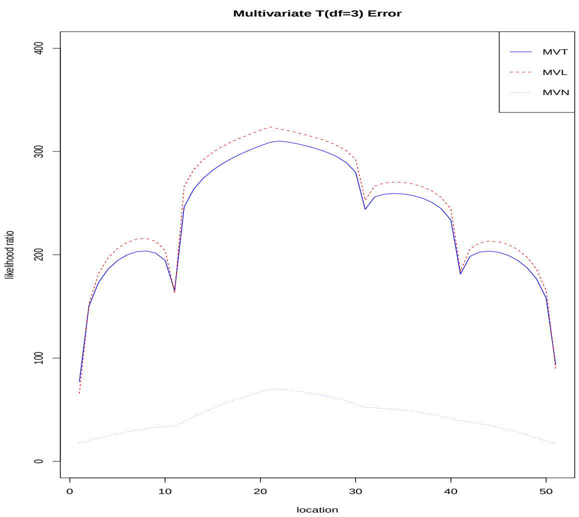
The recombination fraction is calculated based on the map distance that is converted by the Haldane map function. We simulated our data by a developmental trait with nine equally spaced time points under a 0.4 heritability level ($H^2 = 0.4$). The sample size of $n = 100$ is used in the simulation study. For each case, the simulation is repeated 100 times. Then we consider four different error distribution cases: (1) $\varepsilon \sim t_3$, t distribution with 3 degrees of freedom; (2) $\varepsilon \sim t_5$, t distribution with 5 degrees of freedom; (3) $\varepsilon \sim$ Laplace distribution with mean 0 and covariance I; (4) $\varepsilon \sim N(0, I)$.

Both case (1) and (2) are heavy tailed distributions and are often used in the literatures to mimic the outlier situations. Under case (3), the procedure will provide the MLE of unknown parameters, which would serve as a baseline to evaluate the performance of other estimation procedures. Case (4) is often used to evaluate the efficiency of different estimation methods compared to the traditional MLE when error is exactly multivariate normally distributed and there are no outliers.

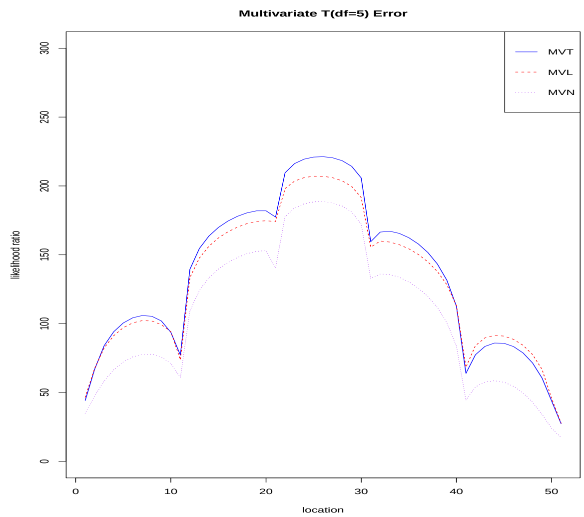
The performance of using multivariate Laplace distribution (MVL), multivariate t distribution (MVT) and multivariate normal distribution (MVN), is compared via the likelihood ratio (LR) statistics across the simulated genetic linkage group in Figure 4.1 and the estimates of model parameters are reported in Table 4.1, 4.2, 4.3 and 4.4.

In real application, the LR is used as the indicator of a QTL signal. The larger the LR value at a genomic position, the stronger the evidence of a QTL at that position. Figure 4.1 explicitly displays the difference in LR values by applying different distribution methods. Clearly, one can see the proposed method performs better than the others when the error distribution has heavy tail. The normal distribution method cannot indicate QTL position easily when there're many outliers existing. As the data gets close to normal data, the proposed method performs closely well as the multivariate t distribution method. If the error distribution is multivariate normal, all three methods can detect the QTL position, which indicates the proposed method and the multivariate t distribution are as efficient as the traditional normal distribution method.

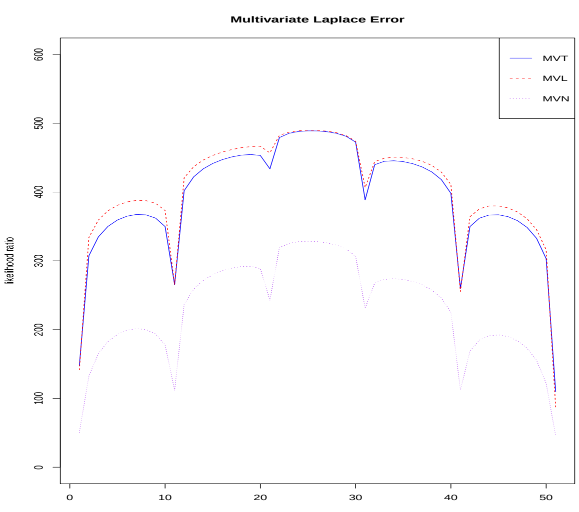
On the other hand, we also report the MLEs and standard errors of the model parameters and the QTL position by applying different distribution methods under four different error distribution cases. The results are similar to LR plots. The proposed method outperforms or is at least comparable to any other methods except for multivariate normal distribution error, which the proposed method's estimates are slightly larger than the true values. However, when the error distribution has heavier tail, the proposed method are comparable to multivariate t distribution method.



