Longitudinal data analysis with covariates measurement error

by

Md Erfanul Hoque

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Statistics
University of Manitoba
Winnipeg

Abstract

Longitudinal data occur frequently in medical studies and covariates measured by error are typical features of such data. Generalized linear mixed models (GLMMs) are commonly used to analyse longitudinal data. It is typically assumed that the random effects covariance matrix is constant across the subject (and among subjects) in these models. In many situations, however, this correlation structure may differ among subjects and ignoring this heterogeneity can cause the biased estimates of model parameters. In this thesis, following Lee et al. (2012), we propose an approach to properly model the random effects covariance matrix based on covariates in the class of GLMMs where we also have covariates measured by error. The resulting parameters from this decomposition have a sensible interpretation and can easily be modelled without the concern of positive definiteness of the resulting estimator. Performance of the proposed approach is evaluated through simulation studies which show that the proposed method performs very well in terms biases and mean square errors as well as coverage rates. The proposed method is also analysed using a data from Manitoba Follow-up Study.

KEY WORDS: Cholesky decomposition, Longitudinal data, Measurement error, Monte Carlo Expectation-maximization algorithm, Random effects

Acknowledgement

First and foremost, my utmost gratitude to my advisor Dr. Mahmoud Torabi for his invaluable support, guidance and mentorship throughout the course of the research and writing of this thesis. I would like to thank him for his extreme patience in the face of numerous obstacles about my thesis.

I would also like to thank Dr. Liqun Wang and Dr. Robert B. Tate for serving as my thesis committee members. I also wish to thank Dr. Robert B. Tate for providing the Manitoba Follow-up Study data set.

My earnest thanks also goes to Dr. Wasimul Bari, whose continuous support, helpful suggestions and cooperation has helped me a lot over the last few years.

Many thanks go to the faculty members and staffs in the Department of Statistics, University of Manitoba.

I owe many thanks to my classmates and friends, who have made my life as a student colorful and enjoyable. In addition, I am deeply indebted to my friends in Winnipeg, who take care of me as a family, especially Nahid, Niaz, Sajal, Aminul, Bappi and many others for their help and inspiration through-out this research work.

Finally, and most especially, I would like to express my heartfelt thanks to my family and my love, my beloved princess Fahmida Sultana for their constant support and inspiration during the whole time of the study. I love all of you a lot.

Dedication

This work is dedicated to my Father and Mother.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Longitudinal Studies

Longitudinal studies are very common in practice such as in health and life science, epidemiology, medical studies, and biomedical research. A study is known to be a longitudinal study when each individual in the study is followed over the period of time and for each individual, data are collected at multiple point of times. For example, blood pressure measurement of each individual may be measured repeatedly over time, or multiple quizzes may be taken for each student throughout the semester. Thus, the basic feature of such study is successive measurements on each of a number of individuals. A major benefit of longitudinal studies over cross-sectional studies is that one can study the changes of variable over time in longitudinal studies while in cross-sectional one can't.

Covariates in the longitudinal studies may be classified into two categories such as time-dependent covariates and time-independent covariates. The variables which vary over time within individuals are known to be time-dependent covariates. On the other hand, time-independent covariates are the factors which don't changes over time such as an individual's gender, race and other baseline factors. The fundamental goal in longitudinal studies is to investigate the effect of important predictor variables on individual response over time period.

The repeated measurements on the individuals e.g. from the same family in longitudinal studies are likely to be correlated. For example, members from the familial ancestry are genetically connected and their health conditions are generally correlated. Unlike the univariate case, in order to make valid inferences it is most important to take care of the correlation among the repeated measurements when analysing the data from these studies. Hence, there are two sources of variability in longitudinal data: within-individual variation, i.e., the random variation among the repeated measurements with each individual; and between-individual variation, i.e., the random variation in the data between different individuals. Moreover, the number of measurements and measurement times on each individual may vary individuals to individuals, i.e, the observed data are often unbalanced. Furthermore, in longitudinal studies, there may be early drop out on some individuals (subjects) for various reasons such as side effects of treatment. It is also known that blood pressures are usually measured with errors i.e, the observed values may vary from the actual (unobserved) values. These features in longitudinal studies indicate that the observed data are often complex or incomplete. Therefore, because of all these special characteristics of longitudinal data, statistical methods for analysing such data need special attention.

1.1.1 Methods for Analysis Longitudinal Data

Many model approaches have been developed to analyse longitudinal data and all these models can be classified into three broad categories: marginal models, transitional models, conditional models which includes random effect (mixed models).

Suppose that data contains n independent subjects. Let Y_{ij} be the response variable at time point j for subject $i, i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, m$. Let $\mathbf{X}_{ij}$ be the covariates whose effects are of interest. Generalized Linear Model(GLM) can be used to fit the data if the observations are independent of each others as follows

$$g(\mu_{ij}) = \mathbf{X}_{ij}^T \boldsymbol{\beta}, \tag{1.1}$$

where $\mu_{ij} = E[y_{ij}|\mathbf{X}_{ij}]$, $g(\cdot)$ is the link function that connects μ_{ij} to the linear predictors and $\boldsymbol{\beta}$ is the vector of regression parameters. For continuous responses, the link function is the identity link $g(\mu) = \mu$. For binary responses, the commonly used link functions are the logit link $g(\mu) = \log \frac{\mu}{1-\mu}$, log-log link $g(\mu) = \log\{-\log(1-\mu)\}$, and the probit link $g(\mu) = \Phi^{-1}(\mu)$, where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal variable. However, GLMs are no longer an appropriate solution in the presence of within-individual associations. In this section, we provide an overview of commonly used three general classes of models to analyse longitudinal data.

Marginal models

Marginal model approach has been widely used in longitudinal studies when inference about the population average are the main focus. In this case, firstly we model the marginal expectation of the response as a function of covariates without conditioning on other outcomes or unobserved random components and then the covariance structure of the response, i.e., the variance and the correlation of the response measurements are modelled separately. Marginal models are generally free of full parametric assumption for the joint distribution of the multivariate responses. A set of estimating equations, called generalized estimating equations (GEEs), introduced by Liang and Zeger (1986) are used to estimate the parameters in marginal model where the mean parameters are of the primary interest and the association between outcomes is considered as a nuisance feature. As a result, the marginal mean of the responses is modelled by the GEE approach assuming a common correlation structure across all the individuals.

Transition models

A Markov structure for the longitudinal process is assumed to model the within-individual correlation in a transition model. The idea is to assume that the present response observation depends on the past response observations given observed data. The previous response can be treated as additional covariates. A transition model of first-order Markov model for a longitudinal process may be written as

$$E(y_{ij}) = g(y_{i,j-1}, \mathbf{x}_i, \boldsymbol{\beta}),$$

where $g(\cdot)$ is a known link function. A class of marginalized models which specify a conditional model for the data generation underlying process is discussed by Zeger and Heagerty (2000). This models allow the estimation of marginal mean parameters. Heagerty (2002) generalized the model of Azzalini (1994) to a broad class of marginalized transition models (MTM) for the case of binary data. Chen et al., (2009) extended Heagerty's (2002) model for categorical data and this models permits marginal regression analysis and allows a general pth order dependence structure.

Mixed effects models

The mixed effects model approach assumes that the outcome is a function of covariates with regression parameters varying from one subject to another subject. In this case, random effects for each individual are introduced to incorporate the within-individual correlation and between-individual variation in the data. The measurement within the individual are correlated as each individual shares the same random effects. The two sources of variation in longitudinal data are specially incorporated by the mixed effects model. Thus, in addition to standard population-average inference, a mixed effects model allows individual-specific inference. A generalized linear mixed model (GLMM) is used which has the following form

$$g(\mu_{ij}^{u}) = \mathbf{X}_{ij}^{T}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{T}\mathbf{u}_{i}, \tag{1.2}$$

where $\boldsymbol{\beta}$ is the vector of fixed-effects parameters, $\mathbf{u}_i$ is the vector of random-effects associated with covariates $\mathbf{Z}_{ij}$, and μ_{ij}^{u} is the conditional expected value of Y_{ij}, the jth measured response on the ith subject. The vector of random effects $\mathbf{u}_i$ has a certain distribution, say, $f(\mathbf{u}_i)$ with variance σ_u. The main goal of statistical inference is to estimate the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_u)^T$, with primary interest in $\boldsymbol{\beta}$ (McCulloch and Searle, 2001). In case of continuous response and identity link function, a linear mixed model (LMM) is given by

$$Y_{ij} = \mathbf{X}_{ij}^T\boldsymbol{\beta} + \mathbf{Z}_{ij}^T\mathbf{u}_i + \epsilon_{ij}, \tag{1.3}$$

where the random error ϵ_{ij} is often assumed to have a normal distribution with mean 0 and variance σ_ϵ, and the random effects $\mathbf{u}_i$ that vary between subjects are assumed to follow a normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{D}(\boldsymbol{\theta})$.

For estimating the fixed-effects parameters and the parameters associated with random effects, the straightforward way is to use the maximum likelihood (ML) method based on marginal distribution of the observations. However, computing the likelihood function which involves integration over the random components is often difficult for GLMMs as it is not in closed form in most of the cases. In LMMs with normal variance components, some authors proposed iterative algorithms for computing the ML estimates or restricted maximum likelihood (REML) estimates (Harville, 1977; Fellner, 1986). The algorithm of Harville (1977) has been adopted by Schall (1991) to yield approximate ML or REML estimates in GLMMs. In standard applications e.g., the example of analysing the effect of air pollutants on pulmonary function development in children considered by Laird and Ware (1982), the random effects are assumed to be independent of covariates in these models. However, it is shown by Neuhaus and McCulloch (2006) that if there is correlation between the random effects and one of the covariates then naively fitting a GLMM ignoring this correlation leads to inconsistent estimators. To reduce the bias, the authors introduced conditional ML method by partitioning the covariates into between- and within-individual components. Moreover, a verity of methods have been proposed for fitting the GLMMs, including Monte-Carlo EM (McCulloch, 1994), Laplace approximations (Liu and Pierce, 1993; Breslow and Lin, 1995), Penalized Quasilikelihood (Breslow and Clayton, 1993), Corrected Penalized Quasilikelihood (Lin and Breslow, 1996), data cloning (Lele et al., 2010) and Bayesian procedures including EM-type algorithms (Stiratelli, Laird and Ware, 1984) and Gibbs sampler (Zeger and Karim, 1991).

1.2 Measurement Error

Measurement error is a common concern in many scientific research. In case of analysing correlated data, the effect of covariate measurement error is a serious problem. It is known that the variables obtained from self-reported questionnaires contain error, e.g., dietary intake, and nutritional consumption. One of the major sources of measurement error in data is self-reported bias. It has been well established in the literature that many covariates of medical interests, such as blood pressure (Carrol, Ruppert, and Stefanski, 1995), urinary sodium chloride (USC) level (Wang, Carrol, and Liang, 1996), and exposure to indoor or outdoor pollutants (Tosteson, Stefanski, and Schafer, 1989) are often subject to measurement error. In the presence of covariates measurement errors, naive estimators for the model parameters are often inconsistent (Fuller, 2009; Cook and Stefanski, 1994; and Prentice, 1982).

We often need to assume an error structure to address the covariate measurement errors. Let W_{ij} be the observed covariate value for the individual i at time point j, with possible measurement errors, and let X_{ij} be the corresponding unobserved true covariate value. There are two way of defining the relationship between X_{ij} and W_{ij} (Carroll et al., 2006): one models the dependence of X_{ij} on W_{ij}, and the other models the dependence of W_{ij} on X_{ij}, given the other variables.

The *Classical additive measurement error model* assumes that

$$W_{ij} = X_{ij} + e_{ij},$$

where e_{ij} is the measurement error for individual i at time point j and independent and identically distributed (i.i.d) with mean 0 and variance σ_e^2, and are independent of X_{ij}. Often, e_{ij} also assume to follow a multivariate normal distribution.
The *Berkson additive measurement error model* (Berkson, 1950) assumes that

$$X_{ij} = W_{ij} + e_{ij},$$

where e_{ij} is the measurement error for individual i at time point j and independent and identically distributed (i.i.d) with mean 0 and variance σ_e^2, and are independent of W_{ij}. Often, e_{ij} assume to follow a multivariate normal distribution.
The *Multiplicative measurement error model* is given by

$$W_{ij} = X_{ij}e_{ij}.$$

One can also define,

$$X_{ij} = W_{ij}e_{ij}.$$

The choice of additive or multiplicative measurement error model depends on how much variation around the mean come into the observation. This may well come in multiplicatively or additively. Log-sacle of multiplicative model converts to the additive model which is convenient to deal with. In this case the log of error variable, e_{ij}, has normal distribution with mean 0 and variance σ_e^2 but in multiplicative model this is not an obvious case. However, sometimes researchers are not interested to work in log-scale.

The *Structural measurement error model* assumes a strong parametric assumption about the distribution of the true covariate X_{ij}. That means in this case X_{ij}'s are random variables. Structural model can be viewed as an empirical Bayes method (Whittemore, 1989). Likelihood methods and Bayesian methods are commonly used in this case.
On the other hand, the *Functional measurement error model* assumes no distributional assumption on true covariates X_{ij}, that means X_{ij}'s are fixed constant. Regression calibration and simulation extrapolation (SIMEX) are the most common used methods in this case.

The classical error model is the mostly used in practice. However, it is important to note that choice of error model depends only on the data at hand. Carroll et al. (2006) provided the following suggestions: we can use classical measurement error model if the error-prone covariate is measured uniquely to an individual such as blood pressure measurements; on the other hand, if same value of the error-prone covariate are provided to all the individuals in a group but the true covariate value is specific to an individual, such as a person's actual exposure to air pollution in a city of people exposed to the same level of air pollution in that city, then Berkson measurement error model can be used. In other words, if the observed covariate over-estimates the true covariate, we use the classical measurement error model, otherwise, one can use the Berkson measurement error model.

Various types of measurement error can arise in practice. It is important to distinguish between differential and non-differential measurement error. It can be said that the error in the observed value W is non-differential if W contains no additional information about Y with respect to X. In this case, W is said to be a surrogate for X if the conditional independence of Y and W exist, that is, Y and W are independent given $X = x$. Otherwise the measurement error is said to be differential.

1.2.1 Effect of Measurement error in parameter estimation

The naive inference procedure leads to biased estimates of regression coefficients in the presence of measurement error. To demonstrate the impact of measurement error on the parameter estimation we conduct a simple simulation study. We consider a classical simple linear regression model with an error-prone covariate as follows

$$Y_i = \beta_0 + \beta_x X_i + \epsilon_i; \quad i = 1, 2, \ldots, n = 100, \tag{1.4}$$

where $X_i \sim \mathrm{N}(\mu_x, \sigma_x^2)$ and $\epsilon_i \sim \mathrm{N}(0, \sigma_\epsilon^2)$. Let X_i be the error-prone covariate and let us consider a classical additive structural measurement error model for X_i as follows

$$W_i = X_i + e_i; \quad i = 1, 2, \ldots, n, \tag{1.5}$$

where W_i is the surrogate of variable X_i and e_i's are assumed to be independent and identically distributed with mean 0 and variance σ_e^2. Also, we assume that X_i, W_i and ϵ_i are mutually independent.

Now for simulation purpose, we generate X_i, $i = 1, 2, \ldots, 100$ from $\mathrm{N}(\mu_x = 0, \sigma_x^2 = 1)$ and Y_i from (1.4) for which we consider the true value of regression coefficients are $\beta_0 = 0$, $\beta_x = 1$ and $\epsilon_i \sim \mathrm{N}(0, 1/4)$. Moreover, we generate $e_1, e_2, \ldots, e_{100}$ from $\mathrm{N}(0, \sigma_e^2 = 1)$ for Figure 1.1 and consider the additive error model (1.5) for X_i. We get the naive estimate of the linear regression model to the observed data $\{(Y_i, W_i); i = 1, 2, \ldots, 100\}$.

Firstly, we regress Y on X and Y on W and then compare them to see the effects of measurement error by Figure 1.1. Based on Figure 1.1, we can see the bias as the slope for the error-prone data is too small in absolute value. Moreover, the variability about the line is much larger which indicate the excess variance of the error-prone data and definitely excess variance indicates the loss of power.

