GENERALIZED LINEAR MIXED MODEL ANALYSIS USING QUASI-LIKELIHOOD

AMANDA LAMBERT

Generalized Linear Mixed Model Analysis using Quasi-Likelihood

by

©Amanda Lambert

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirement for the Degree of Master of Applied Statistics

Department of Mathematics and Statistics
Memorial University of Newfoundland

November, 2007

St. John's Newfoundland Canada

Abstract

When investigating the relationship between two or more variables, regression is a commonly used method of analysis. Linear regression, in particular, is used when the expected value of the response is a linear function of the explanatory variables. If it is not a linear function, generalized linear regression is used. Furthermore, when the data is not independent, mixed models are used. There are various ways to analyze linear mixed models and generalized linear mixed models. In this thesis, we focus on the moment method of analysis, simulated approaches and the quasi-likelihood method of analysis. Analysis is conducted on simulated data for a linear mixed model, simulated data for a generalized linear mixed model and on a real data set. The real data set is a clustered data set of the number of times a person visits a physician in a given year.

Acknowledgements

I would like to acknowledge my practicum supervisors: Dr. Gary Sneddon and Dr. Brajandra Sutradhar. Their guidance, support and patience throughout the course of this work has been invaluable.

Financial support was provided by Memorial University of Newfoundland's School of Graduate Studies and both of my supervisors during my Master's program. For this support I am greatly appreciative.

Finally, I would like to thank my family for their love and support over the past four years. The endless encouragement of my husband and parents gave me the strength and perseverance that I needed to accomplish this piece of work. My thesis is dedicated to them.

Contents

Abstract ii

Acknowledgements iii

List of Tables viii

List of Figures x

1 Background 1

1.1 Introduction 1

1.2 Moment Method 5

1.3 Quasi-Likelihood Method 8

1.4 Other Approaches 9

1.5 Outline of Thesis 10

2 Simulation Data Analysis - Linear Mixed Model 11

2.1 Introduction 11

2.2 Moment Method Analysis 13

2.3 Simulated Method Analysis 13

2.3.1 Simulation Method 1 15

2.3.2 Simulation Method 2 19

2.3.3 Simulation Method 3 23

2.3.4 Simulation Method 4 27

2.4 Quasi-Likelihood Method Analysis 32

3 Simulation Data Analysis - Generalized Linear Mixed Model 41

3.1 Introduction 41

3.2 Moment Method Analysis 42

3.3 Simulation Method Analysis 51

3.4 Quasi-Likelihood Method Analysis 59

4 Real Data Analysis 71

4.1 Introduction 71

4.2 Exploratory Analysis 72

4.2.1 Summary Statistics 72

4.2.2 Graphs 74

4.3 Data Analysis 82

4.3.1 Fixed-Effects Logistic Regression 83

4.3.2 Moment Method Analysis 88

4.3.3 Quasi-Likelihood Method Analysis 90

List of Tables

2.1 Moment Method Estimates (LMM) 15

2.2 Moment Method Variances (LMM) 16

2.3 Simulated Method 1 Estimates (LMM) 19

2.4 Simulated Method 1 Variances (LMM) 20

2.5 Simulated Method 2 Estimates (LMM) 23

2.6 Simulated Method 2 Variances (LMM) 23

2.7 Simulated Method 3 Estimates (LMM) 27

2.8 Simulated Method 3 Variances (LMM) 28

2.9 Simulated Method 4 Estimates (LMM) 32

2.10 Simulated Method 4 Variances (LMM) 33

2.11 Quasi-Likelihood Method Estimates (LMM) 39

2.12 Quasi-Likelihood Method Variances (LMM) 40

3.1 Moment Method Estimates (GLMM) 48

3.2 Moment Method Variances (GLMM) 49

3.3 Moment Method Estimates: Effects of Changes in X (GLMM) 51

3.4 Moment Method Variances: Effects of Changes in X (GLMM) 52

3.5 Simulated Method Estimates (GLMM) 54

3.6 Simulated Method Variances (GLMM) 57

3.7 Simulated Method Estimates: Effects of Changes in X (GLMM) . . . 58

3.8 Simulated Method Variances: Effects of Changes in X (GLMM) . . . 59

3.9 Quasi-Likelihood Method Estimates (GLMM) 65

3.10 Quasi-Likelihood Method Variances (GLMM) 68

3.11 Quasi-Likelihood Method Estimates: Effects of Changes in X (GLMM) 69

3.12 Quasi-Likelihood Method Variances: Effects of Changes in X (GLMM) 70

4.1 Annual Number of Visits to a Physician 72

4.2 Age 73

4.3 Education Level 73

4.4 Number of Chronic Conditions 74

4.5 Moment Method Analysis: Binary dependent variable: Physician Visit Indicator for 1985 to 1990 89

4.6 Quasi-Likelihood Method Analysis: Binary dependent variable: Physician Visit Indicator for 1985 to 1990 91

List of Figures

2.1 Moment Method Histograms (LMM): Simulation 1 14

2.2 Simulated Method 1 Histograms (LMM): Simulation 1 18

2.3 Simulated Method 2 Histograms (LMM): Simulation 1 22

2.4 Simulated Method 3 Histograms (LMM): Simulation 1 26

2.5 Simulated Method 4 Histograms (LMM): Simulation 1 31

3.1 Moment Method Histograms (GLMM): Simulation 1 46

3.2 Moment Method Histograms (GLMM): Simulation 9 47

3.3 Simulated Method Histograms (GLMM): Simulation 7 55

3.4 Simulated Method Histograms (GLMM): Simulation 9 56

3.5 Quasi-Likelihood Method Histograms (GLMM): Simulation 1 66

3.6 Quasi-Likelihood Method Histograms (GLMM): Simulation 7 67

4.1 1985 Physician Visit Indicator Variable Histogram 75

4.2 Explanatory Variable Histograms 76

4.3 Gender by the 1985 Physician Visit Indicator Variable Spineplot . . . 77

4.4 Age by the 1985 Physician Visit Indicator Variable Spineplot 79

4.5 Education Level by the 1985 Physician Visit Indicator Variable Spineplot 80

4.6 Number of Chronic Conditions by the 1985 Physician Visit Indicator Variable Spineplot 81

Chapter 1

Background

1.1 Introduction

Regression is a commonly used method of analysis when investigating a relationship between two or more variables in a nondeterministic fashion (Devore, 2000). It demonstrates how the conditional distribution of the response y differs across subpopulations, which is determined by the predictor or predictors values (Cook and Weisberg, 1999).

There are situations in which one can assume that the relationship between the dependent and the independent variables is a linear function of the parameters. In these situations, linear regression may be used to evaluate this relationship. Specifically, linear regression assumes that the expected value of the response y is a linear function of the predictor values x. The simple linear regression model is defined in (1.1):

$$y = \beta_0 + \beta_1 x + \epsilon. \tag{1.1}$$

In this model, y is the response variable, x is the predictor variable, β_0 and β_1 are unknown fixed effects (fixed but unknown parameters), and ϵ is the error term. The error term is assumed to be normally distributed with $E(\epsilon) = 0$ and $V(\epsilon) = \sigma^2$. The error terms are also assumed to be independent across observations (Devore, 2000).

Simple linear regression assesses the linear relationship between the dependent variable and one independent variable. However, there are situations in which one would like to assess the relationship between the dependent variable and more than one independent variable. In these situations, multiple linear regression may be used (Ramsey and Schafer, 1997). The multiple linear regression model is defined in (1.2):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon. \tag{1.2}$$

As with the linear regression model, y is the response variable, $x_1, x_2, \cdots x_k$ are the predictor variables, $\beta_0, \beta_1, \cdots, \beta_k$ are unknown fixed effects, and ϵ is the error term. The error term is assumed to be normally distributed with $E(\epsilon) = 0$ and $V(\epsilon) = \sigma^2$. The error terms are also assumed to be independent across observations (Devore, 2000).

The linear regression models assume that the dependent variable, y, is normally distributed. There are many cases in which the dependent variable is not normally distributed. In these situations the above models cannot be applied. Instead, we consider generalized linear models (GLM). Generalized linear models may be used when the y variable comes from an exponential family other than normal. Using linear regression, we assume that the expected value of y is a linear function of x. When using a generalized linear model, we assume that some function of the expected value of y is a linear function of x. The function utilized is called a link function. The particular link function that is used will depend on the dependent variable. The type of dependent variable and link function used in the generalized linear regression model determines which type of generalized linear model is appropriate, such as logistic or log-linear (McCullagh and Nelder, 1989).

For example, a logistic generalized linear model would need to be used if the dependent variable is binary. An example would be the person's gender, which can only be either male or female. For a logistic regression model, the link function employed is called the logit. The logit is defined in (1.3):

$$\pi = \frac{e^{\beta_0 + \beta_1 x + \cdots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x + \cdots + \beta_k x_k}}. \tag{1.3}$$

This can be rewritten as a linear function of the parameters.

$$logit(\pi) = log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k. \tag{1.4}$$

In this model, $\mu[Y|X_1, \cdots, X_k] = \pi$ and $Var\{Y|X_1, \cdots, X_k] = \pi(1-\pi)$ (Ramsey and Schafer, 1997) where Y is binary with $P(Y = 1) = p$.

Furthermore, the regression models described above assume that the data are independent. However, there are many situations in which the data may not be independent. For example, the data may exist in clusters, which occur when the data is not distributed independently and identically but occur in homogenous clusters. Homogenous clusters would be considered correlated (Mendenhall, Ott, Scheaffer, 1996). For example, suppose we have data on a number of families. For each family, we know the number of times each member visits a physician in a given year. In this case, the data is clustered into groups of families. Another example would be data collected from high school students on their opinions about a school issue. The opinions of students in particular classrooms may be correlated.

When data is correlated within clusters, the analysis needs to account for this correlation. Using a mixed-effects model is the most common way to account for such a correlation (McGilchrist, 1994). Models that have both fixed effects and random effects are called mixed-effects models (Fox, 2002). The linear mixed-effects model is defined as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \epsilon. \tag{1.5}$$

In this model, $\mathbf{y}$ is a response variable, $\mathbf{X}$ is a matrix of predictor variables, β is a vector of fixed effect coefficients, $\mathbf{Z}$ is a known matrix, γ is a vector of random effect coefficients, and ϵ is a vector of error terms. The γ vector and ϵ vector are distributed independently with means $\mathbf{0}$. The covariance matrices of γ and ϵ are D and $\sigma^2 I$, respectively (Prasad and Rao, 1990).

Equivalently, we can define the linear mixed model as:

$$y_{ij} = x_{ij}^T\beta + \gamma_i + \epsilon_{ij}, \qquad j = 1 \ldots n_i, i = 1 \ldots I, \tag{1.6}$$

where y_{ij} is the response variable for the jth member in the ith family, $x_{ij} = (x_{ij1}, \ldots, x_{ijk})^T$ is a $k \times 1$ vector of corresponding explanatory variables, $\beta = (\beta_1, \ldots, \beta_k)^T$ is a $k \times 1$ vector of unknown parameters, and n_i is the number of members in the ith family.

The random effects γ_i are assumed to be independent $N(0, \sigma_\gamma^2)$ and the error terms ϵ_{ij} are assumed to be independent $N(0, \sigma^2)$. The random effects and error terms are independent of each other.

Finally, if the response variable is binary and grouped in clusters that are correlated, we need to use another model called a generalized linear mixed effects model (Sutradhar and Rao, 2001). The generalized linear mixed effects model is defined as:

$$y_{ij}|\gamma_i \sim \text{Binomial } (1, \pi_{ij}). \tag{1.7}$$

where

$$\pi_{ij} = P(y_{ij} = 1|\gamma_i) = \frac{e^{x_{ij}^T \beta + \sigma_\gamma \gamma_i}}{1 + e^{x_{ij}^T \beta + \sigma_\gamma \gamma_i}}, \qquad i = 1 \cdots I, j = 1 \cdots n_i.$$

In this model, x_{ij}^T and β are as defined previously, and the random effects γ_i are assumed to be independent $N(0, 1)$.

When using a linear mixed effects model like (1.6), the estimation of the variance of the random effects (σ_γ^2) may not be difficult since the estimators can often be written in a closed form. These formulas become more complicated for a generalized linear model.

1.2 Moment Method

A moment method for estimating the variance of random effects for a linear mixed model was proposed by Prasad and Rao (1990). This method uses the general theory of Henderson (1975) for a mixed linear model. That is, a two-stage estimator (or predictor) of a small-area mean under each model is obtained and then the variance components in the estimator are replaced with their estimators. The small area mean is obtained by first deriving the best linear unbiased estimator (or predictor) assuming that the variance components that determine the variance-covariance matrix are known.

For the linear mixed model (1.6), Prasad and Rao (1990) present unbiased quadratic estimators of σ^2 and σ_γ^2 as the following:

- To solve for β:

$$\hat{\beta} = \left(\sum_{i=1}^{I} x_i' \Sigma_i^{-1} x_i\right)^{-1} \left(\sum_{i=1}^{I} x_i' \Sigma_i^{-1} y_i\right). \tag{1.8}$$

  where x_i and y_i are as defined previously, and

$$\Sigma_i^{-1} = \begin{bmatrix} \hat{\sigma}^2 + \hat{\sigma}_\gamma^2 & \hat{\sigma}_\gamma^2 & \cdots & \hat{\sigma}_\gamma^2 \\ \hat{\sigma}_\gamma^2 & \hat{\sigma}^2 + \hat{\sigma}_\gamma^2 & \cdots & \hat{\sigma}_\gamma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_\gamma^2 & \hat{\sigma}_\gamma^2 & \cdots & \hat{\sigma}^2 + \hat{\sigma}_\gamma^2 \end{bmatrix}^{-1}.$$

- To solve for σ^2:

$$\hat{\sigma}^2 = \frac{1}{nI - I - (p-1) + \lambda} \sum_{i=1}^{I} \sum_{j=1}^{n} \hat{e}_{ij}^2, \tag{1.9}$$

  where p is the number of parameters in β. We use $\lambda = 0$ if the model has no intercept term and $\lambda = 1$ otherwise. The $\hat{e}_{ij}$'s are the residuals from the ordinary least squares regression of $y_{ij} - \bar{y}_i$ on $\{x_{ij1} - \bar{x}_{i.1} \ldots x_{ij(p-1)} - \bar{x}_{i.(p-1)}\}$, where $\bar{y}_i = \frac{\sum_{j=1}^{n} y_{ij}}{n}$, and $\bar{x}_{i.1} = \frac{\sum_{j=1}^{n} x_{ij1}}{n}$.

- To solve for σ_γ^2:

$$\hat{\sigma}_\gamma^2 = \frac{1}{n_*}\left[\sum_{i=1}^{I} \sum_{j=1}^{n} \hat{u}_{ij}^2 - (nI - (p-1))\hat{\sigma}^2\right] \tag{1.10}$$

where

$$n_* = n - tr\left[(X'X)^{-1}\sum_{i=1}^{I} n^2 \bar{x}_i \bar{x}_i'\right], X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}, \bar{x}_i = \begin{bmatrix} x_{i.1} \\ x_{i.2} \\ \vdots \\ \bar{x}_{i.(p-1)} \end{bmatrix},$$

and the $\hat{u}_{ij}$'s are the residuals from the ordinary least squares regression of y_{ij} on $x_{ij1} \cdots x_{ij(p-1)}$.

For cases in which data is non-Gaussian and correlated, it is computationally difficult to make inferences. Jiang (1998) discusses a method to find estimators that are both computationally feasible and consistent for a generalized linear mixed model. The method is based on simulated moments.

For the model given in (1.7), Jiang (1998) presents the following joint moment equation to solve for β and σ_γ:

$$w - E(w) = 0, \tag{1.11}$$

where $w = (W_1, W_2)^T$ and W_1 and W_2 are defined as follows:

$$\begin{aligned} W_1 &= \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} y_{ij}, \\ W_2 &= \sum_{i=1}^{m}\sum_{j \neq k} y_{ij} y_{ik}. \end{aligned}$$

We cannot solve (1.11) explicitly for β or σ_γ, therefore we need to estimate these parameters by solving (1.11) with Newton's method. Initial estimates were chosen and used to start a Newton-Raphson iteration. Suppose $\hat{\beta}_M$, $\hat{\sigma}_{\gamma,M}$ denote solutions to the above equation. Then, at iteration $(r+1)$:

$$\begin{bmatrix} \hat{\beta}_{M(r+1)} \\ \hat{\sigma}_{\gamma,M(r+1)} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_{M(r)} \\ \hat{\sigma}_{\gamma,M(r)} \end{bmatrix} + (P^T)_r^{-1}(w - E(w))_r.$$

where

$$P^T = \begin{bmatrix} \frac{\partial E(W_1)}{\partial \beta^T} & \frac{\partial E(W_2)}{\partial \beta^T} \\ \frac{\partial E(W_1)}{\partial \sigma_\gamma} & \frac{\partial E(W_2)}{\partial \sigma_\gamma} \end{bmatrix},$$

w is defined as above, and $E(w) = (E(W_1), E(W_2))^T$.

As some of these expectations are very difficult to find, they can be approximated.

1.3 Quasi-Likelihood Method

The quasi-likelihood method of estimation, unlike the maximum likelihood approach, does not require specification of the distribution of the response variable (Ramsey and Schafer, 1997).

The general quasi-likelihood equation that can be used for both the linear and generalized linear model to estimate β and the variance components is as follows (Sutradhar, 2004):

$$\sum_{i=1}^{I} \frac{\partial M_i^T(\theta)}{\partial \theta} V_i^{-1}(S_i - M_i(\theta)) = 0. \tag{1.12}$$

where $S_i = (y_i^T, u_i^T)$, with $y_i = (y_{i1}, \cdots, y_{in})$, $u_i = (u_{i1}^T, u_{i2}^T)^T$, $u_{i1} = (y_{i1}^2, \cdots, y_{in}^2)^T$, and $u_{i2} = (y_{i1}y_{i2}, \cdots, y_{i,n-1}y_{i,n})^T$. Also $M_i(\theta) = E(S_i)$ and $V_i(\theta) = cov(S_i)$ where $\theta = (\beta^T, \sigma)^T$.

1.4 Other Approaches

When using the logistic regression model, the estimation of β and σ_z is more difficult. This is because there are no closed mathematical forms for the estimates. There are many approaches suggested in the literature to conduct this estimation.

Schall (1991) discussed the estimation of random effects in a generalized linear model. He presented an algorithm for estimating in a generalized linear mixed model the fixed effects, random effects, and components of the dispersion. He discussed various conditions under which his method yielded approximate maximum or quasi maximum likelihood estimates of the fixed effects and dispersion components as well as approximate empirical Bayes estimates of the random effects.

In the hierarchical model, Breslow and Clayton (1993) concluded that the PQL (penalized quasi-likeilhood) method of estimation for the parameters and random effects is useful. The PQL method of estimation, when applied to clustered binary data, underestimates the variance components and fixed effects. The method does improve in situations in which the binomial observations have denominations greater than one.

Breslow and Lin (1995) derived formulas for the asymptotic biases of regression coefficients and variance components, for small variance component values, using three estimators. These were estimated in generalized linear mixed models with canonical link function and a group of random effects by using the first and second order Laplace expansions of the integrated likelihood as well as using the PQL method of estimation. The PQL and first order Laplace expansion produced biased estimates, especially when used on data that are binary and correlated. A corrected PQL and the second order Laplace expansion produce good estimators for variance components that are small and very good for those that are large.

Kuk (1995) proposed a method of adjusting initially defined estimates by an iterative bias correction to produce estimates that are asymptotically unbiased and consistent. This method can be applied to any parametric model and the estimates produced are almost unbiased with the standard errors only somewhat inflated.

1.5 Outline of Thesis

In this thesis we have evaluated mixed models using a moment method, quasi-likelihood and a simulation approach that will be discussed in Chapter 2. Chapter 2 deals with the evaluation of these approaches for a linear mixed model. In Chapter 3 we evaluate these methods of estimation for the logistic model. Finally, in Chapter 4 we apply these methods to a clustered data set. This data set contains information on 180 people from 48 families. The information obtained includes the number of times they visited a physician each year over the years 1985 - 1990, their age, their gender, the number of chronic conditions they had, and their education level.

Chapter 2

Simulation Data Analysis - Linear Mixed Model

2.1 Introduction

For this chapter, all analyses used a linear mixed model with the aim of estimating β, σ^2 and σ_γ^2. We do this first using the moment method, then with a simulation method which treats random effects as fixed effects, and finally with the quasi-likelihood method.

As a reminder, the linear mixed model that we are using for this chapter is the following:

$$y = X\beta + Z\gamma + \epsilon, \tag{2.1}$$

where all terms are as previously defined in Chapter 1.

All simulations used the following assumptions and parameters:

- $i = 1, \ldots, I = 100$.
- $j = 1, \ldots, n_i = n = 4$.
- $\beta = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.
- $X_i = \begin{cases} \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix} & \text{if } i = 1, \cdots, 50; \\ \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} & \text{if } i = 51, \cdots, 100. \end{cases}$

The methods of estimation were studied with all 16 combinations of the following σ^2 and σ_γ^2 values: $\sigma^2 = (0.1, 1, 2, 4)$ and $\sigma_\gamma^2 = (0.1, 1, 2, 4)$. Five further simulations were conducted under the following situations:

- $\beta_1 = 1.0$, $\beta_2 = 0.1$, $\sigma^2 = 0.1$, $\sigma_\gamma^2 = 0.1$.
- $\beta_1 = 1.0$, $\beta_2 = 0.1$, $\sigma^2 = 0.1$, $\sigma_\gamma^2 = 1.0$.
- $\beta_1 = 1.0$, $\beta_2 = 5.0$, $\sigma^2 = 0.1$, $\sigma_\gamma^2 = 0.1$.
- $\beta_1 = 1.0$, $\beta_2 = 5.0$, $\sigma^2 = 0.1$, $\sigma_\gamma^2 = 1.0$.
- $\beta_1 = 1.0$, $\beta_2 = 2.0$, $\sigma^2 = 0.01$, $\sigma_\gamma^2 = 0.01$.

There were 500 simulated datasets used in each simulation.

2.2 Moment Method Analysis

We begin with a discussion of the moment estimates of β, σ^2 and σ_ζ^2. This method was discussed in Section 1.2.

Table 2.1 shows the estimated values for β_1, β_2, σ^2 and σ_ζ^2 for all of the simulations conducted. Table 2.2 shows the variances of the estimated values for β_1, β_2, σ^2 and σ_ζ^2.

This moment method performed well as the estimates of β_1, β_2, σ^2 and σ_ζ^2 appear unbiased. The evidence is that all of the estimated values are close to the true values for all parameters in all simulations, which can be seen in Table 2.1. Also, all of the variances of these estimates are very small, which can be seen in Table 2.2. Graphical evidence of the skewness of the estimates can be seen when they are plotted. Refer to Figure 2.1 for an example. This figure shows the histograms of $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\sigma}^2$ and $\hat{\sigma}_\zeta^2$ for simulation 1. As we can see all of the estimates appear to follow a normal distribution approximately. Graphs of the estimates for the other simulations are not shown as they are similar to those for simulation 1.

2.3 Simulated Method Analysis

We wish to investigate if one can treat the random effects as fixed effects (in some sense) and use this assumption to estimate σ^2 and σ_ζ^2 in the same manner as β. This was implemented in four different ways. Each of these estimation procedures will be described in more detail in the following four sections.