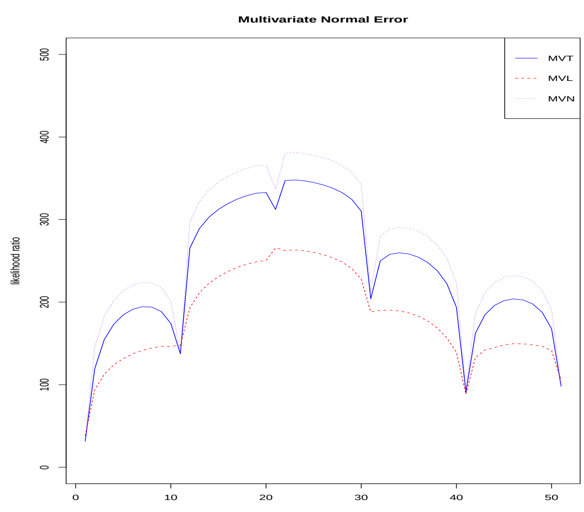
Error Distribution: (1)



Error Distribution: (2)



Error Distribution: (3)



Error Distribution: (4)

Figure 4.1: The LR profile plots

Error distribution is multivariate T distribution with degrees of freedom 3.

		MVL method	MVT method	MVN method
QTL position	l	47.91(3.03)	48.32(3.55)	48.24(4.17)
Qq parameters	$\hat{\eta}_{10}$	1.15(0.03)	1.16(0.07)	1.19(0.11)
	$\hat{\eta}_{20}$	6.80(0.17)	6.91(0.14)	7.28(0.19)
	$\hat{\eta}_{30}$	12.59(0.23)	12.61(0.25)	12.77(0.26)
	$\hat{\eta}_{40}$	7.03(0.24)	7.05(0.20)	7.20(0.38)
	$\hat{\eta}_{50}$	7.02(0.13)	7.05(0.18)	7.15(0.38)
QQ parameters	$\hat{\eta}_{11}$	1.18(0.03)	1.21(0.06)	1.24(0.10)
	$\hat{\eta}_{21}$	7.20(0.14)	7.46(0.14)	7.69(0.17)
	$\hat{\eta}_{31}$	11.61(0.21)	11.25(0.25)	11.81(0.24)
	$\hat{\eta}_{41}$	6.57(0.16)	6.41(0.18)	6.73(0.23)
	$\hat{\eta}_{51}$	6.66(0.17)	6.55(0.17)	6.79(0.24)
Covariance parameters	$\hat{\psi}$	0.94(0.01)	0.95(0.02)	0.94(0.04)
	$\hat{\sigma}^2$	0.36(0.01)	0.21(0.02)	0.56(0.67)

Table 4.1: The MLEs and standard errors (in the parenthesis) of the model parameters and the QTL position.

Error distribution is multivariate T distribution with degrees of freedom 5.

		MVL method	MVT method	MVN method
QTL position	l	46.66(2.66)	47.68(2.72)	47.82(2.43)
Qq parameters	$\hat{\eta}_{10}$	1.15(0.07)	1.17(0.06)	1.20(0.07)
	$\hat{\eta}_{20}$	6.82(0.11)	6.94(0.15)	7.26(0.16)
	$\hat{\eta}_{30}$	12.53(0.26)	12.54(0.24)	12.75(0.22)
	$\hat{\eta}_{40}$	7.00(0.23)	7.01(0.20)	7.16(0.21)
	$\hat{\eta}_{50}$	7.03(0.13)	7.05(0.17)	7.14(0.20)
QQ parameters	$\hat{\eta}_{11}$	1.18(0.03)	1.20(0.06)	1.22(0.07)
	$\hat{\eta}_{21}$	7.20(0.14)	7.42(0.14)	7.66(0.14)
	$\hat{\eta}_{31}$	11.58(1.20)	11.32(0.26)	11.81(0.25)
	$\hat{\eta}_{41}$	6.55(0.20)	6.43(0.20)	6.72(0.20)
	$\hat{\eta}_{51}$	6.65(0.16)	6.57(0.16)	6.79(0.18)
Covariance parameters	$\hat{\psi}$	0.93(0.01)	0.94(0.02)	0.94(0.02)
	$\hat{\sigma}^2$	0.28(0.03)	0.21(0.02)	0.32(0.04)

Table 4.2: The MLEs and standard errors (in the parenthesis) of the model parameters and the QTL position.

Error distribution is multivariate Laplace distribution.