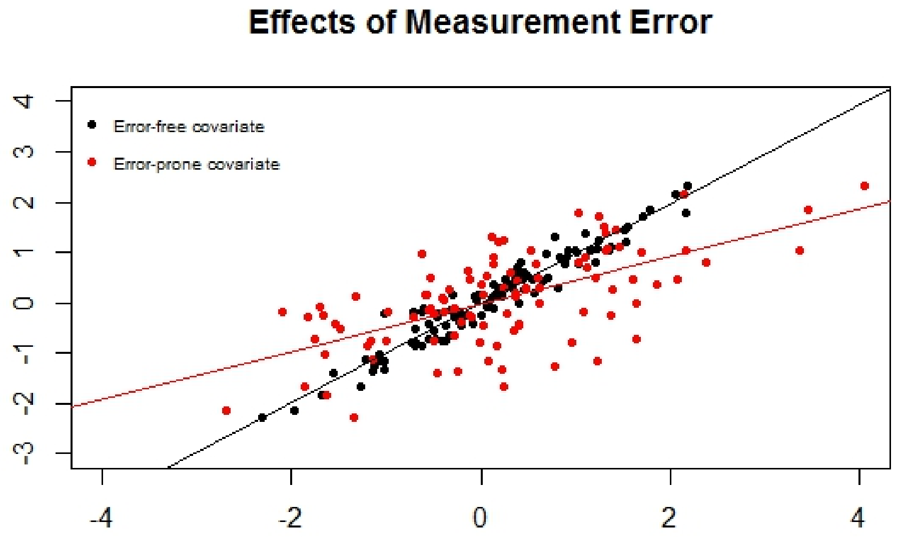

Figure 1.1: Illustration of the impact of with and without measurement error on data

Secondly, we vary σ_e within $[0, 2]$ to reflect the variation of the degree of measurement error. we simulate the data for 200 times for each value of σ_e and record the empirical averages of the naive estimator and finally plot them in Figure 1.2. It is obvious from Figure 1.2 that the measurement error causes bias the naive estimator towards zero. This is known as so-called attenuation phenomenon. This attenuation factor depends on the magnitude of the measurement error variance. If the variance is larger, the attenuation will increase. With the large error in the reliability ratio (Carroll et al., 2006) $\frac{\sigma_x^2}{(\sigma_x^2 + \sigma_e^2)} \leq 0.5$ with σ_x^2 denoting the variance of X_i, the bias of the naive estimator compared to the true estimator is over 60%.

At the same time it can be observed from Figure 1.2 that the classical measurement error causes loss of power. It means that the increase of variance of measurement error causes the power loss.

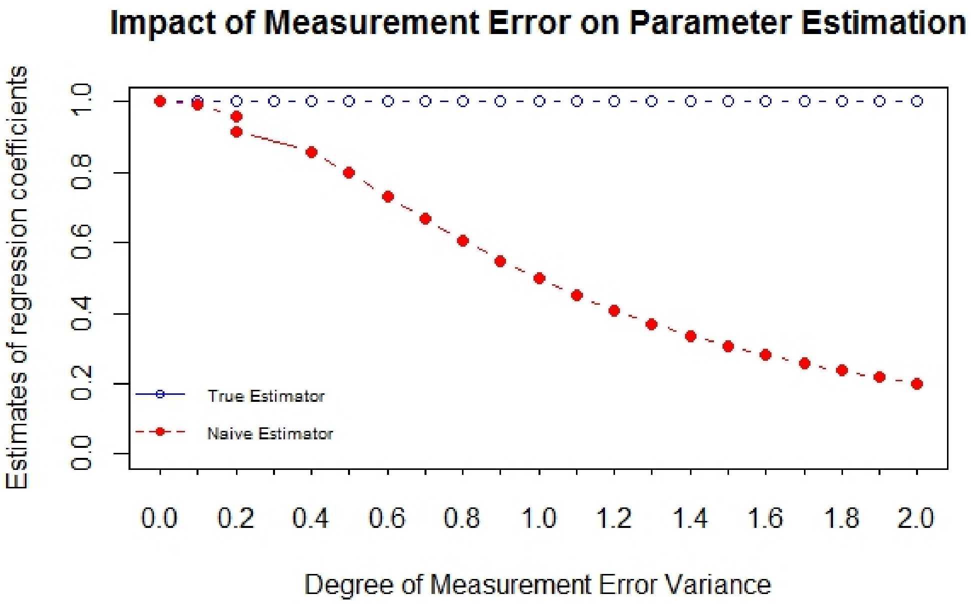

Figure 1.2: Demonstration of the effect of measurement error on regression coefficients

The attenuation phenomenon is confirmed theoretically as follows. Under the non-differential measurement error assumption and classical additive error model, naively fitting the linear model (1.4) to the observed data $\{(Y_i, W_i);\, i = 1, 2, \ldots, 100\}$ tends to a misspecified model

$$Y_i = \beta_0^* + \beta_x^* X_i + \epsilon_i^*; \quad i = 1, 2, \ldots, n, \tag{1.6}$$

where β_0^* and β_x^* are regression coefficients under the wrong model and $\epsilon_i^* \sim$

$\mathrm{N}(0, \sigma_\epsilon^*)$. The naive least square estimate for β_x is then given by

$$\hat{\beta}_x^* = \frac{\sum_{i=1}^n (W_i - \bar{W})(Y_i - \bar{Y})}{\sum_{i=1}^n (W_i - \bar{W})^2},$$

where $\bar{W} = \frac{\sum_{i=1}^n W_i}{n}$ and $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$.

After some algebra we can write the naive least square estimator as

$$\begin{aligned}
\hat{\beta}_x^* &= \frac{\sum_{i=1}^n (W_i - \bar{W})\{\beta_x(X_i - \bar{X}) + (\epsilon_i - \bar{\epsilon})\}}{\sum_{i=1}^n (W_i - \bar{W})^2} \\
&= \beta_x \frac{\sum_{i=1}^n (W_i - \bar{W})(X_i - \bar{X})}{\sum_{i=1}^n (W_i - \bar{W})^2} + \frac{\sum_{i=1}^n (W_i - \bar{W})(\epsilon_i - \bar{\epsilon})}{\sum_{i=1}^n (W_i - \bar{W})^2} \\
&= \beta_x \frac{\sum_{i=1}^n (X_i - \bar{X} + e_i - \bar{e})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X} + e_i - \bar{e})^2} + \frac{\sum_{i=1}^n (W_i - \bar{W})(\epsilon_i - \bar{\epsilon})}{\sum_{i=1}^n (W_i - \bar{W})^2} \\
&= \beta_x \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (X_i - \bar{X})(e_i - \bar{e})}{\sum_{i=1}^n (X_i - \bar{X})^2 + 2\sum_{i=1}^n (X_i - \bar{X})(e_i - \bar{e}) + \sum_{i=1}^n (e_i - \bar{e})^2} \\
&+ \frac{\sum_{i=1}^n (W_i - \bar{W})(\epsilon_i - \bar{\epsilon})}{\sum_{i=1}^n (W_i - \bar{W})^2} \\
&\xrightarrow{p} \beta_x \Big(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_e^2} \Big), \quad \text{as } n \to \infty,
\end{aligned}$$

where independence between X_i and e_i and independence between W_i and ϵ_i are assumed for the convergence in probability. Therefore, the regression coefficient estimate with the mismeasured covariate using the naive analysis procedure tends to be attenuated estimate. This attenuation increases with the magnitude of the variance of the measurement error.

However, Berkson error model causes little or no bias in the estimates of regression parameters compared to classical additive error model in case of linear regression. The reason for this is the measurement error e_i is simply imbibed into ϵ_i in the response model. That is,

$$\begin{aligned} Y_i &= \beta_0 + \beta_x W_i + (\beta_x e_i + \epsilon_i) \quad \text{as } X_i = W_i + e_i, \text{ under Berkson error model} \\ &= \beta_0 + \beta_x W_i + \epsilon_i^*, \end{aligned}$$

where ϵ_i^* has variance $(\beta_x^2 \sigma_e^2 + \sigma_\epsilon^2)$.

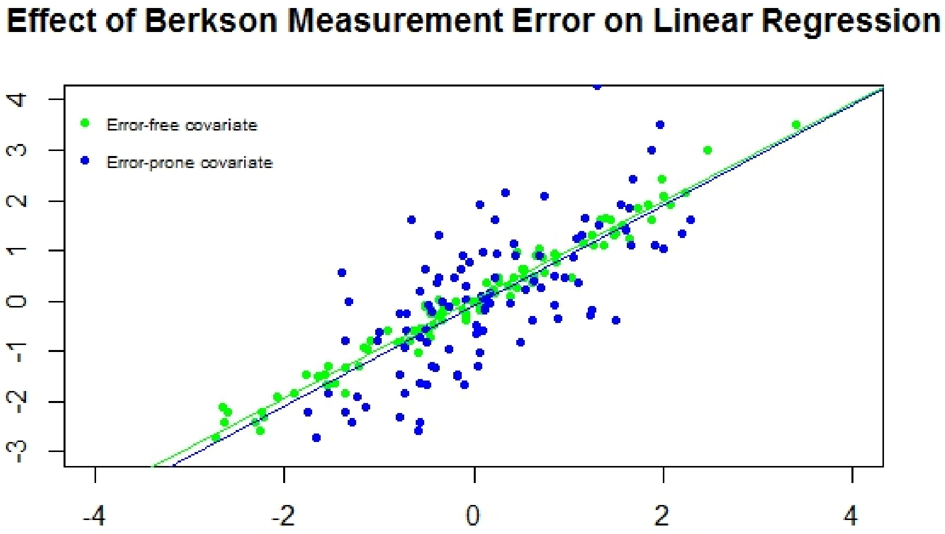

Figure 1.3: Impact of Berkson measurement error on simple linear regression

This can be obviously seen from Figure (1.3) which illustrates the unbiasedness of the regression parameter. That is the fit of Y_i on W_i and Y_i on X_i are, in fact, almost similar which indicates that the measurement error is ignorable in this situation. Figure 1.3 is drawn using the same simulated data used for classical measurement error under Section 1.2.1. The Berkson error model also decreases the power of a study because of the random error variance inflation.

1.2.2 Methods for handling measurement error

There have been numerous methods for the correction of bias caused by measurement error in literatures. These approaches are referred as functional methods and structural methods based on the distributional assumption of the unobserved variable, availability of additional data about the unobserved variable and the parametric or non-parametric assumption of the approach. Comprehensive reviews of the covariate measurement error methods are provided in Fuller (2009), Gustafson (2004), and Carroll et al. (2006). Some popular used methods for covariate measurement errors in regression include the corrected scores approach of Nakamura (1990, 1992), regression calibration, the simulation-extrapolation (SIMEX) approach (Cook and Stefanski, 1994), likelihood methods, approximation methods and Bayesian methods. When it is important to specify a marginal distribution of the error-prone covariate or it is of interest to study the marginal behaviour of error variables, then the structural methods are needed. Minimal distributional assumption of the unobserved covariates are needed for the regression calibration and SIMEX method. On the other hand, likelihood methods and Bayesian methods make strong distributional assumption on the unobserved covariates, so they are more efficient if the specification of the covariate distribution is correct.

1.2.3 Mixed measurement error models

Mixed effects models have become increasingly popular to model cluster dependence in which the response can be defined as a function of two effects: fixed effects and unobserved cluster specific random effects and error terms. The data within the same cluster are correlated as they share common random effects. Mixed effects models contain two types of parameters: fixed effects which associate with the mean effects of predictors on the response; random effects which indicate the cluster effects on the repeated measurements in corresponding clusters.

In linear mixed effects models, the fixed effects and the random effects have a linear combination to the response. For longitudinal data, classical linear regression is inappropriate because of the correlation of the observation within each cluster. Linear mixed effects models can obtain from classical linear regression models by introducing the random effects to incorporate the within-cluster correlation and between-cluster variation. The magnitude of the random effects measure the between and within cluster-variations. As the linear mixed models (LMM) incorporate two types of variation, it can be interpreted as a hierarchical two-stage model where within-individual variation can be specified in first stage and between-individual variation in second stage (Wu, 2009).

In a generalized linear mixed models given the random effects, the responses are assumed to have a distribution (normal, binomial, etc.) and mean of the responses is related with the covariates through a generalized linear models (GLMs). Hence, GLMMs are an extension of linear mixed models to allows response variables from different distribution. Moreover, it can be said that the GLMMs are the extension of the GLMs by including both the fixed and random effects.

For example, let Y_{ij} be the response of systolic blood pressure for individual i at the j yearly visit in a clinic and Z_{ij} be the age of that individual. This is a linear relationship between these two variables with a subject-specific intercept and slope. Then a suitable LMM can be written as

$$Y_{ij} = \beta_0 + Z_{ij}\beta_z + \mathbf{A}_{ij}\mathbf{u}_i + \epsilon_{ij} \tag{1.7}$$

where $\mathbf{A}_{ij} = (1, Z_{ij})$ and $\mathbf{u}_i = (u_{0,i}, u_{1,i})^T$. Here, β_0 and β_z are the intercept and slope across the population and $u_{0,i}$ and $u_{z,i}$ are the deviations of intercept and slope from the average for the subject i. The matrix of covariance components is

$$\mathbf{D}(\boldsymbol{\theta}) = \begin{bmatrix} \mathrm{var}(u_{0,i}) & \mathrm{cov}(u_{0,i}, u_{1,i}) \\ \mathrm{cov}(u_{0,i}, u_{1,i}) & \mathrm{var}(u_{1,i}) \end{bmatrix}.$$

Now suppose that the response variable Y_{ij} is related to a true nutrition intake, X_{ij}, over the past year. If we consider the regression coefficient of X_{ij} as fixed effect, that is, independent of the subject, then the LMM can be used to model the relationship between Y_{ij}, X_{ij} and Z_{ij} as follows

$$Y_{ij} = \beta_0 + X_{ij}\beta_x + Z_{{ij}}\beta_z + \mathbf{A}_{ij}\mathbf{u}_i + \epsilon_{ij}. \tag{1.8}$$

If we assume that the true intake is unobserved and the observed intake is W_{ij}, then we have a linear mixed measurement error model (LMMeM). The LMMeM can be fitted by using standard methods for LMM given W_{ij}. Now the GLMMs of Y_{ij} given X_{ij} and Z_{ij} can be written as

$$g(\mu_{ij,x}^{u_i}) = \beta_0 + X_{ij}\beta_x + Z_{ij}\beta_z + \mathbf{A}_{ij}\mathbf{u}_i, \tag{1.9}$$

Here $\mu_{ij,x}^{u_i}$ is the expected value of Y_{ij} given the random effects, and the $\mathbf{u}_i$ are i.i.d normal random vectors with mean zero and covariance matrix $\mathbf{D}(\boldsymbol{\theta})$. For the measurement error effects on a GLMM for Y_{ij}, X_{ij} and Z_{ij}, a new GLMM relating Y_{ij} to W_{ij} and Z_{ij}, by assuming the additive measurement error and normal structural model. This new models are known as generalized linear measurement error model (GLMMeM) (Wang et al., 1998). To illustrate this, we assume X_{ij} are mutually independent and follows normal distribution with μ_x and σ_x^2, and independent of Z_{ij}. And then the models in Wang et al. (1998) are obtained under the assumption of classical structural additive errors, that is,

$$W_{ij} = X_{ij} + e_{ij},$$

where the e_{ij} are independent and normally distributed with mean zero and covariance matrix σ_e^2. One can also consider the multiple version of variables in the above equations.

1.3 Modelling the Random Effects Covariance Matrix in Mixed Model

Random effects (mixed) models are a common class of models used frequently to analyse longitudinal data. These models offer many advantages such as ability to handle various observation times across individuals and permit non-stationary covariance structures. However, little attention has been taken to modelling the random effects covariance matrix in these models. In particular, in modelling, it is often neglected whether this variance-covariance matrix is the same for the all individuals/subjects or whether it varies from subject to subject. Hence, in these models, typically it is assumed that the random effects covariance matrix is constant across the subjects.

For analysing discrete longitudinal data, GLMMs are commonly used and in these models, biased estimates of the fixed effects can result by ignoring this heterogeneity (Heagerty and Kurland, 2001). For continuous longitudinal data, the inferences of the parameters and the standard errors for the fixed and random effects will be incorrect. Moreover, in the presence of missing data or covariate measurement errors, incorrectly modelling the covariance structure can result in biased estimates of fixed effects.

