A CLASS OF MIXED-DISTRIBUTION MODELS WITH APPLICATIONS IN FINANCIAL DATA ANALYSIS

Name: Anqi Tang
Department: Department of Statistics
Major Professor: Xufeng Niu
Degree: Doctor of Philosophy
Term Degree Awarded: Spring, 2011

Statisticians often encounter data in the form of a combination of discrete and continuous outcomes. A special case is zero-inflated longitudinal data where the response variable has a large portion of zeros. These data exhibit correlation because observations are obtained on the same subjects over time. In this dissertation, we propose a two-part mixed distribution model to model zero-inflated longitudinal data. The first part of the model is a logistic regression model that models the probability of nonzero response; the other part is a linear model that models the mean response given that the outcomes are not zeros. Random effects with AR(1) covariance structure are introduced into the both parts of the model to allow serial correlation and subject specific effect.

Estimating the two-part model is challenging because of high dimensional integration necessary to obtain the maximum likelihood estimates. We propose a Monte Carlo EM algorithm for estimating the maximum likelihood estimates of parameters. Through simulation study, we demonstrate the good performance of the MCEM method in parameter and standard error estimation.

To illustrate, we apply the two-part model with correlated random effects and the model
with autoregressive random effects to executive compensation data to investigate potential determinants of CEO stock option grants.
A CLASS OF MIXED-DISTRIBUTION MODELS WITH APPLICATIONS IN FINANCIAL DATA ANALYSIS

By

ANQI TANG

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, 2011
The members of the Committee approve the Dissertation of Anqi Tang defended on Mar 16, 2011.

Xufeng Niu
Professor Directing Dissertation

Yingmei Cheng
University Representative

Wei Wu
Committee Member

Fred Huffer
Committee Member

Approved:

Dan McGee, Chair
Department of Statistics

Joseph Travis, Dean, College of Arts and Sciences

The Office of Graduate Studies has verified and approved the above named committee members.
The members of the Committee approve the Dissertation of Anqi Tang defended on Mar 16, 2011.

Xufeng Niu  
Professor Directing Dissertation

Yingmei Cheng  
Outside Committee Member

Wei Wu  
Committee Member

Fred Huffer  
Committee Member

The Office of Graduate Studies has verified and approved the above named committee members.
Dedicated to my parents
ACKNOWLEDGEMENTS

I would first like to acknowledge my gratitude to my major professor Dr. Xufeng Niu for his advice guidance and encouragement. I am thankful for his time and ideas which made my research experience fulfilling and productive.

I would like to acknowledge Professor Utset for his advice and for his crucial contribution to my Ph.D. work. His originality inspired and motivated me to find and solve the statistical problems that later became my dissertation. I benefited a lot from the long discussions I had with him.

I am also grateful to Dr. Fred Huffer, Dr. Wei Wu and Dr. Yimei Cheng, who served as my committee members. I appreciate their time and support. Without their help, I would not have accomplished what I have been able to accomplish. My thanks also goes to Dr. Dan McGee, Dr. Anuj Srivastava, Dr. Debajyoti Sinha, Dr. Florentina Bunea and Dr. Jayaram Sethuraman. I am fortunate that I had taken their classes and had received so much help from them.

My sincere thanks also goes to the current staff members and those who left the Department of Statistics. I would like to thank Ms. Pam McGhee, Ms Jennifer Rivera, Ms Megan Trautman, Mr. Chauncey Richburg, Ms. Dominique Grant, Ms. Evangelous Robinson and Mr. James Stricherz for helping me whenever I needed assistance.

I owe my gratitude to the Professors and staff members in the Department of Family Medicine and Rural Health at Florida State University especially to Dr. Askar Chukmahtov, Dr. Leslie Beitsch and Dr. Henry Carreta. It has been an unforgettable experience working with them.

My dear friends helped me stay focused thoughout my Ph.D. years. Their help and care helped me overcome difficulties I encountered in my graduate study. I would sincerely thank my friends Shuva, Wenting, Yu, Xiaoyun, Wei, Vernon and all my colleagues in the Department of Statistics at Florida State University.

Last but not the least, none of these would have been possible without the love and
support from my family. My family has always been a constant source of love, support and encouragement. I have to thank my parents for being most supportive and encouraging.

— Anqi Tang
# TABLE OF CONTENTS

List of Tables ................................................................. vii

List of Figures ................................................................. ix

Abstract ................................................................................ xi

1. Motivation and Literature Review ........................................ 1
   1.1 Longitudinal Data Analysis ............................................. 1
   1.2 Zero-Inflated Longitudinal Data ...................................... 2
   1.3 Literature Review .......................................................... 2
   1.4 Overview of the Study ..................................................... 5

2. Models and Properties ....................................................... 6
   2.1 Linear Mixed-Effects Model ............................................. 6
   2.2 Mixed Distribution Models ............................................. 15
   2.3 Mixed-Distribution Models with Random Effects and Serial Correlation .................................................. 20

3. Simulation Study ............................................................. 35
   3.1 Simulation Set-up .......................................................... 35
   3.2 Simulation Results ........................................................ 36

4. Applications ................................................................. 43
   4.1 APPLICATION TO FINANCE .......................................... 43
   4.2 Analysis ........................................................................ 49

5. Discussion and Future Work ................................................ 83

APPENDICES ........................................................................... 87

A. Results from a simplified model .......................................... 87

B. Overall Effects of Covariates ............................................. 89

REFERENCES .......................................................................... 91

BIOGRAPHICAL SKETCH ....................................................... 94
LIST OF TABLES

3.1 Dataset 1: the Two-part mixed model with $\phi_1=0$ ........................................... 38
3.2 Simulation Results for the Two-part mixed model with $\phi_1=0.5$ ...................... 39
3.3 Simulation Results for the Two-part mixed model (Tooze) with correlated random effects ................................................................. 40
3.4 Simulation Results: Set One ................................................................. 40
3.5 Simulation Results: Set Two ................................................................. 41
3.6 Simulation Results: Set Three ............................................................... 41
3.7 Simulation Results: Set Four ................................................................. 42
3.8 Simulation Results: Set Five ................................................................. 42
4.1 Frequency distribution of restatement announcement by year .......................... 44
4.2 Distribution of restating firms by year ....................................................... 45
4.3 Frequency of restatement announcements .................................................. 45
4.4 Most Frequent Restatement Reasons ........................................................ 46
4.5 Distribution of restating firms across industry ........................................... 48
4.6 Components of CEO compensations in misreported years and normal years .... 53
4.7 Descriptive Summary of Firm Characteristics ............................................. 57
4.8 Parameter Estimates of the Full Model: Correlated Random Intercept ........... 62
4.9 Parameter Estimates of the Reduced Model: Correlated Random Intercept .. 64
4.10 Results from fitting the Full Model ......................................................... 70
4.11 Results from fitting a Reduced Model with fewer covariates ....................... 75
4.12 Results from the final Model with AR(1) random effect ............................ 77
LIST OF FIGURES

2.1 Distribution of log-transformed option ........................................... 16
4.1 Estimated Value of in-the-money unexercised exercisable options ........ 52
4.2 Random Sample ................................................................. 54
4.3 Market to Book Ratio by year .................................................. 56
4.4 Distribution of log-transformed option ........................................... 59
4.5 Distribution of arcsine-transformed percentage of option: zeros excluded . 60
4.6 Normal qq Plot: random effects ................................................. 65
4.7 Histogram: random effects ....................................................... 66
4.8 Subject-Specific Residuals ....................................................... 66
4.9 Population-Averaged Residuals .................................................. 67
4.10 Random Effects in the first part of the Model without AR(1) Effect ...... 71
4.11 Random Effects second part of the Model without AR(1) Effect .......... 71
4.12 Random Effects in the first part of the Model with AR(1) Effect .......... 72
4.13 Random Effects in the second part of the Model with AR(1) Effect ..... 72
4.14 Residual Plots of the Model without AR(1) Effect ......................... 73
4.15 Residual Plots of the Model with AR(1) Effect ............................. 73
4.16 Random Effects in the first part of the Model without AR(1) Effects .... 78
4.17 Random Effects in the second part of the Model without AR(1) Effects . 78
4.18 Residual Plots of the Model without AR(1) Effect: further reduced model 79
4.19 Residual Histograms of the Model without AR(1) Effect: further reduced model ................................. 79
4.20 Residual QQ-Plots of the Model without AR(1) Effect: further reduced model 80
4.21 Random Effects in the first part of the Model with AR(1) Effects: further reduced model 80
4.22 Random Effects in the second part of the Model with AR(1) Effects: further reduced model 81
4.23 Residual Plots of the Model with AR(1) Effect: further reduced model 81
4.24 Residual Histograms of the Model with AR(1) Effect: further reduced model 82
4.25 Residual QQ-Plots of the Model with AR(1) Effect: further reduced model 82
ABSTRACT

Statisticians often encounter data in the form of a combination of discrete and continuous outcomes. A special case is zero-inflated longitudinal data where the response variable has a large portion of zeros. These data exhibit correlation because observations are obtained on the same subjects over time. In this dissertation, we propose a two-part mixed distribution model to model zero-inflated longitudinal data. The first part of the model is a logistic regression model that models the probability of nonzero response; the other part is a linear model that models the mean response given that the outcomes are not zeros. Random effects with AR(1) covariance structure are introduced into the both parts of the model to allow serial correlation and subject specific effect.

Estimating the two-part model is challenging because of high dimensional integration necessary to obtain the maximum likelihood estimates. We propose a Monte Carlo EM algorithm for estimating the maximum likelihood estimates of parameters. Through simulation study, we demonstrate the good performance of the MCEM method in parameter and standard error estimation.

To illustrate, we apply the two-part model with correlated random effects and the model with autoregressive random effects to executive compensation data to investigate potential determinants of CEO stock option grants.
CHAPTER 1

Motivation and Literature Review

1.1 Longitudinal Data Analysis

Longitudinal data models have become increasingly popular in modern empirical and theoretical research. A data set that contains observation on a single individual over time is called time series. A data set that contains observations on a group of individuals at a given time point is cross-sectional. A Longitudinal or longitudinal data set is one that consists of repeated observations on the same cross section of individuals over a given time span. For instance, two of the famous Longitudinal data sets in the financial field are the Lifelong Labor Market Database and the University of Michigan’s Longitudinal Study of Income Dynamics (PSID).

Longitudinal data analysis has enabled researchers to undertake studies in a variety of fields. In economics, Longitudinal data is usually used to study firm characteristics and wage. In health related research, it is used to study characteristics of group of people followed over time. It is also widely used in psychology and political science etc. Longitudinal data usually allow researches to solve some problems that cannot be answered using time-series data or cross-sectional data alone. Longitudinal data give researchers more observations, therefore, higher degrees of freedoms and higher efficiency of estimates.

The goal of many empirical studies is to establish a causal relationship between independent variables and a dependent variable, then further estimate the effect. With Longitudinal data, researches can perform regression analysis with both spatial dimension and temporal dimension.
1.2 Zero-Inflated Longitudinal Data

Zero-inflation problem is common in many areas. For example, health care expenditure data often present two separate parts, one part consists of all zeros (zero expense) and a part consists of continuous positive numbers. Another example is the data that motivate this study: executive compensation. There are several components in executive compensation: salary, bonus, option and etc. Options are performance-based compensation, therefore, in theory, executives who receive more options should work harder. In order to examine if options work effectively, many researchers conduct studies to investigate the relationship between the amount of option grants CEOs received each year and their company’s performance. Most commonly applied statistical models in some studies include generalized linear model or simple linear regression model. One problem overlooked in these studies is that executives do not receive any options in many fiscal years, therefore, the normality assumption for linear models do not hold anymore. In some other studies, Tobit model[34] is applied because of the excessive amount of zeros in CEO’s stock options data. As a special case of censored regression model, Tobit model solves the unobservability problem by introducing a latent variable into the model. However, the zeros in CEO stock option grants result from the choice of the board of director or CEOs themselves. These zeros are observed true values, therefore, Tobit model may not be applicable for our study.

The purpose of this study is to propose a statistical model that models CEO stock options in two parts: the first part models the probability of granting options and the second part examines economic determinants associated with the positive amount of options granted given that there are options granted that year.