		MVL method	MVT method	MVN method
QTL position	l	47.24(2.32)	47.96(2.62)	48.18(2.43)
	$\hat{\eta}_{10}$	1.14(0.02)	1.16(0.04)	1.19(0.06)
	$\hat{\eta}_{20}$	6.71(0.10)	6.83(0.12)	7.27(0.13)
Qq parameters	$\hat{\eta}_{30}$	12.82(0.27)	12.81(0.31)	12.75(0.24)
	$\hat{\eta}_{40}$	7.13(0.16)	7.14(0.17)	7.16(0.16)
	$\hat{\eta}_{50}$	7.10(0.14)	7.12(0.15)	7.12(0.15)
	$\hat{\eta}_{11}$	1.17(0.02)	1.22(0.04)	1.24(0.05)
	$\hat{\eta}_{21}$	7.05(0.04)	7.57(0.10)	7.71(0.12)
QQ parameters	$\hat{\eta}_{31}$	11.97(0.23)	10.98(0.24)	11.77(0.20)
	$\hat{\eta}_{41}$	6.74(0.12)	6.28(0.15)	6.73(016)
	$\hat{\eta}_{51}$	6.80(0.11)	6.45(0.13)	6.80(0.14)
Covariance	$\hat{\psi}$	0.93(0.01)	0.94(0.02)	0.93(0.02)
parameters	$\hat{\sigma}^2$	0.21(0.03)	0.10(0.03)	0.22(0.02)

Table 4.3: The MLEs and standard errors (in the parenthesis) of the model parameters and the QTL position.

Error distribution is multivariate normal distribution.

		MVL method	MVT method	MVN method
QTL position	l	46.76(7.64)	48.11(2.83)	48.06(2.78)
	$\hat{\eta}_{10}$	1.16(0.13)	1.18(0.05)	1.21(0.05)
	$\hat{\eta}_{20}$	6.85(0.70)	6.99(0.11)	7.27(0.11)
Qq parameters	$\hat{\eta}_{30}$	12.53(1.29)	12.47(0.23)	12.79(0.22)
	$\hat{\eta}_{40}$	7.01(0.73)	6.99(0.17)	7.19(0.16)
	$\hat{\eta}_{50}$	7.01(0.73)	7.01(0.17)	7.15(0.16)
	$\hat{\eta}_{11}$	1.18(0.13)	1.20(0.06)	1.23(0.06)
	$\hat{\eta}_{21}$	7.20(0.74)	7.38(0.12)	7.67(0.12)
QQ parameters	$\hat{\eta}_{31}$	11.56(1.19)	11.41(0.23)	11.80(0.21)
	$\hat{\eta}_{41}$	6.54(0.69)	6.47(0.19)	6.71(0.18)
	$\hat{\eta}_{51}$	6.64(0.69)	6.59(0.15)	6.78(0.15)
Covariance	$\hat{\psi}$	0.93(0.10)	0.94(0.02)	0.94(0.02)
parameters	$\hat{\sigma}^2$	0.20(0.03)	0.20(0.01)	0.22(0.01)

Table 4.4: The MLEs and standard errors (in the parenthesis) of the model parameters and the QTL position.

4.4 Real Data Analysis

In this section, we applied the proposed method to a real data to identify QTLs position of rice tiller number development over 10 days. The detailed description of the study and data set can be found in **?** and **?**. There are two inbred lines, semidwarf IR64 and tall Azucena, crossing to generate an F_1 progeny population. In summary, there are 40 isozyme and RAPD markers, and 135 RFLP markers constructing a genetic linkage map of total length $2005cM$ across 12 rice chromosomes. The observations of tiller numbers were measured every 10 days and there are nine developmental measurements were recorded for each rice.

We used the LR as the indicator of a QTL signal and performed a genome-wide linkage scan at every $2cM$ across 12 rice chromosomes. Figure 4.2 shows the LR profile plots, where the results obtained with the multivariate Laplace distribution method and the multivariate t distribution method. A 5% genome-wide permutation threshold indicated by the horizontal dashed lines based on 1,000 permutations. The results show that those two distribution methods are comparable, but multivariate Laplace distribution can generate higher LR values in many positions. Both methods indicate one QTL location in chromosome 3 between marker RZ519 and Pgi-1, which is identical to the results reported in existing studies, **?**, **?**, and **?**.

However, the estimate of degrees of freedom with multivariate t distribution is around 10. That indicates the original data distribution is close to normal. For comparison purpose, we did a sensitivity study to check the robustness of using multivariate Laplace distribution, multivariate t distribution, and multivariate normal distribution by adding two same artificial extreme values, $(4, 14, 27, 33, 30, 28, 27, 20, 19)'$, into the first two observations. We also provide a boxplot of original data and artificial extreme values, where those values are five standard deviation away from the mean. From Figure 4.3 and Figure 4.4 (Pooled Chromosomes Graphs), we can see that the LR profile plots do not vary too much from both t and Laplace distribution methods, indicating that those two methods possess certain robustness. On the other hand, the normal distribution method provides different LR profile plot when outliers present. We can observe the same conclusion from the parameters estimates in table 4.5 and table 4.6. It's also worth mentioning that the multivariate Laplace distribution method is comparable to multivariate t distribution method in most of cases, but it takes more time to estimate the degrees of freedom for t distribution to fit the data and brings more computations during the process. Thus, multivariate Laplace distribution can be more applicable as a new robust method in practical problems.

