A Review and Comparison of Models and Estimation Methods for Multivariate Longitudinal Data of Mixed Scale Type

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Casey L. Codd, M.A.

Graduate Program in Psychology

The Ohio State University

2014

Dissertation Committee:

Robert Cudeck, Advisor

Michael C. Edwards

Minjeong Jeon
Abstract

Models for the joint analysis of multiple outcome variables which are of possibly different scale types are useful because they allow researchers to answer questions related to the association between the trajectories of several variables and also provide a way to evaluate the change in the association between variables over time. There are several types of models that can handle multivariate longitudinal data, but one common approach involves using generalized linear mixed models with correlated random effects. This type of model is quickly growing in popularity in fields such as medicine and biostatistics, and the potential applications in the behavioral sciences are also quite broad. The limiting feature has been that estimation of the model parameters involves integration over the random effects in order to obtain the marginal distribution of the data.

In this dissertation, a review of several widely-available estimation methods and models for multivariate longitudinal data is provided. To evaluate the performance of the estimation methods within the multivariate generalized linear mixed model framework, a simulation study was conducted. The particular estimation methods of interest were adaptive Gaussian quadrature (AGQ), Laplace approximation (LA), penalized quasi-likelihood (PQL), and marginal quasi-likelihood (MQL). Results indicated that although AGQ and LA typically estimate the parameter estimates with less bias, PQL and MQL
tend to produce more stable estimates for the covariance matrix of random effects.
Furthermore, even though PQL performed quite poorly in many conditions, the bias of
parameter estimates tended to decrease as the correlation between the random effects
increased. Data from the National Longitudinal Study of Youth are used to illustrate the
applicability of the multivariate generalized linear mixed model to behavioral data. A
summary of the major findings in the literature review and also of numerical studies is
given.
Acknowledgments

First, I would like to thank my advisor, Bob Cudeck, who patiently offered encouragement and guidance throughout my graduate education. Thank you for sharing your knowledge of psychometric models, for offering many enjoyable conversations, and for believing in my potential to succeed. I would also like to thank Dr. Michael Edwards and Dr. Minjeong Jeon for their feedback on this dissertation and willingness to serve on my committee.

Lastly, I would like to express my gratitude to my family. Without the support and encouragement of my husband, Terrance, in particular, this process would have been much more difficult. I am glad that we got to go through it together. I also wish to acknowledge my son, Logan, who helps keep things in perspective. His laughs and hugs make each day better.
Vita

2007.................................................B.S. Psychology, Southern Nazarene University

2011..................................................M.A. Psychology, The Ohio State University

2008 to present.................................Graduate Assistant, Department of Psychology, The Ohio State University

Publications


Fields of Study

Major Field: Psychology
Table of Contents

Abstract ............................................................................................................................... ii
Acknowledgments.............................................................................................................. iv
Vita...................................................................................................................................... v
Table of Contents ............................................................................................................... vi
List of Tables ..................................................................................................................... ix
List of Figures .................................................................................................................... xi
Chapter 1: Introduction ...................................................................................................... 1
  1.1 Introduction to Dataset .............................................................................................. 5
Chapter 2: Generalized Linear Mixed Models for a Single Variable ......................... 11
  2.1 Linear Mixed-Effects Models for One Dependent Variable ................................... 11
  2.2 Generalized Linear Mixed Models .......................................................................... 14
    2.2.1 Continuous Variables ....................................................................................... 17
    2.2.2 Binary Variables ............................................................................................... 21
Chapter 3: Estimation Methods for Generalized Linear Mixed Models ..................... 25
  3.1 Methods Based on Linearization ............................................................................. 28
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.1 Penalized Quasi-Likelihood</td>
<td>28</td>
</tr>
<tr>
<td>3.1.2 Marginal Quasi-Likelihood</td>
<td>31</td>
</tr>
<tr>
<td>3.2 Methods Based on Integral Approximation</td>
<td>31</td>
</tr>
<tr>
<td>3.2.1 Adaptive Gaussian Quadrature</td>
<td>32</td>
</tr>
<tr>
<td>3.2.2 Laplace</td>
<td>34</td>
</tr>
<tr>
<td>3.3 Comparison of Estimation Methods</td>
<td>36</td>
</tr>
<tr>
<td>Chapter 4: Multivariate Longitudinal Data Analysis</td>
<td>40</td>
</tr>
<tr>
<td>4.1 Models for Multivariate Longitudinal Data</td>
<td>41</td>
</tr>
<tr>
<td>4.1.1 Conditional and Marginal Models</td>
<td>43</td>
</tr>
<tr>
<td>4.1.2 Multivariate Mixed Models</td>
<td>48</td>
</tr>
<tr>
<td>4.1.3 Dimension Reduction</td>
<td>55</td>
</tr>
<tr>
<td>4.1.4 Second Order Latent Growth Curves</td>
<td>56</td>
</tr>
<tr>
<td>4.2 Multivariate Generalized Linear Mixed Models</td>
<td>60</td>
</tr>
<tr>
<td>4.2.1 Example</td>
<td>62</td>
</tr>
<tr>
<td>Chapter 5: Simulation Study</td>
<td>67</td>
</tr>
<tr>
<td>5.1 Design of Simulation Study</td>
<td>69</td>
</tr>
<tr>
<td>5.2 Results</td>
<td>77</td>
</tr>
<tr>
<td>Chapter 6: Summary and Future Directions</td>
<td>100</td>
</tr>
<tr>
<td>References</td>
<td>107</td>
</tr>
</tbody>
</table>
Appendix A: NLSY Example Results ............................................................................ 115

Appendix B: Sample Syntax ........................................................................................... 116

Appendix C: Additional Tables ...................................................................................... 118
List of Tables

Table 1. Frequencies for the number of repeated observations for each outcome variable 8
Table 2. Frequencies for the number of observations at each age by outcome variable .... 8
Table 3. Data for the first 10 cases ..................................................................................... 9
Table 4. The distributions and link functions for common variable types ...................... 16
Table 5. Parameter estimates and standard errors for the GLMM for the BMI data .... 20
Table 6. Parameter estimates and standard error for GLMM for television rules data .... 24
Table 7. Parameter estimates and standard errors for multivariate GLMM for television
rules and BMI data............................................................................................................ 65
Table 8. Number of models that did not converge ............................................................ 78
Table 9. Number of models with non-positive definite covariance matrix for the random-effects ........................................................................................................ 79
Table 10. Percentage of confidence intervals (α=.05) for \( \phi_{12} \) that contain 0 .............. 99
Table 11. Parameter estimates and run times (min:sec) for multivariate GLMM for
television rules and BMI data .......................................................................................... 115
Table 12. Mean and standard deviation of parameter estimates for simulated models .. 119
Table 13. Average bias and mean square error (MSE) of parameter estimates for
simulated models ............................................................................................................. 123
Table 14. Percentage of confidence intervals (α=.05) that contain the true value of the parameter.

Table 15. Mean and average bias of parameter estimates for runs that converged successfully and where Φ was positive definite for all estimation methods.
List of Figures

Figure 1. Plot of median BMI-for-age for males (blue, solid line) and females (red, dashed line) ......................................................................................................................... 7

Figure 2. Classification of models for multivariate longitudinal data ......................... 43

Figure 3. Path diagram for a second order latent growth curve model ......................... 59

Figure 4. Boxplots for parameter estimates of the intercept of the binary variable ($\beta_0$) over the three variance and three correlation conditions for each estimation method ...... 84

Figure 5. Boxplots for parameter estimates of the slope of the binary variable ($\beta_1$) over the three variance and three correlation conditions for each estimation method .......... 86

Figure 6. Percentage of confidence intervals ($\alpha=.05$) that contain the true value of $\beta_0$ .. 88

Figure 7. Percentage of confidence intervals ($\alpha=.05$) that contain the true value of $\beta_1$ ... 89

Figure 8. Boxplots for parameter estimates of the variance of the random effect for the slope of the binary variable ($\phi_{11}$) over the three variance and three correlation conditions for each estimation method ................................................................. 92

Figure 9. Percentage of confidence intervals ($\alpha=.05$) that contain the true value of $\phi_{11}$ . 94

Figure 10. Boxplots for parameter estimates of the covariance of the random-effects ($\phi_{12}$ ) over the three variance and three correlation conditions for each estimation method ... 96
Figure 11. Percentage of confidence intervals ($\alpha=.05$) that contain the true value of $\phi_{11}^{12}$

Figure 12. Percentage of confidence intervals ($\alpha=.05$) for $\phi_{11}^{12}$ that contain 0 when $\rho^{12} = .30$
Chapter 1: Introduction

The study of change is a topic of interest in many disciplines. As a result, longitudinal designs are often used by applied researchers to study the trajectory of an outcome variable over time and to answer questions related to the impact of covariates on that trajectory. The challenge for methodological researchers has been to develop statistical models that are flexible enough to accommodate increasingly complex research questions and experimental designs.

One feature that makes longitudinal data unique is that repeated measures are taken on the same set of individuals over time. This allows for assessment of within-individual change, something that is not possible in cross-sectional designs. The repeated measures also create clustering within individuals. By clustering I mean that the repeated observations for each individual tend to be more similar than the observations between individuals. This complicates the analysis because the clustering must be accounted for in the model.

Mixed-effects models have become a popular approach for the analysis of clustered data. They are also often referred to as mixed models, hierarchical models, and random coefficient models. In the context of longitudinal data analysis, they provide a way to model the population-level trajectory through a set of fixed-effect regression
parameters. The defining feature of mixed models, however, is the ability to allow those regression parameters to vary randomly from individual-to-individual. These subject-specific regression coefficients are called random-effects. The variability of the random-effects provides a way to quantify the heterogeneity in the population, which allows researchers to answer questions related to differences between individuals.

By far, the most common mixed model uses a linear combination of regression parameters to model the change in a single continuous variable over time. This type of model has been widely reviewed by several authors (Laird & Ware, 1982; Pinheiro & Bates, 2000; Fitzmaurice, Laird & Ware, 2004; Weiss, 2005), and also applied in a diverse set of fields. Three limitations of this type of model, however, are that 1) not all change-processes are linear, 2) not all variables are continuous, and 3) most longitudinal designs involve more than a single outcome variable. Next I will discuss ways that the traditional linear mixed-effects model has been extended in order to compensate for each of these limitations.

First, several researchers have suggested ways to accommodate a nonlinear trajectory in mixed models (Lindstrom & Bates, 1990; Vonesh & Carter, 1992; Davidian & Giltinan, 1995). This flexibility is obviously appealing, but it comes at the price of increased complexity with regard to estimation. A wide variety of estimation methods have been proposed. These methods include individual-at-a-time approaches (Davidian & Giltinan, 1995), linearization methods (Lindstrom & Bates, 1990; Beal & Sheiner, 1982), adaptive quadrature approximations (Pinheiro & Bates, 1995; Davidian & Galant, 1993), and several other approaches. Of course, the estimation methods vary in terms of both
computational demand and accuracy. Many algorithms for estimating nonlinear mixed models have been implemented in mainstream software packages. For example SAS (SAS Institute, 2011a), R (R Core Team, 2013), Stata (StataCorp, 2013), and Mplus (Mplus, 2012) all have procedures for estimating at least some version of a nonlinear mixed model. As a result, it is becoming increasingly common to see applications of these models in the literature.

In recent years there has also been an increased interest in the analysis of variables that are not continuous. Generalized linear models (Nelder and Wedderburn, 1972; McCullagh & Nelder, 1989) were proposed as a way to unify regression models for different variable types, such as dichotomous, count variables, ordinal scales, and nominal measurements. These models were then merged with linear mixed-effects models in order to handle clustered variables. The resulting model was labeled as a generalized linear mixed model (GLMM). GLMMs have been discussed extensively by many authors including McCulloch and Searle (2001), Vonesh (2012) and Stroup (2013). For many of the same reasons as nonlinear mixed-effects models, GLMMs are also difficult to estimate. Similar estimation methods are used for both models, but the performance of the estimators is not always comparable between nonlinear mixed-effects models and GLMMs. This is because in addition to being nonlinear, GLMMs can also be used for data that is highly discrete. Some of the estimation methods that perform well with nonlinear mixed-effects models tend break down with certain types of data in the GLMM framework.
Finally, even though most models for longitudinal data are designed to handle only a single outcome variable, the majority of longitudinal studies measure several outcome variables over time. Researchers are often interested in answering questions that involve the relationships between these variables, so models must be able to handle several dependent variables simultaneously. Furthermore, because the outcome variables are typically not all measured on the same scale, the models must also be flexible enough to handle a combination of distributions. These types of models have been called multivariate generalized linear mixed models. Recent reviews on an entire class of models for multivariate longitudinal data have been given by Verbeke, Fieuws, Molenberghs and Davidian (2012), Bandyopadhyay, Ganguli and Chatterjee (2011), and Verbeke and Davidian (2009), illustrating the growing popularity of this type of approach.

Given that estimation of univariate GLMMs is difficult, it is not hard to imagine that estimation of multivariate GLMMs is even more complex. Several estimation methods have been compared in the univariate context, but little work has been done to compare their performance with multivariate models. The multivariate models have several interesting features that do not exist (or are at least not as prominent) in univariate models, so it is of interest to know how well they can be explored with each estimation method.

The overall goal of this thesis is to investigate how well longitudinal data with multiple dependent variables that are possibly of different scale types can be analyzed using estimation methods that are currently popular in the analysis of GLMMs for a
single dependent variable. In order to do this, it is first important to understand the estimation methods and how each performs under various conditions within the univariate GLMM framework. It is also important to understand the models that are used in the analysis of multivariate longitudinal data, know what features of the data can be investigated with each approach, and determine which framework seems to be the most useful in the context of multiple variables measured repeatedly over time on possibly different scales.

The remaining part of Chapter 1 will be spent introducing a dataset that will be referenced throughout this dissertation. In Chapter 2, I will define the GLMM for a single outcome variable. Particular attention will be given to handling continuous and binary variables within this framework, and I will illustrate how it can be applied in the analysis of behavioral data. In Chapter 3, several different estimation methods will be reviewed and their performance will be compared. Chapter 4 will begin with a general introduction to the analysis of multivariate longitudinal data. Then I will outline the multivariate GLMM framework and illustrate how it can be applied to behavioral data. In Chapter 5, I will present a simulation study which compares the performance of the various estimation methods on multivariate GLMMs. Finally, a discussion of conclusions and future research questions will be given in Chapter 6.

1.1 Introduction to Dataset

In order to solidify the theoretical ideas contained in this dissertation, a concrete example will be referenced throughout. The data for this example is taken from the
National Longitudinal Study of Youth 1979 (NLSY79) Child and Young Adult survey. This study is directed by the U.S. Bureau of Labor Statistics and the National Institute for Child Health and Human Development. Interviews are conducted every two years by the National Opinion Research Center at the University of Chicago and the study is managed by the Center for Human Resource Research at the Ohio State University. The interviews for the Child and Young Adult survey began in 1986. However, only data from the waves between 1988 and 2010 will be used here.

The first variable that will be analyzed is the child’s response to the question, “In your home, are there any rules about watching television?” Possible response options are “yes” or “no”, so the scale of this variable is dichotomous. This particular question was part of the Child Self-Administered Supplement which was only given to children between the ages of 10 and 14.

The second variable that is of interest is the child’s body mass index (BMI). This variable is calculated from the child’s weight (in pounds) and height (in inches) using the following formula:

\[ \text{BMI} = \frac{\text{weight}}{\text{height}^2} \times 703. \]

The Centers for Disease Control and Prevention provide national trends for BMI-for-age broken down by gender (Centers for Disease Control & Prevention, 2000), which are displayed in Figure 1. The national trends are developed using five cross-sectional health examination studies, with years ranging from 1963 to 1994. BMI is a continuous variable.

Since the question about television rules was only administered to children between the ages of 10 and 14, the data used here will also only include those age ranges.
A subset of 4,135 individuals with at least one measurement occasion on one of the two variables is included in the dataset. As with any longitudinal study, there is a large amount of missing data. A frequency table for the number of repeated observations for each variable is given in Table 1. Since the data is only collected every two years, the maximum number of repeated observations for any individual is two or three depending on his or her age at the first measurement occasion. Also, a frequency table for the number of observations at each age is given in Table 2. Again, no individual contributes data at all five ages.

Figure 1. Plot of median BMI-for-age for males (blue, solid line) and females (red, dashed line)
Table 1. Frequencies for the number of repeated observations for each outcome variable

<table>
<thead>
<tr>
<th>BMI</th>
<th>Television Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 26 7 0</td>
</tr>
<tr>
<td>1</td>
<td>209 602 111 2</td>
</tr>
<tr>
<td>2</td>
<td>62 379 1477 68</td>
</tr>
<tr>
<td>3</td>
<td>6 78 558 550</td>
</tr>
</tbody>
</table>

Table 2. Frequencies for the number of observations at each age by outcome variable

<table>
<thead>
<tr>
<th>Age</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Television Rules</td>
<td>923</td>
<td>1636</td>
<td>1595</td>
<td>1577</td>
<td>1520</td>
</tr>
<tr>
<td>BMI</td>
<td>1676</td>
<td>1779</td>
<td>1692</td>
<td>1710</td>
<td>1615</td>
</tr>
</tbody>
</table>
The data for the first 10 cases is shown in Table 3. They are illustrative of the number of variables and the different types that are the subject of this thesis.

<table>
<thead>
<tr>
<th>id</th>
<th>age</th>
<th>BMI</th>
<th>TV</th>
</tr>
</thead>
<tbody>
<tr>
<td>201</td>
<td>11</td>
<td>23.0223</td>
<td>1</td>
</tr>
<tr>
<td>202</td>
<td>10</td>
<td>28.689</td>
<td>.</td>
</tr>
<tr>
<td>302</td>
<td>11</td>
<td>17.1631</td>
<td>1</td>
</tr>
<tr>
<td>302</td>
<td>13</td>
<td>16.4196</td>
<td>1</td>
</tr>
<tr>
<td>303</td>
<td>10</td>
<td>17.358</td>
<td>.</td>
</tr>
<tr>
<td>303</td>
<td>12</td>
<td>17.7123</td>
<td>.</td>
</tr>
<tr>
<td>303</td>
<td>14</td>
<td>18.8521</td>
<td>0</td>
</tr>
<tr>
<td>401</td>
<td>10</td>
<td>18.5917</td>
<td>0</td>
</tr>
<tr>
<td>801</td>
<td>12</td>
<td>22.1299</td>
<td>0</td>
</tr>
<tr>
<td>801</td>
<td>14</td>
<td>28.1889</td>
<td>0</td>
</tr>
<tr>
<td>802</td>
<td>11</td>
<td>15.8175</td>
<td>1</td>
</tr>
<tr>
<td>802</td>
<td>13</td>
<td>17.678</td>
<td>0</td>
</tr>
<tr>
<td>803</td>
<td>10</td>
<td>18.0813</td>
<td>.</td>
</tr>
<tr>
<td>803</td>
<td>12</td>
<td>18.0212</td>
<td>1</td>
</tr>
<tr>
<td>803</td>
<td>14</td>
<td>23.0209</td>
<td>0</td>
</tr>
<tr>
<td>1603</td>
<td>10</td>
<td>23.902</td>
<td>.</td>
</tr>
<tr>
<td>1603</td>
<td>12</td>
<td>21.2051</td>
<td>0</td>
</tr>
<tr>
<td>1603</td>
<td>14</td>
<td>19.4835</td>
<td>0</td>
</tr>
<tr>
<td>2501</td>
<td>10</td>
<td>21.2727</td>
<td>.</td>
</tr>
<tr>
<td>2501</td>
<td>12</td>
<td>23.8012</td>
<td>0</td>
</tr>
<tr>
<td>2501</td>
<td>14</td>
<td>30.1756</td>
<td>1</td>
</tr>
<tr>
<td>3001</td>
<td>14</td>
<td>23.4131</td>
<td>.</td>
</tr>
</tbody>
</table>

Table 3. Data for the first 10 cases
Investigating change over time in two variables that may be selected from a much larger data archive that has potentially dozens of candidate measurements may not initially seem like such a major conceptual or statistical extension when compared to a more traditional single variable repeated measures analysis. However, the statistical complexity associated with a GLMM for even two variables is an order of magnitude more difficult than is fitting a model to a single variable. Among other things, this is due to the fact that maximum likelihood estimation of most models for multivariate GLMMs requires dealing with numerical integration to obtain the marginal distribution of the variables. Since the dimension of the integration is equal to the number of random-effects in the overall multivariate model, the difficulty increases with each additional random effect.

Models of this kind have only become feasible in the last decade, and their availability in widely distributed commercial software is even more recent. Although these developments make the use of these models possible, there has been very little systematic investigation of the performance of these methods. It is known that some estimators based on approximate methods are comparatively fast, and that other estimators based on direct and intensive calculations to obtain maximum likelihood estimates are theoretically of high quality. However, it is not known under what conditions the various estimators perform adequately in the multivariate GLMM context.
Chapter 2: Generalized Linear Mixed Models for a Single Variable

The primary focus of this chapter will be GLMMs. However, because GLMMs are an extension of linear mixed-effects models, it is useful to begin with a brief introduction to that framework. This will provide an opportunity to introduce notation and the foundational ideas of mixed models without the additional layers of complexity that are present in GLMMs.

2.1 Linear Mixed-Effects Models for One Dependent Variable

The linear mixed-effects model for a single variable measured repeatedly over time will be described in this section. For additional background on this type of model see Laird and Ware (1982). Let \( y_i \) be the \((n_i \times 1)\) vector of the observed outcome variable for individual \( i \), for all individuals \( i=1,\ldots,m \). The linear mixed model can be written as

\[
y_i = X_i \beta + Z_i u_i + e_i,
\]

where \( X_i \) is an \((n_i \times p)\) design matrix for the fixed-effects, \( \beta \) is a \((p \times 1)\) vector of fixed-effects, \( Z_i \) is an \((n_i \times k)\) design matrix for the random-effects, \( u_i \) is a \((k \times 1)\) vector of random-effects, and \( e_i \) is an \((n_i \times 1)\) vector of residuals. The fixed-effects are a set of
parameters which define the population-level trajectory. The random-effects are subject-specific terms which, along with the fixed-effects, define a trajectory for individual $i$.

In addition to showing that the model in (1) and the definitions that follow are subject-specific, the $i$ subscript also indicates that the linear mixed-effects model can accommodate different numbers of repeated measures and varying measurement occasions for each individual. This allows for considerable flexibility in the patterns of available data that can be considered in the use of this general model. As long as the missingness is ignorable (i.e., missing at random or missing completely at random), then valid inferences can be obtained with no additional accommodations for the missing data (Little & Rubin, 2002, Ch. 6).

Several assumptions are generally made about the random terms in the model. First, the random-effects are assumed to follow a multivariate normal distribution

$$u_i \sim N(0, \Phi),$$

(2)

where $\Phi$ is a $(k \times k)$ covariance matrix. Some work has been done with models that relax this assumption (Codd & Cudeck, 2014), but most of the time normality is both a convenient and reasonable assumption. The residuals are also assumed to follow a multivariate normal distribution with

$$e_i \sim N(0, \Lambda_i).$$

(3)
Here, $\mathbf{\Lambda}_i$ is an $(n_i \times n_i)$ covariance matrix. The random-effects and residuals are also assumed to be uncorrelated. In other words, $\text{cov}(\mathbf{e}_i, \mathbf{u}_i) = \mathbf{0}$.

