Mixed-Effects Models for Count Data with Applications to Educational Research

Jihyung Shin
MIXED-EFFECTS MODELS FOR COUNT DATA WITH APPLICATIONS TO EDUCATIONAL RESEARCH.

By

JIHYUNG SHIN

A Dissertation submitted to the Department of Statistics in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Degree Awarded: Spring Semester, 2012
Jihyung Shin defended this dissertation on March 27, 2012.
The members of the supervisory committee were:

Xufeng Niu
Professor Directing Dissertation

Shouping Hu
University Representative

Stephanie AlOtaiba
Committee Member

Daniel McGee
Committee Member

Wei Wu
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with the university requirements.
I dedicate this dissertation to my family.
ACKNOWLEDGMENTS

I express my appreciation to many colleagues and professors from statistics department at the Florida State University. I am indebted to my adviser Dr. Xufeng Niu for his guidance, patient and encouragement. I also would like to thank my committee members, Dr. McGee, Dr. Wei Wu and Dr. Shouping Hu for generous support in completing this dissertation. I would like to acknowledge Dr. Stephanie Al Otaiba. Her guidance and advise on educational research gave me inspiration on the dissertation topic.

I also would like to thank Dr. Ramsier for his guidance in statistics education. I had great experience in teaching with his support. I also like to thank Dr. Henry Carreta for the invaluable experience working with him as a statistician in the Department of Family Medicine and Rural Health at College of Medicine of Florida State University.

Finally, I cannot thank enough to my family for helping me to finish this journey.
# TABLE OF CONTENTS

List of Tables ................................................................. vii
List of Figures ................................................................. ix
Abstract ................................................................. x

1 Introduction ................................................................. 1
   1.1 Motivation of the Study ....................................... 2
   1.2 Summary of Chapters ....................................... 6

2 Statistical Models and Estimation ........................................ 8
   2.1 Mixed-effects Linear Models ............................... 8
   2.2 Mixed-effects and Mixed-distribution Models ............... 12
   2.3 Zero-inflated Poisson Models ............................... 15
   2.4 Proposed Model .............................................. 19
       2.4.1 The Models .............................................. 19
       2.4.2 The Estimation .......................................... 21
       2.4.3 Model Comparisons and Selection Criteria ............ 23

3 Simulation study ......................................................... 26
   3.1 Simulation Design .............................................. 26
   3.2 Simulation Results ............................................. 28

4 Application to letter-sound fluency of kindergarten children .... 40
   4.1 Data Description .............................................. 40
   4.2 Model Specification ........................................... 46
   4.3 Summary of the results ....................................... 47

5 Conclusion and Future direction ....................................... 51

Appendices ................................................................. 55

A Additional results .................................................. 55

B SAS Codes ................................................................. 60

C IRB Approval memorandums ........................................... 63

References ................................................................. 67
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>True parameter values of simulated data sets, eight cases</td>
<td>28</td>
</tr>
<tr>
<td>3.2</td>
<td>Simulation results for the proposed model using 100 sets of simulated data sets ($n = 3$) from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 1</td>
<td>29</td>
</tr>
<tr>
<td>3.3</td>
<td>Simulation results for the proposed model using 100 sets of simulated datasets ($n = 3$) from the model with correlated random effects given in Section 2.4 with true parameter values of case 1</td>
<td>30</td>
</tr>
<tr>
<td>3.4</td>
<td>Simulation results for the proposed model using 100 sets of simulated datasets with $m = 100$ subjects and different $n$ from the model with correlated random effects given in Section 2.4 with true parameter values of case 1</td>
<td>31</td>
</tr>
<tr>
<td>3.5</td>
<td>Simulation results for the proposed model using 100 sets of simulated datasets ($n = 10$) from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 1</td>
<td>32</td>
</tr>
<tr>
<td>3.6</td>
<td>Simulation results for the proposed model using 100 sets of simulated datasets ($n = 10$) from the model with correlated random effects given in Section 2.4 with true parameter values of case 1</td>
<td>33</td>
</tr>
<tr>
<td>3.7</td>
<td>Simulation results for the proposed model using 100 sets of simulated datasets ($n = 3$) from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 2</td>
<td>33</td>
</tr>
<tr>
<td>3.8</td>
<td>Simulation results for the proposed model using 100 sets of simulated datasets ($n = 3$) from the model with correlated random effects given in Section 2.4 with true parameter values of case 2</td>
<td>34</td>
</tr>
<tr>
<td>3.9</td>
<td>Simulation results for the proposed model using 100 sets of simulated datasets ($n = 10$) from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 2</td>
<td>34</td>
</tr>
<tr>
<td>3.10</td>
<td>Simulation results for the proposed model using 100 sets of simulated datasets ($n = 10$) from the model with correlated random effects given in Section 2.4 with true parameter values of case 2</td>
<td>35</td>
</tr>
</tbody>
</table>
3.11 Simulation results for the proposed model using 100 sets of simulated datasets $(n = 3)$ from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 3 .............................................. 35
3.12 Simulation results for the proposed model using 100 sets of simulated datasets $(n = 3)$ from the model with correlated random effects given in Section 2.4 with true parameter values of case 3 .............................................. 36
3.13 Simulation results for the proposed model using 100 sets of simulated datasets $(n = 10)$ from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 3 .............................................. 36
3.14 Simulation results for the proposed model using 100 sets of simulated datasets $(n = 10)$ from the model with correlated random effects given in Section 2.4 with true parameter values of case 3 .............................................. 37
3.15 Simulation results for the proposed model using 100 sets of simulated datasets $(n = 3)$ from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 4 .............................................. 37
3.16 Simulation results for the proposed model using 100 sets of simulated datasets $(n = 3)$ from the model with correlated random effects given in Section 2.4 with true parameter values of case 4 .............................................. 38
3.17 Simulation results for the proposed model using 100 sets of simulated datasets $(n = 10)$ from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 4 .............................................. 38
3.18 Simulation results for the proposed model using 100 sets of simulated datasets $(n = 10)$ from the model with correlated random effects given in Section 2.4 with true parameter values of case 4 .............................................. 39

4.1 Descriptive statistics of the sample. .................................................. 41
4.2 Pearson Correlation Coefficients of AIMSWeb LSF measured at fall, winter and spring, $n = 461$ ................................................................. 42
4.3 Parameter estimates of the full model. ............................................. 48
4.4 Parameter estimates of the reduced model. ..................................... 49
A.1 Parameter estimates for Zero-Inflated Poisson Models. ................. 56
A.2 Parameter estimates for zero-inflated Poisson models with the shape parameter $\tau$. ................................................................. 57
A.3 Parameter estimates of the full model with the shape parameter $\tau$. .... 57
A.4 Parameter estimates of the reduced model with the shape parameter $\tau$. ... 58
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Matthew Effect in Reading (Excerpt from Wren (2000) [70])</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Distribution of all observed AIMS Letter Sound Fluency (LSF) test scores of</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>461 children with 3 measurements per child</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>AIMS score of Fall, Winter and Spring test for four subjects [(a)-(d)]</td>
<td>44</td>
</tr>
<tr>
<td>4.2</td>
<td>Distribution of AIMS score of Fall, Winter, and Spring.</td>
<td>45</td>
</tr>
<tr>
<td>4.3</td>
<td>Histograms and qq plots of random effects from the model with correlated</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>random effects.</td>
<td></td>
</tr>
<tr>
<td>A.1</td>
<td>Histograms and qq plots of random effects from the model with correlate ran-</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>andom effects and $\tau$</td>
<td></td>
</tr>
</tbody>
</table>
ABSTRACT

This research is motivated by an analysis of reading research data. We are interested in modeling the test outcome of ability to fluently recode letters into sounds of kindergarten children aged between 5 and 7. The data showed excessive zero scores (more than 30% of children) on the test. In this dissertation, we carefully examine the models dealing with excessive zeros, which are based on the mixture of distributions, a distribution with zeros and a standard probability distribution with non negative values. In such cases, a log normal variable or a Poisson random ($\lambda$) variable is often observed with probability ($p$) from semicontinuous data or count data. The previously proposed models, mixed-effects and mixed-distribution models (MEMD) by Tooze(2002) et al. for semicontinuous data and zero-inflated Poisson (ZIP) regression models by Lambert(1992) for count data are reviewed.

We apply zero-inflated Poisson models to repeated measures data of zero-inflated data by introducing a pair of possibly correlated random effects to the zero-inflated Poisson model to accommodate within-subject correlation and between subject heterogeneity. The model describes the effect of predictor variables on the probability of nonzero responses (occurrence) and mean of nonzero responses (intensity) separately. The likelihood function is maximized using dual quasi-Newton optimization of an approximated by adaptive Gaussian quadrature. The maximum likelihood estimates are obtained through standard statistical software package. Using different model parameters, the number of subject, and the number of measurements per subject, the simulation study is conducted and the results are presented.

The dissertation ends with the application of the model to reading research data and future research. We examine the number of correct letter sound counted of children collected over 2008 -2009 academic year. We find that age, gender and socioeconomic status are significantly related to the letter sound fluency of children in both parts of the model. The model provides better explanation of data structure and easier interpretations of parameter values, as they are the same as in standard logistic models and Poisson regression models. The model can be extended to accommodate serial correlation which can be observed in longitudinal data. Also, one may consider multilevel zero-inflated Poisson model. Although the multilevel model was proposed previously, parameter estimation by penalized quasi likelihood methods is questionable, and further examination is needed.
CHAPTER 1

INTRODUCTION

In data analyses, sometimes observations have a nested structure. In other words, the observations nested in person, that person can be also nested in group and so on. This structure can be seen in many different research areas. For example, in organizational studies where a researcher is interested in the different workplace characteristics affecting worker productivity. Another example can be found in public health studies where fertility rates are studied across counties nested in different state level.

Not only multiple individuals but also multiple measurements can be nested within higher level of cluster unit. In other words, the nested structure can be observed when there are multiple measurements or observations of one individual at multiple time points. In this case, the measurements are naturally nested in person. When such data are analyzed, we often investigate the changes over time which measures the characteristics of the individuals repeatedly. For example, in medical research, the observation might be a size of tumor, in biology, the number of salmonellae can be counted to investigate the growth of salmonellae depending on under conditions possibly affecting food product contamination, or in education, a researcher may be interested in a linear or a quadratic growth rate of learning ability on a student.

Over decades, statisticians have developed models that can distinguish and both between and within variability of each individual or a group through various approaches, such as multivariate regression, analysis of variance or repeated measure analysis of variance. One popular approach for analysis of such data is building a hierarchical linear model which can also be rewritten as the two-stage models by Laird and Ware (1982) [31]. With general covariance structure, it is preferred because the application of the two-stage models for unbalanced data is easier than using multivariate models [31].

As mentioned above, such hierarchical data structure can be observed in many research areas, but one popular example might come from an educational research, where data are naturally structured in a form of a hierarchy. Students are grouped in classes, classes are belonged to schools and schools are within a certain geographic district and so on. In educational research, where multiple level of hierarchy can be naturally observed in data, hierarchical models allow general statistical framework for many researches. Although there is no limitation in terms of the number of levels defined in hierarchical models, the two-level and three-level models are the most widely used ones. For example, as early in 1980’s, Braun et al. (1983)[6] used hierarchical linear models to examine the prediction of later
academic success depending on standardized test scores among different graduate business schools. Since there were only few minority students which were subjects of research interest, differentiating the prediction models for each school played a essential part of analysis rather than using pooled data across schools.

As well as two-level hierarchical models with nested subjects within schools, three-level hierarchies are often faced as a researcher. For example, Nye et al. (2000) [47] applied three-level hierarchical linear models to predict reading and math scores of each students, in purpose of investigating classroom size. Students scores were collected within classrooms within schools in Tennessee. In larger scale, Raudenbush (1999) [55] analyzed United States National Assessment of Educational Progress data to investigate state level variation in mathematics achievements. The data contains mathematics scores of students nested within schools nested within states. Despite multiple number of levels in hierarchy, all the essential statistical features can be found in the basic two-level models (Raudenbush & Bryk’s (2002))[54], which is reviewed in detail in chapter 2.

### 1.1 Motivation of the Study

This study is motivated by the analysis of repeatedly measured letter-sound fluency (LSF) of 461 kindergarten children. The data was originally collected by Al Otaiba et al. (2011) as part of a cluster-randomized control field trial using Individualized Student Instruction for Kindergarten on reading ability [2]. In large scale of learning behavior, "reading" is the foundation for any other learning activities. In other words, people cannot comprehend anything unless they read fluently what they are reading [14]. Among many measurements on reading skills, it has been shown that LSF plays an important role. In a study by Stage et el. (2001) [63], LSF of kindergarten was a significant predictor of first-grade reading growth. Moreover, Speece et al. (2005)[56] argued that LSF could enhance early literacy skill development as early as kindergarten. Recently, Al Otaiba et al. (2010)[3] showed that LSF had a strong relationship with a spelling ability and suggested an ongoing curriculum-based progress monitoring of LSF and suggested the LSF as a mean of identifying students at risk for leaning disabilities.

Another aspect of importance on reading education at early age can be explained by looking at the "Matthew Effect" [68] [64], which is used to describe the phenomenon, such that a student with good foundation skills later performs better academically while a student without such skills keep performing poorly. Throughout years of researches, the idea of Matthew Effects gained empirical supports in reading ability. The term, Matthew Effect, derived from a passage from the New Testament;

> For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath. (Matthew 25:29)

While the Matthew Effect and its effects on learning disability has been observed and concerned [61], it was demonstrated that early detection and intervention can improve reading ability of children with risk of reading disability[8]. As a recently published report by the Early Grade Learning Community of Practice was titled by Early Reading: Igniting
Figure 1.1: Matthew Effect in Reading (Excerpt from Wren (2000) [70])

Education for All [14], the learning process would be awaken at the moment when a child was able to read, and the direct objective of reading comprehension is learning. In this report, the authors conclude that, in developing countries, even the nation’s social and economic growth depends on the reading education. As the Matthew effect implies that lack of early intervention and investment on reading education would cause cumulative effects on overall leaning. It would ultimately lead the larger inequalities over the period, efficient instruction on early reading ability was urged not only in local scale, but also in global scale.

In the original study by Al Otaiba et al. (2011) [2], multiple aspects of students’ language and conventional literacy skills were measured using different norm-referenced tools. These include letter-sound correspondence, expressive vocabulary growth, word reading skills, and ability to decode pseudo words \(^1\). Among these assessments, our particular interest was an investigation of the relationship between the letter-sound fluency and the characteristics of individuals. The students’ letter-sound correspondence was measured by AIMSWeb Letter Sound Fluency \(^2\) (Shinn & Shinn (2004) [60]). The test protocol present 10 rows of 10 lowercase letters and children were asked to say the sound of the letters make in one minute. However, if a child could not produce any correct sound in the first ten, the test was discontinued. The number of correct sounds were recorded. When the child finished naming

\(^1\)Different tools were used for measurements of each items. For details, refer Al Otaiba et al. (2011) [2]
\(^2\)Alternate form reliability is .90
the letter sound within one minute, the number of correct pronunciation were re-scaled per sixty seconds.

In exploratory data analysis, we found that there were significant amounts of students who scored a zero on the test resulting highly skewed distribution of scores, and a hierarchical linear model was not applicable. As we can see in Figure 1.2, there is a peak at zeros on the distribution of the observed count of correctly pronounced letters measured for one minute. It means that a child was not able to read or recognize the letter.

![Distribution of aims](image)

**Figure 1.2**: Distribution of all observed AIMS Letter Sound Fluency (LSF) test scores of 461 children with 3 measurements per child

A similar situation with non-ignorable observation of excessive zeros often arose in econometric, and the observed zeros can even be very meaningful. For instance, zero expenses or costs can be easily observed in medical expenditures. Another example may be found in the number of claims on insurance per person. A person may not make any insurance claim after purchase, or make multiple claims. There have been studies to deal with these problems including nonparametric approaches. However, nonparametric approaches are only considering the rank of the observed data value, so it is hard to adapt the approach when
there is a meaningful observed value of zeros. Another consideration would be a two-part model proposed by Duan, Maning, Morris and Newhouse (1987) [12]. Although this two-part model is very popular in econometrics, the model separated the observed values in two part and does not deal with the dependency among the observations of data in longitudinal setting. In analyses of data with many zeros in longitudinal data, we would want to accommodate for within subject correlation and between subject heterogeneity.

For repeated measures data with many zeros, a two-part random-effect models have been proposed by Olsen and Schafer (2001) [48]. Later, Tooze (2002) [66] proposed mixed-effects and mixed-distribution models based on papers by Lachenburch (1976 [29], 1992 [30]) for cross-sectional data and Grunwald and Jones (2000) [16] for time series data. The original work was designed to develop a model which can be used in analysis of monthly medical expenditure data with clumping at zero. Basically, the model considers observed values as a mixture of distributions, logistic models for the occurrence, and normal or log normal models for intensity. The first part of the model estimates the probability of having non-zero response, then the second part of the model deals with the magnitude of non-zero responses. A pair random effects are allowed in the model for within person variations. The relationship between observations from each of distributions is explained by allowing correlation in between random effects.

With a discrete random variable of non-ignorable excessive zeros, a hurdle model by Mullahy (1986) [44] and a zero-inflated Poisson (ZIP) regression by Lambert (1992) [33] are popular approaches to analyze such data. The hurdle model regards the observed value with clumps as a hurdle. Then the model is developed in two parts, the probability of clearing the hurdle, and the zero-truncated form of a discrete distribution. In many application, Poisson or negative binomial distribution are popular choices. This model has advantage of dealing with both zero-inflated data and zero-deflated data.

On the other hand, the ZIP regression model is based on the independent counts with a mixture of Bernoulli and Poisson distributions. The hurdle model and the ZIP model are closely related as the ZIP regression model can be reparameterized into the hurdle model. However, they are not equivalent to each other as different parameters are modeled in regression context. Later, the ZIP regression model has been extended into transition models, which are extensions of generalized linear models by Dunson and Haseman (1999) [13]. This model explains a conditional distribution of the response as a function of past responses and covariates to accommodate a longitudinal data. It was applied to prediction of carcinogenicity from animal studies with counts of the number of tumors present over time.

The zero-inflated count data can also be observed in repeated measures data or longitudinal setting. In general, the within subject variations are specified by introducing random effects in the model. In previous literature, the extension of zero-inflated Poisson model with random effects was initiated with the model proposed by Hall (2000) [18]. Hall extended the model to an upper bounded count situation with added random effects into the Poisson part of the model, so that the model can handle repeated measure data. The model was applied to both analysis of Lambert’s printed wiring board data from his original paper, and the whitefly data from horticultural experiment.

As there are two parts in the models with mixture of distributions, two independent random effects in the zero-inflated Poisson model were considered by Yau and Lee (2001)
The models were applied to occupational injury program data and the length of stay for hospital patients data respectively. The zero-inflated Poisson models with random effects were also applied to health care outcomes research. Hur et al. (2002) applied the zero-inflated Poisson models with a pair of independent random effects to the number of complications data of patients within medical facilities.

Min and Agresti (2005) [43] introduced correlated random effects in the model considering zero-inflated count data. They chose Hurdle models [44], which can be regarded as models with reparametrization of the zero-inflated Poisson model. In the Hurdle models, the positive part of the response variable is assumed to follow a truncated count distribution with probability mass function with the mean value from the untruncated count distribution. Min and Agresti conducted a simulation study and showed the superiority of the Hurdle models over zero-inflated Poisson models in dealing with zero-inflated count data. However, as the major difference between the Hurdle model and the zero-inflated Poisson model lies on the interpretation of the logit component of the model, the choice of model should depend on the question of interest. More recently, Gschlößl and Gzado (2008) [45] extended the zero-inflated Poisson model to spatially correlated data by specifying random effects with Gaussian conditional autoregressive model.