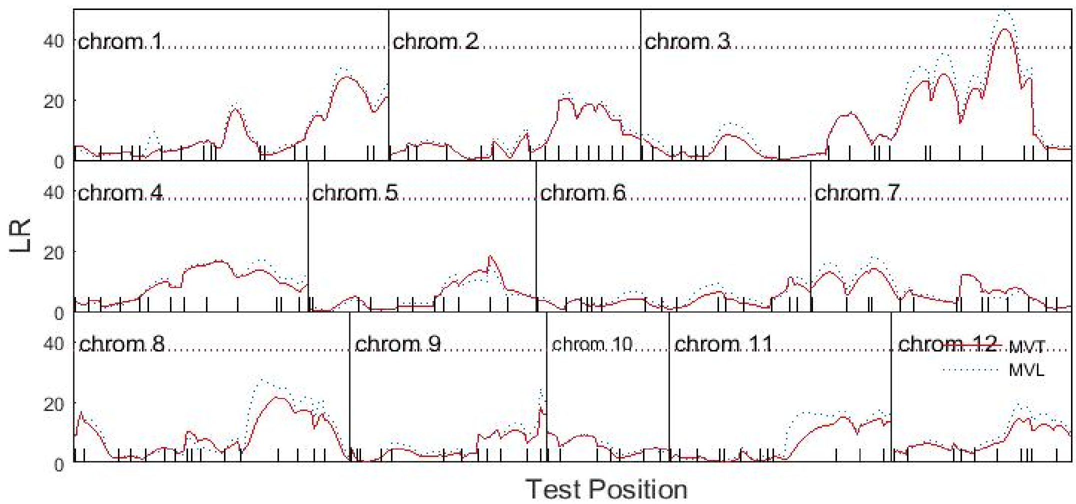


Figure 4.2: The LR profile plot across 12 chromosomes. The proposed multivariate Laplace distribution method is solid curve and the multivariate t distribution method is dash-dotted curve. The 5% genome-wide threshold value for claiming the existence of a QTL is given as the horizontal dotted line.

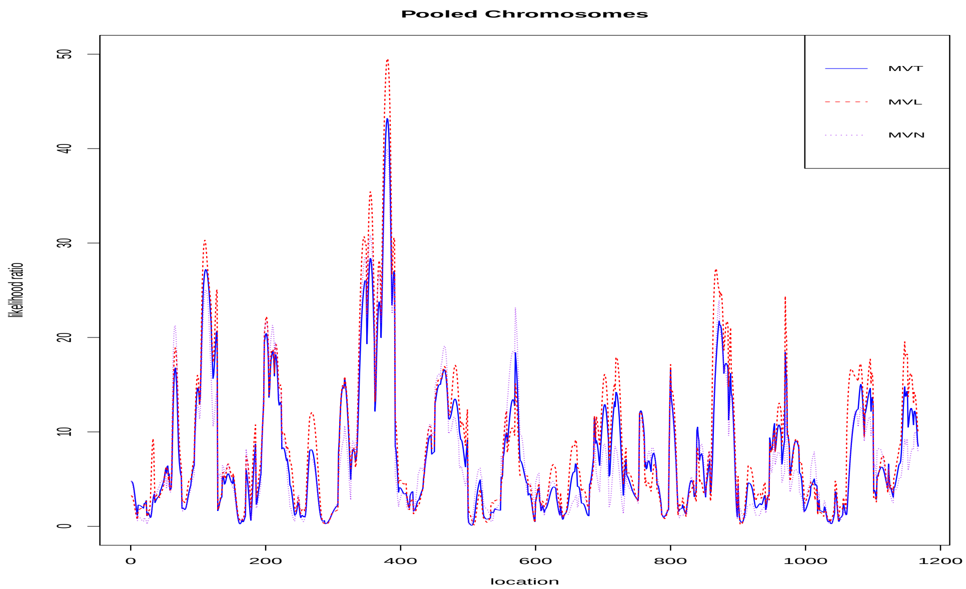


Figure 4.3: Pooled graph for real data: the LR profile plot across 12 chromosomes.