Many authors have discussed the issue of accounting the heterogeneity in covariance matrix. Chiu et al. (1996) model the covariance matrix using a log matrix parametrization in marginal models and obtain estimates using estimating equations. Using the modified Cholesky decomposition, Pourahmadi (1999, 2000) developed random effects covariance matrix depending on subject-specific covariates. Following the idea of modified Cholesky decomposition, Daniels and Pourahmadi (2002) develop a class of dynamic conditionally mixed models by allowing to vary the marginal covariance matrix across subjects. However, they consider the random effects covariance matrix to be fixed across individuals. Daniels and Zhao (2003) and Pourahmadi and Daniels (2002) propose the similar type of modelling for random effects covariance matrix. Lin et al. (1997) examine heterogeneity in the within-individual variance in linear mixed models. To deal with unbalanced longitudinal data, Pourahmadi's (Pourahmadi , 1999, 2000) method is generalized by Pan and Mackenzie (2003, 2006). Recently, Lee et al. (2012) develop a heterogeneous random effects covariance matrix for GLMMs which depends on covariates. The modified Cholesky decomposition is used in their work to obtain the random effects covariance matrix. However, to our knowledge no work has been done on modelling the random effects covariance matrix for GLMMs with covariate measurement errors.

1.4 Contribution of the Dissertation

In many medical studies, longitudinal data or repeated measurement data occur frequently where often changes in a particular characteristic in the participating individuals are investigated by observing repeatedly over time. The GLMMs are commonly practiced to analyse such data (Breslow and Clyton, 1993; McCulloch and Searle, 2001; Diggle et al., 2002) and it enables us to account for between and within individuals heterogeneity. It is an important requirement that variables are perfectly measured for the validity of inferential methods. However, in practice, longitudinal data are prone to be not perfect and seriously biased results can be led by ignoring this.

Covariate measurement error is a common typical feature of longitudinal study (Carroll et al., 2006). There have been considerable works for the analysis of longitudinal data with missing values in the literature (Ibrahim et al., 1999, 2001; Yi and Cook, 2002; Molenberghs and Kenward, 2007; Wu et al., 2009). Recently, the attention has been increased to address the effects of covariate measurement error in the analysis of longitudinal data (Carroll et al., 2006, ch 11).

Extensive work has been done on covariates measurement error (Cook and Stefanski, 1994; Carroll et al., 1995; Lin et al., 1996; Wang et al., 1998, Fuller, 2009; Torabi, 2013). The GLMMs framework are adopted to make the inference procedure in most of the cases. In these models, it is assumed that the random effects covariance matrix is constant across the subject and also the high dimensionality and positive definite constraints make the structure of random effects covariance matrix restricted. However, the covariance matrix may vary by measured covariates in many situation, and biased estimates of the fixed effects may result by ignoring this heterogeneity (Heagerty and Kurland, 2001). In 2011, Yi et al. presented a fairly general framework to make inference for longitudinal data with covariates measurement error and missing responses simultaneously by adopting the framework of GLMMs. They have employed the EM algorithm to conduct inference for parameters of interest (Meng and Van Dyk, 1998). However, in their work the random effect covariance matrix has been left unspecified. Recently, Lee et al. (2012) introduced a heterogeneous random effects covariance matrix for GLMMs by using modified Cholesky decomposition (Pourahmadi, 1999, 2000) which depends on covariates.

In this thesis, our goal is to properly model the random effects covariance matrix under the framework of GLMMs with covariates measurement errors. For this purpose, we extend the model introduced by Lee et al. (2012) for the random effects covariance matrix for the GLMMs to the case when the covariates are also subject to measurement error. An important benefit of using random effects covariance matrix by defining covariates and Cholesky decomposition is that each subject has specific covariance structure and also our proposed model accounts for the covariates measured with error.

1.5 Outline of the Thesis

The outline of the thesis is as follows. In Chapter 2, the Proposed approach to model the random effects covariance matrix for GLMMs with covariates measurement error is given. Also the general framework is provided to make the inference procedure to estimate model parameters. Chapter 3 contains the simulation study to assess the performance of the different methods (Proposed method, Naive 2 method and Naive 1 method). In Chapter 4, a real data set from Manitoba Follow-up Study is analysed. We make some conclusions based on the Proposed method and discuss some future work direction in Chapter 5.

Chapter 2

Theory and Methods

2.1 Notation and Model Specification

The model of interest specifies repeated measures on each of n subjects with responses that follow a generalized linear model with random intercepts for each subjects, with time-specific covariates that is subject to error and time-specific error free covariates. Suppose that Y_{ij} be the response variable at time point j for subject i. Let $\mathbf{X}_{ij}$ be the vector of error-prone covariates, and $\mathbf{Z}_{ij}$ be the vector of error-free covariates, $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, m$. Further, let $\mathbf{W}_{ij}$ be an observed version of $\mathbf{X}_{ij}$. Denote the response vector for the ith subject by $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im})^T$ and also denote $\mathbf{X}_i = (\mathbf{X}_{i1}^T, \ldots, \mathbf{X}_{im}^T)^T$, $\mathbf{Z}_i = (\mathbf{Z}_{i1}^T, \ldots, \mathbf{Z}_{im}^T)^T$, and $\mathbf{W}_i = (\mathbf{W}_{i1}^T, \ldots, \mathbf{W}_{im}^T)^T$.

2.2 Response Process

We assume that Y_{ij} follows a conditional distribution in the exponential family given the random effects $\mathbf{u}_i$, taking the form

$$f(y_{ij}|\mathbf{u}_i;\eta_{ij},\phi) = \exp\{(y_{ij}\eta_{ij} - b(\eta_{ij}))/a(\phi) + c(y_{ij},\phi)\}, \tag{2.1}$$

where, $a(\cdot)$, $b(\cdot)$ and $c(\cdot,\cdot)$ are known functions and parameters η_{ij} can be further modelled to accommodate within-subject variability. The ϕ is the dispersion parameter that is known or to be estimated such as (e.g.) $\phi = 1$ for binary response. To emphasize on the estimation of parameters of interest we treat ϕ as known here. We assume $\mathbf{Y}_i$ follows the GLMM and model a transformation of the mean as a linear model in both fixed and random factors:

$$\mu_{ij} = E[Y_{ij}|\mathbf{u_i}]$$

$$g(\mu_{ij}) = \beta_0 + \mathbf{X}_{ij}^T\boldsymbol{\beta}_x + \mathbf{Z}_{ij}^T\boldsymbol{\beta}_z + \mathbf{u_i}, \tag{2.2}$$

where $g(\cdot)$ is the link function and $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_x^T, \boldsymbol{\beta}_z^T)^T$ is the fixed vector of regression parameters.

Also suppose that the random effects, $\mathbf{u}_i$, are independent and identically distributed, and independent of the explanatory variables. Here,

$$\mathbf{u}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i),$$

$\mathbf{u}_i = (u_{i1}, \ldots, u_{im})^T$ is a $m \times 1$-dimensional vector of random effects and random effects covariance matrix $\boldsymbol{\Sigma}_i$ is indexed by subject i.

We also assume, $f(\mathbf{y}_i|\mathbf{u}_i, \mathbf{x}_i, \mathbf{w}_i, \mathbf{z}_i) = f(\mathbf{y}_i|\mathbf{u}_i, \mathbf{x}_i, \mathbf{z}_i)$, which is related to the usual non-differential error mechanism (Carroll et al., 2006, p. 36), but is different due to the dependence on the random effects.

2.3 Measurement Error Process

To feature the measurement error process we employ a multiple regression model (Carroll et al., 2006, p. 27) as follows:

$$\mathbf{W}_{ij} = \gamma_0 + \mathbf{\Gamma}_x \mathbf{X}_{ij} + \mathbf{\Gamma}_z \mathbf{Z}_{ij} + \mathbf{e}_{ij}\,, \tag{2.3}$$

where error terms $\mathbf{e}_{ij}^T$s are independent of $\mathbf{X}_{ij}$, $\mathbf{Z}_{ij}$, and the responses as well as random effects $\mathbf{u_i}$. Also, $\mathbf{e}_{ij}$ follow a distribution, $f(\mathbf{e}_{ij}, \boldsymbol{\tau})$, where $\boldsymbol{\tau}$ is the associated parameters. It is often assumed that $\mathbf{e}_{ij}$ has zero mean. Let $\gamma_0 = (\gamma_{01}, \ldots, \gamma_{0p})^T$ be the vector of intercept coefficients, $\mathbf{\Gamma}_x = (\Gamma_{x1}, \ldots, \Gamma_{xp})^T$ and $\mathbf{\Gamma}_z = (\Gamma_{z1}, \ldots, \Gamma_{zq})^T$ denote the vector of regression coefficients, respectively. Here, p and q are the dimensions of $\mathbf{W}_{ij}$ and $\mathbf{Z}_{ij}$, respectively. Also let $\gamma = (\mathbf{\Gamma}_0^T, \mathbf{\Gamma}_x^T, \mathbf{\Gamma}_z^T)^T$ be the vector including all the regression coefficients. By setting $\mathbf{\Gamma}_0 = 0, \mathbf{\Gamma}_z = 0 \,\&\, \mathbf{\Gamma}_x = \mathbf{I}_p$, where $\mathbf{I}_p$ is the $p \times p$ identity matrix, the above model (2.3) can be written as classical additive error model (Carrol et al., 2006, p.27):

$$\mathbf{W}_{ij} = \mathbf{X}_{ij} + \mathbf{e}_{ij}.$$

Another class of measurement error models named Berksn error models can be obtained by notationally switching $\mathbf{W}_{ij}$ and $\mathbf{X}_{ij}$ in (2.3). That is, $\mathbf{X}_{ij} = \mathbf{W}_{ij} + \mathbf{e}_{ij}$.

2.4 A Model for the Random Effects Covariance Matrix

The main contribution in this thesis is to use a specified model for the random effects covariance matrix for GLMMs with covariates measurement error where heterogeneity of the random effects covariance matrix is ignored. For this purpose we employ the structure of the random effects covariance matrix given by Lee et al. (2012). They propose a heterogeneous random effects covariance matrix for GLMMs which depends on subject-specific covariates. Actually, to model $\mathbf{\Sigma}_i$, random effects covariance matrix, they decompose the random effects covariance matrix based on the modified Cholesky decomposition (Pourahmadi, 1999, 2000) which results in a set of dependence parameters, generalized autoregressive parameters (GARPs) and a set of variance parameters, innovation variances (IVs). The basic idea of this proposed structure is that the covariance matrix $\mathbf{\Sigma}_i$ of the random effects vector $\mathbf{u}_i$, can be diagonalized by a lower triangular matrix which is constructed from the regression coefficients when u_{ij} is regressed on its predecessors $u_{i1}, u_{i2}, \ldots, u_{ij-1}$. More specifically, it can be written as:

$$u_{i1} = \epsilon_{i1} \tag{2.4}$$

$$u_{ij} = \sum_{t=1}^{j-1} \phi_{i,jt} u_{it} + \epsilon_{ij}, \quad \text{for } j = 2, 3, \ldots, m, \tag{2.5}$$

where, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{im})^T \sim N(\mathbf{0}, \mathbf{D}_i)$ with $\mathbf{D}_i = \text{diag}(\sigma^2_{i1}, \sigma^2_{i2}, \ldots, \sigma^2_{im})$. Then for the $j = 2, 3, \ldots, m$ we can write (2.4) and (2.5) in matrix form as follows:

$$\mathbf{T}_i \mathbf{u}_i = \boldsymbol{\epsilon}_i, \tag{2.6}$$

where $\mathbf{T}_i$ is a unit lower triangular matrix having ones on its diagonal and $-\phi_{i,jt}$ in the (j,t)th position for $2 \leq j \leq m$ and $t = 1, 2, \ldots, j-1$.

From (2.6) we can write

$$\mathbf{T}_i \boldsymbol{\Sigma}_i \mathbf{T}_i^T = \text{var}(\boldsymbol{\epsilon}_i) = \mathbf{D}_i.$$

The GARPs are denoted by ϕ and σ_{ij}^2 represents the IVs. Time-and/or subject-specific covariate vectors can be used to model the GARPs and IVs by setting

$$\phi_{i,jt} = \mathbf{k}_{i,jt}^T \delta, \quad log\left(\sigma_{ij}^2\right) = \mathbf{h}_{i,j}^T \boldsymbol{\lambda}, \tag{2.7}$$

where δ and $\boldsymbol{\lambda}$ are $a \times 1$ and $b \times 1$ vectors of unknown dependence parameters, respectively. The design vectors $\mathbf{k}_{i,jt}$ and $\mathbf{h}_{i,j}$ are covariates to model the GARP/IV parameters as a function of the subject specific covariates (Pourahmadi, 2000; Pourahmadi and Daniels, 2002; Deniels and Zhao, 2003; Lee et al., 2012). The parametrization of GARPs or IVs has various advantages. Firstly, we can model the random effects covariance matrix in terms of covariates because of unconstrained character of GARPs and IVs. Secondly, as in (2.7) there is a linear combination of covariates, the parameters have a reasonable interpretation and easy to model (Deniels and Zhao, 2003; Lee et al., 2012) and the positive definiteness of $\boldsymbol{\Sigma}_i$ is guaranteed because of the positive σ_{ij}^2. Also we can have specific $\boldsymbol{\Sigma}_i$ for each subject i. The proposed approach also covers the AR(1) as a special case.

2.5 The General Inference Method

To derive the likelihood function for the GLMM with covariates measurement error, let us define the parameters as $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \gamma^T, \boldsymbol{\tau}^T, \delta^T, \boldsymbol{\lambda}^T)^T$. The complete data likelihood function of $\mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i$ for subject i can be written as:

$$\begin{aligned} L_i(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i) &= f(\mathbf{y}_i, \mathbf{x}_i, \mathbf{w}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\theta}) \\ &= f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}) f(\mathbf{x}_i, \mathbf{w}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\theta}) \\ &= f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}) f(\mathbf{x}_i | \mathbf{w}_i, \mathbf{z}_i; \gamma, \boldsymbol{\tau}) f(\mathbf{u}_i; \delta, \boldsymbol{\lambda}), \end{aligned}$$

where $f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta})$ belongs to the exponential family (2.1) given the random effects. Also we can assume,

$$f(\mathbf{x}_i | \mathbf{w}_i, \mathbf{z}_i; \gamma, \boldsymbol{\tau}) \sim N(\boldsymbol{\mu}_f, \sigma_f^2),$$

where the details of $\boldsymbol{\mu}_f, \sigma_f^2$ are given in Section (2.6) and independent of $\mathbf{u}_i$. Also assuming that the random effects, $\mathbf{u}_i$, are independent and identically distributed, and independent of the explanatory variables:

$$\mathbf{u}_i \sim f(\mathbf{u}_i | \mathbf{x}_i, \mathbf{z}_i; \delta, \boldsymbol{\lambda}) = f(\mathbf{u}_i; \delta, \boldsymbol{\lambda})$$

and the $f(\mathbf{u}_i; \delta, \boldsymbol{\lambda})$ has a multivariate normal density with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_i$ and this can be simplified based on the proposed structure of the random effect covariance matrix (Lee et al., 2012) as follows:

$$f(\mathbf{u}_i; \delta, \boldsymbol{\lambda}) = (2\pi)^{-m/2} \left[\prod_{j=1}^{m} \left(\sigma_{ij}^2 \right)^{-1/2} \right] \exp\left(-\frac{1}{2} \sum_{j=1}^{m} \frac{\epsilon_{ij}^2}{\sigma_{ij}^2} \right) \text{ with } \quad \epsilon_{i1} = u_{i1}.$$

Then we can write the complete data log-likelihood as:

$$
\begin{aligned}
l_c(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \log L_i(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i) \\
&= \sum_{i=1}^{n} \{ \log f(\mathbf{y}_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}) + \log f(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \gamma, \boldsymbol{\tau}) + \log f(\mathbf{u}_i; \delta, \boldsymbol{\lambda}) \} \,. \quad (2.8)
\end{aligned}
$$