1.2 Summary of Chapters

Followed by this introduction chapter, the previously developed models are discussed in chapter 2. The models include the basic model for repeated measures data, the mixed-effects linear models. The basic structure of the mixed-effects linear model and estimation methods including maximum likelihood estimation and restricted maximum likelihood estimation are detailed. Then, we review the models dealing with excessive zeros. The mixed-effects and mixed-distribution model proposed by Tooze (2002) [66] is based on the repeated measure data with clumping at zero. For zero-inflated count data, Hurdle model by John and Mullahy [44] and zero-inflated Poisson models by Lambert [33] are widely used. We carefully review the mixed-effects and mixed-distribution model by Tooze (2002) and the zero-inflated Poisson regression by Lambert with discussion on their properties and estimation methods. Then we propose the ZIP with correlated random effects structures. In the proposed model, the correlated random effects are considered as in the mixed-effects and mixed-distribution model with semicontinuous data in section 2.2. The estimation of the proposed model was done by approximating integral of likelihood function by adaptive Gaussian quadrature, then optimized by dual quasi Newton method.

The proposed models are validated through simulation studies. For the estimation of parameters, the likelihood function of the model was approximated by adaptive Gaussian quadrature method to carry out a dual quasi Newton method for optimization. Depending on the number of subjects (100 versus 500) and the number of measurements per subject (3 versus 10), we explore the predictability and estimability of parameters of proposed model. Both categorical and continuous explanatory variables were simulated and included in the model. In this simulation study, we simulated data resembles our reading research data for the application. For example, similar parameter values to the application results were used in data simulation. Additional data simulation was also conducted with arbitrary
true parameter values. Our proposed model consists of two parts, so we consider two different cases of correlated random effects and uncorrelated random effects. In each case, 100 simulated trials were conducted. The detailed procedures of simulation study and summary of results were presented in Chapter 3.

The applicability of proposed model to reading research data is demonstrated in chapter 4. The descriptive statistics and the estimated parameters with model specifications are presented. The primary goal is a modeling letter-sound fluency of kindergarten children aged in between 4 and 6 based on the characteristics of individuals. The number of correct letter sound counted of children were collected three times in winter, spring and fall in 2008-2009 academic year. About 30% students was not able to name the sound of at least one lower case alphabet letter, the ZIP regression model with correlated random effects are fitted. We find that age, gender and socioeconomic status are significantly related to the letter sound fluency of children in both parts of the model. The model provides better explanation of data structure and easier interpretations of parameter values, as they are the same as in standard logistic models and Poisson regression models.

At last, we conclude this study with the future work and limitations which are discussed in chapter 5. There are few aspects we would like to extend the model further. Since the gain cumulative knowledge from education, we also would like to extend the model with serial correlation in the random effects accounting time dependency. Another consideration is multilevel model for zero-inflated data. In this dissertation, we only considered two level, person level and within person level. However, as naturally nested data structure is often observed in educational research, one may consider extending the model accounting multilevel data. Previously, Lee et al. (2006) [35] proposed the multilevel zero inflated Poisson regression models with parameter estimation by penalized quasi likelihood method, but further examination of this model remains.
CHAPTER 2
STATISTICAL MODELS AND ESTIMATION

One popular basic approach on the analysis of data in educational research with repeated measures on each student is using mixed-effects linear models, which this chapter begins with. Although there has been many studies dealing with non-ignorable zeros, most of the models developed based on the assumption of a mixture distributions in data. Zeros can be observed in both continuous random variables and discrete random variables, and each type of random variables should be treated differently in the analysis. We first review mixed-distribution mixed-effects models by Tooze (2002) in dealing with mixture distribution of zeros and normal or log-normal data. Then, we take a closer look at models dealing with excessive zeros in count data, zero-inflated Poisson (ZIP) regression by Lambert (1992). We propose the zero-inflated Poisson regression model with correlated random effects. The models and estimation methods are detailed at the end of this section.

2.1 Mixed-effects Linear Models

Based on the concept of distinguishing and identifying individual and population characteristics of unbalanced data, Laird & Ware (1982) extended Harville’s research and defined a family of models for serial measurements, a two-stage model. Following closely on details in Laird & Ware (1982), the structure of two-stage model is explained. In the model, we assume the each response vector, \( y_i, i = 1, \ldots, m \), is multivariate normal with mean \( \mu_i(n_i \times 1) \) and an arbitrary covariance matrix \( \Sigma \).

At Stage 1, the model introduces population parameter, individual effects, and within-person variation. Let \( \alpha \) denote a \( Q \times 1 \) vector of unknown population parameter and \( b_i \) be a \( P \times 1 \) unknown inherent characteristics of the subjects. The vectors, \( \alpha \) and \( b_i \), are linked to \( y_i \) by design matrices, \( X_i (n_i \times Q) \) and \( Z_i (n_i \times P) \), respectively. We regard \( \alpha \) as a fixed effect, \( b_i \) as a random effect.

\[
y_i = X_i \alpha + Z_i b_i + e_i,
\]  

(2.1)

Again, we assume that the error term, \( e_i \), is independent to each other and distributed normally with mean 0 and an \( n_i \times n_i \) positive-definite covariance matrix \( R_i \).

At Stage 2, the between-person variation is explained. The \( b_i \)’s are assumed to be distributed normally with mean 0 and \( P \times P \) positive-definite covariance matrices \( D \) and independent of \( e_i \). The marginal distributions of \( y_i \), are independent normals with mean of \( X_i \alpha \) and covariance matrices \( Z_i D Z_i' + R_i \).
In a way that the two-stage model explains both between and within person variations, the two level data with a nested data structure can be also fitted in the following manner. Raudenbush & Bryk’s (2002) [54] adopted the term, hierarchical linear model, where the nesting unit coefficients are estimated. Let the individuals are naturally nested in certain groups, we define individuals as level one, groups as level two. We are considering $N = \sum_{i=1}^{I} n_i$ samples within $i = 1, \ldots, I$ groups at level 2 and each group consists of $j = 1, \ldots, n_i$ individuals at level 1 with the response variable $y_{ij}$. Based on the description given by Raudenbush & Bryk (2002) [54], the formal structure of the simple two-level model at each level follows:

a) **Level 1**

$$
\begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_{n_i}
\end{bmatrix}_{(n_i \times 1)} =
\begin{bmatrix}
  1 & z_{11}^i & z_{12}^i & \cdots & z_{1p-1}^i \\
  1 & z_{21}^i & z_{22}^i & \cdots & z_{2p-1}^i \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  1 & z_{n_i1}^i & z_{n_i2}^i & \cdots & z_{n_ip-1}^i
\end{bmatrix}_{(n_i \times P)}
\begin{bmatrix}
  \beta_{0i} \\
  \beta_{1i} \\
  \vdots \\
  \beta_{P-1i}
\end{bmatrix}_{(P \times 1)} +
\begin{bmatrix}
  e_{1i} \\
  e_{2i} \\
  \vdots \\
  e_{n_i}
\end{bmatrix}_{(n_i \times 1)}
$$

$$y_i = Z_i \beta_i + e_i,$$

b) **Level 2**

$$
\begin{bmatrix}
  \beta_{0i} \\
  \beta_{1i} \\
  \vdots \\
  \beta_{P-1i}
\end{bmatrix}_{(P \times 1)} =
\begin{bmatrix}
  \omega'_i & 0 & \cdots & 0 \\
  0 & \omega'_i & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & \omega'_i
\end{bmatrix}_{(P \times Q \times P)}
\begin{bmatrix}
  \alpha_0^0 \\
  \alpha_1^0 \\
  \vdots \\
  \alpha_{P-1}^0
\end{bmatrix}_{(Q \times P)} +
\begin{bmatrix}
  b_{1i} \\
  b_{2i} \\
  \vdots \\
  b_{P-1i}
\end{bmatrix}_{(P \times 1)}
$$

$$\beta_i = W_i \alpha + b_i,$$

where $\omega'_i = \begin{bmatrix} 1 & w_{i1}^1 & w_{i2}^1 & \cdots & w_{iQ-1}^1 \end{bmatrix}$.

In the model described above, $z_{ij}^p$ and $\omega'_i$ are $p = 1, \ldots, (P-1)$ level-1 and $q = 1, \ldots, (Q-1)$ level-2 predictors respectively. $\beta_{pi}$ are level-1 coefficients which can be of fixed coefficients, non-randomly varying coefficients or random coefficients. However, at level 2, $\alpha_p^q$ denote
fixed effects. The random effects at each level-1 and level-2 are denoted by \( e_{ij} \) and \( b_{pj} \) respectively. Typically, we assume that \( e_i \) is normally distributed with a mean of zero and variance-covariance matrix of \( \sigma^2 \mathbf{I} \), where \( \mathbf{I} \) is an \( n_j \) by \( n_j \) identity matrix.

Although \( \beta_{pi} \), the level-1 coefficients in the model (2.2), can represent both fixed and random coefficients, when we assume that \( \beta_{pi} \) is representing a random component, simple two-level hierarchical linear models can be written as a form of two-stage models proposed by Liard and Ware (1982) \([31]\).

\[
y_i = Z_i \beta_i + e_i = Z_i (W_i \alpha + b_i) + e_i = Z_i W_i \alpha + Z_i b_i + e_i. \tag{2.4}
\]

Let,

\[
X_i = Z_i W_i = \begin{bmatrix}
\omega_i' & \omega_i' Z_{1i} & \cdots & \omega_i' Z_{pi} \\
\omega_i & \omega_i Z_{1i} & \cdots & \omega_i Z_{pi} \\
\vdots & \vdots & \ddots & \vdots \\
\omega_i & \omega_i Z_{ni} & \cdots & \omega_i Z_{pi}
\end{bmatrix} \in (n_i \times Q \cdot P), \tag{2.5}
\]

then,

\[
y_i = X_i \alpha + Z_i b_i + e_i, \tag{2.6}
\]

as in the model (2.1).

The marginal density of \( y_i \) is,

\[
f(y_i) = \int f(y_i|b_i)f(b_i)db_i \tag{2.7}
\]

where \( f(b_i) \) is a density function of the random effect \( b_i \).

Then, we have the likelihood function, such that

\[
L(\theta|X) = \prod_{i=1}^{n_i} \frac{1}{2 \pi} \frac{1}{2} |V_i(\gamma)|^{-1} \exp(-\frac{1}{2}(y_i - X_i \alpha)' V_i(\gamma)^{-1}(y_i - X_i \alpha)) \tag{2.8}
\]

where \( \theta = (\alpha', \gamma') \), \( V_i = Z_i D Z_i^T + R_i \) and \( \gamma' \) is a vector of all parameters in the covariance matrices.

Then, the log-likelihood, \( \Lambda \), for the data, \( y_1, \ldots, y_m \), is

\[
\Lambda(\theta) = \sum_{i=1}^{m} \frac{n_i}{2} log(2\pi) - \frac{1}{2} \sum_{i=1}^{m} log|V_i| - \frac{1}{2} \sum_{i=1}^{m} (y_i - X_i \alpha)' V_i^{-1}(y_i - X_i \alpha), \tag{2.9}
\]

By taking partial derivatives with respect to each of the parameters, the estimates for each of parameters are,

\[
\hat{\alpha} = \left( \sum_{i=1}^{m} X_i' V_i^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{m} X_i' V_i^{-1} y_i \right), \tag{2.10}
\]
\[ \sigma^2 = \frac{1}{N} \sum_{i=1}^{m} (y_i - X_i \hat{\alpha})' V_i^{-1} (y_i - X_i \hat{\alpha}), \quad (2.11) \]

where \( N = \sum n_i \).

With known covariance matrices, \( R_i, D \), generalized least squares is appropriate method for the estimation. When the \( R_i \) and \( D \) are known and at least to proportionality and \( D^{-1} \) and \( R_i^{-1} \) exist, the solution to the equation \((2.12)\) is the BLUE(best linear unbiased estimates) of \( \alpha \) and the BLUP(best linear unbiased prediction) of \( b_i \), it is commonly known as Henderson’s mixed model equations (1986) [21].

\[
\begin{bmatrix}
X' \hat{R}^{-1} X & X' \hat{R}^{-1} Z \\
Z' \hat{R}^{-1} X & Z' \hat{R}^{-1} X + \hat{D}^{-1}
\end{bmatrix} \begin{bmatrix}
\hat{\alpha} \\
\hat{b}
\end{bmatrix} = \begin{bmatrix}
X' \hat{R}^{-1} y \\
Z' \hat{R}^{-1} y
\end{bmatrix}
\] (2.12)

However, when the covariance matrices are not known, we also need to estimated the covariance matrices. In terms of the estimation methods, Jennrich and Schlucher (1986) [25] detailed Newton-Raphson, Fisher scoring and the generalized EM algorithms with the application to the growth model fitting. They concluded that when the number of estimated parameter is small, any method would be feasible with preference of fast and clean convergence of Newton-Raphson algorithm. When the number of estimated parameters is large, EM algorithm is preferred, because of its low cost per iteration.

For covariance matrix estimators, the likelihood-based methods, maximum likelihood (ML) and restricted maximum likelihood (REML), are discussed in many papers previously (Hartley and Rao (1967) [19], Patterson and Thompson (1971) [49] Harville (1977) [20] Laird and Ware (1982) [31] Jennrich and Schlucher (1986) [25]), two likelihood-based estimators are favored by simulation evidence over other estimators (Swallow and Monahan(1984) [65]). Laird & Ware (1982) [31] put the ML and the REML as leading candidates for the estimators using the EM algorithm with two reasons. The use of either of these estimates unifies both estimation methods and EM algorithm computation with the estimation of other parameters in the model in addition to the capability of accommodating the data missing at random (Rubin (1976) [57], Little (1995) [40]).

When \( V_i \) is not known, it is replaced by maximum likelihood estimates of \( V_i, \hat{V}_i \), however, it is not an unbiased estimator. The restricted maximum likelihood estimation was originally brought by Bartlett in late 30’s, then its usage for variance component estimation was proposed by Patterson and Thompson [49] in early 70’s. Following Harville (1977) [20], the restricted maximum log-likelihood, \( \Lambda_R \), is written as

\[
\Lambda_R = -\frac{N-p}{2} \log(2\pi) + \frac{1}{2} \sum_{i=1}^{m} \log |X_i' X_i| - \frac{1}{2} \sum_{i=1}^{m} \log |V_i| - \frac{1}{2} \sum_{i=1}^{m} [X_i' \hat{V}_i^{-1} X_i] - \frac{1}{2} \sum_{i=1}^{m} (y_i - X_i \hat{\alpha})' \hat{V}_i^{-1} (y_i - X_i \hat{\alpha}),
\] (2.13)

where \( \hat{\alpha} \) is of the form given in 2.10.

As an alternative to the maximum likelihood method, there is a minimum variance quadratic unbiased estimation (MIVQUE) which was proposed by Rao (1971) [53]. However,
Harville (1977) [20] argued that, strictly speaking, the quadratic estimators without any assumptions on parameter spaces are not estimators. So the estimators are not comparable with other likelihood based estimators. He also showed that the ML or REML estimators derived under the normality assumptions are reliable estimators even the any assumptions on the distribution of random effects($b_i$) and error terms ($e_i$) are not specified.

2.2 Mixed-effects and Mixed-distribution Models

The mixed-effects linear models can accommodate data from different settings such as clustered, longitudinal, and repeated measurements. With advantage of the practicality to distinguish two sources of variations, in between groups, and within groups, the mixed-effects linear models are widely used. However, in statistical model fitting practice, one cannot ignore the underlying assumptions on the statistical models. Probably the most common case of assumption violation would be the normality. One may try to transform the data and normalize, but with excessive observation of certain values, many observations with the same value are still preserved in the transformed data. Moreover, when the observations are skewed due to peaked value of zeros, transformation may not be feasible choice of normalizing data set.

On the analysis of data with excessive zeros, there have been many models developed in statistical literature. The most intuitive approach would be treating the data set as a mixture of distributions which consist two-part, a binary part as a occurrence part and a positive part as an intensity part. In dealing with zeros from continuous variable, Olsen and Schafer (2001) [48] introduced the term “semicontinuous”. The term, semicontinuous variable is defined as a continuous variable, often skewed, with point masses at one or more locations (typically 0). Olsen and Schafer first extended the 2-part model to the longitudinal data, then followed by Tooze (2002) [66] who developed mixed-effects and mixed-distribution models. The mixed-effects and mixed-distribution models define a mixture of distributions that takes the general form

$$f(y) = \begin{cases} P(Y = 0), & y = 0, \\ (1 - Pr(Y = 0))h(y), & y > 0, \\ 0, & y < 0. \end{cases}$$

(2.14)

where $h(y)$ is a probability density defined when $y > 0$.

Let $y_{ij}$ be an observation from the $j^{th}, j = 1, \ldots, n$ measurement on the $i^{th}, i = 1, \ldots, m$ subject, where the $y_{ij}$ are all nonnegative. Then, we define $R_{ij}$, which represent the occurrence variable where

$$R_{ij} = \begin{cases} 0, & \text{if } Y_{ij} = 0, \\ 1, & \text{if } Y_{ij} > 0. \end{cases}$$

(2.15)

A conditional probability of $R_{ij}$ is defined by,

$$Pr(R_{ij} = r_{ij}|\theta_1) = \begin{cases} 1 - p_{ij}(\theta_1), & \text{if } r_{ij} = 0, \\ p_{ij}(\theta_1), & \text{if } r_{ij} = 1. \end{cases}$$

(2.16)
where $\theta_1 = [\alpha_1', b_1']$ is a vector of fixed occurrence effects $\alpha_1$, and random unit occurrence effect $b_{1i}$.

The first part of mixed-effects and mixed-distribution model is an occurrence part ($y_{ij} > 0$) and involves logistic model for occurrence so that

$$\text{Logit}(p_{ij}(\theta_1)) = \log \left( \frac{p_{ij}(\theta_1)}{1 - p_{ij}(\theta_1)} \right) = X_{1ij}'\alpha_1 + b_{1i}$$

(2.17)

where $X_{1ij}$ is a vector of covariates for occurrence.