Conditional on the random-effects, the distribution of $y_i$ is

$$y_i | \mathbf{u}_i \sim \mathcal{N}(\mathbf{X}_i \beta + \mathbf{Z}_i \mathbf{u}_i, \mathbf{\Lambda}_i).$$

Finally, the marginal distribution of $y_i$ is

$$f(y_i) = \int f(y_i | \mathbf{u}_i) f(\mathbf{u}_i) d\mathbf{u}_i.$$  (5)

In the context of linear mixed-effects models, $f(y_i)$ can actually be simplified even further (Davidian & Giltinan, 1995, Ch. 3). This is due to the assumptions in (2) and (3), and the fact that the model in (1) is linear in the random-effects. The closed-form expression for the marginal distribution of $y_i$ is

$$y_i \sim \mathcal{N}(\mathbf{X}_i \beta, \mathbf{Z}_i \mathbf{\Phi} \mathbf{Z}_i' + \mathbf{\Lambda}_i).$$  (6)

Maximum likelihood estimation of $\beta$, $\Phi$, and $\Lambda_i$ when $y_i$ is normally distributed has been described by many writers (Demidenko, 2004, Ch. 2; Vonesh & Chinchilli, 1997, Ch. 6; Davidian & Giltinan, 1995, Ch. 3) and is implemented in many computer programs including SAS (SAS Institute, 2011a), SPSS (IBM Corp., 2012), Stata (StataCorp, 2013), Mplus (Mplus, 2012), HLM (Raudenbush, Bryk, & Congdon, 2004), SuperMix (Hedeker & Gibbons, 2008), and several packages within R (R Core Team, 2013) including lme4 (Bates, Maechler, Bolker, & Walker, 2013).
2.2 Generalized Linear Mixed Models

The linear mixed model defined in the previous section is designed to handle only continuous, normally distributed response variables. In reality, researchers are often confronted with many other types of variables. The GLMM framework is a way to unify mixed models for variables from the exponential family of distributions, which includes distributions that are suitable for handling continuous, dichotomous, ordinal, nominal, and count variables, among others. More details about the exponential family of distributions will be given below. The GLMM will be defined in this section. Additional background on this framework can be found in McCulloch and Searle (2001), Vonesh (2012), and Stroup (2013).

GLMMs can be described as a kind of latent variable model. Let $y_i$ be an observed or manifest variable on which participant $i$ is actually measured. Repeated measures produce the multivariate structure, $y_i = (y_{i1}, \ldots, y_{in_i})'$. As in the previous section, consider a linear mixed model analogous to (1) that is the sum of independent variables including design variables or subject level measurements weighted by fixed-effects plus possibly different design variables or subject level measurements weighted by random-effects

$$\eta_i = X_i \beta + Z_i u_i.$$  \hfill (7)
Here, $\eta$ is an ($n_i \times 1$) vector which defines the systematic component or linear predictor of the model. There are four main components in the GLMM framework, and this is the first.

The second component is the distribution of the random-effects. This distribution is assumed to be normal, as was the case in (2) for the linear mixed model.

Because many variables are not normally distributed as in (4), a linear model is not always appropriate. Therefore, the third component of a GLMM is a link function which converts the expected value of the conditional distribution of the data given the random effects to the linear predictor, $\eta$. Define the expected value of the conditional distribution of the data given the random effects as $\mu_i = E(y_i | u_i)$. The link function is

$$g(\mu_i) = \eta_i = X_i\beta + Z_iu_i.$$  \hspace{1cm} (8)

Similarly, (8) can be written as

$$\mu_i = g^{-1}(X_i\beta + Z_iu_i).$$  \hspace{1cm} (9)

The particular form of $g(\cdot)$ or $g^{-1}(\cdot)$ is determined by the type of variable contained in $y_i$. For example, the major variable types and their link functions are described in Table 4. One can see that a linear mixed model can be expressed as a GLMM using an identity link function, something that will be illustrated in further detail in the following section.
<table>
<thead>
<tr>
<th>Variable Type</th>
<th>Distribution</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary</td>
<td>Bernoulli</td>
<td>logit or probit</td>
</tr>
<tr>
<td>Continuous</td>
<td>Normal</td>
<td>identity</td>
</tr>
<tr>
<td>Count</td>
<td>Poisson</td>
<td>logarithm</td>
</tr>
<tr>
<td>Nominal</td>
<td>Multinomial</td>
<td>generalized logit</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Multinomial</td>
<td>cumulative logit</td>
</tr>
</tbody>
</table>

Table 4. The distributions and link functions for common variable types

The final part of the model is the random component which specifies the distribution of $y_{ij}$ conditional on the random-effects. Although the random-effects are assumed to be normally distributed, the conditional distribution $f(y_{ij} \mid u_i)$ is not necessarily normal. The only requirements for $f(y_{ij} \mid u_i)$ are that it be a member of the exponential family. The exponential family of distributions has a probability density function with the following general form

$$f(y_{ij} \mid u_i) = \exp \left\{ \frac{y_{ij} \mu_{ij} - b(\mu_{ij})}{a_i(\phi)} + c(y_{ij}, \phi) \right\}$$ (10)

where $b(\mu_{ij})$, $a_i(\phi)$, and $c(y_{ij}, \phi)$ are specific functions whose values depend on the particular distribution that is used, and $\phi$ is a dispersion parameter. For additional background on the exponential family of distributions and how they fit within the generalized linear model framework, see McCullagh and Nelder (1989, Ch. 2).
Combining each of the four components of the GLMM, the marginal distribution of $y_{ij}$ has the same form as (5). The difference is that for GLMMs $f\left(y_i\right)$ does not generally have a closed-form expression as it did for the linear mixed model in (6).

To summarize, the four components that are required in the specification of a GLMM for variable $y_i$ are:

1. A function for $\eta_i$ which is a linear combination of covariates and fixed and random-effects, as in (7)
2. A distribution for the random-effects, as in (2)
3. A link function relating the conditional mean of the data to the linear predictor, as in (9)
4. A distribution for $y_i | u_i$ which is from the exponential family, as in (10)

Several common distributions are members of the exponential family and are therefore encompassed within the GLMM framework. Some of the most common variable types include continuous, dichotomous, ordinal and count variables. The following sections will illustrate how continuous and binary data can be modeled using GLMMs with normal and Bernoulli distributions, respectively.

2.2.1 Continuous Variables

When a response variable is continuous, it is generally modeled using a normal distribution with an unknown mean vector and covariance matrix that are specified as
parts of a statistical model to be estimated. The probability density function for Gaussian data is written as

\[ f(y_i | u_i) = (2\pi)^{-m/2} |\Lambda_i|^{-1/2} \exp\left( -\frac{1}{2} (y_i - \mu_i)' \Lambda_i^{-1} (y_i - \mu_i) \right), \]

where \(\Lambda_i\) is the residual variance matrix, as was defined in (3). In other words,

\[ \text{var}(y_i | u_i) = \Lambda_i. \]

In order to be as general as possible, assume that the linear predictor has the form of (7), and that the distribution of the random-effects is as in (2). Because the range of \(y_i\) is boundless and does not need to be restricted, an identity link function can be used. This means that

\[ \mu_i = g^{-1}(X_i\beta + Z_iu_i) = \eta_i = X_i\beta + Z_iu_i. \]

To summarize, the components of the GLMM for continuous, normally distributed data are:

1. Linear predictor: \(\eta_i = X_i\beta + Z_iu_i\)
2. Random-effects distribution: \(u_i \sim N(0, \Phi)\)
3. Link function: \(\mu_i = \eta_i = X_i\beta + Z_iu_i\)
4. Distribution for \(y_i | u_i\): \(y_i | u_i \sim N(\mu_i, \Lambda_i)\)

Based on this specification, one can see that a GLMM for a normally distributed continuous variable is equivalent to the linear mixed model described in Chapter 2.1.
Example

The BMI data that was introduced in Chapter 1 will be used here to illustrate the GLMM for a continuous variable. Although BMI actually changes nonlinearly as one ages (as shown in Figure 1), over the range of ages included in this sample the trajectory is fairly linear. It also appears that there might be slight differences in BMI for males and females, however, models including sex as a covariate indicated that this was not actually the case. Therefore, a linear model with only fixed and random-effects for the intercept and slope will be used here.

The linear predictor for the BMI data is

\[
\eta_i = \begin{pmatrix} \eta_{i1} \\ \vdots \\ \eta_{in} \end{pmatrix} = \begin{pmatrix} 1 & \text{age}_{i1} - 10 \\ \vdots & \vdots \\ 1 & \text{age}_{in} - 10 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 1 & \text{age}_{i1} - 10 \\ \vdots & \vdots \\ 1 & \text{age}_{in} - 10 \end{pmatrix} \begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix}.
\]

(11)

\[
\begin{align*}
\eta_i &= X_i \beta + Z_i u_i.
\end{align*}
\]

Age is shifted here so that the parameter estimate for the intercept, \( \beta_0 \), reflects the predicted BMI at age 10. The fixed-effect parameter for the slope, \( \beta_1 \), represents the average change in BMI for each one-year increase in age. The random-effects for the intercept, \( u_{0i} \), and slope, \( u_{1i} \), are assumed to be uncorrelated, a decision which is based on parameter estimates from a series of models that were previously fit to the data. The distribution of the random-effects is, therefore,
The identity link function is used, meaning that $\mu_i = \eta_i$. Finally, the conditional distribution of $y_i$ given the random-effects is

$$y_i | u_i \sim \mathcal{N}(\eta_i, \Lambda_i),$$

where $\Lambda_i = \sigma_e^2 I$. This is a relatively common structure to use for the residual variance. It indicates that the variance of the residuals is constant over time and that the residuals are uncorrelated between time points. There are, however, also many other options for residual structures. See Davis (2002, Ch. 6) or Verbeke and Molenberghs (2000, Ch. 8) for additional examples. This model was estimated using adaptive Gaussian quadrature in SAS PROC GLIMMIX. The results are presented in Table 5.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>18.86</td>
<td>.08</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>.88</td>
<td>.02</td>
</tr>
<tr>
<td>$\phi_{00}$</td>
<td>17.04</td>
<td>.72</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>.33</td>
<td>.05</td>
</tr>
<tr>
<td>$\sigma_e^2$</td>
<td>4.34</td>
<td>.25</td>
</tr>
</tbody>
</table>

Table 5. Parameter estimates and standard errors for the GLMM for the BMI data.
The results indicate that the average BMI at age 10 is estimated to be 18.86, and that BMI increases at an average rate of 0.88 per year. The variances for the random effect intercept ($\hat{\phi}_{00} = 17.04$) and slope ($\hat{\phi}_{11} = .33$) are large compared to their standard errors. This suggests that there are individual differences in both initial BMI and change in BMI over time.

2.2.2 Binary Variables

When a response variable is binary, it is generally modeled using a Bernoulli distribution. In keeping with the GLMM notation, the probability density function for Bernoulli data at time point $y_{ij}$ is written as

$$f(y_{ij}|u_i) = \mu_{ij}^{y_{ij}}(1-\mu_{ij})^{1-y_{ij}}.$$  

Conditional on the random-effects, the data are assumed to be independent between time points. This means that

$$f(y_i|u_i) = f(y_{i1}|u_i) \cdots f(y_{in_i}|u_i).$$

Once again, assume that the linear predictor has the form of (7), and that the distribution of the random-effects is as in (2). There are two link functions that are commonly used with binary data. These are the probit and logit links. Here, the logit will be used meaning that
\[
\text{logit}(\mu_i) = \log \left( \frac{\mu_i}{1 - \mu_i} \right) = \eta_i = X_i \beta + Z_i u_i.
\]

Solving for \(\mu_i\) term by term,

\[
\mu_i = \text{logit}^{-1}(\eta_i) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} = \frac{\exp(X_i \beta + Z_i u_i)}{1 + \exp(X_i \beta + Z_i u_i)}. 
\]

To summarize, the components of the GLMM for binary data with a Bernoulli distribution are:

1. Linear predictor: \(\eta_i = X_i \beta + Z_i u_i\)
2. Random-effects distribution: \(u_i \sim N(0, \Phi)\)
3. Link function: \(\mu_i = \text{logit}^{-1}(X_i \beta + Z_i u_i) = \frac{\exp(X_i \beta + Z_i u_i)}{1 + \exp(X_i \beta + Z_i u_i)}\)
4. Distribution for \(y_{ij} | u_i: y_{ij} | u_i \sim \prod_{j=1}^{n_i} \text{Bernoulli}(\mu_{ij})\)

Example

To further illustrate the analysis of binary data in the GLMM framework, consider modeling the change in the television rules variable over age. The model used here will include fixed-effects for the intercept and slope, and a random effect for the intercept. Once again, age will be used as the indicator of time. A random effect for the slope is not
included because previous models indicated that there was not substantial variability between individuals in terms of change over time.

The linear predictor for this model is

$$\begin{pmatrix} \eta_i \\ \vdots \\ \eta_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & \text{age}_{i, 10} \\ \vdots & \vdots \\ 1 & \text{age}_{i, 10} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} u_{0i}$$

and the distribution of the random effect is

$$u_{0i} \sim N(0, \psi_{00}).$$

The logit link function is used, meaning

$$\mu_i = \frac{\exp(X_i \beta + Z_i u_i)}{1 + \exp(X_i \beta + Z_i u_i)}. \quad (13)$$

The distribution of $y_{ij}$ conditional on the random-effects is Bernoulli with the probability parameter equal to the $j^{th}$ element of (13). This model was estimated using adaptive Gaussian quadrature in SAS PROC GLIMMIX. The results are presented in Table 6.

The parameter estimates indicate that the log-odds of having rules about television at age 10 are approximately $\hat{\beta}_0 = .53$. For this parameter, the log-odds are defined as $\hat{\beta}_0 = \text{logit}(p)$, where $p$ is the probability of having rules about television at age.
Conversely, the probability can be determined directly as $p = \logit^{-1}(\hat{\beta}_0)$. In this example, the probability of having rules about television at age 10 is 0.63. Again, in terms of log-odds this rate decreases as one ages ($\hat{\beta}_1 = -.26$), reaching a final predicted probability of having rules about television of approximately 0.37 at the age of 14. The variance of the random effect for the intercept is also large ($\hat{\varphi}_{00} = 2.19$), indicating that there are individual differences in the initial likelihood of having rules about television at age 10.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>.53</td>
<td>.06</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-.26</td>
<td>.02</td>
</tr>
<tr>
<td>$\varphi_{00}$</td>
<td>2.19</td>
<td>.22</td>
</tr>
</tbody>
</table>

Table 6. Parameter estimates and standard error for GLMM for television rules data
Chapter 3: Estimation Methods for Generalized Linear Mixed Models

The generalized linear model is a significant development in regression because it extends the class of variables for which regression models are appropriate. Similarly, GLMMs are an important enhancement to the class of mixed-effects models for repeated measures data. In spite of the progress represented by this class of methods, their use in practical research was limited by the complexity of the estimation problems until these methods were incorporated into popular commercial software. Even with the availability of commercial software, however, estimation of GLMMs can still be difficult. Depending on the complexity of the model, successful estimation may involve considerable run times and require fairly accurate starting values. A consequence of the obstacle that the computational problems pose is the fact that several estimation methods have been presented that attempt to estimate GLMMs more quickly. In this section I review four of the most widely-used methods that have been suggested to this point.

The marginal distribution of \( y_i \) is defined as

\[
f(y_i) = \int f(y_i | u_i) f(u_i) du_i ,
\]

(14)
where \( f(y_i | u_i) \) is the distribution of \( y_i \) conditional on the random-effects as defined in (10), and \( f(u_i) \) is the distribution of the random-effects as defined in (2).

The parameters of GLMMs are usually estimated with maximum likelihood of the marginal distribution of \( y_i \). Define \( \Sigma = vech(\Phi)' \) and \( \Gamma = vech(\Lambda_i)' \), where \( \Sigma (q \times 1) \) and \( \Gamma (s \times 1) \) are column vectors of order \( q = k(k+1)/2 \) and \( s = n_i(n_i+1)/2 \) with \( k \) and \( n_i \) the order of \( \Phi \) and \( \Lambda_i \), respectively. The \( vech \) operator stacks the lower diagonal elements of a symmetric matrix into a column vector. For example,

\[
vech\left(\begin{array}{ccc}
\varphi_{11} \\
\varphi_{21} & \varphi_{22}
\end{array}\right) = \begin{pmatrix}
\varphi_{11} \\
\varphi_{21} \\
\varphi_{22}
\end{pmatrix}.
\]

Then, define \( \theta \) to be a vector of all model parameters. In other words, \( \theta = (\beta', \Sigma, \Gamma)' \). The likelihood function is

\[
L(\theta) = \prod_{i=1}^{N} \int f(y_i | u_i) f(u_i) du_i.
\]

(15)

Even though the form of \( f(y_i | u_i) \) and \( f(u_i) \) are known, in general there is not a closed-form expression for (15). This is because the link function typically causes the random-effects to enter the model nonlinearly. Consequently, approximate methods must be used to obtain estimates for \( \theta \).
In some special cases (15) can simplify to a function which does not involve integration. For example, in a GLMM for a continuous variable where both $f(y_i | u_i)$ and $f(u_i)$ are normal and an identity link function is used, the random-effects enter the model linearly. Therefore, $f(y_i)$ will also follow a normal distribution, and the integral in (14) can be avoided (Davidian & Giltinan, 1995). This is generally not the case, however.

Several methods for approximating (15) have been suggested. The most common approaches can be divided into two classes. The first class of methods linearizes the link function so that estimation methods for linear mixed models can be used (Zeger, Liang, & Albert, 1988; Breslow & Clayton, 1993; Wolfinger & O’Connell, 1993). The second class of methods attempts to approximate the integral directly using numerical integration. In the following sections, several approaches will be described in detail. The performance of each of the estimation methods will be compared in the final section of this chapter.

It is worth pointing out that the estimation methods covered in this chapter are not the only methods that have been proposed for GLMMs. Other methods include several variations of the EM algorithm (McCulloch, 1994; Steele, 1996; Booth & Hobert, 1999), an Iterative Bias Correction method (Kuk, 1995), Gibbs sampling procedures (Karim & Zeger, 1992), and nonparametric approaches (Atkin, 1996), among others. The approaches that are covered in this chapter were selected because of their accessibility for
applied researchers. In the major software packages for GLMMs, the linearization and numerical integration approaches are by far the most widely-implemented.

3.1 Methods Based on Linearization

The penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL) methods discussed in this section can be motivated from several different perspectives. Essentially equivalent procedures have been suggested by Zeger, Liang, and Albert (1988) using generalized estimating equations, Breslow and Clayton (1993) using an elaboration of the Laplace approximation, and Wolfinger and O’Connell (1993) using Taylor series expansion. In the context of nonlinear mixed models, Lindstrom and Bates (1990), Beal and Sheiner (1982), and Vonesh and Carter (1992) also suggested similar procedures. The development in this section will follow that of Wolfinger and O’Connell (1993).

3.1.1 Penalized Quasi-Likelihood

The PQL method has been outlined by several writers (Vonesh, 2012; Wolfinger & O’Connell, 1993; Demidenko, 2004, Ch. 7), and will be reviewed in this section. Recall from Chapter 2.2 that

$$E(y_i | u_i) = \mu_i = g^{-1}(X_i \beta + Z_i \xi_i) = g^{-1}(\eta_i).$$

(16)
A first-order Taylor series expansion of (16) about $\hat{\beta}$ and $\bar{u}_i$, the current, fixed values of $\beta$ and $u_i$, has the following form

$$
g^{-1}(\eta_i) \approx g^{-1}(X_i\hat{\beta} + Z_i\bar{u}_i) + \tilde{\Delta}_i \left[ X_i(\beta - \hat{\beta}) + Z_i(u_i - \bar{u}_i) \right].$$

Here, $\tilde{\Delta}_i$ is a diagonal matrix of derivatives of $E(y_i | u_i)$ evaluated at the expansion points, $\hat{\beta}$ and $\bar{u}_i$. In other words,

$$\tilde{\Delta}_i = \frac{\partial (g^{-1}(\eta_i))}{\partial \eta_i} \bigg|_{\hat{\beta}, \bar{u}_i}.$$

The model for the data, $y_i$, can then be defined as

$$y_i = g^{-1}(X_i\hat{\beta} + Z_i\bar{u}_i) + \tilde{\Delta}_i \left[ X_i(\beta - \hat{\beta}) + Z_i(u_i - \bar{u}_i) \right] + \varepsilon_i. \quad (17)$$

From (17) a pseudo response vector, $\bar{y}_i$, can be formed as

$$\bar{y}_i = \tilde{\Delta}_i^{-1} (y_i - g^{-1}(X_i\hat{\beta} + Z_i\bar{u}_i)) + X_i\hat{\beta} + Z_i\bar{u}_i. \quad (18)$$

Next, define a weight matrix, $\tilde{W}_i$, as

$$\tilde{W}_i = \tilde{V}_i^{-1}\tilde{\Delta}_i^{-2},$$

where $\tilde{V}_i$ is a matrix of the diagonal elements of $var(\varepsilon_i | \hat{\beta}_i, \bar{u}_i)$. 

29
A linear mixed-effects model of the form

$$\tilde{y}_i = X_i \beta + Z_i u_i + \tilde{W}_i^\dagger \epsilon_i^*.$$  

(19)

can be fit to the pseudo data, assuming that $\epsilon_i^* \sim N(0, \Lambda^*)$. Because (19) is linear in the random-effects, the likelihood function has a closed-form solution. The model can be fit using either pseudo likelihood (i.e., full maximum likelihood based on the pseudo data) or restricted pseudo-likelihood (i.e., restricted maximum likelihood based on the pseudo data). Because the model fit to the pseudo data is technically a linear mixed model, the distinction between full maximum likelihood and restricted maximum likelihood is the same as it is within that framework. Specifically, restricted maximum likelihood adjusts the variance and covariance parameters for uncertainty that arises as a result of estimating the fixed-effects. For additional information on the distinction between restricted maximum likelihood and full maximum likelihood see Raudenbush and Bryk (2002, Ch. 3). Once the parameter estimates from the model in (19) are obtained, $\tilde{\beta}$ and $\tilde{u}_i$ are updated and the next iteration is initiated.

This method is a doubly iterative procedure, where the Taylor series expansion about the current estimate of $\tilde{\beta}$ and $\tilde{u}_i$ is the first step, and the fitting of a linear mixed model to the pseudo data is the second step. Iteration between these steps continues until the difference between the parameter estimates in successive iterations is sufficiently small.
3.1.2 Marginal Quasi-Likelihood

The penalized quasi-likelihood (PQL) routine outlined in the previous section is often described as a subject-specific method since the Taylor series expansion is taken around the fixed and random-effects. Marginal quasi-likelihood (MQL), in contrast, is a population-averaged approach. In the mixed model literature this method was developed by Beal and Sheiner (1982). An equivalent procedure was proposed by Breslow and Clayton (1993) for the GLMM framework. The basic algorithm is equivalent to that of PQL, except for instead of expanding around the fixed and random-effects, $\beta^*$ and $\bar{u}_i$, the Taylor series expansion is taken about the fixed-effect and the average random effect. Since the random-effects are normally distributed with $E(u_i)=0$, this is equivalent to expanding about the fixed-effects only. In other words, estimation using MQL is closely related to the routine described in Section 3.1.1, except for the Taylor series expansion is about $\tilde{\beta}$ only.