1.3 Literature Review

In the past few decades, the growth in American chief executive officer (CEO) compensation has become increasingly controversial and attracted close attention from media, investors, politicians and researchers. CEO pay has increased dramatically in comparison to the growth rate of average workers’ wage. Top CEOs have faced criticism for their excessive pay.

Followed the Enron scandal in 2001, an extensive amount of studies have been focusing on CEO’s incentives to manipulate their pay. Who sets CEO pay? How is CEO pay determined? Are CEO’s compensation package efficient? Do CEOs get paid as much when the companies
perform poorly? What gives CEOs incentive to engage in fraudulent activities? When do CEOs commit fraud? These are the questions that concern most of the people. To answer these questions, let’s first take a closer look at components of CEO compensation.

CEO’s pay contract usually includes the following: base salary, bonus, long-term incentive plan, option grant and restricted stock. Base salary for CEOs is usually determined through competitive benchmarking generally based on industry salary surveys. The surveys report certain quantiles, structure of base salary usually adjusted for firm size and industry[24]. In the process of reviewing how base salary was determined, Murphy found the surveys do not reflect certain factors that are considered relevant to pay level. For example, CEO’s experience, skills, education, and more importantly, firm performance. When these factors are not incorporated into salary, some part of CEO salary will not be explained by firm performance, in other words, CEOs will be rewarded for luck[5]. Moreover, using a peer group does not always lead to optimal CEO pay.

Base salary is one of the key components in CEO pay package. First, it makes up about 50% of CEO’s total compensation. Second, it is the fixed component in CEO pay contract. Risk-averse executives will prefer more salary than other type of compensation such as option. Third, other components of CEO pay usually depend on base salary[24].

Bonus is another important component of executive compensation. Although it does not make the biggest proportion of pay contract, a common hypothesis is that bonus is one rationale for earnings manipulations as bonus is usually tie to company short-term performance [10]. The actual amount of bonus awarded each year depends on certain achievement of performance. According to the executive compensation study by Kevin Murphy, executive bonus plans can be categorized in terms of three components: performance measure, performance standards, and the structure of the pay-performance relation. Most companies do not pay executives bonus unless a minimum targeted firm performance is achieved. Among all performance measures, operating income is the most common one and most companies use single performance measure. Murphy listed the following performance-standard determination process given a certain type of performance measure used by the company: “Budget”, “Prior-Year”, “Peer Group”, “Tuneless Standards” and “Cost of Capital”. The target bonus is usually expressed as a multiple of base salary and determined by companies’ compensation committee. It has been proved that on average 80% of the bonus can be explained by company performance (Hay, 1991), however, a portion of bonus
still depends on individual performance, which is mostly difficult to observe and evaluate. Moreover, target bonus is usually tied to base salary, which is not always explained by firm performance. Therefore, it is not always clear if the bonus scheme works effectively. The incentive bonus creates will be discussed in the next section. We will examine the sum of base salary and annual bonus since they together are the “cash compensation” each executive gets.

Other than the “Cash Component”, stock options make up a large proportion of executive pay. Stock options are contract that allow the recipient to buy stock at exercise price that usually equals the market price as of the day of grant. Most stock options mature in ten years and typically become exercisable (vested) in 3 years. The rapid growth in stock options in executive compensation has drawn close attention from researchers. Core and Guay (2002) found that changes in the value of executive stock and stock options were as much as fifty times as large as annual changes in cash compensation by the late 1990s. Yermack (1994) reported that the enormous growth in U.S. executives’ compensation resulted largely from stock options. There are two schools of opinions about growth in stock options. Many researchers argue that it is necessary to adopt stock options to align managers’ incentives with shareholders’ interest [32]. Linking CEO’s pay to the stock market has been considered a natural mean to mitigate conflict between managers and shareholders. However, late studies suggest that stock options do not work effectively.

Studies have shown that CEOs compensation are not entirely tied to their performance[5]. Articles show that CEO get bonus even when their companies perform poorly.[13] This shows that the portions of CEOs’ compensation that are tied to annual performance may also be tied to goals. For example, signaling to the market that the company is performing well, or retaining the current CEO. We are interested in the determinants of CEOs’ performance-based compensation, in particular, stock options.

Financial restatement is the release to the public of previously issued financial statement amended with new information. Restatements do not necessarily result from fraud, however, a restatement usually has negative impact on a company which include negative publicity, negative market reaction, profit loss, or downgrades in credit ratings. As the number of restatement announcement increases significantly over recent years, Government Accountability Office (GAO) released two reports on financial restatements result from fraud and accounting and audit issues in year 2002 and 2006[25][26]. The reports discussed the
reasons for restatements and the impact of restatements on restating company’s stock prices. Researchers generally agree that investors react negatively to restatement\cite{4,26} and we expect it to affect firm value. And restatements also force a lot of companies to replace their CEOs or CFOs, according to report, the portion of CEOs at restating companies who resigned rose from 8.8 percent in 2005 to 11.3 percent in 2006; the percentage who were dismissed jumped from 0.9 percent to 1.2 percent\cite{1}. We are interested in examining the differences in option grants between restating companies and non-restating companies and how restatements influence CEO compensation.

1.4 Overview of the Study

The remainder of the study is as follows: Chapter 2 provides a literature review of statistical models used in longitudinal data analysis. These methods include classic mixed model and a two-part mixed-distribution model with correlated random effects. We review different covariance structure and later in Chapter 2, we propose a two-part model with AR(1) random effects that allows serial correlation within subjects. We also discuss MCEM algorithm and model estimation in detail.

In Chapter 3, we present a simulation study. We generate data from the two-part model with AR(1) random effects and examine the estimates obtained from MCEM method. We also compare the results with estimates obtained from two-part model without serial correlation. The executive compensation data is analyzed in Chapter 4. We focus on modeling stock options newly granted in each fiscal year and present descriptive statistics. At the end of this chapter, we fit Tooze’s two-part model and the model with autoregressive random effects to the executive compensation data. We compare parameter estimates and their interpretation obtained from these two models and model goodness-of-fit.
CHAPTER 2
Models and Properties

2.1 Linear Mixed-Effects Model

2.1.1 Model Formulation

To formulate linear mixed-effect models for longitudinal data, we assume that for each subject $i$, $y_{it}$ is the response variable and $X_i = \{1, x_{1i}, \cdots, x_{Ki}\}$ are the predictor variables of the $i^{th}$ subject at time $t$. We can write the model as following:

$$
\begin{align*}
y_{it} &= \alpha_0 + \alpha_1 x_{1it} + \alpha_2 x_{2it} + \cdots + \alpha_K x_{Kit} + b_{i1} z_{1it} + b_{i2} z_{2it} + \cdots + b_{iQ} z_{Qit} + \epsilon_{it}, \\
i &= 1, \ldots, N; t = 1, \ldots, n_i
\end{align*}
$$

We can rewrite the equation in the following form:

$$
\begin{align*}
Y_i &= X_i \alpha + Z_i b_i + \epsilon_j, \\
i &= 1, \cdots, N,
\end{align*}
$$

i.e.,

$$
\begin{pmatrix}
y_{i1} \\
\vdots \\
y_{in_j}
\end{pmatrix} =
\begin{pmatrix}
1 & x_{1i1} & \cdots & x_{Ki1} \\
\vdots & \vdots & \ddots & \vdots \\
1 & x_{1in_j} & \cdots & x_{Kin_j}
\end{pmatrix}
\begin{pmatrix}
\alpha_0 \\
\vdots \\
\alpha_K
\end{pmatrix}
+ \begin{pmatrix}
z_{1i1} & \cdots & z_{Qi1} \\
\vdots & \ddots & \vdots \\
z_{1in_j} & \cdots & z_{Qin_j}
\end{pmatrix}
\begin{pmatrix}
b_{i1} \\
\vdots \\
b_{iQ}
\end{pmatrix}
+ \begin{pmatrix}
\epsilon_{i1} \\
\vdots \\
\epsilon_{in_j}
\end{pmatrix},
$$

where $Z_i$'s are known covariates, $\alpha = \{\alpha_1, \ldots, \alpha_K\}'$ are fixed-effect coefficients, which stay the same over time and cross section; and $b_i = \{b_{i1}, \ldots, b_{iQ}\}$ are random-effect coefficients, which depend on $i$. The $b_i$'s are also considered as subject-specific effect that remain constant over time within each subject. Furthermore, $b_1, \cdots, b_n$ are assumed to be independent and distributed as $N(0, D)$, $N$ represents Normal; $b_i$ and $b_j$ are independent. The $\epsilon_i$'s are assumed to be independent with distribution $N(0, \Sigma_j)$; $\{\epsilon_1, \cdots, \epsilon_N\}$ are independent of $\{b_1, \cdots, b_n\}$. A variable is considered as a fixed effect variable if the effect that is specific to individual
stays constant over time, or the effect specific to every time period is the same for all subjects. The fixed-effect model is viewed as one in which the researchers draw inference for those levels in the sample. If the levels of a variable included in the study represent a random sample from the studied population, the effect of the variable should be treated as a random effect.

\( \mathbf{Y}_i \) is the response vector for subject \( i \). Given the assumptions we make, it can be easily shown that \( E(\mathbf{Y}_i) = \mathbf{X}_i \alpha \) and \( \text{Var}(\mathbf{Y}_i) = V_i = Z_i D Z_i' + \Sigma_i \). This implies the marginal model

\[
\mathbf{Y}_i \sim N(\mathbf{X}_i \alpha, Z_i D Z_i' + \Sigma_i). \tag{2.3}
\]

Conditional on the random effect \( \mathbf{b}_i \), \( \mathbf{Y}_i \) is normally distributed with mean \( \mathbf{X}_i \alpha + Z_i \mathbf{b}_i \) and covariance matrix \( \Sigma_i \). Let \( f(\mathbf{b}_i) \) be the density function of \( \mathbf{b}_i \). We can write the marginal density of \( \mathbf{Y}_i \) as:

\[
f(\mathbf{y}_i) = \int f(\mathbf{y}_i | \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i.
\]

Let \( \theta \) denote the vector of parameters in the mixed linear model. \( \theta = (\alpha', \gamma')' \), where \( \gamma \) is the vector of all parameters in the covariance matrices. Let \( \Theta \) denote the parameter space for \( \theta \). The traditional approach to draw inference based on the sample is to maximize the marginal likelihood function over the parameter space \( \Theta \). The marginal likelihood function for the general linear mixed model is:

\[
L(\theta | \mathbf{X}) = \prod_{i=1}^{N} \{(2\pi)^{-n_i/2} |V_i(\gamma)|^{-1/2} \exp(-\frac{1}{2}(\mathbf{Y}_i - \mathbf{X}_i \alpha)'V_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i \alpha))\}. \tag{2.4}
\]

When \( \gamma \) is known, the maximum likelihood estimate of fixed-effects coefficient \( \alpha \) can be easily derived as follows [21]:

\[
l(\alpha) = \sum_{i=1}^{N} -\frac{n_i}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{N} \ln|V_i(\gamma)| + \sum_{i=1}^{N} (-\frac{1}{2}(\mathbf{Y}_i - \mathbf{X}_i \alpha)'V_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i \alpha)). \tag{2.5}
\]

The equation \( \frac{\partial l}{\partial \alpha} = 0 \) reduces to

\[
\sum_{i=1}^{N} \mathbf{Y}_i'V_i^{-1}\mathbf{X}_i = \alpha' \sum_{i=1}^{N} \mathbf{X}_i'V_i^{-1}\mathbf{X}_i.
\]
Therefore,
\[
\hat{\alpha} = \left( \sum_{i=1}^{N} X_i' V_i^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{N} X_i' V_i^{-1} Y_i \right).
\] (2.6)

Here we assume that whenever a matrix inverse appear in an equation, the inverse exists. This is true when \( \gamma \) is known. When the covariance matrices are unknown, we will use an estimate of \( \gamma \) and replace \( V_i(\gamma) \) with \( \hat{V}_i = V_i(\hat{\gamma}) \). The common methods used to estimate \( \gamma \) will be introduced in the next section.

### 2.1.2 Estimation of Variance Components

- **Maximum Likelihood Estimation**

  Methods of estimating variance components are extensive, mostly in ANOVA model setting and can only be applied to balanced design. In Harville’s paper (1977) [16], he reviewed several common approaches to variance component estimation. Among the approaches, maximum likelihood estimators remain popular because of its attractive features.