The EM algorithm is employed to evaluate the above log-likelihood function because of its intractable form (Meng and Van Dyk, 1998). The E-step can be written as (at iteration l):

$$
\begin{aligned}
Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(l)}) &= \mathrm{E}\left[l_c(\boldsymbol{\theta}) | \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i; \boldsymbol{\theta}^{(l)} \right] \\
&= \sum_{i=1}^{n} \int\int \Big\{ \log f(\mathbf{y}_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}) + \log f(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \gamma, \boldsymbol{\tau}) \\
&+ \log f(\mathbf{u}_i; \delta, \lambda) \Big\} f\left(\mathbf{x}_i, \mathbf{u}_i | \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)} \right) \, d\mathbf{x}_i d\mathbf{u}_i \\
&= I_1 + I_2 + I_3, \quad (2.9)
\end{aligned}
$$

where $f\left(\mathbf{x}_i, \mathbf{u}_i | \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)} \right)$ is the conditional density of missing components $(\mathbf{x}_i, \mathbf{u}_i)$ given the observed data $(\mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i)$. As in the above $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l)})$ function, the multiple integrals are not in closed form for computation, it is not often possible to evaluate this expectation directly. Hence Monte Carlo EM (MCEM) algorithm can be employed (McCulloh and Searl, 2001, Sect. 10.3). To do this, we need to generate a sample from $f(\mathbf{x}_i, \mathbf{u}_i | \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)})$ for each i. Gibbs sampling or Metroplois-Hasting algorithm can be used to accomplish this (McCulloch, 1997; Levine and Casella, 2011; Fort and Moulines, 2003; Caffo et al., 2005). Essentially, we can iteratively sample from $f(\mathbf{x}_i|\mathbf{y}_i, \mathbf{u}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}), f(\mathbf{u}_i|\mathbf{y}_i, \mathbf{x}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)})$. These conditional distributions have the form respectively as follows:

$$f(\mathbf{x}_i|\mathbf{y}_i, \mathbf{u}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}) \propto f(\mathbf{y}_i|\mathbf{x}_i, \mathbf{u}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}) f(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)})$$

$$f(\mathbf{u}_i|\mathbf{y}_i, \mathbf{x}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}) \propto f(\mathbf{y}_i|\mathbf{x}_i, \mathbf{u}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}) f(\mathbf{u}_i; \boldsymbol{\theta}^{(l)})$$

Suppose that, we take a pseudo-random sample of size M, $\Big\{ (\mathbf{x}_i^{(1)}, \mathbf{u}_i^{(1)}), (\mathbf{x}_i^{(2)}, \mathbf{u}_i^{(2)}), \ldots, (\mathbf{x}_i^{(M)}, \mathbf{u}_i^{(M)}) \Big\}$ for individual i, from the joint distribution $f\big(\mathbf{x}_i, \mathbf{u}_i|\mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}\big)$ via the Metropolis-Hasting algorithm. Then the E-step at the $(l+1)$th EM iteration can be written as:

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l)}) &\approx \sum_{i=1}^{n} \Big\{ \frac{1}{M} \sum_{k=1}^{M} l_c \left(\boldsymbol{\theta}^{(l)}; \mathbf{y}_i, \mathbf{x}_i^{(k)}, \mathbf{u}_i^{(k)} \right) \Big\} \\ &= \sum_{i=1}^{n} \sum_{k=1}^{M} \frac{1}{M} \log f \left(\mathbf{y}_i, |\mathbf{x}_i^{(k)}, \mathbf{z}_i, \mathbf{u}_i^{(k)}; \boldsymbol{\beta} \right) \\ &+ \sum_{i=1}^{n} \sum_{k=1}^{M} \frac{1}{M} \log f \left(\mathbf{x}_i^{(k)} | \mathbf{w}_i, \mathbf{z}_i; \gamma, \boldsymbol{\tau} \right) + \sum_{i=1}^{n} \sum_{k=1}^{M} \frac{1}{M} \log f \left(\mathbf{u}_i^{(k)}; \delta, \boldsymbol{\lambda} \right) \\ &= Q^{(1)}(\boldsymbol{\beta}; \boldsymbol{\theta}^{(l)}) + Q^{(2)}(\gamma, \boldsymbol{\tau}; \boldsymbol{\theta}^{(l)}) + Q^{(3)}(\delta, \boldsymbol{\lambda}; \boldsymbol{\theta}^{(l)}). \end{aligned} \tag{2.10}$$

In the M-step of MCEM algorithm, an optimization procedure can be employed to maximize $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(l)})$ with respect to $\boldsymbol{\theta}$ to produce an updated estimate $\boldsymbol{\theta}^{(l+1)}$. These E and M steps will continue until convergence and then the current values of $\boldsymbol{\theta}$ will be declared as MLEs of $\boldsymbol{\theta}$.
By the following steps we can describe the algorithm:

1. Choose the starting values,$\boldsymbol{\theta}^{(0)} = \big(\boldsymbol{\beta}^{(0)}, \gamma^{(0)}, \boldsymbol{\tau}^{(0)}, \delta^{(0)}, \boldsymbol{\lambda}^{(0)}\big)$. Set $l = 1$.

2. Generate M vectors $\left(\mathbf{x}_i^{(k)}, \mathbf{u}_i^{(k)}\right), \quad \left(k = 1, 2, \ldots, M\right)$, from the conditional distribution $\left(\mathbf{x}_i, \mathbf{u}_i\right)$ given $\left(\mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i\right)$ with unknown parameters *vs* in the distribution replaced by the current estimate $\boldsymbol{\theta}^{(l-1)}$ using the Metroplis-Hasting algorithm.

   a. Calculate $\boldsymbol{\beta}^{(l)}$ as the value that maximizes

   $$\sum_{i=1}^{n}\sum_{k=1}^{M}\frac{1}{M}\log f\left(\mathbf{y}_i|\mathbf{x}_i^{(k)}, \mathbf{z}_i, \mathbf{u}_i^{(k)}; \boldsymbol{\beta}\right)$$

   b. Calculate $\boldsymbol{\gamma}^{(l)}$ and $\boldsymbol{\tau}^{(l)}$ as those values that maximize

   $$\sum_{i=1}^{n}\sum_{k=1}^{M}\frac{1}{M}\log f\left(\mathbf{x}_i^{(k)}|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\gamma}, \boldsymbol{\tau}\right)$$

   c. Calculate $\boldsymbol{\delta}^{(l)}$ and $\boldsymbol{\lambda}^{(l)}$ as those values that maximize

   $$\sum_{i=1}^{n}\sum_{k=1}^{M}\frac{1}{M}\log f\left(\mathbf{u}_i^{(k)}; \boldsymbol{\delta}, \boldsymbol{\lambda}\right)$$

   d. Set $l = l + 1$

3. If convergence is achieved, the current values can be declared as the ML estimates, otherwise return to step (2).

Standard errors of model parameters estimate

The standard errors of the MLEs cannot be automatically obtained from EM algorithm. To obtain the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, we can use the formula given by McLachlan and Krishnan (2007) which is

$$\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\theta}}) \approx \left[\sum_{i=1}^{n}\sum_{k=1}^{M}\frac{1}{M}\mathbf{S}_{ik}(\hat{\boldsymbol{\theta}})\mathbf{S}_{ik}^{T}(\hat{\boldsymbol{\theta}})\right]^{-1}, \tag{2.11}$$

where, $\mathbf{S}_{ik}(\hat{\boldsymbol{\theta}}) = \left.\frac{\partial l_c\left(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{x}_i^{(k)},\mathbf{u}_i^{(k)}\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. Then by taking the square root of the diagonal element of $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\theta}})$, the standard errors of the MLEs can be obtained. One can also use an optimization procedure to get the standard error. For example, **optim** function in *R-program* provides the hessian matrix which can be used to calculate the standard errors (R core team, 2016).

2.6 An Illustration

As an illustration of MCEM algorithm method, let us consider the longitudinal binary data and assume the following logistic mixed model

$$\operatorname{logit}\{P(Y_{ij} = 1 \mid X_{ij}, Z_{ij}, u_{ij})\} = \beta_0 + \beta_x X_{ij} + \beta_z Z_{ij} + u_{ij}.$$

And the density function for y_{ij} is given by

$$f(y_{ij}|\mathbf{x}_i,\mathbf{z}_i,\mathbf{u}_i;\boldsymbol{\beta}) = P_{ij}(u_{ij})^{y_{ij}}(1 - P_{ij}(u_{ij}))^{1-y_{ij}},$$

$$i = 1,2,\ldots,n \text{ and } j = 1,2,\ldots,m$$

with

$$P_{ij}(u_{ij}) = P(Y_{ij} = 1|X_{ij}, Z_{ij}, u_{ij})$$

$$= \frac{\exp(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij})}{1+\exp(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij})} \quad \text{where} \quad \mathbf{x}_{ij} = (X_{ij}, Z_{ij})^T \text{ and } \boldsymbol{\beta}^T = (\beta_0, \beta_x, \beta_z)$$

and $1 - P_{ij}(u_{ij}) = \dfrac{1}{1+\exp(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij})}$.

Then we can write the following density function:

$$\begin{aligned} f(\mathbf{y}_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}) &= \frac{\exp\{\sum_{j=1}^m y_{ij}\big(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij}\big)\}}{\prod_{j=1}^m\{1+\exp\big(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij}\big)\}} \\ &= \exp\Bigg[\sum_{j=1}^m y_{ij}\big(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij}\big) - \Bigg(\sum_{j=1}^m \log\big\{1+\exp\big(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij}\big)\big\}\Bigg)\Bigg]. \end{aligned} \tag{2.12}$$

Also, $f(\mathbf{u}_i)$ has a multivariate normal density with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_i$ which can be written as

$$f(\mathbf{u}_i; \delta, \boldsymbol{\lambda}) = (2\pi)^{-m/2}\Bigg[\prod_{j=1}^m \left(\sigma_{ij}^2\right)^{-1/2}\Bigg]\exp\left(-\frac{1}{2}\sum_{j=1}^m \frac{\epsilon_{ij}^2}{\sigma_{ij}^2}\right) \text{ with } \quad \epsilon_{i1} = u_{i1}. \tag{2.13}$$

Moreover, let us consider classical additive structural measurement error model as

$$W_{ij} = X_{ij} + e_{ij},$$

where $e_{ij} \sim N(0, \tau)$ and X_{ij} be the vector of covariates subject to error and has normal distribution $N(\mu_x, \sigma_x^2)$. We can write $W_{ij}|X_{ij} \sim N(X_{ij}, \tau)$. Then the conditional distribution of $X_{ij}|W_{ij}$ can be written as follows:

$$X_{ij}\Big|W_{ij}; \mu_x, \sigma_x^2, \tau \sim N(\mu_f, \sigma_f^2),$$

$$\text{where,} \quad \mu_f = \frac{\mu_x\tau + W_{ij}\sigma_x^2}{\tau + \sigma_x^2} \text{ and } \sigma_f^2 = \frac{\sigma_x^2\tau}{\tau+\sigma_x^2}. \tag{2.14}$$

Therefore, we can write the complete data log-likelihood following (2.8) as:

$$
\begin{aligned}
l_c(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \log L_i(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i) \\
&= \sum_{i=1}^{n} \left[\sum_{j=1}^{m} y_{ij}(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_{ij}) - \left(\sum_{j=1}^{m} \log \left\{ 1 + \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_{ij}) \right\} \right) \right] \\
&+ \sum_{i=1}^{n} \left[-\frac{m}{2} \log \left(2\pi \frac{\sigma_x^2 \tau}{\tau + \sigma_x^2} \right) - \frac{1}{2} \sum_{j=1}^{m} \frac{\left(X_{ij} - \dfrac{\mu_x \tau + W_{ij}\sigma_x^2}{\tau + \sigma_x^2} \right)^2}{\dfrac{\sigma_x^2 \tau}{\tau + \sigma_x^2}} \right] \\
&+ \sum_{i=1}^{n} \left[-\frac{m}{2} \log (2\pi) - \sum_{j=1}^{m} \frac{1}{2} \log \sigma_{ij}^2 - \frac{1}{2} \sum_{j=1}^{m} \frac{\epsilon_{ij}^2}{\sigma_{ij}^2} \right], \qquad (2.15)
\end{aligned}
$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^T, \gamma^T = (\mu_x, \sigma_x^2)^T, \boldsymbol{\tau}^T, \delta^T, \boldsymbol{\lambda}^T \right)^T$ is the associated parameters to develop EM algorithm. We can write the observed data likelihood as follows

$$
\begin{aligned}
L(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i) = \int \int & f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}) f(\mathbf{x}_i | \mathbf{w}_i; \gamma, \boldsymbol{\tau}) f(\mathbf{u}_i; \delta, \boldsymbol{\lambda}) \\
& f\left(\mathbf{x}_i, \mathbf{u}_i | \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}\right) \, d\mathbf{x}_i d\mathbf{u}_i \qquad (2.16)
\end{aligned}
$$

This likelihood function is not in closed form to express and we rely on MCEM algorithm to evaluate this. Here we consider the parameters $\boldsymbol{\beta}, \delta, \boldsymbol{\lambda}, \gamma, \boldsymbol{\tau}$ for estimation. The ML estimators of $\boldsymbol{\beta}, \delta, \boldsymbol{\lambda}, \gamma, \boldsymbol{\tau}$ can be obtained by solving the following estimating equations:

$$
\sum_{i=1}^{n} \mathrm{E} \left\{ \frac{\partial \log f(\mathbf{y}_i \mid \mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \middle| \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i \right\} = 0,
$$

$$\sum_{i=1}^{n} \mathrm{E}\left\{ \frac{\partial \log f\big(\mathbf{x}_i; \gamma, \boldsymbol{\tau}\big)}{\partial \gamma} \middle| \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i \right\} = 0,$$

$$\sum_{i=1}^{n} \mathrm{E}\left\{ \frac{\partial \log f\big(\mathbf{x}_i; \gamma, \boldsymbol{\tau}\big)}{\partial \boldsymbol{\tau}} \middle| \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i \right\} = 0,$$

$$\sum_{i=1}^{n} \mathrm{E}\left\{ \frac{\partial \log f\big(\mathbf{u}_i; \delta, \boldsymbol{\lambda}\big)}{\partial \delta} \middle| \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i \right\} = 0,$$

$$\sum_{i=1}^{n} \mathrm{E}\left\{ \frac{\partial \log f\big(\mathbf{u}_i; \delta, \boldsymbol{\lambda}\big)}{\partial \boldsymbol{\lambda}} \middle| \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i \right\} = 0,$$

where the conditional expectations are with respect to the conditional distribution of missing components $\big(\mathbf{x}_i, \mathbf{u}_i\big)$ given the observed data $\big(\mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i\big)$. Here for the parameters $\boldsymbol{\beta}, \delta, \boldsymbol{\lambda}, \gamma, \boldsymbol{\tau}$ score functions for individual i can be expressed as:

$$\frac{\partial \log L(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i)}{\partial \boldsymbol{\beta}} = \frac{1}{L(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i)} \int\int \frac{\partial f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$$

$$f(\mathbf{x}_i | \mathbf{w}_i, \mathbf{z}_i; \gamma, \boldsymbol{\tau}) f(\mathbf{u}_i; \delta, \boldsymbol{\lambda}) d\mathbf{x}_i d\mathbf{u}_i \qquad (2.17)$$

where,

$$\frac{\partial f(\mathbf{y}_i|\mathbf{x}_i,\mathbf{z}_i,\mathbf{u}_i;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \exp\Bigg[\sum_{j=1}^{m} y_{ij}\big(\mathbf{x}_{ij}^T\boldsymbol{\beta}+u_{ij}\big) - \Bigg(\sum_{j=1}^{m}\log\big\{1+\exp\big(\mathbf{x}_{ij}^T\boldsymbol{\beta}+u_{ij}\big)\big\}\Bigg)\Bigg]$$

$$\Bigg[\sum_{j=1}^{m} y_{ij}\mathbf{x}_{ij} - \frac{1}{1+\exp\big(\mathbf{x}_{ij}^T\boldsymbol{\beta}+u_{ij}\big)}\exp\big(\mathbf{x}_{ij}^T\boldsymbol{\beta}+u_{ij}\big)\,\mathbf{x}_{ij}\Bigg]$$

$$= \exp\Bigg[\sum_{j=1}^{m} y_{ij}\big(\mathbf{x}_{ij}^T\boldsymbol{\beta}+u_{ij}\big) - \Bigg(\sum_{j=1}^{m}\log\big\{1+\exp\big(\mathbf{x}_{ij}^T\boldsymbol{\beta}+u_{ij}\big)\big\}\Bigg)\Bigg]$$

$$\Bigg[\sum_{j=1}^{m}\Bigg(y_{ij} - P_{ij}(u_{ij})\Bigg)\mathbf{x}_{ij}\Bigg]$$

Also,

$$\frac{\partial\log L(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{x}_i,\mathbf{u}_i)}{\partial\boldsymbol{\tau}} = \frac{1}{L(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{x}_i,\mathbf{u}_i)}\int\int f(\mathbf{y}_i|\mathbf{x}_i,\mathbf{z}_i,\mathbf{u}_i;\boldsymbol{\beta})$$

$$\frac{\partial f(\mathbf{x}_i|\mathbf{w}_i;\gamma,\boldsymbol{\tau})}{\partial\boldsymbol{\tau}} f(\mathbf{u}_i;\delta,\boldsymbol{\lambda})d\mathbf{x}_i d\mathbf{u}_i \tag{2.18}$$

where,

$$\frac{\partial f(\mathbf{x}_i|\mathbf{w}_i;\gamma,\boldsymbol{\tau})}{\partial\boldsymbol{\tau}} = \frac{\partial}{\partial\boldsymbol{\tau}}\Bigg[\exp\big\{\log f(\mathbf{x}_i|\mathbf{w}_i;\gamma,\boldsymbol{\tau})\big\}\Bigg]$$

$$= f(\mathbf{x}_i|\mathbf{w}_i;\gamma,\boldsymbol{\tau})\frac{\partial}{\partial\boldsymbol{\tau}}\big\{\log f(\mathbf{x}_i|\mathbf{w}_i;\gamma,\boldsymbol{\tau})\big\},$$

and,

$$\frac{\partial\log L(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{x}_i,\mathbf{u}_i)}{\partial\gamma} = \frac{1}{L(\theta;\mathbf{y}_i,\mathbf{x}_i,\mathbf{u}_i)}\int\int f(\mathbf{y}_i|\mathbf{x}_i,\mathbf{z}_i,\mathbf{u}_i;\boldsymbol{\beta})$$

$$\frac{\partial f(\mathbf{x}_i|\mathbf{w}_i;\gamma,\boldsymbol{\tau})}{\partial\gamma} f(\mathbf{u}_i;\delta,\boldsymbol{\lambda})d\mathbf{x}_i d\mathbf{u}_i \tag{2.19}$$

where,

$$\frac{\partial f(\mathbf{x}_i|\mathbf{w}_i;\gamma,\boldsymbol{\tau})}{\partial\gamma} = \frac{\partial}{\partial\gamma}\Big[\exp\big\{\log f(\mathbf{x}_i|\mathbf{w}_i;\gamma,\boldsymbol{\tau})\big\}\Big]$$

$$= f(\mathbf{x}_i|\mathbf{w}_i;\gamma,\boldsymbol{\tau})\frac{\partial}{\partial\gamma}\big\{\log f(\mathbf{x}_i|\mathbf{w}_i;\gamma,\boldsymbol{\tau})\big\}.$$