3.2 Methods Based on Integral Approximation

Adaptive Gaussian quadrature (AGQ) and the Laplace approximation have been applied to many different types of models. In the context of GLMMs, they have been reviewed by Pinheiro and Chao (2006), Demidenko (2004), and Vonesh (2012). In contrast to the PQL and MQL approaches described in the previous section, these methods attempt to approximate the integral in (15) numerically. Additionally, they are
singly-iterative procedures whereas both MQL and PQL were doubly-iterative. Both AGQ and Laplace’s method will be described in the following sections.

3.2.1 Adaptive Gaussian Quadrature

Quadrature methods approximate an integral through a weighted average of the integrand evaluated at a set of nodes. For GLMMs, adaptive Gauss-Hermite quadrature is a popular approach. This method has the following general form

$$\int_{-\infty}^{\infty} k(x)\exp\left(-x^2\right)dx \approx \sum_{q=1}^{Q} w_q k(n_q),$$

where $w_q$ are the weights and $n_q$ are the nodes for a Q-point method (Krommer & Ueberhuber, 1994, Section 4.2).

Writing the likelihood function from (15) in the form of (20) requires a change of variables. However, the fact that $f(u_i)$ is assumed to be normal ensures that obtaining the appropriate form will always be possible. Maximum likelihood estimation can then be used to minimize the function and obtain parameter estimates.

For example, in a GLMM with a single random effect, the distribution of the random effect is

$$f(u_i) = \left(2\pi\phi\right)^{-1/2} \exp\left(-\frac{u_i^2}{2\phi}\right).$$
Let $v_i = u_i / \sqrt{2\phi}$. Conversely, $u_i = v_i \sqrt{2\phi}$ with first derivative $\frac{\partial u_i}{\partial v_i} = \sqrt{2\phi}$. Using this transformation, the density of $v_i$ can be calculated based on the density of $u_i$ evaluated at $u_i(v_i) = v_i \sqrt{2\phi}$ as follows

$$h(v_i) = f(u_i(v_i)) \frac{\partial u_i}{\partial v_i}$$

$$= (2\pi\phi)^{1/2} \exp \left( \frac{-(v_i \sqrt{2\phi})^2}{2\phi} \right) \sqrt{2\phi}.$$  

$$= \exp(-v_i^2) \frac{\sqrt{\pi}}{\sqrt{\pi}}$$

The marginal density of $y_i$ in (14) can then be calculated as

$$f(y_i) = \int f(y_i | u_i) f(u_i) \partial u_i$$

$$= \int h(y_i | v_i) f(v_i) \partial v_i$$

$$= \frac{1}{\sqrt{\pi}} \int h(y_i | v_i) \exp(-v_i^2) \partial v_i$$

(22)

where $f(y_i | u_i)$ is the distribution of the outcome variable conditional on the original random effect, $u_i$, which comes from the exponential family, and $h(y_i | v_i)$ is found by substituting $u_i = v_i \sqrt{2\phi}$ into $f(y_i | u_i)$. One can see that (22) has the form of (20), so the integral can be written as

$$f(y_i) \approx \sum_{q=1}^{Q} w_q^* h(y_i | v_{iq})$$

where $w_q^* = w_q / \sqrt{\pi}$, $w_q$ are the quadrature weights, and $v_{iq}$ are the quadrature nodes.
The approximation in (20) becomes more accurate as the number of quadrature points increases. At the same time, the computational demand also increases at the rate of $Q_v$ for $v$-dimensional integration. Therefore, selecting an appropriate number of quadrature points is a balancing act between accuracy and feasibility. In adaptive quadrature, the number of quadrature points is determined by fitting a series of models with an increasing number of quadrature points until the increase in quadrature points no longer provides substantial improvement in performance (SAS Institute, 2011a). Additionally, adaptive routines place the quadrature points in areas with the highest density.

### 3.2.2 Laplace

Laplace’s method is closely related to adaptive Gaussian quadrature (Stroup, 2012; Demidenko, 2004). Rather than evaluate the integrand at several quadrature points, however, the Laplace approximation evaluates the integrand at the single point where the function takes its maximum value. The general form of the approximation is as follows

$$\int_{-\infty}^{\infty} \exp(h(x)) dx \approx (2\pi)^{1/2} \exp(h(x_{\text{max}})) \left| \frac{-\partial^2 h(x)}{\partial x^2} \right|_{x=x_{\text{max}}}^{1/2},$$

(23)

where $x_{\text{max}}$ is the value of $x$ that maximizes $h(x)$. Since the first derivative of $h(x)$ will be zero at its maximum value, the second derivative is used to find $x_{\text{max}}$. The second derivative should be negative, ensuring that a maximum value has, in fact, been found.
In a GLMM with a single random effect, the integral that needs to be approximated has the following general form

\[ f(y_i) = \int_{u_i} f(y_i | u_i) f(u_i) du_i , \]

where \( f(y_i | u_i) \) is the distribution of the outcome variable conditional on the random effect which is a member of the exponential family of distributions, and \( f(u_i) \) is the distribution of the random effect. In order to format this appropriately for the Laplace approximation in (23), one can simply set \( h(x) = \ln[f(y_i | u_i) f(u_i)] \).

It has been demonstrated that the Laplace approximation is equivalent to AGQ with a single quadrature point (e.g., SAS Institute, 2011a). However, there are computational differences between the two approaches that make them better suited for different types of problems. In particular, the Laplace approximation is simpler computationally and tends to be more efficient than AGQ with one quadrature point (Stroup, 2013).

In summary, it seems that currently the four most widely-applied methods for estimating GLMMs are the following: 1) PQL, 2) MQL, 3) AGQ, and 4) Laplace. Although these are all intended to be accurate and broadly applicable methods for estimating this class of models, the methods have different features when examined algebraically. Based on characteristics of their derivation, these differ in computational requirements and sophistication of the approximation to the model and the likelihood function.
3.3 Comparison of Estimation Methods

Given that each of these methods is just a way to approximate the likelihood function, it is not surprising that they are all subject to some amount of inaccuracy. The real topic of interest is in determining how well they perform in general, and in knowing under what conditions they begin to break down. There have been a few studies comparing estimation methods for GLMMs for univariate dependent variables, and the results from those studies will be outlined in this section. Since there is also some overlap with estimation for nonlinear mixed models, some results from that domain will also be included.

In general, each of the four estimation approaches outlined in the previous section tends to perform well when there are a large number of subjects (Breslow & Lin, 1995), a large number of observations per subject (Breslow & Clayton, 1993), the range of observed values is large (Vonesh, 2012), and the variances of the random-effects are small (Rodriguez & Goldman, 2001). Obviously, that set of restrictions is fairly limiting. More importantly, real data problems rarely meet all of those requirements.

Overall, the variance parameters for the random-effects are more difficult to estimate than the fixed effect parameters. In fact, estimation of the random effect variances is often what differentiates the methods (Pinheiro & Chao, 2006; Dey & Lim, 2013). Nevertheless, under certain conditions even the fixed-effects can be difficult to estimate and can lead to biased results. Also, the methods tend to differ only in terms of bias, not in terms of the variability of the parameter estimates (Pinheiro & Chao, 2006).
Based on the theoretical derivation of MQL, one might suspect that the accuracy of this approach would decrease as the variance of the random-effects increases. This is due to the fact that the Taylor series expansion is taken about the fixed-effects and the mean of the random-effects, which is assumed to be 0. Specifically, as the variances of the random-effects increase (i.e. as the subject-specific terms move further from 0), the MQL parameter estimates tend to be biased toward zero (Rodriguez & Goldman, 1995).

When the number of subjects or the number of observations per subject decreases, all of the methods tend to be less accurate. The bias tends to be largest for MQL, but it is also substantial for PQL (Breslow & Lin, 1995; Lin & Breslow, 1996). In both cases, the parameter estimates are biased toward zero when the number of repeated measures is small (Breslow & Clayton, 1993). Laplace tends to be less biased than PQL, and AGQ less biased than Laplace (Vonesh, 2012).

When it comes to comparing estimation methods for different variable types researchers have focused on what has been described as the “worst case scenario”. That is, a GLMM for a binary variable. This model is considered to be the worst case scenario because binary data is highly discrete, which means there is not much variability within or between individuals. All of the methods perform worse with GLMMs for binary data (Vonesh, 2012). This is especially true for MQL and PQL.

In addition to advising against using MQL or PQL with binary data, researchers have also cautioned against using these methods altogether when the random effect variances are large (Goldstein & Rasbash, 1996). Additionally, some researchers suggest using the restricted pseudo likelihood versions of MQL and PQL rather than the
maximum pseudo likelihood versions because they tend to produce less biased estimates (Breslow & Lin, 1995; Lin & Breslow, 1996). Other studies, however, have found the differences between parameters estimates using restricted pseudo likelihood and maximum pseudo likelihood to be quite small (Pinheiro & Chao, 2006).

One additional disadvantage of PQL and MQL is that they are estimated using pseudo likelihood rather than maximum likelihood. As a result, nested models cannot be compared using likelihood ratio tests or other standard measures of model fit. Some alternative measures of fit have been suggested by Vonesh, Chinchilli, and Pu (1996), which are similar in spirit to the standard $R^2$ measure used in regression analysis. However, these approaches are not as natural as the likelihood-based measures typically used for mixed models.

Despite each of these findings, MQL and PQL are the methods most frequently used by analysts (Rodriguez & Goldman, 2001). PQL in particular is the default estimation method in many GLMM software packages. For example, PQL using RPL is the default approach in SAS PROC GLIMMIX (SAS Institute, 2011a), HLM (Raudenbush, Bryk, & Congdon, 2004) and SPSS (IBM Corp., 2012). MQL and PQL converge very quickly and, as was mentioned before, perform well under many conditions. Therefore, these approaches should not be disregarded altogether. The problem is that because one does not typically know the parameter estimates before fitting a model, it is not always obvious whether you are in a condition where PQL or MQL is known to perform well.
AGQ is generally the least biased estimation method. However, it is not always a feasible approach for more complicated problems. One study found that there is little difference between using AGQ with 5 quadrature points and AGQ using 7 quadrature points in GLMMs with a repeated measures binary variable (Pinheiro & Chao, 2006). Based on this finding, these authors concluded that 5 quadrature points is often sufficient for GLMMs, which reduces the computational burden of AGQ slightly. In many cases Laplace provides a more efficient alternative to AGQ. For example, Joe (2008) recommends using Laplace for quick comparison of competing models (e.g., to determine the number of covariates or the structure of the random effect covariance matrix), then using AGQ to estimate a final model. Similarly, Rodriguez and Goldman (2001) suggest fitting GLMMs using several estimation methods and comparing the results.
Given the large investment that is required to complete a longitudinal study, it is rare for researchers to only collect information on a single variable over time. Instead, there is a tendency to collect an entire battery of measurements at each occasion. When it comes to analysis of this type of data, an intuitive question is whether the outcome variables should be modeled independently or simultaneously.

Obviously, a major factor in this decision is the research questions one hopes to answer with the analysis. Many questions can be answered with univariate analyses. When this is the case, it is not always beneficial to use a multivariate model. For example, researchers have found that there are often minimal efficiency gains in fitting multivariate models. This is especially true when the sample size is large (Lesaffre & Molenberghs, 1991), relationships between the outcomes are not strong (Gueorguieva & Agresti, 2001), and there is not missing data (Fitzmaurice & Laird, 1997; Gueorguieva & Sanacora, 2006). Additionally, Fieuws, Verbeke, Maes and Vanrenterghem (2008) showed that models which include several uncorrelated repeated measures variables are no better than univariate models at producing accurate predictions.

Longitudinal data often does not fit into these categories, however. Perhaps the most obvious violation of the previous conditions is the fact that missing data is particularly common in longitudinal studies. When this is the case (or when any of the
other conditions are violated) multivariate models can be an appealing approach. This
point will be discussed in further detail in the following sections.

Additionally, there are certain research questions that can only be answered with
models for multivariate longitudinal data. The questions of interest in multivariate studies
have been described as dealing with the “evolution of the associations” and also the
“association of the evolutions” (Fieuws & Verbeke, 2004). For example, researchers have
used multivariate models to assess the association between smoking cessation and weight
gain (Liu, Daniels, & Marcus, 2009), the relationship between motivation and perceived
competence throughout one’s first semester of high school (Ferrer & McArdle, 2003), the
association between patterns of change in several hormones during a stressful situation
(MacCallum, Kim, Malarkey & Kiecolt-Glaser, 1997), and the relationship between the
rates of hearing loss at different frequencies (Fieuws & Verbeke, 2006).

Of course, there are many different ways to approach the analysis of multivariate
longitudinal data. In this chapter, several approaches will be summarized. The advantages
and disadvantages of each approach will be discussed. Finally, the multivariate GLMM
will be introduced in the final section of this chapter.

4.1 Models for Multivariate Longitudinal Data

Recent reviews of multivariate longitudinal data analysis have been written by
Bandyopadhyay, Ganguli, and Chatterjee (2011), Verbeke and Davidian (2009), Faes,
Geys, and Catalano (2009), and Verbeke, Fieuws, Molenberghs and Davidian (2012).
Additionally, a special issue of Statistical Methods in Medical Research was entirely
devoted to the topic ("Longitudinal Analysis Methods for Multivariate Data", 2007). Each of these reviews and the references contained within will be summarized in this section.

The classification system used here will follow that of Verbeke et al. (2012), simply because it seems to be the most comprehensive approach. They organize methods for the analysis of multivariate longitudinal data into four categories. These can be summarized as in Figure 2. Their classification scheme depends on whether the model involves a latent variable on the outcome dimension and/or on the time dimension.

The term “latent variable” is used in the traditional sense when referring to the outcome dimension. However, their use of the term latent variable to describe the time dimension is a bit unconventional. By this they mean to identify models where the repeated measurements can be viewed as a manifestation of time, which is evolving latently. A more practical way of looking at this is models that can handle data that is unbalanced between outcome variables (i.e., that which differs either in the number of observations per individual or in the particular observation times for each outcome variable), are considered to have a latent variable for the time dimension. Models that require data to be balanced between outcome variables are not considered to have a latent time dimension.
4.1.1 Conditional and Marginal Models

The goal of conditional and marginal models is to specify directly the joint density of the outcome variables. Some of the methods involve conditioning on various parts of the data. This set of methods is referred to as conditional models. Methods which do not involve conditioning, and instead directly specify the marginal distribution of the variables are referred to as marginal models. Both sets of approaches will be discussed in the following sections.

Conditional Models

The most common type of conditional model reduces the dimension of the dependent variable by specifying the distribution of one longitudinal variable, conditional on all other longitudinal variables. In other words a univariate model is specified for one longitudinal outcome where all of the other longitudinal variables are included as time-
varying covariates. For example, this type of model for two continuous outcome variables could have a form similar to the following

\[ y_{ij}^1 = \beta_0 + \beta_1 t_{ij} + \beta_2 y_{ij}^2 + e_{ij} \tag{24} \]

Here, the \( j \)th observation of the first outcome variable for individual \( i \) \( (y_{ij}^1) \) is a linear function of time \( (t_{ij}) \) and the second outcome variable at the same time point \( (y_{ij}^2) \).

Although this approach is commonly used in practice, it has obvious drawbacks. Primarily, the parameter estimates and the interpretation of the model are affected by the choice of dependent variable. To be concrete, consider two outcome variables, \( y^1 \) and \( y^2 \), measured over time. There are two possible conditional models for the variables

\[ y^1 = g(t_i, y^2; \beta, u_i) + e_i \tag{25} \]

and

\[ y^2 = g(t_i, y^1; \beta, u_i) + e_i \tag{26} \]

In both (25) and (26) \( g(\cdot) \) represents a linear function of covariates and parameters. The model for \( y^1 \) in (25) includes \( y^2 \) as a time-varying covariate. The model for \( y^2 \) in (26) includes \( y^1 \) as a time varying covariate. Even though using one variable to predict the other might not align with substantive theory, a conditional model requires that researchers make such a decision. Additionally, the models in (25) and (26) will produce different sets of parameters, and it will be difficult to make inferences about the marginal evolutions of \( y^1 \) and \( y^2 \) based on those parameters. Finally, if the variables are measured
at different time points or if there is missing data present, the researcher will also be forced to make decisions about how to handle that component. When more than two outcome variables are modeled simultaneously, these complications are compounded even further.

In longitudinal data analysis specifically, another type of conditional model is transition models. In transition models the observation at a given time point is modeled conditional on all other observations and also previous time points. For example, a possible transition model for two outcome variables could be similar to (24) except for it could also include the previous observation of the dependent variable as a predictor. Specifically,

\[ y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 y_{ij} + \beta_3 y_{i,j-1} + \epsilon_{ij}. \]  

(27)

Like the model in (24) this model involves a linear function of time and the second outcome variable; however it also uses a linear function of the dependent variable at the previous time point to predict the value of the dependent variable at the current time point. This is a relatively simple structure. Transition models can involve much more complicated structures, including several previous time points or more elaborate functions of the previous data. For more details on this type of model, see Davis (2002) or Fitzmaurice and Molenberghs (2009).

Transition models are obviously designed for the analysis of data that is collected over time. However, a number of features limit their utility in longitudinal data analysis and instead make them better suited for time series analysis. Specifically, transition
models are best suited for data where the repeated measures are equally spaced and missing data is not present. Additionally, the parameters of transition models are particularly sensitive to the model for time dependence. Because the parameters for the covariates are adjusted for a subject’s response history, it is also sometimes difficult to interpret the parameters.

Marginal Models

In contrast to conditional models, marginal models provide direct inference for the marginal evolutions of the longitudinal variables. When the data are balanced in terms of measurement occasions and the outcome variables are all normally distributed, a marginal model that can be used is a multivariate fixed-effects model. Verbeke et al. (2012) refer to these models as multivariate regression models. However, the term multivariate fixed-effects model seems to align better with the methods that are used in the analysis of longitudinal data. To be concrete about what is meant by these terms, the model will be described briefly.

The basic idea is that each response is allowed to follow its own trajectory, but the residuals (both over time and over outcome variables) are allowed to be correlated. The model is similar to the mixed-effects model covered in Chapter 2.1, except that it does not include any random-effects and it is able to accommodate more than one dependent variable. For example, a multivariate fixed-effects model for \( r \) outcome variables can be written as

\[
\text{...}
\]
\[ y_i^1 = X_i^1 \beta_i + e_i^1 \]
\[ \vdots \]
\[ y_i^r = X_i^r \beta_i + e_i^r \]

(28)

where \( i \) indexes individuals and the term in the superscript indexes the outcome variables.

The design matrix for the \( r^{th} \) outcome variable is \( X_i^r \), the parameter vector for the \( r^{th} \) outcome variable is \( \beta_i^r \), and the outcome-specific residual vector for the \( i^{th} \) individual is \( e_i^r \).

Let \( y_i = \left( y_i^1, \ldots, y_i^r \right)' \), \( e_i = \left( e_i^1, \ldots, e_i^r \right)' \), \( X_i = \left( X_i^1, \ldots, X_i^r \right)' \), and \( \beta = \left( \beta_i^1, \ldots, \beta_i^r \right)' \). Because there are no random-effects in this model, \( \text{cov}(y_i) = \text{cov}(e_i) = \Lambda_i \).

As in the mixed-effects model, \( y_i \) is assumed to be normally distributed. In this case, that distribution has the following form

\[ y_i \sim N(\mu_i, \Lambda_i). \]

(29)

The covariance matrix \( \Lambda_i \) provides information about both the association within each outcome variable over time and also the association between the outcome variables.

One potential drawback of a model like this is the large number of parameters that are required. If each variable is measured at \( n \) time points and \( \Lambda_i \) is unstructured, it will have \( rn(rn+1)/2 \) parameters, where \( r \) is again the number of outcome variables. Of course, these parameters are also in addition to the fixed effect parameters that are used to model each outcome variable. There are ways to structure \( \Lambda_i \) in order to reduce the
number of parameters. Many of these parameterizations can also aid in interpretation of
the model. However, in the end one is still often left with a very large number of
parameters.

In addition to the large number of parameters that are often required to fit this
kind of model, dealing with missing data and non-continuous variables is often
problematic. Some work has been done on similar models that can be used with non-
continuous variables. For examples see Ten Have and Morabia (1999) or Molenberghs
and Lesaffre (1994).

4.1.2 Multivariate Mixed Models

The term multivariate mixed model is used to describe models like those
discussed in Chapter 2.2, but that have been extended to accommodate multiple outcome
variables. These models can also be seen as extensions of the multivariate fixed-effects
model that was just described to include random-effects. The general idea is that a mixed
model is defined for each outcome variable, and then the outcome variables are related
through the random-effects. This is in contrast to the multivariate fixed-effects model
where the outcomes are related through the residual covariance matrix.

The relationship between the outcome variables can be induced through the
random-effects in one of two ways. The first approach is to include random-effects that
are common between the outcomes. These models are referred to as shared random-
effects models. The second approach is to include unique random-effects for each
outcome variable, and then model the relationship between the outcome variables through the covariances of the random-effects. These are referred to as correlated random-effects models. Both of these approaches will be discussed in more detail in the following sections. However, I will first discuss the advantages and disadvantages of this general class of methods.

There are several advantages in using multivariate mixed models to analyze multivariate longitudinal data. As many authors have noted, when all of the outcomes are continuous, multivariate mixed models can be implemented in standard software for mixed models (Thiebaut, Jacqmin-Gadda, Chene, Leport, & Commenges, 2002; Fieuws & Verbeke, 2004; Snijders & Bosker, 1999). This is done by stacking the outcomes in a single vector (as if it were one outcome) and then using dummy-coded variables to assign only certain parameters to each outcome.

Another benefit of multivariate mixed models has to do with missing data. As Schafer and Yucel (2002) pointed out, it is much easier to accommodate missing data on a dependent variable than it is to accommodate missing data on a predictor. This is because most mixed models are estimated with some form of full information maximum likelihood (FIML). Because an individual’s contribution to the likelihood function is calculated one subject at a time, it is possible to work only with the data that is present for that subject, ignoring everything that is missing. As long as the data are missing at random (MAR) valid inferences can be obtained even though there is not complete information.
The association of multivariate mixed-models with missing data goes further than this, however. Recall that data are defined to be MAR if the probability of missingness depends on a value or variable that is observed, not missing (Little & Rubin, 2002, Ch. 1). This means that if the probability of missingness on $y^1$ depends on $y^2$, an analysis of $y^1$ excluding $y^2$ would violate the MAR assumption. However, if $y^2$ is included in the model then the missingness on $y^1$ is considered MAR and FIML can be used (McCulloch, 2008).

Multivariate mixed models are also often used specifically in the context of missing data when the missingness is non-ignorable. Two types of models that fall into this category are selection and pattern-mixture models. Both have been widely reviewed in the missing data literature. Introductions to these approaches can be found in Molenberghs and Fitzmaurice (2009) or Little (2009).