The second part of model is an intensity part and involves the non zero part of the observed values. Let $S_{ij} \equiv [Y_{ij} | R_{ij} = 1]$ be the intensity variable with probability density function $f(s_{ij}|\theta_2)$ for $s_{ij} > 0$ and mean $E(S_{ij}|\theta_2) = \mu_{S_{ij}}(\theta_2)$ where $\theta_2 = [\alpha_2', b_2']$ is a vector of fixed intensity effects $\alpha_2$, and random unit occurrence effect $b_{2i}$. For skewed distribution, a lognormal model for intensity is assumed so that

$$\text{Log}(S_{ij}|\theta_2) \sim N(X_{2ij}'\alpha_1 + b_{2i}, \sigma^2_e)$$

(2.18)

where $X_{2ij}$ is a vector of covariates for intensity. It is assumed that the random effects in 2.17 and 2.18 are jointly normal and possibly correlated such that,

$$\begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} \sim BVN \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_1 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma^2_2 \end{bmatrix} \right).$$

(2.19)

Under this assumption (2.19), the subject-specific mean intensity is

$$E(S_{ij}|\theta_2) = \exp(X_{2ij}'\alpha_1 + b_{2i} + \frac{\sigma^2_e}{2})$$

(2.20)

and the marginal mean intensity is

$$E(S_{ij}|\theta_2) = \exp(X_{2ij}'\alpha_1 + \frac{\sigma^2_e}{2} + \frac{\sigma^2_e}{2})$$

(2.21)

The probability density function of $Y_{ij}$ is

$$f(y_{ij}|\theta) = f(y_{ij}|R_{ij} = 0|\theta)P(R_{ij} = 0|\theta) + f(y_{ij}|R_{ij} = 0|\theta)P(R_{ij} = 0|\theta)$$

$$= \Pr(R_{ij} = 0|\theta_1)\delta_0(y_{ij}) + f(s_{ij}|\theta_2)Pr(R_{ij} = 1|\theta_1)$$

$$= [1 - p_{ij}(\theta_1)]\delta_0(y_{ij}) + f(s_{ij}|\theta_2)p_{ij}(\theta_1),$$

(2.22)

where $\theta = [\theta_1, \theta_2]$ and $\delta_1(y)$ is a Dirac delta function[52] such that

$$\left\{\begin{array}{ll}
\int_{-\infty}^{\infty} \delta_0(y)dy_{ij} = 1, \\
\delta_0(y) = 0, & \text{when } y_{ij} \neq 0.
\end{array}\right.$$  

(2.23)
The conditional expectation of $Y_{ij}$ is:

$$E(Y_{ij}|\theta) = \int f(y_{ij}|\theta)y_{ij}dy = \int f(s_{ij}|\theta_2)p_{ij}(\theta_1)s_{ij}ds = p_{ij}(\theta_1)E(S_{ij}|\theta_2) = p_{ij}(\theta_1)\mu_{s_{ij}}(\theta_2),$$  \hspace{1cm} (2.24)

and the conditional variance is:

$$Var(Y_{ij}|\theta) = E((Y_{ij}-E(Y_{ij}|\theta))^2|\theta)$$

$$= E(Y_{ij}^2|\theta) + E(Y_{ij}|\theta)^2 - 2E(Y_{ij}|\theta)^2$$

$$= E(Y_{ij}^2|\theta) - E(Y_{ij}|\theta)^2$$

$$= p_{ij}(\theta_1)E(S_{ij}^2|\theta_2) - p_{ij}(\theta_1)^2\mu_{s_{ij}}(\theta_2)^2$$

$$= p_{ij}(\theta_1)\text{var}(S_{ij}|\theta) + p_{ij}(\theta_1)\mu_{S_{ij}}(\theta_2)^2 - p_{ij}(\theta_1)^2\mu_{s_{ij}}(\theta_2)^2$$

$$= p_{ij}(\theta_1)\text{var}(S_{ij}|\theta_2) + p_{ij}(\theta_1)(1-p_{ij}(\theta_1))\mu_{s_{ij}}(\theta_2)^2.$$  \hspace{1cm} (2.25)

The contribution to the likelihood for the $i^{th}$, $i = 1, \ldots, m$ subject is,

$$L_i(\alpha_1, \alpha_2, \sigma_1, \sigma_2, \sigma_e, \rho|y_{i1}, \ldots, y_{in_i})$$

$$= \frac{\int_{b_{1i}} \int_{b_{2i}} f(y_{i1}, \ldots, y_{in_i}|b_{1i}, b_{2i}, \alpha_1, \alpha_2, \sigma_1, \sigma_2, \sigma_e, \rho)db_{1i}db_{2i}}{\int_{b_{1i}} \int_{b_{2i}} f(y_{i1}, \ldots, y_{in_i}|\theta_1, \theta_2, \sigma_1, \sigma_2, \sigma_e, \rho)\prod_{j=1}^{n_i} f(y_{ij}|\alpha_1, \alpha_2, b_{1i}, b_{2i})f(b_{1i}, b_{2i}|\sigma_1, \sigma_2, \sigma_e, \rho)db_{1i}db_{2i}}$$  \hspace{1cm} (2.26)

Then, the likelihood function of correlated mixed-effects and mixed-distribution models is

$$L(\alpha_1, \alpha_2, \sigma_1, \sigma_2, \sigma_e, \rho|Y) = \prod_{i=1}^{m} \int_{b_{1i}} \int_{b_{2i}} f(y_{i1}, \ldots, y_{in_i}|b_{1i}, b_{2i}, \alpha_1, \alpha_2, \sigma_1, \sigma_2, \sigma_e, \rho)db_{1i}db_{2i}$$

$$= \prod_{i=1}^{m} \int_{b_{1i}} \int_{b_{2i}} \prod_{j=1}^{n_i} \int \prod_{j=1}^{n_i} [1-p_{ij}(\alpha_1, b_{1i})]^{1-r_{ij}}[p_{ij}(\alpha_1, b_{1i})]^{r_{ij}}$$

$$\times f(s_{ij}|\alpha_2, b_{2i})f(b_{1i}, b_{2i}|\sigma_1, \sigma_2, \sigma_e, \rho)db_{1i}db_{2i}.$$  \hspace{1cm} (2.27)

With the assumptions that $b_{1i}$ and $b_{2i}$ are independent, ($\rho = 0$), the likelihood function of uncorrelated mixed-effects and mixed-distribution models can be factored into two parts such that

$$L(\alpha_1, \alpha_2, \sigma_1, \sigma_2, \sigma_e|Y) = \prod_{i=1}^{m} \int_{b_{1i}} \prod_{j=1}^{n_i} [1-p_{ij}(\theta_1)]^{1-r_{ij}}[p_{ij}(\theta_1)]^{r_{ij}} f(b_{1i}|\sigma_1)db_{1i}$$

$$\times \prod_{i=1}^{m} \int_{b_{2i}} \prod_{j=1}^{n_i} \int f(s_{ij}|\theta)f(b_{2i}|\sigma_2, \sigma_e)db_{2i}$$  \hspace{1cm} (2.28)
In the paper by Tooze et al. [66], the model fitting was done using SAS PROC NLMIXED procedure (SAS Institute, Cary, NC, Version 9) [24]. Default estimation methods utilize an approximation of likelihood by adaptive Gaussian quadrature (Pinheiro and Bates (1995) [50]), and quasi-Newton optimization (Davidon, 1991 [9]).

Once the model is fitted, standard regression diagnostics can be used for the assessment of the goodness of fit of the model. Also, one can calculate the residuals for the intensity variable, \( S_{ij} \), which are given by \( \ln(s_{ij}) - X_{2ij}' \alpha_1 + b_{2i} \). Construction of quantile-quantile plot of residuals can be used to check the normality assumption and constant variances. Also, the quantile-quantile plots of the random effects for both occurrence and intensity parts, \( \hat{b}_{1i} \) and \( \hat{b}_{2i} \), can be constructed to assess if each of the estimates follow normal distribution as the model assumes.

### 2.3 Zero-inflated Poisson Models

In some situation, zero-inflated data occur in discrete random variables, such as in count data. In this case, one of popular approaches is using zero-inflated count models. Lambert (1992) [33] approached the zero-inflated count data with the mixtures of two distributions, zeros and Poisson distribution. In zero-inflated Poisson regression (ZIP), \( Y = (Y_1, \ldots, Y_m)^T \) are independent and

\[
Y_i \sim \begin{cases} 
0, & \text{with probability } 1 - p_i, \\
\text{Poisson}(\lambda_i), & \text{with probability } p_i.
\end{cases}
\]  

so that

\[
Pr(Y_i = y_i) = \begin{cases} 
(1 - p_i) + p_i e^{-\lambda_i}, & y_i = 0, \\
p_i \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, & y_i > 0, \, 0 \leq p_i \leq 1.
\end{cases}
\]  

Then, we have,

\[
\text{Logit}(p_i) = \log \left( \frac{p_i}{1 - p_i} \right) = X_{1i} \alpha_1 \text{ and } \]  
\[
\text{Log}(\lambda_i) = X_{2i} \alpha_2
\]

for covariate vectors \( X_{1i} \) and \( X_{2i} \). In the model, \( p_i \) is considered to be a mixing parameter for the mixture of two distributions, binary and Poisson processes. Both of \( p_i \) and \( \lambda_i \) depend on the characteristics of subject, but the covariates \( X_{1i} \) and \( X_{2i} \), which influence the parameters, \( p_i \) and \( \lambda_i \), are necessarily the same in the model. In the model, one may interpret the parameters \( \alpha_1 \) and \( \alpha_2 \) as in the standard logistic and Poisson regression models. The difference between \( \alpha_1 \) and \( \alpha_2 \) associated with possibly different covariates, \( X_{1i} \) and \( X_{2i} \), may be the ultimate interest.
The conditional expectation and conditional variance of \( y_i \) are given by

\[
E(Y_i|\alpha) = \sum_y Pr(y_i|\alpha_1, \alpha_2)y_i = \sum_{y_i>0} Pr(y_i|\alpha_1, \alpha_2)y_i
\]

\[
= \sum_{y_i>0} p_i \frac{e(-\lambda_i)\lambda_i^{y_i}}{y_i!} = p_i \sum_{y_i>0} \frac{e(-\lambda_i)\lambda_i^{y_i}}{y_i!} = p_i \lambda_i,
\]

(2.33)

\[
Var(Y_i|\alpha) = E((Y_i-E(Y_i|\alpha))^2|\alpha)
\]

\[
= E(Y_i^2|\alpha) + E(Y_i|\alpha)^2 - 2E(Y_i|\alpha)^2
\]

\[
= E(Y_i^2|\alpha) - E(Y_i|\alpha)^2
\]

\[
= \sum_y Pr(y_i|\alpha)y_i^2 - (\sum_y Pr(y_i|\alpha)y_i)^2
\]

\[
= \sum_{y_i>0} Pr(y_i|\alpha_1)\alpha_1 y_i - (\sum_{y_i>0} Pr(y_i|\alpha_1)\alpha_1 y_i)^2
\]

\[
= \sum_{y_i>0} p_i \frac{e(-\lambda_i)\lambda_i^{y_i}}{y_i!} y_i^2 - (\sum_{y_i>0} p_i \frac{e(-\lambda_i)\lambda_i^{y_i}}{y_i!} y_i)^2
\]

\[
= p_i(\lambda_i + \lambda_i^2) - p_i^2 \lambda_i^2 = \lambda_i p_i (1 + \lambda_i (1 - p_i)),
\]

(2.34)

where \( \alpha = [\alpha_1, \alpha_2] \).

With the probability mass function given in (2.44), we have the likelihood function of,

\[
L(\alpha|Y) = \prod_{i=1}^m I_{y_i=0} \left( 1 - p_i + p_i e^{-\lambda_i} \right) + I_{y_i>0} \left( \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \right)
\]

\[
= \prod_{i=1}^m I_{y_i=0} \left( \frac{1}{1 + e^{X_i_1\alpha_1}} + \frac{e^{X_i_1\alpha_1}}{1 + e^{X_i_1\alpha_1}} e^{X_2\alpha_2} \right)
\]

\[
+ I_{y_i>0} \left( \frac{e^{X_i_1\alpha_1} \exp(-e^{X_2\alpha_2} e^{X_2\alpha_2 y_i})}{1 + e^{X_i_1\alpha_1}} \right)
\]

\[
= \prod_{i=1}^m I_{y_i=0} \left( \frac{1 + \exp(X_i_1\alpha_1 - e^{X_2\alpha_2})}{1 + e^{X_i_1\alpha_1}} \right)
\]

\[
+ I_{y_i>0} \left( \frac{\exp(X_i_1\alpha_1 - e^{X_2\alpha_2} + y_i X_2\alpha_2)}{(1 + e^{X_i_1\alpha_1})y_i!} \right)
\]

(2.35)

where \( I \) is an indicator function.

Then, the log-likelihood function is

\[
\Lambda(\alpha_1, \alpha_2|Y) = \sum_{y_i=0} \log(e^{X_i_1\alpha_1} + \exp(-e^{X_2\alpha_2})) + \sum_{y_i>0} (y_i X_2\alpha_2 - e^{X_2\alpha_2})
\]

\[
- \sum_{i=1}^n \log(1 + e^{X_i_1\alpha_1}) - \sum_{y_i>0} \log(y_i!)
\]

(2.36)
This model would be the best fit when a distribution of population is considered as a mixture of one set of only zero response and the other set of zero and other responses. In other words, the model does not separate the observed values into two subsets of data, zeros and non-zeros. For example, the random variable \( Y_i \) can be redefined as \( Y_i' = 0 \) from \( Y_i = 0 \), and \( Y_i' = 1 \) when \( Y_i \) is from the Poisson distribution. Suppose that we can distinguish \( Y_i' \) from \( Y_i \), the log-likelihood would be,

\[
\Lambda(\alpha_1, \alpha_2 | Y) = \sum_{i=1}^{n} \log(f(Y_i' | \alpha_2)) + \sum_{i=1}^{n} \log(f(Y_i | Y_i', \alpha_1))
\]

\[
= \sum_{i=1}^{n} ((1 - Y_i')X_2i\alpha_2 - \log(1 + e^{X_2i\alpha_2})) + \sum_{i=1}^{n} Y_i'(Y_iX_1i\alpha_1 - e^{X_1i\alpha_1})
\]

\[-\sum_{i=1}^{n} Y_i' \log(y_i!) \]

\[
= L(\alpha_2) + L(\alpha_1) - \sum_{i=1}^{n} Y_i' \log(y_i!) \tag{2.37}
\]

where \( X_{1i} \) and \( X_{2i} \) are the \( i^{th} \) row of \( X_1 \) and \( X_2 \). It is worth note that the log-likelihood function of zero-inflated Poisson model can be written as summation of two log-likelihood functions respect to each of parameters, \( \alpha_1 \) and \( \alpha_2 \). Since each of the log-likelihood functions can be maximized separately, the maximization of log-likelihood function becomes easy in this case. The parameter estimates can be obtained by iterative maximization of the log-likelihood function using the EM algorithm by Dempster et al. (1977) [10]. Lambert (1992) [33] showed that without heavily relying on initial guess of \( \alpha \), the EM algorithm converged. Although there are other algorithms, such as Newton-Raphson method, which are faster than the EM algorithm, the Newton-Raphson algorithm failed to converge in this case. Also, the EM algorithm is simpler to program in general.

In many application, the parameters, \( \lambda_i \) and \( p_i \) are often related in the ZIP regression. In other words, the mean of Poisson distribution depends on the number of zeros from the other distribution of being zero. With the shape parameter \( \tau \), the same covariates were used in \( \lambda_i \) and \( p_i \), which are defined by

\[
\text{logit}(p_i) = -\tau X_{2i}\alpha_2, \tag{2.38}
\]

\[
\log(\lambda_i) = X_{2i}\alpha_2. \tag{2.39}
\]

It is worth to note that when the shape parameter \( \tau > 0 \), the zero-inflation decreases and as \( \tau \to 0 \), the zero-inflation increases. It is obvious, we have less parameters in the model with the shape parameter \( \tau \).
The likelihood function is

\[ L(\alpha_1, \tau | Y) = \prod_{i=1}^{m} I_{Y_i=0} \left( 1 - p_i + p_i e^{-\lambda_i} \right) + I_{Y_i>0} \left( p_i e^{-\lambda_i} \sum_{Y_i} \frac{y_i!}{y_i!} \right) \]

Then, the log-likelihood function is,

\[ \Lambda(\alpha_1, \tau) = \sum_{Y_i=0} \log \left( 1 - p_i + p_i e^{-\lambda_i} \right) + \sum_{Y_i>0} \left( p_i e^{-\lambda_i} \sum_{Y_i} \frac{y_i!}{y_i!} \right) \]

\[ - \sum_{Y_i=0} \log(1 + e^{-\tau X_i}) - \sum_{Y_i>0} \log(y_i!). \] (2.41)

Without any prior information on \( \tau \), the EM algorithm is not useful method for parameter estimation. Although the Newton-Raphson algorithm failed to converge for the model without the shape parameter, the algorithm converged for the model with the shape parameter in the simulation example [33]. With large number of sample size, the maximum likelihood estimates of \( \alpha_1 \) and \( \alpha_2 \) for ZIP regression and \( \tau \) for ZIP regression with \( \tau \) are approximately normal with means of \( \alpha_1 \), \( \alpha_2 \), and \( \tau \), respectively. Also, variances can be obtained from the inverse observed information matrices. The asymptotic distribution of such information matrices are detailed in Lambert (1992) [33].

The assessment of goodness-of-fit of the model can be done by using likelihood ratio test. The twice of difference of log-likelihood under the null hypothesis and alternative hypothesis is asymptotically chi-squared [46]. Therefore, for nested zero-inflated Poisson models, the models can be compared by the log likelihood ratio test. However, when the zero-inflated Poisson models are compared to other non-nested models for count data, such as Poisson regression, negative Binomial regression, or Hurdle regression [44], the use of a Vuong statistic (1989) [67] are suggested. Vuong statistic is a model selection test criteria using the Kullback-Leibler information criterion [28] which is based on log likelihood ratio test. The Vuong statistics for testing hypothesis, \( E(m_i = 0) \), is given by,

\[ V = \frac{\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} m_i \right)}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (m_i - \bar{m})^2}} \] (2.42)

where \( m_i = \log \left( \frac{P_1(Y_i|X_i)}{P_2(Y_i|X_i)} \right) \) and \( P_n(Y_n|X_n) \) is the predicted probability of observed count for case \( i \) from model \( n \) [41].
2.4 Proposed Model

In this section, we introduce a pair of possibly correlated random effects to the zero-inflated Poisson models to accommodate within-subject correlation and between subject heterogeneity. After specifying the models, the estimation methods and model diagnostic and comparison methods follow.

2.4.1 The Models

We hold the same assumption as in the 2.2, except the (2.18), where we introduce the Poisson distribution in the model for intensity part of the two-parts, mixed-distribution, model. Let $y_{ij}$ be an observation from the $j^{th}, j = 1, \ldots, n_i$ measurement on the $i^{th}, i = 1, \ldots, m$ subject, where the $y_{ij}$ are all non negative,

$$Y_{ij} \sim \begin{cases} 0, & \text{with probability } 1 - p_{ij}, \\ \text{Poisson}(\lambda_{ij}), & \text{with probability } p_{ij}. \end{cases} \quad (2.43)$$

so that

$$Pr(Y_{ij} = y_{ij}) = \begin{cases} (1 - p_{ij}) + p_{ij}e^{-(\lambda_{ij})}, & y_{ij} = 0, \\ p_{ij}\frac{e^{(-\lambda_{ij})}\lambda_{ij}^{y_{ij}}}{y_{ij}!}, & y_{ij} > 0, 0 \leq p_{ij} \leq 1. \end{cases} \quad (2.44)$$

Then, we have,

$$Logit(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = X_{1ij}\alpha_1 + b_{1i} \quad \text{and} \quad (2.45)$$

$$Log(\lambda_{ij}) = X_{2ij}\alpha_2 + b_{2i} \quad (2.46)$$

for covariate vector $X_{1ij}$ and $X_{2ij}$ and random effects, $b_{1i}$ and $b_{2i}$.

The zero-inflated Poisson model with random effects is based on subject-specific models. The distribution of response variable is modeled in terms of covariates of each subject, the variances of random effects are specified to each of the subject. In the model, $\alpha_1$, and $\alpha_2$ measure the change in the conditional logit of $p_{ij}$ and the conditional log of $\lambda_{ij}$ with the covariates $X_{1ij}$ and $X_{2ij}$ for measurements in each subject described by $b_{1i}$ and $b_{2i}$, respectively.

We also assume that the random effects, $b_{1i}$ and $b_{2i}$, are drawn from normal distributions and possibly correlated as in (2.19). The normal distribution is chosen because it is the only well-established multivariate distribution and the choice of correlation structure is also versatile. Although other models such as Poisson-Gamma models can be made for mathematical simplicity, normally distributed random effects are more natural choice and more practical [36].

Similarly to (2.33), we have conditional expectation and conditional variance of $y_{ij}$, which are given by,

$$E(Y_{ij}|\alpha) = p_{ij}\lambda_{ij}, \quad (2.47)$$

$$Var(Y_{ij}|\alpha) = \lambda_{ij}p_{ij}(1 + \lambda_{ij}(1 - p_{ij})), \quad (2.48)$$

19
where $\mathbf{\alpha} = [\alpha_1, \alpha_2]$.