Figure 2.1: Moment Method Histograms (LMM): Simulation 1

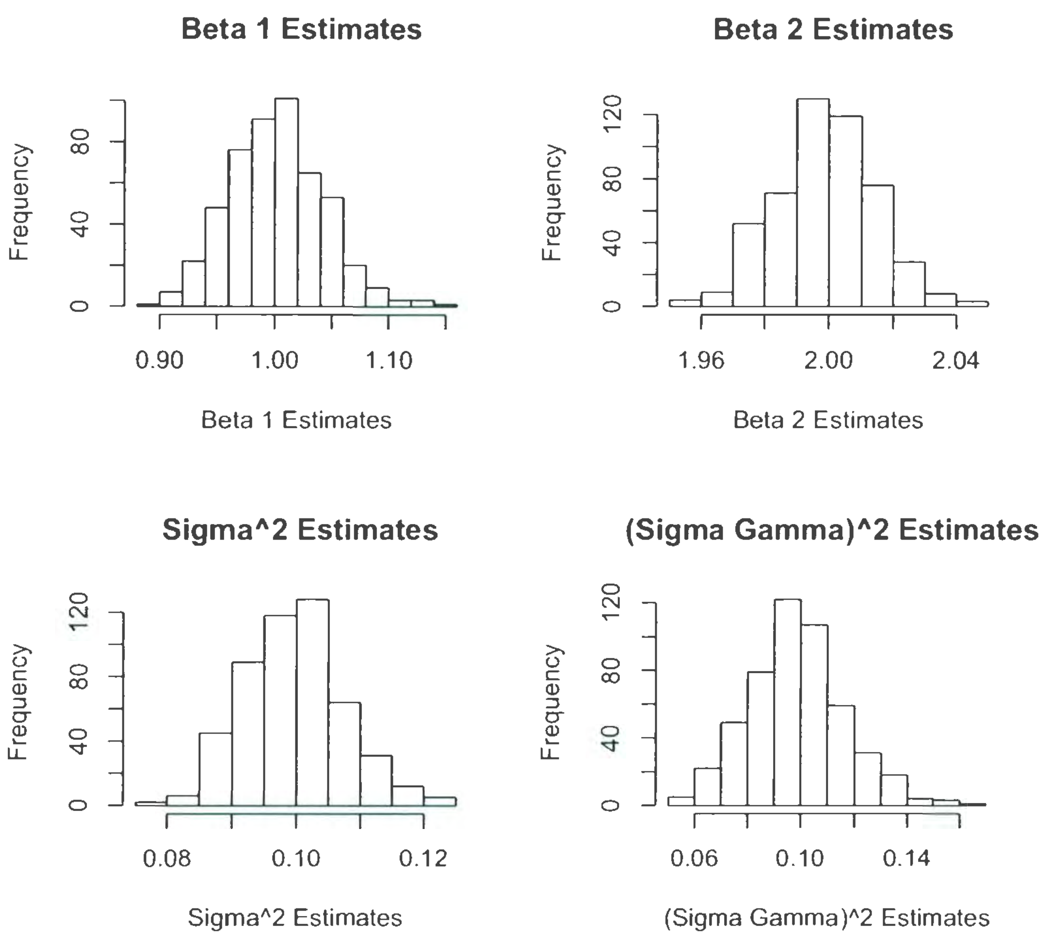


Table 2.1: Moment Method Estimates (LMM)

Sim	β_1	β_2	σ^2	σ_γ^2	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$	$\hat{\sigma}_\gamma^2$
1	1.0	2.0	0.1	0.1	1.00186	1.99858	0.09970	0.09850
2	1.0	2.0	0.1	1.0	1.00249	1.99789	0.09970	0.98641
3	1.0	2.0	0.1	2.0	1.00236	1.99787	0.09970	1.97322
4	1.0	2.0	0.1	4.0	1.00215	1.99788	0.09970	3.94701
5	1.0	2.0	1.0	0.1	1.00140	1.99928	0.99697	0.09812
6	1.0	2.0	1.0	1.0	1.00588	1.99550	0.99697	0.98496
7	1.0	2.0	1.0	2.0	1.00711	1.99111	0.99697	1.97113
8	1.0	2.0	1.0	4.0	1.00778	1.99371	0.99697	3.94102
9	1.0	2.0	2.0	0.1	1.00098	1.99982	1.99393	0.09796
10	1.0	2.0	2.0	1.0	1.00614	1.99550	1.99393	0.98444
11	1.0	2.0	2.0	2.0	1.00832	1.99364	1.99393	1.96992
12	1.0	2.0	2.0	1.0	1.01005	1.99209	1.99393	3.94225
13	1.0	2.0	4.0	0.1	1.00048	2.00050	3.98786	0.09781
14	1.0	2.0	1.0	1.0	1.00571	1.99613	3.98786	0.98307
15	1.0	2.0	4.0	2.0	1.00868	1.99364	3.98786	1.96829
16	1.0	2.0	4.0	1.0	1.01176	1.99101	3.98786	3.93984
17	1.0	0.1	0.1	0.1	1.00186	0.09858	0.09970	0.09850
18	1.0	0.1	0.1	1.0	1.00249	0.09789	0.09970	0.98641
19	1.0	5.0	0.1	0.1	1.00186	4.99858	0.09970	0.09850
20	1.0	5.0	0.1	1.0	1.00249	4.99789	0.09970	0.98641
21	1.0	2.0	0.01	0.01	1.00059	1.99955	0.00997	0.00985

2.3.1 Simulation Method 1

For the first method, data was simulated using the model described by (1.6). This method estimates σ and σ_γ by treating them the same as β. Therefore the x matrix will be changed to include the x values along with the γ_i and ϵ_{ij} terms. The x matrix and y vector have the following forms:

Table 2.2: Moment Method Variances (LMM)

Sim	β_1	β_2	σ^2	σ_γ^2	$Var(\hat\beta_1)$	$Var(\hat\beta_2)$	$Var(\hat\sigma^2)$	$Var(\hat\sigma_\gamma^2)$
1	1.0	2.0	0.1	0.1	0.00169	0.000238	0.000060	0.000322
2	1.0	2.0	0.1	1.0	0.01098	0.000335	0.000060	0.021750
3	1.0	2.0	0.1	2.0	0.02111	0.000347	0.000060	0.085025
4	1.0	2.0	0.1	4.0	0.04130	0.000355	0.000060	0.336198
5	1.0	2.0	1.0	0.1	0.00603	0.001392	0.006045	0.002912
6	1.0	2.0	1.0	1.0	0.01692	0.002385	0.006045	0.032192
7	1.0	2.0	1.0	2.0	0.02770	0.002765	0.006045	0.104114
8	1.0	2.0	1.0	4.0	0.04853	0.003075	0.006045	0.372568
9	1.0	2.0	2.0	0.1	0.01054	0.002554	0.024180	0.009024
10	1.0	2.0	2.0	1.0	0.02239	0.003993	0.024180	0.047177
11	1.0	2.0	2.0	2.0	0.03384	0.004770	0.024180	0.128769
12	1.0	2.0	2.0	4.0	0.05540	0.005530	0.024180	0.446457
13	1.0	2.0	4.0	0.1	0.01943	0.004860	0.096720	0.034372
14	1.0	2.0	4.0	1.0	0.03235	0.006685	0.096720	0.087597
15	1.0	2.0	4.0	2.0	0.04178	0.007985	0.096720	0.188709
16	1.0	2.0	4.0	4.0	0.06768	0.009539	0.096720	0.515077
17	1.0	0.1	0.1	0.1	0.00169	0.000238	0.000060	0.000322
18	1.0	0.1	0.1	1.0	0.01098	0.000335	0.000060	0.021750
19	1.0	5.0	0.1	0.1	0.00169	0.000238	0.000060	0.000322
20	1.0	5.0	0.1	1.0	0.01098	0.000335	0.000060	0.021750
21	1.0	2.0	0.01	0.01	0.00017	0.000024	0.000001	0.000003

$$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in} \end{pmatrix} \quad \text{and } x_i^* = \begin{pmatrix} x_{i11} & \cdots & x_{i1p} & \gamma_i & \epsilon_{i1} \\ \vdots & & \vdots & \vdots & \\ x_{in1} & \cdots & x_{inp} & \gamma_i & \epsilon_{in} \end{pmatrix}. \tag{2.2}$$

where γ_i is generated from $N(0, \sigma_\gamma^2)$ and ϵ_{ij} is generated from $N(0, \sigma^2)$.

To estimate σ, σ_γ and β, β^* is defined as follows:

$$\beta^* = \begin{pmatrix} \beta \\ \sigma_\zeta \\ \sigma \end{pmatrix}.$$

We use (1.8) to estimate β^*, which gives an estimate for β, σ and σ_ζ:

$$\hat{\beta}^* = \left(\sum_{i=1}^{I} x_i^{*'} x_i^*\right)^{-1} \left(\sum_{i=1}^{I} x_i^{*'} y_i\right). \tag{2.3}$$

The previous simulations were repeated using this procedure. Table 2.3 shows the estimated values for β_1, β_2, σ^2 and σ_ζ^2. Table 2.4 shows the variances of the estimated values for β_1, β_2, σ^2 and σ_ζ^2.

This method performed well in giving unbiased estimates for β_1 and β_2 as can be seen in Table 2.3. Also, the variances of all of these estimates are all very small, which can be seen in Table 2.4. Graphical evidence of the skewness of the estimates can be seen when they are plotted. Refer to Figure 2.2 for an example. The top two histograms in this figure are of $\hat{\beta}_1$ and $\hat{\beta}_2$ for simulation 1. As we can see, both of these plots follow a normal distribution approximately. Graphs of all of the other β estimates are not shown as they are similar to those for simulation 1.

This method of estimation did not perform well for σ^2 and σ_ζ^2. The estimates for these parameters were all very close to zero and not close to the original values, which can be seen in Table 2.3. Also, the variances were all small, which can be seen in Table 2.4. Graphical evidence of the skewness of the estimates can be seen when they are plotted. Refer to Figure 2.2 for an example. The bottom two histograms in this figure are of $\hat{\sigma}^2$ and $\hat{\sigma}_\zeta^2$ for simulation 1. As we can see, both of these plots are skew to the right. Graphs of all of the other σ^2 and σ_ζ^2 estimates are not shown as they are similar to those for simulation 1.

Figure 2.2: Simulated Method 1 Histograms (LMM): Simulation 1

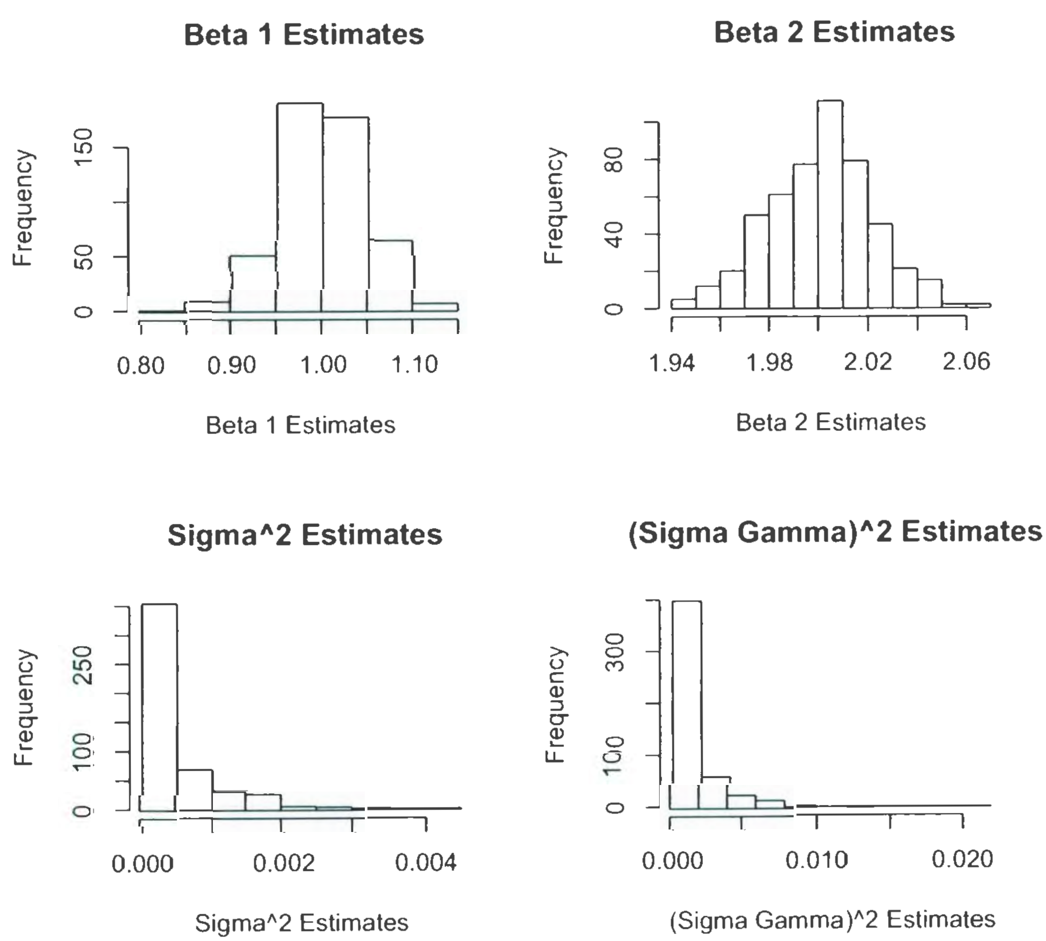


Table 2.3: Simulated Method 1 Estimates (LMM)

Sim	β_1	β_2	σ^2	σ_z^2	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$	$\hat{\sigma}_z^2$
1	1.0	2.0	0.1	0.1	0.999832	2.000971	1.0700e-06	3.9871e-06
2	1.0	2.0	0.1	1.0	1.000122	2.002091	1.1634e05	2.5149e-05
3	1.0	2.0	0.1	2.0	1.000298	2.002773	2.3526e-05	4.6827e-05
4	1.0	2.0	0.1	4.0	1.000546	2.003732	1.7420e-05	8.8898e-05
5	1.0	2.0	1.0	0.1	0.999178	2.001960	8.0055e-07	1.0866e-05
6	1.0	2.0	1.0	1.0	0.999468	2.003080	1.0700e-05	3.9871e-05
7	1.0	2.0	1.0	2.0	0.999641	2.003759	2.2190e-05	6.6302e-05
8	1.0	2.0	1.0	4.0	0.999893	2.004718	4.5515e-05	0.00011509
9	1.0	2.0	2.0	0.1	0.998782	2.002557	6.5630e-07	1.6675e-05
10	1.0	2.0	2.0	1.0	0.999072	2.003677	1.0154e-05	5.0432e-05
11	1.0	2.0	2.0	2.0	0.999248	2.004356	2.1400e-05	7.9742e-005
12	1.0	2.0	2.0	4.0	0.999497	2.005315	1.4381e-05	0.00013260
13	1.0	2.0	4.0	0.1	0.998218	2.003402	4.7673e-07	2.7007e-05
14	1.0	2.0	1.0	1.0	0.998512	2.004522	9.0546e-06	6.7484e-05
15	1.0	2.0	4.0	2.0	0.998688	2.005200	2.0308e-05	0.00010086
16	1.0	2.0	4.0	4.0	0.998936	2.006160	4.2801e-05	0.00015948
17	1.0	0.1	0.1	0.1	0.999832	0.100974	1.0700e-06	3.9871e-06
18	1.0	0.1	0.1	1.0	1.000122	0.102094	1.1634e-05	2.5148e-05
19	1.0	5.0	0.1	0.1	0.999832	5.000974	1.0700e-06	3.9871e-06
20	1.0	5.0	0.1	1.0	1.000122	5.002094	1.16341e-05	2.5148e-05
21	1.0	2.0	0.01	0.01	0.999947	2.000308	1.0700e-07	3.9871e-07

2.3.2 Simulation Method 2

For the second method, data was simulated without any random effects. That is, the following linear model was used:

$$y_{ij} = x_{ij}^T\beta + \epsilon_{ij}, \qquad i = 1\cdots I, j = 1\cdots n_i. \tag{2.4}$$

where all terms are defined as in equation (1.6).

Table 2.4: Simulated Method 1 Variances (LMM)

Sim	β_1	β_2	σ^2	σ_γ^2	$Var(\hat\beta_1)$	$Var(\hat\beta_2)$	$Var(\hat\sigma^2)$	$Var(\hat\sigma_\gamma^2)$
1	1.0	2.0	0.1	0.1	0.002099	0.000456	1.5255e-07	3.7999e-06
2	1.0	2.0	0.1	1.0	0.017033	0.003594	1.3458e-05	0.00025528
3	1.0	2.0	0.1	2.0	0.033574	0.007083	4.9539e-05	0.00099333
4	1.0	2.0	0.1	4.0	0.066618	0.014065	0.00019065	0.00391242
5	1.0	2.0	1.0	0.1	0.005825	0.001414	1.4790e-05	3.0003e-05
6	1.0	2.0	1.0	1.0	0.020987	0.004565	4.5255e-05	0.00037999
7	1.0	2.0	1.0	2.0	0.037665	0.008044	9.9913e-05	0.00123302
8	1.0	2.0	1.0	4.0	0.070904	0.015012	0.00027594	0.00440020
9	1.0	2.0	2.0	0.1	0.009913	0.002546	5.4570e-05	0.00008693
10	1.0	2.0	2.0	1.0	0.025212	0.005656	0.00010458	0.00054755
11	1.0	2.0	2.0	2.0	0.075331	0.016089	0.00039965	0.00493209
12	1.0	2.0	2.0	4.0	0.075331	0.016089	0.00039965	0.00493209
13	1.0	2.0	4.0	0.1	0.018050	0.004752	0.00020974	0.00028482
14	1.0	2.0	4.0	1.0	0.033544	0.007848	0.00029804	0.00097835
15	1.0	2.0	4.0	2.0	0.050424	0.011313	0.00041831	0.00219022
16	1.0	2.0	4.0	4.0	0.083948	0.018260	0.00072407	0.00607989
17	1.0	0.1	0.1	0.1	0.002099	0.000457	4.5255e-07	3.7999e-06
18	1.0	0.1	0.1	1.0	0.017033	0.003594	1.3458e-05	0.00025528
19	1.0	5.0	0.1	0.1	0.002099	0.000456	1.5255e-07	3.7999e-06
20	1.0	5.0	0.1	1.0	0.017033	0.003594	1.3458e-05	0.00025528
21	1.0	2.0	0.01	0.01	0.000210	4.5649e-05	4.5255e-09	3.7999e-08

Following Simulation Method 1, we define y_i and x_i as:

$$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in} \end{pmatrix} \quad \text{and } x_i^* = \begin{pmatrix} x_{i11} & \cdots & x_{i1p} & e_{i1} \\ \vdots & & \vdots & \\ x_{in1} & \cdots & x_{inp} & e_{in} \end{pmatrix}. \qquad (2.5)$$

where e_{ij} is generated from $N(0, \sigma^2)$.

To estimate σ and β, β^* was defined as follows:

$$\beta^* = \begin{pmatrix} \beta \\ \sigma \end{pmatrix}.$$

We use (1.8) to estimate β^*, which gives an estimate for β and σ:

$$\hat{\beta}^* = \left(\sum_{i=1}^{I} x_i^{*'} x_i^*\right)^{-1} \left(\sum_{i=1}^{I} x_i^{*'} y_i\right). \tag{2.6}$$

The previous simulations were redone (without any random effects) using this procedure. Table 2.5 shows the estimated values for β_1, β_2 and σ^2. Table 2.6 shows the variances of the estimated values for β_1, β_2 and σ^2.

This method performed well in giving unbiased estimates for β_1 and β_2, as can be seen in Table 2.5. Also, the variances of all of these estimates are very small, which can be seen in Table 2.6. Graphical evidence of the skewness of the estimates can be seen when they are plotted. Refer to Figure 2.3 for an example. The top two histograms in this figure are of $\hat{\beta}_1$ and $\hat{\beta}_2$ for simulation 1. As we can see, both of these plots follow a normal distribution approximately. Graphs of all of the other β estimates are not shown as they are similar to those for simulation 1.

This method of estimation did not perform well in estimating σ^2. These estimates were all very close to zero and not close to the original values, which can be seen in Table 2.5. Also, the variances were all small, which can be seen in Table 2.6. Graphical evidence of the skewness of the estimates can be seen when they are plotted. Refer to Figure 2.3 for an example. The bottom histogram in this figure is of σ^2 for simulation 1. As we can see, both of these plots are skew to the right. Graphs of all of the other σ^2 estimates are not shown as they are similar to those for simulation 1.

Figure 2.3: Simulated Method 2 Histograms (LMM): Simulation 1

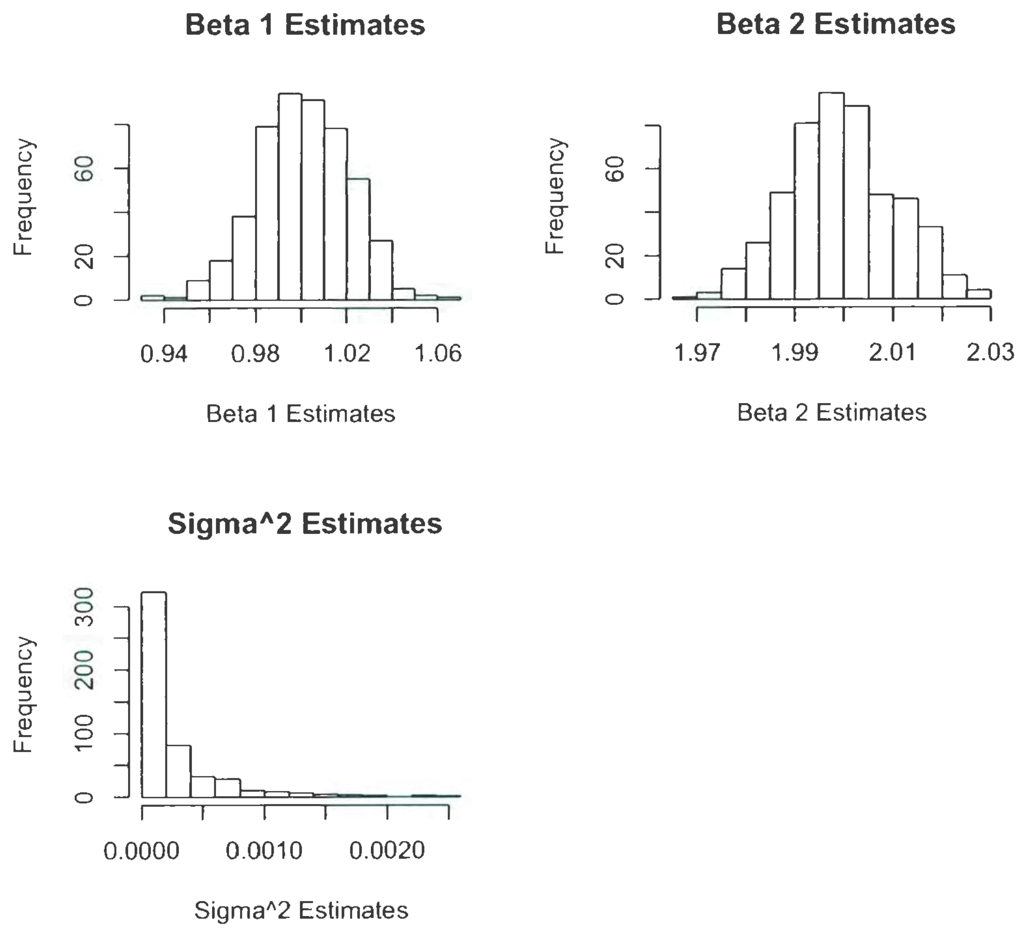


Table 2.5: Simulated Method 2 Estimates (LMM)

Sim	β_1	β_2	σ^2	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
1	1.0	2.0	0.1	1.000919	1.999491	6.471533e-07
2	1.0	2.0	1.0	1.002907	1.998392	6.471533e-06
3	1.0	2.0	2.0	1.004111	1.997725	1.294907e-05
4	1.0	2.0	4.0	1.005814	1.996783	2.589813e-05
5	1.0	0.1	0.1	1.000919	0.099491	6.471533e-07
6	1.0	5.0	0.1	1.000919	1.999491	6.471533e-07
7	1.0	2.0	0.01	1.000291	1.999839	6.471533e-08

Table 2.6: Simulated Method 2 Variances (LMM)

Sim	β_1	β_2	σ^2	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$	$Var(\hat{\sigma}^2)$
1	1.0	2.0	0.1	0.000400538	0.0001172487	1.351570e-07
2	1.0	2.0	1.0	0.00400538	0.001172497	1.351570e-05
3	1.0	2.0	2.0	0.00801076	0.002344994	5.406278e-05
4	1.0	2.0	4.0	0.01602152	0.004689987	0.0002162511
5	1.0	0.1	0.1	0.000400538	0.0001172497	1.351570e-07
6	1.0	5.0	0.1	0.000400538	0.0001172497	1.351570e-07
7	1.0	2.0	0.01	0.0000400538	1.172497e-05	1.351570e-09

2.3.3 Simulation Method 3

For the third method, data was simulated using the model described by (1.6). This method estimates σ_ϵ by treating it the same as β.