Other general benefits of multivariate mixed models are that they provide a more complete picture of the separate and joint effects of variables and help avoid multiple testing by allowing for overall tests of the effects of predictors (e.g., through likelihood ratio tests as opposed to using ad hoc methods) (McCulloch, 2008). Although the models discussed in this section will be for continuous variables only, it is possible to extend them to include variables of different types. This is because the relationship between the variables is manifested through the random-effects rather than the residuals. The name for this particular kind of model is multivariate generalized linear mixed models. These will be discussed in more detail at the end of this chapter.
Shared Random-effects

The first approach to assessing the relationship between dependent variables using a multivariate mixed model involves introducing random-effects which are shared between the dependent variables. For example, a model for two normally distributed outcome variables measured at time $j$ for individual $i$ assuming linear change in each is:

$$
y_{ij}^{1} = \left( \beta_{0i}^{1} + u_{0i} \right) + \left( \beta_{1i}^{1} + u_{1i} \right) t_{ij} + e_{ij}^{1}
$$
$$
y_{ij}^{2} = \left( \beta_{0i}^{2} + u_{0i} \right) + \left( \beta_{1i}^{2} + u_{1i} \right) t_{ij} + e_{ij}^{2}.
$$

In this model, the outcome variables have separate linear trajectories defined by the variable-specific fixed-effects ($\beta_{0i}^{1}, \beta_{1i}^{1}, \beta_{0i}^{2},$ and $\beta_{1i}^{2}$). The two outcome variables are linked by shared random-effects for the intercept ($u_{0i}$) and slope ($u_{1i}$). The random-effects are generally assumed to be normally distributed, as in the mixed-effects model for a single outcome variable. The form of this distribution is similar to that in (2). Of course, complete specification of this model would also require distributional assumptions for the residuals. In general, this will be similar to (3) except for the dimension of $\Lambda_i$ will be $(rn \times rn)$, at least in the case of data that is measured at the same number of time points for each individual and each outcome variable. In the two variable example presented above, $r = 2$.

A benefit of the shared random-effects approach is that when it comes to estimation of the model, the dimension of integration does not increase. Recall from Chapter 3 that the dimension of integration is equal to the number of random-effects.
Since the random-effects are shared between the variables, the number of random-effects in the model does not increase with the number of outcome variables. To be thorough, it is worth pointing out that this benefit actually does not apply to linear mixed models since the marginal distribution can be obtained directly. However, it can reduce the computational problem quite substantially with other types of variables. Finally, one drawback of the shared random-effects approach is that it implies that there is a perfect positive correlation between the shared random-effects (McCulloch, 2008). For obvious reasons, this is not typically a very realistic assumption.

**Correlated Random-effects**

A second approach is to include separate random-effects for each dependent variable. The association between the dependent variables is then assessed by allowing the random-effects to be correlated. For example, a model for two normally distributed outcome variables measured at time \( j \) for individual \( i \) could look something like this:

\[
\begin{align*}
    y_{ij}^1 &= (\beta_0^1 + u_{i0}^1) + (\beta_1^1 + u_{i1}^1) t_{ij} + e_{ij}^1 \\
    y_{ij}^2 &= (\beta_0^2 + u_{i0}^2) + (\beta_1^2 + u_{i1}^2) t_{ij} + e_{ij}^2
\end{align*}
\]  

(30)

As in the shared random-effects model, the population-level trajectories for the outcome variables are modeled using distinct fixed-effects (\( \beta_0^1, \beta_1^1, \beta_0^2, \) and \( \beta_1^2 \)). However, the random-effects for each outcome variable are now distinct and allowed to be correlated. For example the distribution of the random-effects from (30) could be specified as
\[
\begin{pmatrix}
u_{0i}^1 \\
u_{ii}^1 \\
u_{0i}^2 \\
u_{ii}^2 
\end{pmatrix} \sim \mathcal{N}
\begin{pmatrix}
0 \\
0 \\
0 \\
0 
\end{pmatrix}
, 
\begin{bmatrix}
\varphi_{00}^1 & 0 & 0 \\
\varphi_{10}^1 & \varphi_{11}^1 & 0 \\
0 & \varphi_{10}^{12} & \varphi_{00}^2 \\
0 & \varphi_{11}^{12} & \varphi_{11}^{11} 
\end{bmatrix}
\]

In this notation, the two elements in the subscript refer to the type of random effect (0=intercept, 1=slope), and the elements in the superscript refer to the number of the outcome variable. For example, the parameter \( \varphi_{00}^{12} \) represents the covariance of the intercepts for the first and second outcome variables. The parameter \( \varphi_{10}^2 \) represents the covariance of the slope and intercept for the second outcome variable. This structure indicates that the within-outcome intercepts and slopes (\( \varphi_{10}^1 \) and \( \varphi_{10}^2 \)), between-outcome intercepts (\( \varphi_{00}^{12} \)), and between-outcome slopes (\( \varphi_{11}^{12} \)) are allowed to be correlated. The distribution of the residuals is again assumed to be similar to (3) except for the dimension of \( \Lambda_i \) will be \( rn \times rn \), at least in the case of data that is measured at the same number of time points for each individual and each outcome variable. For this example specifically, the residual structure could look something like this.
In contrast to the shared random-effects approach, this model allows the correlation between random-effects to be positive or negative. By examining these correlations, one can quickly get an idea of the relationship between the trajectories of the dependent variables. For example, a strong positive correlation between the slopes for two outcome variables could indicate that individuals who learn math quickly also tend to learn reading quickly.

It is also possible to calculate a measure of the “evolution of association” (i.e. how the correlation changes over time) based on the parameter estimates from the model (Fieuws & Verbeke, 2004). This is done by calculating the correlations of the outcome variables at specific time points using the parameter estimates from the model rather than the raw data. For the model given in (30), (31) and (32), the marginal correlation between the two variables as a function of time is

\[
\rho_{ij} = \frac{\varphi_{00}^{12} + t_y \varphi_{11}^{12}}{\sqrt{\varphi_{00}^{2} + 2t_y \varphi_{10}^{2} + t_y^2 \varphi_{11}^{2}} + \sigma^2_{12} \sqrt{\varphi_{00}^{2} + 2t_y \varphi_{10}^{2} + t_y^2 \varphi_{11}^{2} + \sigma^2_{12}}}.
\]
One other advantage of correlated random-effects models is that they can lead to more accurate predictions when random-effects are allowed to be correlated (Hall & Clutter, 2004; Fieuws, Verbeke, Maes, & Vanrenterghem, 2008). Compared to the shared random-effects approach, however, one disadvantage of this approach is that the dimension of integration quickly increases as more random-effects are added to the model. The linearization methods that were discussed in Chapter 3.1 can lessen the computational burden a little, but it is still difficult to estimate models that have a very large number of random-effects. In fact, some work has specifically been done on high dimensional multivariate mixed-effects models (Fieuws, Verbeke, Boen & Delecluse, 2006; Fieuws & Verbeke, 2009; Fieuws & Verbeke, 2006). In these models, the authors suggest using a pairwise fitting approach, where all combinations of outcome variables are fit two at a time using multivariate mixed models and then parameter estimates are combined to create a single set of results.

4.1.3 Dimension Reduction

Another set of approaches that can be used in the analysis of longitudinal data is dimension reduction methods. The general idea of these approaches is to summarize the outcome variables with a single composite score and assess the change in the composite score over time. For example, Oort (2001) used factor analysis to reduce the dimension of the dependent variable and then evaluated change over time through the change in the mean of the latent variable. Another approach is to use principal components analysis (Timmerman & Kiers, 2003).
The interpretation of this type of model is quite different from the multivariate mixed models discussed previously. Here, parameter estimates will yield a description of change in the composite variable over time, not information on the association between the trajectories of the variables. There are also a number of ways to create a composite score, especially when dealing with different variable types. Different approaches will lead to differences in final model estimates. An additional disadvantage is that it can be difficult to deal with missing data with some of the dimension reduction approaches.

4.1.4 Second Order Latent Growth Curves

Second order latent growth curve models are also a way to approach multivariate longitudinal data. This approach combines factor analysis and latent growth curve modeling by forming a latent variable at each time point and then assessing the change in that latent variable over time.

Define a model for three variables measured each measured at $n$ time points to be
Here, \( \eta_1, \ldots, \eta_n \) represent the latent variables that are formed at each of the measurement occasions. The matrix \( \Delta_i \) is a factor loading matrix which defines the relationship between the manifest and latent variables. The model in (33) defines what is referred to as the measurement model. The structural model defines the trajectory for the latent variable. This is written as

\[
\begin{pmatrix}
  y_{1i}^1 \\
y_{1i}^2 \\
y_{1i}^3 \\
\vdots \\
y_{ni}^1 \\
y_{ni}^2 \\
y_{ni}^3
\end{pmatrix} =
\begin{pmatrix}
  1 \\
  \delta_i^2 \\
  \delta_i^3 \\
  \vdots \\
  1
\end{pmatrix}
\begin{pmatrix}
  \eta_1 \\
  \vdots \\
  \eta_n
\end{pmatrix} +
\begin{pmatrix}
  e_{1i}^1 \\
  e_{1i}^2 \\
  e_{1i}^3 \\
  \vdots \\
  e_{ni}^1 \\
  e_{ni}^2 \\
  e_{ni}^3
\end{pmatrix}.
\]

\( \mathbf{y}_i = \Delta_i \mathbf{\eta} + \mathbf{e}_i \) (33)

where \( \mathbf{T} \) is a design matrix which, in this model, specifies a linear trajectory for the latent variable. The latent variables at each of the measurement occasions also have a residual, \( \varsigma_j \). The residuals for the observed and latent variables, \( \mathbf{e}_i \) and \( \varsigma \), respectively, are assumed to be normally distributed with means equal to zero and unknown variances. The latent variables in the structural model, \( \mathbf{\xi} \), are assumed to follow a multivariate normal distribution. Similar to fixed-effects, the means of that distribution are the parameter
estimates for the intercept and slope of the latent variable. Similar to random-effects, the variances of $\xi$ give an indication of the variability between individuals on the intercept and slope, and the covariance gives an indication of the association between the two terms. A path diagram for this model is given in Figure 3.

This type of model can be estimated in most structural equation modeling software. The interpretation is similar to the composite score approach, but the way in which the outcome variables are combined is more empirical and a bit more appealing intuitively. Models such as this are not uncommon in the latent variable literature (Proust, Jacqmin-Gadda, Taylor, Ganiayre, Commenges, 2006; Harring, 2009; Hancock, Kuo, & Lawrence, 2001).

Like the dimension reduction methods, however, the second order latent growth curve models don’t provide answers to the questions that are often asked in the analysis of multivariate longitudinal data. Specifically, they provide only a measure of the change in the latent variable over time, not the change in the individual variables. Conceptually, if the latent variable structure is reasonable, changes in the observed variables will be manifested through changes in the latent variable. However, this overall measure of change will not provide a way to determine how similar or distinct the trajectories for each outcome variable are. They also do not provide any direct measure of the association between the trajectories of the observed variables.
Figure 3. Path diagram for a second order latent growth curve model
4.2 Multivariate Generalized Linear Mixed Models

In the same way that the multivariate mixed model extends the linear mixed model to accommodate multiple continuous dependent variables, the multivariate generalized linear mixed model extends the generalized linear mixed model to handle multiple dependent variables that are possibly of different scale types. This can be done by stacking the repeated measures of each of the response variables into a single column vector. Define the composite response vector as

$$y_i = \begin{bmatrix} y_i^1 \\ \vdots \\ y_i^r \end{bmatrix},$$

(35)

where, for example, $y_i^1$ is the $(n_i^1 \times 1)$ vector of responses for individual $i$ on the first response variable. The dimension of $y_i$ is $(g_i \times 1)$ where $g_i = \sum_{c=1}^{r} n_i^c$. This allows each individual to be measured at a different number of occasions for each response variable. In the case when all response variables are measured at the same number of occasions for each individual, the dimension of $y_i$ is $(rn \times 1)$.

The linear predictor for the multivariate model is again defined by concatenating the vectors from the GLMM in (7). This can be written as
\[
\eta_i = X_i \beta + Z_i u_i
\]

Also define \( \mu_i \) in a similar way as

\[
\mu_i = g^{-1}(X_i \beta + Z_i u_i)
\]

The relationship between the dependent variables is induced through the distribution of the random-effects. As in the univariate GLMM, the random-effects are assumed to follow a multivariate normal distribution. If the covariances between the random-effects for different dependent variables are fixed at zero, then the dependent variables will be independent. Otherwise, any covariances that are allowed to be freely estimated will provide information about the relationship between the trajectories of the two dependent variables.

Conditional on the random-effects, the outcome variables are assumed to be independent. This means that the marginal distribution of \( y_i \) can be specified as
\[
\begin{align*}
 f\left(y_1, \ldots, y_i \right) &= \int f\left(y_1, \ldots, y_i | u_i \right) f\left(u_i \right) du_i \\
 &= \int f\left(y_1 | u_i \right) \cdots f\left(y_i | u_i \right) f\left(u_i \right) du_i
\end{align*}
\]  

(38)

where \( f\left(y_1 | u_i \right), \ldots, f\left(y_i | u_i \right) \) are defined as they were in (10).

The estimation for these models can be carried out using the same methods that are appropriate for univariate GLMMs, many of which were described in Chapter 3. If the link functions and marginal distributions are the same for each outcome variable, it is not difficult to implement the multivariate GLMM in standard GLMM software. However, not all software packages are flexible enough to handle the multivariate GLMM when the link functions and/or marginal distributions are not the same for each outcome variable.

4.2.1 Example

The BMI and television variables from the NLSY dataset will now be used to illustrate the multivariate GLMM framework. Let \( Y_1 \) be the \((n_1 \times 1)\) response vector for the BMI data and \( Y_2 \) be the \((n_2 \times 1)\) response vector for the television data. Using the same models that were defined in Chapter 2.2.1 for the BMI data and Chapter 2.2.2 for the television data, the linear predictor for the multivariate GLMM can be written as
\[
\eta_i = X_i \beta + Z_i u_i
\]

\[
\begin{pmatrix}
\eta_i^1 \\
\eta_i^2
\end{pmatrix} = \begin{pmatrix}
X_i & 0 \\
0 & X_i^2
\end{pmatrix} \begin{pmatrix}
\beta^1 \\
\beta^2
\end{pmatrix} + \begin{pmatrix}
Z_i & 0 \\
0 & Z_i^2
\end{pmatrix} \begin{pmatrix}
u_i^1 \\
u_i^2
\end{pmatrix}
\]

\[
\begin{pmatrix}
\eta_{1i} \\
\eta_{2i}
\end{pmatrix} = \begin{pmatrix}
1 & \text{age}_{1i} - 10 & 0 & 0 \\
0 & 0 & 1 & \text{age}_{1i} - 10
\end{pmatrix} \begin{pmatrix}
\beta_0^1 \\
\beta_1^1 \\
\beta_0^2 \\
\beta_1^2
\end{pmatrix} + \begin{pmatrix}
1 & \text{age}_{1i} - 10 & 0 \\
0 & 0 & 1 \\
0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
u_{0i}^1 \\
u_{1i}^1 \\
u_{0i}^2
\end{pmatrix}
\]

In (39) the intercept and slope for \( y_i^1 \) are both random, the intercept for \( y_i^2 \) is random, but the slope for \( y_i^2 \) is fixed only.

The distribution of the random-effects, \( f(u_i) \), is

\[
\begin{pmatrix}
\nu_{0i}^1 \\
\nu_{1i}^1 \\
\nu_{0i}^2
\end{pmatrix} \sim N \begin{pmatrix}
0 \\
0 \\
0
\end{pmatrix}, \begin{pmatrix}
\phi_{00}^1 & \phi_{01}^1 & \phi_{02}^1 \\
\phi_{10}^1 & \phi_{11}^1 & \phi_{12}^1 \\
\phi_{20}^1 & \phi_{21}^1 & \phi_{22}^1
\end{pmatrix}
\]

This structure is based on the univariate GLMMs from Chapter 2, and also on a series of multivariate models that were fit to assess the relationship between the random-effects for BMI and television rules. The random-effect intercepts for each variable are allowed to
be correlated, which induces a relationship between the outcome variables. However, the covariance between the random effect intercept and slope for the BMI variable is constrained to be 0 as it was in the univariate model. The covariance between the random effect slope for the BMI variable and the random effect intercept for the television rules variable is also constrained to be 0.

Once again, the identity link function is used for \( y_i^1 \) and the logit link function is used for \( y_i^2 \), meaning that

\[
\begin{align*}
\mu_i &= g^{-1}(X_i \beta + Z_i u_i) \\
\begin{pmatrix}
\mu_i^1 \\
\mu_i^2
\end{pmatrix} &= 
\begin{pmatrix}
X_i \beta^1 + Z_i u_i^1 \\
\logit^{-1}(X_i \beta^2 + Z_i u_i^2)
\end{pmatrix}
\end{align*}
\]  

(41)

The marginal distribution of \( y_i \) is

\[
f(y_i) = \int f(y_i^1 | u_i) f(y_i^2 | u_i) f(u_i) \, du_i ,
\]

(42)

where \( f(y_i^1 | u_i) \sim N(\mu_i^1, \Lambda_i) \), \( f(y_i^2 | u_i) \sim \prod_{j=1}^{n_i} \text{Bernoulli}(\mu_{ij}^2) \) and \( f(u_i) \) is as defined in (40). The residual covariance matrix for the BMI variable is diagonal with variance parameter \( \sigma_e^2 \). Once again, this model was fit using adaptive Gaussian quadrature in SAS PROC GLIMMIX. The parameter estimates are given in Table 7. The parameter estimates and run times for this model using several different estimation methods are also provided in Table 11 in Appendix A.
Table 7. Parameter estimates and standard errors for multivariate GLMM for television rules and BMI data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta^1_0$</td>
<td>18.86</td>
<td>.08</td>
</tr>
<tr>
<td>$\beta^1_1$</td>
<td>.88</td>
<td>.02</td>
</tr>
<tr>
<td>$\beta^2_0$</td>
<td>.54</td>
<td>.06</td>
</tr>
<tr>
<td>$\beta^2_1$</td>
<td>-.26</td>
<td>.02</td>
</tr>
<tr>
<td>$\varphi^1_{00}$</td>
<td>17.05</td>
<td>.72</td>
</tr>
<tr>
<td>$\varphi^1_{11}$</td>
<td>.33</td>
<td>.05</td>
</tr>
<tr>
<td>$\varphi^2_{00}$</td>
<td>2.19</td>
<td>.22</td>
</tr>
<tr>
<td>$\varphi^{12}_{00}$</td>
<td>-.37</td>
<td>.17</td>
</tr>
<tr>
<td>$\sigma^2_e$</td>
<td>4.34</td>
<td>.25</td>
</tr>
</tbody>
</table>

The parameter estimates for both variables are relatively unchanged from those presented for the univariate models. The covariance between the intercepts for the BMI and television data is $\hat{\varphi}_{00}^{12} = -.37$. This is large relative to its standard error indicating that there is some systematic variation in these two random-effects. Specifically those that are more likely to have rules about watching television at the age of 10 tend to have lower BMIs, and vice versa. This covariance corresponds to a correlation of -.06, indicating that although there is a relationship, it is not very strong. This could explain why the parameter estimates from the multivariate model are not very different from the parameter estimates from the univariate models.
A model which allowed both the intercepts and slopes between the two outcome variables to be estimated rather than constrained to zero was also fit. The results for this model looked fairly similar to those that were presented in Table 7 except for the covariance between the slopes was not significant. Although these results are not presented in detail here, it is worth noting that there does not appear to be a relationship between how quickly rules for television are removed and change in BMI.

One of the main reasons for the multivariate approach is to gain information about the random effect covariances. Although the covariance, \( \hat{\phi}_{12} \), is not large in the example that was presented, it is one thing to estimate the parameter and observe that it is small, and quite another thing to just assume it is zero without investigating. In other situations, ignoring the covariance between random-effects between two variables that may exist and that may be sizeable introduces bias to the estimation of all components of the model. The major benefit of the multivariate approach is that it offers a considerable improvement in the estimation of complicated repeated measures designs.
Chapter 5: Simulation Study

The goal of this simulation is to investigate how well multivariate GLMMs can be estimated using estimation methods that are currently popular in the analysis of GLMMs for a single dependent variable. Given what is known about the performance of the estimation methods with univariate GLMMs (see Chapter 3.3), there are several questions that are of particular interest in the multivariate context.

First, it is known that the variance parameters in GLMMs are typically the most difficult to estimate (Pinheiro & Chao, 2006). Although one hopes that all parameters can be estimated accurately, in univariate models the variance parameters are often of secondary interest. Many researchers simply hope to determine their nominal significance, using the variance parameters only to inform decisions about whether a random effect should be included in the model or not. In multivariate GLMMs, on the other hand, the variance and covariance parameters are often the focus of the study. In particular, a primary purpose of the multivariate structure is in determining the association between the trajectories for the dependent variables. This is done through the covariances of the random-effects. In this setting, researchers are interested not just in nominal significance, but also in the magnitude of the parameter estimate. Therefore, one
goal of this study is to determine how well the covariances between random-effects from different dependent variables can be recovered.

Another question that is of interest is how the presence of multiple dependent variables can impact the quality of the estimation. Simulations for univariate GLMMs have shown that parameter estimates for binary variables can be quite biased (Vonesh, 2012) but that a large number of observations within clusters can improve the results (Breslow & Clayton, 1993). I am interested in determining whether the presence of an additional dependent variable can also impact the results and whether the strength of the relationships between the random-effects plays a role in this result. For example, can the estimates for a binary variable be improved by simultaneously modeling a highly correlated continuous variable? Or, conversely, can the estimates for a continuous variable be negatively impacted by the presence of a highly correlated binary variable.

Lastly, an important consideration in fitting any model is balancing accuracy with practical computational issues. I am interested in how each of the methods performs with regard to convergence. In particular, are some of the estimation methods less likely to reach convergence? Do certain methods tend to produce non-positive definite covariance matrices?

The design of the simulation study will be discussed in the following section. The results of the simulation will be presented after that.
5.1 Design of Simulation Study

The design of this simulation study is modeled after those conducted by Breslow and Clayton (1993) and Vonesh (2012). The goal of both of those studies was to compare the performance of various estimation methods in the context of a GLMM for a binary variable measured over time. Again, comparisons of estimation methods for GLMMs have tended to focus on binary variables for two reasons. First, binary variables offer the challenge of being highly discrete. Second, binary variables are present in many disciplines, and as a result mixed-effects logistic regression models are often used.