The marginal parameters on each subject over multiple measurements can be obtained from the subject specific parameters by contribution to the likelihood for the $i^{th}$, $i = 1, \ldots, m$, subject. It is given by,

$$L_i(\alpha_1, \alpha_2, \sigma_1, \sigma_2, \rho | y_{i1}, \ldots, y_{in_i})$$

$$= \int_{b_{1i}} \int_{b_{2i}} f(y_{i1}, \ldots, y_{in_i} | b_{1i}, b_{2i}, \alpha_1, \alpha_2, \sigma_1, \sigma_2, \rho) db_{1i} db_{2i}$$

$$= \int_{b_{1i}} \int_{b_{2i}} f(y_{i1}, \ldots, y_{in_i} | b_{1i}, b_{2i}, \alpha_1, \alpha_2, \sigma_1, \sigma_2, \rho) \times f(b_{1i}, b_{2i} | \alpha_1, \alpha_2, \sigma_1, \sigma_2, \rho) db_{1i} db_{2i}$$

$$= \int_{b_{1i}} \int_{b_{2i}} \prod_{j=1}^{n_i} f(y_{ij} | \alpha_1, \alpha_2, b_{1i}, b_{2i}) f(b_{1i}, b_{2i} | \sigma_1, \sigma_2, \rho) db_{1i} db_{2i}$$

$$= \int_{b_{1i}} \int_{b_{2i}} \prod_{j=1}^{n_i} ((1 - p_{ij}) + p_{ij} e^{(-\lambda_{ij})}) (1 - r_{ij}) \left( p_{ij} \frac{e^{(-\lambda_{ij})} \lambda_{ij}^{y_{ij}}}{y_{ij}!} \right) r_{ij}$$

$$\times f(b_{1i}, b_{2i} | \sigma_1, \sigma_2, \rho) db_{1i} db_{2i}$$

(2.49)

where $r_{ij}$ is defined in equations (2.14) and (2.15).

Then, the likelihood function is,

$$L(\alpha_1, \alpha_2, \sigma_1, \sigma_2, \rho | \mathbf{Y})$$

$$= \prod_{i=1}^{m} L_i(\alpha_1, \alpha_2, \sigma_1, \sigma_2, \rho | y_{i1}, \ldots, y_{in_i})$$

$$= \prod_{i=1}^{m} \int_{b_{1i}} \int_{b_{2i}} \prod_{j=1}^{n_i} ((1 - p_{ij}) + p_{ij} e^{(-\lambda_{ij})}) (1 - r_{ij}) \left( p_{ij} \frac{e^{(-\lambda_{ij})} \lambda_{ij}^{y_{ij}}}{y_{ij}!} \right) r_{ij}$$

$$\times f(b_{1i}, b_{2i} | \sigma_1, \sigma_2, \rho) db_{1i} db_{2i}$$

$$= \prod_{i=1}^{m} \int_{b_{1i}} \int_{b_{2i}} \prod_{j=1}^{n_i} f(r_{ij} | b_{1i}, b_{2i}) f(r_{ij} | y_{ij} | b_{2i}) f(b_{1i}, b_{2i} | \sigma_1, \sigma_2, \rho) db_{1i} db_{2i}$$

(2.50)

Again, the likelihood function can be factored into two parts, when there is no correlation between random effects, $b_{1i}$ and $b_{2i}$.

As in section 2.3, we can also consider the case when $p_{ij}$ and $\lambda_{ij}$ are related by adding the shape parameter $\tau$ in the model. The model can be obtained by replacing the equation (2.45) by

$$Logit(p_{ij}) = \log \left( \frac{p_{ij}}{1 - p_{ij}} \right)$$

$$= -\tau X_{2ij} \alpha_2 + b_{2i}$$

(2.51)
When we introduce the shape parameter $\tau$ in the model, the random effects term is not restricted, but the same covariates are shared in both logistic and Poisson parts of the model. Unlike in the model (2.45) where the covariates from logit part and Poisson part are not necessarily the same, the same covariates are shared in the model with the shape parameter $\tau$ as in (2.51). However, one can take advantage of fitting more parsimonious model with the shape parameter $\tau$ than the model without the shape parameter. The interpretation of the shape parameter in the model is the same as in the zero-inflate Poisson model without random effects described in section 2.3.

2.4.2 The Estimation

As Hall (2000) \cite{18} adopted the EM algorithm to obtain maximum likelihood estimates, it is a natural approach to use such method as it was already used in Lambert’s \cite{33} original model which include fixed effects only. In the EM algorithm it involves maximization of log likelihood function with respect to the parameters in the model. However, as the integration with respect to the parameter on the random effect is complicated, it was approximated using Gaussian quadrature.

However, with a pair of random effects in the model, different estimation methods was used by Yau and Lee (2001) \cite{71}, and Wang et al. (2002) \cite{26}. In both studies, the penalized quasi-likelihood approach is used to fit the model. The log-likelihood function is divided into two parts, one part of conditional log-likelihood on random effects ($\Lambda_1$), and the other part of joint density function of random effects ($\Lambda_2$), so that the log-likelihood function ($\Lambda = \Lambda_1 + \Lambda_2$) can be viewed as a penalized log-likelihood function. The variance component are obtained using restricted maximum likelihood estimates.

With the advantage of a separate parametrization, one would be able to fit the model efficiently with the penalized log-likelihood. Yau and Lee \cite{71} adopted the Newton-Raphson algorithm, whereas Wang et al. \cite{26} did the EM algorithm to ensure the convergence and stability. However, in model fitting process, the estimators fitted using the penalized quasi-likelihood approach are likely to be biased and inconsistent for highly non-normal responses with large variances of random effects \cite{7}, \cite{43}. Thus, the models with binary outcomes as in the occurrence part of the zero-inflated Poisson model, using the penalized quasi-likelihood approach in model fitting may not be adequate.

Unlike the penalized quasi-likelihood approach which estimate the parameters separately for each part of the model, Hur et al. (2002) \cite{22} fitted the zero-inflated Poisson models with uncorrelated random effects simultaneously. For quicker convergence, Newton-Raphson method was used with numerically approximated integrals of log-likelihood. Although Min and Agresti \cite{43} favored Hurdle model over zero-inflated Poisson models, the same methods were used to obtain maximum likelihood estimates of the model. Allowing correlated random effects in the model, with the parametric approach, the adaptive Gaussian quadrature for the integral approximation and Fisher scoring methods for maximization were used to obtain the maximum likelihood estimates.

In all models reviewed previously assumed that the random effects are drawn from a normal distribution. To relax this assumption, Min and Agresti \cite{43}, and Lam et al. \cite{32} proposed the use of nonparametric approach in model fitting using nonparametric maximum likelihood estimation methods \cite{38}, \cite{39} and the sieve maximum likelihood estimation.
method, respectively. In later models developed by Gschlößl and Gzado (2008) [45], the model was fitted in Bayesian framework by proper prior distribution for the spatial random effects.

To fit the zero-inflated Poisson model with random effects, first one should obtain the marginal likelihood estimation. However, marginal likelihoods are generally difficult to compute and it is true it this case as well. Therefore, among many numerical approximation of integral, such as adaptive Gaussian quadrature, the Monte Carlo EM algorithm, Markov chain Monte Carlo, penalized quasi likelihood, and Laplace approximation, the first three methods are known for their convergence to the maximum likelihood estimate [43].

The inferences about parameters were based on the marginal likelihood function, which is the integrated likelihood function over the random effects. The integral is not analytically intractable, therefore numerical approximation is carried out using adaptive Gaussian quadrature [50]. Gaussian quadrature is adequate to numerical evaluation of integrals against probability measure [34]. Moreover, in recent simulation study, the adaptive Gaussian quadrature approximation provide efficient and accurate estimation in multi-level mixed models [51]. Also, as the number of quadrature point increases, desired accuracy of approximation was obtained. The approximation is done by replacing integral by the summation of weighted conditional distribution of the data given the random effects over the specified quadrature points, denoted by \( q_1 = 1, \ldots, B_{q_1}, q_2 = 1, \ldots, B_{q_2} \), at chosen abscissas \( b_{1iq} \) and \( b_{2iq} \).

It is given by,

\[
L_{i}^{GQ}(\alpha_1, \alpha_2, \sigma_1, \sigma_2, \rho|y_{1i}, \ldots, y_{ni}) = h_{1i} \\
= \int_{b_{1i}}^{B_{q_1}} \int_{b_{2i}}^{B_{q_2}} \prod_{j=1}^{n_i} f(y_{ij} | \alpha_1, \alpha_2, b_{1j}, b_{2j}) f(b_{1j}, b_{2j} | \sigma_1, \sigma_2, \rho) \, db_{1j} \, db_{2j} \\
\approx \sum_{q_1}^{B_{q_1}} \sum_{q_2}^{B_{q_2}} \prod_{j=1}^{n_i} f(y_{ij} | \alpha_1, \alpha_2, b_{1iq}, b_{2iq}) A(B_{q_1}, B_{q_2}) \\
= \sum_{q_1}^{B_{q_1}} \sum_{q_2}^{B_{q_2}} l(y_{ij} | B_{q_1}, B_{q_2}) A(B_{q_1}, B_{q_2})
\]  

(2.52)

where \( A(B_{q_1}, B_{q_2}) \) is the standard Gauss-Hermite weights. Then the approximated full log-likelihood is,

\[
\Lambda = \sum_{i=1}^{m} h_{1i}.
\]

(2.54)

The first and second derivatives of \( \Lambda \) with respect to arbitrary parameter vector \( \theta \) are

\[
\frac{\partial \Lambda}{\partial \theta} = \sum_{i=1}^{m} \frac{1}{h_{1i}} \frac{\partial h_{1i}}{\partial \theta}
\]

(2.55)

and

\[
\frac{\partial^{2} \Lambda}{\partial \theta \partial \theta'} = \sum_{i=1}^{m} \frac{1}{h_{1i}^2} \left[ h_{1i} \frac{\partial^2 h_{2i}}{\partial \theta \partial \theta'} - h_{2i} \frac{\partial h_{2i}}{\partial \theta} \right],
\]

(2.56)
where the first and second derivatives of $h_i$ with respect to arbitrary parameter vector $\theta$ are

$$\frac{\partial h_{1i}}{\partial \theta} = \sum_{q_1} \sum_{q_2} \frac{\partial \log l(y_{ij} | B_{q_1}, B_{q_2})}{\partial \theta} l(y_{ij} | B_{q_1}, B_{q_2}) A(B_{q_1}, B_{q_2})$$

(2.57)

and

$$\frac{\partial^2 h_{1i}}{\partial \theta \partial \theta'} = \frac{\partial h_{2i}}{\partial \theta}$$

$$= \sum_{q_1} \sum_{q_2} \left[ \frac{\partial \log l(y_{ij} | B_{q_1}, B_{q_2})}{\partial \theta} \frac{\partial l(y_{ij} | B_{q_1}, B_{q_2})}{\partial \theta'} + \frac{\partial^2 \log l(y_{ij} | B_{q_1}, B_{q_2})}{\partial \theta \partial \theta'} l(y_{ij} | B_{q_1}, B_{q_2}) \right]$$

$$\times A(B_{q_1}, B_{q_2})$$

$$= \sum_{q_1} \sum_{q_2} \left[ \left( \frac{\partial \log l(y_{ij} | B_{q_1}, B_{q_2})}{\partial \theta} \right) \left( \frac{\partial \log l(y_{ij} | B_{q_1}, B_{q_2})}{\partial \theta'} \right) + \frac{\partial^2 \log l(y_{ij} | B_{q_1}, B_{q_2})}{\partial \theta \partial \theta'} \right]$$

$$\times l(y_{ij} | B_{q_1}, B_{q_2}) A(B_{q_1}, B_{q_2}).$$

(2.58)

The solutions of the full log-likelihood function given in (2.54) can be obtained iteratively using the dual quasi-Newton algorithm. The dual quasi-Newton method (Dennies (1977)) uses the gradient instead of computing second-order derivatives as they are approximated (2.57). Among many optimization techniques, the dual quasi-Newton method is preferred because it balances between the speed and stability (24) of the parameter estimation. This approximation is implemented in the PROC NLMIXED procedure (SAS Institute, Cary, NC, Version 9.3) as a default option (24).

There is one more option which a user can specify in terms of integration approximation, the first-order method by Beal and Sheiner (1982, 1988) [4], [5] and Sheiner and Beal (1985) [59]. Users also can specify other optimization techniques, which are trust region method, Newton-Raphson method with line search, Newton-Raphson method with ridging, double-dogleg method, conjugate gradient methods, and Nelder-Mean simplex method. The details on the approximation and optimization techniques can be found from the SAS users’ manual (24).

2.4.3 Model Comparisons and Selection Criteria

Although there are many tools and methods for model selection, we will examine four criteria, Akaike’s Information Criterion (AIC), corrected Akaike’s Information Criterion (AICC), consistent Akaike’s Information Criterion (CAIC), and Schwartz’s Bayesian Criterion (SBC) under the principle of choosing a model as parsimony as possible. First, we take a look at Akaike’s Information Criterion (AIC [1]) proposed by Akaike for model selection. It is defined by,

$$AIC(k) = D(\Lambda) + 2k,$$

(2.59)

where $k$ is the number of independently adjusted parameters within the model and $D(\Lambda)$ is twice the negative log-likelihood, deviance. This criteria is basically deviance plus the
second term, which puts the penalty on the number of parameters in a model. Later, Hurvich and Tsai (1993)\textsuperscript{[23]} proposed the corrected version of AIC, which can adjust a small sample bias by adding the additional penalty term,

\[ AICC(k) = D(\Lambda) + 2(k + 1)(k + 2)/(N - k - 2). \]  

Alternatively, one can use Schwartz’s Bayesian Criterion (SBC, also known as Bayesian information criterion, BIC)\textsuperscript{[58]}, which is developed for the same purpose as of AIC, a mathematical formulation of the principle of parsimony in model building. The procedure differs from AIC only in the penalty of dimension like in other criteria. It is given by,

\[ SBC(k) = D(\Lambda) + \log(N)k, \]  

where \( N \) is the number of observations.

Bozdogan (1987) proposed consistent AIC similar to the SBC by adding one term,

\[ CAIC(k) = D(\Lambda) + (\log(N) + 1)k, \]  

Although all methods aims for selecting parsimonious model, since \( \log(n) > 2 \) for \( n > 10 \), the SBC put more penalty on the number of parameters in a model, the model selected using AIC and SBC may be markedly different from each other for large \( n \). Gurka (2006) \textsuperscript{[17]} pointed out that one should discuss the model selection criteria in terms of its large-sample notion of efficiency and consistency. We consider the efficiency when the model is estimated in finite dimension when the true model is in infinite dimension. The true model in finite dimension is chosen with probability of one by consistency criteria. Theoretically speaking, AIC and AICC are efficient and SBC and CAIC is consistent in terms of model selection.

A researcher should be the one who picks and uses the correct criterion on model selection depends on the field of the study. Gurka (2006) \textsuperscript{[17]} performed a simulation study of selecting a correct model based on four different criteria mentioned above. In his study, it is concluded that the SBC performed slightly better although all of the criteria resulted in more than 90\% of accuracy rates, when estimating both fixed and random effect of the mixed model. These criteria are useful in comparison of different models whether or not the one is nested to another, but they do not perform any test of the comparison.

As well as the model selection criteria described above, one can also not ignore well-known statistics for model comparison, especially when one model is nested to the other model. In application, we will fit two different cases, which are the model with correlated random effects and the model without. When we assume correlated random effect, we will have one more parameter to estimate, which is the correlation between two random effects, \( \rho \). Although we can compare the model selection criterion, the most intuitive approach would be the log likelihood ratio statistics by Nelder and Wedderburn (1972) \textsuperscript{[46]}. It is known the sampling distribution of the deviance \( (D_{LR}) \) is,

\[ D_{LR} = 2 \left[ l(b_{model1}; y) - l(b_{model2}; y) \right] \sim \chi^2_{(df=p_1-p_2)}, \]  

where \( b \) is the maximum likelihood estimator of the parameters, and \( p_1 \) and \( p_2 \) are the number of parameters in the models. In the application section of this study, we will
compare AIC, AICC, BIC and log-likelihood test. We will compare the nested models in probabilistic approach by specifically using log likelihood ratio statistics.
CHAPTER 3

SIMULATION STUDY

This chapter presents the simulation study of the models from the previous chapter. A series of simulation was conducted to examine the impact of different parameter values, the number of subject, and the number of measurement per subject in different model settings. Both correlated and non-correlated random effects of models were considered. All the procedures from data generation to the model fitting was done by using SAS (SAS Institute, Cary, NC, Version 9.3). Simulation design and estimation methods are detailed and the chapter ends with the simulation results.

3.1 Simulation Design

The simulation was consist of the following steps.

**Step 1** Generate \( y_{ij} \) following (3.1) and define \( r_{ij} \).
- Calculate \( p_{ij} \) and \( \lambda_{ij} \)
- Generate \( y_{ij} \sim Bernoulli(p_{ij}) \)
  - If \( y'_{ij} = 1 \), then \( y_{ij} \sim Poisson(\lambda_{ij}) \), else \( y'_{ij} = y_{ij} = 0 \)
  - Define \( r_{ij} \); If \( y_{ij} = 0 \) then \( r_{ij} = 0 \), else \( r_{ij} = 1 \)

**Step 2** Fit Binomial regression model using \( r_{ij} \) and Poisson regression model using \( y_{ij} > 0 \).

**Step 3** Using the parameter estimates from Step 3, Fit the final model.

**Step 4** Repeat Step 1 through Step 3 100 times, then calculate mean and standard deviation of each parameter estimates.

The simulated data contain one dichotomized and one continuous explanatory variables. Since, the proposed model was motivated by the applications to reading research data, the similar setting was considered. A covariate, \( x_{1i} \), is following Bernoulli distribution with \( p = 0.2 \), binary outcome represent categorical variable such as gender and race. The continuous variable, \( t_{ij} \) is drawn from uniform distribution on interval \([4,6]\), representing time at the measurement or the ages of children at the measurement time. To simplify the simulation study, we assume that design matrices \( X_1 \) and \( X_2 \) are identical, and we only considered the example when the number of measurement of each subject is equal, i.e. \( n_i = n \).
With specified design and true parameter values, from the proposed model, the observations were generated such that,

\[
Pr(Y_{ij} = y_{ij}) = \begin{cases} 
(1 - p_{ij}) + p_{ij}e^{(-\lambda_{ij})}, & y_i = 0, \\
p_{ij}e^{(-\lambda_{ij})}y_{ij}^{y_{ij}}, & y_{ij} > 0, 0 \leq p_{ij} \leq 1. 
\end{cases}
\]

where \(p_{ij}\) and \(\lambda_{ij}\) are defined as,

\[
\text{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \alpha_{10} + \alpha_{11}t_{ij} + \alpha_{12}x_{1ij} + b_{1i}\ 	ext{and (3.2)}
\]

\[
\text{log}(\lambda_{ij}) = \alpha_{20} + \alpha_{21}t_{ij} + \alpha_{22}x_{1ij} + b_{2i}\ 	ext{and (3.3)}
\]

and the random effects, \(b_{1i}\) and \(b_{2i}\) are possibly correlated. The log likelihood function is,

\[
L(\alpha_1, \alpha_2, \sigma_1, \sigma_2, \rho | Y) = \prod_{i=1}^{m} L_i(\alpha_1, \alpha_2, \sigma_1, \sigma_2, \rho | y_{i1}, \ldots, y_{im}) = \prod_{i=1}^{m} \int_{b_{1i}} \int_{b_{2i}} \int_{y_{i1}}^{y_{im}} \int_{y_{ij}}^{y_{ij}} L_i(\alpha_1, \alpha_2, \sigma_1, \sigma_2, \rho | y_{i1}, \ldots, y_{im}) \times f(b_{1i}, b_{2i} | \sigma_1, \sigma_2, \rho) db_{1i} db_{2i}
\]

where \(r_{ij}\) is defined in equations (2.14) and (2.15).