  Under ML, we compute estimates that jointly maximize the likelihood function (2.4). The likelihood function we described in section 2.1.1 assesses the joint probability of simultaneously observing all sample data actually obtained. Note that when we estimate variance components using ML, we actually treat fixed-effects of the model as known values. ML approach fails to allocate some degrees of freedom to estimation of the fixed-effects and therefore leads to biased estimates when the sample size is small [31]. Because of the reasons mentioned above, statisticians developed restricted maximum likelihood approach, which we will introduce in the next section.

- **Restricted Maximum Likelihood Estimation**

  In general, the restricted maximum likelihood (REML) estimation method is preferred over ML method because ML method does not take account of the loss in degrees of freedom resulting from the estimation of the model’s fixed effects and therefore may provide biased estimates. In contrast, the REML estimators are not only invariant to the fixed effects of the model but also free of the estimates of the fixed effects [9]. Thompson [37] first suggested the REML estimators for covariance matrix components.
for balanced data. Corbeil and Searle further extended the method so that it can be applied to both balanced and unbalanced data designs. To derive the REML for mixed-effects models, we first define

\[ Y = X\alpha + Zb + \epsilon, \]  

\[(2.7)\]

define

\[
Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}, 
\quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix}, 
\quad X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}, 
\quad Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_N \end{pmatrix},
\]

where \( Y \) follows \( N(X\alpha, V) \), \( V = \text{diag}(V_1, \cdots, V_N) = \Sigma + ZDZ' \), \( \Sigma = \text{diag}(\Sigma_1, \cdots, \Sigma_N) \), \( D = \text{diag}(D, \cdots, D) \).

The REML estimators are obtained by maximizing the likelihood of all error contrasts \( Y^* = L'Y \) which are linear combinations of the data that have zero expectation \([28]\), in other words, \( L \) is any full-rank matrix such that \( E(L'Y) = 0 \), that is \( L'X = 0 \) (\( L \) does not depend on \( \alpha \) or covariance matrix components). It is clear that \( Y^* \) follows \( N(0, LV(\gamma)L') \).

The estimation procedure follows the transformation used by Patterson and Thompson \([28]\) which consists of partitioning the log-likelihood function \((2.5)\) into two parts, and one part does not depend on the fixed effects in the model. The transformation being used is :

\[ Y^* = AY. \]  

\[(2.8)\]

In \((2.8)\) \( Y^* \) and \( A \) can be partitioned as:

\[ Y^* = \begin{pmatrix} Y^*_1 \\ Y^*_2 \end{pmatrix}, \quad A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \]

where \( A_1 = I - X(X'X)^{-1}X' \), \( A_2 = X'V^{-1} \). Note that \( A_1 \) is a \( (N - \text{rank}(X)) \times n \) matrix whose rows are any \( N - \text{rank}(X) \) linearly independent rows of the form \( I - X(X'X)^{-1}X' \).

Then we have \( Y \sim N(AX\alpha, AVA') \).

\[
AX\alpha = \begin{pmatrix} I - X(X'X)^{-1}X' \\ X'V^{-1} \end{pmatrix} X\alpha = \begin{pmatrix} 0 \\ X'V^{-1}X\alpha \end{pmatrix}
\]

\[
AVA' = \begin{pmatrix} A_1 \\ X'V^{-1} \end{pmatrix} V(A'_1, V^{-1}X) = \begin{pmatrix} A_1V'A'_1 \\ 0 \\ X'V^{-1}X \end{pmatrix}
\]
Now we have $Y_1^*$ and $Y_2^*$ uncorrelated and the distribution of $Y_1^*$ does not depend on the fixed effect $\alpha$ in the model. $E(Y_1^*) = 0, Var(Y_1^*) = A_1V A_1'$. The likelihood of $Y_1^*$ is the basis of estimators of the variance components. Now we see that the inference of $\gamma$ is based on likelihood function associated with $N - \text{rank}(X)$ linearly independent error contrasts rather than the one associated with the full data $Y$. $A_1$ is a symmetric, idempotent, and singular matrix of rank $N - \text{rank}(X)$. REML estimator is a value of $\gamma$ that maximizes the value of the REML function $L_{REML}$:

$$L_{REML}(\theta) = \left| \sum_{i=1}^{N} X_i'V_i(\gamma)X_i \right|^{-\frac{1}{2}}L_{ML}(\theta). \quad (2.9)$$

Harville [15] showed that the likelihood of error contrasts can be expressed in the following form:

$$L(\gamma) = (2\pi)^{-\frac{(n - \text{rank}(X))}{2}}\left| \sum_{i=1}^{N} X_i'X_i \right|^\frac{1}{2} \times \left| \sum_{i=1}^{N} X_i'V_i^{-1}X_i \right|^{-\frac{1}{2}} \prod_{i=1}^{N} |V_i|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^{N} (Y_i - X_i\hat{\alpha})'V_i^{-1}(Y_i - X_i\hat{\alpha}) \right\}, \quad (2.10)$$

where $\hat{\alpha}$ is obtained from (2.6). Therefore, the REML estimator $\hat{\gamma}$ is independent of the choice of $A$. Further, Harville argued that the full ML estimator depends on $Y$ only through a set of $n - \text{rank}(X)$ linearly independent error contrasts, in other words, as a function of $A^1Y$, it depends only on $A_1Y$ but not on $A_2Y$. Other alternative formulations of transformation were also suggested by other authors[9].

ML and REML estimators share some desirable properties: (i) consistency. (ii) asymptotically normal distribution. (iii) asymptotically efficiency. As we mentioned before, ML estimators may be biased when the sample size is small. However, neither approach is uniformly superior over the other one. Based on Kreft and Leeuw’ review (1998) [19] of simulation studies, it is still unclear which approach is more correct. But it is worth noting that ML method provides estimate for fixed-effects while REML method does not. REML method captures only the stochastic portion of the model. In Wolfinger’s paper (1996) [40], he reviewed some of the advantages of REML. One appealing property is its ability to accommodate data which are missing at random or intentionally incomplete. It can also be used to select appropriate covariance models for the data, which will be introduced in the next section. It has been suggested by Singer and Willett [31] that when we compare models that differ only in their
variance components, we can use either method; and when we compare models that differ in both fixed effects and variance components, we must use the full information model. Therefore, when sample size is sufficiently large, the choice depends on where one’s interest is.

2.1.3 Covariance Structure

In most research, we are primarily interested in fixed effects in the model because they provide us answers to our research questions. So why should we bother to model the covariance structure in longitudinal studies? A lot of researchers use classic normal and independence assumptions when their data are actually longitudinal. Note that when such assumptions are made, the correlation between observations is ignored and therefore may result in biased results. We must consider appropriate approaches for modeling the correlation/time dependence among the repeated measures obtained on the same subjects. In addition, when missing data is present in the dataset, correct covariance model is usually a requirement for obtaining valid parameter estimation. A unstructured covariance model seems to be the most straightforward one can easily think of, however, it is not always possible to model longitudinal data using an unstructured covariance matrix. Many times, we need to make certain assumptions of covariance structure of random effects to make models more parsimonious. Using a covariance structure that is not close to the correct covariance may cause confidence intervals to be incorrectly wide or narrow. Choosing a proper covariance model improves the efficiency of the fixed effect estimates and will allow for more accurate confidence intervals and hypothesis tests [39]. In this section, we are going to review some of the common covariance structures and their extensions [40] used in longitudinal study literature. We will follow Weiss’ approach and assume that data are taken at equally spaced time points $t_i = t = (1, 2, 3, 4, 5)$ with no missing observations.

- Unstructured Covariance Model
  
  The unstructured covariance model is the most general model among all. It is appealing since it places no restrictions on the covariance coefficients. The only restriction is that the covariance matrix should be symmetric and positive-definite. This set up is feasible when the number of measurement occasions is relatively small and all subjects are measured at the same set of occasions. However, for a $5 \times 5$ covariance matrix, the
number of unknown parameters in the covariance matrix is 15. And for a \( n \times n \) matrix, we will have to estimate \( \frac{n(n+1)}{2} \) unknown parameters. This will take too many degrees of freedom. Therefore, even though unstructured covariance model seems straightforward, it is not very desirable. In most research studies, researchers prefer to select the most efficient inferences based on the most parsimonious covariance structure that fits the data reasonably well [40].

- **Compound Symmetry (CS)**

In compound symmetry covariance model, we assume that each observation has constant variance across time:

\[
\text{Var}(Y_{it}) = \sigma^2;
\]

And correlation between any two observations from the same subject is constant:

\[
\text{Corr}(Y_{it}, Y_{il}) = \rho, \quad j \neq l
\]

The algebraic form of the covariance matrix for data with 5 observations is

\[
\text{Var}(Y_i | \sigma^2, \rho) = \sigma^2 \begin{pmatrix}
1 & \rho & \rho & \rho & \rho \\
\rho & 1 & \rho & \rho & \rho \\
\rho & \rho & 1 & \rho & \rho \\
\rho & \rho & \rho & 1 & \rho \\
\rho & \rho & \rho & \rho & 1
\end{pmatrix}
\]  

(2.11)

The key feature of CS model is that the correlation between two observations from the same subject does not depend on time gap between the observations within subject. It says the best predictor of \( Y_{it} \) is a linear function of the average of all previous observations. This is unlikely when data are collected from subjects over a long period of time because we would expect the correlation between pairs to decline as they become temporally remote within subject. Compound symmetry covariance structure is common when the rate of change do not differ much across subjects.

The compound symmetry model is equivalent to random intercept model. The random intercept model assumes that \( Y_{it} \) has two sources of variation. One is the variation of subject-specific intercepts around their population mean, the other is the variation of observations around the subject-specific mean. If we write the model \( Y_{it} = b_i + \epsilon_{it} \), where \( b_j \sim N(\tau, D) \) and \( \epsilon_{it} \sim N(0, \sigma^2) \), we know that given \( b_{it} \), \( Y_{it} \) follows \( N(0, \sigma^2) \).
And variance of $Y_{it}$ is $Var(Y_{it}) = D + \sigma^2$; covariance of any two of the observations from the same subject is $Cov(Y_{it}, Y_{il}) = Cov(b_i + \epsilon_{it}, b_i + \epsilon_{il}) = Var(b_i) = D$. Therefore, the covariance matrix is

$$
\begin{pmatrix}
\sigma^2 + D & D & \cdots & D \\
D & \sigma^2 + D & \cdots & \\
\vdots & \vdots & \ddots & \vdots \\
D & \cdots & \sigma^2 + D
\end{pmatrix}
$$

As we can see, this is just an alternative way of writing the compound symmetry covariance matrix.