Following Simulation Method 1, we find the y_i and x_i matrices as:

$$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in} \end{pmatrix} \quad \text{and } x_i^* = \begin{pmatrix} x_{i11} & \cdots & x_{i1p} & \gamma_i \\ \vdots & & \vdots & \\ x_{in1} & \cdots & x_{inp} & \gamma_i \end{pmatrix}, \tag{2.7}$$

where γ_i is generated from $N(0, \sigma_\gamma^2)$.

To estimate σ, σ_γ and β, β^* was defined as follows:

$$\beta^* = \begin{pmatrix} \beta \\ \sigma_\gamma \end{pmatrix}.$$

We use (1.8) to estimate β^*, which gives an estimate for β and σ_γ:

$$\hat{\beta}^* = \left(\sum_{i=1}^{I} x_i^{*\prime} x_i^*\right)^{-1} \left(\sum_{i=1}^{I} x_i^{*\prime} y_i\right). \tag{2.8}$$

Then, to estimate σ^2, we use the following equation:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{I} \left(y_i - x_i^* \hat{\beta}^*\right)^{\prime} \left(y_i - x_i^* \hat{\beta}^*\right)}{nI}. \tag{2.9}$$

where $\hat{\beta}^*$ is defined in (2.8).

The previous simulations were redone using this procedure. Table 2.7 shows the estimated values for β_1, β_2, σ^2 and σ_γ^2. Table 2.8 shows the variances of the estimated values for β_1, β_2, σ^2 and σ_γ^2.

This method performed well in giving unbiased estimates for β_1 and β_2 as can be seen in Table 2.7. Also, the variances of all of these estimates are all very small, which can be seen in Table 2.8. Graphical evidence of the skewness of the estimates can be seen when they are plotted. Refer to Figure 2.1 for an example. The top two histograms in this figure are of $\hat{\beta}_1$ and $\hat{\beta}_2$ for simulation 1. As we can see, both of these plots follow a normal distribution approximately. Graphs of all of the other β estimates are not shown as they are similar to those for simulation 1.

This method of estimation did not perform well for σ^2, but it did do better than the previous two simulation methods. Most of the estimates are not close to their original values. However, some simulations performed well.

The cases with better estimates of σ^2 are found to be the cases where σ^2 is large relative to σ_γ^2. All variances of the simulated estimates can be see in Table 2.8. Graphical evidence of the skewness of the estimates can be seen when they are plotted. Refer to Figure 2.1 for an example. The bottom left histogram in this figure is of σ^2 for simulation 1. As we can see, this plot follows a normal distribution approximately and thus is not skew. Graphs of all of the other σ^2 estimates are not shown, as they are similar to those for simulation 1.

This method of estimation did not perform well for estimating σ_γ^2. The estimates were all very close to zero and not close to the original values, which can be seen in Table 2.7. Also, the variances were all small, which can be seen in Table 2.8. Graphical evidence of the skewness of the estimates can be seen with plots of the estimates. An example of one such plot is in Figure 2.1. The bottom right histogram in this figure is of $\hat{\sigma}_\gamma^2$ for simulation 1. As we can see, this plot is skew to the right. Graphs of all of the other σ_γ^2 estimates are not shown, as they are similar to those for simulation 1.

Figure 2.4: Simulated Method 3 Histograms (LMM): Simulation 1

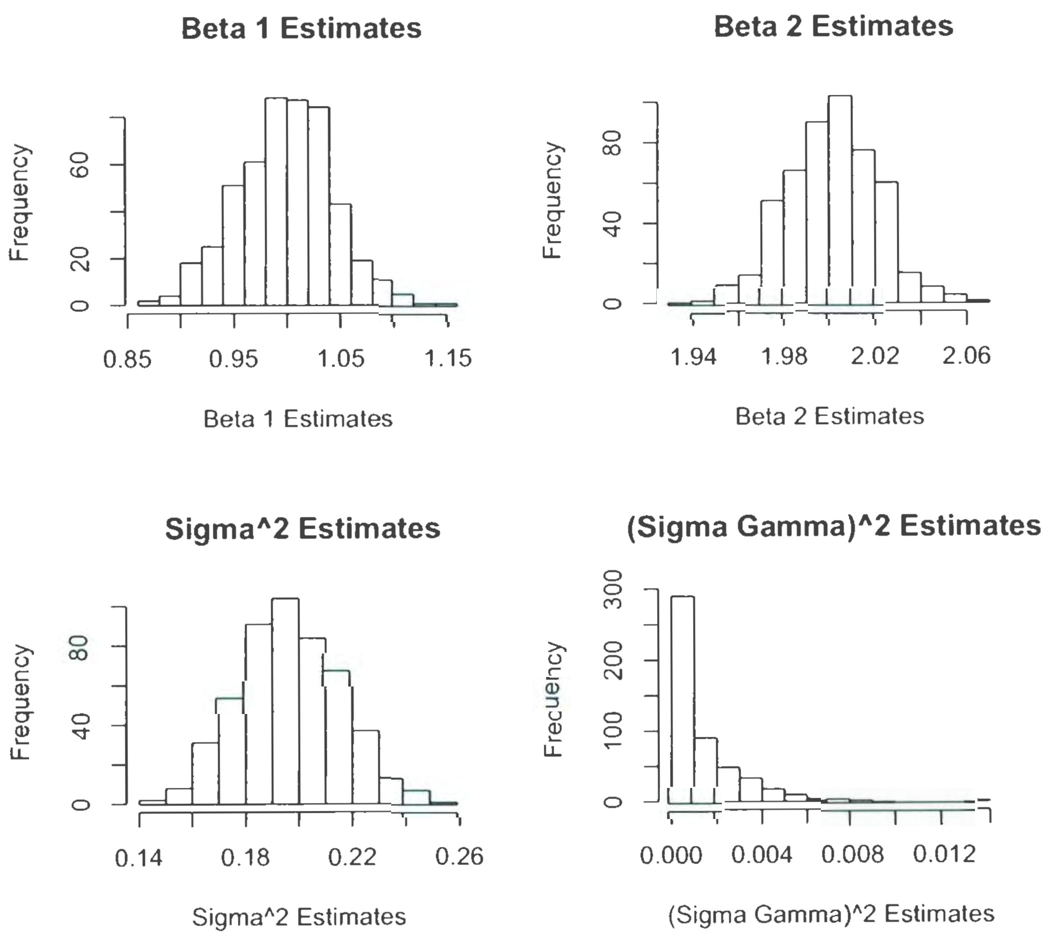


Table 2.7: Simulated Method 3 Estimates (LMM)

Sim	β_1	β_2	σ^2	σ_γ^2	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$	$\hat{\sigma}_\gamma^2$
1	1.0	2.0	0.1	0.1	0.992053	2.000934	0.196695	1.141692e-06
2	1.0	2.0	0.1	1.0	0.999174	2.001832	1.073182	2.148716e-05
3	1.0	2.0	0.1	2.0	0.998904	2.002377	2.047069	1.618845e-05
4	1.0	2.0	0.1	1.0	0.998743	2.003147	3.994852	9.7064e-05
5	1.0	2.0	1.0	0.1	0.997675	2.002053	1.090513	3.535568e-08
6	1.0	2.0	1.0	1.0	0.997487	2.002952	1.966949	1.141692e-05
7	1.0	2.0	1.0	2.0	0.997373	2.003496	2.940805	3.068778e-05
8	1.0	2.0	1.0	1.0	0.997212	2.004266	1.888544	7.388147e-05
9	1.0	2.0	2.0	0.1	0.996748	2.0027315	2.083656	9.009974e-07
10	1.0	2.0	2.0	1.0	0.996560	2.003630	2.960062	6.852437e-06
11	1.0	2.0	2.0	2.0	0.996446	2.004175	3.933899	2.283385e-05
12	1.0	2.0	2.0	1.0	0.996285	2.004945	5.881611	6.137556e-05
13	1.0	2.0	1.0	0.1	0.995437	2.003691	4.069950	1.103366e-06
14	1.0	2.0	4.0	1.0	0.995249	2.004590	1.946312	2.375446e-06
15	1.0	2.0	4.0	2.0	0.995135	2.005134	5.920123	1.370487e-05
16	1.0	2.0	4.0	4.0	0.994974	2.005904	7.867798	1.566769e-05
17	1.0	0.1	0.1	0.1	0.999205	0.100934	0.196695	1.141692e-06
18	1.0	0.1	0.1	1.0	0.999017	0.101832	1.073182	2.148716e-05
19	1.0	5.0	0.1	0.1	0.999205	5.000934	0.196695	1.141692e-06
20	1.0	5.0	0.1	1.0	0.999017	5.001832	1.073182	2.14716e-05
21	1.0	2.0	0.01	0.01	0.999755	2.000295	0.019670	1.141692e-07

2.3.4 Simulation Method 4

For the fourth method, data was simulated using the model described by (1.6). This method estimates σ_γ by treating it the same as β, as does the third method. Therefore, the y_i and x_i^* matrices are:

$$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in} \end{pmatrix} \quad \text{and } x_i^* = \begin{pmatrix} x_{i11} & \cdots & x_{i1p} & \gamma_i \\ \vdots & & \vdots & \\ x_{in1} & \cdots & x_{inp} & \gamma_i \end{pmatrix}, \qquad (2.10)$$

Table 2.8: Simulated Method 3 Variances (LMM)

Sim	β_1	β_2	σ^2	σ_γ^2	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$	$Var(\hat{\sigma}^2)$	$Var(\hat{\sigma}_\gamma^2)$
1	1.0	2.0	0.1	0.1	0.0019967	0.00039607	0.00036385	3.026078e-06
2	1.0	2.0	0.1	1.0	0.0156879	0.00301013	0.02169234	0.00018046
3	1.0	2.0	0.1	2.0	0.0308821	0.00591439	0.08410268	0.00070509
4	1.0	2.0	0.1	4.0	0.0612577	0.01172273	0.33088020	0.00280313
5	1.0	2.0	1.0	0.1	0.0061976	0.00134560	0.00583203	3.359525e-05
6	1.0	2.0	1.0	1.0	0.0199670	0.00396072	0.03638510	0.00030261
7	1.0	2.0	1.0	2.0	0.0352086	0.00686563	0.10970750	0.00091293
8	1.0	2.0	1.0	4.0	0.0656114	0.01267488	0.37951830	0.00314263
9	1.0	2.0	2.0	0.1	0.0108470	0.00240039	0.02063819	0.00010511
10	1.0	2.0	2.0	1.0	0.0246638	0.00501615	0.06079727	0.00048423
11	1.0	2.0	2.0	2.0	0.0399340	0.00792145	0.14547410	0.00121043
12	1.0	2.0	2.0	4.0	0.0704172	0.01373125	0.43882990	0.00365173
13	1.0	2.0	4.0	0.1	0.0201330	0.00450979	0.07798875	0.00036802
14	1.0	2.0	4.0	1.0	0.0340168	0.00712646	0.13661470	0.00096740
15	1.0	2.0	4.0	2.0	0.0493276	0.01003231	0.24318910	0.00193691
16	1.0	2.0	4.0	4.0	0.0798681	0.01584289	0.58180620	0.00484173
17	1.0	0.1	0.1	0.1	0.0019967	0.00039607	0.00036369	3.026078e-06
18	1.0	0.1	0.1	1.0	0.0156879	0.00301013	0.02169234	0.00018046
19	1.0	5.0	0.1	0.1	0.0019967	0.00039607	0.00036369	3.026078e-06
20	1.0	5.0	0.1	1.0	0.0156879	0.00301013	0.02169234	0.00018046
21	1.0	2.0	0.01	0.01	0.0001997	3.960723e-05	3.636851e-06	3.026078e-08

where γ_i is generated from $N(0, \sigma_\gamma^2)$.

To estimate σ, σ_γ and β, β^* was defined as follows:

$$\beta^* = \begin{pmatrix} \beta \\ \sigma_\gamma \end{pmatrix}.$$

We use (1.8) to estimate β^*, which gives an estimate for β, and σ_γ.

$$\hat{\beta}^* = \left(\sum_{i=1}^{I} x_i^{*\prime} x_i^*\right)^{-1} \left(\sum_{i=1}^{I} x_i^{*\prime} y_i\right) = \begin{pmatrix} \hat{\beta} \\ \hat{\sigma}_\gamma \end{pmatrix}. \tag{2.11}$$

Let us define a new reduced β vector as the following:

$$\hat{\beta}_r^* = \hat{\beta}.$$

Then, to estimate σ^2, we use the following equation:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{I}\left(y_i - x_i^*\hat{\beta}_r^*\right)'\left(y_i - x_i^*\hat{\beta}_r^*\right)}{nI}. \tag{2.12}$$

The $\hat{\beta}_r$ vector used in (2.12) to estimate σ does not include the estimate of σ_γ obtained, which is the sole difference between estimation method 3 and estimation method 4. Essentially we include σ_γ in β^* so that when we estimate β we are also estimating σ_γ. As a result, $\hat{\beta}^*$ is a vector of the β and σ_γ estimates. We use, in (2.12), only the part of the $\hat{\beta}^*$ vector that has the β estimates.

The previous simulations were redone using this procedure. Table 2.9 shows the estimated values for β_1, β_2, σ^2 and σ_γ^2. Table 2.10 shows the variances of the estimated values for β_1, β_2, σ^2 and σ_γ^2.

This method performed well in giving unbiased estimates for β_1 and β_2, as can be seen in Table 2.9. Also, the variances of all of these estimates are all very small, which can be seen in Table 2.10. Graphical evidence of the skewness of the estimates can be seen when they are plotted. Refer to Figure 2.5 for an example. The top two histograms in this figure are of $\hat{\beta}_1$ and $\hat{\beta}_2$ for simulation 1. As we can see, both of these plots follow a normal distribution approximately. Graphs of all of the other β

estimates are not shown as they are similar to those for simulation 1.

This method of estimation did not perform well for σ^2 but, similar to method 3, it did do better than methods 1 and 2. Most of the estimates are not close to their original values. However, some simulations performed well.

The estimates of σ^2 tend to be better when σ^2 is large relative to σ_γ^2. They are also better when σ^2 and σ_γ^2 are small and equal. All variances of the simulation estimates can be seen in Table 2.10. Graphical evidence of the skewness of the estimates can be seen when they are plotted. Refer to Figure 2.5 for an example. The bottom left histogram in this figure is of σ^2 for simulation 1. As we can see, this plot follows a normal distribution approximately and thus is not skew. Graphs of all of the other σ^2 estimates are not shown, as they are similar to those for simulation 1.

This method of estimation did not perform well for estimating σ_γ^2. The estimates were all very close to zero and not close to the original values, which can be seen in Table 2.9. Also, the variances were all small, which can be seen in Table 2.10. Graphical evidence of the skewness of the estimates can be seen when they are plotted. Refer to Figure 2.5 for an example. The bottom right histogram in this figure is of $\hat{\sigma}_\gamma^2$ for simulation 1. As we can see, this plot is skew to the right. Graphs of all of the other σ_γ^2 estimates are not shown, as they are similar to those for simulation 1.

Figure 2.5: Simulated Method 4 Histograms (LMM): Simulation 1

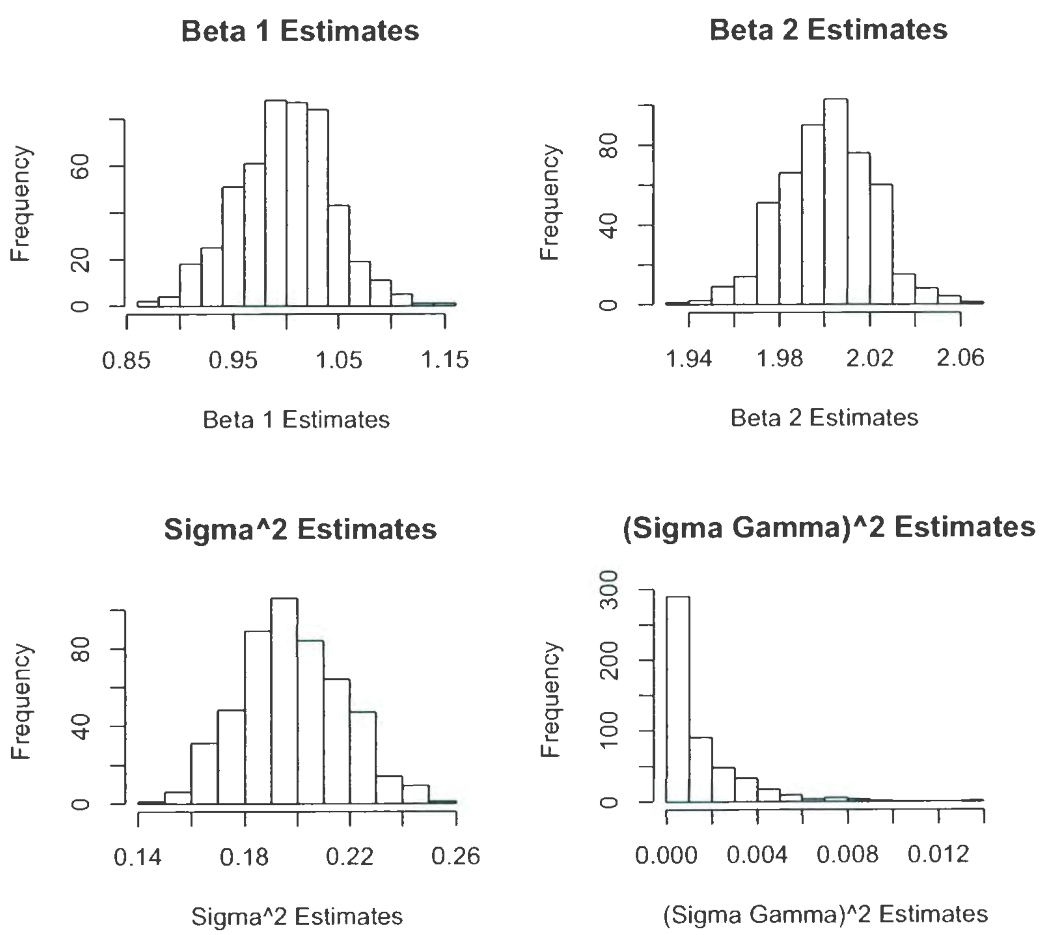


Table 2.9: Simulated Method 1 Estimates (LMM)

Sim	β_1	β_2	σ^2	σ_γ^2	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$	$\hat{\sigma}_\gamma^2$
1	1.0	2.0	0.1	0.1	0.9992053	2.0009335	0.198028	1.111692e-06
2	1.0	2.0	0.1	1.0	0.9990171	2.0018323	1.083730	2.118716e-05
3	1.0	2.0	0.1	2.0	0.99890035	2.0023767	2.067802	1.618815e-05
4	1.0	2.0	0.1	4.0	0.9987425	2.0031467	1.035916	9.706120e-05
5	1.0	2.0	1.0	0.1	0.9976749	2.0020532	1.094401	3.535568e-08
6	1.0	2.0	1.0	1.0	0.9974870	2.0029520	1.980282	1.111692e-05
7	1.0	2.0	1.0	2.0	0.9973730	2.0031960	2.964163	3.068778e-05
8	1.0	2.0	1.0	1.0	0.9972120	2.0042660	1.932730	7.388147e-05
9	1.0	2.0	2.0	0.1	0.9967177	2.0027315	2.090328	9.009974e-07
10	1.0	2.0	2.0	1.0	0.9965598	2.0036303	2.976319	6.852437e-06
11	1.0	2.0	2.0	2.0	0.9964460	2.0011750	3.960565	2.283385e-05
12	1.0	2.0	2.0	4.0	0.9962850	2.0019450	5.928925	6.137556e-05
13	1.0	2.0	4.0	0.1	0.9954366	2.0036908	4.082154	1.103366e-06
14	1.0	2.0	1.0	1.0	0.9952187	2.0015896	4.968298	2.375416e-06
15	1.0	2.0	1.0	2.0	0.9951348	2.0011340	5.952637	1.370487e-05
16	1.0	2.0	1.0	1.0	0.9949738	2.0059040	7.921130	1.566769e-05
17	1.0	0.1	0.1	0.1	0.9992053	0.1009335	0.198028	1.111692e-06
18	1.0	0.1	0.1	1.0	0.9990171	0.1018323	1.083730	2.118716e-05
19	1.0	5.0	0.1	0.1	0.9992053	5.0009335	0.198028	1.111692e-06
20	1.0	5.0	0.1	1.0	0.9990174	5.0018323	1.083730	2.118716e-05
21	1.0	2.0	0.01	0.01	0.9997187	2.0002952	0.019803	1.111692e-07

2.4 Quasi-Likelihood Method Analysis

The quasi-likelihood method discussed in Section 1.3 may also be used for the analysis of a linear mixed model.