For random effects, \(b_{1i}\) and \(b_{2i}\) were generated from a bivariate normal distribution with both correlated and uncorrelated each other. With different parameter values, we considered the following cases including two random effects with high correlation and low correlation. The true parameter values are listed in Table 3.1. The chosen parameter values for Case 1 and Case 2 are suggested which resembles the application results.

In case 1 and case 3, correlation between two random effects, \(\rho\), is calculated to be about 0.71 resulting high correlation. On the other hand, in case 2 and case 4, the random effects are not highly correlated each other with \(\rho \approx 0.28\). As listed in Table 3.1, two different set of parameter values were considered in this simulation study. To compare the performance of proposed model, different sample sizes and measurement sizes per subject were considered. In each case, data with one hundred individuals and five hundreds individuals with short serial measurements of three, and long serial measurements of ten per subject were generated. The data simulation and model fitting was done 100 times for each cases, the simulation results is presented in the following section 3.2.

The model fitting process was done by using PROC NLMIXED procedure (SAS Institute, Cary, NC, Version 9.3) [24]. This procedure maximized the likelihood function using quasi-Newton optimization approximated by adaptive Gaussian quadrature by default option. It is allowing users to specify a general likelihood, in our case, one of the form (3.4). One drawback of this procedure is that the parameter estimation result can be very sensitive on initial starting values. The default setting of the initial value of parameter estimates for
Table 3.1: True parameter values of simulated data sets, eight cases

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Case 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occurence (Logistic)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>-8.0</td>
<td>-8.0</td>
<td>-3.0</td>
<td>-3.0</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>2.0</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>0.2</td>
<td>0.5</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Intensity (Poisson)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{20}$</td>
<td>-3.0</td>
<td>-3.0</td>
<td>-6.0</td>
<td>-6.0</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>1.3</td>
<td>1.3</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-1.5</td>
<td>-1.5</td>
</tr>
<tr>
<td>$\sigma_2^2$</td>
<td>0.4</td>
<td>1.0</td>
<td>0.4</td>
<td>1.0</td>
</tr>
<tr>
<td>$\rho\sigma_1\sigma_2$</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>$(\rho)$</td>
<td>0.7</td>
<td>0.3</td>
<td>0.7</td>
<td>0.3</td>
</tr>
</tbody>
</table>

any kind of model is 1 for every parameter. Also, it also have difficulty determining the number of quadrature points when the initial values are far from the optimum values. To overcome these, we first fit the logistic model with $r_{ij}$, then Poisson regression model with positive $y_{ij}$ values without any random effects. These estimation processes for initial value searching were also done by using SAS, PROC GENMOD procedure. Then, the parameter estimates were used as the initial values in PROC NLMIXED. For $\sigma_1^2$, $\sigma_2^2$, and $\rho\sigma_1\sigma_2$, 0.5 was given as the initial value. Each simulation was replicated 100 times, and the parameter estimates were compared.

### 3.2 Simulation Results

In each Table presented below, the first two column are the simulated parameters and their true values, then the next two columns are the mean and standard deviation of estimated parameters from 100 simulations with 100 generated subjects. The last two columns present the simulation results from the generated data with 500 subjects. In each case, four tables were presented, with different number of measurements of three and ten per subject, and correlated/uncorrelated random effects from two parts of the model.

The Tables 3.2, 3.3, 3.4, 3.5, and 3.6 present the result of simulation with true parameter values of case 1. With only 3 number of repeated measurements from each subject, the variance estimates of random effects were somewhat over estimated when the random effects from logit parts and Poisson parts were correlated with small number of subjects. The estimated value was closer to the true parameter value when the sample size gets larger. That is, comparing two simulation results with $m = 100$, and $m = 500$, the result from data set with $m = 500$ performed much better than the other one, with smaller standard deviation and the mean of estimated parameter values closer to the true value. So, we did
Table 3.2: Simulation results for the proposed model using 100 sets of simulated data sets \((n = 3)\) from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 1

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True Values</th>
<th>(m = 100)</th>
<th>Mean</th>
<th>Std.Dev</th>
<th>(m = 500)</th>
<th>Mean</th>
<th>Std.Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_{10})</td>
<td>-8.0</td>
<td>-8.424</td>
<td>2.014</td>
<td>-8.140</td>
<td>0.760</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha_{11})</td>
<td>2.0</td>
<td>2.101</td>
<td>0.439</td>
<td>2.031</td>
<td>0.163</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha_{12})</td>
<td>-5.0</td>
<td>-5.482</td>
<td>2.445</td>
<td>-5.098</td>
<td>0.341</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma^2_1)</td>
<td>0.2</td>
<td>0.325</td>
<td>0.426</td>
<td>0.215</td>
<td>0.207</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textbf{Occurrence (Logistic)}

\textbf{Intensity (Poisson)}

\(\alpha_{20}\) | -3.0 | -2.974 | 0.177 | -2.998 | 0.073 |
\(\alpha_{21}\) | 1.3 | 1.294 | 0.029 | 1.300 | 0.012 |
\(\alpha_{22}\) | -0.5 | -0.480 | 0.383 | -0.482 | 0.143 |
\(\sigma^2_2\) | 0.4 | 0.389 | 0.071 | 0.399 | 0.028 |

another investigation of the accuracy of estimates depending on the number of subject, as presented in table 3.4. Since, the table 3.3 and 3.6 show the simulation result when the number of measurement per subject is equal to 3 and 10 respectively, the additional simulation with the number of measurements per subject in between 3 and 10 were conducted. Table 3.4 shows the parameter estimates depending on different number of measurements per subject, when there are 100 simulated subjects with correlated random effects. You can easily observe that as the number of measurement increases, the estimation of parameters becomes more precise with smaller standard deviation and closer to the true parameter value. Among all parameters estimated, you can see this trend from the estimates of \(\sigma^2_1\) the most significantly.

However, with larger number of measurements per subject of 10, the simulation results show unbiased estimates of both variance and covariance values, as the simulation results were presented in 3.5 and 3.6. Also, the parameter estimates of categorical variable (\(\alpha_{12}\) and \(\alpha_{22}\)) were more variable than the estimates of continuous variable (\(\alpha_{11}\) and \(\alpha_{21}\)) in both parts of the model.
Table 3.3: Simulation results for the proposed model using 100 sets of simulated datasets \( (n = 3) \) from the model with correlated random effects given in Section 2.4 with true parameter values of case 1

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True ( m = 100 )</th>
<th>( m = 500 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_{10} )</td>
<td>-8.0 -8.540 2.018</td>
<td>-8.126 0.844</td>
</tr>
<tr>
<td>( \alpha_{11} )</td>
<td>2.0 2.118 0.443</td>
<td>2.028 0.182</td>
</tr>
<tr>
<td>( \alpha_{12} )</td>
<td>-5.0 -5.256 0.743</td>
<td>-5.034 0.335</td>
</tr>
<tr>
<td>( \sigma^2_1 )</td>
<td>0.2 0.453 0.450</td>
<td>0.260 0.202</td>
</tr>
<tr>
<td>( \rho \sigma_1 \sigma_2 )</td>
<td>0.2 0.273 0.169</td>
<td>0.208 0.063</td>
</tr>
<tr>
<td>( (\rho) )</td>
<td>0.7 0.628 0.649</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.4: Simulation results for the proposed model using 100 sets of simulated datasets with $m = 100$ subjects and different $n$ from the model with correlated random effects given in Section 2.4 with true parameter values of case 1

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True Values</th>
<th>$n = 4$</th>
<th>$n = 5$</th>
<th>$n = 6$</th>
<th>$n = 7$</th>
<th>$n = 8$</th>
<th>$n = 9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simulated</td>
<td>Mean</td>
<td>Std.Dev</td>
<td>Mean</td>
<td>Std.Dev</td>
<td>Mean</td>
<td>Std.Dev</td>
</tr>
<tr>
<td>Occurrence (Logistic)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>2.0</td>
<td>2.042</td>
<td>0.381</td>
<td>2.001</td>
<td>0.296</td>
<td>2.036</td>
<td>0.285</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>-5.0</td>
<td>-5.240</td>
<td>0.646</td>
<td>-5.186</td>
<td>0.593</td>
<td>-5.127</td>
<td>0.482</td>
</tr>
<tr>
<td>$\sigma^2_1$</td>
<td>0.2</td>
<td>0.379</td>
<td>0.351</td>
<td>0.309</td>
<td>0.224</td>
<td>0.269</td>
<td>0.232</td>
</tr>
<tr>
<td>Intensity (Poisson)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{20}$</td>
<td>-3.0</td>
<td>-3.019</td>
<td>0.172</td>
<td>-2.970</td>
<td>0.186</td>
<td>-2.998</td>
<td>0.113</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>1.3</td>
<td>1.301</td>
<td>0.026</td>
<td>1.295</td>
<td>0.026</td>
<td>1.298</td>
<td>0.017</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>-0.5</td>
<td>-0.530</td>
<td>0.347</td>
<td>-0.445</td>
<td>0.310</td>
<td>-0.484</td>
<td>0.254</td>
</tr>
<tr>
<td>$\sigma^2_2$</td>
<td>0.4</td>
<td>0.404</td>
<td>0.068</td>
<td>0.411</td>
<td>0.072</td>
<td>0.390</td>
<td>0.068</td>
</tr>
<tr>
<td>$\rho \sigma_1 \sigma_2$</td>
<td>0.2</td>
<td>0.217</td>
<td>0.118</td>
<td>0.221</td>
<td>0.130</td>
<td>0.211</td>
<td>0.105</td>
</tr>
<tr>
<td>$(\rho)$</td>
<td>0.7</td>
<td>0.555</td>
<td>0.620</td>
<td>0.653</td>
<td>0.660</td>
<td>0.660</td>
<td>0.647</td>
</tr>
</tbody>
</table>
Table 3.5: Simulation results for the proposed model using 100 sets of simulated datasets \((n = 10)\) from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 1

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True Values</th>
<th>m = 100</th>
<th>m = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.Dev</td>
<td>Mean</td>
</tr>
<tr>
<td>Occurrence (Logistic)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha_{10})</td>
<td>-8.0</td>
<td>-8.178</td>
<td>-7.931</td>
</tr>
<tr>
<td>(\alpha_{11})</td>
<td>2.0</td>
<td>2.036</td>
<td>1.987</td>
</tr>
<tr>
<td>(\alpha_{12})</td>
<td>-5.0</td>
<td>-5.112</td>
<td>-5.008</td>
</tr>
<tr>
<td>(\sigma_1^2)</td>
<td>0.2</td>
<td>0.178</td>
<td>0.193</td>
</tr>
<tr>
<td>Intensity (Poisson)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha_{20})</td>
<td>-3.0</td>
<td>-3.009</td>
<td>-2.998</td>
</tr>
<tr>
<td>(\alpha_{21})</td>
<td>1.3</td>
<td>1.300</td>
<td>1.300</td>
</tr>
<tr>
<td>(\alpha_{22})</td>
<td>-0.5</td>
<td>-0.505</td>
<td>-0.490</td>
</tr>
<tr>
<td>(\sigma_2^2)</td>
<td>0.4</td>
<td>0.388</td>
<td>0.398</td>
</tr>
</tbody>
</table>

The simulation results with the true parameter values of case 2 were summarized in Tables 3.7, 3.8, 3.9, and 3.10. Compare to the case 1, the estimated coefficients are the same except the elements of covariance matrix of random effects. With larger values of variances on random effects, they are weakly correlated. In comparison of Table 3.2 and 3.7, the standard deviation of 100 estimate parameters of case 2 are bigger than the ones with true parameters of case 1. As presented in Tables 3.7 and 3.8, some variability were observed from the estimated parameters when there were only three measurements per subject. Overall, the parameter values were well estimated when there were strong correlation in between random effects.

Tables 3.11, 3.12, 3.13, and 3.14 present the simulation results with the true parameter values of case 3. Although the highly correlated between random effects were set like in case 1, the model does not have relatively large intercept values to estimate in logistic part. The result summary statistics showed similar estimation behaviors as in case 1, they were slightly more reliable. As we observed in case 1, large value of standard deviation on hundred estimated coefficients on categorical variable from logistic part of the model, \(\alpha_{12}\) was observed. As in all other cases, the larger number of subject and measurements per subject, the more reliable estimation was performed.

The summary results of simulation of case 4 were presented in Tables 3.15, 3.16, 3.17, and 3.18. As we have seen in previous cases, relatively large standard deviations of estimated parameters were observed again with small number of measurements per subject (3.15 and 3.16). In case of uncorrelated random effects, the estimated parameters of logistic part were less variable comparing to case 2, whereas the estimated parameters of Poisson part were more variable. Compared to the simulation results of case 3, the similar patterns were observed as we compared the simulation results of case 1 and case 2. The estimated parameters were unbiased and reliable with more number of subjects and measurements as you can see well estimated parameters in Tables 3.17 and 3.18.
Table 3.6: Simulation results for the proposed model using 100 sets of simulated datasets ($n = 10$) from the model with correlated random effects given in Section 2.4 with true parameter values of case 1

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True Values</th>
<th>$m = 100$</th>
<th>$m = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Std.Dev</td>
</tr>
<tr>
<td><strong>Occurrence (Logistic)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>-8.0</td>
<td>-8.013</td>
<td>0.953</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>2.0</td>
<td>2.005</td>
<td>0.208</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>-5.0</td>
<td>-5.021</td>
<td>0.356</td>
</tr>
<tr>
<td>$\sigma^2_1$</td>
<td>0.2</td>
<td>0.219</td>
<td>0.150</td>
</tr>
<tr>
<td><strong>Intensity (Poisson)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{20}$</td>
<td>-3.0</td>
<td>-2.995</td>
<td>0.099</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>1.3</td>
<td>1.299</td>
<td>0.014</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>-0.5</td>
<td>-0.524</td>
<td>0.227</td>
</tr>
<tr>
<td>$\sigma^2_2$</td>
<td>0.4</td>
<td>0.392</td>
<td>0.059</td>
</tr>
<tr>
<td>$\rho \sigma_1 \sigma_2$</td>
<td>0.2</td>
<td>0.205</td>
<td>0.079</td>
</tr>
<tr>
<td>$(\rho)$</td>
<td>0.7</td>
<td>0.698</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.7: Simulation results for the proposed model using 100 sets of simulated datasets ($n = 3$) from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 2

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True Values</th>
<th>$n = 100$</th>
<th>$n = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Std.Dev</td>
</tr>
<tr>
<td><strong>Occurrence (Logistic)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>-8.0</td>
<td>-8.066</td>
<td>2.138</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>2.0</td>
<td>2.025</td>
<td>0.466</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>-5.0</td>
<td>-5.444</td>
<td>2.462</td>
</tr>
<tr>
<td>$\sigma^2_1$</td>
<td>0.5</td>
<td>0.616</td>
<td>0.543</td>
</tr>
<tr>
<td><strong>Intensity (Poisson)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{20}$</td>
<td>-3.0</td>
<td>-2.993</td>
<td>0.169</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>1.3</td>
<td>1.301</td>
<td>0.023</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>-0.5</td>
<td>-0.465</td>
<td>0.523</td>
</tr>
<tr>
<td>$\sigma^2_2$</td>
<td>1.0</td>
<td>0.972</td>
<td>0.169</td>
</tr>
</tbody>
</table>

33
Table 3.8: Simulation results for the proposed model using 100 sets of simulated datasets \((n = 3)\) from the model with correlated random effects given in Section 2.4 with true parameter values of case 2

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True Values</th>
<th>(m = 100)</th>
<th>(m = 500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_{10})</td>
<td>-8.0</td>
<td>-7.850</td>
<td>1.973</td>
</tr>
<tr>
<td>(\alpha_{11})</td>
<td>2.0</td>
<td>1.976</td>
<td>0.423</td>
</tr>
<tr>
<td>(\alpha_{12})</td>
<td>-5.0</td>
<td>-5.512</td>
<td>3.385</td>
</tr>
<tr>
<td>(\sigma_1^2)</td>
<td>0.5</td>
<td>0.541</td>
<td>0.582</td>
</tr>
<tr>
<td>(\rho_{\sigma_1\sigma_2})</td>
<td>0.2</td>
<td>0.187</td>
<td>0.264</td>
</tr>
<tr>
<td>((\rho))</td>
<td>0.3</td>
<td>0.252</td>
<td>0.308</td>
</tr>
</tbody>
</table>

Table 3.9: Simulation results for the proposed model using 100 sets of simulated datasets \((n = 10)\) from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 2

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True Values</th>
<th>(m = 100)</th>
<th>(m = 500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_{10})</td>
<td>-8.0</td>
<td>-7.894</td>
<td>1.973</td>
</tr>
<tr>
<td>(\alpha_{11})</td>
<td>2.0</td>
<td>1.982</td>
<td>0.240</td>
</tr>
<tr>
<td>(\alpha_{12})</td>
<td>-5.0</td>
<td>-5.085</td>
<td>0.429</td>
</tr>
<tr>
<td>(\sigma_1^2)</td>
<td>0.5</td>
<td>0.508</td>
<td>0.244</td>
</tr>
<tr>
<td>(\rho_{\sigma_1\sigma_2})</td>
<td>0.2</td>
<td>0.187</td>
<td>0.264</td>
</tr>
<tr>
<td>((\rho))</td>
<td>0.3</td>
<td>0.252</td>
<td>0.308</td>
</tr>
</tbody>
</table>
Table 3.10: Simulation results for the proposed model using 100 sets of simulated datasets ($n = 10$) from the model with correlated random effects given in Section 2.4 with true parameter values of case 2

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True Values</th>
<th>$m = 100$</th>
<th>Mean</th>
<th>Std.Dev</th>
<th>$m = 500$</th>
<th>Mean</th>
<th>Std.Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Occurrence (Logistic)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>$-8.0$</td>
<td>$-7.954$</td>
<td>$1.010$</td>
<td>$-8.000$</td>
<td>$0.505$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$2.0$</td>
<td>$1.988$</td>
<td>$0.216$</td>
<td>$2.000$</td>
<td>$0.108$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>$-5.0$</td>
<td>$-5.050$</td>
<td>$0.406$</td>
<td>$-5.012$</td>
<td>$0.196$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>$0.5$</td>
<td>$0.460$</td>
<td>$0.205$</td>
<td>$0.477$</td>
<td>$0.086$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intensity (Poisson)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>$-3.0$</td>
<td>$-2.975$</td>
<td>$0.218$</td>
<td>$-2.985$</td>
<td>$0.125$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$1.0$</td>
<td>$1.299$</td>
<td>$0.022$</td>
<td>$1.297$</td>
<td>$0.019$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>$-0.5$</td>
<td>$-0.480$</td>
<td>$0.388$</td>
<td>$-0.489$</td>
<td>$0.142$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>$1.0$</td>
<td>$0.982$</td>
<td>$0.141$</td>
<td>$1.006$</td>
<td>$0.084$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho \sigma_1 \sigma_2$</td>
<td>$0.2$</td>
<td>$0.193$</td>
<td>$0.145$</td>
<td>$0.205$</td>
<td>$0.057$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\rho)$</td>
<td>$0.3$</td>
<td>$0.287$</td>
<td>$0.296$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.11: Simulation results for the proposed model using 100 sets of simulated datasets ($n = 3$) from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 3