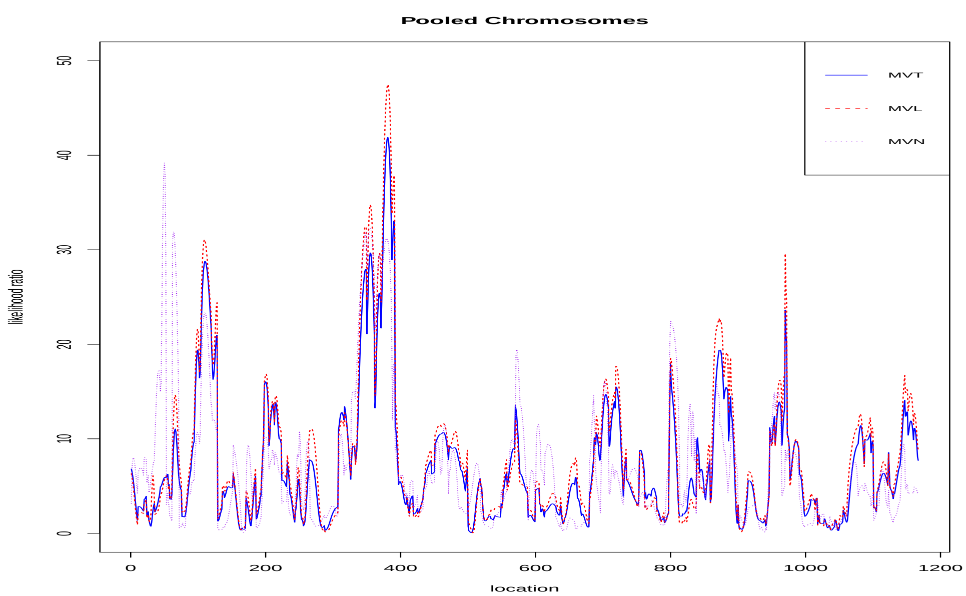


Figure 4.4: Pooled graph for sensitivity studies: the LR profile plot across 12 chromosomes with 10 artificial extreme observations.

		MVL method	MVT method	MVN method
QTL position	l	$264cM$	$264cM$	$264cM$
Qq parameters	$\hat{\eta}_{10}$	1.166	1.177	1.221
	$\hat{\eta}_{20}$	6.959	7.115	7.549
	$\hat{\eta}_{30}$	12.029	12.202	12.692
	$\hat{\eta}_{40}$	6.219	6.375	6.696
	$\hat{\eta}_{50}$	6.477	6.551	6.761
QQ parameters	$\hat{\eta}_{11}$	1.157	1.171	1.206
	$\hat{\eta}_{21}$	7.062	7.264	7.520
	$\hat{\eta}_{31}$	10.759	11.041	11.505
	$\hat{\eta}_{41}$	6.372	6.515	6.824
	$\hat{\eta}_{51}$	6.505	6.611	6.835
Covariance parameters	$\hat{\psi}$	0.788	0.782	0.808
	$\hat{\sigma}^2$	1.483	1.034	1.324

Table 4.5: QTL location and MLEs of estimated parameters with SAD(1) covariance structure

		MVL method	MVT method	MVN method
QTL position	l	$266cM$	$266cM$	$86cM$
Qq parameters	$\hat{\eta}_{10}$	1.166	1.173	1.223
	$\hat{\eta}_{20}$	6.946	7.063	7.531
	$\hat{\eta}_{30}$	12.018	12.147	12.682
	$\hat{\eta}_{40}$	6.216	6.332	6.705
	$\hat{\eta}_{50}$	6.470	6.527	6.763
QQ parameters	$\hat{\eta}_{11}$	1.169	1.176	1.266
	$\hat{\eta}_{21}$	7.166	7.304	7.855
	$\hat{\eta}_{31}$	10.900	11.089	12.000
	$\hat{\eta}_{41}$	6.480	6.564	7.164
	$\hat{\eta}_{51}$	6.084	6.660	7.115
Covariance parameters	$\hat{\psi}$	0.793	0.788	0.890
	$\hat{\sigma}^2$	1.677	1.034	1.636

Table 4.6: Sensitivity Studies: replace first two observations by artificial extreme values.