From (1.12), we can solve explicitly for β:

$$\beta = \left(\sum_{i=1}^{I} x_i^T \Sigma_i^{-1} x_i \right)^{-1} \sum_{i=1}^{I} x_i^T \Sigma_i^{-1} y_i.$$

Table 2.10: Simulated Method 4 Variances (LMM)

Sim	β_1	β_2	σ^2	σ_γ^2	$Var(\beta_1)$	$Var(\beta_2)$	$Var(\sigma^2)$	$Var(\hat{\sigma}_\gamma^2)$
1	1.0	2.0	0.1	0.1	0.001996702	0.000396723	0.0003705961	3.026078e-06
2	1.0	2.0	0.1	1.0	0.01568793	0.003010129	0.02207376	0.0001804578
3	1.0	2.0	0.1	2.0	0.03088214	0.005914388	0.0855426	0.0007058088
4	1.0	2.0	0.1	4.0	0.06125772	0.01172273	0.3364106	0.002803130
5	1.0	2.0	1.0	0.1	0.006197615	0.001345603	0.00587593	3.359525e-05
6	1.0	2.0	1.0	1.0	0.01996702	0.003960723	0.03705961	0.0003026078
7	1.0	2.0	1.0	2.0	0.03520858	0.006865626	0.1118031	0.0009129322
8	1.0	2.0	1.0	4.0	0.06565111	0.01267488	0.3865209	0.003142628
9	1.0	2.0	2.0	0.1	0.01084703	0.002400389	0.02074139	0.0001051068
10	1.0	2.0	2.0	1.0	0.02466379	0.005016153	0.0618266	0.0004842267
11	1.0	2.0	2.0	2.0	0.03993401	0.007921447	0.1482384	0.001210431
12	1.0	2.0	2.0	4.0	0.07041717	0.01373125	0.4472137	0.003654729
13	1.0	2.0	4.0	0.1	0.02013302	0.004509786	0.07826675	0.0003680217
14	1.0	2.0	4.0	1.0	0.03401675	0.007126462	0.1383692	0.000967397
15	1.0	2.0	4.0	2.0	0.04932758	0.01003231	0.2473064	0.001936907
16	1.0	2.0	4.0	4.0	0.07986808	0.01584289	0.05929538	0.004841725
17	1.0	0.1	0.1	0.1	0.001996702	0.0003960723	0.0003705961	3.026078e-06
18	1.0	0.1	0.1	1.0	0.01568793	0.003010129	0.02207376	0.0001804578
19	1.0	5.0	0.1	0.1	0.001996702	0.0003960723	0.0003705961	3.026078e-06
20	1.0	5.0	0.1	1.0	0.01568793	0.00301029	0.02207376	0.0001804578
21	1.0	2.0	0.01	0.01	0.0001996702	3.960723e-05	3.705961e-06	3.026078e-08

where x_i and y_i are as previously defined in Chapter 1, and

$$\Sigma_i^{-1} = \begin{bmatrix} \hat{\sigma}^2 + \hat{\sigma}_\gamma^2 & \hat{\sigma}_\gamma^2 & \cdots & \hat{\sigma}_\gamma^2 \\ \hat{\sigma}_\gamma^2 & \hat{\sigma}^2 + \hat{\sigma}_\gamma^2 & \cdots & \hat{\sigma}_\gamma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_\gamma^2 & \hat{\sigma}_\gamma^2 & \cdots & \hat{\sigma}^2 + \hat{\sigma}_\gamma^2 \end{bmatrix}^{-1}$$

As we can see, the equation for $\hat{\beta}$ for the quasi-likelihood method is the same as the moment method estimator. We cannot solve (1.12) explicitly for $\sigma^{2*} = (\sigma^2, \sigma_\gamma^2)$, therefore we need to estimate σ^2 and σ_γ^2 by solving (1.12) numerically with Newton's method. Initial estimates were chosen and used to start a Newton-Raphson iteration. Suppose $\hat{\sigma}^{2*}$ is a solution to (4.12); then, at iteration $(r+1)$:

$$\begin{pmatrix} \hat{\sigma}^2_{(r+1)} \\ \hat{\sigma}^2_{\gamma,(r+1)} \end{pmatrix} = \begin{pmatrix} \hat{\sigma}^2_{(r)} \\ \hat{\sigma}^2_{\gamma,(r)} \end{pmatrix} + \left(\sum_{i=1}^{I} \frac{\partial \lambda_i^T}{\partial \hat{\sigma}^{2*}} \Omega_i^{-1} \frac{\partial \lambda_i}{\partial \hat{\sigma}^{2*}} \right)_r^{-1} \left(\sum_{i=1}^{I} \frac{\partial \lambda_i^T}{\partial \sigma^{2*}} \Omega_i^{-1} (U_i - \lambda_i) \right)_r ,$$

where

$$U_i = \begin{bmatrix} y_{i1}^2 \\ y_{i2}^2 \\ \vdots \\ y_{in}^2 \\ y_{i1}y_{i2} \\ y_{i1}y_{i3} \\ \vdots \\ y_{i,n-1}y_{i,n} \end{bmatrix},$$

$$\lambda_i = E(U_i) = \begin{bmatrix} E(y_{i1}^2) \\ E(y_{i2}^2) \\ \vdots \\ E(y_{in}^2) \\ E(y_{i1}y_{i2}) \\ E(y_{i1}y_{i3}) \\ \vdots \\ E(y_{i,n-1}y_{i,n}) \end{bmatrix} = \begin{bmatrix} \sigma^2 + \sigma_\gamma^2 + (x_{i1}^T\beta)^2 \\ \sigma^2 + \sigma_\gamma^2 + (x_{i2}^T\beta)^2 \\ \vdots \\ \sigma^2 + \sigma_\gamma^2 + (x_{in}^T\beta)^2 \\ \sigma_\gamma^2 + x_{i1}^T\beta x_{i2}^T\beta \\ \sigma_\gamma^2 + x_{i1}^T\beta x_{i3}^T\beta \\ \vdots \\ \sigma_\gamma^2 + x_{i,n-1}^T\beta x_{i,n}^T\beta \end{bmatrix},$$

$$\frac{\partial \lambda_i^T}{\partial \sigma_\gamma^{2*}} = \frac{\partial}{\partial \sigma_\gamma^{2*}} = \begin{bmatrix} \frac{\partial \lambda_i}{\partial \sigma^2} & \frac{\partial \lambda_i}{\partial \sigma_\gamma^2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}, \text{ and}$$

$$\Omega_i = Var(U_i) = Var \begin{pmatrix} y_{i1}^2 \\ y_{i2}^2 \\ \vdots \\ y_{in}^2 \\ y_{i1}y_{i2} \\ y_{i1}y_{i3} \\ \vdots \\ y_{i,n-1}y_{in} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix}.$$

We shall further define Ω_i using Σ_{11}, Σ_{12}, and Σ_{22} as follows:

$$\Sigma_{11} = Var \begin{pmatrix} y_{i1}^2 \\ y_{i2}^2 \\ \vdots \\ y_{in}^2 \end{pmatrix}$$

$$= \begin{bmatrix} Var(y_{i1}^2) & Cov(y_{i1}^2, y_{i2}^2) & \cdots & Cov(y_{i1}^2, y_{i,n-1}^2) & Cov(y_{i1}^2, y_{in}^2) \\ Cov(y_{i2}^2, y_{i1}^2) & Var(y_{i2}^2) & \cdots & Cov(y_{i2}^2, y_{i,n-1}^2) & Cov(y_{i2}^2, y_{in}^2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Cov(y_{i,n-1}^2, y_{i1}^2) & Cov(y_{i,n-1}^2, y_{i2}^2) & \cdots & Var(y_{i,n-1}^2) & Cov(y_{i,n-1}^2, y_{in}^2) \\ Cov(y_{in}^2, y_{i1}^2) & Cov(y_{in}^2, y_{i2}^2) & \cdots & Cov(y_{in}^2, y_{i,n-1}^2) & Var(y_{in}^2) \end{bmatrix},$$

where

$$\begin{aligned} Var(y_{ij}^2) &= (\sigma_\gamma^2 + \sigma^2)(4(x_{ij}^T\beta)^2 + 2(\sigma_\gamma^2 + \sigma^2)), \\ Cov(y_{ij}^2, y_{ik}^2) &= 4x_{ij}^T\beta x_{ik}^T\beta\sigma_\gamma^2 + 2\sigma_\gamma^4 \quad j \neq k. \end{aligned}$$

$$\Sigma_{12} = \begin{bmatrix} Cov(y_{i1}^2, y_{i1}y_{i2}) & Cov(y_{i1}^2, y_{i1}y_{i3}) & \cdots & Cov(y_{i1}^2, y_{i,n-1}y_{in}) \\ Cov(y_{i2}^2, y_{i1}y_{i2}) & Cov(y_{i2}^2, y_{i1}y_{i3}) & \cdots & Cov(y_{i2}^2, y_{i,n-1}y_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(y_{i,n-1}^2, y_{i1}y_{i2}) & Cov(y_{i,n-1}^2, y_{i1}y_{i3}) & \cdots & Cov(y_{i,n-1}^2, y_{i,n-1}y_{in}) \\ Cov(y_{in}^2, y_{i1}y_{i2}) & Cov(y_{in}^2, y_{i1}y_{i3}) & \cdots & Cov(y_{in}^2, y_{i,n-1}y_{in}) \end{bmatrix},$$

where

$$Cov(y_{ij}^2, y_{ij}y_{ik}) = \begin{cases} 2(x_{ij}^T\beta)^2\sigma_\gamma^2 + 2x_{ij}^T\beta x_{ik}^T\beta\sigma_\gamma^2 + 2x_{ij}^T\beta x_{ik}^T\beta\sigma^2 + 2\sigma_\gamma^2\sigma^2 + 2\sigma_\gamma^4 & j < k \\ Cov(y_{ij}^2, y_{ik}y_{ij}) & k < j. \end{cases}$$

$$Cov(y_{ij}^2, y_{ik}y_{il}) = 2x_{ij}^T\beta x_{ik}^T\beta\sigma_\gamma^2 + 2x_{ij}^T\beta x_{il}^T\beta\sigma_\gamma^2 + 2\sigma_\gamma^4 \quad j \neq k \neq l, k < l,$$

$$\Sigma_{22} = \begin{bmatrix} Var(y_{i1}, y_{i2}) & Cov(y_{i1}y_{i2}, y_{i1}y_{i3}) & \cdots & Cov(y_{i1}y_{i2}, y_{i,n-1}y_{in}) \\ Cov(y_{i1}y_{i3}, y_{i1}y_{i2}) & Var(y_{i1}, y_{i3}) & \cdots & Cov(y_{i1}y_{i3}, y_{i,n-1}y_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(y_{i,n-2}y_{in}, y_{i1}y_{i2}) & Cov(y_{i,n-2}y_{in}, y_{i1}y_{i3}) & \cdots & Cov(y_{i,n-2}y_{in}, y_{i,n-1}y_{in}) \\ Cov(y_{i,n-1}y_{in}, y_{i1}y_{i2}) & Cov(y_{i,n-1}y_{in}, y_{i1}y_{i3}) & \cdots & Var(y_{i,n-1}y_{in}) \end{bmatrix}.$$

where

$$\begin{aligned} Var(y_{ij}y_{ik}) &= (x_{ij}^T\beta)^2(\sigma_\zeta^2 + \sigma^2) + (x_{ik}^T\beta)^2(\sigma_\zeta^2 + \sigma^2) + 2x_{ij}^T\beta x_{ik}^T\beta\sigma_\zeta^2 + \\ &\quad 2\sigma_\zeta^4 + 2\sigma_\zeta^2\sigma^2 + \sigma^4 \quad j < k. \end{aligned}$$

$$Cov(y_{ij}y_{ik}, y_{ij}y_{im}) = \begin{cases} x_{ij}^T\beta x_{ik}^T\beta\sigma_\zeta^2 + x_{ij}^T\beta x_{il}^T\beta\sigma_\zeta^2 + x_{ik}^T\beta x_{il}^T\beta\sigma_\zeta^2 + (x_{ij}^T\beta)^2\sigma_\zeta^2 + \\ \quad x_{ik}^T\beta x_{il}^T\beta\sigma^2 + \sigma_\zeta^2\sigma^2 + 2\sigma_\zeta^4 \quad j < k, j < m, k \neq m \\ Cov(y_{ij}y_{ik}, y_{im}y_{ij}) \quad j < k, m < j, k \neq m \\ Cov(y_{ik}y_{ij}, y_{ij}y_{im}) \quad k < j, j < m, k \neq m \\ Cov(y_{ik}y_{ij}, y_{im}y_{ij}) \quad k < j, m < j, k \neq m. \end{cases}$$

$$\begin{aligned} Cov(y_{ij}y_{ik}, y_{il}y_{im}) &= x_{ij}^T\beta x_{il}^T\beta\sigma_\zeta^2 + x_{ij}^T\beta x_{im}^T\beta\sigma_\zeta^2 + x_{ik}^T\beta x_{il}^T\beta\sigma_\zeta^2 + x_{ik}^T\beta x_{im}^T\beta\sigma_\zeta^2 + 2\sigma_\zeta^4 \\ &\quad j < k, l < m, j \neq l, k \neq m. \end{aligned}$$

The previous simulations were redone using this procedure. Table 2.11 shows the estimated values for β_1, β_2, σ^2 and σ_ζ^2. Table 2.12 shows the variances of the estimated values for β_1, β_2, σ^2 and σ_ζ^2.

This method performed well in giving unbiased estimates for β_1 and β_2, as can be seen in Table 2.11. Also, the variances of all of these estimates are all very small, which can be seen in Table 2.12. The only exception to this is simulation 9, which has a higher variance for both β_1 and β_2 compared to the rest. If we look at the estimates for these parameters for this simulation, we see that they are the furthest from the true parameters but the estimates are still not bad.

The quasi-likelihood method performed well for σ^2 and σ_γ^2, though not as well as it did for β. Some estimates appear biased, particularly when σ^2 or σ_γ^2 is large. The variances of all of these estimates are all very small when σ^2 and σ_γ^2 are unbiased, which can be see in Table 2.12. However, the simulations that did not produce good estimates also seem to have an issue with outliers.

The estimates do not appear to follow any trend of biasedness. That is, they do not become unbiased or biased as a value of σ^2 or σ_γ^2 becomes larger or smaller.

For σ^2 it appears that the variances are becoming larger as the values for σ_γ^2 become larger as well as for when the values of σ^2 become larger. In addition, the variances for σ_γ^2 appear to become larger as the values for σ_γ^2 become larger.

It should also be noted that there are a couple of simulations in which the estimated value for σ_γ^2 is negative. This is because there is nothing in the iterative estimation scheme that prevents $\hat{\sigma}_\gamma^2$ from becoming negative.

Table 2.11: Quasi-Likelihood Method Estimates (LMM)

Sim	β_1	β_2	σ^2	σ_γ^2	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$	$\hat{\sigma}_\gamma^2$
1	1.0	2.0	0.1	0.1	1.000572	1.999609	0.1023606	0.08147847
2	1.0	2.0	0.1	1.0	1.002572	1.997817	0.09746832	0.9965456
3	1.0	2.0	0.1	2.0	1.002501	1.997758	0.1013545	1.956483
4	1.0	2.0	0.1	4.0	1.002242	1.997801	0.1002508	3.943108
5	1.0	2.0	1.0	0.1	1.024620	1.980704	0.9164412	-0.8892235
6	1.0	2.0	1.0	1.0	1.004274	1.996789	0.9887554	1.040084
7	1.0	2.0	1.0	2.0	1.007549	1.994053	0.9447538	2.525565
8	1.0	2.0	1.0	4.0	1.005565	1.995476	0.9465685	1.318274
9	1.0	2.0	2.0	0.1	1.094606	1.924918	2.061436	2.704438
10	1.0	2.0	2.0	1.0	0.99796 47	2.0020388	2.020512	0.8674288
11	1.0	2.0	2.0	2.0	1.007250	1.994494	1.936994	2.262527
12	1.0	2.0	2.0	4.0	1.008259	1.993523	1.889096	1.489014
13	1.0	2.0	4.0	0.1	1.016196	1.987931	3.990002	0.05606692
14	1.0	2.0	4.0	1.0	1.010156	1.992572	1.044534	0.895599
15	1.0	2.0	4.0	2.0	1.014622	1.991283	1.026259	1.888344
16	1.0	2.0	4.0	4.0	1.014579	1.991454	3.956583	1.672387
17	1.0	0.1	0.1	0.1	1.0020566	0.0984206	0.09945787	0.1001519
18	1.0	0.1	0.1	1.0	1.0025607	0.0978259	0.1004505	0.9851484
19	1.0	5.0	0.1	0.1	1.008150	1.993546	0.1035344	0.086501
20	1.0	5.0	0.1	1.0	1.002358	1.997988	0.1001424	0.9874656
21	1.0	2.0	0.01	0.01	1.000889	1.999310	0.01067947	-0.004296529

Table 2.12: Quasi-Likelihood Method Variances (LMM)

Sim	β_1	β_2	σ^2	σ_γ^2	$Var(\beta_1)$	$Var(\beta_2)$	$Var(\hat{\sigma}^2)$	$Var(\hat{\sigma}_\gamma^2)$
1	1.0	2.0	0.1	0.1	0.001802491	0.0002857912	0.0001225541	0.03652485
2	1.0	2.0	0.1	1.0	0.01101498	0.0003389412	0.0008130844	0.08523364
3	1.0	2.0	0.1	2.0	0.02116491	0.0003495881	0.003558799	0.4026576
4	1.0	2.0	0.1	4.0	0.04132562	0.0003566885	0.005873911	1.377193
5	1.0	2.0	1.0	0.1	0.1576340	0.09614026	9.698067	156.79910
6	1.0	2.0	1.0	1.0	0.01748851	0.002723732	0.02418229	0.2868234
7	1.0	2.0	1.0	2.0	0.02984284	0.003943186	0.1390198	16.028941
8	1.0	2.0	1.0	4.0	0.05184079	0.005376615	0.1801387	14.895175
9	1.0	2.0	2.0	0.1	4.629121	2.963241	0.3588191	2968.44620
10	1.0	2.0	2.0	1.0	0.0479069	0.02052437	0.1272192	2.679157
11	1.0	2.0	2.0	2.0	0.03380364	0.005026768	0.5756753	14.414851
12	1.0	2.0	2.0	4.0	0.05584598	0.00611066	2.162666	54.736158
13	1.0	2.0	1.0	0.1	0.03257327	0.01289022	0.1254139	1.054371
14	1.0	2.0	4.0	1.0	0.03357349	0.007437036	0.2216968	0.6071359
15	1.0	2.0	4.0	2.0	0.04580061	0.00851557	0.2022703	0.7232518
16	1.0	2.0	1.0	1.0	0.06804563	0.009817614	0.3159631	201.83902
17	1.0	0.1	0.1	0.1	0.001696515	0.0002381026	0.0001105949	0.0006835622
18	1.0	0.1	0.1	1.0	0.01098391	0.0003356688	0.0001734860	0.02613172
19	1.0	5.0	0.1	0.1	0.009231578	0.00500337	0.001253035	5.190114
20	1.0	5.0	0.1	1.0	0.01098292	0.0003355242	0.0002568756	0.05849566
21	1.0	2.0	0.01	0.01	0.0003084314	0.0001089242	0.0000909913	0.08701333

Chapter 3

Simulation Data Analysis - Generalized Linear Mixed Model

3.1 Introduction

For this chapter, all analyses are done using a logistic model with fixed and random effects, which is an example of a generalized linear mixed model (GLMM).

For all analyses of the generalized linear model, we wish to estimate β and σ_γ. We first do this using the moment method, then with the simulated method introduced in the previous chapter, and finally with the quasi-likelihood method.

As a reminder, the generalized linear mixed model that we are using for this chapter is the following:

$$y_{ij}|\gamma_i \sim \text{Binomial}\,(1, \pi_{ij}). \tag{3.1}$$

where

$$\pi_{ij} = P(y_{ij}=1|\gamma_i) = \frac{e^{x_{ij}^T\beta+\sigma_\gamma\gamma_i}}{1+e^{x_{ij}^T\beta+\sigma_\gamma\gamma_i}}, \qquad i=1\cdots I,\ j=1\cdots n_i.$$

All simulations used the following initial conditions and parameters:

- $i = 1, \ldots, I = 100.$
- $j = 1, \ldots, n = 4.$
- $\beta = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$
- $X_i = \begin{cases} \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} & \text{if } i = 1, \cdots, 50; \\ \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} & \text{if } i = 51, \cdots, 100. \end{cases}$
- y_{ij} are created from the model defined in (1.7).

The generalized linear methods of estimation were studied for the following situations. First, for the β given earlier, we used σ_γ=0.1, 0.3, 0.5, 0.7, 1.0, 2.0, 4.0. Then we used $(\beta_1, \beta_2, \sigma_\gamma) = (1, 0.1, 0.1)$ and $(1, 5, 0.1)$.

3.2 Moment Method Analysis

Jiang (1998) introduced a method of estimating the fixed effects and variance components in a generalized linear mixed model that was based on simulated moments.

This method was suggested for it's computational feasibility and the consistency of its estimators. The method, however, can produce inefficient moment estimators.

To combat the problem of inefficient moment estimators, Jiang and Zhang (2001) proposed robust methods for estimating the parameters of interest in an extended generalized linear mixed model. A first step estimator is calculated by solving a system of estimating equations. This estimator is consistent. Next, a second step estimator is calculated by solving a system of optimal estimating equations. This second step estimator maintains the asymptotic optimality and produces much better results than the first step estimator.

We begin with discussion of the results using the moment method of Jiang (1998), discussed in Section 1.2.

As mentioned in Chapter 1, some of the expectations used for this method are very difficult to find for the logistic model and they can be approximated. For example, we need to evaluate expectations such as:

$$E\left[\frac{e^{\mu+\sigma_{\gamma_i}\gamma_i}}{1+e^{\mu+\sigma_{\gamma_i}\gamma_i}}\right] = \int\left[\frac{e^{\mu+\sigma_{\gamma_i}\gamma_i}}{1+e^{\mu+\sigma_{\gamma_i}\gamma_i}}\right] f(\gamma_i)d\gamma_i.$$

It is not possible to evaluate this integral explicitly. To overcome this problem, we generated 500 γ_i values from the standard normal distribution and approximated the integral by calculating:

$$\int\left[\frac{e^{\mu+\sigma_{\gamma_i}\gamma_i}}{1+e^{\mu+\sigma_{\gamma_i}\gamma_i}}\right] f(\gamma_i)d\gamma_i \approx \frac{1}{w}\sum_{k=1}^{w}\frac{e^{\mu+\sigma_{\gamma_{i,w}}\gamma_{i,w}}}{1+e^{\mu+\sigma_{\gamma_{i,w}}\gamma_{i,w}}}.$$

The value for w needs to be large. We used $w = 500$ in our studies. We also used $w = 1000$ in some cases, but found little difference from the results when $w = 500$.

We further define $E(w)$ using $E(W_1)$ and $E(W_2)$:

$$
\begin{aligned}
E(W_1) &\approx \frac{1}{500}\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} e^{x_{ij}^T\beta} \sum_{w=1}^{500} \frac{e^{\sigma_\gamma \gamma_{i,w}}}{1 + e^{x_{ij}^T\beta + \sigma_\gamma \gamma_{i,w}}}, \\
E(W_2) &\approx \frac{1}{250}\sum_{i=1}^{m}\sum_{j\neq k} e^{x_{ij}^T\beta + x_{ik}^T\beta} \sum_{w=1}^{500} \frac{e^{2\sigma_\gamma \gamma_{i,w}}}{(1 + e^{x_{ij}^T\beta + \sigma_\gamma \gamma_{i,w}})(1 + e^{x_{ik}^T\beta + \sigma_\gamma \gamma_{i,w}})}.
\end{aligned}
$$

We further define P^T using $\frac{\partial E(W_1)}{\partial \beta^T}$, $\frac{\partial E(W_2)}{\partial \beta^T}$, $\frac{\partial E(W_1)}{\partial \sigma_{\gamma}}$ and $\frac{\partial E(W_2)}{\partial \sigma_\gamma}$.

$$
\begin{aligned}
\frac{\partial E(W_1)}{\partial \beta^T} &\approx \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} x_{ij}^T e^{x_{ij}^T\beta} \Bigg\{ \frac{1}{500}\sum_{w=1}^{500} \frac{e^{\sigma_\gamma \gamma_{i,w}}}{1 + e^{x_{ij}^T\beta + \sigma_\gamma \gamma_{i,w}}} \\
&\quad - e^{x_{ij}^T\beta} \left(\frac{1}{500}\sum_{w=1}^{500} \frac{e^{2\sigma_\gamma \gamma_{i,w}}}{(1 + e^{x_{ij}^T\beta + \sigma_\gamma \gamma_{i,w}})^2} \right) \Bigg\}, \\
\frac{\partial E(W_1)}{\partial \sigma_{\gamma}} &\approx \frac{1}{500}\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} e^{x_{ij}^T\beta} \left(\sum_{w=1}^{500} \frac{\gamma_{i,w} e^{\sigma_\gamma \gamma_{i,w}}}{(1 + e^{x_{ij}^T\beta + \sigma_\gamma \gamma_{i,w}})^2} \right), \\
\frac{\partial E(W_2)}{\partial \beta^T} &\approx \frac{1}{250}\sum_{i=1}^{m}\sum_{j\neq k} \Bigg\{ (x_{ij}^T + x_{ik}^T) e^{x_{ij}^T\beta + x_{ik}^T\beta} \sum_{w=1}^{500} \frac{e^{2\sigma_\gamma \gamma_{i,w}}}{(1 + e^{x_{ij}^T\beta + \sigma_\gamma \gamma_{i,w}})(1 + e^{x_{ik}^T\beta + \sigma_\gamma \gamma_{i,w}})} \\
&\quad - e^{x_{ij}^T\beta + x_{ik}^T\beta} \sum_{w=1}^{500} \Bigg[\frac{x_{ij}^T e^{x_{ij}^T\beta + 3\sigma_\gamma \gamma_{i,w}} + x_{ik}^T e^{x_{ik}^T\beta + 3\sigma_\gamma \gamma_{i,w}} +}{(1 + e^{x_{ij}^T\beta + \sigma_\gamma \gamma_{i,w}})^2 (1 + e^{x_{ik}^T\beta + \sigma_\gamma \gamma_{i,w}})^2} \\
&\quad + \frac{(x_{ij}^T + x_{ik}^T) e^{x_{ij}^T\beta + x_{ik}^T\beta + 4\sigma_\gamma \gamma_{i,w}}}{(1 + e^{x_{ij}^T\beta + \sigma_\gamma \gamma_{i,w}})^2 (1 + e^{x_{ik}^T\beta + \sigma_\gamma \gamma_{i,w}})^2} \Bigg] \Bigg\}, \\
\frac{\partial E(W_2)}{\partial \sigma_\gamma} &\approx \frac{1}{250}\sum_{i=1}^{m}\sum_{j\neq k} e^{x_{ij}^T\beta + x_{ik}^T\beta} \sum_{w=1}^{500} \frac{2\gamma_{i,w} e^{2\sigma_\gamma \gamma_{i,w}} + \gamma_{i,w} e^{x_{ij}^T\beta + 3\sigma_\gamma \gamma_{i,w}} + \gamma_{i,w} e^{x_{ik}^T\beta + 3\sigma_\gamma \gamma_{i,w}}}{(1 + e^{x_{ij}^T\beta + \sigma_\gamma \gamma_{i,w}})^2 (1 + e^{x_{ik}^T\beta + \sigma_\gamma \gamma_{i,w}})^2}.
\end{aligned}
$$

Table 3.1 shows the estimated values for β_1, β_2 and σ_γ for all of the simulations conducted. Table 3.2 shows the variances of the estimated values for β_1, β_2 and σ_γ.