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True Values</th>
<th>$m = 100$</th>
<th>Mean</th>
<th>Std.Dev</th>
<th>$m = 500$</th>
<th>Mean</th>
<th>Std.Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Occurrence (Logistic)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>$-3.0$</td>
<td>$-3.137$</td>
<td>$1.603$</td>
<td>$-3.120$</td>
<td>$0.841$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$1.0$</td>
<td>$1.043$</td>
<td>$0.344$</td>
<td>$1.029$</td>
<td>$0.172$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>$-5.0$</td>
<td>$-5.360$</td>
<td>$0.769$</td>
<td>$-5.084$</td>
<td>$0.338$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>$0.2$</td>
<td>$0.370$</td>
<td>$0.408$</td>
<td>$0.225$</td>
<td>$0.234$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intensity (Poisson)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>$-6.0$</td>
<td>$-5.975$</td>
<td>$0.139$</td>
<td>$-5.996$</td>
<td>$0.060$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$2.0$</td>
<td>$1.997$</td>
<td>$0.023$</td>
<td>$1.999$</td>
<td>$0.010$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>$-1.5$</td>
<td>$-1.585$</td>
<td>$0.518$</td>
<td>$-1.530$</td>
<td>$0.205$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>$0.4$</td>
<td>$0.395$</td>
<td>$0.079$</td>
<td>$0.399$</td>
<td>$0.028$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.12: Simulation results for the proposed model using 100 sets of simulated datasets ($n = 3$) from the model with correlated random effects given in Section 2.4 with true parameter values of case 3

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True Values</th>
<th>$m = 100$</th>
<th>Std.Dev</th>
<th>$m = 500$</th>
<th>Std.Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Occurrence (Logistic)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>-3.0</td>
<td>-3.226</td>
<td>1.828</td>
<td>-2.933</td>
<td>0.858</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>1.0</td>
<td>1.062</td>
<td>0.377</td>
<td>0.987</td>
<td>0.181</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>-5.0</td>
<td>-5.664</td>
<td>3.409</td>
<td>-5.055</td>
<td>0.428</td>
</tr>
<tr>
<td>$\sigma_{1}^{2}$</td>
<td>0.2</td>
<td>0.461</td>
<td>0.446</td>
<td>0.209</td>
<td>0.149</td>
</tr>
<tr>
<td><strong>Intensity (Poisson)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{20}$</td>
<td>-6.0</td>
<td>-6.006</td>
<td>0.197</td>
<td>-5.995</td>
<td>0.182</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>2.0</td>
<td>1.997</td>
<td>0.029</td>
<td>1.997</td>
<td>0.026</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>-1.5</td>
<td>-1.534</td>
<td>0.506</td>
<td>-1.501</td>
<td>0.278</td>
</tr>
<tr>
<td>$\sigma_{2}^{2}$</td>
<td>0.4</td>
<td>0.413</td>
<td>0.071</td>
<td>0.393</td>
<td>0.055</td>
</tr>
<tr>
<td>$\rho \sigma_{1} \sigma_{2}$</td>
<td>0.2</td>
<td>0.216</td>
<td>0.165</td>
<td>0.191</td>
<td>0.092</td>
</tr>
<tr>
<td>$(\rho)$</td>
<td>0.7</td>
<td>0.495</td>
<td></td>
<td>0.667</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.13: Simulation results for the proposed model using 100 sets of simulated datasets ($n = 10$) from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 3

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True Values</th>
<th>$m = 100$</th>
<th>Std.Dev</th>
<th>$m = 500$</th>
<th>Std.Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Occurrence (Logistic)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>-3.0</td>
<td>-2.966</td>
<td>0.197</td>
<td>-2.999</td>
<td>0.436</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>1.0</td>
<td>0.999</td>
<td>0.186</td>
<td>0.999</td>
<td>0.089</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>-5.0</td>
<td>-5.099</td>
<td>0.415</td>
<td>-5.010</td>
<td>0.176</td>
</tr>
<tr>
<td>$\sigma_{1}^{2}$</td>
<td>0.2</td>
<td>0.231</td>
<td>0.177</td>
<td>0.245</td>
<td>0.121</td>
</tr>
<tr>
<td><strong>Intensity (Poisson)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{20}$</td>
<td>-6.0</td>
<td>-5.970</td>
<td>0.206</td>
<td>-5.953</td>
<td>0.147</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>2.0</td>
<td>1.996</td>
<td>0.029</td>
<td>1.995</td>
<td>0.015</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>-1.5</td>
<td>-1.571</td>
<td>0.258</td>
<td>-1.505</td>
<td>0.122</td>
</tr>
<tr>
<td>$\sigma_{2}^{2}$</td>
<td>0.4</td>
<td>0.402</td>
<td>0.065</td>
<td>0.414</td>
<td>0.042</td>
</tr>
</tbody>
</table>
Table 3.14: Simulation results for the proposed model using 100 sets of simulated datasets \((n = 10)\) from the model with correlated random effects given in Section 2.4 with true parameter values of case 3

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True Values</th>
<th>(m = 100)</th>
<th>Mean</th>
<th>Std.Dev</th>
<th>(m = 500)</th>
<th>Mean</th>
<th>Std.Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_{10})</td>
<td>-3.0</td>
<td>-2.962</td>
<td>0.848</td>
<td>-3.065</td>
<td>0.398</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha_{11})</td>
<td>1.0</td>
<td>0.997</td>
<td>0.179</td>
<td>1.012</td>
<td>0.084</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha_{12})</td>
<td>-5.0</td>
<td>-5.115</td>
<td>0.347</td>
<td>-4.991</td>
<td>0.200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma^2_1)</td>
<td>0.2</td>
<td>0.282</td>
<td>0.221</td>
<td>0.229</td>
<td>0.120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\rho \sigma_1 \sigma_2)</td>
<td>0.2</td>
<td>0.206</td>
<td>0.072</td>
<td>0.233</td>
<td>0.096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>((\rho))</td>
<td>0.7</td>
<td>0.616</td>
<td></td>
<td></td>
<td>0.758</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.15: Simulation results for the proposed model using 100 sets of simulated datasets \((n = 3)\) from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 4

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True Values</th>
<th>(m = 100)</th>
<th>Mean</th>
<th>Std.Dev</th>
<th>(m = 500)</th>
<th>Mean</th>
<th>Std.Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_{10})</td>
<td>-3.0</td>
<td>-2.985</td>
<td>1.647</td>
<td>-3.006</td>
<td>0.716</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha_{11})</td>
<td>1.0</td>
<td>0.993</td>
<td>0.352</td>
<td>0.998</td>
<td>0.146</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha_{12})</td>
<td>-5.0</td>
<td>-5.457</td>
<td>2.539</td>
<td>-4.971</td>
<td>0.434</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma^2_1)</td>
<td>0.5</td>
<td>0.550</td>
<td>0.608</td>
<td>0.501</td>
<td>0.286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\rho \sigma_1 \sigma_2)</td>
<td>0.2</td>
<td>0.532</td>
<td>0.087</td>
<td>-5.903</td>
<td>0.398</td>
<td></td>
<td></td>
</tr>
<tr>
<td>((\rho))</td>
<td>0.7</td>
<td>1.992</td>
<td>1.987</td>
<td>0.555</td>
<td>0.055</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma^2_2)</td>
<td>1.0</td>
<td>0.997</td>
<td>0.158</td>
<td>0.984</td>
<td>0.093</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.16: Simulation results for the proposed model using 100 sets of simulated datasets \((n = 3)\) from the model with correlated random effects given in Section 2.4 with true parameter values of case 4

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True Values</th>
<th>(m = 100) Mean</th>
<th>Std.Dev</th>
<th>(m = 500) Mean</th>
<th>Std.Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_{10})</td>
<td>-3.0</td>
<td>-3.359</td>
<td>1.798</td>
<td>-2.979</td>
<td>0.784</td>
</tr>
<tr>
<td>(\alpha_{11})</td>
<td>1.0</td>
<td>1.084</td>
<td>0.365</td>
<td>1.007</td>
<td>0.167</td>
</tr>
<tr>
<td>(\alpha_{12})</td>
<td>-5.0</td>
<td>-6.013</td>
<td>4.137</td>
<td>-5.075</td>
<td>0.419</td>
</tr>
<tr>
<td>(\sigma_1^2)</td>
<td>0.5</td>
<td>0.707</td>
<td>0.780</td>
<td>0.578</td>
<td>0.317</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intensity (Poisson)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_{20})</td>
<td>-6.0</td>
<td>-5.994</td>
<td>0.446</td>
<td>-5.971</td>
<td>0.230</td>
</tr>
<tr>
<td>(\alpha_{21})</td>
<td>2.0</td>
<td>1.997</td>
<td>0.072</td>
<td>1.996</td>
<td>0.031</td>
</tr>
<tr>
<td>(\alpha_{22})</td>
<td>-1.5</td>
<td>-1.416</td>
<td>0.651</td>
<td>-1.548</td>
<td>0.279</td>
</tr>
<tr>
<td>(\sigma_2^2)</td>
<td>1.0</td>
<td>0.984</td>
<td>0.184</td>
<td>0.975</td>
<td>0.086</td>
</tr>
<tr>
<td>(\rho \sigma_1 \sigma_2)</td>
<td>0.2</td>
<td>0.226</td>
<td>0.247</td>
<td>0.224</td>
<td>0.098</td>
</tr>
<tr>
<td>((\rho))</td>
<td>0.3</td>
<td>0.271</td>
<td></td>
<td></td>
<td>0.298</td>
</tr>
</tbody>
</table>

Table 3.17: Simulation results for the proposed model using 100 sets of simulated datasets \((n = 10)\) from the model with uncorrelated random effects given in Section 2.4 with true parameter values of case 4

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>True Values</th>
<th>(m = 100) Mean</th>
<th>Std.Dev</th>
<th>(m = 500) Mean</th>
<th>Std.Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_{10})</td>
<td>-3.0</td>
<td>-3.127</td>
<td>0.903</td>
<td>-3.085</td>
<td>0.416</td>
</tr>
<tr>
<td>(\alpha_{11})</td>
<td>1.0</td>
<td>1.017</td>
<td>0.187</td>
<td>1.005</td>
<td>0.085</td>
</tr>
<tr>
<td>(\alpha_{12})</td>
<td>-5.0</td>
<td>-4.968</td>
<td>0.433</td>
<td>-4.876</td>
<td>0.265</td>
</tr>
<tr>
<td>(\sigma_1^2)</td>
<td>0.5</td>
<td>0.573</td>
<td>0.314</td>
<td>0.653</td>
<td>0.260</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intensity (Poisson)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_{20})</td>
<td>-6.0</td>
<td>-5.753</td>
<td>0.522</td>
<td>-5.794</td>
<td>0.395</td>
</tr>
<tr>
<td>(\alpha_{21})</td>
<td>2.0</td>
<td>1.971</td>
<td>0.064</td>
<td>1.991</td>
<td>0.049</td>
</tr>
<tr>
<td>(\alpha_{22})</td>
<td>-1.5</td>
<td>-1.572</td>
<td>0.396</td>
<td>-1.490</td>
<td>0.216</td>
</tr>
<tr>
<td>(\sigma_2^2)</td>
<td>1.0</td>
<td>1.007</td>
<td>0.139</td>
<td>0.990</td>
<td>0.056</td>
</tr>
</tbody>
</table>
Table 3.18: Simulation results for the proposed model using 100 sets of simulated datasets ($n = 10$) from the model with correlated random effects given in Section 2.4 with true parameter values of case 4

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
<th>$m = 100$</th>
<th>$m = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.Dev</td>
<td>Mean</td>
</tr>
<tr>
<td>Occurrence (Logistic)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>-3.0</td>
<td>-3.097</td>
<td>0.958</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>1.0</td>
<td>1.019</td>
<td>0.198</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>-5.0</td>
<td>-5.089</td>
<td>0.438</td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>0.5</td>
<td>0.480</td>
<td>0.194</td>
</tr>
<tr>
<td>Intensity (Poisson)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{20}$</td>
<td>-6.0</td>
<td>-5.756</td>
<td>0.627</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>2.0</td>
<td>1.967</td>
<td>0.078</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>-1.5</td>
<td>-1.503</td>
<td>0.410</td>
</tr>
<tr>
<td>$\sigma_2^2$</td>
<td>1.0</td>
<td>0.900</td>
<td>0.222</td>
</tr>
<tr>
<td>$\rho \sigma_1 \sigma_2$</td>
<td>0.2</td>
<td>0.241</td>
<td>0.174</td>
</tr>
<tr>
<td>($\rho$)</td>
<td>0.3</td>
<td>0.366</td>
<td></td>
</tr>
</tbody>
</table>

Overall, with long series of measurement showed better performance in terms of parameter estimations even with smaller number of sample sizes of 100. Our simulation study show that sample size of 100 and more than three measurement per subject was enough to produce unbiased estimation of true parameters with either weak or strong correlated random effects. The number of observation was critical in terms of estimation of parameters in the model. Also, the parameter associated with dichotomous variable estimated worse than the parameter related to continuous variable. In comparison of the model with highly correlated random effects and weakly correlated random effects, the model with high correlation was more precisely estimated. Since we estimate the parameters from both parts simultaneously, higher correlation was well detected.
CHAPTER 4

APPLICATION TO LETTER-SOUND FLUENCY OF KINDERGARTEN CHILDREN

This chapter presents the application of our proposed model from the previous chapter 2.4. Data measuring letter sound fluency of 461 children collected during 2008-2009 academic year was used in the model. A dependent variable was a number of counts of correctly pronounced letter sound, and demographic information including age were used as independent variables. Each student had three measurements of letter sound fluency, fall, winter, and spring. The proposed model predicted outcome in two parts, which was a probability of positive number of correct pronunciation, and then number of correct pronunciation per minute. After detailed data description in section 4.1, both full and reduced model are specified in section 4.2 and the results are presented in section 4.3.

4.1 Data Description

As stated in chapter 1.1, the data used in this study was a part of a large-scaled cluster-randomized control field trial by Al Otaiba et al. (2011) [2]. The aim of the original study was identifying students with reading disabilities using Response to Intervention (RTI) approaches [2]. RTI is a general education framework involving instructional intervention and monitoring of student progress by universal screening. Recently, among the RTI models, the multitiered models in early literacy intervention were demonstrated to be effective in terms of identifying individuals with reading disability. In study by Al Otaiba et al. (2011) [2], a specially designed cluster-randomized control field trial was conducted to examine the effects of multitiered interventions, classroom instruction (Tier 1) and targeted and differentiated small group interventions (Tier 2) to the beginning literacy trajectories of students.

There were total 14 schools participated from northern Florida, the schools were randomly assigned to two groups, control group and treatment group. The assignment of schools to groups were based on several criteria, proportion of students who received free or reduced-priced lunch, Title I and Reading First participation, and the proportion of students who passed the Florida high-stakes reading test at third grade, as a proxy for socioeconomic status of students, extra resources of schools, and overall response to reading instruction of students, respectively.
Table 4.1: Descriptive statistics of the sample.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean ± Std.Dev.</th>
<th>Frequency (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>5.15 ± 0.29</td>
<td></td>
</tr>
<tr>
<td>Winter</td>
<td>5.54 ± 0.51</td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>5.80 ± 0.46</td>
<td></td>
</tr>
<tr>
<td>Scores</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>8.14 ± 9.87</td>
<td></td>
</tr>
<tr>
<td>Winter</td>
<td>25.04 ± 15.22</td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>38.90 ± 17.54</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>252(54.66%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>209(45.34%)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>4(0.87%)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>4(0.87%)</td>
<td></td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>1(0.22%)</td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>268(58.13%)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>145(31.45%)</td>
<td></td>
</tr>
<tr>
<td>Multiracial</td>
<td>26(5.64%)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>13(2.82%)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>28(6.07%)</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>433(93.93%)</td>
<td></td>
</tr>
<tr>
<td>Eligible for free and reduced-priced lunch (FARL)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>275(67.24%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>134(32.76%)</td>
<td></td>
</tr>
<tr>
<td>Intervention Status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>214(46.42%)</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>247(53.58%)</td>
<td></td>
</tr>
</tbody>
</table>

*Note that Black (African American) accounts for 58.13%, in the analysis, this variable was dichotomized.*

In year of 2007-2008, teachers in both groups received common professional development\(^1\). In addition to the professional development, teachers in treatment group received

\(^1\)(1) a researcher-delivered summer day-long workshop on RTI and individualized instruction, (2) materials and games for center activities, and (3) data on students’ reading performance provided through Florida’s Process Monitoring Reporting Network (PMRN), the state’s Web-accessed database. For details, refer Al Otaiba at el. (2011) \(^2\)
additional training on Individualized Student Instruction for Kindergarten (ISI-K)\textsuperscript{2} using A2i software\textsuperscript{3}. In following year, 2008-2009, the control group also received the ISI-K intervention training, and the treatment group received the second year training of ISI-K.

In this study, we applied our model to the data collected from 2008-2009. Among student measures, our interest was focused on the relationship between the letter-sound correspondence and characteristics of students. The kindergarten children’s progress in early letter-sound correspondence was assessed using the AIMSWeb Letter Sound Fluency. It included 10 rows of 10 lower case letters, and the number of letter with correct sound was counted during 60 seconds on each child. The testing was discontinued if a child can not produce any correct sounds in the first 10. Also, if a child was able to pronounce all 100 lower case letters within 60 seconds, the raw score was rescaled into 60 seconds. For example, if a child finished reading 100 lower case letters in 52 seconds, the score of 100 is multiplied by 60/52. The collected data contains excessive amount of zeros, which means, there are students who were not able to pronounce a correct sound.

With consents of both parents and teacher, the data were collected from 461 students aged between 5 and 7 years old. The collected data include various demographic information such as birth date, gender, race, and free or reduced-priced lunch recipient status without any identifications of students. Basic descriptive statistics are presented in Table 4.1. Each student was tested on AIMSWeb Letter-Sound Fluency (LSF) for three times, fall, winter and spring during the academic year 2008-2009. Since the data were collected from 13 different schools, measurement times were not equally spaced.

The scores of each student tend to increase as they learn, the positive correlation was observed among the three scores collected over time. Pearson correlation coefficients were calculated and presented in Table 4.2. However, it is also noted that not all students scored higher than the previous test. Average score increased in winter compared to fall test score was 16.90, 21(4.56\%) children did not scored better than fall test. In spring, student scored on average 13.86 points higher than the previous test, although there were 53(11.50\%) children who scored worse than the winter test.

Table 4.2: Pearson Correlation Coefficients of AIMSWeb LSF measured at fall, winter and spring, \(n = 461\)

<table>
<thead>
<tr>
<th></th>
<th>(y_{i1})</th>
<th>(y_{i2})</th>
<th>(y_{i3})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_{i1})</td>
<td>1</td>
<td>0.54</td>
<td>0.46</td>
</tr>
<tr>
<td>(y_{i2})</td>
<td>0.54</td>
<td>1</td>
<td>0.72</td>
</tr>
<tr>
<td>(y_{i3})</td>
<td>0.49</td>
<td>0.72</td>
<td>1</td>
</tr>
</tbody>
</table>

Overall, throughout the study period, exam scores were increased 30.76 points on average comparing fall and spring score. Only four students did not show any improvement.

\textsuperscript{2}Designed by Connor and colleagues, the ISI-K intervention includes A2i software, ongoing teacher professional development, and in-class support \cite{2}.

\textsuperscript{3}A2i software calculates the amounts and types of instruction for each child based on the test scores, month in school, and desired end of year outcome \cite{2}.
on the exam score, however, looking at individually, without any dramatic changes on the scores. For example, one student (subject (d) in 4.1) scored 28, 22 and 25 on each assessment.

It is also observed that more than 30% of children scored a zero on the first test in Fall. The percentage of students who scored zeros on the test decreased over time. 5.64% and 1.31% of students scored zero on the test in Winter and Spring respectively. As presented in Figure 4.2, the distribution of score is highly skewed to the right.
Figure 4.1: AIMS score of Fall, Winter and Spring test for four subjects [(a)-(d)]
Figure 4.2: Distribution of AIMS score of Fall, Winter, and Spring.
4.2 Model Specification

AIMS scores measured in fall, winter and spring were used as a response variable. Each student’s age at the time of the test taken was used as an explanatory variable to estimate the number of letters with correct sound of each person. In the full model, age, gender, intervention status (control group or treatment group), eligibility for free or reduced-priced lunch, race, and ethnicity were included as predictors. We excluded ethnicity as there were only few students who were Hispanic.