In addition to the repeated measures binary variable that was analyzed in previous simulations, the present study will also incorporate a second dependent variable. This second variable will be continuous and normally distributed. In order to be as parsimonious as possible, each of the dependent variables is modeled using two fixed-effects (intercept and slope) and one random effect (slope). The random-effects for the slopes of the two dependent variables are allowed to be correlated. Individuals are measured at 7 time points, \( j = 1, 2, 3, 4, 5, 6, 7 \). However, time is centered in the model so that \( t_{ij} = -3, -2, -1, 0, 1, 2, 3 \). To be concrete, the linear predictor for the model will have the following form
\[ \eta_i = X_i \beta + Z_i \ u_i \]

\[
\begin{pmatrix}
\eta_{11}^i \\
\eta_{12}^i \\
\eta_{13}^i \\
\eta_{14}^i \\
\eta_{15}^i \\
\eta_{16}^i \\
\eta_{17}^i
\end{pmatrix} = 
\begin{pmatrix}
X_{1i} & 0 \\
0 & X_{2i}^2
\end{pmatrix}
\begin{pmatrix}
\beta_{1}^i \\
\beta_{2}^i
\end{pmatrix} + 
\begin{pmatrix}
0 \\
Z_{1i}^2
\end{pmatrix}
\begin{pmatrix}
u_{1i}^1 \\
u_{2i}^2
\end{pmatrix}
\]

\[ \eta_{11}^i = \begin{pmatrix} 1 & -3 & 0 & 0 \end{pmatrix} \begin{pmatrix} -3 & 0 \end{pmatrix} \\
\eta_{12}^i = \begin{pmatrix} 1 & -2 & 0 & 0 \end{pmatrix} \begin{pmatrix} -2 & 0 \end{pmatrix} \\
\eta_{13}^i = \begin{pmatrix} 1 & -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 \end{pmatrix} \\
\eta_{14}^i = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \end{pmatrix} \\
\eta_{15}^i = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} \\
\eta_{16}^i = \begin{pmatrix} 1 & 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_{0}^1 \end{pmatrix} \begin{pmatrix} 2 & 0 \end{pmatrix} \\
\eta_{17}^i = \begin{pmatrix} 1 & 3 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_{1}^i \end{pmatrix} \begin{pmatrix} 3 & 0 \end{pmatrix} + \begin{pmatrix} u_{1i}^1 \\
u_{2i}^2
\end{pmatrix}
\]

(43)

where \( \eta_{1i}^1 \) is the linear predictor for the binary variable and \( \eta_{1i}^2 \) is the linear predictor for the continuous variable. The covariance structure for the random-effects is

\[
\begin{pmatrix}
u_{1i}^1 \\
u_{2i}^2
\end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\
0 \end{pmatrix}, \begin{pmatrix} \varphi_{11}^1 & \varphi_{12}^1 \\
\varphi_{21}^1 & \varphi_{22}^2
\end{pmatrix} \right)
\]

(44)
where $\varphi_{11}^1$ is the variance of the random effect for the slope of the binary variable, $\varphi_{11}^2$ is the variance of the random effect for the slope of the continuous variable, and $\varphi_{11}^{12}$ is the covariance between the two random-effects.

Once again, the logit link function is used for the binary outcome variable $(y_i^1)$ and the identity link function is used for the continuous outcome variable $(y_i^2)$. To be concrete,

$$\mu_i = g^{-1}(X_i\beta + Z_iu_i)$$

$$\begin{pmatrix} \mu_i^1 \\ \mu_i^2 \end{pmatrix} = \begin{pmatrix} \text{logit}^{-1}(X_i\beta^1 + Z_iu_i^1) \\ X_i^2\beta^2 + Z_i^2u_i^2 \end{pmatrix}$$

The marginal distribution of $y_i$ is

$$f(y_i) = \int f(y_i^1 | u_i) f(y_i^2 | u_i) f(u_i) \, du_i,$$

where $f(y_i^1 | u_i) \sim \prod_{j=1}^7 \text{Bernoulli}(\mu_{ij}^1)$, $f(y_i^2 | u_i) \sim \text{Normal}(\mu_i^2, \Lambda_i)$, and $f(u_i)$ is as defined in (44). The residual covariance matrix for $y_i^2$ is diagonal with variance parameter $\sigma_e^2$.

There are 100 subjects in each sample. The fixed effect parameter values are $\beta_0^1 = -2.5$ (binary variable intercept), $\beta_1^1 = 1$ (binary variable slope), $\beta_0^2 = 15$ (continuous variable intercept) and $\beta_1^2 = 3$ (continuous variable slope). The sample size, number of time points, the fact the $t_{ij}$ is centered, and the parameter values for $\beta_0^1$ and $\beta_1^1$ were each
selected in order to aid in comparison with the previous studies by Breslow and Clayton (1993) and Vonesh (2012). These pieces all correspond to components of their simulations with a single binary variable measured over time.

There are two variance parameters that are constant throughout the simulation. These are the variance of the random effect for the continuous variable which has a value of $\varphi_{11}^2 = 3$ and the residual variance which has a value of $\sigma^2_e = 2$.

There are two conditions that will be varied in this simulation. The first condition is the variance of the random effect for the slope of the binary variable. This condition is relevant because previous studies (Vonesh, 2012) found that when the variance of the random-effects is small, the estimation methods perform better overall and also more similarly to each other. There will be three variance conditions corresponding to $\varphi_{11}^1 = .25$, $\varphi_{11}^1 = .50$, and $\varphi_{11}^1 = 1.0$.

The second condition is the correlation between the random-effects for the slopes of the binary and continuous variables. This condition is of interest because it will help to answer the question of whether the parameter estimates for either dependent variable can be affected (either positively or negatively) by an association with another variable. This condition will also help answer questions related to how well multivariate GLMMs can estimate the correlations between random-effects for different variables. Because the major focus of multivariate GLMMs is often on these parameters, it is important to know how accurately they can be recovered.

There will be three correlation conditions. These correspond to $\rho^{12} = 0$ (no relationship), $\rho^{12} = .3$ (moderate relationship) and $\rho^{12} = .7$ (strong relationship). In terms
of the model parameters, $\rho^{12} = \frac{\phi^{12}_{11}}{\sqrt{\phi^{11}_{11} \phi^{22}_{11}}}$. Although the conditions for the simulation are defined in terms of correlations for the random-effects, the model will actually be estimated using covariances for the parameters. Because the covariance depends on the variances of the random-effects which also change depending on the condition, the models are defined in terms of correlations so that the strength of the relationship between the random-effects remains consistent.

Each of the three variance conditions will be crossed with each of the three correlation conditions, resulting in nine unique cells. There will be 300 replications within each cell.

The estimation methods that will be compared are penalized quasi-likelihood using restricted maximum likelihood (R-PQL), penalized quasi-likelihood using maximum likelihood (M-PQL), marginal quasi-likelihood using restricted maximum likelihood (R-MQL), marginal quasi-likelihood using maximum likelihood (M-MQL), Laplace approximation (LA), and adaptive Gaussian quadrature with 3 (AGQ3) and 7 (AGQ7) quadrature points. These particular estimation methods were selected because they are the most widely-available approaches. As was mentioned previously, all of the major mixed model software packages implement at least a subset of these methods. Additionally, the two variations of AGQ are used because previous studies indicated that the improvement in performance when going from 5 quadrature points to 7 quadrature points was minimal in univariate GLMMS for a binary variable (Pinheiro & Chao, 2006). I would like to evaluate the impact of the number of quadrature points within the multivariate framework. Also, both the restricted maximum likelihood and maximum
likelihood versions of PQL and MQL are used because there have been mixed reports on whether restricted maximum likelihood results in an improvement in estimates for GLMM models (Breslow & Lin, 1995; Lin & Breslow, 1996; Pinheiro & Chao, 2006).

In this simulation, SAS PROC GLIMMIX will be used for estimation. This decision is based on accessibility, ease of estimating the multivariate models, and the availability of a wide-range of built-in estimation methods. Although a number of software packages for mixed models can be used to estimate GLMMs, very few are flexible enough to allow for estimation of models with multiple dependent variables with different distributions and link functions. Syntax for running a basic multivariate GLMM with two dependent variables in PROC GLIMMIX will be included in Appendix B.

It is worth noting that there is, of course, a difference between an estimation method and the implementation of that estimation method. Although I have tried to be clear about the details of the methods that are being used, the fact remains that the estimates for this study are inevitably influenced by the fact that they are obtained using the implementation of these procedures within SAS. In fairness, however, this is true of all studies of this nature, whether the estimation is done in SAS, another software package, or using procedures that are written entirely by the researcher. In general, one would hope that even though factors such as run times may vary across implementations, final model estimates should be comparable. For a comparison of software packages for generalized linear mixed models see Zhang et al. (2011) or Zhou, Perkins, and Hui (1999).
5.1.1 Criteria for evaluating performance

The estimation methods will be compared based on a number of factors. The first factor is the stability of the estimation routine. One thing that will be considered within this factor is how often the model converges. The quasi-Newton method is used for optimization, and convergence is determined when one of three conditions are met: 1) the relative change in the gradient between successive iterations is less than 1e-12, 2) the absolute value of the largest gradient element is less than 1e-5, or 3) the relative change in the function value between successive iterations is less than some quantity determined by the machine precision.

Determining convergence for the linearization methods is a bit more complicated since they are doubly-iterative procedures. The two sets of iterations are referred to as the inner and outer iterations. The inner iteration produces estimates for the linear mixed model based on the pseudo data. Convergence for this part of the model is determined by the criteria that were mentioned above. The outer iteration determines whether the parameter estimates that were produced are stable over cycles of the procedure or if additional iterations through the linearization process should be completed. The convergence criteria for this part of the model looks at the relative change in the parameter estimates between iterations. If the largest change for any of the parameters is less than 1e-8, then the estimation process terminates.

Another issue that will be considered with regard to the stability of the estimation method is how many of the models that do converge result in estimates that do not seem
reliable. For example, a common problem in GLMMs is covariance matrices that are not-positive definite.

The second factor that will be considered is the accuracy of the parameter estimates. The parameter estimates will be compared in terms of bias, variability, and parameter recovery. To elaborate, measures of bias will help determine whether certain methods have a tendency to overestimate or underestimate specific types of parameters. For example, it could be the case that all methods estimate the fixed-effects fairly well, but some tend to underestimate the variance parameters. For each condition and estimation method, the average of the parameter estimates will be presented in addition to the average bias. The average bias is defined as \( \frac{\sum (\hat{\theta}_i - \theta)}{N} \), where \( \theta \) represents any parameter and \( N \) is the number of models that converged and had positive definite covariance matrices for the given condition and estimation method.

In addition to bias, it is also important to look at how variable the parameter estimates are. Bias gives a measure of average performance, but small variability indicates that the estimation method performs well consistently. For each condition and estimation method the standard deviation of the parameter estimates and the average mean square error (MSE) will also be reported. The average MSE is defined as

\[ \frac{\sum (\hat{\theta}_i - \theta)^2}{N}. \]

A final way of looking at the performance of the estimation methods is through the confidence intervals that are produced for each parameter. Specifically, it is of interest to know how often the confidence intervals cover the true value of the parameter.
This is important because it gives an overall indication of how often one would make correct inferences based on the results of the model. It is worth noting that procedures for testing elements of the random-effects covariance matrix are more complicated than those used for testing the fixed-effects. More specifically, the issue has to do with testing parameters that are on the boundary of the parameter space (e.g., $H_0: \varphi_{11} = 0$). When evaluating parameters of this sort, a Satterthwaite approximation is used to construct the limits of the confidence interval. The resulting upper and lower bounds are not necessarily symmetric about the point estimate of the parameter. For additional information see Burdick and Graybill (1992) or SAS Institute (2011b, p. 3130-3131).

5.2 Results

5.2.1 Stability of Estimation Methods

Of the 2700 multivariate GLMMs that were run, 31 did not converge. The estimation methods and conditions under which convergence did not occur are listed in Table 8. It appears that the integral approximation methods (AGQ7, AGQ3 and LA) have a more difficult time reaching convergence when the correlation between random-effects is large but the variance of the random-effects is small to moderate. Even so, the rates of non-convergence are still fairly low, with the largest cell having 6 models (2%) that did not converge. Therefore, it seems to me that this issue is probably not of much concern.
Even though the majority of the models did converge, several had problems estimating the covariance matrix of the random-effects. In particular, $\Phi$ was non-positive definite (NPD) in several cases. The number of models where $\Phi$ was NPD are displayed in Table 9.

In each case, the matrix was NPD because the variance of the random effect for the binary variable $(q_{11})$ was fixed at zero after it began to go negative in the estimation routine. This problem appears to occur for the AGQ7, AGQ3 and LA estimation methods almost exclusively. The rates of NPD covariance matrices get worse as the correlation between the random-effects increases, but improve as the variance of the random effect increases.

### Table 8. Number of models that did not converge. Cells that were equal to 0 were omitted

<table>
<thead>
<tr>
<th>Variance Condition</th>
<th>Correlation Condition</th>
<th>AGQ7</th>
<th>AGQ3</th>
<th>LA</th>
</tr>
</thead>
<tbody>
<tr>
<td>.25</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.25</td>
<td>.30</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.25</td>
<td>.70</td>
<td>2</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>.50</td>
<td>.70</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

78
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>.25</td>
<td>0</td>
<td>47</td>
<td>46</td>
<td>52</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.25</td>
<td>.30</td>
<td>185</td>
<td>192</td>
<td>200</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.25</td>
<td>.70</td>
<td>229</td>
<td>235</td>
<td>234</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.50</td>
<td>0</td>
<td>12</td>
<td>16</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.50</td>
<td>.30</td>
<td>86</td>
<td>98</td>
<td>106</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.50</td>
<td>.70</td>
<td>62</td>
<td>69</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.30</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.70</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9. Number of models with non-positive definite covariance matrix for the random-effects. Cells that are equal to 0 are suppressed.

The issue with NPD covariance matrices has been identified by other researchers. Vonesh (2012) found that approximately 25% of binary GLMMs resulted in NPD covariance matrices when the variance of the random-effects was small. This condition corresponded to a variance of .25 for the random effect intercept in his simulation. These rates dropped off considerably to less than 1% when the variance of the random effect was large. This condition corresponded to a variance of 1.0 in his simulation. He found these results to be true for several estimation methods including AGQ, Laplace, MQL and PQL.

In the present study, if attention is restricted to the conditions with $p^{12} = 0$, the rates found here are slightly lower than those found by Vonesh. For example, in the .25 variance / 0 correlation condition, approximately 17% of replications resulted in NPD
covariance matrices when using AGQ7, AGQ3 or Laplace. The most notable difference between these results and those presented by Vonesh, however, is the fact that the NPD problem was not present for the linearization methods in the current study.

A small simulation was conducted in order to investigate this discrepancy further. In this simulation I replicated the univariate binary GLMM that Vonesh used, keeping all of the conditions exactly the same. I used the same estimation methods as Vonesh, which were AGQ, Laplace, MQL and PQL. Vonesh also used SAS PROC GLIMMIX, so the implementation of the estimation methods should also be the same. My results from this simulation produced 51 out of 200 replications (25.5\%) that had NPD covariance matrices in the .25 variance condition. This rate was the same for all four estimation methods. Vonesh reported that 41 out of 200 replications were NPD in this condition. In my simulation I also found that 2 out of 200 replications (1\%) resulted in NPD covariance matrices in the 1.0 variance condition. Again, this was true for all four estimation methods. The NPD rate in this condition matches Vonesh exactly. I was satisfied with the degree to which these results matched those of Vonesh, indicating that I had replicated his study both in terms of simulating the data and estimating the models.

Based on these results, it seems that the discrepancy in the NPD rates between the original simulation and the replication is most likely attributable to one of two things. First, in my original study I did not use the default convergence criteria in SAS PROC GLIMMIX. Instead, I specified more conservative criteria which required the models to complete more iterations before determining convergence. The second difference between my original simulation and that of Vonesh is that my model is multivariate.
To investigate this further, I re-ran the Vonesh data using the more conservative convergence criteria. The rates of NPD covariance matrices based on this specification were exactly the same as those using the default convergence criteria. Therefore, it seems to me that the multivariate structure must have an impact on the rates of NPD covariance matrices. Specifically, it seems that having a continuous dependent variable in addition to the binary variable can substantially lower the NPD rates for the linearization methods.

It is interesting to note that the NPD rates get worse as the correlation between the random-effects increases in the low variance condition ($\phi_{11} = .25$). The rates change from approximately 15% when $\rho^{12} = 0$ to 65% when $\rho^{12} = .30$ to 75% when $\rho^{12} = .7$. In the moderate variance condition ($\phi_{11}^{1} = .5$), the NPD rates are again lowest when $\rho^{12} = 0$ (around 5%) but are largest when $\rho^{12} = .30$ (around 32%). In the large variance condition ($\phi_{11}^{1} = 1.0$), the NPD rates drop off considerably and are less than 1% in all conditions.

Overall, it is certainly disconcerting that the rates of NPD $\Phi$ matrices are so high. When one considers the types of multivariate GLMMs that applied researchers are interested in fitting, the model that this simulation is designed around is fairly simple. In addition to its basic design, the data should also be well-behaved compared to real data examples. By this I mean that the simulated data does not have any missing values, individuals are all measured at the same time points, and the true model is actually known. All of these factors should indicate that, in general, the performance of the estimation methods with the simulated data will likely be better than their performance with a real data example. The fact that so many do not even converge to reasonable estimates is a concern.
Since the model becomes unstable when $\Phi$ is NPD, it is difficult to know how to compare the estimation methods in an unbiased way. It seems that there are three options. First, all of the runs can be compared regardless of the NPD status. Second, only those runs that were not problematic can be compared, meaning that in some conditions comparisons between estimation methods will be based on different numbers of replications. Finally, only those replications that were not problematic for any estimation method can be compared. In my opinion, the first approach seems problematic since it does not discriminate between results that are known to be flawed and those which might be reasonable. Since researchers do not tend to present results of models that have NPD covariance matrices, it also seems best to exclude those models from this comparison. For the most part the results will be presented using the second approach. In the cases where the second and third approaches produce discrepant conclusions, the differences will be noted.

5.2.2 Evaluation of Fixed Effect Estimates

As would be expected based on results from previous studies and knowledge of estimation for GLMMs, in general the parameters for the binary variable are estimated much less accurately than the parameters for the continuous variable. The means and standard deviations of the parameter estimates for each estimation method in each condition can be found in Table 12 in Appendix C. The mean bias and average MSE can be found in Table 13, also in Appendix C.
The average parameter estimate for the fixed effect intercept for the continuous variable was within .01 of the true value ($\beta_0^2 = 15.00$) for every estimation method in all conditions. The standard deviation of the estimates was small and did not fluctuate systematically over conditions, hovering right around .05 in all cases. Similarly, the average MSE was near zero in all conditions. There was more variability in the parameter estimates for the fixed effect slope for the continuous variable ($\beta_1^2$), with an average standard deviation of around .18. The largest average MSE for any of the conditions was .03, which is still quite small. Even though the parameter estimates for $\beta_1^2$ were more variable than those for $\beta_0^2$, the average parameter estimate was still within .03 of the true value ($\beta_1^2 = 3.00$) in every condition.

The fixed effect parameter estimates for the binary variables vary much more between conditions and estimation methods. In general, the integral approximation methods (AGQ7, AGQ3 and LA) tend to be more accurate than the linearization methods (M-PQL, M-MQL, R-PQL, R-MQL).

The boxplots in Figure 4 summarize the parameter estimates for the fixed-effect intercept of the binary variable over the three variance and three correlation conditions for each estimation method. From this, one can see that the accuracy and variability of $\beta_0^1$ depends not only on the estimation method, but also on the correlation condition and variance condition. In the low variance condition ($\varphi_{11}^1 = .25$) the estimation methods tend to perform more similarly, although the integral approximation methods still tend to be more accurate than the linearization methods. This finding was also verified by Vonesh

83
(2012). As the variance of the random effect for the slope of the binary variable increases, the discrepancy between the estimation methods becomes larger. It is interesting to point out that the intercept does not even have a random component in this model. Nevertheless, the estimation of this parameter is impacted in a non-trivial way by the increased variability.

![Boxplots for parameter estimates of the intercept of the binary variable ($\beta^1_0$) over the three variance and three correlation conditions for each estimation method. The true value of the parameter is denoted by the gray horizontal line.](image)

Figure 4. Boxplots for parameter estimates of the intercept of the binary variable ($\beta^1_0$) over the three variance and three correlation conditions for each estimation method. The true value of the parameter is denoted by the gray horizontal line.
The impact of the correlation condition on the estimation of $\beta_0^{\dagger}$ seems to depend on the variance condition and the estimation method. The exception is the M-MQL and R-MQL methods which are relatively unaffected by the correlation. In the low variance condition ($\phi_{11}^{\dagger} = .25$), the parameter estimates have a tendency to decrease as the correlation increases across all of the estimation methods. This brings the M-PQL and R-PQL methods closer to the true value of the parameter, but moves the integral approximation methods further away. In the moderate variance condition ($\phi_{11}^{\dagger} = .50$), the estimation methods seem to be less affected by changes in the correlation. However, there still appears to be a small shift toward the true parameter value for M-PQL and R-PQL methods as the correlation increases. The correlation does not seem to impact the parameter estimates in the high variance ($\phi_{11}^{\dagger} = 1.0$) condition.

The same general findings also hold true for the parameter estimates for the slope of the binary variable ($\hat{\beta}_1^{\dagger}$). Boxplots for this parameter are displayed in Figure 5. The one difference is that the linearization methods tend to underestimate the slope, whereas they had a tendency to overestimate the intercept. This could be attributable to the fact that the intercept was negative and the slope is positive. Therefore, in both cases the linearization methods tend to be bias the fixed effect estimates toward zero.
Figure 5. Boxplots for parameter estimates of the slope of the binary variable ($\beta_1$) over the three variance and three correlation conditions for each estimation method. The true value of the parameter is denoted by the gray horizontal line.

Another way of looking at the accuracy of the fixed effect parameter estimates is as the percentage of confidence intervals that contain the true model parameter. These
percentages are displayed in Table 14 in Appendix C for 95% confidence intervals. The coverage rates for $\beta_0^l$ and $\beta_1^l$ are also displayed in Figure 6 and Figure 7, respectively. The coverage is around the nominal rate for AGQ7, AGQ3 and Laplace for the fixed-effect parameters of both the continuous and binary variables. The variance and covariance conditions do not seem to have an impact on this result. The coverage is also around the nominal rate for the other estimation methods for the fixed-effect parameters for the continuous variable. The coverage for the fixed-effect parameters for the binary variable, on the other hand, varies quite considerably over the conditions for the linearization methods.

In particular, the coverage rates drop considerably as the variance of the random effect for the slope of the binary variable increases. For example, focusing only on M-PQL and the condition where $\rho^{12} = 0$, 87% of the confidence intervals for $\beta_0^l$ contain the true model parameter when $\phi_{11}^l = .25$, 64% when $\phi_{11}^l = .50$, and 31% when $\phi_{11}^l = 1.0$. These rates are slightly higher for $\beta_1^l$, but the general pattern is the same. In some cases, the coverage rates are better for R-PQL, but not substantially so. The coverage rates have the same pattern for M-MQL and R-MQL, but are much lower. For example, again focusing only on the condition where $\rho^{12} = 0$, the coverage rates for $\beta_0^l$ using M-MQL are 85% when $\phi_{11}^l = .25$, 48% when $\phi_{11}^l = .50$, and 6% when $\phi_{11}^l = 1.0$. 

87
Figure 6. Percentage of confidence intervals (α=.05) that contain the true value of $\beta_0$. The dashed line denotes 95%, the nominal coverage rate for the given significance level.
Figure 7. Percentage of confidence intervals ($\alpha=.05$) that contain the true value of $\beta_1$. The dashed line denotes 95%, the nominal coverage rate for the given significance level.
Within each variance condition, higher correlations tend to lead to better coverage rates for both M-PQL and R-PQL. For example, when $\phi_{11} = 1.0$ the coverage rates for $\beta_1$ using M-PQL are 53% when $\rho^{12} = 0$, 59% when $\rho^{12} = .30$, and 76% when $\rho^{12} = .70$. The correlation between the random-effects does not appear to have an impact when using either M-MQL or R-MQL.