Included in Table 3.1 is a column containing the number of simulations (out of 500) that broke a Newton-Raphson iteration due to the σ_γ value being too large (larger than 5). The number of breaks is in the table under the column NA. This method can produce a number of negative estimates, particularly when σ^2 is small.

This moment method performed well in estimating β_1 and β_2, giving unbiased results. Evidence supporting this is that all of the estimates of β_1 and β_2 are close to their original values, which can be seen in Table 3.1. However, bias tends to increase as σ_γ increases. The variability of the estimates also increases with increasing σ_γ. Also, the variances of all of these estimates are very small, which can be see in Table 3.2.

Finally, if we look at the medians, means and trimmed means of all β_1 and β_2 estimates along with the histograms of these estimates, we see that there are some sets of estimates that do not appear approximately normally distributed. To show an example of the histograms in which $\hat{\beta}_1$ and $\hat{\beta}_2$ perform poorly, we can refer to the top two histograms in Figure 3.2. For an example of the histograms in which $\hat{\beta}_1$ and $\hat{\beta}_2$ perform well, we can see the top two histograms in Figure 3.1.

This moment method did not perform well in estimating σ_γ in all cases. We see high bias in $\hat{\sigma}_\gamma$ for extreme values of σ_γ ($\sigma_\gamma = 0.1, 2, 4$). For σ_γ, it appears that the variances are becoming larger as the values for σ_γ become larger.

We made one more change to our model assumptions before proceeding to the next method of analysis. For all of the simulations conducted up to this point, the following

Figure 3.1: Moment Method Histograms (GLMM): Simulation 1

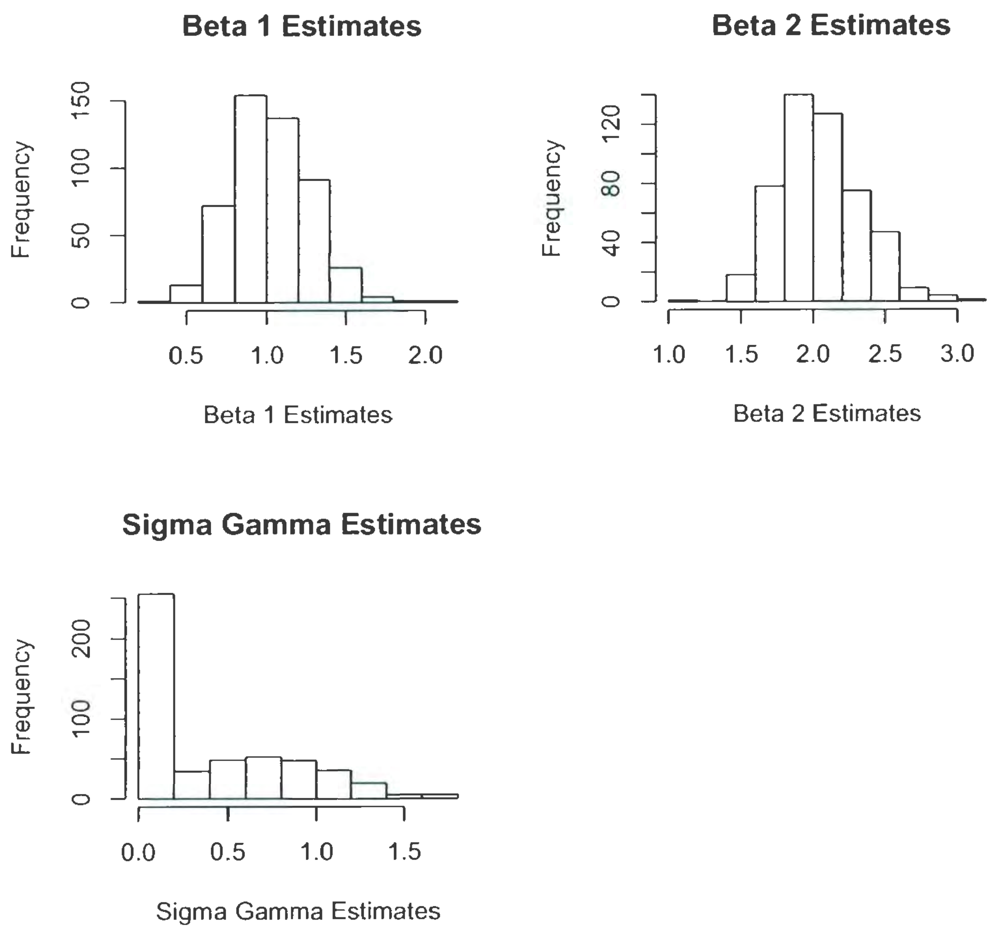


Figure 3.2: Moment Method Histograms (GLMM): Simulation 9

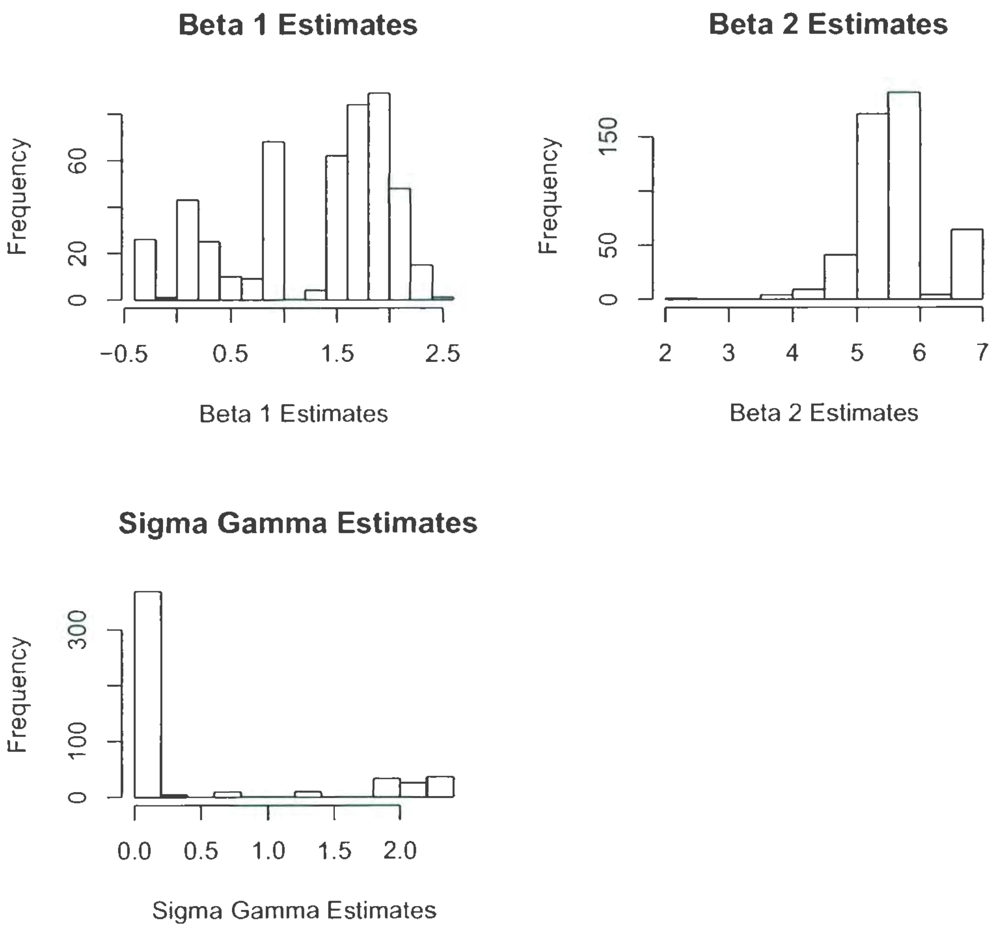


Table 3.1: Moment Method Estimates (GLMM)

Sim	β_1	β_2	σ_γ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_\gamma$	NA
1	1.0	2.0	0.1	1.025084	2.046246	0.390453	1
2	1.0	2.0	0.3	1.013614	2.040114	0.3614617	0
3	1.0	2.0	0.5	1.043679	2.066509	0.5118357	0
4	1.0	2.0	0.7	1.032868	2.058424	0.6675309	1
5	1.0	2.0	1	1.028650	2.070387	0.9736604	0
6	1.0	2.0	2	1.042971	2.133546	2.090974	12
7	1.0	2.0	4	0.9657727	1.7976482	3.531782	141
8	1.0	0.1	0.1	1.0086046	0.0978195	0.1939328	0
9	1.0	5.0	0.1	1.295802	5.655304	0.457599	15

X matrix was used:

$$X_i = \begin{cases} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} & \text{if } i = 1, \cdots, 50; \\ \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} & \text{if } i = 51, \cdots, 100. \end{cases}$$

We would like to see if the method is sensitive to the choice of the design matrix X. In particular, we evaluated the moment method using the following four X matrices:

- Case 1:

  $X_{ij1} \sim$ Uniform(0,1)

  $X_{ij2} \sim$ Uniform(0,0.5)

- Case 2:

Table 3.2: Moment Method Variances (GLMM)

Sim	β_1	β_2	σ_γ	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$	$Var(\hat{\sigma}_\gamma)$
1	1.0	2.0	0.1	0.05856356	0.0791706	0.2023941
2	1.0	2.0	0.3	0.05341902	0.07392654	0.1905228
3	1.0	2.0	0.5	0.06318721	0.09258111	0.2367752
4	1.0	2.0	0.7	0.0666329	0.09961748	0.2633809
5	1.0	2.0	1.0	0.0723449	0.1165150	0.3111153
6	1.0	2.0	2.0	0.1428336	0.3075629	0.7232923
7	1.0	2.0	1.0	0.269333	0.1967347	0.5703742
8	1.0	0.1	0.1	0.02467374	0.006208123	0.06385224
9	1.0	5.0	0.1	0.5498386	0.1659572	0.6941932

$$X_i = \begin{cases} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} & \text{if } i = 1, \cdots, 50; \\ \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} & \text{if } i = 51, \cdots, 100. \end{cases} \quad \text{(the same } X \text{ previously utilized)}$$

- Case 3:

$$X_i = \begin{cases} \begin{bmatrix} 1 & x_{ij} \\ 1 & x_{ij} \\ 1 & x_{ij} \\ 1 & x_{ij} \end{bmatrix} & \text{if } i = 1, \cdots, 50, \text{ where } x_{ij} \sim \text{Uniform}(0,1); \\ \begin{bmatrix} 1 & x_{ij} \\ 1 & x_{ij} \\ 1 & x_{ij} \\ 1 & x_{ij} \end{bmatrix} & \text{if } i = 51, \cdots, 100, \text{ where } x_{ij} \sim \text{Uniform}(-0.5,0). \end{cases}$$

- Case 4:

$$X_i = \begin{cases} \begin{bmatrix} 1 & x_{ij} \\ 1 & x_{ij} \\ 1 & x_{ij} \\ 1 & x_{ij} \end{bmatrix} & \text{if } i = 1, \cdots, 50, \text{ where } x_{ij} \sim \text{Uniform}(0,1); \\ \begin{bmatrix} 1 & x_{ij} \\ 1 & x_{ij} \\ 1 & x_{ij} \\ 1 & x_{ij} \end{bmatrix} & \text{if } i = 51, \cdots, 100, \text{ where } x_{ij} \sim \text{Uniform}(0,1). \end{cases}$$

All four of these design matrices were combined with the following four sets of parameter values, leaving us with 16 simulations to examine:

- $\beta_1 = 1.0$, $\beta_2 = 2.0$, $\sigma_\varepsilon = 0.25$
- $\beta_1 = 1.0$, $\beta_2 = 2.0$, $\sigma_\varepsilon = 0.5$
- $\beta_1 = 0.1$, $\beta_2 = 0.2$, $\sigma_\varepsilon = 0.25$
- $\beta_1 = 0.1$, $\beta_2 = 0.2$, $\sigma_\varepsilon = 0.5$

Table 3.3 shows the estimated values for β_1, β_2 and σ_ε and Table 3.4 shows the variances of the estimated values for β_1, β_2 and σ_ε.

Looking at Table 3.3, we can see that the choice of X has an effect on the estimation of β_1, β_2 and σ_ε. Also, we can see that whether β is large or small has an effect on the results.

Looking at Table 3.4, the lowest variances of β_1 occur in case 2 and case 3. For β_2, the lowest variances occur in case 2. For σ_ε, the lowest variances occur in case 3.

Overall, the program does give varied results depending on the X matrix chosen; however, all programs performed well in estimating β_1, β_2 and σ_γ.

Table 3.3: Moment Method Estimates: Effects of Changes in X (GLMM)

Sim	Case	β_1	β_2	σ_γ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_\gamma$
1	1	0.1	0.2	0.25	0.1179	0.1609	0.2768
2	1	0.1	0.2	0.5	0.1066	0.1709	0.4757
3	1	1.0	2.0	0.25	0.9803	2.0314	0.3363
4	1	1.0	2.0	0.5	1.0132	1.9761	0.4775
5	2	0.1	0.2	0.25	0.0870	0.2033	0.2364
6	2	0.1	0.2	0.5	0.0878	0.2026	0.4539
7	2	1.0	2.0	0.25	1.0112	2.0671	0.3956
8	2	1.0	2.0	0.5	1.0437	2.0665	0.5118
9	3	0.1	0.2	0.25	0.0991	0.1821	0.2115
10	3	0.1	0.2	0.5	0.1027	0.2073	0.4548
11	3	1.0	2.0	0.25	1.0025	1.9939	0.2471
12	3	1.0	2.0	0.5	1.0008	2.0169	0.4350
13	4	0.1	0.2	0.25	0.1072	0.1942	0.2351
14	4	0.1	0.2	0.5	0.1062	0.2015	0.4566
15	4	1.0	2.0	0.25	1.0051	2.0083	0.3012
16	4	1.0	2.0	0.5	0.9978	2.0098	0.4526

3.3 Simulation Method Analysis

As utilized for the linear mixed model, the simulated approach to estimation was applied. For the generalized linear mixed model we are only estimating β_1, β_2 and σ_γ. As such, we only used one of the four simulation methods proposed for the linear mixed model.

In particular, we employed Simulation Method 3. For this method, data was simulated using the model described in (1.7). This method estimates σ_γ by treating it the same

Table 3.4: Moment Method Variances: Effects of Changes in N (GLMM)

Sim	Case	β_1	β_2	σ_z	$Var(\beta_1)$	$Var(\beta_2)$	$Var(\sigma_z)$
1	1	0.1	0.2	0.25	0.0822	0.2821	0.0720
2	1	0.1	0.2	0.5	0.0862	0.3099	0.0853
3	1	1.0	2.0	0.25	0.1071	0.1553	0.1285
4	1	1.0	2.0	0.5	0.1291	0.3997	0.1583
5	2	0.1	0.2	0.25	0.0165	0.0049	0.0520
6	2	0.1	0.2	0.5	0.0191	0.0053	0.0570
7	2	1.0	2.0	0.25	0.0598	0.0907	0.2170
8	2	1.0	2.0	0.5	0.0632	0.0926	0.2368
9	3	0.1	0.2	0.25	0.0108	0.0515	0.0512
10	3	0.1	0.2	0.5	0.0127	0.0616	0.0586
11	3	1.0	2.0	0.25	0.0190	0.1187	0.0688
12	3	1.0	2.0	0.5	0.0212	0.1117	0.0885
13	4	0.1	0.2	0.25	0.0371	0.1027	0.0192
14	4	0.1	0.2	0.5	0.0370	0.1135	0.0582
15	4	1.0	2.0	0.25	0.0686	0.2581	0.0995
16	4	1.0	2.0	0.5	0.0713	0.2717	0.1103

as β. Following Simulation Method 1 for the linear mixed model, we find the y_i and x_i matrices as:

$$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in} \end{pmatrix} \quad \text{and } x_i^* = \begin{pmatrix} x_{i11} & \cdots & x_{i1p} & z_i \\ \vdots & & \vdots & \vdots \\ x_{in1} & \cdots & x_{inp} & z_i \end{pmatrix}, \tag{3.2}$$

where z_i is generated from $N(0, \sigma_z^2)$.

To estimate σ_z and β, β^* was defined as follows:

$$\beta^* = \begin{pmatrix} \beta \\ \sigma_\gamma \end{pmatrix}.$$

We use logistic regression to estimate β^*, which gives an estimate for β, and σ_γ. For this estimation we are including σ_γ in β^* so that when we estimate β we are also estimating σ_γ. As a result, $\hat{\beta}$ is a vector of the β and σ_γ estimates. All simulations were run and Table 3.5 shows the estimated values for β_1, β_2, and σ_γ. Table 3.6 shows the variances of the estimated values for β_1, β_2, and σ_γ.

Results were reasonable for β_1 and β_2. As we can see in Table 3.5, the program performs better for small values of σ_γ and for small values of β_2. The simulations that did not perform well had high values of σ_γ (2 and 4) or a high value for β_2 (5). Surprisingly, the variances, which can be seen in Table 3.6, do not get worse as σ_γ gets larger. In contrast, the variances for β_1 and β_2 get smaller as σ_γ becomes larger.

Finally, if we look at the histograms of the estimates of β_1 and β_2, we see that there are a few simulations that do not appear approximately normal, but this is due to outliers. For an example of the histograms in which β_1 and β_2 perform poorly, we can refer to the top two histograms in Figure 3.4. For an example of the histograms in which β_1 and β_2 perform well, we can see the top two histograms in Figure 3.3.

For the estimation of the β's, this method appears to be sensitive to the choice of the β's since it performed very well for $\beta_1 = 1.0$, $\beta_2 = 0.1$, and $\sigma_\gamma = 0.1$ but not very well for $\beta_1 = 1.0$, $\beta_2 = 5.0$, and $\sigma_\gamma = 0.1$.

The method did not perform well in estimating σ_γ. As we can see in Table 3.5, none of the estimates are close to the original value. However, the variances for $\hat{\sigma}_\gamma$, which can be seen in Table 3.6, are not high. The only exception to this is the variance for

$\hat{\sigma}_\gamma$ in simulation 9, which is quite high. However, the small variances do not imply that this method is performing well since the estimates are poor. The histograms for $\hat{\sigma}_\gamma$ for all but simulation 9 appear approximately normal. The histogram for simulation 9 is skew right. The estimates appear to follow a trend of biasedness. That is, the estimation becomes worse as the value of σ_γ becomes larger.

Overall, it appears that this method performed poorly in estimating all parameters.

Table 3.5: Simulated Method Estimates (GLMM)

Sim	β_1	β_2	σ_γ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_\gamma$
1	1.0	2.0	0.1	1.1164	2.1235	0.0110
2	1.0	2.0	0.3	1.1209	2.1130	0.0054
3	1.0	2.0	0.5	1.0845	2.0180	0.0003
4	1.0	2.0	0.7	1.0463	1.9581	0.0107
5	1.0	2.0	1.0	0.9659	1.8156	0.0166
6	1.0	2.0	2.0	0.7209	1.3683	0.0160
7	1.0	2.0	4.0	0.1024	0.8034	0.0043
8	1.0	0.1	0.1	1.0074	0.1019	-0.00632
9	1.0	5.0	0.1	5.6541	14.7550	0.1676

We did attempt the second modification performed on the moment method of analysis. That is, we ran this program using different X matrices to see if different results are produced, implying that the program is sensitive to the X matrix chosen.