The proposed model specified in section 2.4 was applied to the data. Let $y_{ij}$ the $j^{th}$ measurement of $i^{th}$ students on LSF, the formal structure of the full model follows by:

$$Pr(Y_{ij} = y_{ij}) = \begin{cases} 
(1 - p_{ij}) + p_{ij}e^{-\lambda_{ij}}, & y_i = 0, \\
p_{ij}e^{-\lambda_{ij}} y_{ij}^{-1}, & y_{ij} > 0, 0 \leq p_{ij} \leq 1.
\end{cases} \quad (4.1)$$

where $p_{ij}$ and $\lambda_{ij}$ are defined as,

$$\text{logit}(p_{ij}) = \alpha_{10} + \alpha_{11} \cdot \text{Age}_{ij} + \alpha_{12} \cdot \text{Female}_{ij} + \alpha_{13} \cdot \text{FARL}_{ij} + \alpha_{14} \cdot \text{Treatment}_{ij} + \alpha_{15} \cdot \text{Non} - \text{Black}_{ij} + b_{1i} \quad (4.2)$$

$$\text{log}(\lambda_{ij}) = \alpha_{20} + \alpha_{21} \cdot \text{Age}_{ij} + \alpha_{22} \cdot \text{Female}_{ij} + \alpha_{23} \cdot \text{FARL}_{ij} + \alpha_{24} \cdot \text{Treatment}_{ij} + \alpha_{25} \cdot \text{Non} - \text{Black}_{ij} + b_{2i} \quad (4.3)$$

with $i = 1, 2, ..., 461$ individuals and $j = 1, 2, 3$ measurements of each individuals. Male, African American, non free or reduced-priced lunch recipient and control group were used as reference categories.

In terms of random effects, both correlated and uncorrelated random effects were considered and fitted. Both models were fitted using PROC NLMIXED procedure (SAS Institute, Cary, NC, Version 9.3). As described in section 2.4, the initial staring values were fitted before fitting the proposed model. The starting value search process involved redefined observed values of $r_{ij}$,

$$R_{ij} = \begin{cases} 
0, & \text{if } Y_{ij} = 0, \\
1, & \text{if } Y_{ij} > 0.
\end{cases} \quad (4.4)$$

For the proposed model with uncorrelated random effects, the following process was carried out. First, the logistic regression was fitted using $r_{ij}$ without any random effects, then fixed effect Poisson regression was fitted with $S_{ij} = [Y_{ij}|R_{ij} = 1]$. Then the parameter estimates from two separate preliminary estimation were used as starting values. For random effects, 0.5 was used for estimation of $\sigma_1^2$ and $\sigma_2^2$. The same starting values were used to fit the proposed model with correlated random effects. For starting value of $\sigma_{12}$, the starting value was calculated so that the correlation between two random effects to be 0.5. For example, if the estimated values of $\sigma_1^2$ and $\sigma_2^2$ were 0.48 and 0.38 respectively, then the starting value was set to be $\rho \sigma_1 \sigma_2 = 0.5 \cdot 0.48 \cdot 0.38 \approx 0.09$. It was chosen because we assume moderate amount of positive correlation of two random effects.

Variables were selected by backward selection methods. Finally, the model was reduced to three predictors, age, gender, and eligibility for free or reduced-price lunch. Therefore,
the $p_{ij}$ and $\lambda_{ij}$ are defined as,

$$
\text{logit}(p_{ij}) = \alpha_{10} + \alpha_{11} \cdot \text{Age}_{ij} + \alpha_{12} \cdot \text{Female}_{ij} + \alpha_{13} \cdot \text{FARL}_{ij} + b_{1i}
$$

$$
\log(\lambda_{ij}) = \alpha_{20} + \alpha_{21} \cdot \text{Age}_{ij} + \alpha_{22} \cdot \text{Female}_{ij} + \alpha_{23} \cdot \text{FARL}_{ij} + b_{2i}
$$

with $i = 1, 2, \ldots, 461$ individuals and $j = 1, 2, 3$ measurements of each individuals. Graphical analysis of the data was done to check if there was any potential interaction effects among the variables in reduced model, Age, Female, and FARL, but no pattern was found. The fitted models were compared based on a likelihood ratio test and AIC, AICC, and BIC. Parameter estimates with model fit statistics were described in the following section.

### 4.3 Summary of the results

Both models with and without correlated random effects were fitted and the results are presented in Table 4.3 and 4.4. In the full model, the intervention of status of ISI-K training and race variable were not significant in both occurrence and intensity part of the model. As we selected the variables by backward selection method, non significant covariates were removed from the model one at a time. Then our final model includes age, gender and the free or reduced-price lunch recipient status. In this application, the covariates in the logistic part of the model and the Poisson part of the model were the same, although theoretically they are not necessarily the same.

The students in teachers with ISI-K group outperformed students in teachers with professional development only group in 2007-2008 academic year. As presented in Table 4.3, this was not observed in the following year. We suspect that the treatment variable is not significant, since all the teachers in the group already received a professional development and ISI-K training in 2008-2009. In both parts of the model, the logistic part and the intensity part, an older student scored better. As students gets one year older, they are 6.6 times more likely being able to name the sound of letter. Among those who are able to name the sound, they are likely to name 3 more letters correctly in a given minute.

In terms of gender, a female student was more likely to pronounce correct sound of letters and get higher scores. The odds of recognizing at least one lower case alphabet sound of female students are 1.74. On average, female students are predicted to name one more letter correctly in a given minute. The eligibility of free or reduced-price lunch was significantly associated with letter-sound fluency. A student who did not receive free or reduced-priced lunch was not only less likely get zero score, but also predicted to get higher scores on the test. Difference between two parts of the model was magnitudes of the estimates. Generally, estimates were smaller in the intensity part of the model than the occurrence part of the model. However, the direction of parameters from both of the part were the same. Furthermore, in this particular case, the variables significant in the occurrence part were also significant in intensity part of the model.

Random effects on each part of the model showed different results. A variance estimate on random effects for the occurrence part of the model was not significant in the model. In other words, probability of scoring non-zeros on the test was not different among the children. However, random effects variance for intensity part of the model was highly significant, meaning that some children got much better scores than their classmates when the
Table 4.3: Parameter estimates of the full model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Uncorrelated Random Effects</th>
<th></th>
<th></th>
<th>Correlated Random Effects</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std Err</td>
<td>Prob &gt;</td>
<td>Estimate</td>
<td>Std Err</td>
<td>Prob &gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( t )</td>
<td></td>
<td></td>
<td>( t )</td>
</tr>
<tr>
<td>Occurence (Logistic)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-7.983</td>
<td>1.491</td>
<td>&lt; .0001</td>
<td>-7.991</td>
<td>1.349</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>age</td>
<td>1.915</td>
<td>0.288</td>
<td>&lt; .0001</td>
<td>1.888</td>
<td>0.258</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Female</td>
<td>0.564</td>
<td>0.211</td>
<td>0.008</td>
<td>0.559</td>
<td>0.207</td>
<td>0.007</td>
</tr>
<tr>
<td>FARL</td>
<td>-0.603</td>
<td>0.264</td>
<td>0.023</td>
<td>-0.609</td>
<td>0.261</td>
<td>0.020</td>
</tr>
<tr>
<td>Treatment</td>
<td>-0.057</td>
<td>0.207</td>
<td>0.783</td>
<td>-0.064</td>
<td>0.202</td>
<td>0.750</td>
</tr>
<tr>
<td>Non-Black</td>
<td>0.209</td>
<td>0.240</td>
<td>0.384</td>
<td>0.204</td>
<td>0.237</td>
<td>0.389</td>
</tr>
<tr>
<td>( \sigma_1^2 )</td>
<td>0.432</td>
<td>0.409</td>
<td>0.291</td>
<td>0.414</td>
<td>0.451</td>
<td>0.359</td>
</tr>
<tr>
<td>Intensity (Poisson)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.923</td>
<td>0.129</td>
<td>&lt; .0001</td>
<td>-1.921</td>
<td>0.128</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>age</td>
<td>0.937</td>
<td>0.019</td>
<td>&lt; .0001</td>
<td>0.934</td>
<td>0.018</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Female</td>
<td>0.131</td>
<td>0.065</td>
<td>0.044</td>
<td>0.136</td>
<td>0.063</td>
<td>0.033</td>
</tr>
<tr>
<td>FARL</td>
<td>-0.217</td>
<td>0.081</td>
<td>0.007</td>
<td>-0.213</td>
<td>0.079</td>
<td>0.007</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.019</td>
<td>0.066</td>
<td>0.770</td>
<td>0.019</td>
<td>0.064</td>
<td>0.768</td>
</tr>
<tr>
<td>Non-Black</td>
<td>-0.096</td>
<td>0.077</td>
<td>0.216</td>
<td>-0.092</td>
<td>0.076</td>
<td>0.224</td>
</tr>
<tr>
<td>( \sigma_2^2 )</td>
<td>0.403</td>
<td>0.032</td>
<td>&lt; .0001</td>
<td>0.385</td>
<td>0.029</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>( \rho \sigma_1 \sigma_2 )</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.399</td>
<td>0.069</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

AIC       13024.36  12964.41  
AICC      13024.71  12964.81  
BIC       13080.55  13024.62  
-2 Loglikelihood 12996.36  12934.41  
Diff in -2ll 61.95  < .0001

\(^a\chi^2_{df=1}^2\) test was used.

The child scored more than zero point on the test. However, when the estimates of variances on random effects are closer to zero, it is recommended to compare the models with and without random effects by likelihood ratio test. It is because the normality approximation will be extremely poor in many cases, so Wald test on the variances can be highly misleading \([54],[22]\). The quantile-quantile plots and histograms of random effects were presented in Figure 4.3, non-normal behavior of random effects from logistic part was observed. However, straight line of quantile-quantile plot of random effects from Poisson part indicates that the normality assumption was not violated. One approach of testing the heterogeneity among the subject is comparing the model without random effects to the model with random effects. In the present case, the difference in twice of negative log-likelihood values between the model with and without random effects is relatively large (18599.46 vs. 12939.39, respectively)\(^4\), so it is clear that we prefer the model with random effects.

\(^4\)For parameter estimates of the model without random effects, see Appendix A, Table A.1.
Table 4.4: Parameter estimates of the reduced model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Uncorrelated Random Effects</th>
<th>Correlated Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std Err</td>
</tr>
<tr>
<td><strong>Occurrence (Logistic)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-8.020</td>
<td>1.504</td>
</tr>
<tr>
<td>age</td>
<td>1.920</td>
<td>0.293</td>
</tr>
<tr>
<td>Female</td>
<td>0.560</td>
<td>0.209</td>
</tr>
<tr>
<td>FARL</td>
<td>-0.484</td>
<td>0.226</td>
</tr>
<tr>
<td>$\sigma^2_1$</td>
<td>0.387</td>
<td>0.391</td>
</tr>
<tr>
<td><strong>Intensity(Poisson)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.933</td>
<td>0.122</td>
</tr>
<tr>
<td>age</td>
<td>0.937</td>
<td>0.019</td>
</tr>
<tr>
<td>Female</td>
<td>0.129</td>
<td>0.065</td>
</tr>
<tr>
<td>FARL</td>
<td>-0.270</td>
<td>0.069</td>
</tr>
<tr>
<td>$\sigma^2_2$</td>
<td>0.405</td>
<td>0.032</td>
</tr>
<tr>
<td>$\rho\sigma_1\sigma_2$</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

| AIC    | 13017.81 | 12961.30 |
| AICC   | 13017.99 | 12961.51 |
| BIC    | 13057.95 | 13005.45 |
| -2 Loglikelihood | 12997.81 | 12939.30 |
| Diff in -2ll | 58.51    | < .0001 |

$^a$ $\chi^2_{df=1}$ test was used.

Also, a correlation between the random effects from each part of the model was significant for both full and reduced models. The estimate on the correlation between the occurrence and intensity part of the model is positive. It indicates that students who were able to pronounce correctly at least one letter, they scored higher than who were not. In other words, after accounting for covariate differences, children tended to have a non-zero count on LSF AIMSWeb test, also tended to earn higher mean number of counts on the test. Although there were subtle differences in estimated parameters between the models with and without correlated random effects, the model with correlated random effects found to be better than the one without it based on a likelihood ratio test and other model selection criteria for both full and reduced model.

In addition to the models specified above, we also fitted additional models when $p_{ij}$ and $\lambda_{ij}$ were related as a function of each other with the shape parameter $\tau$. Similar to the models with unrelated $p_{ij}$ and $\lambda_{ij}$, we preferred the model with random effects over the model with only fixed effects$^5$. Also, the correlated random effects from the occurrence part and the intensity part was favored based on log likelihood ratio test$^6$.

---

$^5$ $\chi^2_{df=2} = 5606.92$, $p < 0.0001$ for full model, and $\chi^2_{df=2} = 5632.11$, $p < 0.0001$ for reduced model

$^6$ $\chi^2_{df=1} = 32.94$, $p < 0.0001$ for full model, and $\chi^2_{df=1} = 31.50$, $p < 0.0001$ for reduced model
Although there were less parameter values to estimate in the model with the shape parameter, all model comparison statistics including AIC, BIC and log likelihood ratio test favored the model without the shape parameter ($\chi^2 = 60.02, df = 3, p < 0.001$ for the reduced model with random effects). Moreover, in the model with covariates in the reduced model, age, gender, and the free or reduced-price lunch recipient status, when the shape parameter and the correlation between random effects were considered in the model, gender was not a significant predictor in the model. However the model without gender did not provide better explanation of data structure and was not preferred after comparing the model fit statistics. The summary of parameter estimates for the additional models are presented in Appendix A.
For repeated measures data in longitudinal setting, fitting mixed-effects linear models is a natural approach, since the data has a nested structure of observations within each subject. With brief introduction of previously developed models, motivating example of reading research data with excessive zeros collected in longitudinal setting was introduced in chapter 1. Various literature were reviewed in chapter 2 including mixed-effects linear models, mixed-effects and mixed-distribution models, and zero-inflated Poisson regression. Then, we propose extended mixed-effects and mixed-distribution models for count data to the zero-inflated model with a pair of correlated random effects. The basic properties and estimation procedures of the model were discussed. In chapter 3, we study the behavior of estimated parameters through simulation study. It is noted that both the number of measurement per subject and the number of subject were factors related the accuracy of estimated parameters. In models both with and without correlation on random effects, the unbiased estimated parameters were fitted.

At last, we explored varying impact of age, gender, socioecomonic status and race on educational achievement and the results are presented in chapter 4. With excessive zeros in the observed data, it was a natural to consider the data as a mixture of two distributions. In our model building process with reading research data, our proposed model was superior than any other model. Based on the model fit statistics, the model with a pair of correlated random effects was the most parsimonious model even without the shape parameter. The added pair of random effects was best explained between subject heterogeneity accounting possible correlation between the logistic part and the Poisson of the model. Moreover, it is very attractive that the interpretation of the estimated parameters are the same as in the standard logistic model and Poisson regression model in application of this model in other research areas.

As the socioeconomic status of the student’s family has been the strongest single predictor of educational achievement and academic outcomes and it has been well-established in previous literature reviews [62], our results are consistent with the previous findings. Our findings suggest that there are not any inherited unmeasured random effects on the initial reading ability comparing to significant between person variability on the mean scores after controlling for person level characteristics. Another finding is a significant correlation among the random effects meaning that a child with non-zero score on the test tend to get higher mean scores. Therefore, it implies that the introduction on alphabet letters might
be solely related to the start point of the reading education (age) and the socioeconomic background (FARL) of a student.

In this study, within person and between person variations were only concerns in terms of fitting and applying the proposed model to the reading research data. Although this study is limited to the analysis of effects of person level characteristics on their educational outcome, the neighborhood effects were suggested to be significant factors on educational outcomes of young children [27], [37]. Therefore, higher level hierarchical structure could also be considered in the model. In repeated measures data analysis, such as the reading research data used in chapter 4, one might suspect a serial correlation among the measurements, then the error structures are needed to be modeled adequately. Also, since the data were collected from 13 different schools, school differences could be also considered in the model.

For instance, zero-inflated Poisson model was extended to adapt multilevel hierarchical data structure by Lee et al. (2006)[35] as a multilevel zero-inflated Poisson regression for correlated count data. Specifically, three-level structure with repeated measures was considered. To account the correlated measurements within subject, autoregressive error structure was fitted in the model. Details on multilevel ZIP regression model, autoregressive model and three-level zero-inflated Poisson model with autoregressive correlation are following.

Multi-level ZIP regression model

There have been many further developments on ZIP regression to accommodate more complex data structure. For example, as introduced previously, Wang et al. (2002)[69] and Hur et al. (2002) [22] extended ZIP regression to dependent observations within groups, and Yau et al. (2001) [71] proposed a two part conditional model for repeated measured data analysis. Based on these models, Lee et al. (2006) [35] extended the zero-inflated Poisson regression to multi-level data. For example, the model considered the three-level hierarchical data where \( y_{ijk} \) denoted \( k^{th} \) measurement of the \( j^{th} \) subject within the \( i^{th} \) group \( (i = 1, 2, \ldots, m; j = 1, 2, \ldots, n_i \text{ and } k = 1, 2, \ldots, n_{ij}) \).

Three-level hierarchical zero-inflated Poisson model is specified as,

\[
\begin{align*}
\logit(p_{ijk}) &= \log \left( \frac{p_{ijk}}{1-p_{ijk}} \right) = X_{1ijk}^T \alpha_1 + b_{1i} + u_{1ij} \quad \text{and} \\
\log(\lambda_{ijk}) &= X_{2ijk}^T \alpha_2 + b_{2i} + u_{2ij} 
\end{align*}
\]

(5.1)

(5.2)

where \( X_{1ijk} \) and \( X_{2ijk} \) are covariates appearing in the logistic and Poisson part of the model, respectively, and \( \alpha_1 \) and \( \alpha_2 \) are the corresponding coefficients.

Three-level zero-inflated Poisson model with autoregressive correlation

A general AR\((p)\) model of a stochastic process \( \{e_{it}\} \) is defined by following relationship.

\[
\begin{align*}
e_{it} - \phi_1 e_{i(t-1)} - \phi_2 e_{i(t-2)} - \cdots - \phi_p e_{i(t-p)} &= C + a_{it} \\
(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) e_{it} &= C + a_{it}
\end{align*}
\]

(5.3)
where \( \{a_{it}, t = 0, 1, \ldots, \} \) is a white-noise process with mean zero and variance \( \sigma^2_{it} \).

The autocovariance function of \( \{e_{it}\} \) is

\[
R_{ik} = \text{Cov}(e_{it}, e_{i(t+k)}) = E(e_{it}e_{i(t+k)}) - \mu_i^2
\]

\[
= \phi_i1\gamma_{i(k-1)} + \phi_i2\gamma_{i(k-2)} + \cdots + \phi_ip\gamma_{i(k-p)} - \mu_i^2
\]

and the autocorrelation function of \( \{e_{it}\} \) is

\[
\rho_{ik} = \frac{R_{ik}}{\mu_i^2}, \ k = 0, \pm 1, \ldots \tag{5.5}
\]

At any time point, \( t \), when \( \{e_{it}\} \) has a same mean, \( \mu_i \), and variance, \( \sigma^2_i \), and \( \mathbf{R}_i(k) = \text{Cov}(e_{it}, e_{i(t+k)}) \) for any integer \( k \), the process is called a stationary process. To suffice such a condition, all the roots of \( \phi_i(B) = 1 - \phi_i1B - \phi_i2B^2 \cdots - \phi_ipB^p = 0 \) fall outside the unit circle. Suppose that a stochastic process \( \{e_{it}\} \) is stationary with \( E(e_{it}) = \mu_i \). In this case, the AR(\( p \)) process is causal and can be expressed in the form:

\[
e_{it} = \mu_i + \sum_{j=0}^{\infty} \psi_{ij}a_{i(t-j)}, \ t = 0, \pm 1, \pm 2, \ldots \tag{5.6}
\]

and \( E(e_{it}a_{i(t+k)}) = 0 \), for \( k > 0 \). For a causal process, future random noises and current or past observations are independent of each other.