Furthermore,

$$\frac{\partial\log L(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{x}_i,\mathbf{u}_i)}{\partial\delta} = \frac{1}{L(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{x}_i,\mathbf{u}_i)}\int\int f(\mathbf{y}_i|\mathbf{x}_i,\mathbf{z}_i,\mathbf{u}_i;\boldsymbol{\beta})$$

$$f(\mathbf{x}_i|\mathbf{w}_i,\mathbf{z}_i;\gamma,\boldsymbol{\tau})\frac{\partial f(\mathbf{u}_i;\delta,\boldsymbol{\lambda})}{\partial\delta}d\mathbf{x}_i d\mathbf{u}_i \qquad (2.20)$$

where,

$$f(\mathbf{u}_i;\delta,\boldsymbol{\lambda}) = (2\pi)^{-m/2}\left[\prod_{j=1}^{m}\left(\sigma_{ij}^2\right)^{-1/2}\right]\exp\left(-\frac{1}{2}\sum_{j=1}^{m}\frac{\epsilon_{ij}^2}{\sigma_{ij}^2}\right) \text{ with } \quad \epsilon_{i1} = u_{i1},$$

$$\phi_{i,jt} = \mathbf{k}_{i,jt}^T\delta, \quad log\left(\sigma_{ij}^2\right) = \mathbf{h}_{i,j}^T\boldsymbol{\lambda}$$

$$u_{ij} = \sum_{t=1}^{j-1}\phi_{i,jt}u_{it} + \epsilon_{ij}, \quad \text{for } j = 2,3,\ldots,m$$

Then we can write

$$\frac{\partial f(\mathbf{u}_i;\delta,\boldsymbol{\lambda})}{\partial\delta} = \frac{\partial}{\partial\delta}\Big[\exp\big\{\log f(\mathbf{u}_i;\delta,\boldsymbol{\lambda})\big\}\Big]$$

$$= f(\mathbf{u}_i;\delta,\boldsymbol{\lambda})\frac{\partial}{\partial\delta}\big\{\log f(\mathbf{u}_i;\delta,\boldsymbol{\lambda})\big\}$$

$$= -f(\mathbf{u}_i;\delta,\boldsymbol{\lambda})\sum_{j=1}^{m}\frac{\epsilon_{ij}}{\sigma_{ij}^2}\frac{\partial\epsilon_{ij}}{\partial\delta}$$

$$\text{with} \quad \frac{\partial\epsilon_{i1}}{\partial\delta} = 0 \quad \text{as } \epsilon_{i1} = u_{i1} \quad \text{and } \frac{\partial\epsilon_{ij}}{\partial\delta} = -\sum_{t=1}^{j-1}u_{it}k_{i,jt}.$$

Similarly, we can write

$$\frac{\partial \log L(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i)}{\partial \boldsymbol{\lambda}} = \frac{1}{L(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i)} \int \int f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}) \\ f(\mathbf{x}_i | \mathbf{w}_i, \mathbf{z}_i; \gamma, \boldsymbol{\tau}) \frac{\partial f(\mathbf{u}_i; \delta, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} d\mathbf{x}_i d\mathbf{u}_i, \tag{2.21}$$

where,

$$\frac{\partial f(\mathbf{u}_i; \delta, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \frac{\partial}{\partial \boldsymbol{\lambda}} \Big[\exp\{\log f(\mathbf{u}_i; \delta, \boldsymbol{\lambda})\} \Big].$$

And after some algebra, we can write

$$\frac{\partial f(\mathbf{u}_i; \delta, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = f(\mathbf{u}_i; \delta, \boldsymbol{\lambda}) \sum_{j=1}^{m} \left(\frac{\epsilon_{ij}^2}{\sigma_{ij}^2} - 1 \right) h_{i,j}.$$

The MCEM can be then used to get the approximation of the integrals in (2.17)-(2.21).

We follow the equation (2.9) to get the E-step of the MCEM. To evaluate E-steps we generate sample from these conditional distributions using Metropolis-Hasting (M-H) algorithm. We illustrate the algorithm for the following case:

$$f(\mathbf{x}_i | \mathbf{y}_i, \mathbf{u}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}) \propto f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{u}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}) f(\mathbf{x}_i | \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}).$$

That is, when $\mathbf{x}_i$ is missing we can write the algorithm as follows to generate data for each iteration. Here,

$$f(X) = f(\mathbf{x}_i|\mathbf{y}_i, \mathbf{u}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)})$$

$$\propto f(\mathbf{y}_i|\mathbf{x}_i, \mathbf{u}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}) f(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)})$$

$$\propto \frac{\exp\{\sum_{j=1}^{m} y_{ij}(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij})\}}{\prod_{j=1}^{m}\{1+\exp(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij})\}} \times \exp\left\{-\sum_{j=1}^{m}\frac{(X_{ij}-\mu_f)^2}{2\sigma_f^2}\right\}(\text{following } 2.14)$$

$$g(X) = \exp\left\{-\sum_{j=1}^{m}\frac{(X_{ij}-\mu_f)^2}{2\sigma_f^2}\right\}.$$

To do M-H, we first generate initial sample $\mathbf{x}_0 = (x_{01}, \ldots, x_{0m})$ from $N(\mu_f, \sigma_f^2)$. Then we generate a sample x from $N(\mu_f, \sigma_f^2)$ and v from uniform $(0, 1)$. After that we compute the acceptance ratio as:

$$A_k(x_{i-1}, x) = \min\left\{\frac{f(x_{i-1})g(x)}{f(x)g(x_{i-1})}, 1\right\}.$$

Now, if

$$v < A_k(x_{i-1}, x), \quad \text{then } x_i = x$$

$$v > A_k(x_{i-1}, x), \quad \text{then } x_i = x_{i-1}.$$

Similarly, we can generate samples from the other conditional distributions as well. And after generating samples we evaluate the E-steps and M-steps following the algorithm given in the general inference method section.

Chapter 3

Simulation

3.1 Simulation Set-up

We conduct a simulation study to evaluate the performance of our Proposed approach. In particular, we make a comparison of the efficiency of our Proposed method to the method where covariates measurement error are ignored (Naive 1). Moreover, we also compare the efficiency of the Proposed method to the method for GLMMs with covariates measurement error where constant random effects covariance matrix is considered across the subjects (Naive 2). We use the following model to simulate data:

$$\text{logit}\{P_{ij}\} = \beta_0 + \beta_x X_{ij} + \beta_z Z_{ij} + \beta_{z^*} Z_i^* + u_{ij}, \tag{3.1}$$

$$i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, m,$$

where $P_{ij} = P(Y_{ij} = 1 | X_{ij}, \mathbf{Z}_{ij}^+)$, with $\mathbf{Z}_{ij}^+ = (Z_{ij}, Z_i^*)$ and Z_i^* equals to 0 or 1 with an equal sample size per group. The error-free covariate Z_{ij} is generated as the random variables $v_i + \zeta_{ij}$, where v_i's and ζ_{ij}'s are independently identically distributed following $N(0, 0.5^2)$, and they are independent. We consider classical additive model for the structural measurement error for which we generate surrogate variable W_{ij} as

$$W_{ij} = X_{ij} + e_{ij},$$

where e_{ij}'s are independently identically distributed following $N(0, \sigma^2)$, where σ^2 indicates the measurement error variation in covariate X_{ij}. The true covariate X_{ij} is generated from model $X_{ij} = \mu_x + a_i + \xi_{ij}$, where $\mu_x = 1$, a_i's and ξ_{ij}'s are independently identically distributed following $N(0, 1)$.

We consider random effects $\mathbf{u}_i = (u_{i1}, \ldots, u_{im}) \sim N(\mathbf{0}, \mathbf{\Sigma}_i)$. To simplify variance components of $\mathbf{\Sigma}_i$, the parameters of the random effects covariance matrix are defined as:

$$\phi_{i,jt} = \delta_0 I(|j-t| = 1) + \delta_1 I(|j-t| = 1) Z_i^* \quad \text{and}$$

$$log(\sigma_{ij}^2) = \lambda_0 + \lambda_1 Z_i^* \tag{3.2}$$

Here, we consider $n = 100$ and two cases with $m = 5$, $m = 10$. The initial values of $\boldsymbol{\beta} = (\beta_0, \beta_x, \beta_z, \beta_{z^*}) = (1, 2, 2, 2)$. We set three different values $(0, 0.4, 0.8)$ for σ to see the impact of varying degree of measurement error on estimation. The initial values of the parameters of the random effects covariance matrix are $\boldsymbol{\delta} = (\delta_0, \delta_1) = (0.5, 0.3)$, and $\boldsymbol{\lambda} = (\lambda_0, \lambda_1) = (0.1, 0.2)$.

We generate $B = 200$ data sets. Then we fit three models: the first model (Proposed Model) is the proposed method; the second model (Naive 2) is the model which considers homogeneous covariance matrix, that is, $u_{ij} = u_{i0} \sim N(0, \sigma_u^2)$, where σ_u^2 is the average of the σ_{ij}^2 defined in the equation (3.2)and covariate X_{ij} with measurement error and the third model (Naive 1) is the model which ignores covariate measurement error and assuming homogeneous covariance matrix.

3.2 Simulation Results

In the following, we report the simulation results of empirical biases (Bias), root mean square errors (RMSE), and coverage rates (CR) for the 95% confidence intervals (CI) of the model parameters estimate where e.g. for the β_0 (fixed intercept) we have:

$$\text{Bias}_{\beta_0} = \frac{1}{B}\sum_{b=1}^{B} \hat{\beta}_0^{(b)} - \beta_0,$$

where $\hat{\beta}_0^{(b)}$ is the estimated value in each simulation run b and β_0 is the true value of this parameter. Also,

$$\text{RMSE}_{\beta_0} = \sqrt{\text{Bias}_{\beta_0}^2 + \text{Var}_{\beta_0}^2},$$

$$95\%\ \text{CI}_{\beta_0}^{(b)} = \hat{\beta}_0^{(b)} \pm 1.96\sqrt{\text{Var}^{(b)}(\hat{\beta}_0^{(b)})},$$

where the Var_{β_0} are the average of model-based variances $\text{Var}(\hat{\beta_0})$ over B simulation runs, and CR is the proportion of times (out of $B = 200$) that the true parameter falls in the corresponding 95% CI.

The following Tables represent the results of the three methods (Naive 1, Naive 2, and Proposed) for the two cases $m = 5$ and $m = 10$ under different measurement error variations (no error [$\sigma = 0.0$], moderate error [$\sigma = 0.4$] and severe error [$\sigma = 0.8$]).

	Bias	RMSE	CR
β_0	-0.0012	0.2216	0.96
β_x	0.0461	0.2253	0.97
β_z	0.0326	0.3351	0.96
β_{z^*}	0.0611	0.3979	0.97
σ_u^2	-0.0208	0.2338	0.61

Table 3.1a: Simulation results for Naive 1 with m=5 and no error ($\sigma = 0.0$)

	Bias	RMSE	CR
β_0	0.0049	0.2010	0.97
β_x	0.0250	0.1810	0.96
β_z	-0.0137	0.2571	0.97
β_{z^*}	0.0042	0.2932	0.98
σ_u^2	-0.0533	0.1996	0.31

Table 3.1b: Simulation results for Naive 2 with m=5 and no error ($\sigma = 0.0$)

	Bias	RMSE	CR
β_0	0.0063	0.0647	1.00
β_x	-0.0046	0.0449	1.00
β_z	0.0080	0.0798	1.00
β_{z^*}	-0.0085	0.1002	1.00
δ_0	0.5831	1.2232	0.09
δ_1	-0.1502	0.2056	0.09
λ_0	0.0280	0.1344	0.12
λ_1	-0.0069	0.1893	0.08

Table 3.1c: Simulation results for Proposed Method with m=5 and no error ($\sigma = 0.0$)

Tables 3.1a, 3.1b and 3.1c show the results of the fixed effect and random effects for the three methods with the absence of measurement errors when number of follow-ups is 5 $(m = 5)$. It is evident form the results that the proposed method works well under this situation in terms of bias, RMSE and coverage rate for 95% CI. In Naive 1 and Naive 2, we can see considerable bias in fixed effect estimates $(\beta_x = 0.0461, \beta_z = 0.0326, \beta_{z^*} = 0.0611$ for Naive 1), $(\beta_x = 0.0250, \beta_z = -0.0137, \beta_{z^*} = 0.0042$ for Naive 2) whereas in Proposed method, except β_0, the estimates show fairly small biases $(\beta_x = -0.0046, \beta_z = 0.0080, \beta_{z^*} = -0.0085)$. The proposed method shows the good coverage rate for 95% CI. The RMSEs are also smaller for proposed method compared to the other two methods. We can see considerable bias in fixed effect parameters for misspecification of the distribution of random effects. In random effect covariance matrix parameters, the GARPs (δ_0, δ_1) indicate relatively large bias but the IVs (λ_0, λ_1) parameters have small bias.

Tables 3.2a, 3.2b and 3.2c present the results for moderate measurement error variation. It is obvious here that ignoring the measurement error in data results considerable biases. The biases tend to increase with the increase of magnitude of measurement error.

	Bias	RMSE	CR
β_0	0.1264	0.2464	0.94
β_x	-0.2216	0.2942	0.73
β_z	0.0032	0.3236	0.95
β_{z^*}	-0.0802	0.3976	0.95
σ_u^2	-0.0208	0.2338	0.61

Table 3.2a: Simulation results for Naive 1 with m=5 and moderate error $(\sigma = 0.4)$

	Bias	RMSE	CR
β_0	-0.0269	0.1802	0.98
β_x	0.0547	0.1901	0.99
β_z	0.0457	0.3001	0.97
β_{z^*}	0.0688	0.3378	0.98
σ_u^2	-0.0275	0.1753	0.25

Table 3.2b: Simulation results for Naive 2 with m=5 and moderate error $(\sigma = 0.4)$

	Bias	RMSE	CR
β_0	0.0006	0.0484	1.00
β_x	0.0051	0.0416	1.00
β_z	0.0044	0.0660	1.00
β_{z^*}	0.0045	0.0918	1.00
δ_0	0.4493	1.1386	0.13
δ_1	-0.1555	0.2045	0.17
λ_0	0.0197	0.1402	0.17
λ_1	-0.0170	0.1671	0.13

Table 3.2c: Simulation results for Proposed Method with m=5 and moderate error $(\sigma = 0.4)$

The results indicate the higher biases for Naive 1 and Naive 2, while the Proposed method shows considerably smaller biases in fixed effects parameters. Moreover, the proposed method indicates smaller RMSEs as well as good coverage rate for 95% CI in the estimates of fixed effects compared to the other two methods. For example, the estimates of the coefficient of the measurement error variable (β_x) has bias -0.2216 with 73% coverage rate for Naive 1, 0.0547 with 99% coverage rate for Naive 2 while in the proposed method the bias reduces to 0.0051 with 100% coverage rate.