5.2.3 Evaluation of Variance and Covariance Estimates

As was expected, in general there is much more bias and variability in the parameter estimates for the variance and covariance terms than there was for the fixed-effects. The quality of the parameter estimates vary greatly depending on the particular parameter being estimated, the estimation method that is used, and particular variance or correlation condition being evaluated. The results for each of the variance and covariance parameters will be discussed in this section.

The first parameter to be evaluated is the variance of the random effect for the slope of the binary variable ($\phi_{11}$). This is a parameter that varies by condition in this simulation. When $\phi_{11}$ is small, AGQ7, AGQ3 and Laplace all tend to be more biased when the correlation between the random-effects is large compared to when the correlation between the random effects is small. The impact of the correlation is actually the opposite for M-PQL and R-PQL. Specifically, when $\rho^{12}$ is large M-PQL and R-PQL are less biased than when $\rho^{12}$ is small. In addition to the information presented in Table 12 and Table 13 in Appendix C, this result can also be seen in Figure 8. This particular
result is even more prominent when the runs that did not converge or resulted in NPD solutions for the other estimation methods are excluded from the calculations. The average parameter estimate and mean bias for these runs are presented in Table 15 in Appendix C. For example, when using M-PQL and $\phi_{11}^1 = .25$, the average bias for $\hat{\phi}_{11}^1$ is -.12 when $\rho_{12} = 0$, -.11 when $\rho_{12} = .30$, and .03 when $\rho_{12} = .70$. The standard deviations and average MSE are not reported for this set of replications because the values did not differ substantially from the complete set of replications for any of the estimation methods or conditions. The size of $\rho_{12}$ does not seem to impact M-MQL or R-MQL.

Finally, as the variance of the random effect increases, the linearization methods tend to become more biased toward zero.

There do not appear to be any trends in the variability of the estimates for $\phi_{11}^1$ when using AGQ7, AGQ3 or Laplace. The variability of the parameter estimates does increase drastically for the linearization methods as the variance of the random effect increases. However, the size of $\rho_{12}$ does not seem to have an impact on the variability of the parameter estimates for $\hat{\phi}_{11}^1$. 

91
Figure 8. Boxplots for parameter estimates of the variance of the random effect for the slope of the binary variable ($\phi_{11}$) over the three variance and three correlation conditions for each estimation method. The true value of the parameter is denoted by the gray horizontal line.
Looking again at the percentage of 95% confidence intervals that cover the true value of the model parameter, which are displayed in Table 14 in Appendix C as well as in Figure 9, one can see that when $\phi_{11}^{1} = .5$ or $\phi_{11}^{1} = 1.0$ AGQ7, AGQ3 and Laplace all have coverage rates near the nominal level. However, when $\phi_{11}^{1} = .25$ the coverage rates decrease as $\rho^{12}$ increases. For M-PQL, R-PQL, M-MQL, and R-MQL the coverage rates all decrease as $\phi_{11}^{1}$ increases. Regardless of the variance condition, the size of $\rho^{12}$ seems to have an impact when using M-PQL or R-PQL. Specifically, when $\rho^{12} = .70$ the coverage rates are much higher. In this sense, R-PQL does tend to perform a little better than M-PQL but not by much.

The next parameter to evaluate is the variance of the random effect for the slope of the continuous variable ($\phi_{11}^{2}$). All of the estimation methods are able to estimate this parameter fairly accurately. There is one condition ($\phi_{11}^{1} = .25$, $\rho^{12} = .30$) where it seems that AGQ7, AGQ3 and Laplace perform poorly compared to the other estimation methods. However, when the runs that did not converge or resulted in NPD matrices for any of the estimation methods are excluded, the performance of the linearization methods is more similar to that of AGQ7, AGQ3 and Laplace. This can be seen in Table 15 in Appendix C.
Figure 9. Percentage of confidence intervals ($\alpha = .05$) that contain the true value of $\varphi_{11}$. The dashed line denotes 95%, the nominal coverage rate for the given significance level.
The residual variance ($\sigma^2_e$) is also estimated very well by all of the estimation methods. The bias is near zero and the variability is small in all of the conditions for all of the estimation methods. These results are shown in the last panel of Table 12.

The final parameter to investigate is the covariance between the random-effects ($\varphi_{11}^{12}$). In the multivariate setting, this parameter is often of particular interest. Boxplots for the estimates of this parameter by estimation method and condition are presented in Figure 10. When $\rho^{12} = 0$, there is very little bias for any of the estimation methods. However, as $\rho^{12}$ increases so does the bias. This is true for all of the estimation methods when $\varphi_{11}^{1} = .25$ or $\varphi_{11}^{1} = .50$, but not true of AGQ7, AGQ3 or Laplace when $\varphi_{11}^{1} = 1.0$. When $\rho^{12} = .30$ or $\rho^{12} = .70$, it is also true that the performance of AGQ7, AGQ3, and Laplace improves as $\varphi_{11}^{1}$ increases, but the performance of M-PQL, M-MQL, R-PQL, and R-MQL deteriorates as $\varphi_{11}^{1}$ increases. Based on the results in Table 13, it appears that M-PQL and R-PQL show the least amount of bias on this parameter when $\varphi_{11}^{1} = .25$. However, when the runs that did not converge or resulted in a NPD covariance matrix for the other estimation methods are excluded, this result is no longer present. This information can be seen in Table 15. The variability of the parameter estimates increases slightly as $\varphi_{11}^{1}$ increases, but the variability does not seem to be impacted by changes in $\rho^{12}$. 
Figure 10. Boxplots for parameter estimates of the covariance of the random-effects ($\phi_{11}$) over the three variance and three correlation conditions for each estimation method. The true value of the parameter is denoted by the gray horizontal line.
When looking at the confidence intervals for $\varphi_{11}$, which are displayed in Table 14 in Appendix C as well as in Figure 11, there does not appear to be any systematic change in coverage rates for AGQ7, AGQ3 or Laplace for any of the variance or covariance conditions. For M-PQL, M-MQL, R-PQL and R-MQL, however, the coverage rates decrease both as $\varphi_{11}$ increases and as $\rho_{12}$ increases. This means that for all four of the linearization methods the lowest coverage rates occur in the condition where $\varphi_{11} = 1.0$ and $\rho_{12} = .70$. For M-PQL this rate is 82%. It is slightly higher for R-PQL at 84%. The coverage rates for both M-MQL and R-MQL, however, are disturbingly low at 11% and 13%, respectively.

Another important consideration for the covariance parameter is how often the estimates from the model would indicate that there is not a relationship between the random-effects. This can be determined by looking at the percentage of confidence intervals that contain 0. This information is presented in Table 10. When $\rho_{12} = 0$, one hopes that a large percentage of the confidence intervals will contain zero since that is the true value of the model parameter. All of the methods perform around the nominal rate when $\rho_{12} = 0$, regardless of the variance condition. Conversely, when $\rho_{12} = .7$, all of the methods also perform well in the sense that a very small percentage of the confidence intervals contain 0. When $\rho_{12} = .3$, however, the percentage of confidence intervals that contain 0 is much larger than one would hope. This is displayed in Figure 12 in addition to Table 10. In practical terms, this means that the parameter estimates indicate that there is not a relationship between the random-effects when in fact there actually is a relationship. The percentage of confidence intervals that contain 0 when $\rho_{12} = .3$
decreases as $\varphi_{11}$ increases. Also, when $\varphi_{11}$ is small the linearization methods tend to perform better in this respect, meaning that fewer confidence intervals contain 0.

Figure 11. Percentage of confidence intervals ($\alpha=.05$) that contain the true value of $\varphi_{11}$. The dashed line denotes 95%, the nominal coverage rate for the given significance level.
<table>
<thead>
<tr>
<th>Condition</th>
<th>$\varphi_{11}^1$</th>
<th>$\rho^{12}$</th>
<th>$\varphi_{11}^{12}$</th>
<th>AGQ7</th>
<th>AGQ3</th>
<th>LA</th>
<th>M-PQL</th>
<th>M-MQL</th>
<th>R-PQL</th>
<th>R-MQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>98.81%</td>
<td>98.81%</td>
<td>98.79%</td>
<td>96.99%</td>
<td>96.32%</td>
<td>97.67%</td>
<td>96.32%</td>
</tr>
<tr>
<td>0.25</td>
<td>0.3</td>
<td>.26</td>
<td>0</td>
<td>83.33%</td>
<td>85.05%</td>
<td>86.73%</td>
<td>58.86%</td>
<td>56.52%</td>
<td>58.86%</td>
<td>57.19%</td>
</tr>
<tr>
<td>0.25</td>
<td>0.7</td>
<td>.61</td>
<td>0</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.33%</td>
<td>0.33%</td>
<td>0.33%</td>
<td>0.33%</td>
</tr>
<tr>
<td>0.50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>97.22%</td>
<td>97.89%</td>
<td>97.84%</td>
<td>95.67%</td>
<td>95.33%</td>
<td>95.67%</td>
<td>95.33%</td>
</tr>
<tr>
<td>0.50</td>
<td>0.3</td>
<td>.37</td>
<td>0</td>
<td>50.47%</td>
<td>52.24%</td>
<td>51.03%</td>
<td>44.33%</td>
<td>42.00%</td>
<td>45.00%</td>
<td>43.00%</td>
</tr>
<tr>
<td>0.50</td>
<td>0.7</td>
<td>.86</td>
<td>0</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>1.00</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>97.00%</td>
<td>97.32%</td>
<td>97.32%</td>
<td>98.00%</td>
<td>97.33%</td>
<td>98.00%</td>
<td>97.33%</td>
</tr>
<tr>
<td>1.00</td>
<td>0.3</td>
<td>.52</td>
<td>0</td>
<td>35.57%</td>
<td>35.35%</td>
<td>35.91%</td>
<td>35.67%</td>
<td>35.00%</td>
<td>36.33%</td>
<td>35.00%</td>
</tr>
<tr>
<td>1.00</td>
<td>0.7</td>
<td>1.2</td>
<td>0</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Table 10. Percentage of confidence intervals ($\alpha=.05$) for $\varphi_{11}^{12}$ that contain 0

Figure 12. Percentage of confidence intervals ($\alpha=.05$) for $\varphi_{11}^{12}$ that contain 0 when $\rho^{12} = .30$
Longitudinal studies are prevalent in a number of disciplines. As part of these studies, researchers often collect data on a large number of variables at each measurement occasion. Although many research questions can be answered using separate models for each outcome variable, there are many benefits to using models that can handle several outcomes simultaneously. Among other things, multivariate models can be more efficient, they can be more flexible when it comes to handling missing data, and they can be used to answer a wider range of research questions.

In particular, some important research questions can only be answered through a multivariate longitudinal model. For example, one might be interested in how the relationship between two variables changes over time. Alternatively, one could also be interested in how the trajectories for two variables are related. A flexible model that is able to accommodate many different types of data is the multivariate generalized linear mixed model.

Even though this framework is widely applicable in the analysis of multivariate longitudinal data, one challenge of the multivariate GLMM is estimation. Many of the same methods that are used for univariate GLMMs can be used in the multivariate model, but it is important to know how these estimation methods perform in the multivariate
context. In particular, it is important to know how well the parameters can be recovered and how often one will reach the correct conclusions about the relationships between the variables when various estimation methods are used.

In the simulation study that was conducted as part of this thesis, the performance of several estimation methods was compared under a number of different conditions. The estimation methods that were used can be grouped into two categories. These are numerical approximation methods (AGQ7, AGQ3, and Laplace) and linearization methods (M-MQL, R-MQL, M-PQL, and R-PQL). The model that was evaluated was a multivariate GLMM for a binary and continuous variable measured over time. The two conditions that were varied in the simulation were the variance of the random effect for the slope of the binary variable and the covariance between the random-effects for the slopes of the binary and continuous variables.

From the start, one major difference between the estimation methods was the frequency with which models resulted in covariance matrices for the random-effects that were not positive definite. This problem was restricted almost exclusively to the AGQ7, AGQ3 and Laplace estimation methods, and it seemed to be the worst when the variance of the random-effects was small but the covariance between them was large. The NPD solutions pose a problem for applied researchers since they lead to obviously flawed estimates even when the model being fit is appropriate for the data at hand.

Results from previous simulation studies on the performance of various estimation methods within the univariate GLMM framework have shown that variance parameters are more difficult to estimate than fixed-effects (Pinheiro & Chao, 2006; Dey
Lim, 2013), the parameters for binary variables are more difficult to estimate than the parameters for other types of variables (Rodriguez & Goldman, 2001), and that adaptive Gaussian quadrature outperforms most other estimation methods (Vonesh, 2012). With regard to the specific estimation methods that were considered here, previous research has also shown that 5 quadrature points is often sufficient for GLMM problems (Pinheiro & Chao, 2006), and that the benefit of using residual pseudo likelihood over maximum likelihood is often small (Breslow & Lin, 1995; Lin & Breslow, 1996; Pinheiro & Chao, 2006). For the most part, the results of the current simulation study agree with all of these findings.

Although AGQ7, AGQ3 and Laplace consistently perform better than the other estimation methods, one of the most notable findings of the current simulation is that both M-PQL and R-PQL benefit from the multivariate structure. In particular, the accuracy of the parameter estimates for both of these methods was best when the correlation between the random-effects was large. This is an encouraging finding, especially given the problems that were present for the numerical approximation methods with regard to NPD solutions.

Of course, an improvement in performance is always a relative measure, though. Although M-PQL and R-PQL performed better when the correlation was large, they still performed quite poorly overall when the variance of the random-effects was large. For example, the percentage of confidence intervals that contain the true value of the parameter was quite low in the largest variance condition for these estimation methods. It seems to be the case that M-PQL and R-PQL produce more stable estimates for the
covariance matrix of random effects than do the numerical approximation methods, but they do so with greater bias.

The other linearization methods (M-MQL and R-MQL) also perform quite poorly when the variance of the random-effects is large. Moreover, they do not seem to benefit from the correlation between the random-effects. Namely, there is no systematic change in the accuracy of the parameter estimates over the various correlation conditions.

When the variance of the random-effects is small, AGQ7, AGQ3 and Laplace actually perform more poorly as the correlation between the random-effects increases. This trend does not seem to be present when the variance of the random-effects is larger.

One question that was of particular interest in this study was how well the parameter for the covariance between the random-effects could be recovered. This parameter is noteworthy because it is used to determine the relationship between the trajectories of the variables, which is often the primary goal of a multivariate analysis. When there is not a relationship between the random-effects (i.e., $\rho_{12} = 0$) all of the methods estimate this parameter well. At the opposite extreme, when the relationship between the random-effects is strong (i.e., $\rho_{12} = .7$), the integral approximation methods recover the parameter well but the linearization methods tend to underestimate the relationship. This problem is especially bad for M-MQL and R-MQL, where in the high variance condition less than 15% of the confidence intervals contain the true model parameter.

The condition that is perhaps the most interesting, however, is when there is a moderate relationship between the random-effects. Within this condition, a very large
percentage of the confidence intervals for this parameter contain 0, indicating that there is not a relationship between the random-effects. This is true for all of the estimation methods. It is concerning that, based on these parameter estimates and significance tests, one would so frequently make the incorrect conclusion that there is not relationship between the random-effects. Although a correlation of .3 is not strong, it is in line with many of the correlations that are seen in the social and behavioral sciences.

As with any simulation study, it is important to remember that these results depend heavily on the structure of the model being fit and the values of the parameter estimates that were chosen. As such, there are several conditions that would be interesting to evaluate in future studies.

First, the current study used 7 time points to allow for more direct comparison with previous simulations that had been conducted (Breslow & Clayton, 1993; Vonesh, 2012). Although 7 repeated measures is not large in every discipline, in my opinion it is larger than what is often seen in the social and behavioral sciences. It would be interesting to see how the estimation methods perform when there are fewer time points, since this is perhaps a more realistic scenario and since in the univariate GLMM context other researchers have found that fewer time points leads to less accurate estimates (Breslow & Clayton, 1993).

Second, one of the difficulties of multivariate GLMMs is that they often involve a large number of random-effects. Although it wasn’t much of a factor in the current simulation, one of the major limitations of the Laplace approximation and the quadrature-based methods is that they become very computationally demanding as the number of...
random-effects increases. Comparing the differences in performance when there are several random-effects would be a useful extension. In particular, it would be helpful to know how additional random-effects impact the convergence rates in addition to the accuracy of the parameter estimates.

Third, the present study looked at a multivariate model for a continuous and binary variable. It would be interesting to see how performance changes when other variable types are paired together, or when 3 or 4 outcome variables are modeled simultaneously. This is related to the previous point in the sense that additional outcome variables would also result in additional random-effects.

Finally, the present study used complete data where individuals were all measured at the same number of time points. Since missing data is so prevalent in longitudinal studies, and the ability to handle data that is MAR is one of the advantages of the multivariate GLMM framework, it would be worthwhile to compare the estimation methods under various missing data scenarios.

In the end, it is clear that multivariate GLMMs are a useful approach in the analysis of multivariate longitudinal data. With regard to estimation, there is not a universal recommendation that can be made about which method to use. In general, the integral approximation methods estimate the model parameters more accurately but they also have issues with convergence and NPD solutions. There is some indication that both R-PQL and M-PQL perform better in the multivariate context than they do in the univariate context, but under many conditions they still perform quite poorly. The difficult part about real data is that you never know what condition you are in. Therefore,
like others have recommended (Joe, 2008; Rodriguez & Goldman, 2001), it seems best to
fit the models using a few different estimation methods keeping in mind the relative
strengths and weaknesses of each.
References


StataCorp. (2013). Stata Statistical Software (Release 13) [Computer software]. College Station, TX: StataCorp LP.


### Appendix A: NLSY Example Results

<table>
<thead>
<tr>
<th>Parameter Estimate</th>
<th>AGQ7</th>
<th>AGQ3</th>
<th>Laplace</th>
<th>R-PQL</th>
<th>M-PQL</th>
<th>R-MQL</th>
<th>M-MQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0^1$</td>
<td>18.86</td>
<td>18.86</td>
<td>18.86</td>
<td>18.86</td>
<td>18.86</td>
<td>18.86</td>
<td>18.86</td>
</tr>
<tr>
<td>$\beta_1^1$</td>
<td>.88</td>
<td>.88</td>
<td>.88</td>
<td>.88</td>
<td>.88</td>
<td>.88</td>
<td>.88</td>
</tr>
<tr>
<td>$\beta_0^2$</td>
<td>.54</td>
<td>.53</td>
<td>.49</td>
<td>.40</td>
<td>.40</td>
<td>.38</td>
<td>.38</td>
</tr>
<tr>
<td>$\beta_1^2$</td>
<td>-.26</td>
<td>-.25</td>
<td>-.23</td>
<td>-.19</td>
<td>-.19</td>
<td>-.18</td>
<td>-.18</td>
</tr>
<tr>
<td>$\phi_{00}^1$</td>
<td>17.05</td>
<td>17.05</td>
<td>17.05</td>
<td>17.05</td>
<td>17.05</td>
<td>17.05</td>
<td>17.05</td>
</tr>
<tr>
<td>$\phi_{11}^1$</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>$\phi_{00}^2$</td>
<td>2.19</td>
<td>1.96</td>
<td>1.15</td>
<td>.70</td>
<td>.69</td>
<td>.58</td>
<td>.58</td>
</tr>
<tr>
<td>$\phi_{12}^2$</td>
<td>-.37</td>
<td>-.36</td>
<td>-.34</td>
<td>-.28</td>
<td>-.28</td>
<td>-.27</td>
<td>-.27</td>
</tr>
<tr>
<td>$\sigma_e^2$</td>
<td>4.34</td>
<td>4.34</td>
<td>4.34</td>
<td>4.34</td>
<td>4.34</td>
<td>4.34</td>
<td>4.34</td>
</tr>
</tbody>
</table>

Table 11. Parameter estimates and run times (min:sec) for multivariate GLMM for television rules and BMI data
Appendix B: Sample Syntax

In order to run a multivariate GLMM in SAS PROC GLIMMIX, the data must be formatted properly. This involves stacking the repeated observations for each individual and each outcome variable in a single column. The researcher must also define variables that represent the columns of the design matrices for the fixed and random-effects and a variable that contains the distribution that should be used to model each outcome. As an example, a portion of the data that was presented in Table 3 is given below. Here, it is formatted appropriately for the multivariate GLMM analysis.