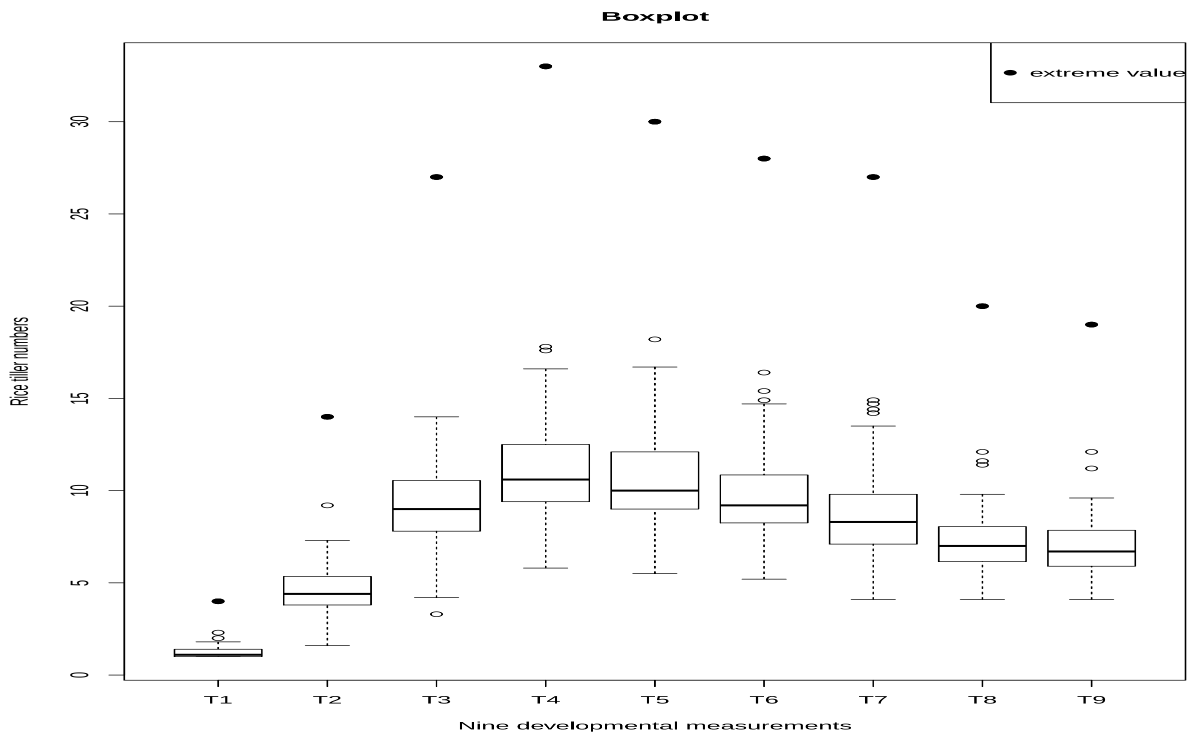


Figure 4.5: Boxplot of original data and artificial extreme values.

Chapter 5

Discussion

5.1 Summary

In this dissertation, we proposed a robust estimation procedure for two mixture multivariate regression models, the classic multivariate linear regression and the mixed effect linear regression models, by assuming a multivariate Laplace distributions for the regression errors and the random effects. Similar to t-distribution, the multivariate Laplace distribution is a scale mixture of the multivariate normal distribution, which enables us to construct an efficient EM algorithm to estimate the unknown parameters in the model.

On theoretical side, the ascent properties are proved for the proposed EM estimation procedures, and the performance of the proposed robust estimation procedures is evaluated by extensive simulation studies and sensitivity studies in both statistical models. Simulation results show that the proposed estimation procedures outperforms the estimation procedures based on the normality assumption. Comparing to the existing robust multivariate t-estimation procedure, the proposed method maintains robustness, and is computationally more efficient.

Finally, the proposed robust estimation procedure is applied to identify the QTL underlying a functional mapping framework with dynamic traits of agricultural or biomedical interest. Both simulation studies and real data analysis indicates that the proposed method performs better than the t and the normal based method when the data follow heavy-tailed distributions.

5.2 Future Work

The only concern raised in implementing the proposed robust estimation procedure is the unboundedness of the multivariate Laplace density function at the origin. In fact, for $p > 1$, $f_U(u) \to \infty$ as $u \to \mu$. This unboundedness makes the proposed EM algorithm potentially unstable since the weights, such as δ_{ij}, τ_{ij}, might become very volatile when Q_{ij} values are close to 0.

Recently we have identified some alternative definitions of the multivariate Laplace distribution which might be free from this issue. For example, the multivariate Laplace distribution defined in **?** has the density function

$$f_U(u) = \frac{|\Sigma|^{-1/2}}{2^p \pi^{(p-1)/2} \Gamma(\frac{p+1}{2})} \exp\left\{ -\sqrt{Q(u; \mu, \Sigma)} \right\}.$$

It is easy to see that the above density function is bounded at the origin. More importantly, the density function can be also written as a scale mixture of multivariate normal distribution with a inverse gamma scalar random variable, which is the key feature for designing an EM algorithm. This multivariate Laplace distribution certainly presents us some hope to remove the computational instability as seen in the proposed estimation procedures in this dissertation. This will be one of the topics in our future research along with this direction.

Appendix A

Proof of The Ascent Property: Theorem 1

Proof. Suppose the latent class variable G has probability mass function $P(G = j) = \pi_j, j = 1, 2, \ldots, g$. Let $\psi(v)$ be the density function of the exponential variable V. $\psi(v|y, x; \theta)$ as the conditional density function of V given $Y = y, X = x$, and $w(j, v|y, x, \theta)$ as the conditional joint mass-density function of G and V given $Y = y, X = x$, and θ. Then we have

$$w(j, v|y, x, \theta) = \pi_j \phi(y|\beta_j' x, v\Sigma_j)\psi(v)/f(y|x, \theta),$$

where $f(y|x, \theta)$ is defined in (2.1.2).