- (first-order) Autoregressive (AR)

In Autoregressive covariance model, we still assume $Var(Y_{it}) = \sigma^2$. But instead of constant correlation, we assume declining correlation when the time period between two observations increases (we mainly discuss AR(1) model here). That is,

$$Corr(Y_{it}, Y_{il}) = \rho^{|t_{it} - t_{il}|}$$

In our example, the covariance matrix is

$$Var(Y_i|\sigma^2, \rho) = \sigma^2 \begin{pmatrix}
1 & \rho & \rho^2 & \rho^3 & \rho^4 \\
\rho & 1 & \rho & \rho^2 & \rho^3 \\
\rho^2 & \rho & 1 & \rho & \rho^2 \\
\rho^3 & \rho^2 & \rho & 1 & \rho \\
\rho^4 & \rho^3 & \rho^2 & \rho & 1
\end{pmatrix}$$

(2.12)

AR model suggests that the last observation times correlation is the best predictor of next observation. In AR model, we estimate two parameters $\sigma$ and $\rho$, which means it is a parsimonious model. But in the mean time, the assumptions are quite constrained. Same as in time series analysis, AR(1) model can be written in the following form:

$$Y_{it} - x_{it}\alpha = \rho(Y_{i,t-1} - x_{i,t-1}\alpha) + \delta_{it}$$

where $\delta_{it} \sim N(0, \sigma^2(1 - \rho^2))$. This is so since we assume that $Y_{it}$ has variance $\sigma^2$. The AR(1) model can be extended to AR(p) model, which relates the current observation $Y_{it}$ to the previous $p$ observations. We will discuss generalization of AR model in the following section.
• Heterogeneous (first-order) Autoregressive (ARH) Similar to the CSH model we discussed before, the heterogeneous autoregressive model is a relaxed version of AR model. It also assumes that data are equally spaced. We will write the covariance matrix below:

\[
\text{Var}(Y_i) = \begin{pmatrix}
\sigma_1^2 & \sigma_1 \sigma_2 \rho & \sigma_1 \sigma_3 \rho^2 & \sigma_1 \sigma_4 \rho^3 & \sigma_1 \sigma_5 \rho^4 \\
\sigma_2 \sigma_1 \rho & \sigma_2^2 & \sigma_2 \sigma_3 \rho & \sigma_2 \sigma_4 \rho^2 & \sigma_2 \sigma_5 \rho^3 \\
\sigma_3 \sigma_1 \rho^2 & \sigma_3 \sigma_2 \rho & \sigma_3^2 & \sigma_3 \sigma_4 \rho & \sigma_3 \sigma_5 \rho^2 \\
\sigma_4 \sigma_1 \rho^3 & \sigma_4 \sigma_2 \rho^2 & \sigma_4 \sigma_3 \rho & \sigma_4^2 & \sigma_4 \sigma_5 \rho \\
\sigma_5 \sigma_1 \rho^4 & \sigma_5 \sigma_2 \rho^3 & \sigma_5 \sigma_3 \rho^2 & \sigma_5 \sigma_4 \rho & \sigma_5^2 \\
\end{pmatrix}
\]  

(2.13)

• Autoregressive Moving Average
The autoregressive moving average (ARMA) model is a variation of the AR model. There is one more correlation parameter \(\gamma\) in the model. Recall that in the AR(1) model, correlation between \(Y_{it}\) and \(Y_{i,t-k}\) is \(\rho^k\), in the ARMA(1,1) model, we have

\[
\text{Corr}(Y_{it}, Y_{i,t-k}) = \gamma \rho^{k-1}.
\]

The full covariance matrix is

\[
\text{Var}(Y_i) = \sigma^2 \begin{pmatrix}
1 & \gamma & \gamma \rho & \gamma \rho^2 & \gamma \rho^3 \\
\gamma & 1 & \gamma & \gamma \rho & \gamma \rho^2 \\
\gamma \rho & \gamma & 1 & \gamma & \gamma \rho \\
\gamma \rho^2 & \gamma \rho & \gamma & 1 & \gamma \\
\gamma \rho^3 & \gamma \rho^2 & \gamma \rho & \gamma & 1 \\
\end{pmatrix}
\]  

(2.14)

Note that if \(\rho = 1\), the model is the same as the compound symmetry model; if \(\rho = \gamma\), the ARMA(1,1) becomes AR(1) model. The ARMA(1,1) model can be written in terms of previous observation and error term:

\[
Y_{it} = \rho Y_{i,t-1} + \beta \delta_{i,t-1} + \delta_{it}
\]

The covariance structures we have presented are only a part of the covariance models family. It should be noted that when choosing a covariance structure, one should always choose the one that is appropriate for the data and study design. We can use either graphs to help us choose covariance structure visually or compare common goodness-of-fit statistics.
2.2 Mixed Distribution Models

2.2.1 Introduction

While the conventional mixed-effect linear model is widely applied in many studies, there are situations where linear mixed models are not adequate for making inferences. The most common case is when the assumption of normality does not hold. In medical and economics research, it is not uncommon to encounter longitudinal data sets that have many zeros and positive continuous responses. The data are usually measures of an outcome. The distribution is highly skewed and there is a peak at zero (can be other values as well). Examples include health insurance claim data, environment pollution data and climate data. Applying commonly used transformation methods do not solve problem as there will still be a large number of identical values. These data present challenges because it is not only necessary to model associations within subject, heterogeneity between subjects but also to take into account the non-normality present in the data. If the data are treated as if they are normally distributed, the estimations and predictions will not be reliable since the model assumptions do not hold.

One straightforward way to resolve the problem is to divide the data into two parts: one part with zeros only and the other part with positive values. The second part can be modeled using models we discussed in the last section. However, this approach ignores zero values and potential associations between zeros and continuous responses. And in many cases, the zeros may contain valuable information if they are “true values”. Remember that in longitudinal studies, it is suggested that we should not delete observations due to potential loss of useful information. Two-part models are introduced when the response is “semi-continuous”. They model the occurrence process and intensity process. Two-part models combine probability of occurrence of a positive value and the probability distribution of the positive responses.

Our main goal is to model the dynamics in performance-based CEO compensation, in other words, mostly option grants. Applying the classic mixed-effect model to executive option compensation may result in biased estimates because an excessive amount of zeros violates the normality assumption. Figure 2.1 presents distribution of option grants without any transformation. Clearly, it is highly skewed and does not follow a normal distribution. Therefore, mixed-effects model will fit our data poorly. In the context of executive compensation, one is modeling a sequential decision making process: in the first step,
board of directors decide whether or not to grant options to a CEO; if it decides to grant them, then decides how many to grant.

The problem of fitting “semi-continuous” data was first addressed by Aitchison[2]. Two-part models have been applied in many econometric studies as well as medical studies[11][12]. Duan (1983)[12] adopted the two-part model to analyze health insurance cost by decomposing health care expenditures into the event of positive costs and the amount of expenditures. This was done by fitting two models (logit and linear regression) separately. The general procedure was summarized by Lachenbruch[20]: define the response variable as $y = (x, d)$, where $d = 1$ if $y \neq 0$ and $d = 0$ otherwise. $x = \text{response}$ when $d = 1$. The general form of this mixed-distribution model is:

$$f(y) = \begin{cases} 
P(Y = 0) & \text{if } y = 0 \\
(1 - P(Y = 0))h(y) & \text{if } y > 0 \\
0 & \text{otherwise}
\end{cases}$$  \hspace{1cm} (2.15)

The second part of this model models $Y$ conditional on nonzero observations. Because $d$ is binary, the model can also be written as $f(y) = p^{1-d}((1 - p)h(x))^d$, where $p$ is the probability of the event $y > 0$ and $h(.)$ is the density function of $x$. In general, $h(.)$ is
assumed to be normal distribution, but alternative specifications can be made as well. We will discuss it in the next section. This general model is often applied to cross-sectional data and it was extended to longitudinal data by Olsen and Schafer[27]. Tooze et. al. [35] also develop a similar random-effects mixed-distribution model.

Two-part models are attractive because they provide insight into determinants of occurrence and level of intensity; on the other hand, results obtained from classical mixed-effects models may be driven by the excessive amount of zeros. It is easier to interpret when the zeros are valid responses rather than proxies for missing or censored data (e.g. value of the response is set to 0 when the actual observed value is lower than a certain limit.)[12]. Sample selection models, such as Tobit model[34] and Heckman Selection model[18] are widely used in literature to model limited or censored responses.

2.2.2 Mixed-Distribution Models with Correlated Random Effects

In this section, we follow Tooze et. al.(2002)[35] and Olsen et. al (2001)[27] and introduce the mixed-distribution models for longitudinal data.

For a response variable $Y_{it}$ of the $i^{th}$ subject at time $t$, let $R_{it}$ denote the occurrence variable where

$$R_{it} = \begin{cases} 
0 & \text{if } Y_{it} = 0 \\ 1 & \text{if } Y_{it} > 0 
\end{cases} \quad (2.16)$$

First, we model the occurrence variable $R_{it}$ with a logistic regression model with random effects; $R_{it}$ has conditional probabilities:

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

where we assume that $p_{it}(\theta_1)$ follows a logistic regression model. The occurrence model is:

$$\logit(p_{it}(\theta_1)) = \log\left(\frac{p_{it}(\theta_1)}{1 - p_{it}(\theta_1)}\right) = \eta_{it}(\theta_1). \quad (2.17)$$

$$\eta_{i}(\theta_1) = X_{1i}\beta_1 + Z_{1i}b_{1i}, \quad (2.18)$$

where $\eta_{i} = (\eta_{i1}, \cdots, \eta_{in_i})'$ and $\theta_1 = (\beta_1, b_{1i})$. $i = 1, \cdots, N; t = 1, \cdots, n_i$; $\beta_1$ and $b_{1i}$ are vectors of the fixed effects and random effects in the occurrence model. $X_1$ and $Z_1$ are the matrices of covariates pertaining to the fixed and random effects in the occurrence process.
Next, we specify the model for positive continuous observations \( y_{it} \). Define \( V_{it} = [h(Y_{it}) | R_{it} = 1] \), where \( h(.) \) is a monotone increasing function such that \( V_{it} \) is approximately normal\(^1\); examples of \( h(.) \) include log or identity. Let \( V_i = (V_{i1}, \cdots, V_{in_i}) \), the second model for continuous positive response is:

\[
V_i = X_{2i} \beta_2 + Z_{2i} b_{2i} + \epsilon_i. \tag{2.19}
\]

Similarly to the first step, let \( \theta_2 = (\beta_2, b_{2i}) \), where \( \beta_2 \) and \( b_{2i} \) are fixed and random effects in the second model. We assume that \( V_{it} \) has density function \( f(.) \). The residuals \( \epsilon_i \)'s are assumed to be \( N(0, \sigma_e I) \), \( X_{2i} \) and \( Z_{2i} \) are matrices of covariates pertaining to fixed and random effects in the intensity process. The random effects from the two parts are assumed to be jointly normally distributed and possibly correlated. The assumption of potential correlation between the two parts is crucial because it is likely that zero values and positive values within the same subjects are correlated. If \( \Psi_{12} = 0 \), then the occurrence process and intensity process are two independent parts. It essentially means presence of zero values has no influence on the amount of option grants granted at another time point. However, it is found by Olsen and Schafer\(^{27} \) that this condition typically does not hold; random effects from two parts are usually strongly correlated.

\[
b_i = \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N \left( 0, \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix} \right). \tag{2.20}
\]

Covariates in both models can be different and time varying covariates can be included in either process. To make models simpler, we follow Tooze’s approach by assuming the only random effect in each model is the intercept. The random effects can be written as \( (b_{1i}, b_{2i}) \), where \( b_{1i} \) is random effect in the first part and \( b_{2i} \) is from the second part. Furthermore, the joint distribution of random effects can be simplified to

\[
\begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N \left( 0, \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \right). \tag{2.21}
\]

\(^{1}\)Note that due to certain computational limitations in SAS, which we will be using to fit our models, it is sometimes preferred to adopt a lognormal model for non-transformed data instead of a normal model for log-transformed data.
Let $\theta = (\theta_1, \theta_2)$, the pdf of $Y_{it}$ can be written as the following:

$$f(y_{it}|\theta) = f(y_{it}|R_{it} = 0, \theta)P(R_{it} = 0|\theta) + f(y_{it}|R_{it} = 1, \theta)P(R_{it} = 1|\theta)$$

$$= \delta_0(y_{it})P(R_{it} = 0|\theta_1) + f(V_{it}|\theta_2)P(V_{it} = 1|\theta_2)$$

$$= (1 - p_{it}(\theta_1))\delta_0(y_{it}) + p_{it}(\theta_1)f(V_{it}|\theta_2). \quad (2.22)$$

where

$$\begin{cases}
  f_+^+ \delta_0(y)dy = 1 \\
  \delta_0y = 0 \quad \text{when } y \neq 0
\end{cases}$$

The conditional expectation of $Y_{it}$ is:

$$E(Y_{it}|\theta) = \int f(y_{it}|\theta)y_{it}dy = p_{it}(\theta_1)\int f(v_{it}|\theta_2)v_{it}dv$$

$$= p_{it}(\theta_1)E(V_{it}|\theta_2) = p_{it}(\theta_1)\mu_{V_{it}}(\theta_2). \quad (2.23)$$

And similarly, the conditional variance of $Y_{it}$ is:

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

$$= E(Y_{it}^2|\theta) - E(Y_{it}|\theta)^2 = p_{it}(\theta_1)E(V_{it}^2|\theta_2) - p_{it}(\theta_1)^2\mu_{V_{it}}(\theta_2)^2$$

$$= p_{it}(\theta_1)Var(V_{it}|\theta_2) + p_{it}(\theta_1)\mu_{V_{it}}(\theta_2)^2 - p_{it}(\theta_1)^2\mu_{V_{it}}(\theta_2)^2$$

$$= p_{it}(\theta_1)Var(V_{it}|\theta_1) + p_{it}(\theta_1)(1 - p_{it}(\theta_1))\mu_{V_{it}}(\theta_2)^2 \quad (2.24)$$