	Bias	RMSE	CR
β_0	0.3271	0.3806	0.68
β_x	-0.7088	0.7228	0.01
β_z	-0.1506	0.3182	0.90
β_{z^*}	-0.3409	0.4899	0.84
σ_u^2	-0.0208	0.2338	0.61

Table 3.3a: Simulation results for Naive 1 with m=5 and severe error ($\sigma = 0.8$)

	Bias	RMSE	CR
β_0	0.0190	0.1816	0.97
β_x	0.0194	0.1550	0.98
β_z	0.0433	0.2553	0.96
β_{z^*}	-0.0062	0.2791	0.99
σ_u^2	-0.0353	0.2078	0.28

Table 3.3b: Simulation results for Naive 2 with m=5 and severe error ($\sigma = 0.8$)

	Bias	RMSE	CR
β_0	0.0087	0.0474	1.00
β_x	-0.0042	0.0472	1.00
β_z	-0.0077	0.0664	1.00
β_{z^*}	-0.0109	0.0815	1.00
δ_0	0.5940	1.2768	0.20
δ_1	-0.1533	0.2055	0.19
λ_0	0.0153	0.1163	0.20
λ_1	-0.0308	0.1591	0.21

Table 3.3c: Simulation results for Proposed Method with m=5 and severe error ($\sigma = 0.8$)

In case of severe measurement error variation and m=5 (number of follow-ups), the performance of the three methods is shown in Tables 3.3a, 3.3b and 3.3c. It is clear from the results that the performance of Naive 1 is noticeably affected with the increase of magnitude of measurement error. Based on the results, it can be seen that there is considerably large finite-sample biases in fixed effects estimates, $\beta_0 = 0.3271, \beta_x = -0.7088, \beta_z = -0.1506, \beta_{z^*} = -0.3409$ and very low coverage rates for the 95% CI $(68\%, 1\%, 90\%, 84\%)$ respectively for Naive 1 approach. The biases become smaller for Naive 2 method such as $\beta_0 = 0.0190, \beta_x = 0.0194, \beta_z = 0.0433, \beta_{z^*} = -0.0062$. However, the proposed approach seems to perform very well with respect to biases, RMSEs as well as the coverage rates. The biases $(\beta_0 = 0.0087, \beta_x = -0.0042, \beta_z = -0.0077, \beta_{z^*} = -0.0109)$ are fairly small with 100% coverage rates. The RMSEs obtained from the methods Naive 1 and Naive 2 are much bigger than those obtained from the the Proposed approach. Also, the RMSEs for the method Naive 2 are much smaller than the corresponding values of the method Naive 1.

	Bias	RMSE	CR
β_0	0.0041	0.1585	0.96
β_x	0.0160	0.1521	0.95
β_z	0.0187	0.2264	0.94
β_{z^*}	0.0226	0.2670	0.95
σ_u^2	0.0213	0.2486	0.62

Table 3.4a: Simulation results for Naive 1 with m=10 and no error $(\sigma = 0.0)$

	Bias	RMSE	CR
β_0	-0.0010	0.0635	1.00
β_x	0.0033	0.0533	1.00
β_z	0.0057	0.0705	1.00
β_{z^*}	0.0148	0.1100	1.00
σ_u^2	-0.0203	0.1044	0.28

Table 3.4b: Simulation results for Naive 2 with m=10 and no error $(\sigma = 0.0)$

	Bias	RMSE	CR
β_0	0.0002	0.0214	1.00
β_x	0.0005	0.0157	1.00
β_z	-0.0014	0.0247	1.00
β_{z^*}	0.0005	0.0343	1.00
δ_0	-0.0816	0.2501	0.30
δ_1	-0.0466	0.1411	0.32
λ_0	0.0158	0.0769	0.34
λ_1	0.0070	0.1271	0.30

Table 3.4c: Simulation results for Proposed Method with m=10 and no error ($\sigma = 0.0$)

Tables 3.4a, 3.4b and 3.4c represents the results of the estimates of the coefficients of fixed and random effects for the three approaches under no measurement error and 10 number of follow-ups ($m = 10$). In terms of bias, RMSEs and coverage rates of 95% CI, it is apparent from the results that the Proposed approach works perfectly well compared to the other two approaches for fixed effects parameters estimate. The parameters of the random effect covariance matrix (GARPs, IVs) also shows smaller and higher coverage rates compared to the 5 number of follow-ups ($m = 5$). As expected, the RMSEs of the model parameters estimate for the Proposed method are smaller than the corresponding values of the both methods Naive 1 and Naive 2; and also the method Naive 2 has smaller RMSEs compared to the method Naive 1.

	Bias	RMSE	CR
β_0	0.0402	0.1572	0.98
β_x	-0.2643	0.2913	0.44
β_z	-0.1517	0.2557	0.85
β_{z^*}	-0.0418	0.2557	0.96
σ_u^2	0.0213	0.2486	0.62

Table 3.5a: Simulation results for Naive 1 with m=10 and moderate error ($\sigma = 0.4$)

	Bias	RMSE	CR
β_0	0.0090	0.0678	1.00
β_x	0.0083	0.0504	1.00
β_z	0.0074	0.0780	1.00
β_{z^*}	0.0059	0.1033	1.00
σ_u^2	-0.0199	0.1016	0.29

Table 3.5b: Simulation results for Naive 2 with m=10 and moderate error ($\sigma = 0.4$)

	Bias	RMSE	CR
β_0	0.0006	0.0201	1.00
β_x	-0.0003	0.0148	1.00
β_z	0.0003	0.0228	1.00
β_{z^*}	0.0002	0.0307	1.00
δ_0	-0.0696	0.2221	0.39
δ_1	-0.0454	0.1248	0.35
λ_0	0.0021	0.0565	0.39
λ_1	-0.0151	0.1009	0.38

Table 3.5c: Simulation results for Proposed Method with m=10 and moderate error ($\sigma = 0.4$)

From Tables 3.5a, 3.5b and 3.5c, it is observed that under moderate measurement errors and 10 number of follow-ups, the Proposed method provides large improvement on biases for fixed effect estimates ($\beta_0 = 0.0006, \beta_x = -0.0003, \beta_z = 0.0003, \beta_{z^*} = 0.0002$) than Naive 1 ($\beta_0 = 0.0402, \beta_x = 0.2643, \beta_z = -0.1517, \beta_{z^*} =$

-0.0418) and moderate improvement than Naive 2 ($\beta_0 = 0.0090, \beta_x = 0.0083, \beta_z = 0.0075, \beta_{z^*} = -0.0059$). These improvement are also true for RMSEs and coverage rates of 95% CI. In this case, we can also see the smaller amount of biases with good coverage rates of random effects covariance parameters in comparison with the 5 number of follow-ups ($m = 5$).

	Bias	RMSE	CR
β_0	0.2444	0.2876	0.64
β_x	-0.6816	0.6892	0.00
β_z	-0.1937	0.2650	0.79
β_{z^*}	-0.2646	0.3761	0.78
σ_u^2	0.0213	0.2486	0.62

Table 3.6a: Simulation results for Naive 1 with m=10 and severe error ($\sigma = 0.8$)

	Bias	RMSE	CR
β_0	0.0012	0.0615	1.00
β_x	0.0010	0.0483	1.00
β_z	0.0006	0.0812	1.00
β_{z^*}	-0.0010	0.0895	1.00
σ_u^2	-0.0114	0.0931	0.27

Table 3.6b: Simulation results for Naive 2 with m=10 and severe error ($\sigma = 0.8$)

The performance of the three methods in the context of the severe measurement errors and $m = 10$ is presented in Tables 3.6a, 3.6b and 3.6c. It is clear from the results that the Proposed approach performs well with small biases and good coverage rates for fixed effects coefficients and random effect covariance coefficients. Also, the RMSEs show the consistency of the performance as other scenarios. On the other hand, the Naive 1 method performs poorly with relatively large amount

	Bias	RMSE	CR
β_0	0.0019	0.0192	1.00
β_x	0.0011	0.0157	1.00
β_z	-0.0003	0.0219	1.00
β_{z^*}	-0.0040	0.0294	1.00
δ_0	-0.1009	0.2390	0.31
δ_1	-0.0521	0.1325	0.32
λ_0	0.0006	0.0571	0.35
λ_1	-0.0161	0.1130	0.32

Table 3.6c: Simulation results for Proposed Method with m=10 and severe error ($\sigma = 0.8$)

of biases and low coverage rates particularly for the coefficient of measurement error variable ($\beta_x = -0.6816$ with 0% coverage rates). The performance of the Proposed method and Naive 2 method are similar in terms of biases but the Proposed method has much smaller RMSEs. The estimates of GARPs and IVs represent fairly small biases with good coverage rates and smaller RMSEs compared to the corresponding values for the 5 number of follow-ups ($m = 5$).

Overall, based on the simulation results, It is evident that the larger biases can occur in the fixed effects parameters by ignoring the measurement error in covariate and also not specifying the distribution of random effects correctly. The simulation results also demonstrate that the Proposed approach performs very well with small biases and RMSEs as well as good coverage rates for 95% CI. Moreover, as expected, the RMSEs for the all methods tend to decrease when the number of follow-ups m increases.

Chapter 4

Application

In this chapter we first provide a brief description of the Manitoba Follow-up Study (MFUS) data set which is used for the application purpose in the thesis. The basic description of this study is obtained from an open web-site:

http://www.mfus.ca/index.php.

Moreover detailed description of this study can be found in Tate et al., (2013, 2015). In section 4.1, we provide a brief description of the MFUS. A detailed background description of the MFUS data set is given in section 4.2. Section 4.3 includes the application of the proposed method using the MFUS data set.

4.1 Manitoba Follow-up Study

The Manitoba Follow-up Study (MFUS) is the longest running study of cardiovascular disease and ageing in Canada. It is believed that the MFSU is the only cohort study in the world which is financed by the members who are being studied. The MFUS cohort consists of 3983 men who were recruits in the Royal Canadian Air

Force during the early years of World War II and was established at the University of Manitoba on July 1, 1948 (Mathewson et al., 1965). Dr. FAL Mathewson was responsible for the initial physical examination to evaluate the fitness of approximately 7000 male air crew recruits for the Royal Canadian Air Force during the World War II. The physical examination of these men include general health assessment, measurement of body weight and blood pressure, medical history of past illness and recording of an electrocardiogram. After the war, these men were invited to take part in a longitudinal study to determine their clinical significance. Thus the seed was planted during the World War II for the MFUS. After the war, about 10% of the study participants were relocating to the United States or overseas and rest of the participants returned to residence throughout the Canada. The mean age of the men in the cohort was around 31 years, with about 90% between age 20 and 39 years.It was declared that all men were free of clinical evidence of ischaemic heart disease. The baseline measurement of systolic and diastolic blood pressure and body mass index (mean $\pm$ standard deviation) were found 121 ± 10 mmHg, 76 ± 8 mmHg and 23.8 ± 2.7kg/m^2, respectively (Tate et al., 2015). The MFUS members were assumed a wide variety of occupations, with approximately half of the cohort members ($\sim$ 2000) remained involved with aviation and about half of those ($\sim$ 1000) remaining as career pilots. The other half began new or renewed civilian occupations (Tate et al., 2015). At present, MFUS continues with its 68th year of uninterrupted study. In addition to continued research in cardiovascular health, the current interest has developed in understanding successful ageing.

The MFUS protocol involves the regular basis medical examination of each study member by his physician. Initially the study members were contacted for examination at 5-years intervals between 1948 to 1963, then 3-years interval between 1964 to 1978, from 1978 to 2003 every year, twice a year from 2004 to 2007 and since 2007 three times each year (Tate et al., 2015). On July 1, 2013 there were 429 members remained in the study with mean age 92 ± 3.2 years and around 91% members living in Canada. A comprehensive information of members with their mean age during calender periods between 1948 and 2013 are described in the following table (Tate et al., 2015).

Calender period	Number of alive at beginning of the period	Mean ± Standard deviation
July 1, 1948	-	-
July 1, 1948 to June 30, 1963	3983	31.1 ± 6.1
July 1, 1963 to June 30, 1978	3786	46.0 ± 5.9
July 1, 1978 to June 30, 1993	3268	60.3 ± 5.2
July 1, 1993 to June 30, 2003	2254	74.2 ± 4.2
July 1, 2003 to June 30, 2013	1287	83.2 ± 3.4

Table 4.1: Description of Manitoba Follow-up Study contacted members during the calender period

There has been a great contribution by the researchers to understand both cardiovascular disease and aging and more than 50 peer-reviewed findings based on this study have been published in different renowned journals.

4.2 Data and Variables

For the purpose of our study, a sub-sample of the MFUS data has been used. For the selection of sample from MFUS participants, approximately 500 men, 1/8th of the cohort was chosen. In particular, from the registry file (one record per MFUS man) a random number in the interval $[0, 1]$ was selected using the *ranuni* function in SAS software (SAS Institute Inc., 2003), and the man retained in our file if < 0.125. The selected people were merged by ID number with the blood pressure and body weight file, and all measurements recorded between July 1, 1948 and July 1, 2008 were kept. Finally, the data set contains 373 members with 10 first follow-up observations for each member has been used in the analyses. High blood pressure is an important risk factor for cardiovascular disease and is one of the major cause of mortality. However, environmental and genetic factors and their interactions may be the cause for complex disorder disease like high blood pressure (Kraft et al., 2003). In this work, it is of interest how hypertension which is based on blood pressure is associated with corresponding risk factors and how individual measurements vary within subjects. We use the the American Heart Association guidelines for cut-off points for blood pressure measurements (AHA, 2003). The cut-off points represent to the clinical classification for hypertension as they relate to the systolic and diastolic blood pressure measurement. For the purpose of analysis, we divide individuals into two categories: having hypertension and not having hypertension. An individual is said to have hypertension if his/her systolic blood pressure is greater than 140 mmHg or his/her diastolic blood pressure is greater than 90 mmHg (Stockwel et al., 1994). The main covariates of interest include age, body mass index (BMI) and ischaemic heart disease status [IHD: Yes, No (ref)]. An individual is said to have IHD if he develops IHD in any of his 10 first follow-ups. These covariate are taken into consideration as they are found to have significant impact on the occurrence of hypertension in many studies (Stamler 1991; Kaufman et al., 1997; Humayun et al., 2009; Zhang et al., 2014; Hoque et al., 2014). In particular, there have been many literatures which show the association between IHD and blood pressure (Rabkin et al., 1978; Tate et al., 1998).

For this study, July 1, 1948 or the closest date from July 1, 1948 is considered as the baseline. Based on the baseline information, it is observed that among the individuals, the mean age is 30.56 years with standard deviation 5.24 and the minimum and maximum age are 20.20 and 52.56, respectively. The mean BMI is 23.87 kg/m^2 with standard deviation 2.66 kg/m^2. It is also noticed that, at baseline all individuals are IHD free. There are 9.1% individuals have IHD upto 10th follow-up and upto July 1, 2008, 25.64% individuals develop IHD.

In many longitudinal studies, when BMI of a person is reported for a certain age, the variable of interest for BMI is actually the long term average values of BMI for that person in that year. BMI is the the ratio of weight (kg) and height (m^2) and weight has a daily as well as seasonal variation. That's why the true and observed BMI differ (Abarin et al., 2014). Moreover, since only the overall weight and height is considered to calculate BMI, there is always an overestimation and underestimation issue of true BMI. Many literatures show that BMI is subject to measurement error (Prentice, 1996; Rothman et al., 2008; O'Neil et al., 2013;

Abarin et al., 2014).

In MFUS, all BMIs are reported by the physicians, however, the BMIs are based on the follow-up weights and heights from the baseline. It means that the physicians in the follow-up times only ask for the weights and then report the BMI using the weights with the heights at the base line. Hence, it can be considered that there exists a variability of any BMI measurement taken at a specific assessment time for an individual. In particular, it can be said that the observed BMI may overestimate the true BMI.

Let the response variable Y_{ij} be the binary response, taking 1 if the subject i has hypertension at assessment j, and 0 otherwise. We can consider the model as follows:

$$\text{logit}\{P_{ij}\} = \beta_0 + \beta_1 \text{Age}_{ij} + \beta_2 \text{BMI}_i + \beta_3 \text{IHD}_i + u_{ij}, \tag{4.1}$$

$$i = 1, 2, \ldots, 373 \,\text{and}\, j = 1, 2, \ldots, 10,$$

where $P_{ij} = P(Y_{ij} = 1 | \text{Age}_{ij}, \text{BMI}_i, \text{IHD}_i, u_{ij})$ and each individual has 10 visits. It is assumed that $\mathbf{u}_i = (u_{i1}, \ldots, u_{ij})$ follow a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_i$.