<table>
<thead>
<tr>
<th>id</th>
<th>age</th>
<th>value</th>
<th>var_name</th>
<th>dist</th>
<th>int1</th>
<th>int2</th>
<th>slope1</th>
<th>slope2</th>
</tr>
</thead>
<tbody>
<tr>
<td>201</td>
<td>11</td>
<td>23.0223</td>
<td>bmi</td>
<td>normal</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>201</td>
<td>11</td>
<td>1</td>
<td>tv</td>
<td>binary</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>202</td>
<td>10</td>
<td>28.689</td>
<td>bmi</td>
<td>normal</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>302</td>
<td>11</td>
<td>17.1631</td>
<td>bmi</td>
<td>normal</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>302</td>
<td>13</td>
<td>16.4196</td>
<td>bmi</td>
<td>normal</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>302</td>
<td>11</td>
<td>1</td>
<td>tv</td>
<td>binary</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>302</td>
<td>13</td>
<td>1</td>
<td>tv</td>
<td>binary</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>303</td>
<td>10</td>
<td>17.358</td>
<td>bmi</td>
<td>normal</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>303</td>
<td>12</td>
<td>17.7123</td>
<td>bmi</td>
<td>normal</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>303</td>
<td>14</td>
<td>18.8521</td>
<td>bmi</td>
<td>normal</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>303</td>
<td>14</td>
<td>0</td>
<td>tv</td>
<td>binary</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>401</td>
<td>10</td>
<td>18.5917</td>
<td>bmi</td>
<td>normal</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>401</td>
<td>10</td>
<td>0</td>
<td>tv</td>
<td>binary</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
The syntax for running a multivariate GLMM is given next. The first line allows
the researcher to specify the dataset and estimation method that should be used. The
second line defines characteristics of the estimation process, such as convergence criteria
and the maximum number of iterations and function calls that should be allowed. The
third line specifies that the observations are clustered by id. The fourth line specifies the
structure of the fixed-effects in the model. The variables that were defined in the data
(int1, int2, slope1 and slope2) are used here to reconstruct the design matrix for the fixed-
effects. The “noint” option tells PROC GLIMMIX not to include an intercept since the
intercept for each variable was included manually through the int1 and int2 variables. The
“dist=byobs(dist)” command specifies that the distribution of the variables should be
determined by observation, and the value is given in the variable “dist”. The commands
“cl”, “solution”, and “covb(details)” request that additional output be created. Similar to
the model line, the random statement sets up the model for the random-effects. In this
model the intercepts for the two variables are random and the covariance matrix is
unstructured. The command “covtest/ cl” requests confidence intervals for the
parameters.

```
proc glimmix data=dataset method=quad;
  nloptions maxiter=100 maxfunc=500 absgconv=1e-5 gconv=1e-12;
class id;
model y = int1 int2 slope1 slope2 / noint dist=byobs(dist) cl
  solution covb(details);
random int1 int2 / subject=id type=un gc;
covtest/cl;
run;
```
| Condition | Parameter | AGQ7 | | | LA | | | M-PQL | | | M-MQL | | | R-PQL | | | R-MQL |
|-----------|-----------|------|--------------------|--------|-------|----------|--------|--------|--------|-------|--------|--------|-------|--------|--------|--------|
|           | $\varphi_{11}$ | $\rho_{12}$ | True Value | Mean | Std | Mean | Std | Mean | Std | Mean | Std | Mean | Std | Mean | Std | Mean | Std |
|           | 0.25 | 0 | -2.5 | -2.55 | 0.21 | -2.54 | 0.21 | -2.51 | 0.20 | -2.38 | 0.18 | -2.35 | 0.19 | -2.39 | 0.18 | -2.35 | 0.19 |
|           | 0.25 | 0.3 | -2.5 | -2.54 | 0.22 | -2.53 | 0.22 | -2.50 | 0.21 | -2.37 | 0.20 | -2.33 | 0.20 | -2.38 | 0.20 | -2.33 | 0.20 |
|           | 0.25 | 0.7 | -2.5 | -2.64 | 0.22 | -2.62 | 0.24 | -2.61 | 0.22 | -2.42 | 0.19 | -2.34 | 0.19 | -2.42 | 0.19 | -2.34 | 0.19 |
|           | 0.50 | 0 | -2.5 | -2.53 | 0.19 | -2.51 | 0.18 | -2.50 | 0.19 | -2.25 | 0.15 | -2.16 | 0.16 | -2.25 | 0.15 | -2.16 | 0.16 |
|           | 0.50 | 0.3 | -2.5 | -2.53 | 0.19 | -2.52 | 0.19 | -2.53 | 0.18 | -2.26 | 0.16 | -2.16 | 0.18 | -2.26 | 0.16 | -2.16 | 0.18 |
|           | 0.50 | 0.7 | -2.5 | -2.54 | 0.20 | -2.54 | 0.20 | -2.53 | 0.20 | -2.33 | 0.18 | -2.16 | 0.18 | -2.34 | 0.18 | -2.16 | 0.18 |
|           | 1.00 | 0 | -2.5 | -2.56 | 0.22 | -2.54 | 0.21 | -2.57 | 0.23 | -2.14 | 0.15 | -1.89 | 0.16 | -2.14 | 0.15 | -1.89 | 0.16 |
|           | 1.00 | 0.3 | -2.5 | -2.52 | 0.19 | -2.50 | 0.19 | -2.52 | 0.20 | -2.13 | 0.14 | -1.88 | 0.17 | -2.13 | 0.14 | -1.88 | 0.17 |
|           | 1.00 | 0.7 | -2.5 | -2.52 | 0.20 | -2.50 | 0.19 | -2.51 | 0.20 | -2.21 | 0.14 | -1.87 | 0.15 | -2.22 | 0.14 | -1.87 | 0.15 |
|           | 0.25 | 15.0 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 |
|           | 0.25 | 15.0 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 |
|           | 0.25 | 15.0 | 14.99 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 |
|           | 0.50 | 15.0 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 |
|           | 0.50 | 15.0 | 14.99 | 0.06 | 15.0 | 0.06 | 14.99 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 |
|           | 0.50 | 15.0 | 15.01 | 0.06 | 15.01 | 0.06 | 15.01 | 0.06 | 15.0 | 0.06 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 |
|           | 1.00 | 15.0 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 |
|           | 1.00 | 15.0 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 | 15.0 | 0.06 |
|           | 1.00 | 15.0 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 | 15.0 | 0.05 |

Table 12. Mean and standard deviation of parameter estimates for simulated models
Table 12 continued

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Condition</th>
<th>AGQ7</th>
<th>AGQ3</th>
<th>LA</th>
<th>M-PQL</th>
<th>M-MQL</th>
<th>R-PQL</th>
<th>R-MQL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\phi_{i1}$</td>
<td>$\rho_{i2}$</td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>$\beta_1^i$</td>
<td>0.25</td>
<td>0</td>
<td>1.00</td>
<td>1.02</td>
<td>0.12</td>
<td>1.02</td>
<td>0.12</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.3</td>
<td>1.00</td>
<td>1.02</td>
<td>0.13</td>
<td>1.02</td>
<td>0.13</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.7</td>
<td>1.00</td>
<td>1.07</td>
<td>0.13</td>
<td>1.07</td>
<td>0.14</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0</td>
<td>1.00</td>
<td>1.01</td>
<td>0.13</td>
<td>1.01</td>
<td>0.13</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.3</td>
<td>1.00</td>
<td>1.01</td>
<td>0.11</td>
<td>1.01</td>
<td>0.11</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.7</td>
<td>1.00</td>
<td>1.03</td>
<td>0.13</td>
<td>1.02</td>
<td>0.13</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0</td>
<td>1.00</td>
<td>1.02</td>
<td>0.15</td>
<td>1.01</td>
<td>0.14</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.3</td>
<td>1.00</td>
<td>1.02</td>
<td>0.15</td>
<td>1.01</td>
<td>0.14</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.7</td>
<td>1.00</td>
<td>1.02</td>
<td>0.14</td>
<td>1.01</td>
<td>0.14</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0</td>
<td>3.00</td>
<td>2.99</td>
<td>0.18</td>
<td>2.99</td>
<td>0.18</td>
<td>2.99</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.3</td>
<td>3.00</td>
<td>3.03</td>
<td>0.18</td>
<td>3.03</td>
<td>0.20</td>
<td>3.03</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.7</td>
<td>3.00</td>
<td>3.01</td>
<td>0.17</td>
<td>2.99</td>
<td>0.17</td>
<td>3.00</td>
</tr>
<tr>
<td>$\beta_2^i$</td>
<td>0.50</td>
<td>0</td>
<td>3.00</td>
<td>3.01</td>
<td>0.18</td>
<td>3.01</td>
<td>0.19</td>
<td>3.01</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.3</td>
<td>3.00</td>
<td>3.02</td>
<td>0.18</td>
<td>3.00</td>
<td>0.18</td>
<td>3.01</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.7</td>
<td>3.00</td>
<td>3.02</td>
<td>0.18</td>
<td>3.00</td>
<td>0.18</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0</td>
<td>3.00</td>
<td>3.01</td>
<td>0.17</td>
<td>3.01</td>
<td>0.17</td>
<td>3.01</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.3</td>
<td>3.00</td>
<td>3.02</td>
<td>0.18</td>
<td>3.02</td>
<td>0.18</td>
<td>3.02</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.7</td>
<td>3.00</td>
<td>3.03</td>
<td>0.17</td>
<td>3.03</td>
<td>0.17</td>
<td>3.03</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Condition</th>
<th>AGQ7</th>
<th>AGQ3</th>
<th>LA</th>
<th>M-PQL</th>
<th>M-MQL</th>
<th>R-PQL</th>
<th>R-MQL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.27</td>
<td>0.15</td>
<td>0.27</td>
<td>0.14</td>
<td>0.23</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.3</td>
<td>0.25</td>
<td>0.30</td>
<td>0.16</td>
<td>0.28</td>
<td>0.15</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.7</td>
<td>0.25</td>
<td>0.40</td>
<td>0.14</td>
<td>0.40</td>
<td>0.13</td>
<td>0.41</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.50</td>
<td>0</td>
<td>0.50</td>
<td>0.53</td>
<td>0.20</td>
<td>0.51</td>
<td>0.19</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.3</td>
<td>0.50</td>
<td>0.56</td>
<td>0.18</td>
<td>0.55</td>
<td>0.17</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.7</td>
<td>0.50</td>
<td>0.57</td>
<td>0.17</td>
<td>0.56</td>
<td>0.16</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0</td>
<td>1.00</td>
<td>1.08</td>
<td>0.35</td>
<td>1.00</td>
<td>0.31</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.3</td>
<td>1.00</td>
<td>1.03</td>
<td>0.31</td>
<td>0.97</td>
<td>0.28</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.7</td>
<td>1.00</td>
<td>1.04</td>
<td>0.31</td>
<td>0.99</td>
<td>0.29</td>
<td>1.04</td>
</tr>
<tr>
<td>$\varphi_{12}$</td>
<td>0.25</td>
<td>0</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.12</td>
<td>0.00</td>
<td>0.12</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.3</td>
<td>0.26</td>
<td>0.17</td>
<td>0.15</td>
<td>0.16</td>
<td>0.14</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.7</td>
<td>0.61</td>
<td>0.76</td>
<td>0.17</td>
<td>0.77</td>
<td>0.16</td>
<td>0.79</td>
</tr>
<tr>
<td>$\varphi_{12}$</td>
<td>0.50</td>
<td>0</td>
<td>0.00</td>
<td>0.01</td>
<td>0.17</td>
<td>0.01</td>
<td>0.17</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.3</td>
<td>0.37</td>
<td>0.35</td>
<td>0.20</td>
<td>0.34</td>
<td>0.19</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.7</td>
<td>0.86</td>
<td>0.92</td>
<td>0.19</td>
<td>0.92</td>
<td>0.18</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0</td>
<td>0.00</td>
<td>0.02</td>
<td>0.21</td>
<td>0.02</td>
<td>0.21</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.3</td>
<td>0.52</td>
<td>0.53</td>
<td>0.23</td>
<td>0.52</td>
<td>0.23</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.7</td>
<td>1.20</td>
<td>1.24</td>
<td>0.28</td>
<td>1.23</td>
<td>0.27</td>
<td>1.25</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Condition</th>
<th>AGQ7</th>
<th>AGQ3</th>
<th>LA</th>
<th>M-PQL</th>
<th>M-MQL</th>
<th>R-PQL</th>
<th>R-MQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{11}$</td>
<td>$\rho_{12}$</td>
<td>True Value</td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>3.00</td>
<td>3.01</td>
<td>0.42</td>
<td>3.00</td>
<td>0.42</td>
<td>3.00</td>
<td>0.45</td>
</tr>
<tr>
<td>0.25</td>
<td>0.3</td>
<td>3.00</td>
<td>2.87</td>
<td>0.45</td>
<td>2.84</td>
<td>0.43</td>
<td>2.83</td>
<td>0.45</td>
</tr>
<tr>
<td>0.25</td>
<td>0.7</td>
<td>3.00</td>
<td>2.99</td>
<td>0.34</td>
<td>3.00</td>
<td>0.37</td>
<td>3.01</td>
<td>0.38</td>
</tr>
<tr>
<td>$\varphi_{11}^2$</td>
<td>0.50</td>
<td>0.3</td>
<td>3.00</td>
<td>2.94</td>
<td>0.46</td>
<td>2.94</td>
<td>0.46</td>
<td>2.95</td>
</tr>
<tr>
<td>0.50</td>
<td>0.3</td>
<td>3.00</td>
<td>2.98</td>
<td>0.41</td>
<td>2.99</td>
<td>0.41</td>
<td>2.96</td>
<td>0.43</td>
</tr>
<tr>
<td>0.50</td>
<td>0.7</td>
<td>3.00</td>
<td>2.94</td>
<td>0.44</td>
<td>2.95</td>
<td>0.44</td>
<td>2.96</td>
<td>0.44</td>
</tr>
<tr>
<td>1.00</td>
<td>0</td>
<td>3.00</td>
<td>2.94</td>
<td>0.43</td>
<td>2.94</td>
<td>0.43</td>
<td>2.94</td>
<td>0.43</td>
</tr>
<tr>
<td>1.00</td>
<td>0.3</td>
<td>3.00</td>
<td>2.99</td>
<td>0.45</td>
<td>2.99</td>
<td>0.45</td>
<td>2.99</td>
<td>0.45</td>
</tr>
<tr>
<td>1.00</td>
<td>0.7</td>
<td>3.00</td>
<td>3.04</td>
<td>0.45</td>
<td>3.04</td>
<td>0.45</td>
<td>3.04</td>
<td>0.45</td>
</tr>
<tr>
<td>$\sigma_e^2$</td>
<td>0.25</td>
<td>0</td>
<td>2.00</td>
<td>1.99</td>
<td>0.11</td>
<td>1.99</td>
<td>0.11</td>
<td>1.99</td>
</tr>
<tr>
<td>0.25</td>
<td>0.3</td>
<td>2.00</td>
<td>1.99</td>
<td>0.12</td>
<td>1.98</td>
<td>0.12</td>
<td>1.99</td>
<td>0.12</td>
</tr>
<tr>
<td>0.25</td>
<td>0.7</td>
<td>2.00</td>
<td>2.01</td>
<td>0.13</td>
<td>2.02</td>
<td>0.13</td>
<td>2.01</td>
<td>0.14</td>
</tr>
<tr>
<td>0.50</td>
<td>0</td>
<td>2.00</td>
<td>2.01</td>
<td>0.13</td>
<td>2.01</td>
<td>0.12</td>
<td>2.01</td>
<td>0.12</td>
</tr>
<tr>
<td>0.50</td>
<td>0.3</td>
<td>2.00</td>
<td>1.99</td>
<td>0.12</td>
<td>2.00</td>
<td>0.12</td>
<td>2.00</td>
<td>0.13</td>
</tr>
<tr>
<td>0.50</td>
<td>0.7</td>
<td>2.00</td>
<td>2.00</td>
<td>0.12</td>
<td>1.99</td>
<td>0.12</td>
<td>2.00</td>
<td>0.12</td>
</tr>
<tr>
<td>1.00</td>
<td>0</td>
<td>2.00</td>
<td>2.01</td>
<td>0.12</td>
<td>2.01</td>
<td>0.12</td>
<td>2.01</td>
<td>0.12</td>
</tr>
<tr>
<td>1.00</td>
<td>0.3</td>
<td>2.00</td>
<td>1.99</td>
<td>0.12</td>
<td>1.99</td>
<td>0.12</td>
<td>1.99</td>
<td>0.12</td>
</tr>
<tr>
<td>1.00</td>
<td>0.7</td>
<td>2.00</td>
<td>2.00</td>
<td>0.12</td>
<td>2.00</td>
<td>0.12</td>
<td>2.00</td>
<td>0.12</td>
</tr>
<tr>
<td>Parameter</td>
<td>Condition</td>
<td>AGQ7</td>
<td>AGQ3</td>
<td>LA</td>
<td>M-PQL</td>
<td>M-MQL</td>
<td>R-PQL</td>
<td>R-MQL</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>$\rho_{12}^{12}$</td>
<td>True Value</td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
<td>MSE</td>
</tr>
<tr>
<td>0.25 0 -2.5</td>
<td>-0.05 0.05</td>
<td>-0.04 0.04</td>
<td>-0.01 0.04</td>
<td>0.12 0.05</td>
<td>0.15 0.06</td>
<td>0.11 0.05</td>
<td>0.15 0.06</td>
<td></td>
</tr>
<tr>
<td>0.25 0.3 -2.5</td>
<td>-0.04 0.05</td>
<td>-0.03 0.05</td>
<td>0 0.04</td>
<td>0.13 0.05</td>
<td>0.17 0.07</td>
<td>0.12 0.05</td>
<td>0.17 0.07</td>
<td></td>
</tr>
<tr>
<td>0.25 0.3 -2.5</td>
<td>-0.14 0.06</td>
<td>-0.12 0.07</td>
<td>-0.11 0.06</td>
<td>0.08 0.04</td>
<td>0.16 0.06</td>
<td>0.08 0.04</td>
<td>0.16 0.06</td>
<td></td>
</tr>
<tr>
<td>0.50 0 -2.5</td>
<td>-0.03 0.04</td>
<td>-0.01 0.03</td>
<td>0 0.04</td>
<td>0.25 0.09</td>
<td>0.34 0.14</td>
<td>0.25 0.08</td>
<td>0.34 0.14</td>
<td></td>
</tr>
<tr>
<td>0.50 0.3 -2.5</td>
<td>-0.03 0.04</td>
<td>-0.02 0.04</td>
<td>-0.03 0.03</td>
<td>0.24 0.08</td>
<td>0.34 0.14</td>
<td>0.24 0.08</td>
<td>0.34 0.14</td>
<td></td>
</tr>
<tr>
<td>0.50 0.7 -2.5</td>
<td>-0.04 0.04</td>
<td>-0.04 0.04</td>
<td>-0.03 0.04</td>
<td>0.17 0.06</td>
<td>0.34 0.15</td>
<td>0.16 0.06</td>
<td>0.34 0.15</td>
<td></td>
</tr>
<tr>
<td>1.00 0 -2.5</td>
<td>-0.06 0.05</td>
<td>-0.04 0.04</td>
<td>-0.07 0.06</td>
<td>0.36 0.15</td>
<td>0.61 0.4</td>
<td>0.36 0.15</td>
<td>0.61 0.4</td>
<td></td>
</tr>
<tr>
<td>1.00 0.3 -2.5</td>
<td>-0.02 0.04</td>
<td>0 0.04</td>
<td>-0.02 0.04</td>
<td>0.37 0.16</td>
<td>0.62 0.42</td>
<td>0.37 0.15</td>
<td>0.62 0.42</td>
<td></td>
</tr>
<tr>
<td>1.00 0.7 -2.5</td>
<td>-0.02 0.04</td>
<td>0 0.04</td>
<td>-0.01 0.04</td>
<td>0.29 0.1</td>
<td>0.63 0.42</td>
<td>0.28 0.1</td>
<td>0.63 0.42</td>
<td></td>
</tr>
</tbody>
</table>

Table 13. Average bias and mean square error (MSE) of parameter estimates for simulated models
Table 13 continued

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Condition</th>
<th>AGQ7</th>
<th>AGQ3</th>
<th>LA</th>
<th>M-PQL</th>
<th>M-MQL</th>
<th>R-PQL</th>
<th>R-MQL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Bias</td>
<td>Bias</td>
<td>Bias</td>
<td>Bias</td>
<td>Bias</td>
<td>Bias</td>
<td>Bias</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSE</td>
<td>MSE</td>
<td>MSE</td>
<td>MSE</td>
<td>MSE</td>
<td>MSE</td>
<td>MSE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1^1$</td>
<td>0.25 0 1.00</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>0.25 0.3</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>0.50 0 1.00</td>
<td>0.07</td>
<td>0.02</td>
<td>0.07</td>
<td>0.02</td>
<td>0.06</td>
<td>0.04</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>0.50 0.3</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>0</td>
<td>-0.13</td>
<td>-0.17</td>
</tr>
<tr>
<td></td>
<td>1.00 0 1.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>-0.19</td>
<td>-0.31</td>
</tr>
<tr>
<td></td>
<td>1.00 0.3</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>-0.14</td>
<td>-0.3</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td>1.00 0.7</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>-0.14</td>
<td>-0.3</td>
<td>-0.14</td>
</tr>
<tr>
<td>$\beta_1^2$</td>
<td>0.25 0 3.00</td>
<td>-0.01</td>
<td>0.03</td>
<td>-0.01</td>
<td>0.03</td>
<td>-0.01</td>
<td>0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>0.25 0.3</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>0.50 0 3.00</td>
<td>0.01</td>
<td>0.03</td>
<td>-0.01</td>
<td>0.03</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.50 0.3</td>
<td>0.02</td>
<td>0.03</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>1.00 0 3.00</td>
<td>0.02</td>
<td>0.03</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>1.00 0.3</td>
<td>0.02</td>
<td>0.03</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>1.00 0.7</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Condition</th>
<th>AGQ7 Bias</th>
<th>AGQ7 MSE</th>
<th>LA Bias</th>
<th>LA MSE</th>
<th>M-PQL Bias</th>
<th>M-PQL MSE</th>
<th>M-MQL Bias</th>
<th>M-MQL MSE</th>
<th>R-PQL Bias</th>
<th>R-PQL MSE</th>
<th>R-MQL Bias</th>
<th>R-MQL MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varphi_{11}$</td>
<td>0.25</td>
<td>0.75</td>
<td>0.02</td>
<td>0.02</td>
<td>-0.02</td>
<td>0.02</td>
<td>-0.12</td>
<td>0.02</td>
<td>-0.15</td>
<td>0.02</td>
<td>-0.15</td>
<td>0.02</td>
<td>-0.15</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.25</td>
<td>0.30</td>
<td>0.05</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.16</td>
<td>0.04</td>
<td>-0.08</td>
<td>0.01</td>
<td>-0.16</td>
<td>0.03</td>
<td>-0.07</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.50</td>
<td>0.00</td>
<td>0.15</td>
<td>0.04</td>
<td>0.15</td>
<td>0.04</td>
<td>0.06</td>
<td>0.04</td>
<td>-0.25</td>
<td>0.07</td>
<td>-0.33</td>
<td>0.11</td>
<td>-0.24</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.50</td>
<td>0.00</td>
<td>0.07</td>
<td>0.03</td>
<td>0.06</td>
<td>0.03</td>
<td>0.05</td>
<td>0.03</td>
<td>-0.17</td>
<td>0.04</td>
<td>-0.34</td>
<td>0.12</td>
<td>-0.16</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>1.00</td>
<td>0.00</td>
<td>0.06</td>
<td>0.04</td>
<td>0.05</td>
<td>0.03</td>
<td>0.02</td>
<td>0.05</td>
<td>-0.49</td>
<td>0.26</td>
<td>-0.73</td>
<td>0.53</td>
<td>-0.47</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>1.00</td>
<td>0.00</td>
<td>0.04</td>
<td>0.01</td>
<td>0.04</td>
<td>0.01</td>
<td>0.04</td>
<td>0.11</td>
<td>-0.36</td>
<td>0.16</td>
<td>-0.73</td>
<td>0.53</td>
<td>-0.35</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.25</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.25</td>
<td>0.00</td>
<td>-0.09</td>
<td>0.03</td>
<td>-0.11</td>
<td>0.03</td>
<td>-0.04</td>
<td>0.02</td>
<td>-0.05</td>
<td>0.01</td>
<td>-0.04</td>
<td>0.02</td>
<td>-0.05</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.25</td>
<td>0.00</td>
<td>0.15</td>
<td>0.05</td>
<td>0.17</td>
<td>0.05</td>
<td>0.19</td>
<td>0.06</td>
<td>-0.06</td>
<td>0.02</td>
<td>-0.13</td>
<td>0.03</td>
<td>-0.05</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.01</td>
<td>0.03</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.02</td>
<td>0.04</td>
<td>-0.02</td>
<td>0.04</td>
<td>-0.01</td>
<td>0.04</td>
<td>-0.08</td>
<td>0.03</td>
<td>-0.12</td>
<td>0.03</td>
<td>-0.07</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
<td>0.04</td>
<td>0.06</td>
<td>0.04</td>
<td>0.07</td>
<td>0.04</td>
<td>-0.11</td>
<td>0.04</td>
<td>-0.26</td>
<td>0.08</td>
<td>-0.1</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.05</td>
<td>0.02</td>
<td>0.04</td>
<td>0.02</td>
<td>0.05</td>
<td>0.02</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.05</td>
<td>0.02</td>
<td>0.05</td>
<td>0.02</td>
<td>0.06</td>
<td>-0.12</td>
<td>0.05</td>
<td>-0.21</td>
<td>0.06</td>
<td>-0.11</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.08</td>
<td>0.02</td>
<td>0.07</td>
<td>0.04</td>
<td>0.08</td>
<td>-0.19</td>
<td>0.09</td>
<td>-0.48</td>
<td>0.24</td>
<td>-0.17</td>
</tr>
</tbody>
</table>