Contribution of the $i^{th}$ subject to the likelihood is:

$$L_i(\beta_1, \beta_2, \sigma_1, \sigma_2, \sigma_e, \rho|y_{i1}, \cdots, y_{in_i}) =$$

$$\int_{b_{1i}} \int_{b_{2i}} f(y_{i1}, \cdots, y_{in_i}|\beta_1, \beta_2, \sigma_1, \sigma_2, \sigma_e, \rho)db_{1i}db_{2i} =$$

$$\int_{b_{1i}} \int_{b_{2i}} f(y_{i1}, \cdots, y_{in_i}|\theta_1, \theta_2, \sigma_1, \sigma_2, \rho)g(b_{1i}, b_{2i}|\beta_1, \beta_2, \cdots)db_{1i}db_{2i} =$$

$$\int_{b_{1i}} \int_{b_{2i}} \prod_{t=1}^{n_i} f(y_{it}|b_{1i}, b_{2i}, \beta_1, \beta_2)g(b_{1i}, b_{2i}|\sigma_1, \sigma_2, \rho, \sigma_e)db_{1i}db_{2i}.$$

Therefore, the likelihood function is:

$$L(\beta_1, \beta_2, \sigma_1, \sigma_2, \sigma_e, \rho|Y) = \prod_{i=1}^{N} L_i$$

$$= \prod_{i=1}^{N} \int_{b_{1i}} \int_{b_{2i}} \prod_{t=1}^{n_i} \left(1 - p_{it}(\beta_1, b_{1i})\right)^{1-r_{it}} \left(p_{it}(\beta_1, b_{1i})\right)^{r_{it}} \times$$

$$f(V_{it}|\beta_2, b_{2i})^{r_{it}} g(b_{1i}, b_{2i}|\sigma_1, \sigma_2, \sigma_e, \rho)db_{1i}db_{2i} \quad (2.25)$$

where $\rho$ is the correlation coefficient of $\sigma_1$ and $\sigma_2$. 

19
Example: Lognormal Model

The example shown here assumes lognormal distribution of $V_{it}$ (Note that this is essentially same as assuming $h(V_{it})$ to be normally distributed) $V_{it} \sim [Y_{it}|R_{it} = 1]$ and $\log(V_{it} | \theta_2) \sim N(X_{2it}\beta_2 + Z_{2it}b_{2i}, \sigma_e^2)$. Same as in the last section, intercept is the only random effect in both models. Now the conditional distribution of $V_{it}$ becomes $N(X_{2it}\beta_2 + b_{2i}, \sigma_e^2)$. The marginal mean of $V_{it}$ is $exp(X_{2it}\beta_2 + \frac{\sigma_e^2}{2} + \frac{\sigma_i^2}{2})$.

2.3 Mixed-Distribution Models with Random Effects and Serial Correlation

2.3.1 Model

The mixed-distribution/two-part models with correlated random effects are useful when modeling semi-continuous longitudinal data. However, these models place strong assumptions on association structure and do not incorporate serial correlation within subjects. In general, ignoring the within subject correlation will lead to biased variances estimates of fixed effects and therefore may lead to wrong conclusions. Albert and Shen[3] proposed a mixed-distribution model with two correlated continuous random processes instead of random effects components to incorporate within subject correlation.

In the context of executive compensation, information is reported once every year at each company’s fiscal year end. In other words, the sampling interval is generally one year. Since we have observations at discrete time, an alternative and more realistic model may be a mixed-distribution with random effects that have discrete AR structure.

In this section, we will adopt similar notations with mild changes. $R_{it}$ is defined to be the same as in the last section.

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

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

$$\logit(p_{it}(\theta_1)) = \log(\frac{p_{it}(\theta_1)}{1 - p_{it}(\theta_1)}) = \eta_{it}(\theta_1).$$
where \( \eta_i = (\eta_{i1}, \cdots, \eta_{in_i})', \theta_1 = (\beta_1, b_{it}). \) \( i = 1, \cdots, N; t = 1, \cdots, n_i. \) As for the intensity process, \( V_{it} \) is defined to be the same as before: \( V_{it} = [h(Y_{it})|R_{it} = 1], \) where \( h(.) \) is a monotone increasing function such that \( V_{it} \) is approximately normal.

The model consists of two parts as follows:

\[
\begin{aligned}
\eta_{it} &= X_{1i} \beta_1 + b_{it} \\
V_{it} &= X_{2i} \beta_2 + c_{it} + \epsilon_{it}
\end{aligned}
\] (2.26)

where \( \beta_1 \) and \( \beta_2 \) are vectors of fixed effects in the occurrence model and intensity model, respectively; \( X_1 \) and \( X_2 \) are the matrices of covariates pertaining to the fixed effects in the model. The residuals \( \epsilon_{it} \) are independent and identically distributed with distribution \( N(0, \sigma_e). \)

We allow the random effects to be correlated and have an AR(1) structure by assuming that

\[
\begin{aligned}
b_{it} &= \phi_1 b_{i,t-1} + Z_{1t}, \quad |\phi_1| < 1 \\
c_{it} &= \phi_2 c_{i,t-1} + Z_{2t}, \quad |\phi_2| < 1
\end{aligned}
\] (2.27)

where \( Z_{1t} \sim N(0, \sigma_1^2) \) and \( Z_{2t} \sim N(0, \sigma_2^2); \) \( Cov(Z_{1t}, b_{1t'}') = 0, Cov(Z_{2t}, b_{2t'}) = 0 \) when \( t \neq t'. \) And we allow \( Z_{1t} \) and \( Z_{2t} \) to be correlated:

\[
Cov(Z_{1j}, Z_{2k}) = \begin{cases} 
0, & j \neq k \\
\delta \sigma_1 \sigma_2, & j = k
\end{cases}
\] (2.28)

\[
\begin{pmatrix} Z_{1t} \\ Z_{2t} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \delta \sigma_1 \sigma_2 \\ \delta \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \right).
\]

Note that \( b_{it} \) and \( c_{it} \) both have AR(1) structure, therefore, we know that they can be written in form of \( Z_{1,t} \) and \( Z_{2,t}, \) respectively.

\[
b_{it} = \sum_{k=0}^{\infty} \phi_1^k Z_{1,t-k}, \\
c_{it} = \sum_{k=0}^{\infty} \phi_2^k Z_{2,t-k}.
\]

Since we assume \( |\phi_1| < 1 \) and \( |\phi_2| < 1, b_{it} \) and \( c_{it} \) are two stationary AR(1) processes. It can
be easily shown that

\[
\begin{align*}
\text{Var}(b_{it}) &= \frac{\sigma_1^2}{1-\phi_1^2} \\
\text{Var}(c_{it}) &= \frac{\sigma_2^2}{1-\phi_2^2} \\
\text{Cov}(b_{it}, b_{it'}) &= \frac{\phi_1^{t'-t}\sigma_1^2}{1-\phi_1^2} \\
\text{Cov}(c_{it}, c_{it'}) &= \frac{\phi_2^{t'-t}\sigma_2^2}{1-\phi_2^2}
\end{align*}
\]

The covariance between \(b_{ij}\) and \(c_{il}\) is:

\[
\text{Cov}(b_{ij}, c_{il}) = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \phi_1^k \phi_2^m \text{Cov}(Z_{1,j-k}, Z_{2,l-m})
\]

\[
= \sum_{k=0}^{\infty} \phi_1^k \phi_2^{l+k-j} \delta \sigma_1 \sigma_2 = \phi_2^{l-j} \frac{\delta \sigma_1 \sigma_2}{1-(\phi_1 \phi_2)^2}
\]

If \(l = j + k\), then the equation simplifies to \(\text{Cov}(b_{ij}, c_{i,j+k}) = \phi_1^k \frac{\delta \sigma_1 \sigma_2}{1-(\phi_1 \phi_2)^2}\). If \(l = j\), then the covariance between the two random effects of the same subject at the same time point is \(\text{Cov}(b_{ij}, c_{ij}) = \frac{\delta \sigma_1 \sigma_2}{1-(\phi_1 \phi_2)^2}\).

Let \(b_i = (b_{i1}, \ldots, b_{in})'\) and \(c_i = (c_{i1}, \ldots, c_{in})'\), \((b_i, c_i)\) is assumed to follow a multivariate normal distribution with mean vector zeros and covariance matrix \(\Sigma\), where

\[
\Sigma = \begin{pmatrix}
\text{Var}(b_i) & \text{Cov}(b_i, c_i) \\
\text{Cov}(b_i, c_i) & \text{Var}(c_i)
\end{pmatrix}
\]

Let \(D_{it} = (b_{it}, c_{it})\), we have \(D_{it} = \phi D_{i,t-1} + Z_t\), where

\[
\phi = \begin{pmatrix}
\phi_1 & 0 \\
0 & \phi_2
\end{pmatrix}, \quad Z_t = \begin{pmatrix}
Z_{1t} \\
Z_{2t}
\end{pmatrix}
\]

Note that the structure of \(D_{it}\) is an extension of single series AR(1). If \(\phi_1 = \phi_2 = 0\), then this model reduces to the correlated random-effects model introduced in the last section. We will assume that \(c_{it} = b_{it}\) and estimate this reduced model in the next section.

### 2.3.2 Estimation

Similarly as before, we assume that given the random effect \(c_{it}\), \(V_{it}\) has distribution function \(f(.)\). Let \(\Theta\) denote the complete unknown parameter space, the likelihood is:

\[
L(\Theta|Y) = \prod_{i=1}^{N} \prod_{t=1}^{n_i} \int_{b_i} \int_{c_i} \prod_{t=1}^{n_i} (1-p_{it})^{1-r\alpha} (p_{it})^{r\alpha} f(V_{it})^{r\alpha} g(b_i, c_i|\Sigma) db_idc_i
\]

(2.29)
where $\mathbf{b}_i = (b_{i1}, \cdots, b_{im_i})'$, $\mathbf{c}_i = (c_{i1}, \cdots, c_{im_i})'$; and $g(\mathbf{b}_i, \mathbf{c}_i | \Sigma)$ is the joint distribution function of $\mathbf{b}_i$ and $\mathbf{c}_i$ with mean vector zero and covariance matrix $\Sigma$.

When $c_{it} = b_{it}$, the likelihood function reduced to a simpler form:

$$L = \prod_{i=1}^{N} L_i = \prod_{i=1}^{N} \int_{b_{i1}t=1}^{n_i} (1 - p_{it})^{1-r_{it}} (p_{it})^{r_{it}} f(V_{it})^{r_{it}} g(\mathbf{b}_i | \Sigma) d\mathbf{b}_i$$  \hspace{1cm} (2.30)

We will first focus on estimation of the shared-random-effects model. The likelihood in equation (2.30) is very difficult to evaluate computationally. The high-order integral cannot be written in a closed form. In the literature, several approaches have been applied to evaluate the integral in (2.30). Tooze and her colleagues [35] estimate their model in SAS, which adopts the adaptive Gaussian quadrature method. Olsen and Schafer [27] used Laplace Approximation and Albert [3] proposed to use MCEM algorithm to estimate their two-part mixed model. Since our model cannot be estimated directly in SAS, we are going to follow Olsen and Schafer’s approach and propose a Monte Carlo EM (MCEM) algorithm to estimate our model. Note that other methods for estimating generalized linear mixed model and nonlinear mixed models can also be used to estimate our model potentially, but due to complexity of the high-order integral in (2.30), no method can be applied without high computational cost.

**Monte Carlo EM Algorithm**

The EM algorithm is an iterative procedure that finds the maximum likelihood estimates. Each iteration consists of an E-step (Expectation) and an M step (Maximization). To develop the EM algorithm in the mixed-distribution two part models setting, we shall first look at the classic EM algorithm.