We evaluated the simulation method using the same X matrices used in the moment method. As before, all of four of these new programs were conducted using the following four simulations, leaving us with 16 simulations in total to compare:

- $\beta_1 = 1.0$, $\beta_2 = 2.0$, $\sigma_\gamma = 0.25$
- $\beta_1 = 1.0$, $\beta_2 = 2.0$, $\sigma_\gamma = 0.5$

Figure 3.3: Simulated Method Histograms (GLMM): Simulation 7

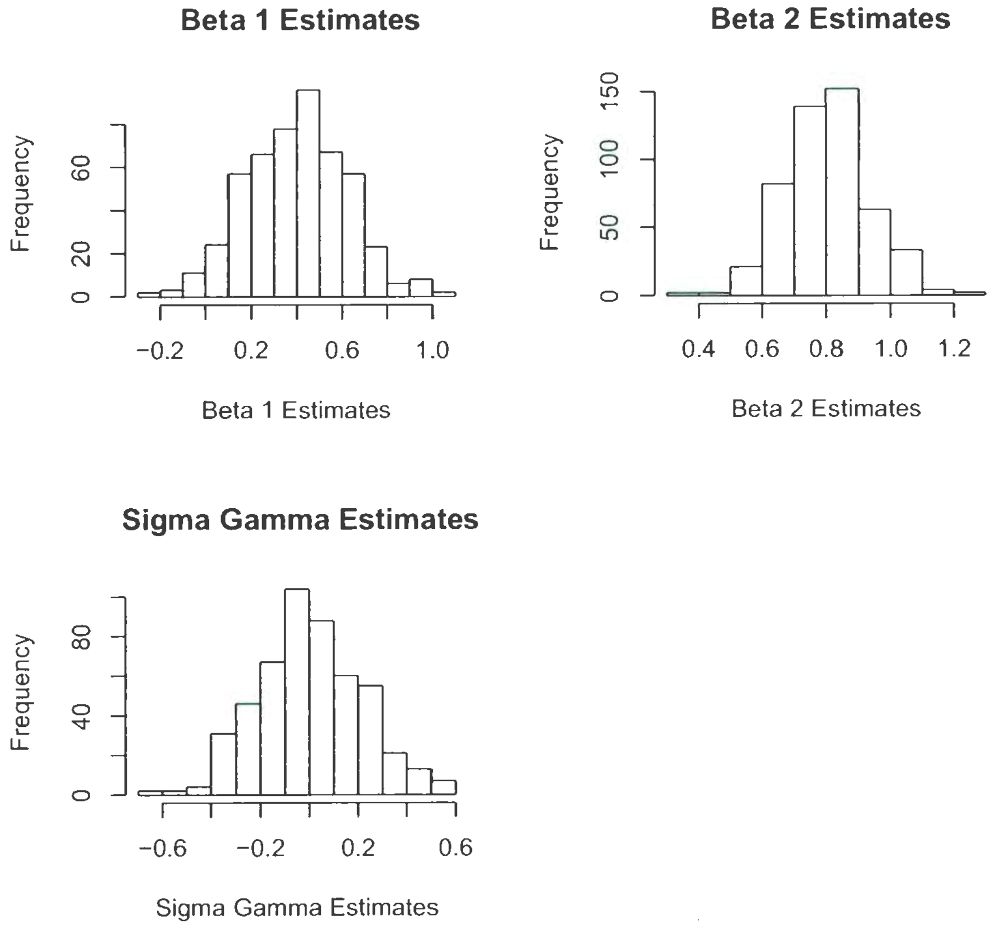


Figure 3.4: Simulated Method Histograms (GLMM): Simulation 9

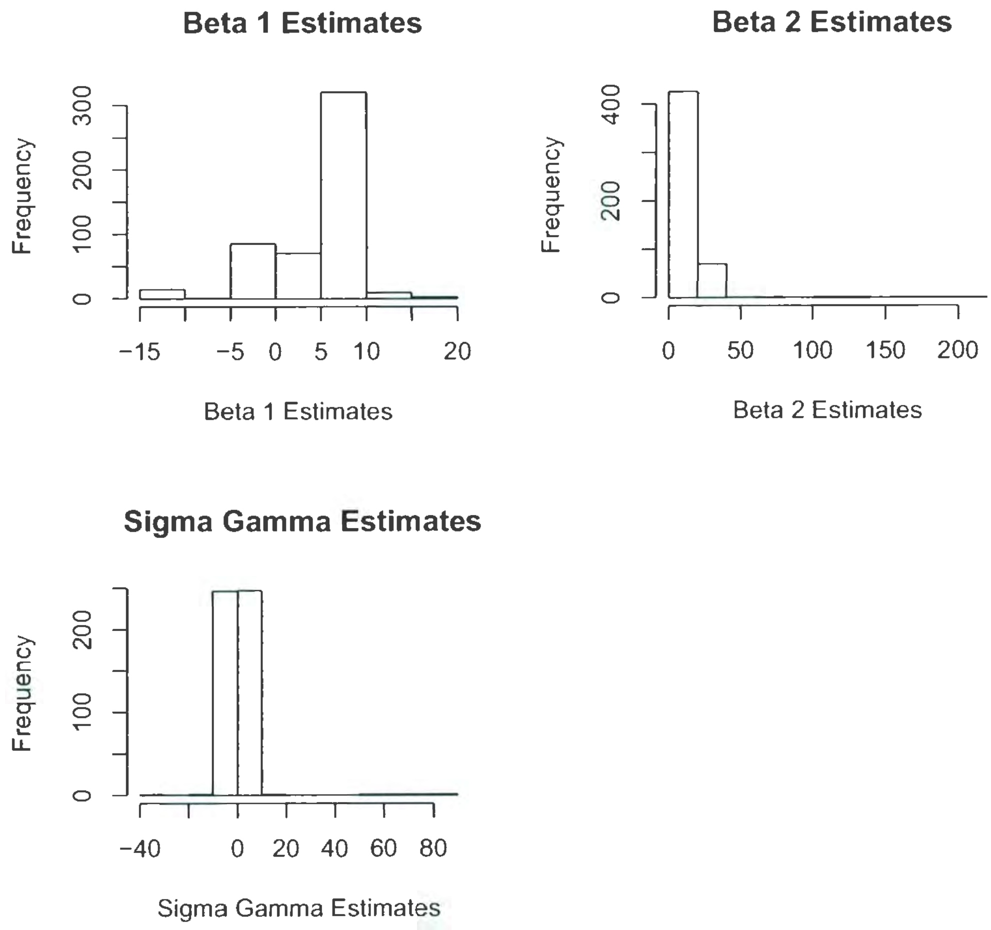


Table 3.6: Simulated Method Variances (GLMM)

Sim	β_1	β_2	σ_γ	$Var(\beta_1)$	$Var(\beta_2)$	$Var(\hat{\sigma}_\gamma)$
1	1.0	2.0	0.1	0.3482	0.3691	0.0411
2	1.0	2.0	0.3	0.4977	0.4668	0.0383
3	1.0	2.0	0.5	0.3508	0.3311	0.0391
4	1.0	2.0	0.7	0.2077	0.1891	0.0107
5	1.0	2.0	1.0	0.0620	0.0108	0.0413
6	1.0	2.0	2.0	0.0546	0.0230	0.0529
7	1.0	2.0	1.0	0.0486	0.0170	0.0452
8	1.0	0.1	0.1	0.0205	0.0061	0.0156
9	1.0	5.0	0.1	27.7602	203.9522	43.7322

- $\beta_1 = 0.1$, $\beta_2 = 0.2$, $\sigma_\gamma = 0.25$
- $\beta_1 = 0.1$, $\beta_2 = 0.2$, $\sigma_\gamma = 0.5$

All simulations were run and Table 3.7 shows the estimated values for β_1, β_2 and σ_γ. Table 3.8 shows the variances of the estimated values for β_1, β_2 and σ_γ. Also shown in both of these tables are the averages of only the positive values of $\hat{\sigma}_\gamma$.

Overall, for all simulations, the estimates of β_1 and β_2 were reasonable. However, as in the original results, all simulations produced poor results for $\hat{\sigma}_\gamma$. This is possible because, when treating σ_γ as a regression parameter, it is not restricted to be non-negative. In fact, all produced negative estimates. Since the average of the estimates were negative, we omitted the negative estimates and recalculated the mean based on the remaining positive estimates. This is shown in the final column of Table 3.7. Again, we can see that none of these produced good estimates.

Table 3.7: Simulated Method Estimates: Effects of Changes in X (GLMM)

Sim	Case	β_1	β_2	σ_γ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_\gamma$	$\hat{\sigma}_\gamma > 0$
1	1	0.1	0.2	0.25	0.1122	0.1879	-0.0091	0.0786
2	1	0.1	0.2	0.5	0.0936	0.2034	-0.0012	0.0891
3	1	1.0	2.0	0.25	1.0185	1.9670	-0.0067	0.0934
4	1	1.0	2.0	0.5	0.9775	1.9085	-0.0039	0.0964
5	2	0.1	0.2	0.25	0.0996	0.1989	-0.0050	0.0883
6	2	0.1	0.2	0.5	0.0926	0.1902	-0.0077	0.0890
7	2	1.0	2.0	0.25	1.1252	2.1229	0.0091	0.1559
8	2	1.0	2.0	0.5	1.0845	2.0480	0.0031	0.1577
9	3	0.1	0.2	0.25	0.1039	0.1865	-0.0069	0.0807
10	3	0.1	0.2	0.5	0.0991	0.1880	-0.0101	0.0846
11	3	1.0	2.0	0.25	1.0012	2.0266	-0.0077	0.1058
12	3	1.0	2.0	0.5	0.9668	1.9629	-0.0054	0.1114
13	4	0.1	0.2	0.25	0.1051	0.1853	-0.0051	0.0860
14	4	0.1	0.2	0.5	0.0991	0.1870	-0.0095	0.0821
15	4	1.0	2.0	0.25	1.0008	2.0286	-0.0051	0.1236
16	4	1.0	2.0	0.5	0.9548	1.9755	-0.0022	0.1273

Table 3.8: Simulated Method Variances: Effects of Changes in X (GLMM)

Sim	Case	β_1	β_2	σ_γ	$Var(\beta_1)$	$Var(\beta_2)$	$Var(\hat{\sigma}_\gamma)$	$Var(\hat{\sigma}_\gamma > 0)$
1	1	0.1	0.2	0.25	0.0776	0.2805	0.0117	0.0012
2	1	0.1	0.2	0.5	0.0686	0.2142	0.0131	0.0015
3	1	1.0	2.0	0.25	0.1069	0.4074	0.0150	0.0049
4	1	1.0	2.0	0.5	0.0953	0.3728	0.0158	0.0052
5	2	0.1	0.2	0.25	0.0178	0.0048	0.0131	0.0041
6	2	0.1	0.2	0.5	0.0206	0.0051	0.0145	0.0048
7	2	1.0	2.0	0.25	0.1886	0.4647	0.0389	0.0153
8	2	1.0	2.0	0.5	0.3508	0.3314	0.0394	0.0141
9	3	0.1	0.2	0.25	0.0107	0.0528	0.0112	0.0045
10	3	0.1	0.2	0.5	0.0113	0.0558	0.0128	0.0045
11	3	1.0	2.0	0.25	0.0163	0.1124	0.0172	0.0062
12	3	1.0	2.0	0.5	0.0169	0.1211	0.0187	0.0076
13	4	0.1	0.2	0.25	0.0378	0.1412	0.0118	0.0045
14	4	0.1	0.2	0.5	0.0397	0.1197	0.0125	0.0041
15	4	1.0	2.0	0.25	0.0687	0.3263	0.0257	0.0099
16	4	1.0	2.0	0.5	0.0749	0.3332	0.0257	0.0102

3.4 Quasi-Likelihood Method Analysis

As previously described, the quasi-likelihood method of estimation, unlike the maximum likelihood approach, does not require specification of the distribution of the response variable. This method uses only the mean and variance to try and estimate μ.

Sutradhar and Rao (2001) proposed an approach to the estimation of parameters of a generalized linear mixed model with two components of dispersion. This method was similar to two-way analysis of variance.

From (1.12), we cannot solve explicitly for β and σ_γ. Therefore, we need to estimate these parameters by solving (4.12) numerically with Newton's method. Initial estimates were chosen and used to start a Newton-Raphson iteration. Suppose $\hat{\beta}_{QL}$ and $\hat{\sigma}_{\gamma QL}$ denote solutions to (4.12); then, at iteration $(r+1)$:

$$\begin{pmatrix} \hat{\beta}_{QL(r+1)} \\ \hat{\sigma}_{\gamma,QL(r+1)} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_{QL(r)} \\ \hat{\sigma}_{\gamma,QL(r)} \end{pmatrix} + \left(\sum_{i=1}^{I} W_i^T \Lambda_i^{-1} W_i \right)_r^{-1} \left(\sum_{i=1}^{I} W_i^T \Lambda_i^{-1} (s_i - \lambda_i) \right)_r ,$$

$$\text{where } s_i = \begin{pmatrix} y_i \\ u_{i2} \end{pmatrix} \text{ where } y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in} \end{pmatrix},\ u_{i2} = \begin{pmatrix} y_{i1}y_{i2} \\ \vdots \\ y_{i,(n-1)}y_{in} \end{pmatrix},$$

$$\lambda_i = E(s_i) = E\begin{pmatrix} y_i \\ u_{i2} \end{pmatrix} = \begin{pmatrix} E(y_{ij}) \\ E(y_{ij}y_{ik})_{j \neq k} \end{pmatrix},$$

$$W_i = \begin{pmatrix} \frac{\partial \lambda_i^T}{\partial \beta^T} \\ \frac{\partial \lambda_i^T}{\partial \sigma_{\gamma}} \end{pmatrix} = \begin{pmatrix} \frac{\partial E(y_{ij})}{\partial \beta^T} \\ \frac{\partial E(y_{ij}y_{ik})_{j \neq k}}{\partial \beta^T} \\ \frac{\partial E(y_{ij})}{\partial \sigma_{\gamma}} \\ \frac{\partial E(y_{ij}y_{ik})_{j \neq k}}{\partial \sigma_{\gamma}} \end{pmatrix}, \text{ and}$$

$$\Lambda_i = Var(s_i) = Var\begin{pmatrix} y_i \\ u_{i2} \end{pmatrix} = \begin{pmatrix} \Omega_i & E_i \\ E_i^T & H_i \end{pmatrix}.$$

We shall further describe λ_i, W_i and Λ_i. Let us define the following:

$$\theta_{ij} = e^{x_{ij}^T \beta} \text{ and } p_i = e^{\sigma_\gamma \gamma_i}.$$

First of all, we will further describe λ_i by defining $E(y_i)$ and $E(y_{ij}y_{ik})_{j \neq k}$ as follows:

$$\begin{aligned} E(y_{ij}) &= \theta_{ij} E\left(\frac{p_i}{1+\theta_{ij}p_i} \right) \\ E(y_{ij}y_{ik})_{j \neq k} &= \theta_{ij}\theta_{ik} E\left(\frac{p_i^2}{(1+\theta_{ij}p_i)(1+\theta_{ik}p_i)} \right) \end{aligned}$$

Next, let us further define W_i by:

$$
\begin{aligned}
\frac{\partial E(y_{ij})}{\partial \beta^T} &= x_{ij}\theta_{ij} E\left(\frac{p_i}{1+\theta_{ij}p_i}\right) - x_{ij}\theta_{ij}^2 E\left(\frac{p_i^2}{(1+\theta_{ij}p_i)^2}\right), \\
\frac{\partial E(y_{ij}y_{ik})}{\partial \beta^T} &= \theta_{ij}\theta_{ik} E\left\{\frac{2\gamma_i p_i^2 + \gamma_i\theta_{ij}p_i^3 + \gamma_i\theta_{ik}p_i^3}{(1+\theta_{ij}p_i)^2(1+\theta_{ik}p_i)^2}\right\}, \\
\frac{\partial E(y_{ij})}{\partial \sigma_{\gamma}} &= \theta_{ij} E\left(\frac{\gamma_i p_i}{(1+\theta_{ij}p_i)^2}\right), \\
\frac{\partial E(y_{ij}y_{ik})}{\partial \sigma_{\gamma_i}} &= \theta_{ij}\theta_{ik}\left\{(x_{ij}+x_{ik}) E\left(\frac{p_i^2}{(1+\theta_{ij}p_i)(1+\theta_{ik}p_i)}\right)\right. \\
&\quad \left. - E\left(\frac{x_{ik}\theta_{ij}p_i^3 + x_{ik}\theta_{ik}p_i^3 + x_{ij}\theta_{ij}\theta_{ik}p_i^4 + x_{ik}\theta_{ij}\theta_{ik}p_i^4}{(1+\theta_{ij}p_i)^2(1+\theta_{ik}p_i)^2}\right)\right\}.
\end{aligned}
$$

Finally, we will further define Λ_i using Ω_i, E_i and H_i as follows:

$$
\Omega_i = Var(y_i) = Var\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in} \end{pmatrix} = \begin{bmatrix} Var(y_{i1}) & Cov(y_{i1}, y_{i2}) & \cdots & Cov(y_{i1}, y_{in}) \\ Cov(y_{i2}, y_{i1}) & Var(y_{i2}) & \cdots & Cov(y_{i2}, y_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(y_{in}, y_{i1}) & Cov(y_{in}, y_{i2}) & \cdots & Var(y_{in}) \end{bmatrix},
$$

$$
\begin{aligned}
Var(y_{ij}) &= \theta_{ij} E\left(\frac{p_i}{1+\theta_{ij}p_i}\right) - \theta_{ij}^2\left[E\left(\frac{p_i}{1+\theta_{ij}p_i}\right)\right]^2, \\
Cov(y_{ij}, y_{ik}) &= \theta_{ij}\theta_{ik} E\left(\frac{p_i^2}{(1+\theta_{ij}p_i)(1+\theta_{ik}p_i)}\right) \\
&\quad - \theta_{ij}\theta_{ik} E\left(\frac{p_i}{1+\theta_{ij}p_i}\right) E\left(\frac{p_i}{1+\theta_{ik}p_i}\right) \qquad j \neq k,
\end{aligned}
$$

$$E_i = Cov(y_i, u_{i2}) = Cov\begin{pmatrix} y_{i1} & y_{i1}y_{i2} \\ y_{i2} & y_{i1}y_{i3} \\ \vdots & \vdots \\ y_{in} & y_{i,(n-1)}y_{in} \end{pmatrix}$$

$$= \begin{bmatrix} Cov(y_{i1}, y_{i1}y_{i2}) & Cov(y_{i1}, y_{i1}y_{i3}) & \cdots & Cov(y_{i1}, y_{i,(n-1)}y_{in}) \\ Cov(y_{i2}, y_{i1}y_{i2}) & Cov(y_{i2}, y_{i1}y_{i3}) & \cdots & Cov(y_{i2}, y_{i,(n-1)}y_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(y_{in}, y_{i1}y_{i2}) & Cov(y_{in}, y_{i1}y_{i3}) & \cdots & Cov(y_{in}, y_{i,(n-1)}y_{in}) \end{bmatrix}.$$

$$Cov(y_{ij}, y_{ij}y_{ik}) = \theta_{ij}\theta_{ik}E\left(\frac{p_i^2}{(1+\theta_{ij}p_i)(1+\theta_{ik}p_i)}\right)$$
$$-\theta_{ij}^2\theta_{ik}E\left(\frac{p_i}{1+\theta_{ij}p_i}\right)E\left(\frac{p_i^2}{(1+\theta_{ij}p_i)(1+\theta_{ik}p_i)}\right) \quad j<k,$$

$$Cov(y_{ij}, y_{ik}y_{il}) = \theta_{ij}\theta_{ik}\theta_{il}E\left(\frac{p_i^3}{(1+\theta_{ij}p_i)(1+\theta_{ik}p_i)(1+\theta_{il}p_i)}\right)$$
$$-\theta_{ij}\theta_{ik}\theta_{il}E\left(\frac{p_i}{(1+\theta_{ij}p_i)}\right)E\left(\frac{p_i^2}{(1+\theta_{ik}p_i)(1+\theta_{il}p_i)}\right) \quad j\neq k\neq l, k<l.$$

$$H_i = Var(u_{i2}) = Var\begin{pmatrix} y_{i1}y_{i2} \\ y_{i1}y_{i3} \\ \vdots \\ y_{i,(n-1)}y_{in} \end{pmatrix}$$

$$= \begin{bmatrix} Var(y_{i1}y_{i2}) & Cov(y_{i1}y_{i2}, y_{i1}y_{i3}) & \cdots & Cov(y_{i1}y_{i2}, y_{i,(n-1)}y_{in}) \\ Cov(y_{i1}y_{i3}, y_{i1}y_{i2}) & Var(y_{i1}y_{i3}) & \cdots & Cov(y_{i1}y_{i3}, y_{i,(n-1)}y_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(y_{i,(n-1)}y_{in}, y_{i1}y_{i2}) & Cov(y_{i,(n-1)}y_{in}, y_{i1}y_{i3}) & \cdots & Var(y_{i,(n-1)}y_{in}) \end{bmatrix}.$$

$$Var(y_{ij}y_{ik}) = \theta_{ij}\theta_{ik}E\left(\frac{p_i^2}{(1+\theta_{ij}p_i)(1+\theta_{ik}p_i)}\right)$$

$$\theta_{ij}^2\theta_{ik}^2\left[E\left(\frac{p_i^2}{(1+\theta_{ij}p_i)(1+\theta_{ik}p_i)}\right)\right]^2 \qquad j < k.$$

$$Cov(y_{ij}y_{ik}, y_{ij}y_{il}) = \begin{cases} \theta_{ij}\theta_{ik}\theta_{il}E\left(\frac{p_i^3}{(1+\theta_{ij}p_i)(1+\theta_{ik}p_i)(1+\theta_{il}p_i)}\right) \\ \quad -\theta_{ij}^2\theta_{ik}\theta_{il}E\left(\frac{p_i^2}{(1+\theta_{ij}p_i)(1+\theta_{ik}p_i)}\right)E\left(\frac{p_i^2}{(1+\theta_{ij}p_i)(1+\theta_{il}p_i)}\right) \\ \quad j < k, j < l, k \neq l, \\ Cov(y_{ij}y_{ik}, y_{il}y_{ij}) \quad j < k, l < j, k \neq l \\ Cov(y_{ik}y_{ij}, y_{ij}y_{il}) \quad k < j, j < l, k \neq l \\ Cov(y_{ik}y_{ij}, y_{il}y_{ij}) \quad k < j, l < j, k \neq l. \end{cases}$$

$$\begin{aligned} Cov(y_{ij}y_{ik}, y_{il}y_{im}) &= \theta_{ij}\theta_{ik}\theta_{il}\theta_{im}E\left(\frac{p_i^4}{(1+\theta_{ij}p_i)(1+\theta_{ik}p_i)(1+\theta_{il}p_i)(1+\theta_{im}p_i)}\right) \\ &\quad -\theta_{ij}\theta_{ik}\theta_{il}\theta_{im}E\left(\frac{p_i^2}{(1+\theta_{ij}p_i)(1+\theta_{ik}p_i)}\right)E\left(\frac{p_i^2}{(1+\theta_{il}p_i)(1+\theta_{im}p_i)}\right) \\ &\quad j < k, l < m, j \neq l, k \neq m. \end{aligned}$$

As with the moment method, some of the expectations in this study cannot be found explicitly. Thus, they are approximated in the same manner as discussed earlier.

A few problems were occurring with the value of $\hat{\sigma}_\gamma$ becoming too large in random simulations and thus resulting in errors. To overcome this problem we needed to add a line of code to terminate the procedure if the value of $\hat{\sigma}_\gamma$ became too large (larger than 5).

All simulations were run and Table 3.9 shows the estimated values for β_1, β_2 and σ_γ. Table 3.10 shows the variances of the estimated values for β_1, β_2 and σ_γ. Included in

Table 3.9 is a column containing the number of simulations (out of 500) that broke a Newton Raphson iteration due to the value of σ_γ being too large (larger than 5). The number of breaks is in the table under the column NA.

First to note is that this program did not run for two of the simulations (simulation 8 - β_1=1, β_2=2, σ_γ=4; simulation 10 - β_1=1, β_2=5, σ_γ=0.1). Errors were produced for both of these simulations that were results of the program attempting to take the inverse of a singular matrix.

For all of the simulations that produced no errors, the method performed well in estimating β_1, as we can see in Table 3.9. We can see, both in the estimates and the variances for $\hat{\beta}_1$ that the method appears to become more biased as the value of σ_γ becomes larger. Especially we note that simulation 7 has 71 estimates that are broken due to $\hat{\sigma}_\gamma$ becoming too large. If we look at the medians, means and trimmed means of all β_1 and β_2 estimates along with the histograms of these estimates, there are a few sets of estimates that do not appear approximately normally distributed.

The method always gives a biased estimate for β_2, as we can see in Table 3.9. All of the variances are small, which can be seen in Table 3.10. The values of $\hat{\beta}_1$ imply that the program produces poorer results as σ_γ becomes larger. If we look at the medians, means and trimmed means of all β_1 and β_2 estimates along with the histograms of these estimates, there are a few sets of estimates that do not appear approximately normally distributed but this is due outliers. For the estimation of β, this method appears to be very sensitive to the choice of β since it performed very well for β_1 1.0, β_2 = 0.1, and σ_γ = 0.1, but not well for β_1 1.0, β_2 5.0, and σ_γ 0.1.

This program did not perform well in estimating σ_γ. There were only a few simulations that did well, as we can see from Table 3.9. The program performed worse when using smaller and larger values of σ_γ. If we look at the number of breaks due to $\hat{\sigma}_\gamma$ being too large, we can see that it becomes an issue for higher values of σ_γ. The program could not calculate estimates for the highest value of σ_γ ($\sigma_\gamma = 4$). The estimation becomes worse as the value of σ_γ becomes larger. If we look at the medians, means and trimmed means of all σ_γ estimates along with the histograms of these estimates, there are a few sets of estimates that do not appear approximately normally distributed.

For an example of the histograms in which $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\sigma}_\gamma$ perform poorly, we can refer to the histograms in Figure 3.6. For an example of the histograms in which $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\sigma}_\gamma$ perform well, refer to the histograms in Figure 3.5.