Note that for given $\theta^{(k)}$, we have $\sum_{j=1}^{g} \int w(j, v|Y_i, X_i, \theta^{(k)})dv = 1$, therefore,

$$\begin{aligned}
L_n(\theta) &= \sum_{i=1}^{n} \log \left[\sum_{j=1}^{g} \pi_j f_\varepsilon(Y_i - \beta_j' X_i, 0, \Sigma_j) \right] = \sum_{i=1}^{n} \left[\sum_{j=1}^{g} \int \log f(Y_i|X_i, \theta) w(j, v|Y_i, X_i, \theta^{(k)})dv \right] \\
&= \sum_{i=1}^{n} \left[\sum_{j=1}^{g} \int \log \left[\pi_j \phi(Y_i|\beta_j' X_i, v\Sigma_j)\psi(v) \right] w(j, v|Y_i, X_i, \theta^{(k)})dv \right] \\
&\qquad - \sum_{i=1}^{n} \left[\sum_{j=1}^{g} \int \log w(j, v|Y_i, X_i, \theta) w(j, v|Y_i, X_i, \theta^{(k)})dv \right] \\
&\hat{=} L_{n1}(\theta) - L_{n2}(\theta).
\end{aligned}$$

Recall the definition $\tau_{ij}^{(k+1)}$ in the E-step, and note that for any i, j, k,

$$\phi(Y_i|\beta_j^{(k)'} X_i, v\Sigma_j^{(k)})\psi(v) = \psi(v|Y_i, X_i, \theta_j^{(k)}) f_\varepsilon(Y_i - \beta_j^{(k)'} X_i, 0, \Sigma_j^{(k)}),$$

we have

$$\frac{\pi_j^{(k)}\phi(Y_i|\beta_j^{(k)'} X_i, v\Sigma_j^{(k)})\psi(v)}{f(Y_i|X_i, \theta^{(k)})} = \tau_{ij}^{(k+1)}\psi(v|Y_i, X_i; \theta_j^{(k)}).$$

Then by the definition of $w(j, v|y, x, \theta)$, we can show that $L_{n1}(\theta)$ is exactly the expression in (2.2.2). Therefore, the M-step in the proposed EM algorithm implies that $n^{-1}L_{n1}(\theta^{(k+1)}) \geq n^{-1}L_{n1}(\theta^{(k)})$.

Therefore, it suffices to show that in probability, $L_{n2}(\theta^{(k+1)}) \leq L_{n2}(\theta^{(k)})$. Note that the difference $L_{n2}(\theta^{(k+1)}) - L_{n2}(\theta^{(k)})$ is equivalent to

$$\begin{aligned}
&\sum_{i=1}^{n}\left[\sum_{j=1}^{g}\int \log\frac{w(j, v|Y_i, X_i, \theta^{(k+1)})}{w(j, v|Y_i, X_i, \theta^{(k)})} w(j, v|Y_i, X_i, \theta^{(k)})dv\right] \\
&= \sum_{i=1}^{n}\left[\sum_{j=1}^{g}\pi_j\int \log\frac{w(v|Y_i, G = j, X_i, \theta^{(k+1)})}{w(v|Y_i, G = j, X_i, \theta^{(k)})} w(v|Y_i, G = j, X_i, \theta^{(k)})dv\right]
\end{aligned}$$

which is less than 0 by the Kullback-Leibler information inequality applied to the conditional density function $w(v|Y_i, G = j, X_i, \theta^{(k)})$ for each i, j, k. □

Appendix B

Proof of The Ascent Property: Theorem 2

Proof. Suppose latent class variable C has probability mass function $P(C = j) = p_j$, $j = 1, 2, ..., G$. Let $\psi(v)$ denote the density function of exponential variable V and $\psi(v|y, x, z, \theta)$ as the conditional density function of V given $Y = y, X = x, Z = z$ and θ, $h(v|y, x)$ the conditional density function of V given $Y = y, X = x$. Then the conditional density functions of G, V, b, denoted by $w(j, v, b|y, x, z, \theta)$ given $Y = y, X = x, Z = z, \theta$, has the form of

$$\begin{aligned} w(j, v, b|y, x, z, \theta) =& \frac{f(C = j, v, b, y, x, z, \theta)}{f(y|x, z, \theta)} = \frac{P(C = j)f(v|C = j)f(y, b|v, C = j)}{f(y|x, z, \theta)} \\ =& \frac{P(C = j)f(v|C = j)f(y|b, v, C = j)f(b|v, C = j)}{f(y|x, z, \theta)} \\ =& \frac{p_j \psi(v)\phi(y|x\beta_j + zb_j, v\Sigma_j)\phi(b_j|0, v\Phi_j)}{f(y|x, z, \theta)}, \end{aligned}$$