In the EM method, random effects $\mathbf{b}_i$'s in the statistical model are assumed to be the missing data and $\mathbf{C} = (\mathbf{Y}, \mathbf{b})$ is the complete data. The $r^{th}$ iteration computes the conditional expectation of the log-likelihood of the complete data $\mathbf{C}$ with respect to missing data $\mathbf{b}$ given observed data $\mathbf{Y}$ and $\Theta^{(r)}$ [38]. In other words, in E-step we find the expectation $Q(\Theta | \Theta^{(r)}) = E(\log(f(\mathbf{Y}, \mathbf{b} | \Theta)) | \mathbf{Y}, \Theta^{(r)})$ where $\Theta$ is the parameter space. In the M-step, we find the $\Theta$ that maximizes $Q(\Theta | \Theta^{(r)})$ and let it be $\Theta^{(t+1)}$. That is, $\Theta^{(t+1)} = \arg \max_{\Theta} Q(\Theta | \Theta^{(r)})$. Note that the expected complete-data likelihood can be written as $Q(\Theta | \Theta^{(r)}) = \int \log(f(\mathbf{Y}, \mathbf{b} | \Theta)) g(\mathbf{b} | \Theta^{(r)}, \mathbf{Y}) d\mathbf{b}$, which cannot be computed analytically. This
is because the conditional density involves \( f(\mathbf{Y}) \), which is an intractable term we try to avoid calculating. However, we may use Monte Carlo Method to generate \( M \) samples from \( g(\mathbf{b}|\mathbf{Y}, \Theta) \). And the expectation can be approximated by \( \frac{1}{M} \sum_{j=1}^{M} \log f(\mathbf{Y}, \mathbf{b}^{(j)}|\Theta^{(r)}) \)

The general MCEM algorithm is as follows:

1. Set initial values for parameters to be estimated: \( \beta_1^{(0)}, \beta_2^{(0)}, \sigma_1^{(0)}, \phi_1^{(0)} \) and \( \sigma_e^{(0)} \); iteration \( l=1 \).
2. Generate \( m \) samples for each subject \( i \), \( \mathbf{b}_i^{(1)}, \cdots, \mathbf{b}_i^{(m)} \), from the posterior density \( f(\mathbf{b}_i|\mathbf{Y}, \beta_1^{(l)}, \beta_2^{(l)}, \phi_1^{(l)}, \sigma_e^{(l)}, \sigma_1^{(l)}) \), where \( i = 1, \cdots, N \).
3. Choose \( \beta_1^{(l+1)}, \beta_2^{(l+1)}, \sigma_1^{(l+1)}, \phi_1^{(l+1)} \) and \( \sigma_e^{(l+1)} \) that maximize the Monte Carlo estimate of \( E(\ln(f(\mathbf{Y}|\mathbf{b}, \beta_1^{(l)}, \beta_2^{(l)}, \sigma_e^{(l)}))) \) and \( E(\ln(g(\mathbf{b}|\phi_1^{(l)}, \sigma_1^{(l)}))) \), respectively. Set iteration \( l=l+1 \).
4. Repeat steps 3 and 4 until convergence is achieved.

This MCEM approach is computationally intensive, however, it is applicable for variety of data structures. We will now demonstrate each step in context of our model in detail.

**Sampling random effect \( \mathbf{b} \)**

To sample from the posterior density of random effects, we first write the function as follows:

\[
\begin{align*}
    f(\mathbf{b}_i|\mathbf{y}_i, \Theta^{(l)}) &= \frac{f(\mathbf{y}_i|\mathbf{b}_i, \Theta^{(l)}) g(\mathbf{b}_i|\Theta^{(l)})}{f(\mathbf{y}_i|\Theta^{(l)})} \\
    &= \frac{\int f(\mathbf{y}_i|\mathbf{b}_i, \beta_1^{(l)}, \beta_2^{(l)}, \sigma_e^{(l)}, \sigma_1^{(l)}) g(\mathbf{b}_i|\phi_1^{(l)}, \sigma_1^{(l)}) db_i}{\int \cdot} \\
    &= \frac{\int f(\mathbf{y}_i|\mathbf{b}_i, \beta_1^{(l)}, \beta_2^{(l)}, \sigma_e^{(l)}) g(\mathbf{b}_i|\phi_1^{(l)}, \sigma_1^{(l)}) db_i}{\int \cdot} \\
    &= \frac{\int f(\mathbf{y}_i|\mathbf{b}_i, \beta_1^{(l)}, \beta_2^{(l)}, \sigma_e^{(l)})}{\int \cdot} \\
    &= \int \cdot \\
\end{align*}
\]

(2.31)

Since the random effect in our model is high dimensional, we will apply both Gibbs Sampling and Metropolis-Hasting Algorithm to sample from this conditional density. In order to obtain a neat form of acceptance function, we let the candidate function take the same form as \( g(\mathbf{b}) \).

Let \( \mathbf{b}_i \) be one draw from the conditional density, and we generate a new value for the \( t^{th} \) component of \( \mathbf{b}_i^{(n)} \) using the candidate function. Let \( \mathbf{b}_i^* = (b_{i1}, \cdots, b_{it}, \cdots, b_{iT}) \), we will accept the new value with probability \( \rho_t \), or else we will use the previous value \( \mathbf{b}_i^{(n)} \). The acceptance probability is given by

\[
\rho_t = \min\left\{ 1, \frac{\int f(\mathbf{y}_i|\mathbf{b}_i^*, \beta_1^{(l)}, \beta_2^{(l)}, \sigma_e^{(l)})}{\int f(\mathbf{y}_i|\mathbf{b}_i^{(n)}, \beta_1^{(l)}, \beta_2^{(l)}, \sigma_e^{(l)})} \right\} \left\{ \frac{g(\mathbf{b}_i^*|\Sigma^{(l)}) g(\mathbf{b}_i^{(n)}|\Sigma^{(l)})}{g(\mathbf{b}_i^{(n)}|\Sigma^{(l)}) g(\mathbf{b}_i^*|\Sigma^{(l)})} \right\}
\]

\[
\rho_t = \min\left\{ 1, \frac{\int f(\mathbf{y}_i|\mathbf{b}_i^*, \beta_1^{(l)}, \beta_2^{(l)}, \sigma_e^{(l)})}{\int f(\mathbf{y}_i|\mathbf{b}_i^{(n)}, \beta_1^{(l)}, \beta_2^{(l)}, \sigma_e^{(l)})} \right\}
\]
Recall that $Q(\theta|\theta^{(r)}) = \int \log f(Y, b|\theta)g(b|\theta^{(r)}, Y)db$ can be approximated by replacing the integral with the Monte Carlo samples we obtain from the posterior distribution of $b_i$: 

$$
\frac{1}{M} \sum_{j=1}^{M} \log f(Y, b^{(j)}|\theta).$$

For the $i$th subject, the joint density can be expressed as:

$$
f(Y_i, b_i|\theta) = f(Y_i|b_i, \theta)g(b_i|\theta) = \prod_{t=1}^{n_i} f(y_{it}|b_i, \theta)g(b_i|\Sigma) \quad (2.32)
$$

Since given the random effect $b$, $y_{it}$s are independent, the complete-data likelihood is:

$$
f(Y, b|\theta) = f(Y|b, \theta)g(b|\theta) = \prod_{i=1}^{N} f(Y_i|b_i, \theta)g(b_i|\Sigma) = \prod_{i=1}^{N} \prod_{t=1}^{n_i} f(y_{it}|b_i, \theta)g(b_i|\Sigma)
$$

and let $\alpha^{(l)} = (\beta_1^{(l)}, \beta_2^{(l)}, \sigma_e^{(l)})$, we can write

$$
f(y_i|b_i^*, \alpha^{(l)}) = \prod_{t=1}^{n_i} \{ \frac{1 - r_{it}}{1 + \exp(X_{1it}^T \beta_1^{(l)} + b_{it}^*)} + \frac{r_{it} \exp(X_{1it}^T \beta_1^{(l)} + b_{it}^*)}{1 + \exp(X_{1it}^T \beta_1^{(l)} + b_{it}^*)} \} \Phi(X_{2it}^T \beta_2^{(l)} + b_{it}^*, \sigma_e^{(l)})
$$

After updating $b^{(n)}$ sequentially, we get the next value $b^{(n+1)}$. In other words, to sample from the conditional density in (2.31), we need the following steps:

1. Generate $b_i^{(0)} \sim N(0, \Sigma^{(l)})$.
2. Update each component in $b_i^{(l)}$ sequentially:
   - For $t = 1, \ldots, n_i$, generate $b_i^{*(t)} = (b_{i1}, \ldots, b_{it}, \ldots, b_{iT})$, where
     
     $$
g(b_{it}^*|b_{i1}, \ldots, b_{i,t-1}, b_{i,t+1}, \ldots, b_{iT}) = a \sim N(\mu^*, \Sigma^*).
     $$

     Assume that $\Sigma_{12}$ is the $t$th row of $\Sigma^{(l)}$ without the $t$th component; $\Sigma_{21} = \Sigma_{12}'$; and $\Sigma_{11}$ is just $\sigma_{tt}$ in $\Sigma$. We can write $\mu^* = \Sigma_{12}^{-1} a; \Sigma^* = \Sigma_{11} - \Sigma_{12}^{'} \Sigma_{22}^{-1} \Sigma_{21}$.

     $$
b_i^{*(t+1)} = \begin{cases} 
     b_i^{*(t)} & \text{with probability } \rho_t \\
     b_i^{(l)} & \text{with probability } 1 - \rho_t 
     \end{cases}
$$

     Repeat the process for $t = 1, \ldots, n_i$.

   - After updating $b_i^{(l)}$ sequentially, we let $b_i^{(1)} = b_i^{*(n_i)}$ and set $l=l+1$. Repeat the updating process until convergence.

**Expected Complete-data Likelihood: E-Step**

Recall that $Q(\theta|\theta^{(r)}) = \int \log f(Y, b|\theta)g(b|\theta^{(r)}, Y)db$ can be approximated by replacing the integral with the Monte Carlo samples we obtain from the posterior distribution of $b_i$: 

$$
\frac{1}{M} \sum_{j=1}^{M} \log f(Y, b^{(j)}|\theta).$$

For the $i$th subject, the joint density can be expressed as:

$$
f(Y_i, b_i|\theta) = f(Y_i|b_i, \theta)g(b_i|\theta) = \prod_{t=1}^{n_i} f(y_{it}|b_i, \theta)g(b_i|\Sigma) \quad (2.32)
$$

Since given the random effect $b$, $y_{it}$s are independent, the complete-data likelihood is:

$$
f(Y, b|\theta) = f(Y|b, \theta)g(b|\theta) = \prod_{i=1}^{N} f(Y_i|b_i, \theta)g(b_i|\Sigma) = \prod_{i=1}^{N} \prod_{t=1}^{n_i} f(y_{it}|b_i, \theta)g(b_i|\Sigma)$$

25
Therefore, the Q function above can be written in a nice form:

\[
Q(\Theta; \Theta^{(r)}) = Q_M(\Theta; \Theta^{(r)}) = \frac{1}{M} \sum_{j=1}^{M} \sum_{i=1}^{N} \log(f(Y_i|b_i^{(j)}, \Theta)g(b_i^{(j)}|\Sigma))
\]

\[
= \frac{1}{M} \sum_{j=1}^{M} \sum_{i=1}^{N} n_i \log(f(y_{it}|b_i^{(j)}, \Theta))) + \frac{1}{M} \sum_{j=1}^{M} \sum_{i=1}^{N} \log(g(b_i^{(j)}|\Sigma)).
\] (2.34)

where \( M \) is the MC sample size and \( b_i^{(j)} \) is the \( j^{th} \) sample of the \( i^{th} \) subject.

\[
\log f(y_{it}|b_i^{(j)}, \Theta) = (1 - r_{it}) \log(1 - p_{it}) + r_{it} \log(p_{it}) + r_{it} \log f(V_{it}|\beta_2, b_i^{(j)})
\]

and

\[
\log g(b_i^{(j)}|\Sigma) = \frac{1}{2} \log(1 - \phi_1^2) - \frac{n_i}{2} \log(2\pi) - \frac{n_i}{2} \log(\sigma_1^2) - \frac{1}{2} b_i^{(j)'\Sigma^{-1}b_i^{(j)}}.
\]

Here \( \Sigma \) is the covariance matrix of \( b_i^{(j)} \), which has an AR(1) structure.