Here BMI$_i$, representing the true body mass index over time for subject i, which cannot be observed in practice and is treated as the error-contaminated covariates. To feature the measurement error variation we employ the following classical structural measurement error model

$$\text{BMI}_{ij} = \text{BMI}_i + e_{ij},$$

where BMI$_{ij}$ is the measurement taken for subject i at assessment time point j and e_{ij}'s are assumed to follow independent normal distribution with mean 0 and variance σ^2. Moreover, it is assumed that

$$BMI_i \sim N(\mu_x, \sigma_x^2).$$

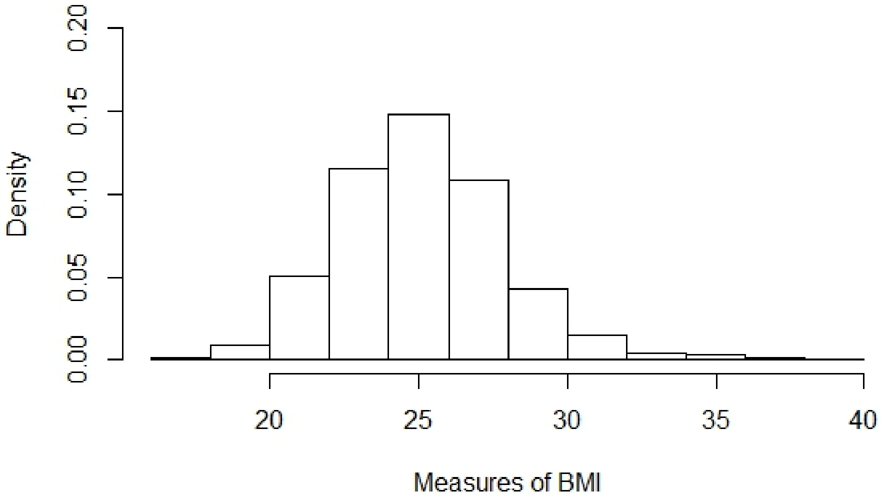

Figure 4.1: BMI measurements from the MFUS data

The Figure 4.1 represents a histogram of measures of BMI of individuals from the MFUS data. It shows that the data for BMI is approximately normally distributed.

4.3 Analysis of Impact of Covariates on Hypertension

We analyse the data with the three approaches - the Proposed approach, Naive 2 approach where constant random effects covariance matrix is considered across the subjects and Naive 1 approach where measurement errors in covariates are also ignored. The results for the methods Naive 1 and Naive 2 are reported in Tables 4.2 and 4.3. In particular, the model parameters estimates, their standard errors and corresponding 95% confidence intervals for the both Naive methods are provided. Table 4.4 represents the results from the proposed method using AR(1) structure of the random effects covariance matrix. The estimates of fixed effect parameters and GARP and IV parameters associated with random effect covariance matrix with standard error and 95% confidence interval are reported (Table 4.4). The GARP and IV parameters are obtained by specifying $k_{i,jt}$ and $h_{i,j}$ as follows:

$$\mathbf{k}_{i,j,j-1} = (1, \mathrm{IHD}_i) \quad \text{and} \quad \mathbf{h}_{i,j} = (1, \mathrm{IHD}_i),$$

and estimated value of $\boldsymbol{\Sigma}_i$ is calculated using $\boldsymbol{\Sigma}_i = \mathbf{T}_i^{-1}\mathbf{D}_i(\mathbf{T}_i^T)^{-1}$, where $\mathbf{D}_i = \mathrm{diag}(\sigma_{i1}^2, \sigma_{i2}^2, \ldots, \sigma_{in_i}^2)$, with $log\left(\sigma_{ij}^2\right) = \mathbf{h}_{i,j}^T\boldsymbol{\lambda}$ and $\mathbf{T}_i$ is a unit lower triangular matrix having ones on its diagonal and $-\phi_{i,jt}$ ($\phi_{i,jt} = \mathbf{k}_{i,jt}^T\delta$) in the (j,t)th position for $2 \leq j \leq 10$. Here we specify the following structure for the parameters of random effects covariance:

$$\phi_{i,jt} = \delta_0 I(|j-t| = 1) + \delta_1 I(|j-t| = 1) IHD_i \quad \text{and}$$

$$log(\sigma_{ij}^2) = \lambda_0 + \lambda_1 IHD_i \tag{4.2}$$

Moreover, the adjusted odds ratio (OR) is used to compare the odd of occurrence of hypertension with covariates. In logistic regression, estimate of OR for covariates x_{ij}, $i = 1, \ldots, n$; $j = 1, \ldots, n_i$, can be obtained as $\widehat{OR} = \exp(\hat{\beta}_j)$. The $100(1 - \alpha)\%$ confidence interval for OR is

$$\widehat{OR} \pm z_{\alpha/2}\sqrt{\text{var}(\widehat{OR})},$$

where $\text{var}(\widehat{OR}) = \left(\widehat{OR}\right)^2 \text{var}(\hat{\beta}_j) = \exp(2\hat{\beta}_j)\text{var}(\hat{\beta}_j)$, using the delta method. The estimated OR and its standard error with 95% confidence interval for the covariates under the all three methods are given in Tables 4.5, 4.6 and 4.7.

4.3.1 Estimation of Parameters

From Table 4.2 (Naive 1), it can be observed that BMI, age and IHD are positively associated with the occurrence of hypertension.

	β_0	β_{BMI}	β_{Age}	β_{IHD}	σ_u^2
Estimate	-13.718	0.278	0.089	1.227	1.982
SE	0.580	0.021	0.006	0.162	0.046
95% LB	-14.855	0.237	0.077	0.909	1.892
95% UB	-12.581	0.319	0.101	1.545	2.072

Table 4.2: Estimate, standard error (SE), and lower bound (LB) and upper bound (UB) 95% CI of model parameters estimate for the method Naive 1

The 95% confidence intervals of these fixed effect estimates indicate the significant effect of these covariates on hypertension.

	β_0	β_{BMI}	β_{Age}	β_{IHD}	σ_u^2	μ_x	σ_x^2	σ^2
Estimate	-13.160	0.269	0.083	1.033	1.959	23.981	6.612	1.380
SE	0.573	0.020	0.006	0.161	0.045	0.042	0.153	0.032
95% LB	-14.283	0.230	0.071	0.717	1.871	23.899	6.312	1.317
95% UB	-12.037	0.308	0.095	1.349	2.047	24.063	6.912	1.443

Table 4.3: Estimate, standard error (SE), and lower bound (LB) and upper bound (UB) 95% CI of model parameters estimate for the method Naive 2

Table 4.3 (Naive 2) reveals that BMI is positively associated with the development of hypertension and this effect is found to be significant. Age and IHD also have significant positive effects on the hypertension. Here, the reliability ratio, $\left[\frac{\sigma_x^2}{(\sigma_x^2+\sigma_e^2)}\right]$, is 0.83 which indicates the measure of amount of error associated with the covariate BMI. Hence, it is clear that there is 17% error associated with the covariate BMI.

The estimates of GARP and IV parameters in Table 4.4 indicate that the covariance matrix varied according to the IHD group. This result demonstrates that the random effect covariance matrix differs by measured covariates and neglecting this heterogeneity can cause the biased estimates of mixed effects (Heagerty and Kurland, 2001). In the estimates of IV, the coefficient of IHD was found significant which indicates that the estimated IV was higher for those individuals who have IHD than the individuals without IHD. By following equation (4.2), the estimated values for $\hat{\mathbf{D}}$ for each group of IHD are $\log\hat{\mathbf{D}}_{(IHD=0)} =$ diag $(1.56, 1.56, 1.56, 1.56, 1.56, 1.56, 1.56, 1.56, 1.56, 1.56)$ and $\log\hat{\mathbf{D}}_{(IHD=1)} =$ diag $(3.53, 3.53, 3.53, 3.53, 3.53, 3.53, 3.53, 3.53, 3.53, 3.53)$.

Fixed Effect parameters				
	β_0	β_{BMI}	β_{Age}	β_{IHD}
Estimate	-12.089	0.208	0.089	0.561
SE	0.588	0.021	0.005	0.180
95% LB	-13.241	0.167	0.079	0.208
95% UB	-10.937	0.249	0.099	0.914
Generalized autoregressive parameters (GARPs: δ)				
	δ_0	δ_1(IHD)		
Estimate	0.146	-0.111		
SE	0.005	0.001		
95% LB	0.136	-0.113		
95% UB	0.156	-0.109		
Innovation Variance parameters (IVs: λ)				
	λ_0	λ_1(IHD)		
Estimate	1.561	1.973		
SE	0.012	0.012		
95% LB	1.537	1.949		
95% UB	1.585	1.997		
Other Associated parameters				
	μ_x	σ_x^2	σ^2	
Estimate	24.034	6.533	1.380	
SE	0.042	0.151	0.032	
95% LB	23.952	6.237	1.317	
95% UB	24.116	6.829	1.443	

Table 4.4: Estimate, standard error (SE), and lower bound (LB) and upper bound (UB) 95% CI of model parameters estimate for the Proposed method

This demonstrates that the estimated IV parameters vary across the status of IHD. The coefficient of IHD in the estimates of GARP is also found significant and this also indicates the substantial variation of GARPs across the status of IHD. As $\hat{\delta}_1$ and $\hat{\lambda}_1$ are statistically significant from zero, the Proposed approach also works better than assuming the same correlation structural AR(1) for the all subjects (Naive 2). By following equation (4.2), the estimated values of $\hat{\mathbf{T}}$ for each group of

IHD are given by

$\hat{\mathbf{T}}_{(IHD=0)} =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.15 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.15 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.15 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.15 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.15 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.15 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.15 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.15 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.15 & 1 \end{bmatrix},$$

$\hat{\mathbf{T}}_{(IHD=1)} =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.04 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.04 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.04 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.04 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.04 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.04 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.04 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.04 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.04 & 1 \end{bmatrix}.$$

The significant estimate of coefficient of BMI indicates that the estimated conditional probability of hypertension given the random effects increases with the increase of individuals BMI. Also, Age was found significant and we know that the conditional probability of hypertension increases as Age increases. The reliability ratio for the Proposed method is also 0.82 which indicates the amount of 18% error involvement with the covariate BMI.

The fixed effects estimates for the all three methods reveal the same nature of covariates effects on hypertension status, but there exists a variation in the magnitudes of the fixed effects for the naive methods compared to the Proposed method, especially in case of IHD covariate. The naive approaches suggest a significant positive IHD effect of individuals who have IHD compared to the individuals without IHD on hypertension status that is nearly more than two times higher than the proposed method estimate. This might be due to ignoring the heterogeneity or temporal dependence on random effect covariance matrix in case of the naive methods. The estimates of BMI effect are also larger in naive analyses than the Proposed method. We also observe that the random effect variance estimate (and its standard error) for the both Naive methods are close to each other, and the variance of measurement error (σ^2) and corresponding mean and variance of BMI (μ_x, σ_x^2) are also similar for the both Naive 2 and Proposed approaches.

4.3.2 Estimation of Odds Ratio

To examine the association of a covariate with the occurrence of hypertension controlling other covariates in the model, one may use adjusted OR. The adjusted OR and its standard error with 95% confidence interval for the covariates under the three approaches are given in Tables 4.5, 4.6 and 4.7.

From Table 4.5 (Naive 1), it is found that the OR for BMI is 1.320 which implies that with one unit increase in BMI, the odds of developing hypertension is expected to increase 32%. It is interesting to observe that the development of hypertension increases with the increase of Age as well (OR=1.09). In case of IHD, it can be

	OR	SE	p-value	95% LB	95% UB
BMI	1.320	0.028	0.000	1.266	1.375
Age	1.093	0.007	0.000	1.080	1.106
IHD	3.411	0.553	0.026	2.328	4.494

Table 4.5: Odds ratio (OR), standard error (SE), p-value, lower bound (LB) and upper bound (UB) of 95% CI of model parameters estimate for the method Naive 1

described that an individual with IHD is 241% more likely to have hypertension compared to an individual without IHD. Note that all the ORs are statistically significant as the p-values are less than 0.05.

	OR	SE	p-value	95% LB	95% UB
BMI	1.309	0.026	0.000	1.257	1.360
Age	1.087	0.007	0.000	1.074	1.099
IHD	2.809	0.452	0.022	1.923	3.696

Table 4.6: Odds ratio (OR), standard error (SE), p-value, lower bound (LB) and upper bound (UB) of 95% CI of model parameters estimate for the method Naive 2

Based on Table 4.6 (Naive 2), it is observed that the estimated OR for BMI is 1.309 with p-value 0.00. It implies that the odds of having hypertension is significantly increased by 31% with one unit increase of BMI. The OR of age also reveals the same nature that means with one unit increase in Age, the odds of developing hypertension is increased by 9%. For the IHD, the OR is 2.809 which means that the individuals with IHD are 181% more likely to develop hypertension than the individuals without IHD. Also, we observe that the all three covariates are statistically significant as their corresponding p-values are less than 0.05.

Same nature can be revealed for BMI, Age and IHD from Table 4.7 (Proposed

	OR	SE	p-value	95% LB	95% UB
BMI	1.231	0.026	0.000	1.181	1.282
Age	1.093	0.005	0.000	1.082	1.104
IHD	1.752	0.315	0.075	1.134	2.371

Table 4.7: Odds ratio (OR), standard error (SE), p-value, lower bound (LB) and upper bound (UB) of 95% CI of model parameters estimate for the Proposed method

method). The OR for BMI indicates the significant increase of odds (1.231) of developing hypertension with the increase of one unit in BMI. The odds of having hypertension is expected to increase 9% with one unit increase in Age. For the IHD, the OR is 1.752 which means that the individuals with the IHD are 75% more likely to develop hypertension than the individuals without the IHD, and the all three covariates are statistically significant from zero.

Chapter 5

Summary and Discussion

There have been considerable research interest in longitudinal data and numerous methods have been proposed to analyse such data with various features. Covariates measurement error are very common problems in longitudinal data. The statistical inference will be biased and misleading without the consideration of the measurement error. Therefore, to obtain the valid statistical inference, it is important to address the measurement error issue.

The GLMMs are commonly used to analyse longitudinal data. In these models, it is typically assumed that the random effects covariance matrix is constant across the subjects. In many situations, however, this correlation structure among subjects may differ and ignoring this heterogeneity can cause the biased estimates of fixed effect parameters. To address this, Lee et al. (2012) proposed a heterogeneous random effects covariance matrix under the GLMMs ,however, they assumed that the covariates are error-free. The main contribution of this thesis is to properly model the random effects covariance matrix under the GLMMs with covariates measurement error. For this purpose, we extend the proposed model by Lee et al. (2012) to model the random effect covariance matrix for the GLMMs to the case when the covariates are subject to measurement error using modified Cholesky decomposition. This covariance matrix is decomposed to the GARPs and IVs parameters and this structure is able to accommodate the heterogeneous covariance matrix depending on subject-specific covariates.

In Chapter 2, we have given a general framework to model the random effects covariance matrix via subject-specific covariates for the GLMMs with measurement error covariates and provided a general inference procedure to estimate the model parameters by exploiting Monte Carlo EM algorithm.

In Chapter 3, simulation studies have been conducted to evaluate the performance of the Proposed method over Naive approaches. From our empirical results, it has been observed that the larger biases can occur in the fixed effects parameters by ignoring the measurement error in covariates and also not specifying the distribution of random effects correctly. The simulation findings also demonstrate that the Proposed approach performs very well in terms of small biases and RMSEs as well as coverage rates of the model parameters estimate. Moreover, as expected, with the increase of number of follow-ups for each subject, the RMSEs for the all methods tended to decrease.

An application of the Proposed method has been shown using the MFUS data set in Chapter 4. Based on the results, it is clear that the random effects covariance matrix differs by IHD and there exists a variation in the magnitudes of the fixed effects for the Proposed method compared to the Naive methods. For the purpose of our analysis, we considered IHD as a time independent variable. However, in MFUS data set, IHD is not a time independent variable. In our data application, the random effects covariance matrix only depended on categorical covariate. However, the Proposed method can also be applied to the continuous covariates.

The Proposed approach for modelling random effects covariance matrix is also computationally attractive and provides parameters which have sensible interpretation for modelling trajectories over time. Especially, the dependence and variability of the random effects can be characterized by the covariance parameters.