where $f(y|x, z, \theta) = \sum_{j=1}^{G} p_j f_{ML}(y - x\beta_j, z\Phi_j z^T + \Sigma_j)$. Note that for given $\theta^{(k)}$, we have $\sum_{j=1}^{G} \iint w(j, v, b|y, x, z, \theta) dv db_j = 1$, therefore,

$$L_m(\theta) = \sum_{i=1}^{m} \log \left[\sum_{j=1}^{G} p_j f_{ML}(y_i - x_i\beta_j, z_i\Phi_j z_i^T + \Sigma_j) \right]$$

$$
\begin{aligned}
&= \sum_{i=1}^{m} \log f(y_i|x_i, z_i, \theta) \sum_{j=1}^{G} \iint w(j, b, v|y_i, x_i, z_i, \theta^{(k)}) dv db_j \\
&= \sum_{i=1}^{m} \sum_{j=1}^{G} \left[\iint \log f(y_i|x_i, z_i, \theta) w(j, b, v|y_i, x_i, z_i, \theta^{(k)}) dv db_j \right] \\
&= \sum_{i=1}^{m} \left\{ \sum_{j=1}^{G} \iint \log \left[p_j \phi(y_i|x_i\beta_j + z_i b_j, v\Sigma_j) \phi(b_j|0, v\Psi_j) \psi(v) \right] \right. \\
&\qquad \left. \times\, w(j, b, v|y_i, x_i, z_i, \theta^{(k)}) dv db_j \right\} \\
&\quad - \sum_{i=1}^{m} \left\{ \sum_{j=1}^{G} \iint \left[\log w(j, b, v|y_i, x_i, z_i, \theta) \right] w(j, b, v|y_i, x_i, z_i, \theta^{(k)}) dv db_j \right\} \\
&\hat{=} L_{m_1}(\theta) - L_{m_2}(\theta).
\end{aligned}
$$

Recall the definition of $p_{ij}^{(k+1)}$ in the E-step, and note that for any i, j, k,

$$
\begin{aligned}
&\phi(y_i|x_i\beta_j^{(k)} + z_i b_j, v\Sigma_j^{(k)}) \phi(b_j|0, v\Psi_j^{(k)}) \psi(v) \\
&\qquad = \phi(b_j|v) \psi(v|y_i, x_i, z_i) f_{ML}(y_i - x_i\beta_j^{(k)}, z_i \Psi_j^{(k)} z_i^T + \Sigma_j^{(k)}),
\end{aligned}
$$

we have

$$
\frac{p_j^{(k)} \phi(y_i|x_i\beta_j^{(k)} + z_i b_j, v\Sigma_j^{(k)}) \phi(b_j|0, v\Psi_j^{(k)}) \psi(v)}{f(y_i|x_i, \theta^{(k)})} = p_{ij}^{(k+1)} \phi(b_j|v, y_i) \psi(v|y_i, x_i, z_i).
$$

Therefore, $L_{m_1}(\theta)$ can be written as

$$
\begin{aligned}
&\sum_{i=1}^{m} \sum_{j=1}^{G} p_{ij}^{(k+1)} \iint \log \left[p_j \phi(y_i|x_i\beta_j + z_i b_j, v\Sigma_j) \phi(b_j|0, v\Psi_j) \psi(v) \right] \cdot \\
&\qquad \phi(b_j|v, y_i) \psi(v|y_i, x_i, z_i) dv db_j.
\end{aligned}
$$

Then by the definition of $w(j, v, b|y, x, z, \theta)$ we can show that $L_{m_1}(\theta)$ is exactly the expression of the conditional expectation of the complete log-likelihood function given the observed data and $\theta^{(k)}$. Therefore, the M-step in the proposed EM algorithm implies that

$L_{m_1}(\theta^{(k+1)}) \geqslant L_{m_1}(\theta^{(k)})$. Therefore, it suffices to show that in probability, $L_{m_2}(\theta^{(k+1)}) \leqslant L_{m_2}(\theta^{(k)})$. Note that the difference $L_{m_2}(\theta^{(k+1)}) - L_{m_2}(\theta^{(k)})$ is equal to

$$\sum_{i=1}^{m}\left[\sum_{j=1}^{G}\iint \log\frac{w(j,b,v|y_i,x_i,z_i,\theta^{(k+1)})}{w(j,b,v|y_i,x_i,z_i,\theta^{(k)})}w(j,b,v|y_i,x_i,z_i,\theta^{(k)})dvdb_j\right],$$

which can be written as

$$\sum_{i=1}^{m}\left[\sum_{j=1}^{G}p_j\iint \log\frac{w(b,v|y_i,x_i,z_i,C=j,\theta^{(k+1)})}{w(b,v|y_i,x_i,z_i,C=j,\theta^{(k)})}w(b,v|y_i,x_i,z_i,C=j,\theta^{(k)})dvdb_j\right].$$

Clearly, this expression is non-positive by applying the Kullback-Leibler information inequality to the conditional density function $w(b,v|y_i,C=j,x_i,z_i,\theta^{(k)})$ for each i,j,k. $\square$