If the time series \( \{e_{it}\} \) can be written in the form,

\[
e_{it} - \sum_{j=1}^{\infty} \pi_{ij}e_{i(t-j)} = C + a_{it} \tag{5.7}
\]

with \( \sum_{j=1}^{\infty} |\pi_{ij}| < \infty \), \( \{e_{it}\} \) is called an invertible process. An AR(\( p \)) process \( \{e_{it}\} \) is always invertible with \( \pi_{ij} = \phi_{ij} \) for \( 1 \leq j \leq p \) and \( \pi_{ij} = 0 \) for \( j > p \) as specified in equation 5.3.

A serial dependence correlation structure, such as autoregressive process for random effects, can be specified in the model. The random effects within the group is represented in \( \mathbf{b}_{11} \) and \( \mathbf{b}_{21} \), with random variations at subject level denoted by \( u_{1ij} \) and \( u_{2ij} \). The first order autoregressive correlation was considered in the model such that \( u_{1ij} \) and \( u_{2ij} \) are distributed as \( N(0, \sigma_{u1}B_{u1}(\rho_{u1})) \) and \( N(0, \sigma_{u2}B_{u2}(\rho_{u2})) \) with a block diagonal matrix with block size depending on the number of measurements of each subjects, and the autocorrelation parameter, \( \rho \), respectively. The penalized log-likelihood is,

\[
L = \sum_{ijk} \left( (1 - Y_{ijk}')X_{1ijk}'\alpha_1 + b_{11} + u_{1ij} \right) - log(1 + e^{X_{1ijk}'\alpha_1 + b_{11} + u_{1ij}})
\]

\[
- \frac{1}{2} (mlog(2\pi\sigma_{b_{1i}}^2) + \sigma_{b_{1i}}^{-2}b_{1i}^Tb_{1i} + Nlog(2\pi\sigma_{u1}^2) + \sigma_{u1}^{-2}u_{1i}^TB_{u1}^{-1}u_{1i})
\]

\[
+ \sum_{ijk} Y_{ijk}'(Y_{ijk}(X_{2ijk}'\alpha_2 + b_{2i} + u_{2ij}) - e^{X_{2ijk}'\alpha_2 + b_{2i} + u_{2ij}} - log(y_{ijk}!))
\]

\[
- \frac{1}{2} (mlog(2\pi\sigma_{b_{2i}}^2) + \sigma_{b_{2i}}^{-2}b_{2i}^Tb_{2i})
\]

53
where $Y'_{ijk}$ are defined as in (2.37). An estimation of parameter can be done using EM algorithm with penalized log-likelihood function[15],[26], while the variance components are estimated in conjunction with restricted maximum likelihood estimates method as proposed by McGilchrist (1994)[42]. However, there are concerns on the use of penalized quasi likelihood on the parameter estimation as written in section 2.4.2, the additional simulation study on model fitting need to be conducted. Also, we can try other estimation methods, such adaptive Gaussian quadrature or semiparametric estimation methods.

As mentioned in chapter 1, Dunson and Haseman (1999)[13] considered zero excessive count data with repeated measure as a transition model with conditional distribution of the responses based on the past response and covariates. Student learn based on what they already know. So with accumulated nature of learning process, one might try a transition model and compare the model fit.
APPENDIX A

ADDITIONAL RESULTS

The zero-inflated Poisson models were fitted as comparison of the model fitted in chapter 4.2. The model is specified as following,

\[
Pr(Y_i = y_i) = \begin{cases} 
(1 - p_i) + p_i e^{(-\lambda_i)}, & y_i = 0, \\
\frac{p_i e^{(-\lambda_i)} \lambda_i^{y_i}}{y_i!}, & y_i > 0, 0 \leq p_i \leq 1.
\end{cases}
\]  

(A.1)

where the parameters \( \lambda_i \) and \( p_i \) are defined by Then, we have,

\[
\text{Logit}(p_i) = \log \left( \frac{p_i}{1 - p_i} \right) = X_{1i} \alpha_1 \quad \text{and} \quad (A.2)
\]

\[
\text{Log}(\lambda_i) = X_{2i} \alpha_2 \quad \text{(A.3)}
\]

for covariate vectors \( X_{1i} \) and \( X_{2i} \).
Table A.1: Parameter estimates for Zero-Inflated Poisson Models.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Full Model</th>
<th>Reduced Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std Err</td>
</tr>
<tr>
<td><strong>Occurrence (Logistic)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-7.581</td>
<td>1.364</td>
</tr>
<tr>
<td>Age</td>
<td>1.809</td>
<td>0.259</td>
</tr>
<tr>
<td>Female</td>
<td>0.535</td>
<td>0.188</td>
</tr>
<tr>
<td>FARL</td>
<td>-0.587</td>
<td>0.238</td>
</tr>
<tr>
<td>Treatment</td>
<td>-0.030</td>
<td>0.184</td>
</tr>
<tr>
<td>Non-Black</td>
<td>0.177</td>
<td>0.215</td>
</tr>
<tr>
<td><strong>Intensity (Poisson)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.519</td>
<td>0.062</td>
</tr>
<tr>
<td>Age</td>
<td>0.341</td>
<td>0.011</td>
</tr>
<tr>
<td>Female</td>
<td>0.089</td>
<td>0.012</td>
</tr>
<tr>
<td>FARL</td>
<td>-0.149</td>
<td>0.014</td>
</tr>
<tr>
<td>Treatment</td>
<td>-0.037</td>
<td>0.012</td>
</tr>
<tr>
<td>Non-Black</td>
<td>-0.058</td>
<td>0.014</td>
</tr>
<tr>
<td>AIC</td>
<td>18596.36</td>
<td></td>
</tr>
<tr>
<td>AICC</td>
<td>18597.62</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>18658.72</td>
<td></td>
</tr>
<tr>
<td>-2 Loglikelihood</td>
<td>18572.36</td>
<td></td>
</tr>
</tbody>
</table>

Additional models with related $p_{ij}$ and $\lambda_{ij}$ are fitted and the results are presented. With the purpose of comparison of models with and without the random effects, we fit the model, without the random effects first, and the estimated parameters are summarized in Table A.2. The model is specified by replacing the equation (A.2) by

$$Logit(p_{ij}) = -\tau X_{2ij} \alpha_2$$ (A.4)

for covariate vectors $X_{2ij}$.

The models with random effects are specified by replacing the equation (4.2) by

$$Logit(p_{ij}) = -\tau X_{2ij} \alpha_2 + b_{2i}$$ (A.5)

for covariate vectors $X_{2ij}$.

As in the model with unrelated $p_{ij}$ and $\lambda_{ij}$, significantly reduced model fit statistics were observed in the models with random effects. This indicates that the models with random effects were preferred with significant variation among subjects. The quantile-quantile plots and histograms of random effects from were presented in Figure A.1. The assumption of normally distributed random effects were satisfied.
Table A.2: Parameter estimates for zero-inflated Poisson models with the shape parameter $\tau$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Full Model</th>
<th>Reduced Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std Err</td>
</tr>
<tr>
<td>Occurrence (Logistic)</td>
<td>$\tau$</td>
<td>-0.614</td>
</tr>
<tr>
<td>Intensity (Poisson)</td>
<td>Intercept</td>
<td>1.503</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>0.344</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>FARL</td>
<td>-0.150</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>-0.037</td>
</tr>
<tr>
<td></td>
<td>Non-Black</td>
<td>-0.057</td>
</tr>
<tr>
<td>AIC</td>
<td>18650.80</td>
<td>18673.02</td>
</tr>
<tr>
<td>AICC</td>
<td>18650.89</td>
<td>18673.07</td>
</tr>
<tr>
<td>BIC</td>
<td>18686.59</td>
<td>18698.58</td>
</tr>
<tr>
<td>-2 Loglikelihood</td>
<td>18636.80</td>
<td>18663.02</td>
</tr>
</tbody>
</table>

Table A.3: Parameter estimates of the full model with the shape parameter $\tau$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Uncorrelated Random Effects</th>
<th>Correlated Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std Err</td>
</tr>
<tr>
<td>Occurrence (Logistic)</td>
<td>$\tau$</td>
<td>-0.752</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_1$</td>
<td>0.514</td>
</tr>
<tr>
<td>Intensity (Poisson)</td>
<td>Intercept</td>
<td>-1.974</td>
</tr>
<tr>
<td></td>
<td>age</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>FARL</td>
<td>-0.248</td>
</tr>
<tr>
<td></td>
<td>Condition</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>-0.075</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_2$</td>
<td>0.408</td>
</tr>
<tr>
<td></td>
<td>$\rho \sigma_1 \sigma_2$</td>
<td>-</td>
</tr>
<tr>
<td>AIC</td>
<td>13047.88</td>
<td>13016.94</td>
</tr>
<tr>
<td>AICC</td>
<td>13048.03</td>
<td>13017.12</td>
</tr>
<tr>
<td>BIC</td>
<td>13084.00</td>
<td>13057.07</td>
</tr>
<tr>
<td>-2 Loglikelihood</td>
<td>13029.88</td>
<td>12996.94</td>
</tr>
<tr>
<td>Diff in -2ll</td>
<td>32.94</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>
Table A.4: Parameter estimates of the reduced model with the shape parameter $\tau$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Uncorrelated Random Effects</th>
<th>Correlated Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$</td>
<td>Estimate</td>
<td>Std Err</td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>0.509</td>
<td>0.371</td>
</tr>
<tr>
<td>$\sigma_2^2$</td>
<td>0.509</td>
<td>0.371</td>
</tr>
</tbody>
</table>

### Occurrence (Logistic)

- Intercept: $-1.984 \pm 0.122$, $p < .0001$
- age: $0.944 \pm 0.019$, $p < .0001$
- Gender: $0.163 \pm 0.063$, $p < .0001$
- FARL: $-0.290 \pm 0.067$, $p < .0001$
- $\rho \sigma_1 \sigma_2$: -

### Intensity (Poisson)

- Intercept: $-1.980 \pm 0.123$, $p < .0001$
- age: $0.944 \pm 0.019$, $p < .0001$
- Gender: $0.128 \pm 0.066$, $p < .0001$
- FARL: $-0.263 \pm 0.070$, $p < .0001$
- $\rho \sigma_1 \sigma_2$: $0.367 \pm 0.069$, $p < .0001$

<table>
<thead>
<tr>
<th>Metric</th>
<th>Uncorrelated Random Effects</th>
<th>Correlated Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>13044.91</td>
<td>13015.41</td>
</tr>
<tr>
<td>AICC</td>
<td>13045.00</td>
<td>13015.53</td>
</tr>
<tr>
<td>BIC</td>
<td>13073.00</td>
<td>13047.52</td>
</tr>
<tr>
<td>-2 Loglikelihood</td>
<td>13030.91</td>
<td>12999.41</td>
</tr>
<tr>
<td>Diff in -2ll</td>
<td>31.50</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>
Figure A.1: Histograms and qq plots of random effects from the model with correlate random effects and $\tau$. 
APPENDIX B
SAS CODES

libname aims 'Y:\Documents\Thesis\Data' ;

data aims.cohort3s;
set aims.cohort3;
if aims=0 then aimsn=0; /* Define r_ij*/
else aimsn=1;
keep aims id age gender farl condition black aimsn;
run;

/* Full Model */
ods output ParameterEstimates=aims.parmsg1full(rename=Parameter=Name) modelfit=aims.modelfitbfull;

proc genmod data=aims.cohort3s descending;
model aimsn=age gender farl condition black /dist=binomial;
run;

data aims.random1;
Parameter='u1v'; Name='Var(Rndm Effect)'; Estimate=0.5;
run;

data aims.random2;
Parameter='u2v'; Name='Var(Rndm Effect)'; Estimate=0.5;
run;

data aims.u12v;
Parameter='u12v'; Name='Cov(Rndm Effect)'; Estimate=0.5;
run;

data aims.start1full;
format Name $20. Parameter $20.;
set aims.parmsg1full aims.random1;
keep estimate name Parameter;
if Name='Intercept' then Parameter = 'alpha10';
if Name='age' then Parameter = 'alpha11';
if Name='Gender' then Parameter = 'alpha12';
if Name='FARL' then Parameter = 'alpha13';
if Name='Condition' then Parameter = 'alpha14';
if Name='black' then Parameter = 'alpha15';
if Name='Scale' then delete;

60
ods output ParameterEstimates=aims.parmsg2full(rename=Parameter=Name) modelfit=aims.modelfitinfull;

proc genmod data=aims.cohort3s(where=(aimsn=1));
model aims=age gender farl condition black /dist=poisson;
run;

data aims.start2full;
format Name $20. Parameter $20.;
set aims.parmsg2full aims.random2;
keep estimate name Parameter;
if Name='Intercept' then Parameter = 'alpha20';
if Name='age' then Parameter = 'alpha21';
if Name='Gender' then Parameter = 'alpha22';
if Name='FARL' then Parameter = 'alpha23';
if Name='Condition' then Parameter = 'alpha24';
if Name='black' then Parameter = 'alpha25';
if Name='Scale' then delete;
run;

data aims.startfull;
set aims.startfull aims.start2full;
run;

data aims.startfullcorr(keep=parameter estimate Name);
set aims.parmsf1full aims.rho;
run;

ods output ParameterEstimates=aims.parmsf1full FitStatistics=aims.fitf1full
ConvergenceStatus=aims.convf1full;
proc nlmixed data=aims.cohort3s ;
parms/data=aims.startfull;
bounds u1v>=0;
bounds u2v>=0;
linpinfl = alpha10 +alpha11*age + alpha12*gender + alpha13*farl + alpha14*condition
 + alpha15*black + u1;
infprob = 1/(1+exp(-linpinfl));
lambda = exp(alpha20 +alpha21*age + alpha22*gender + alpha23*farl + alpha24*condition
 + alpha25*black + u2);
if aims=0 then
  ll = log((1-infprob) + infprob*exp(-lambda));
else ll = log(infprob)- lambda + aims*log(lambda) - lgamma(aims + 1);
model aims ~ general(ll);
random u1 u2 ~ normal([0,0],[u1v,0,u2v]) subject=id;
* random u1 u2 ~ normal([0,0],[u1v,u12v,u2v]) subject=id;
run;
/*************** Models with the shape parameter ***************/

data aims.tau; Parameter='tau'; Name='tau'; Estimate=.5;
run;

data aims.startfulltau;
set aims.start2full aims.random1 aims.tau; /* get starting values from previous models */
run;

data aims.startfulltaucorr(keep= parameter estimate);
set aims.fulltau aims.rho ;
run;

ods output ParameterEstimates=aims.fulltaucorr FitStatistics=aims.fitfulltaucorr;

proc nlmixed data=aims.cohort3s qpoints=25;
parms/data=aims.startfulltaucorr;

bounds u1v>=0;
bounds u2v>=0;

linpinfl = -tau*(alpha20 +alpha21*age + alpha22*gender + alpha23*farl + alpha24*condition + alpha25*black) + u1;
infprob = 1/(1+exp(-linpinfl));

lambda = exp(alpha20 +alpha21*age + alpha22*gender + alpha23*farl + alpha24*condition + alpha25*black + u2);

if aims=0 then ll = log((1-infprob) + infprob*exp(-lambda));
else ll = log(infprob)- lambda + aims*log(lambda) - lgamma(aims + 1);

model aims ~ general(ll);
*random u1 u2 ~ normal([0,0],[u1v,0,u2v]) subject=id; /* uncorrelated r.e.*/
random u1 u2 ~ normal([0,0],[u1v,u12v,u2v]) subject=id;
run;
APPENDIX C

IRB APPROVAL MEMORANDUMS

Office of the Vice President For Research
Human Subjects Committee
Tallahassee, Florida 32306-2742
(850) 644-8673  FAX (850) 644-4392

APPROVAL MEMORANDUM

Date: 10/16/2009

To: Jihyung Shin

Address: 4330
Dept.: ARTS AND SCIENCES, DEANS OFFICE

From: Thomas L. Jacobson, Chair

Re: Use of Human Subjects in Research
Development of Growth model with non-normal error structure and application on reading research data.

The application that you submitted to this office in regard to the use of human subjects in the research proposal referenced above has been reviewed by the Human Subjects Committee at its meeting on 10/14/2009. Your project was approved by the Committee.

The Human Subjects Committee has not evaluated your proposal for scientific merit, except to weigh the risk to the human participants and the aspects of the proposal related to potential risk and benefit. This approval does not replace any departmental or other approvals, which may be required.

If you submitted a proposed consent form with your application, the approved stamped consent form is attached to this approval notice. Only the stamped version of the consent
form may be used in recruiting research subjects.

If the project has not been completed by 10/13/2010 you must request a renewal of approval for continuation of the project. As a courtesy, a renewal notice will be sent to you prior to your expiration date; however, it is your responsibility as the Principal Investigator to timely request renewal of your approval from the Committee.

You are advised that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition, federal regulations require that the Principal Investigator promptly report, in writing any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor is reminded that he/she is responsible for being informed concerning research projects involving human subjects in the department, and should review protocols as often as needed to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

This institution has an Assurance on file with the Office for Human Research Protection. The Assurance Number is IRB00000446.

Cc: Xufeng Niu, Advisor
HSC No. 2009.3341
RE-APPROVAL MEMORANDUM

Date: 9/16/2010

To: Jihyung Shin

Address: 4330
Dept.: ARTS AND SCIENCES, DEANS OFFICE

From: Thomas L. Jacobson, Chair

Re: Re-approval of Use of Human subjects in Research
Development of Growth model with non-normal error structure and application on reading research data.

Your request to continue the research project listed above involving human subjects has been approved by the Human Subjects Committee. If your project has not been completed by 9/7/2011, you are must request renewed approval by the Committee.

If you submitted a proposed consent form with your renewal request, the approved stamped consent form is attached to this re-approval notice. Only the stamped version of the consent form may be used in recruiting of research subjects. You are reminded that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition, federal regulations require that the Principal Investigator promptly report in writing, any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor are reminded of their responsibility for being informed concerning research projects involving human subjects in their department. They are advised to review the protocols as often as necessary to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

Cc: Xufeng Niu, Advisor
HSC No. 2010.4837
Office of the Vice President For Research  
Human Subjects Committee  
Tallahassee, Florida 32306-2742  
(850) 644-8673, FAX (850) 644-4392  

RE-APPROVAL MEMORANDUM

Date: 8/12/2011

To: Jihyung Shin

Address: 4330  
Dept.: ARTS AND SCIENCES, DEANS OFFICE

From: Thomas L. Jacobson, Chair

Re: Re-approval of Use of Human subjects in Research
Development of Growth model with non-normal error structure and application on reading research data.

Your request to continue the research project listed above involving human subjects has been approved by the Human Subjects Committee. If your project has not been completed by 8/8/2012, you are must request renewed approval by the Committee.

If you submitted a proposed consent form with your renewal request, the approved stamped consent form is attached to this re-approval notice. Only the stamped version of the consent form may be used in recruiting of research subjects. You are reminded that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition, federal regulations require that the Principal Investigator promptly report in writing, any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor are reminded of their responsibility for being informed concerning research projects involving human subjects in their department. They are advised to review the protocols as often as necessary to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

Cc: Xufeng Niu, Advisor  
HSC No. 2011.6679
REFERENCES


69


BIOGRAPHICAL SKETCH

Jihyung Shin was born in Seoul, Korea. She received her first Bachelor’s degree in Journalism and Mass Communication with concentration in Advertising from the University of South Carolina, then completed another Bachelor’s degree in Statistics from the Florida State University. Then, she continued her graduated programs and received Masters degree. She defended her PhD dissertation in Spring of 2012.