\[
\Sigma = \frac{\sigma_1^2}{1 - \phi_1^2} \begin{pmatrix}
1 & \phi_1 & \phi_1^2 & \cdots & \phi_1^{n_i-1} \\
\phi_1 & 1 & \cdots & \cdots & \phi_1^{n_i-2} \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
\phi_1^{n_i-1} & \phi_1^{n_i-2} & \cdots & 1 \\
\end{pmatrix}
\]

and its inverse has a very nice form:

\[
\Sigma^{-1} = \frac{1 - \phi_1^2}{\sigma_1^2} \begin{pmatrix}
\frac{1}{1-\phi_1^2} & -\frac{\phi_1^2}{1-\phi_1^2} & \frac{\phi_1}{1-\phi_1^2} & \cdots & -\frac{1}{1-\phi_1^2} \\
-\frac{\phi_1^2}{1-\phi_1^2} & \frac{1}{1-\phi_1^2} & -\frac{\phi_1}{1-\phi_1^2} & \cdots & -\frac{1}{1-\phi_1^2} \\
\frac{\phi_1}{1-\phi_1^2} & -\frac{\phi_1}{1-\phi_1^2} & \frac{1}{1-\phi_1^2} & \cdots & -\frac{1}{1-\phi_1^2} \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
-\frac{1}{1-\phi_1^2} & \frac{1}{1-\phi_1^2} & -\frac{\phi_1}{1-\phi_1^2} & \cdots & \frac{1}{1-\phi_1^2} \\
\end{pmatrix}
\]

\[
= \frac{1}{\sigma_1^2} \begin{pmatrix}
1 & -\phi_1 & -\phi_1^2 & -\phi_1 & \cdots \\
-\phi_1 & 1 + \phi_1^2 & -\phi_1 & 1 + \phi_1^2 & -\phi_1 \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
-\phi_1 & 1 + \phi_1^2 & -\phi_1 & 1 + \phi_1^2 & -\phi_1 \\
\end{pmatrix}
\]

26
Maximization of the Q-function: M-Step

In the M-step, we maximize the Q-function we described before. In general, the M-step is done by solving the score equations for $\Theta$:

$$
\frac{\partial}{\partial \Theta} Q_m(\Theta | \Theta^{(r)}) = 0
$$

where $m$ is the MC sample size. In the setting of mixed-distribution two part model with shared random effects that have AR(1) covariance structure, we need to differentiate (2.33) and solve for each parameter. Note that it is not always possible to get an explicit solution to the score equation, in such situation, we are going to use Newton-Raphson method.

We start the M-step by updating $\beta_1$.

1. Update $\beta_1$

$$
\frac{\partial}{\partial \beta_1} Q_m = \frac{\partial}{\partial \beta_1} \left\{ \frac{1}{M} \sum_j \sum_i \sum_t \left[ (1 - r_{it}) \log(1 - p_{it}^{(j)}) + r_{it} \log(p_{it}^{(j)}) \right] \right\}
$$

$$
= \frac{1}{M} \sum_j \sum_i \sum_t \left[ \frac{r_{it} - p_{it}^{(j)}}{(1 - p_{it}^{(j)})p_{it}^{(j)}} \times \frac{\partial p_{it}^{(j)}}{\partial \beta_1} \right]
$$

$$
= \frac{1}{M} \sum_j \sum_i \sum_t \left[ \frac{r_{it} - p_{it}^{(j)}}{(1 - p_{it}^{(j)})p_{it}^{(j)}} \times (1 - p_{it}^{(j)})p_{it}^{(j)} \times \frac{\partial (X_{it} \beta_1)}{\partial \beta_1} \right]
$$

$$
= \frac{1}{M} \sum_j \sum_i \sum_t (r_{it} - p_{it}^{(j)}) \frac{\partial (X_{it} \beta_1)}{\partial \beta_1}
$$

(2.35)

where

$$
p_{it}^{(j)} = \frac{\exp(X_{it} \beta_1 + b_{it}^{(j)})}{1 + \exp(X_{it} \beta_1 + b_{it}^{(j)})}
$$

and $X_{it}$ is the vector of covariates of the $i^{th}$ subject at time $t$. Assume that $\beta_1$ is a $(Q \times 1)$, vector, $\frac{\partial X_{it} \beta_1}{\partial \beta_1} = X'_{it}$, where $X_{it}$ is a $1 \times Q$ vector. Therefore, (2.35) is equal to:

$$
\frac{1}{M} \sum_j \sum_i \sum_t (r_{it} - p_{it}^{(j)})X'_{it} = \frac{1}{M} \sum_j X(R - P^{(j)}) \triangleq u_1
$$

where $R = (r_1', \ldots, r_N')'$, and each $r_i = (r_{i1}, \ldots, r_{in_i})'$; similarly, $P^{(j)}$ is a vector that consists of the $j^{th}$ MC sample of $p_{it}$'s. $X' = (X_1', \ldots, X_N')'$ and $X = (X_{i1}, \ldots, X_{in_i})'$ $X'$ is a $Q \times (\sum n_i)$ matrix. In order to apply high-dimensional Newton-Raphson Algorithm,
we need get the second order derivatives of $Q_m$ with respect to $\beta_1$. Similarly as before,

\[
\frac{\partial Q_m}{\partial \beta_1 \partial \beta_1} = -\frac{1}{M} \sum_j \sum_i \sum_t \frac{\partial p_{it}^{(j)}}{\partial \beta_1} X_{it} = -\frac{1}{M} \sum_j \sum_i \sum_t p_{it}^{(j)} (1 - p_{it}^{(j)}) X_{it}' X_{it}
\]

\[
= -\frac{1}{M} \sum_j X \text{Diag}\{p_{it}^{(j)} (1 - p_{it}^{(j)})\} X' \triangleq H_1
\] (2.36)

Now we can update $\beta_1$ by applying the following rule:

\[
\beta_1^{(n+1)} = \beta_1^{(n)} - (H_1^{-1} \times u_1)|_{\beta_1^{(n)}}
\]

2 Update $\beta_2$ and $\sigma^2_e$

Now we are going to update $\beta_2$. $\beta_2$ and $\sigma^2_e$ are parameters in the second part of our model. The first and second order derivatives of the Q-function with respect to $\beta_2$ are as follows:

\[
\frac{\partial Q_m}{\partial \beta_2} = \frac{1}{M} \sum_j \sum_i \sum_t \frac{\partial}{\partial \beta_2} (r_{it} \log(f(V_{it}|\beta_2, b_{it}^{(j)})))
\]

\[
= \frac{1}{M} \sum_j \sum_i \sum_t \frac{\partial}{\partial \beta_2} \left[-\frac{1}{2} \log(2\pi) - \log(\sigma_e) - \frac{(V_{it} - X_{it} \beta_2 - b_{it}^{(j)})^2}{2\sigma^2_e}\right] \times r_{it}
\]

\[
= \frac{1}{M\sigma^2_e} \sum_j \sum_i \sum_t r_{it} (V_{it} - X_{it} \beta_2 - b_{it}^{(j)}) \frac{\partial X_{it} \beta_2}{\partial \beta_2}
\]

\[
= \frac{1}{M\sigma^2_e} \sum_j \sum_i \sum_t r_{it} (V_{it} - X_{it} \beta_2 - b_{it}^{(j)}) X_{it}'
\]

\[
= \frac{1}{M\sigma^2_e} \sum_j X \text{Diag}\{V_{it} - X_{it} \beta_2 - b_{it}^{(j)}\} R
\]

\[
\frac{\partial^2 Q_m}{\partial \beta_2 \partial \beta_2} = -\frac{1}{M\sigma^2_e} \sum_j \sum_i \sum_t r_{it} X_{it}' X_{it} = -\frac{1}{M\sigma^2_e} \sum_j X \text{Diag}\{r_{it}\} X'
\]

Let

\[
u_2 = \frac{\partial Q_m}{\partial \beta_2} \quad H_2 = \frac{\partial^2 Q_m}{\partial \beta_2 \partial \beta_2}
\]

Similarly as before,

\[
\beta_2^{(n+1)} = \beta_2^{(n)} - H_2^{-1} \nu_2|_{\beta_2^{(n)}}
\] (2.37)

where $H_2^{-1}$ is the inverse of $H_2$. 

28
After $\beta_2$ is updated, we can use the updated values to update $\sigma_e^2$.

$$
\frac{\partial Q_m}{\partial \sigma_e} = \frac{1}{M} \sum_j \sum_i \sum_t r_{it} \left[ - \frac{1}{\sigma_e} + \frac{(V_{it} - X_{it}\beta_2 - b_{it}^{(j)})^2}{\sigma_e^3} \right] \quad (2.38)
$$

let (2.38) equal to 0 and solve for $\sigma_e^2$, we have

$$
\sigma_e^2 = \frac{1}{M} \sum_j \sum_i \sum_t r_{it} \frac{(V_{it} - X_{it}\beta_2 - b_{it}^{(j)})^2}{\sum_i \sum_t r_{it}} = \frac{1}{M} \frac{WR}{RR'} \quad (2.39)
$$

where $W = (w_{11}, \cdots, w_{\Sigma_n})'$, and $w_{it} = \sum_j (V_{it} - X_{it}\beta_2 - b_{it}^{(j)})^2$.

3 Update $\phi_1$ and $\sigma_1^2$

$\phi_1$ and $\sigma_1^2$ are parameters from the covariance matrix of the random effect $b$. We are updating them at the same time and we will adopt similar approach as above. Before computing the derivatives, we will first introduce some notations. Let

$$
a_i^{(j)} = \exp\left(-\frac{1}{2} b_i^{(j)'\Sigma^{-1}b_i^{(j)}}\right), \quad \alpha = \sigma_1^2
\quad \Sigma^{-1} = \frac{1}{\alpha} B
$$

$$
B = \begin{pmatrix}
1 & -\phi_1 \\
-\phi_1 & 1 + \phi_1^2 & -\phi_1 \\
& -\phi_1 & 1 + \phi_1^2 & -\phi_1 \\
& & & \ddots & \ddots \\
& & & & 1 + \phi_1^2 & -\phi_1 \\
& & & & -\phi_1 & 1
\end{pmatrix}
$$

$$
\frac{\partial B}{\partial \phi_1} = \begin{pmatrix}
0 & -1 \\
-1 & 2\phi_1 & -1 \\
& -1 & 2\phi_1 & -1 \\
& & & \ddots & \ddots \\
& & & & 2\phi_1 & -1 \\
& & & & -1 & 0
\end{pmatrix} \triangleq C
$$

$$
-\frac{1}{2} b_i^{(j)' \frac{\partial \Sigma^{-1}}{\partial \phi_1} b_i^{(j)}} = -\frac{1}{2\alpha} b_i^{(j)' C b_i^{(j)}} \triangleq q_i^{(j)}
$$
\[-\frac{1}{2} b_i^{(j)} \partial^2 \Sigma^{-1} \partial \phi_1 b_i^{(j)} = -\frac{1}{2\alpha} b_i^{(j)} \text{Diag}\{0, 2, \cdots, 2, 0\} b_i^{(j)} \triangleq p_i^{(j)} \]

\[
\frac{\partial \Sigma^{-1}}{\partial \sigma_i^2} = \frac{\partial \Sigma^{-1}}{\partial \alpha} = -\frac{1}{\alpha^2} B
\]

Last, let

\[h_i^{(j)} = \frac{1}{2} b_i^{(j)'} B b_i^{(j)}\]

Then we have,

\[
\frac{\partial Q_m}{\partial \phi_1} = \frac{1}{M} \sum_j \sum_i \log(g(b_i^{(j)} | \Sigma)) = \frac{1}{M} \sum_j \sum_i \left( -\frac{n_i}{2} \log(1 - \phi_i^2) - \frac{n_i}{2} \log(2\pi) - \frac{n_i}{2} \log \sigma_i^2 - \frac{1}{2} b_i^{(j)'} \Sigma^{-1} b_i^{(j)} \right)
\]

\[
= \frac{1}{M} \sum_j \sum_i \left( -\phi_i + q_i^{(j)} \right) = \frac{\tilde{Q} 1}{1 - \phi_i^2} + \frac{1}{M} \tilde{Q} 1
\]

\[
\frac{\partial^2 Q_m}{\partial \phi_i^2} = \frac{1}{M} \sum_j \sum_i \left( -\frac{1}{(1 - \phi_i^2)^2} + p_i^{(j)} \right) = -\frac{N(1 + \phi_i^2)}{(1 - \phi_i^2)^2} + \frac{1}{M} \tilde{P} 1
\]

where \(\tilde{Q} = \sum_{j=1}^M q_1^{(j)}, \cdots, \sum_{j=1}^M q_N^{(j)}\)' and \(1\) is a \(N \times 1\) vector of 1's.