Continued
Table 13 continued

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Condition</th>
<th>AGQ7</th>
<th>AGQ3</th>
<th>LA</th>
<th>M-PQL</th>
<th>M-MQL</th>
<th>R-PQL</th>
<th>R-MQL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\phi_{11}$</td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{e}^2$</td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>3.00</td>
<td>0.01</td>
<td>0.18</td>
<td>0</td>
<td>0.17</td>
<td>0.01</td>
<td>0.2</td>
</tr>
<tr>
<td>0.25</td>
<td>0.3</td>
<td>3.00</td>
<td>-0.13</td>
<td>0.22</td>
<td>-0.16</td>
<td>0.21</td>
<td>-0.17</td>
<td>0.23</td>
</tr>
<tr>
<td>0.25</td>
<td>0.7</td>
<td>3.00</td>
<td>-0.01</td>
<td>0.11</td>
<td>0</td>
<td>0.13</td>
<td>0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>0.50</td>
<td>0</td>
<td>3.00</td>
<td>-0.06</td>
<td>0.21</td>
<td>-0.06</td>
<td>0.22</td>
<td>-0.05</td>
<td>0.22</td>
</tr>
<tr>
<td>0.50</td>
<td>0.3</td>
<td>3.00</td>
<td>-0.02</td>
<td>0.17</td>
<td>-0.01</td>
<td>0.17</td>
<td>-0.04</td>
<td>0.18</td>
</tr>
<tr>
<td>0.50</td>
<td>0.7</td>
<td>3.00</td>
<td>-0.06</td>
<td>0.2</td>
<td>-0.05</td>
<td>0.19</td>
<td>-0.04</td>
<td>0.2</td>
</tr>
<tr>
<td>0.50</td>
<td>0.7</td>
<td>3.00</td>
<td>-0.06</td>
<td>0.19</td>
<td>-0.06</td>
<td>0.19</td>
<td>-0.06</td>
<td>0.19</td>
</tr>
<tr>
<td>1.00</td>
<td>0</td>
<td>3.00</td>
<td>-0.01</td>
<td>0.2</td>
<td>-0.01</td>
<td>0.2</td>
<td>-0.01</td>
<td>0.2</td>
</tr>
<tr>
<td>1.00</td>
<td>0.3</td>
<td>3.00</td>
<td>0.04</td>
<td>0.2</td>
<td>0.04</td>
<td>0.2</td>
<td>0.04</td>
<td>0.2</td>
</tr>
<tr>
<td>1.00</td>
<td>0.7</td>
<td>3.00</td>
<td>0.04</td>
<td>0.2</td>
<td>0.04</td>
<td>0.2</td>
<td>0.04</td>
<td>0.2</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>2.00</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Table 14. Percentage of confidence intervals ($\alpha=0.05$) that contain the true value of the parameter $\phi$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Condition</th>
<th>$\phi_1$</th>
<th>$\phi_2$</th>
<th>True Value</th>
<th>AGQ7</th>
<th>AGQ3</th>
<th>LA</th>
<th>M-POL</th>
<th>M-MQL</th>
<th>R-POL</th>
<th>R-MQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0^1$</td>
<td>.25 0 -2.50</td>
<td>94.84%</td>
<td>94.47%</td>
<td>93.93%</td>
<td>86.96%</td>
<td>85.28%</td>
<td>87.00%</td>
<td>85.28%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0^1$</td>
<td>.25 0.3 -2.50</td>
<td>94.74%</td>
<td>94.39%</td>
<td>92.86%</td>
<td>86.29%</td>
<td>82.27%</td>
<td>86.29%</td>
<td>82.27%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0^1$</td>
<td>.25 0.7 -2.50</td>
<td>95.24%</td>
<td>92.86%</td>
<td>94.23%</td>
<td>90.00%</td>
<td>80.00%</td>
<td>90.00%</td>
<td>80.00%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0^2$</td>
<td>.50 0 -2.50</td>
<td>95.49%</td>
<td>95.77%</td>
<td>94.60%</td>
<td>63.67%</td>
<td>47.67%</td>
<td>64.67%</td>
<td>47.67%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0^2$</td>
<td>.50 0.3 -2.50</td>
<td>96.26%</td>
<td>94.53%</td>
<td>95.36%</td>
<td>62.67%</td>
<td>50.00%</td>
<td>64.00%</td>
<td>50.00%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0^2$</td>
<td>.50 0.7 -2.50</td>
<td>96.49%</td>
<td>96.38%</td>
<td>96.85%</td>
<td>75.00%</td>
<td>47.67%</td>
<td>76.00%</td>
<td>47.67%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1^1$</td>
<td>1.0 0 -2.50</td>
<td>95.00%</td>
<td>95.32%</td>
<td>94.98%</td>
<td>93.98%</td>
<td>93.98%</td>
<td>93.98%</td>
<td>93.98%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1^1$</td>
<td>1.0 0.3 -2.50</td>
<td>96.98%</td>
<td>96.30%</td>
<td>96.98%</td>
<td>28.67%</td>
<td>5.67%</td>
<td>30.00%</td>
<td>5.67%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1^1$</td>
<td>1.0 0.7 -2.50</td>
<td>97.32%</td>
<td>97.32%</td>
<td>97.66%</td>
<td>48.00%</td>
<td>3.67%</td>
<td>49.67%</td>
<td>3.67%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1^2$</td>
<td>.25 0 15.0</td>
<td>94.05%</td>
<td>93.68%</td>
<td>93.52%</td>
<td>93.98%</td>
<td>93.98%</td>
<td>93.98%</td>
<td>93.98%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1^2$</td>
<td>.25 0.3 15.0</td>
<td>97.37%</td>
<td>99.07%</td>
<td>98.98%</td>
<td>94.98%</td>
<td>94.98%</td>
<td>94.98%</td>
<td>94.98%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1^2$</td>
<td>.25 0.7 15.0</td>
<td>95.24%</td>
<td>92.86%</td>
<td>94.23%</td>
<td>93.67%</td>
<td>93.67%</td>
<td>93.67%</td>
<td>93.67%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1^2$</td>
<td>.50 0 15.0</td>
<td>95.49%</td>
<td>95.77%</td>
<td>95.68%</td>
<td>95.67%</td>
<td>95.67%</td>
<td>95.67%</td>
<td>95.67%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1^2$</td>
<td>.50 0.3 15.0</td>
<td>93.46%</td>
<td>93.03%</td>
<td>93.30%</td>
<td>93.00%</td>
<td>93.00%</td>
<td>93.00%</td>
<td>93.00%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1^2$</td>
<td>.50 0.7 15.0</td>
<td>95.18%</td>
<td>95.48%</td>
<td>95.50%</td>
<td>95.00%</td>
<td>95.00%</td>
<td>95.00%</td>
<td>95.00%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1^2$</td>
<td>1.0 0 15.0</td>
<td>95.33%</td>
<td>95.32%</td>
<td>95.32%</td>
<td>95.33%</td>
<td>95.33%</td>
<td>95.33%</td>
<td>95.33%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1^2$</td>
<td>1.0 0.3 15.0</td>
<td>94.63%</td>
<td>94.61%</td>
<td>94.63%</td>
<td>94.67%</td>
<td>94.67%</td>
<td>94.67%</td>
<td>94.67%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1^2$</td>
<td>1.0 0.7 15.0</td>
<td>96.31%</td>
<td>96.31%</td>
<td>96.32%</td>
<td>96.33%</td>
<td>96.33%</td>
<td>96.33%</td>
<td>96.33%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
Table 14 continued

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Condition</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>True Value</th>
<th>AGQ?</th>
<th>AGQ3</th>
<th>LA</th>
<th>M-PQL</th>
<th>M-MQL</th>
<th>R-PQL</th>
<th>R-MQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>(.25)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>98.81%</td>
<td>98.81%</td>
<td>98.79%</td>
<td>96.99%</td>
<td>96.32%</td>
<td>97.67%</td>
<td>96.32%</td>
</tr>
<tr>
<td>(.25)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>87.72%</td>
<td>86.92%</td>
<td>83.67%</td>
<td>91.30%</td>
<td>89.63%</td>
<td>92.64%</td>
<td>90.97%</td>
</tr>
<tr>
<td>(.25)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>95.24%</td>
<td>94.64%</td>
<td>96.15%</td>
<td>90.00%</td>
<td>76.33%</td>
<td>91.00%</td>
<td>78.33%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>97.22%</td>
<td>97.89%</td>
<td>97.84%</td>
<td>95.67%</td>
<td>95.33%</td>
<td>95.67%</td>
<td>95.33%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>91.12%</td>
<td>91.04%</td>
<td>89.69%</td>
<td>90.00%</td>
<td>85.00%</td>
<td>90.67%</td>
<td>85.33%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>98.68%</td>
<td>98.64%</td>
<td>99.55%</td>
<td>84.33%</td>
<td>48.33%</td>
<td>84.67%</td>
<td>50.67%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>97.00%</td>
<td>97.32%</td>
<td>97.32%</td>
<td>98.00%</td>
<td>97.33%</td>
<td>98.00%</td>
<td>97.33%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>92.95%</td>
<td>92.59%</td>
<td>93.62%</td>
<td>85.00%</td>
<td>62.33%</td>
<td>87.33%</td>
<td>64.00%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>94.97%</td>
<td>93.96%</td>
<td>94.31%</td>
<td>82.00%</td>
<td>10.67%</td>
<td>83.67%</td>
<td>13.33%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>87.72%</td>
<td>86.92%</td>
<td>83.67%</td>
<td>91.30%</td>
<td>89.63%</td>
<td>92.64%</td>
<td>90.97%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>95.24%</td>
<td>94.64%</td>
<td>96.15%</td>
<td>90.00%</td>
<td>76.33%</td>
<td>91.00%</td>
<td>78.33%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>97.22%</td>
<td>97.89%</td>
<td>97.84%</td>
<td>95.67%</td>
<td>95.33%</td>
<td>95.67%</td>
<td>95.33%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>91.12%</td>
<td>91.04%</td>
<td>89.69%</td>
<td>90.00%</td>
<td>85.00%</td>
<td>90.67%</td>
<td>85.33%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>98.68%</td>
<td>98.64%</td>
<td>99.55%</td>
<td>84.33%</td>
<td>48.33%</td>
<td>84.67%</td>
<td>50.67%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>97.00%</td>
<td>97.32%</td>
<td>97.32%</td>
<td>98.00%</td>
<td>97.33%</td>
<td>98.00%</td>
<td>97.33%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>92.95%</td>
<td>92.59%</td>
<td>93.62%</td>
<td>85.00%</td>
<td>62.33%</td>
<td>87.33%</td>
<td>64.00%</td>
</tr>
<tr>
<td>(.50)</td>
<td>(.00)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>94.97%</td>
<td>93.96%</td>
<td>94.31%</td>
<td>82.00%</td>
<td>10.67%</td>
<td>83.67%</td>
<td>13.33%</td>
</tr>
</tbody>
</table>
Table 14 continued

<table>
<thead>
<tr>
<th>Parameter ( \Phi_{11} )</th>
<th>Condition True Value</th>
<th>AGQ7</th>
<th>AGQ3</th>
<th>LA</th>
<th>M-PQL</th>
<th>M-MQL</th>
<th>R-PQL</th>
<th>R-MQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi )</td>
<td>.25 0 3.00</td>
<td>94.84%</td>
<td>94.86%</td>
<td>94.74%</td>
<td>95.32%</td>
<td>95.32%</td>
<td>94.67%</td>
<td>94.65%</td>
</tr>
<tr>
<td></td>
<td>.25 0.3 3.00</td>
<td>93.86%</td>
<td>93.46%</td>
<td>92.86%</td>
<td>93.65%</td>
<td>93.65%</td>
<td>93.31%</td>
<td>93.31%</td>
</tr>
<tr>
<td></td>
<td>.25 0.7 3.00</td>
<td>98.41%</td>
<td>98.21%</td>
<td>98.08%</td>
<td>96.00%</td>
<td>96.00%</td>
<td>96.33%</td>
<td>96.33%</td>
</tr>
<tr>
<td></td>
<td>.50 0 3.00</td>
<td>92.01%</td>
<td>91.90%</td>
<td>91.73%</td>
<td>93.00%</td>
<td>93.00%</td>
<td>93.00%</td>
<td>93.00%</td>
</tr>
<tr>
<td></td>
<td>.50 0.3 3.00</td>
<td>94.86%</td>
<td>95.02%</td>
<td>93.81%</td>
<td>95.33%</td>
<td>95.33%</td>
<td>96.00%</td>
<td>96.00%</td>
</tr>
<tr>
<td></td>
<td>.50 0.7 3.00</td>
<td>91.67%</td>
<td>91.86%</td>
<td>91.44%</td>
<td>92.67%</td>
<td>92.67%</td>
<td>92.67%</td>
<td>92.67%</td>
</tr>
<tr>
<td></td>
<td>1.0 0 3.00</td>
<td>91.67%</td>
<td>91.64%</td>
<td>91.64%</td>
<td>93.67%</td>
<td>93.67%</td>
<td>95.00%</td>
<td>95.00%</td>
</tr>
<tr>
<td></td>
<td>1.0 0.3 3.00</td>
<td>92.28%</td>
<td>92.59%</td>
<td>92.28%</td>
<td>93.67%</td>
<td>93.67%</td>
<td>93.33%</td>
<td>93.33%</td>
</tr>
<tr>
<td></td>
<td>1.0 0.7 3.00</td>
<td>93.62%</td>
<td>93.62%</td>
<td>93.65%</td>
<td>96.33%</td>
<td>96.33%</td>
<td>96.33%</td>
<td>96.33%</td>
</tr>
<tr>
<td>( \sigma^2_e )</td>
<td>.50 0 2.00</td>
<td>95.63%</td>
<td>95.65%</td>
<td>95.55%</td>
<td>94.65%</td>
<td>94.65%</td>
<td>95.00%</td>
<td>94.98%</td>
</tr>
<tr>
<td></td>
<td>.50 0.3 2.00</td>
<td>93.86%</td>
<td>92.52%</td>
<td>93.88%</td>
<td>92.98%</td>
<td>92.98%</td>
<td>92.31%</td>
<td>92.31%</td>
</tr>
<tr>
<td></td>
<td>.50 0.7 2.00</td>
<td>88.89%</td>
<td>89.29%</td>
<td>88.46%</td>
<td>94.33%</td>
<td>94.33%</td>
<td>94.00%</td>
<td>94.00%</td>
</tr>
<tr>
<td></td>
<td>1.0 0 2.00</td>
<td>92.71%</td>
<td>92.96%</td>
<td>93.17%</td>
<td>92.67%</td>
<td>92.67%</td>
<td>92.67%</td>
<td>92.67%</td>
</tr>
<tr>
<td></td>
<td>1.0 0.3 2.00</td>
<td>93.93%</td>
<td>93.53%</td>
<td>93.81%</td>
<td>95.00%</td>
<td>95.00%</td>
<td>95.33%</td>
<td>95.33%</td>
</tr>
<tr>
<td></td>
<td>1.0 0.7 2.00</td>
<td>94.30%</td>
<td>94.57%</td>
<td>94.59%</td>
<td>94.33%</td>
<td>94.33%</td>
<td>94.67%</td>
<td>94.67%</td>
</tr>
<tr>
<td></td>
<td>1.0 0 2.00</td>
<td>94.67%</td>
<td>94.65%</td>
<td>94.65%</td>
<td>95.33%</td>
<td>95.33%</td>
<td>95.33%</td>
<td>95.33%</td>
</tr>
<tr>
<td></td>
<td>1.0 0.3 2.00</td>
<td>94.30%</td>
<td>94.28%</td>
<td>94.30%</td>
<td>95.00%</td>
<td>95.00%</td>
<td>95.00%</td>
<td>95.00%</td>
</tr>
<tr>
<td></td>
<td>1.0 0.7 2.00</td>
<td>92.28%</td>
<td>92.28%</td>
<td>92.31%</td>
<td>93.00%</td>
<td>93.00%</td>
<td>93.33%</td>
<td>93.33%</td>
</tr>
<tr>
<td>Parameter</td>
<td>Condition</td>
<td>AGQ7</td>
<td>AGQ3</td>
<td>LA</td>
<td>M-PQL</td>
<td>M-MQL</td>
<td>R-PQL</td>
<td>R-MQL</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>------</td>
<td>------</td>
<td>----</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>$\rho_{12}$</td>
<td>Mean</td>
<td>Bias</td>
<td>Mean</td>
<td>Bias</td>
<td>Mean</td>
<td>Bias</td>
<td>Mean</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>0.25</td>
<td>0.28</td>
<td>0.03</td>
<td>0.27</td>
<td>0.02</td>
<td>0.23</td>
<td>-0.02</td>
</tr>
<tr>
<td>0.25</td>
<td>0.3</td>
<td>0.25</td>
<td>0.30</td>
<td>0.05</td>
<td>0.28</td>
<td>0.03</td>
<td>0.25</td>
<td>0.00</td>
</tr>
<tr>
<td>0.25</td>
<td>0.7</td>
<td>0.25</td>
<td>0.44</td>
<td>0.19</td>
<td>0.43</td>
<td>0.18</td>
<td>0.41</td>
<td>0.16</td>
</tr>
<tr>
<td>0.50</td>
<td>0</td>
<td>0.50</td>
<td>0.54</td>
<td>0.04</td>
<td>0.52</td>
<td>0.02</td>
<td>0.52</td>
<td>0.02</td>
</tr>
<tr>
<td>0.50</td>
<td>0.3</td>
<td>0.50</td>
<td>0.59</td>
<td>0.09</td>
<td>0.56</td>
<td>0.06</td>
<td>0.56</td>
<td>0.06</td>
</tr>
<tr>
<td>0.50</td>
<td>0.7</td>
<td>0.50</td>
<td>0.58</td>
<td>0.08</td>
<td>0.57</td>
<td>0.07</td>
<td>0.55</td>
<td>0.05</td>
</tr>
<tr>
<td>1.00</td>
<td>0</td>
<td>1.00</td>
<td>1.08</td>
<td>0.08</td>
<td>1.00</td>
<td>0.00</td>
<td>1.12</td>
<td>0.12</td>
</tr>
<tr>
<td>1.00</td>
<td>0.3</td>
<td>1.00</td>
<td>1.03</td>
<td>0.03</td>
<td>0.97</td>
<td>-0.03</td>
<td>1.05</td>
<td>0.05</td>
</tr>
<tr>
<td>1.00</td>
<td>0.7</td>
<td>1.00</td>
<td>1.04</td>
<td>0.04</td>
<td>0.99</td>
<td>-0.01</td>
<td>1.04</td>
<td>0.04</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.25</td>
<td>0.3</td>
<td>0.26</td>
<td>0.15</td>
<td>-0.11</td>
<td>0.14</td>
<td>-0.12</td>
<td>0.15</td>
<td>-0.11</td>
</tr>
<tr>
<td>0.25</td>
<td>0.7</td>
<td>0.61</td>
<td>0.80</td>
<td>0.19</td>
<td>0.80</td>
<td>0.19</td>
<td>0.79</td>
<td>0.19</td>
</tr>
<tr>
<td>0.50</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.50</td>
<td>0.3</td>
<td>0.37</td>
<td>0.35</td>
<td>-0.02</td>
<td>0.34</td>
<td>-0.02</td>
<td>0.36</td>
<td>-0.01</td>
</tr>
<tr>
<td>0.50</td>
<td>0.7</td>
<td>0.86</td>
<td>0.93</td>
<td>0.07</td>
<td>0.93</td>
<td>0.07</td>
<td>0.93</td>
<td>0.07</td>
</tr>
<tr>
<td>1.00</td>
<td>0</td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>1.00</td>
<td>0.3</td>
<td>0.52</td>
<td>0.53</td>
<td>0.01</td>
<td>0.52</td>
<td>0.00</td>
<td>0.54</td>
<td>0.02</td>
</tr>
<tr>
<td>1.00</td>
<td>0.7</td>
<td>1.20</td>
<td>1.24</td>
<td>0.03</td>
<td>1.23</td>
<td>0.02</td>
<td>1.25</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 15. Mean and average bias of parameter estimates for runs that converged successfully and where $\Phi$ was positive definite for all estimation methods. Those values that changed by ±.05 or more are shown in bold.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Condition</th>
<th>AGQ7 Mean</th>
<th>AGQ7 Bias</th>
<th>AGQ3 Mean</th>
<th>AGQ3 Bias</th>
<th>LA Mean</th>
<th>LA Bias</th>
<th>M-PQL Mean</th>
<th>M-PQL Bias</th>
<th>M-MQL Mean</th>
<th>M-MQL Bias</th>
<th>R-PQL Mean</th>
<th>R-PQL Bias</th>
<th>R-MQL Mean</th>
<th>R-MQL Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25 0.25</td>
<td>0.25 0.25</td>
<td>3.01 0.01</td>
<td>3.00 0.00</td>
<td>3.00 0.00</td>
<td>3.00 0.00</td>
<td>3.00 0.00</td>
<td>3.00 0.00</td>
<td>3.03 0.03</td>
<td>3.03 0.03</td>
<td>3.03 0.03</td>
<td>3.03 0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25 0.3 0.25</td>
<td>3.01 0.01</td>
<td>2.83 -0.17</td>
<td>2.82 -0.18</td>
<td>2.83 -0.17</td>
<td>2.83 -0.17</td>
<td>2.83 -0.17</td>
<td>2.83 -0.17</td>
<td>2.86 -0.14</td>
<td>2.86 -0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25 0.7 0.25</td>
<td>3.01 0.01</td>
<td>2.99 -0.01</td>
<td>3.01 0.01</td>
<td>3.01 0.01</td>
<td>3.01 0.01</td>
<td>3.01 0.01</td>
<td>3.04 0.04</td>
<td>3.04 0.04</td>
<td>3.04 0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50 0.5 0.5</td>
<td>2.95 -0.05</td>
<td>2.95 -0.05</td>
<td>2.95 -0.05</td>
<td>2.95 -0.05</td>
<td>2.95 -0.05</td>
<td>2.95 -0.05</td>
<td>2.98 -0.02</td>
<td>2.98 -0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50 0.3 0.5</td>
<td>2.97 -0.03</td>
<td>2.97 -0.03</td>
<td>2.96 -0.04</td>
<td>2.96 -0.04</td>
<td>2.96 -0.04</td>
<td>2.96 -0.04</td>
<td>2.99 -0.01</td>
<td>2.99 -0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50 0.7 0.5</td>
<td>2.96 -0.04</td>
<td>2.96 -0.04</td>
<td>2.96 -0.04</td>
<td>2.96 -0.04</td>
<td>2.96 -0.04</td>
<td>2.96 -0.04</td>
<td>2.99 -0.01</td>
<td>2.99 -0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00 1.0 1.0</td>
<td>2.94 -0.06</td>
<td>2.94 -0.06</td>
<td>2.94 -0.06</td>
<td>2.94 -0.06</td>
<td>2.94 -0.06</td>
<td>2.94 -0.06</td>
<td>2.97 -0.03</td>
<td>2.97 -0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00 0.3 1.0</td>
<td>2.99 -0.01</td>
<td>2.99 -0.01</td>
<td>2.99 -0.01</td>
<td>2.99 -0.01</td>
<td>2.99 -0.01</td>
<td>2.99 -0.01</td>
<td>3.02 0.02</td>
<td>3.02 0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00 0.7 1.0</td>
<td>3.04 0.04</td>
<td>3.04 0.04</td>
<td>3.04 0.04</td>
<td>3.04 0.04</td>
<td>3.04 0.04</td>
<td>3.04 0.04</td>
<td>3.08 0.08</td>
<td>3.08 0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 15 continued