Table 3.9: Quasi-Likelihood Method Estimates (GLMM)

Sim	β_1	β_2	σ_γ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_\gamma$	NA
1	1.0	2.0	0.01	1.0013	1.8766	0.1506	0
2	1.0	2.0	0.1	1.0301	1.8756	0.1868	0
3	1.0	2.0	0.3	1.0396	1.8669	0.2945	0
4	1.0	2.0	0.5	1.0615	1.8382	0.4634	0
5	1.0	2.0	0.7	1.0705	1.8211	0.6429	0
6	1.0	2.0	1	1.0481	1.8313	0.9171	0
7	1.0	2.0	2	0.8675	1.8632	1.7926	74
8	1.0	2.0	4		ERROR		
9	1.0	0.1	0.1	0.9415	0.0960	0.0864	0
10	1.0	5.0	0.1		ERROR		

For all of the simulations conducted up to this point, the following X matrix was

Figure 3.5: Quasi-Likelihood Method Histograms (GLMM): Simulation 1

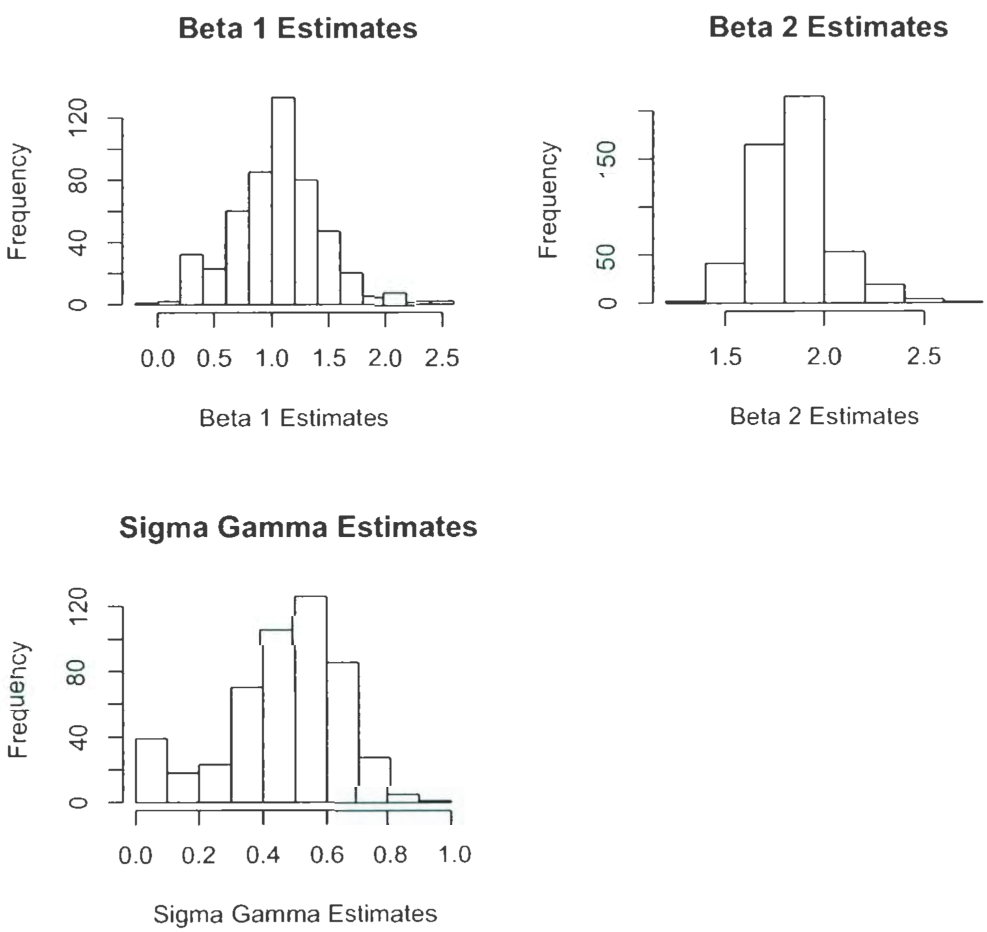


Figure 3.6: Quasi-Likelihood Method Histograms (GLMM): Simulation 7

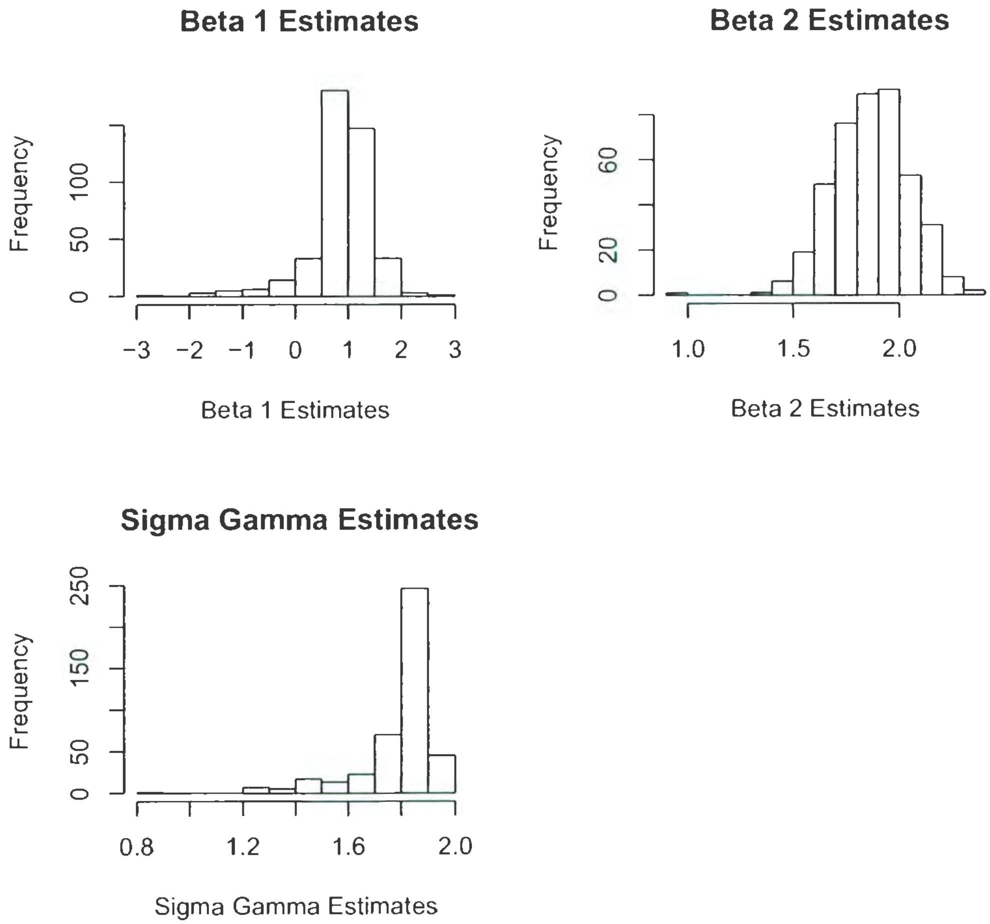


Table 3.10: Quasi-Likelihood Method Variances (GLMM)

Sim	β_1	β_2	σ_γ	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$	$Var(\hat{\sigma}_\gamma)$
1	1.0	2.0	0.01	0.06839	0.03195	0.03119
2	1.0	2.0	0.1	0.07856	0.03678	0.03823
3	1.0	2.0	0.3	0.10135	0.04281	0.04123
4	1.0	2.0	0.5	0.15907	0.03556	0.03552
5	1.0	2.0	0.7	0.21539	0.03503	0.02587
6	1.0	2.0	1	0.19747	0.03133	0.00921
7	1.0	2.0	2	0.35634	0.03103	0.02124
8	1.0	2.0	4		ERROR	
9	1.0	0.1	0.1	0.00353	0.00090	0.00162
10	1.0	5.0	0.1		ERROR	

used:

$$X_i = \begin{cases} \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix} & \text{if } i = 1, \cdots, 50; \\ \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} & \text{if } i = 51, \cdots, 100. \end{cases}$$

As with the moment method, we would like to see if the method is sensitive to the choice of X. We evaluated the method using the same four X matrices as before and each using the same four sets of parameter values, leaving us with 16 simulations in total to compare. Table 3.11 shows the estimated values for β_1, β_2 and σ_γ. Table 3.12 shows the variances of the estimated values for β_1, β_2 and σ_γ.

Looking at Table 3.11, it is interesting to note that the quasi-likelihood performed much better for the X matrix that we have been using all along for our programs.

All other X matrices produced errors for at least one of the four simulations ran. It is also interesting to note that for all four X matrices used, the simulations that had smaller β values consistently gave better results, producing either less biased results or fewer computing problems. Only the X matrix that we have been using all along produced good estimates for the larger β (β_1=1, β_2=2) values. All others either produced an error or had a significant number of breaks due to σ_γ becoming too large, which is indicated by the column NA. Therefore, for β_1, β_2 and σ_γ, we can see that the choice of X and β has an effect on the quality of the estimates.

Table 3.11: Quasi-Likelihood Method Estimates: Effects of Changes in X (GLMM)

Sim	Case	β_1	β_2	σ_γ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_\gamma$	NA
1	1	0.1	0.2	0.25	0.0915	0.1961	0.2228	0
2	1	0.1	0.2	0.5	0.0903	0.1931	0.4522	0
3	1	1.0	2.0	0.25	0.9711	1.8321	0.4080	51
4	1	1.0	2.0	0.5		ERROR		
5	2	0.1	0.2	0.25	0.0981	0.1876	0.2235	0
6	2	0.1	0.2	0.5	0.0956	0.1884	0.4551	0
7	2	1.0	2.0	0.25	1.0359	1.8713	0.2566	0
8	2	1.0	2.0	0.5	1.0615	1.8382	0.4634	0
9	3	0.1	0.2	0.25	0.0938	0.1835	0.2240	0
10	3	0.1	0.2	0.5	0.0944	0.1831	0.4529	0
11	3	1.0	2.0	0.25		ERROR		
12	3	1.0	2.0	0.5		ERROR		
13	4	0.1	0.2	0.25	0.0955	0.1824	0.2263	0
14	4	0.1	0.2	0.5	0.0974	0.1817	0.4539	0
15	4	1.0	2.0	0.25		ERROR		
16	4	1.0	2.0	0.5		ERROR		

Table 3.12: Quasi-Likelihood Method Variances: Effects of Changes in X (GLMM)

Sim	Case	β_1	β_2	σ_γ	$Var(\beta_1)$	$Var(\beta_2)$	$Var(\hat{\sigma}_\gamma)$
1	1	0.1	0.2	0.25	0.0114	0.0465	0.0015
2	1	0.1	0.2	0.5	0.0126	0.0519	0.0005
3	1	1.0	2.0	0.25	0.3574	1.0567	0.1257
4	1	1.0	2.0	0.5		ERROR	
5	2	0.1	0.2	0.25	0.00304	0.00081	0.0017
6	2	0.1	0.2	0.5	0.00347	0.00091	0.00056
7	2	1.0	2.0	0.25	0.09462	0.04388	0.04246
8	2	1.0	2.0	0.5	0.21539	0.03503	0.02587
9	3	0.1	0.2	0.25	0.00163	0.00931	0.00142
10	3	0.1	0.2	0.5	0.00201	0.01176	0.00053
11	3	1.0	2.0	0.25		ERROR	
12	3	1.0	2.0	0.5		ERROR	
13	4	0.1	0.2	0.25	0.0061	0.0175	0.00158
14	4	0.1	0.2	0.5	0.00654	0.0180	0.00054
15	4	1.0	2.0	0.25		ERROR	
16	4	1.0	2.0	0.5		ERROR	

Chapter 4

Real Data Analysis

4.1 Introduction

A longditudinal study was conducted on 180 people from 48 different families by St. John's General Hospital to monitor how many times they visited a physician each year over the years 1985 - 1990. Other information was collected from these individuals at the beginning of the study: their age, their gender, the number of chronic conditions they had, and their education level. The purpose of this analysis is to determine, when accounting for the correlation among family members, which variables, if any, are related to whether a person visits a physician.

As stated above, our response variable is an indicator of whether a subject visited a physician in a particular year. The covariates are gender, age (in years, ranging from 19.9 to 85.2), number of chronic conditions at the start of the study, and education level.

4.2 Exploratory Analysis

Analysis of the physician visit data consisted of looking at summary statistics of all of the data along with some plots.

4.2.1 Summary Statistics

First of all, we looked at some summary statistics of the data. We looked first at the variable of interest, the indicator variable of whether or not a subject visits a physician in a given year, for years 1985 or 1990. As we can see from Table 4.1, the proportion of people that visit a physician is similar each year. Of the subjects in this study, on average about 134 (74%) visit a physician and 46 (26%) do not visit a physician in a given year.

Table 4.1: Annual Number of Visits to a Physician

Year	n	%
1985	141	78
1986	132	73
1987	136	76
1988	139	77
1989	130	72
1990	126	70
Average	134	74

We next looked at the summary statistics for all of the explanatory variables. We can see from Table 4.2 that the youngest person in this study is 19 and the oldest person is 85. The average age is about 39. There are more younger people in this study, since the mean is closer to the younger age than to the older. Also, this is indicated by the median being smaller than the mean, suggesting that the age data is skewed right. Almost half (46%) of the subjects in this study were under the age of 30. Very few were over the age of 75 (2%). Over half (53%) of the subjects included in the study were males.

Table 4.2: Age

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Var.
19.90	22.98	33.20	38.60	51.28	85.20	275.3463

Table 4.3 shows the summary table of the education level variable. The education level of the subjects included in this study seem to be dispersed evenly. Most people (32%) have the lowest level of education. However, the second largest education level group, which includes 26% of the subjects, is at the highest level of education.

Table 4.3: Education Level

	n	%
1 = Lowest	58	32
2	33	18
3	42	23
4 = Highest	47	26

Table 4.4 shows the summary table of the number of chronic conditions. The majority of subjects (80%) included in this study has either none or only one chronic condition in 1985.

The histograms of all of the above data will be discussed in the next section.

Table 4.4: Number of Chronic Conditions

	n	%
0	88	49
1	56	31
2	16	09
3	10	06
4	8	04
5	2	01

4.2.2 Graphs

The next step in the exploratory analysis was to examine different graphs of our dependent and independent variables. Figure 4.1 shows the physician visit indicator variable for 1985. The other years have similar plots and are not shown.

We examined histograms of each of the explanatory variables (Figure 4.2). The majority of people that are included in this study are 20 to 30 years of age. The ages of the rest of the people included follow a normal distribution.

After looking at histograms of the explanatory variables, we looked at spineplots of the explanatory variables by the physician visit indicator for each year from 1985 to 1990. The plot of gender versus the physician visit indicator variable for 1985 is in Figure 4.3. The plots for the other years are similar and are not shown. As we can see, there appears to be a relationship between whether or not a person visits a physician and the gender of that person. The shaded area for visiting a physician (1) is larger for females (2) in all six of the plots, implying that females may visit a physician more than males.

The spineplot of age versus the physician visit indicator variable for 1985 is displayed

Figure 4.1: 1985 Physician Visit Indicator Variable Histogram

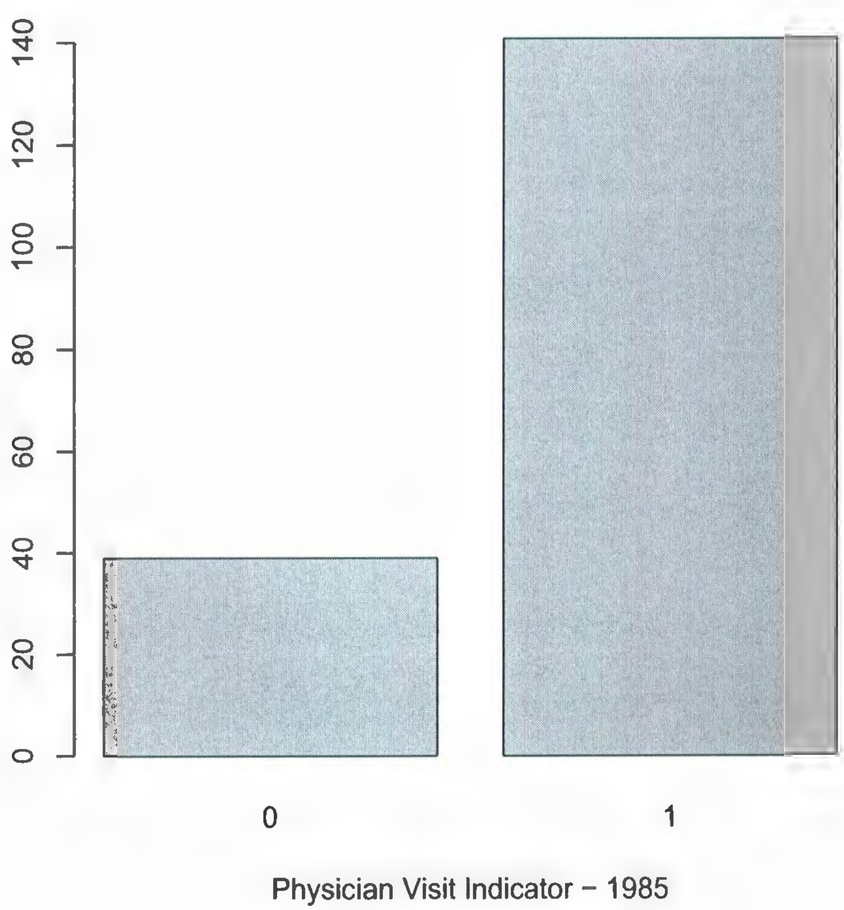


Figure 4.2: Explanatory Variable Histograms

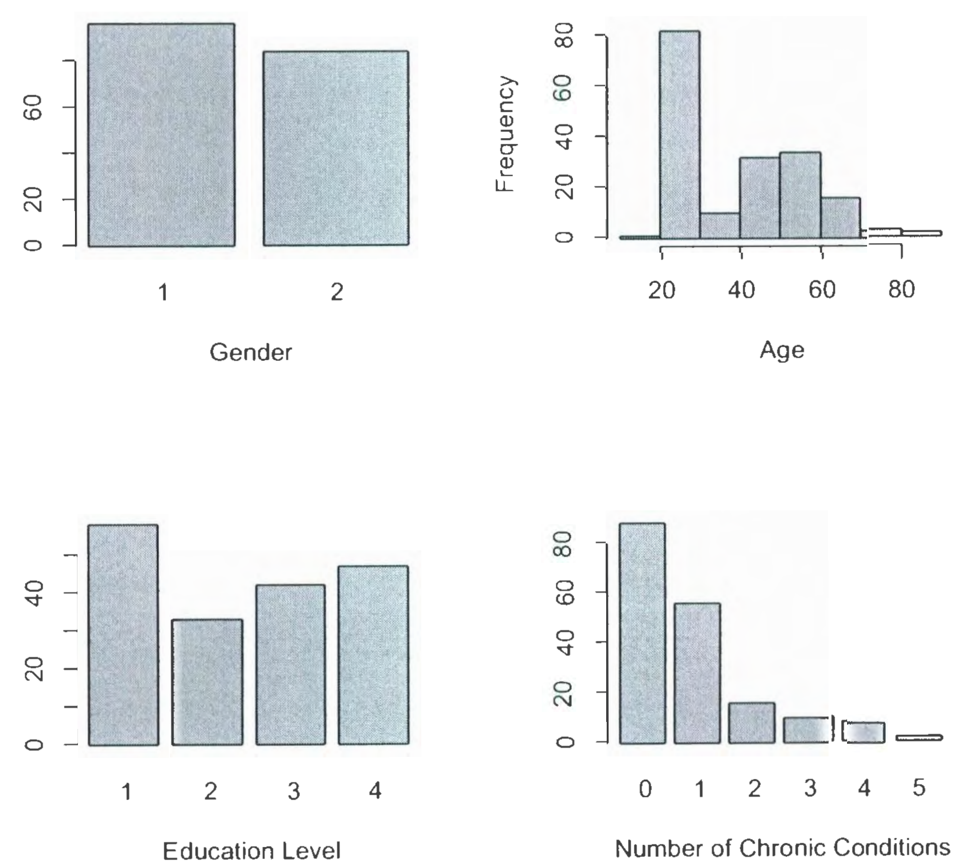
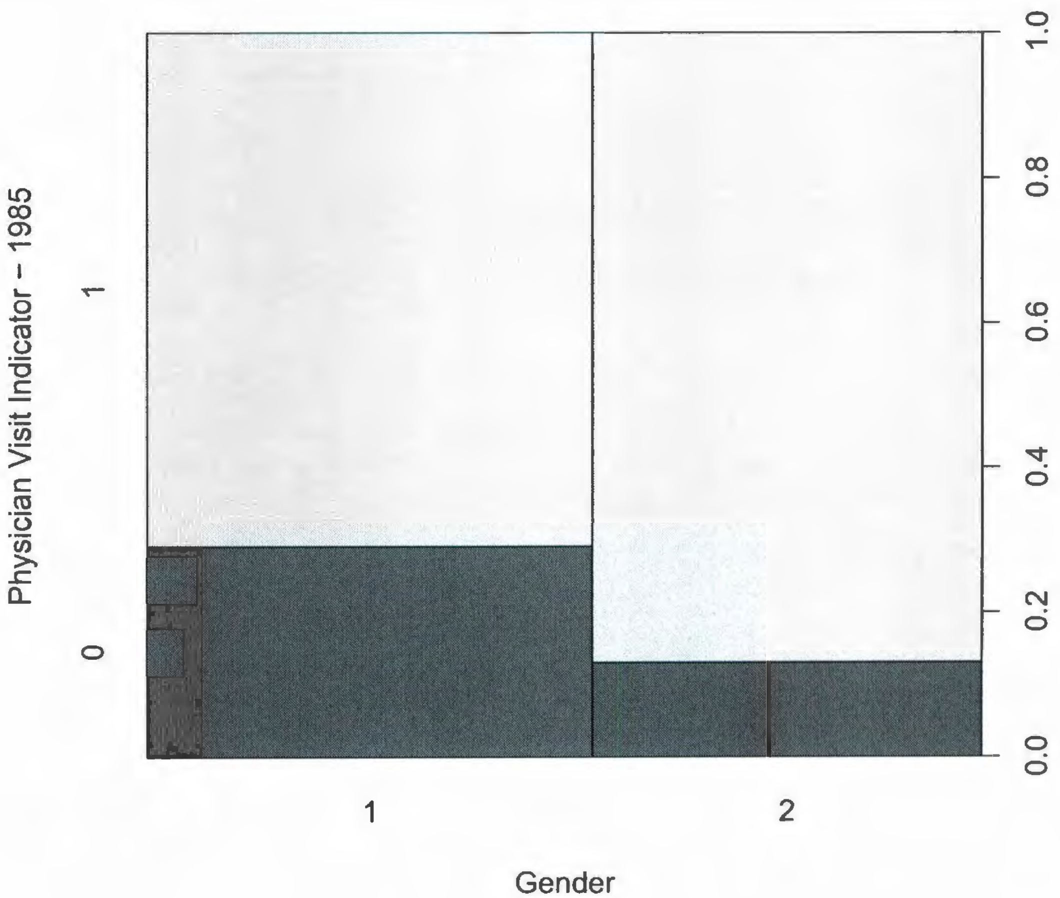


Figure 4.3: Gender by the 1985 Physician Visit Indicator Variable Spineplot



in Figure 4.4. The plots for other years are similar and are not shown. As we can see, it appears that people between the ages of 40 and 50 have more of a chance of visiting a physician at least once in a given year for each of the years from 1985 to 1990.

The spineplot of education level versus the physician visit indicator variable for 1985 is displayed in Figure 4.5. The plots for other years are similar and are not shown. For all plots, people with education level 2 and 3 have a higher tendency to visit a physician at least once in a given year for each of the years from 1985 to 1990.

Finally, the spineplot of the number of chronic conditions that a person has in 1985 versus the physician visit indicator variable for 1985 is displayed in Figure 4.6. The plots for other years are similar and are not shown. The graphs imply that the more chronic conditions a person has, the higher the tendency for that person to visit a physician in a given year for the years from 1985 to 1990.