In the longitudinal data analysis with covariates measurement error, if the main interest is the covariance structure or subject-specific prediction, then proper care needs to be taken in modelling the covariance structure as well as the measurement error. However, taking proper care is important even if these are not the main or direct interest. As an example, if the random effects covariance matrix is not modelled correctly when it is function of subject-specific covariates, then the inference will be incorrect as the standard errors and confidence intervals for the fixed effects as well as variance components will be incorrect.

It should be noted that in measurement error problem, model identifiability is an important issue. In case of measurement error, additional data source such as a validation sub-sample or replicates is needed to perform a measurement error analysis (Carroll et al., 2006). In longitudinal studies, repeated measurements are collected for error-prone variables, and for identifying model parameters these measurements can be used as replicates. In particular, the parameters of an error model are often identifiable if the number of repeated assessment of measurement error covariates is bigger than the number of parameters in the error model (Yi et al., 2011). If the parameters are not identifiable, then in numerical iterative procedures, fast divergence can occur. For example, if there is a non-identifiability problem then the EM algorithm would diverge quickly (Stubbendick and Ibrahim, 2003). Our numerical experience, however, does not indicate that there is an issue with non-identifiability for the models considered in this thesis.

In the past twenty years significant contributions have been made in the area of longitudinal data with covariate measurement errors. However, there are still a lot of interesting and important problems related with this thesis need to be explored in the future. For example, we can propose our random effects subject-specific variance-covariance matrix for the marginalized random effects models as well. Moreover, missing data is a common problem in longitudinal experiments. Missing data can occur due to various reasons such as the desired measurements from individuals are not available, lost to follow-ups or otherwise not taken. As a result, attention has been grown on the analysis of longitudinal data with covariates measurement error and missing responses/covariates (Liu and Wu, 2007; Yi et al., 2011; Yi et al., 2012). In the work of Yi et al. (2011), a general framework was proposed to analyse longitudinal data with covariates measurement error and missing responses. However, in their work the random effects covariance matrix is left unspecified. Hence, our proposed approach can be extended for the case of both covariates measurement error and missing responses.

Furthermore, in longitudinal studies, measurement of response variables may contain error due to imperfect measuring system or other reasons. For example, there always exist some variability in reporting systolic and diastolic blood pressure measurement which can occur because of the digital preference or rounding the value to higher or lower closest one. As an example, let one individual systolic blood pressure measurement is 131.2 and it is usual to round this value as 132 or 130 which entails a variability. In case of MFUS data set, it is observed that on base line in case of systolic blood pressure (SBP), there are $31, 69, 56$ and 24 individuals who report SBP as $110, 120, 130$ and 140, respectively which indicates the preference of rounding the value or digital preference. Same scenario can be observed in case of diastolic blood pressure (DBP), such as, $65, 15, 95, 15,$ and 31 individuals who report DBP as $70, 75, 80, 85$ and 95, respectively which indicates its variability. Hence, it would be an interesting work to model the random effects covariance matrix in case of longitudinal data with response measurement error.

In conclusion, in the presence of covariates measurement error in longitudinal data, incorrectly modelling the random effects covariance matrix has significant effects on inference in model parameters. Hence, it is important to properly model the random effects covariance matrix in the presence of covariates measurement error.

Bibliography

[1] American Heart Association, AHA (2003). http://www.heart.org/HEARTORG/.

[2] ABARIN, T., LI, H., WANG, L., AND BRIOLLAIS, L. On Method of Moments Estimation in Linear Mixed Effects Models with Measurement Error on Covariates and Response with Application to a Longitudinal Study of Gene-Environment Interaction. *Statistics in Biosciences 6*, 1 (2014), 1–18.

[3] AZZALINI, A. Logistic Regression for Autocorrelated Data with Application to Repeated Measures. *Biometrika 81*, 4 (dec 1994), 767.

[4] BERKSON, J. Are there Two Regressions? *Journal of the American Statistical Association 45*, 250 (jun 1950), 164–180.

[5] BRESLOW, N. E., AND CLAYTON, D. G. Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association 88*, 421 (1993), 9–25.

[6] BRESLOW, N. E., AND LIN, X. H. Bias Correction in Generalized Linear Mixed Models with a Single-Component of Dispersion. *Biometrika 82*, 1 (1995), 81–91.

[7] CAFFO, B. S., JANK, W., AND JONES, G. L. Ascent-based Monte Carlo expectation- maximization. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*, 2 (2005), 235–251.

[8] CARROLL, R. J., RUPPERT, D., AND STEFANSKI, L. A. *Measurement error in nonlinear models*. London: Chapman & Hal l/CRC press, 1995.

[9] CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A., AND CRAINICEANU, C. M. *Measurement error in nonlinear models: a modern perspective*. Chapman & Hall /CRC press, 2006.

[10] CHEN, B., YI, G. Y., AND COOK, R. J. Likelihood analysis of joint marginal and conditional models for longitudinal categorical data. *Canadian Journal of Statistics 37*, 2 (jun 2009), 182–205.

[11] CHIU, T. Y. M., LEONARD, T., AND TSUI, K.-W. The Matrix-Logarithmic Covariance Model. *Journal of the American Statistical Association 91*, 433 (1996), 198.

[12] COOK, J. R., AND STEFANSKI, L. A. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association 89*, 428 (1994), 1314–1328.

[13] DANIELS, M. J., AND POURAHMADI, M. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika 89*, 3 (2002), 553–566.

[14] DANIELS, M. J., AND ZHAO, Y. D. Modelling the random effects covariance matrix in longitudinal data. *Statistics in Medicine 22*, 10 (2003), 1631–1647.

[15] DIGGLE, P. *Analysis of longitudinal data*. Oxford University Press, 2002.

[16] FELLNER, W. H. Robust Estimation of Variance Components. *Technometrics 28*, 1 (feb 1986), 51–60.

[17] FORT, G., AND MOULINES, E. Convergence of the Monte Carlo expectation maximization for curved exponential families. *Annals of Statistics 31*, 4 (2003), 1220–1259.

[18] FULLER, W. A. *Measurement error models*, vol. 305. John Wiley & Sons, 2009.

[19] GUSTAFSON, P. *Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments*. Chapman & Hall /CRC press, 2004.

[20] HARVILLE, D. A. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association 72*, 358 (1977), 320–338.

[21] HEAGERTY, P. J. Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics 58*, 2 (jun 2002), 342–351.

[22] HEAGERTY, P. J., AND KURLAND, B. F. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika 88* (2001), 973–985.

[23] HOQUE, E., KHOKAN, M. R., AND BARI, W. "Impact of stature on non-communicable diseases: evidence based on Bangladesh Demographic and Health Survey, 2011 data". *BMC public health 14* (2014), 1007.

[24] HUMAYUN, A., SHAH, A. S., ALAM, S., AND HUSSEIN, H. Relationship of body mass index and dyslipidemia in different age groups of male and female population of Peshawar. *Journal of Ayub Medical College, Abbottabad : JAMC 21* (2009), 141–4.

[25] IBRAHIM, J. G., CHEN, M.-H., AND LIPSITZ, S. R. Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika 88*, 2 (2001), 551–564.

[26] IBRAHIM, J. G., LIPSITZ, S. R., AND CHEN, M.-H. Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61*, 1 (1999), 173–190.

[27] KAUFMAN, J. S., ASUZU, M. C., MUFUNDA, J., FORRESTER, T., WILKS, R., LUKE, A., LONG, A. E., AND COOPER, R. S. Relationship Between Blood Pressure and Body Mass Index in Lean Populations. *Hypertension 30*, 6 (1997), 1511–1516.

[28] KRAFT, P., BAUMAN, L., YUAN, J. Y., AND HORVATH, S. Multivariate variance-components analysis of longitudinal blood pressure measurements from the Framingham Heart Study. *BMC.Genet. 4 Suppl 1:* (2003), S55.

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

[30] LEE, K., LEE, J., HAGAN, J., AND YOO, J. K. Modeling the random effects covariance matrix for generalized linear mixed models. *Computational Statistics & Data Analysis 56*, 6 (2012), 1545–1551.

[31] LELE, S. R., NADEEM, K., AND SCHMULAND, B. Estimability and Likelihood Inference for Generalized Linear Mixed Models Using Data Cloning. *Journal of the American Statistical Association 105*, 492 (2010), 1617–1625.

[32] LEVINE, R. R. A., AND CASELLA, G. Implementations of the Monte Carlo EM Algorithm. *Journal of Computational and Graphical Statistics 10*, 3 (2001), 422–439.

[33] LIANG, K.-L., AND ZEGER, S. Longitudinal Data Analysis using Generalized Linear Models. *Biometrika 73* (1986), 13–22.

[34] LIN, X., RAZ, J., AND HARLOW, S. D. Linear mixed models with heterogeneous within-cluster variances. *Biometrics 53*, 3 (1997), 910–923.

[35] LIN, X. H., AND BRESLOW, N. E. Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association 91*, 435 (1996), 1007–1016.

[36] LIU, Q., AND PIERCE, D. A. Heterogeneity in Mantel-Haenszel-type models. *Biometrika 80*, 3 (1993), 543–556.

[37] LIU, W., AND WU, L. Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. *Biometrics 63*, 2 (2007).

[38] MATHEWSON, F. A., BRERETON, C. C., KELTIE, W. A., AND PAUL, G. I. The University of Manitoba Follow-up Study: A Prospective Investigation of Cardiovascular Disease.I. General Descriptionâ€"Mortality and Incidence of Coronary Heart Disease. *Canadian Medical Association journal 92* (1965), 947–953.

[39] MCCULLOCH, C. E. Maximum Likelihood Variance Components Estimation for Binary Data. *Journal of the American Statistical Association 89:425*, February (1994), 330–335.

[40] MCCULLOCH, C. E. Maximum likelihood algorithms for Generalized Linear Mixed Models. *Journal of the American Statistical Association 92* (1997), 162–170.

[41] MCCULLOCH, C. E., AND SEARLE, S. R. *Generalized, Linear, and Mixed Models*. John Wiley & Sons, 2001.

[42] MCLACHLAN, G., AND KRISHNAN, T. *The EM algorithm and extensions*, vol. 382. John Wiley & Sons, 2007.

[43] MENG, X. L., AND VAN DYK, D. Fast EM-type implementations for mixed effects models. *Journal of the Royal Statistical Society Series B-Statistical Methodology 60* (1998), 559–578.

[44] MOLENBERGHS, G., AND KENWARD, M. *Missing data in clinical studies*, vol. 61. John Wiley & Sons, 2007.

[45] NAKAMURA, T. Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika 77*, 1 (mar 1990), 127–137.

[46] NAKAMURA, T. Proportional hazards model with covariates subject to measurement error. *Biometrics 48*, 3 (1992), 829–838.

[47] NEUHAUS, J. M., AND MCCULLOCH, C. E. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society. Series B: Statistical Methodology 68*, 5 (nov 2006), 859–872.

[48] O'NEILL, D., AND SWEETMAN, O. The consequences of measurement error when estimating the impact of obesity on income. *IZA Journal of Labor Economics 2*, 1 (2013), 3.

[49] PAN, J., AND MACKENZIE, G. On modelling mean-covariance structures in longitudinal studies. *Biometrika 90*, 1 (2003), 239–244.

[50] PAN, J., AND MACKENZIE, G. Regression models for covariance structures in longitudinal studies. *Statistical Modelling 6* (2006), 43–57.

[51] POURAHMADI, M. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika 86*, 3 (1999), 677–690.

[52] POURAHMADI, M. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika 87*, 2 (2000), 425–435.

[53] POURAHMADI, M., AND DANIELS, M. J. Dynamic Conditionally Linear Mixed Models for Longitudinal Data. *Biometrics 58*, 1 (mar 2002), 225–231.

[54] PRENTICE, R. L. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika 69*, 2 (aug 1982), 331–342.

[55] PRENTICE, R. L. Measurement error and results from analytic epidemiology: dietary fat and breast cancer. *Journal of the National Cancer Institute 88*, 23 (1996), 1738–1747.

[56] R CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[57] RABKIN, S. W., MATHEWSON, A. L., AND TATE, R. B. Predicting risk of ischemic heart disease and cerebrovascular disease from systolic and diastolic blood pressures. *Annals of Internal Medicine 88*, 3 (1978), 342–345.

[58] ROTHMAN, K. J. BMI-related errors in the measurement of obesity. *International journal of obesity (2005) 32 Suppl 3* (2008), S56–S59.

[59] SAS INSTITUTE INC. *SAS/STAT Software, Version 9.1*. Cary, NC, 2003.

[60] SCHALL, R. Estimation in generalized linear models with random effects. *Biometrika 78*, 4 (dec 1991), 719–727.

[61] STAMLER, J. Epidemiologic findings on body mass and blood pressure in adults. *Annals of Epidemiology 1*, 4 (1991), 347–362.

[62] STIRATELLI, R., LAIRD, N., AND WARE, J. H. Random-effects models for serial observations with binary response. *Biometrics 40*, 4 (1984), 961–71.

[63] STOCKWELL, D. H., MADHAVAN, S., COHEN, H., GIBSON, G., AND ALDERMAN, M. H. The determinants of hypertension awareness, treatment, and control in an insured population. *American Journal of Public Health 84*, 11 (1994), 1768–1774.

[64] STUBBENDICK, A. L., AND IBRAHIM, J. G. Maximum Likelihood Methods for Nonignorable Missing Responses and Covariates in Random Effects Models. *Biometrics 59*, 4 (2003), 1140–1150.

[65] TATE, R. B., CUDDY, T. E., AND MATHEWSON, F. A. L. Cohort Profile: The Manitoba Follow-up Study (MFUS). *International Journal of Epidemiology 44*, 5 (2015), 1528–1536.

[66] TATE, R. B., MANFREDA, J., AND CUDDY, T. E. The effect of age on risk factors for ischemic heart disease: the Manitoba Follow-Up Study, 1948-1993. *Annals of epidemiology 8*, 7 (1998), 415–21.

[67] TATE, R. B., MICHAELS, L., EDWARD CUDDY, T., AND BAYOMI, D. J. Clinical diagnoses before age 75 and men's survival to their 85th birthday: The manitoba follow-up study. *Gerontologist 53*, 1 (2013), 133–141.

[68] TORABI, M. Likelihood inference in generalized linear mixed measurement error models. *Computational Statistics & Data Analysis 57*, 1 (2013), 549–557.

[69] TOSTESON, T. D., STEFANSKI, L. A., AND SCHAFER, D. W. A measurement-error model for binary and ordinal regression. *Statistics in medicine 8*, 9 (1989), 1139–47; discussion 1149.

[70] WANG, N., CARROLL, R. J., AND LIANG, K. Y. Quasilikelihood estimation in measurement error models with correlated replicates. *Biometrics 52*, 2 (1996), 401–11.

[71] WANG, N., LIN, X., GUTIERREZ, R. G., AND CARROLL, R. J. Bias Analysis and SIMEX Approach in Generalized Linear Mixed Measurement Error Models. *Journal of the American Statistical Association 93*, 441 (mar 1998), 249.

[72] WHITTEMORE, A. S. Errors-in-Variables Regression Using Stein Estimates. *The American Statistician 43*, 4 (1989), 226.

[73] WU, L. *Mixed effects models for complex data*. CRC Press, 2009.

[74] WU, L., LIU, W., AND LIU, J. A longitudinal study of children's aggressive behaviours based on multivariate mixed models with incomplete data. *Canadian Journal of Statistics 37*, 3 (2009), 435–452.

[75] YI, G. Y., AND COOK, R. J. Marginal methods for incomplete longitudinal data arising in clusters. *Journal of the American Statistical Association 97*, 460 (2002), 1071–1080.

[76] YI, G. Y., LIU, W., AND WU, L. Simultaneous Inference and Bias Analysis for Longitudinal Data with Covariate Measurement Error and Missing Responses. *Biometrics 67*, 1 (2011), 67–75.

[77] YI, G. Y., MA, Y., AND CARROLL, R. J. A functional generalized method of moments approach for longitudinal studies with missing responses and covariate measurement error. *Biometrika 99*, 1 (mar 2012), 151–165.

[78] ZEGER, S. L., AND HEAGERTY, P. J. Marginalized Multilevel Models and Likelihood Inference. *Statistical Science 15*, 1 (feb 2000), 1–26.

[79] ZEGER, S. L., AND KARIM, M. R. Generalized Linear Models with Random Effects; a Gibbs Sampling Approach. *Journal of the American Statistical Association 86*, 413 (1991), 79–86.

[80] ZHANG, Y.-X., AND WANG, S.-R. Comparison of blood pressure levels among children and adolescents with different body mass index and waist circumference: study in a large sample in Shandong, China. *European journal of nutrition 53*, 2 (2014), 627–34.