Figure 4.4: Age by the 1985 Physician Visit Indicator Variable Spineplot

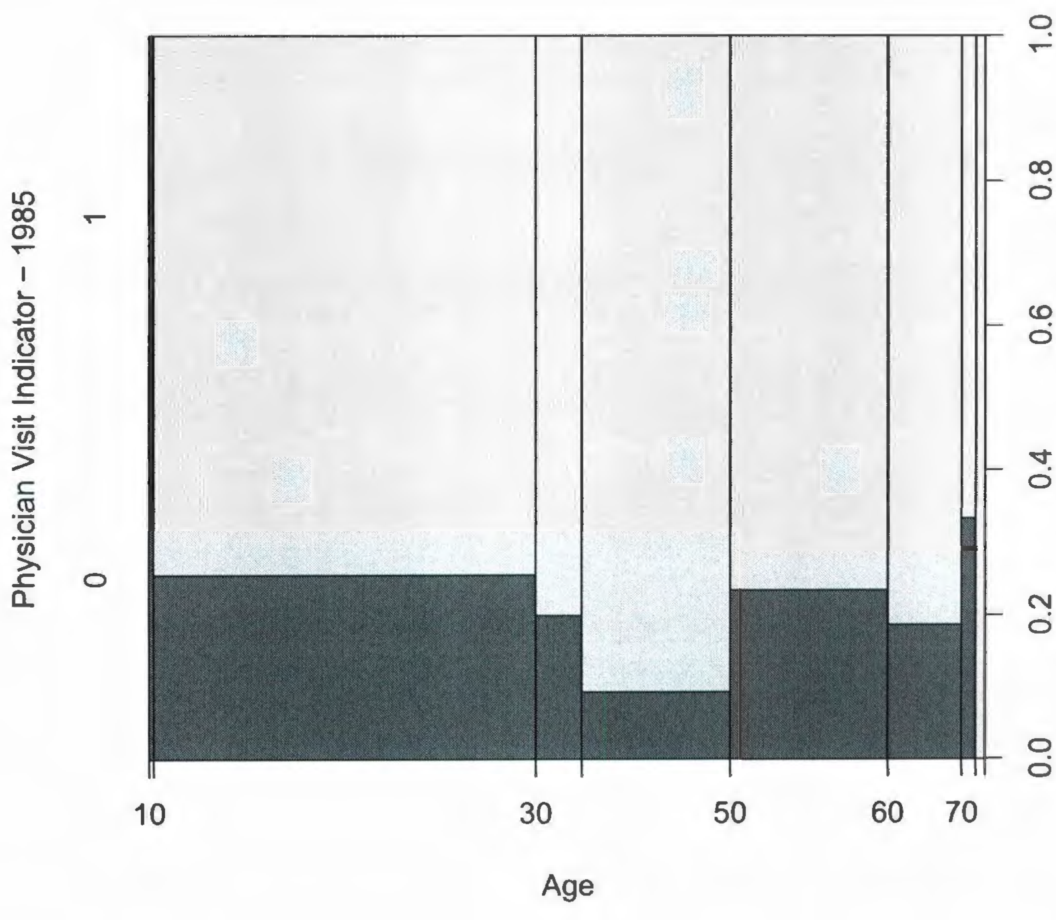


Figure 4.5: Education Level by the 1985 Physician Visit Indicator Variable Spineplot

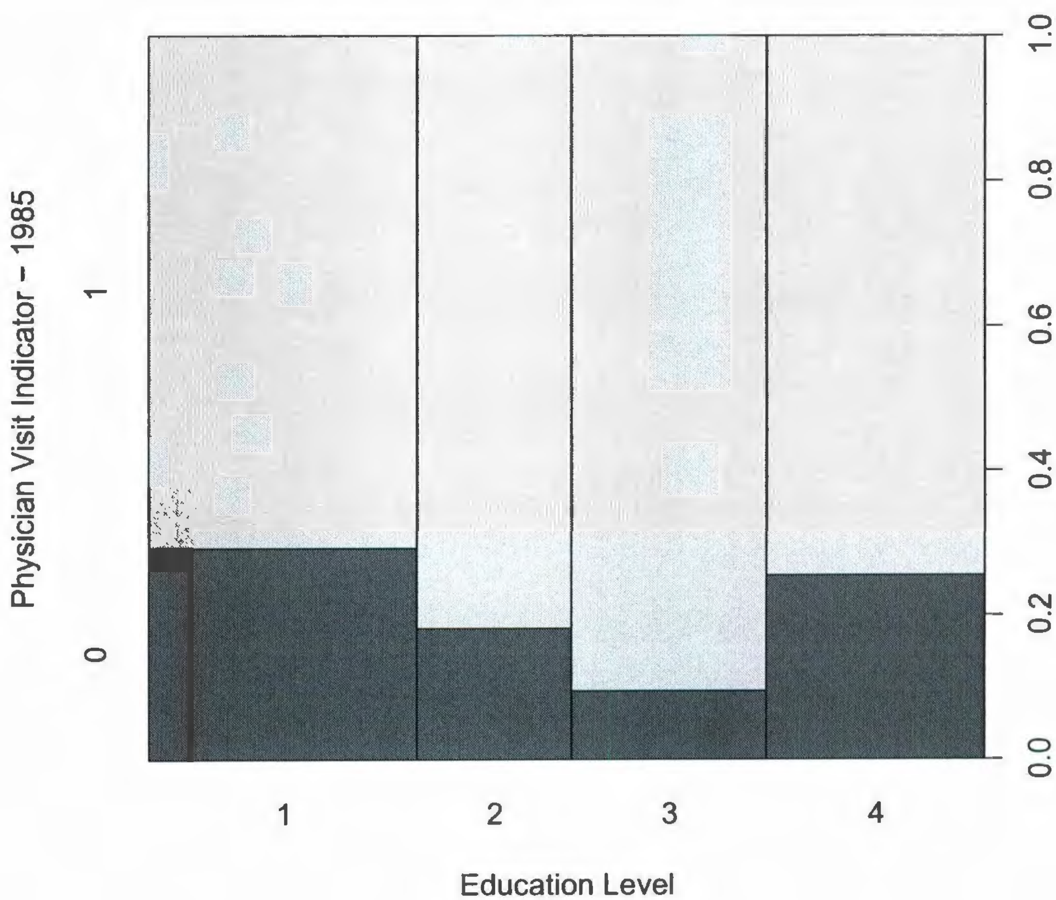
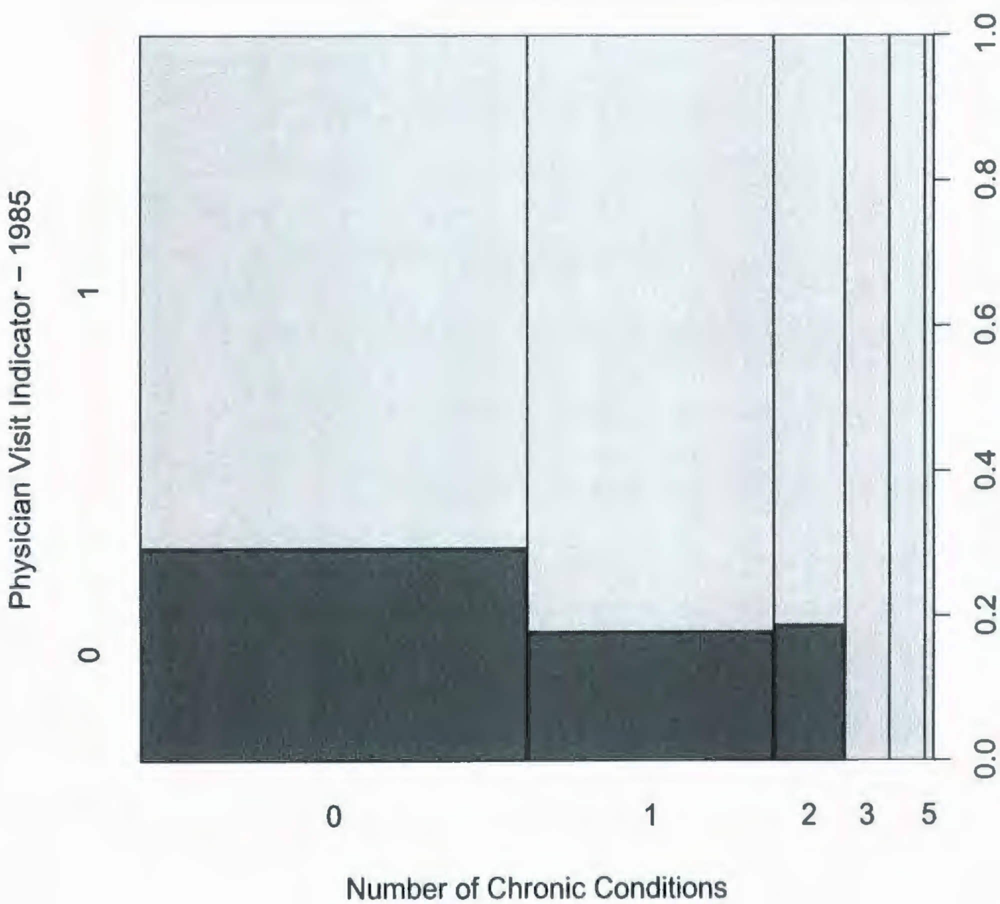


Figure 4.6: Number of Chronic Conditions by the 1985 Physician Visit Indicator Variable Spineplot



4.3 Data Analysis

For this study, we were interested in looking at how our explanatory variables (gender, age, education level, and number of chronic conditions in 1985) are related to whether a person visits a physician in a given year for the years from 1985 to 1990.

As this data is clustered data, there may be a correlation among family members with respect to whether or not they visit a physician. Therefore, we want to account for the correlation that may be present among family members. As such, we need to use a mixed-effects model. We assumed that the correlation between the family members is the same for all pairs of members, and is constant across families. We assume that the families are independent.

Also, as our response variable is an indicator variable, we needed to use a generalized linear model, since the response is binary (1 if the person visited a physician and 0 if the person did not visit a physician). Furthermore, since we have data for whether or not a person visited a physician for six years in a row, we wanted to analyze each year separately.

These data were analyzed using a generalized linear model for each of the six different years of data provided. It was analyzed using three methods for each year to see how the results of these methods compare.

- Fixed-effects logistic regression. This ignores any correlation among family members.
- Moment method described in Chapter 3.
- Quasi-likelihood method described in Chapter 3.

For the fixed-effects logistic regression analysis of this data, we assumed that whether a person visits a physician in a given year follows a Binomial distribution and we fit the following logistic model:

$$\text{logit}(\mu) = \beta_0 + \beta_1 \text{ gen} + \beta_2 \text{ chron} + \beta_3 \text{ educ1} + \beta_4 \text{ educ2} + \beta_5 \text{ educ3} + \beta_6 \text{ age},$$

where μ is the probability that the subject visits a physician. As the education level variable is categorical, it is modelled with three indicator variables. For the moment and quasi-likelihood method we analyzed the mixed model:

$$\text{logit}(\mu) = \beta_0 + \beta_1 \text{ gen} + \beta_2 \text{ chron} + \beta_3 \text{ educ1} + \beta_4 \text{ educ2} + \beta_5 \text{ educ3} + \beta_6 \text{ age} + \gamma_i.$$

4.3.1 Fixed-Effects Logistic Regression

The first analysis that was conducted on the physician visit data was to perform logistic regression for each year in which the data was collected.

Dependent Variable: Physician Visit Indicator for 1985

When looking at the results for the first model, it looked as though we may be able to drop the group of education terms. To see if we could drop terms, we used two different tests. To drop an individual term, we would use a z test and to drop a group of terms, we would use a drop in deviance test. The z test and drop in deviance tests are described below.

z Test to Drop Individual Terms

$H_o: \beta_i = 0$ $H_a: \beta_i \neq 0$	
$z_{obs} = \frac{\beta_i}{s_{\beta_i}}$	p-value $= 2 * P(Z > \lvert z_{obs} \rvert)$, using the z-distribution

Drop in Deviance Test to Drop Groups of Terms

$H_o: \beta_{q+1} = \beta_{q+2} = \cdots = \beta_p = 0$ H_a : at least one $\beta_i \neq 0$	
$F = \frac{(Deviance_R - Deviance_F)/(df_R - df_F)}{Deviance_F/df_F}$	p-value $= P(F > F_{obs})$, using the F distribution with $df_F - df_R$ and df_F degrees of freedom

The test to drop all of the education terms is as follows:

$$H_0: \beta_3 = \beta_4 = \beta_5 = 0$$
$$H_a: \text{at least one of } \beta_3 = \beta_4 = \beta_5 \neq 0$$
$$F_{obs} = \frac{(172.89 - 164.77)/(176 - 173)}{164.77/173} = 2.84186$$
$$\text{p-value} = P(F > F_{obs}) = 0.829 \text{ with 3 and 176 df.}$$

There is little evidence against H_0 since $0.829 > 0.05$, so we should drop all of the proposed terms.

With this new model, we obtain a z-value of -0.277 and a p-value of 0.78, which is greater than 0.05, for the variable age. Therefore, we can also drop this term from our model.

We now have a reduced model that contains only the gender and number of chronic conditions variables. The reduced model has an AIC of 177.22 and is as follows:

$$\boxed{\text{logit}(\mu) = -2.840 + 0.8408 \text{ gen} + 0.5992 \text{ chron}}$$

Therefore, in 1985, we can see that the odds of females visiting a physician is estimated to be $exp(0.8408) = 2.318$ times as large as the odds males visiting a physician.

Also, if the number of chronic conditions increases by one, the odds of visiting a physician are $exp(0.5992) = 1.821$ times higher.

Dependent Variable: Physician Visit Indicator for 1986

As in the analysis of the 1985 data, we get the following initial and final models.

Initial model:
logit(μ) = -1.0010 + 1.2702 gen + 0.1398 chron - 0.0260 educ1 + 0.4772 educ2 + 0.9227 educ3 - 0.0084 age.

Final model:
logit(μ) = -1.0934 + 1.3178 gen + 0.3836 chron.

Therefore, in 1986, females were more likely to see a physician than males. Having more chronic conditions at the beginning of the study is also related to being more likely to visit a physician.

Dependent Variable: Physician Visit Indicator for 1987

As in the analysis of the 1985 data, we get the following initial and final models.

Initial model:

$\text{logit}(\mu) = -0.6839 + 1.0759\ \text{gen} + 0.4928\ \text{chron} - 0.8032\ \text{educ1} + 0.3038\ \text{educ2} + 0.3755\ \text{educ3} + 0.0073\ \text{age}$.

Final model: $\text{logit}(\mu) = -0.5993 + 1.2459\ \text{gen}$.

Therefore, in 1987, females were more likely to see a physician than males.

Dependent Variable: Physician Visit Indicator for 1988

As in the analysis of the 1985 data, we get the following initial and final models.

Initial model:
$\text{logit}(\mu) = -1.4037 + 1.4078\ \text{gen} + 0.3901\ \text{chron} - 0.8270\ \text{educ1} + 0.0279\ \text{educ2} + 0.1030\ \text{educ3} + 0.0188\ \text{age}$.

Final model: $\text{logit}(\mu) = -0.9580 + 1.6047\ \text{gen}$.

Therefore, in 1988, females were more likely to see a physician than males.

Dependent Variable: Physician Visit Indicator for 1989

As in the analysis of the 1985 data, we get the following initial and final models.

Initial model:
$\text{logit}(\mu) = -1.1760 + 0.8494\ \text{gen} + 0.5556\ \text{chron} - 0.1086\ \text{educ1} + 1.3285\ \text{educ2} +$

0.5962 educ3 + 0.0071 age.

Final model: logit(μ) = 0.5062 + 0.6464 chron.

Therefore, in 1989, having more chronic conditions at the beginning of the study related to being more likely to visit a physician.

Dependent Variable: Physician Visit Indicator for 1990

As in the analysis of the 1985 data, we get the following initial and final models.

Initial model:
logit(μ) = -2.0361 + 0.6725 gen + 0.1198 chron + 0.1174 educ1 + 1.9507 educ2 + 1.5875 educ3 + 0.0240 age.

Final model: logit(μ) = -0.4878 + 0.5682 chron + 0.8052 educ1 + 2.0609 educ2 + 1.7389 educ3.

Therefore, in 1990, having more chronic conditions at the beginning of the study related to being more likely to visit a physician. Also, in 1990, having a lower education level is related to being more likely to visit a physician.

Overall

Overall, in most years, gender and/or the number of chronic conditions is related to the probability of seeing a physician.

4.3.2 Moment Method Analysis

After conducting the fixed-effects logistic regression analysis on all six years of data separately, we implemented the moment method of analysis for a generalized linear mixed model that was discussed in Chapter 3. This method takes into account the correlation that exists among family members. The results for all six regressions are included in Table 4.5.

Based on the β estimates, in 1985, females were more likely to visit a physician than males, having more chronic conditions is related to a higher likelihood of visiting a physician, having a higher education level is related to a higher likelihood of visiting a physician, and older people are more likely to visit a physician. Also, since the σ_γ value is large, the correlation among family members is important.

In 1986, females were more likely to visit a physician than males, having more chronic conditions is related to a higher likelihood of visiting a physician, having a higher education level is related to a higher likelihood of visiting a physician, and older people are less likely to visit a physician. Also, since the σ_γ value is large, the correlation among family members is important.

In 1987, females were more likely to visit a physician than males, having more chronic conditions is related to a higher likelihood of visiting a physician, having a higher education level is related to a higher likelihood of visiting a physician, and older people are more likely to visit a physician. Also, since the σ_γ value is large, the correlation among family members is important.

In 1988, females were more likely to visit a physician than males, having more chronic

Table 4.5: Moment Method Analysis: Binary dependent variable: Physician Visit Indicator for 1985 to 1990

Coefficients:								
year	xint	xgen	xchron	xeduc1	xeduc2	xeduc3	age	σ_γ
1985	-0.4130	0.8051	0.7313	-0.9542	0.0855	0.9970	0.0125	0.9055
1986	-1.0332	1.3154	0.4559	-0.0274	0.1990	0.9601	-0.0087	0.4556
1987	-0.7462	1.2006	0.1674	-0.9096	-0.3368	0.1205	0.0083	0.8440
1988	-1.3185	1.3618	0.3282	-0.6849	-0.0481	-0.0693	0.0154	0.0100
1989	-1.4418	0.8215	0.4124	0.0646	1.1152	0.2259	0.0039	0.0100
1990	-2.0699	0.6834	0.4264	0.4248	1.9845	1.6171	0.0244	0.3062

conditions is related to a higher likelihood of visiting a physician, having a higher education level is related to a higher likelihood of visiting a physician, and older people are more likely to visit a physician. Also, since the σ_γ value is small, the correlation among family members is not important.

In 1989, females were more likely to visit a physician than males, having more chronic conditions is related to a higher likelihood of visiting a physician, having a lower education level is related to a higher likelihood of visiting a physician, and older people are more likely to visit a physician. Also, since the σ_γ value is small, the correlation among family members is not important.

In 1990, females were more likely to visit a physician than males, having more chronic conditions is related to a higher likelihood of visiting a physician, having a lower education level is related to a higher likelihood of visiting a physician, and older people are more likely to visit a physician. Also, since the σ_γ value is larger, the correlation among family members may be important.

4.3.3 Quasi-Likelihood Method Analysis

As a comparison to the fixed-effects logistic regression and the moment method of analysis, we also used the quasi-likelihood method of analysis that was discussed in Chapter 3. This method also takes into account the correlation that exists among family members. The results for all six regressions are included in Table 4.6.

The data for 1985 did not produce any results since the estimates became too large in the iterations and thus, produced errors.

In 1986, females were more likely to visit a physician than males, having more chronic conditions is related to a higher likelihood of visiting a physician, having a higher education level is related to a higher likelihood of visiting a physician, and older people are less likely to visit a physician. Also, since the σ_γ value is smaller, the correlation among family members is not important.

In 1987, females were less likely to visit a physician than males, having more chronic conditions is related to a higher likelihood of visiting a physician, and older people are more likely to visit a physician. Also, since the σ_γ value is large, the correlation among family members is important.

In 1988, females were more likely to visit a physician than males, having more chronic conditions is related to a higher likelihood of visiting a physician, and older people are more likely to visit a physician. Also, since the σ_γ value is large, the correlation among family members is important.

In 1989, females were more likely to visit a physician than males, having more chronic conditions is related to a higher likelihood of visiting a physician, and older people

Table 4.6: Quasi Likelihood Method Analysis: Binary dependent variable: Physician Visit Indicator for 1985 to 1990

Coefficients:

year	xint	xgen	xchron	xeduc1	xeduc2	xeduc3	age	σ_s
1985	NA	NA	NA	NA	NA	NA	NA	NA
1986	1.9102	0.5120	1.3491	-0.5924	0.1699	0.7361	-0.1224	0.2760
1987	1.1670	-1.1905	0.1446	-0.8163	-0.7008	-0.9304	0.0516	0.1960
1988	-3.0326	3.0376	0.6784	-1.2921	0.0931	1.0987	0.0611	1.6031
1989	-0.3645	1.1582	0.7325	-0.2776	0.7832	-0.4953	-0.0172	1.0193
1990	-2.1138	0.8212	0.3383	0.0697	1.1168	1.7043	0.0294	0.1878

are less likely to visit a physician. Also, since the σ_s value is large, the correlation among family members is important.

In 1990, females were more likely to visit a physician than males, having more chronic conditions is related to a higher likelihood of visiting a physician, having a higher education level is related to a higher likelihood of visiting a physician, and older people are more likely to visit a physician. Also, since the σ_s value is small, the correlation among family members is not important.

Moment and Quasi-Likelihood Comparison

The moment method produced results for all years, whereas the quasi likelihood method did not produce results for 1985.

The moment method produced all positive results for gender, which is similar to the quasi-likelihood method that only produced one negative gender estimate for year 1987. For the number of chronic conditions and for σ_s, both methods produced positive results for all years. The results for education level and age are different among the two methods.

Bibliography

[1] N. E. Breslow & D. G. Clayton, Approximate Inference in Generalized Linear Mixed Models, *Journal of the American Statistical Association*, vol. 88, no. 421, pp 9–25, Mar 1993.

[2] Norman E. Breslow & Xihong Lin, Bias Correction in Generalised Linear Mixed Models with a Single Component of Dispersion, *Biometrika*, vol. 82, no. 1, pp 81–91, Mar 1995.

[3] R. Dennis Cook & Sanford Weisberg, *Applied Regression Including Computing and Graphics*, John Wiley & Sons, Inc., Toronto, 1999.

[4] Jay L. Devore, *Probability and Statistics for Engineering and the Sciences*, 5th ed. Duxbury Thomson Learning, Canada, 5th edition, 2000.

[5] John Fox, Linear Mixed Models: Appendix to An R and S-PLUS Companion to Applied Regression, May 2002. `http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-mixed-models.pdf`: accessed October 2, 2007.

[6] C. R. Henderson, Best Linear Unbiased Estimation and Prediction Under a Selection Model, *Biometrics*, vol. 31, no. 2, pp 423–448, June 1975.

[7] Jiming Jiang, Consistent Estimators in Generalized Linear Mixed Models, *Journal of the American Statistical Association*, vol. 93, no. 442, pp 720–729, Jun 1998.

[8] Jiming Jiang & Weihong Zhang, Robust Estimators in Generalised Linear Mixed Models, *Biometrika*, vol. 88, no. 3, pp 753–765, Sept 2001.

[9] Anthony Y. C. Kuk, Asymptotically Unbiased Estimation in Generalized Linear Models with Random Effects, *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 2, pp 395–407, 1995.

[10] P. McCullagh & J.A. Nelder, *Generalized Linear Models*, Chapman and Hall, New York, New York, USA, 2nd edition, 1989.

[11] C. A. McGilchrist. Estimation in Generalized Mixed Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 56, no. 1, pp 61–69, 1994.

[12] William Mendenhall, Lyman Ott & Richard L Scheaffer. *Elementary Survey Sampling*. Duxbury Press, Toronto, 5th edition, 1996.

[13] N. G. N. Prasad & J. N. K. Rao, The Estimation of Mean Squared Error of Small Area Estimators. *Journal of the American Statistician Association*, vol. 85, no. 409, pp 163–171, Mar 1990.

[14] Fred L. Ramsey & Daniel W. Schafer. *The Statistical Sleuth: A Course in Methods of Data Analysis*. Duxbury Press, Toronto, 1997.

[15] R. Schall, Estimation in Generalized Linear Models with Random Effects. *Biometrika*, vol. 78, no. 4, pp 719–727, Dec 1991.

[16] B. C. Sutradhar, On Exact Quasi-Likelihood Inference in Generalized Linear Mixed Models, *Sankhya: The Indian Journal of Statistics*, vol. 66, part 2, pp 261–289, 2004.

[17] B. C. Sutradhar & R. P. Rao, On Marginal Quasi-Likelihood Inference in Generalized Linear Mixed Models. *Journal of Multivariate Analysis*, vol. 76, no. 1, pp 1–34, Jan 2001.