\[
\frac{\partial Q_m}{\partial \alpha} = \frac{1}{M} \sum_j \sum_i \left( -\frac{n_i}{2} \log(\alpha) - \frac{1}{2} b_i^{(j)'} \Sigma^{-1} b_i^{(j)} \right)
\]

\[
= -\frac{\Sigma n_i}{2\alpha} + \frac{1}{M\alpha^2} \sum_i \sum_j h_i^{(j)} \triangleq -\frac{\Sigma n_i}{2\alpha} + \frac{1}{M\alpha^2} \tilde{H} 1
\]

\[
\frac{\partial^2 Q_m}{\partial \alpha^2} = \frac{\Sigma n_i}{2\alpha^2} - \frac{2}{M\alpha^3} \sum_i \sum_j h_i^{(j)} = \frac{\Sigma n_i}{2\alpha^2} - \frac{2}{M\alpha^3} \tilde{H} 1
\]

\[
\frac{\partial^2 Q_m}{\partial \alpha \partial \phi_1} = \frac{\partial^2 Q_m}{\partial \phi_1 \partial \alpha} = \frac{1}{M} \sum_j \sum_i -\frac{1}{\alpha} q_i^{(j)} = -\frac{1}{\alpha M} \tilde{Q} 1.
\]

Let

\[
u_3 = \left( \begin{array}{c} \frac{\partial Q_m}{\partial \phi_1} \\ \frac{\partial Q_m}{\partial \alpha} \end{array} \right) \quad H_3 = \left( \begin{array}{ccc} \frac{\partial^2 Q_m}{\partial \phi_1^2} & \frac{\partial^2 Q_m}{\partial \phi_1 \partial \alpha} \\ \frac{\partial^2 Q_m}{\partial \phi_1 \partial \alpha} & \frac{\partial^2 Q_m}{\partial \alpha^2} \end{array} \right)
\]

30
The last piece of our calculation would be
\[
\begin{pmatrix}
\phi_1^{(n+1)} \\
\alpha_1^{(n+1)}
\end{pmatrix} = \begin{pmatrix}
\phi_1^{(n)} \\
\alpha_1^{(n)}
\end{pmatrix} - H_3^{-1} u_3_{(\phi_1^{(n)}, \alpha_1^{(n)})},
\]
(2.45)

### 2.3.3 Estimating Standard Errors

Now we discuss how to estimate the observed information matrix in our model. Recall that in order to estimate parameters, we adopted EM algorithm and therefore, the Fisher Information matrix we obtained was associated with the complete data but not the observed data. In this section, we will follow Louis’ [22] approach to compute the observed information matrix.

Following the same notation, we let \( S \) and \( S^* \) be the gradient vectors of the log likelihood function \( l \) for complete data and \( l^* \) for observed data respectively and \( B \) and \( B^* \) be the negative associated second derivatives. The observed information can be computed as follows:
\[
I(\theta) = E(B) - E(SS^*) + S^* S^* \tag{2.46}
\]
and
\[
S^* = E(S) \tag{2.47}
\]

Equation (2.46) shows that the observed information is the difference between full data information matrix and expected information for the conditional distribution of complete data. Before we apply this formula, let’s review some important facts. The complete-data likelihood of our model is:
\[
f(Y, b| \Theta) = f(Y| b, \Theta)g(b| \Theta) = \prod_{i=1}^{N} f(Y_i| b_i, \Theta)g(b_i| \Sigma) = \prod_{i=1}^{N} \prod_{t=1}^{n_i} f(y_{it}| b_i, \Theta)g(b_i| \Sigma)
\]
and the corresponding log-likelihood function is:
\[
\sum_{i=1}^{N} \log(f(Y_i| b_i, \Theta)g(b_i| \Sigma)) = \sum_{i=1}^{N} \sum_{t=1}^{n_i} \log(f(y_{it}| b_i, \Theta)) + \sum_{i=1}^{N} \log(g(b_i| \Sigma)).
\]

Now we will find the observed information matrix for each parameter in this model. Similarly as before, we take first and second order derivatives of the log-likelihood function with respect to each parameter:
\[
S(\beta_1) = \frac{\partial l}{\partial \beta_1} = \sum_{i=1}^{N} \sum_{t=1}^{n_i} (r_{it} - p_{it}) \frac{X_{it} \beta_1}{\partial \beta_1} = \sum_{i=1}^{N} \sum_{t=1}^{n_i} (r_{it} - p_{it}) X'_{it} = X(R - P)
\]
where $X$ and $R$ are defined same as in the last section while $P = (p_{11}, \cdots, p_{Nn})'$ and $p_{it} = \frac{\exp(X_{it}\beta_1 + b_{it})}{1 + \exp(X_{it}\beta_1 + b_{it})}$. Following (2.47), we can compute $S^*(\beta_1)$ as follows:

$$S^*(\beta_1) = E(S(\beta_1)) = X(R - E(P))$$

Notice that $E(P)$ in the last term in the equation above is difficult to evaluate, therefore similarly as before, we will use the MC samples we drew in the last step of our estimation process to compute this quantity:

$$S^*(\beta_1) \approx X(R - \frac{1}{M} \sum_{j=1}^{M} P^{(j)})$$

where $P^{(j)}$ is the vector of the $j^{th}$ Monte Carlo sample of $p_{it}$'s. And to calculate the second quantity in (2.46), we have:

$$E(S(\beta_1)S(\beta_1)') = E(XRR'X' - XPR'X' - XR'P'X' + XPP'X')$$

$$\approx XRR'X - X \left( \frac{1}{M} \sum_j P^{(j)} \right) R'X' - XR \left( \frac{1}{M} \sum_j P^{(j)} \right) X' + X \left( \frac{1}{M} \sum_j P^{(j)} P^{(j)'} \right) X'$$

and

$$B(\beta_1) = -\frac{\partial S(\beta_1)}{\partial \beta_1} = \sum_i \sum_t p_{it}(1 - p_{it}X'_{it}X_{it}) = X\text{Diag}(P(1 - P))X'.$$

$$E(B(\beta_1)) \approx -\frac{1}{M} \sum_j X\text{Diag}\{P^{(j)}(1 - P^{(j)})\}X'$$

this has the same expression as (2.35).

We can also get the observed information for other parameters using the same method. For $\beta_2$, we have the following formulas:

$$S(\beta_2) = \sum_i \sum_j \frac{\partial}{\partial \beta_2} (r_{it}\log(f(V_{it}|\beta_2, b_i))) = \frac{1}{\sigma^2_e} X\text{Diag}(V - X\beta_2 - b)R$$

And $S^*(\beta_2)$ can be approximated as follows:

$$S^*(\beta_2) = E(S(\beta_2)) \approx \frac{1}{\sigma^2_e} X\text{Diag}\left(V - X\beta_2 - \frac{1}{M} \sum_j b_{it}^{(j)}\right)R$$
\[
E(S^*(\beta_2)S^*(\beta_2)') \\
\approx \frac{1}{\sigma_e^4} \left( X * \text{Diag}(V - X\beta_2)RR' * \text{Diag}(V - X\beta_2)X' - X\text{Diag}(V - X\beta_2)RR'\text{Diag}(\frac{1}{M} \sum_j b^{(j)}X') \right) \\
- \frac{1}{\sigma_e^4} \left( X\text{Diag}(\frac{1}{M} \sum_j b^{(j)}RR')\text{Diag}(V - X\beta_2)X' + X\frac{1}{M} \sum_j \text{Diag}(b^{(j)}RR')\text{Diag}(b^{(j)})X' \right)
\]
and
\[
B(\beta_2) = - \frac{\partial S(\beta_2)}{\partial \beta_2} = - \frac{1}{\sigma_e^2} X\text{Diag}(r_{it})X'.
\]
Note that the expression of \( B(\beta_2) \) does not depend on the random effect \( b \), therefore, 
\( E(B(\beta_2)) = B(\beta_2) \).
Similarly, for the rest of parameters \( \sigma_e^2, \phi_1 \) and \( \sigma_1^2 \), the computation is straightforward. For \( \sigma_e^2 \), we have:
\[
S(\sigma_e^2) = \sum_{i=1}^N \frac{\partial (g(b_i|\Sigma))}{\partial \sigma_e^2} = \frac{1}{2\sigma_e^2} \left( \sum_i \sum r_{it}(V_{it} - X_{it}\beta_2 - b_{it})^2 \right) - \sum_i \sum r_{it} \triangleq \frac{1}{2\sigma_e^2} \left( \tilde{W}R - R'1 \right)
\]
and
\[
S^*(\sigma_e^2) = E(S(\sigma_e^2)) \approx \frac{1}{2\sigma_e^2} \left( \sum_i \sum r_{it}(V_{it} - X_{it}\beta_2 - b_{it})^2 \right) - \sum_i \sum r_{it} \triangleq \frac{1}{2\sigma_e^2} \left( \tilde{W}R - R'1 \right)
\]
\[
E(S(\sigma_e^2)^2) = E \left( \frac{1}{4\sigma_e^2} \left( \tilde{W}R - R'1 \right) \left( \frac{R'\tilde{W}'}{\sigma_e^2} - 1'R \right) \right) \approx \frac{1}{4\sigma_e^2} \left( \frac{1}{M\sigma_e^4} \sum W^{(j)}RR'W^{(j)'} - \frac{2}{\sigma_e^2M} \sum W^{(j)}R1'R + R1'1'R \right)
\]
where \( R = (r_{11}, \ldots, r_{NNN})' \), \( \tilde{W} = (\tilde{w}_{11}, \ldots, \tilde{w}_{NNN})' \), \( \tilde{w}_{it} = (V_{it} - X_{it}\beta_2 - b_{it})^2 \) and \( W = (w_{11}, \ldots, w_{NNN})' \), \( w_{it} = \sum_j (V_{it} - X_{it}\beta_2 - b_{it})^2 \). Next we will compute the information matrix for \( \phi_1 \) and \( \sigma_1^2 \) together.
\[
S(\phi_1, \sigma_1^2) = \begin{pmatrix}
\sum_i \left( -\frac{\phi_1}{1-\phi_1^2} + q_i \right) \\
-\sum_i \frac{n_i}{2a} + \frac{1}{an} \sum_i h_i
\end{pmatrix}
\]
where \( q_i = -\frac{1}{2} \mu_i \frac{\partial \Sigma_{-1}}{\partial \phi_1} b_i \) and \( h_i = \frac{1}{2} b_i^T B b_i \). Here matrix \( B, \tilde{H} \) and \( \tilde{Q} \) have been defined in the last section when we updated \( \phi_1 \) and \( \sigma_1^2 \).
\[
S^*(\phi_1, \sigma_1^2) = E(S(\phi_1, \sigma_1^2)) \approx \begin{pmatrix}
-\frac{N}{1-\phi_1^2} + \frac{1}{M} \tilde{Q}1 \\
-\sum_i \frac{n_i}{2a} + \frac{1}{M} \tilde{H}1 \end{pmatrix} \triangleq \begin{pmatrix}
\left( a + \frac{1}{M} \tilde{Q}1 \right) \\
\left( b + \frac{1}{Ma} \tilde{H}1 \right)
\end{pmatrix}
\]
and next step follows:

\[ E(S(\phi_1, \sigma_1^2) \ast S(\phi_1, \sigma_1^2)'') \approx \left( \begin{array}{cc}
\frac{1}{M} \sum (a + \bar{Q}_1)^2 & \frac{1}{M} \sum (a + \bar{Q}_1)(b + \bar{H}_1) \\
\frac{1}{M} \sum (b + \bar{H}_1)(a + \bar{Q}_1) & \frac{1}{M} (b + \bar{H}_1)
\end{array} \right) \]

And \( E(B(\phi_1, \sigma_1^2)) \) can be approximated by matrix \(-H_3\) in section 2.3.2.
CHAPTER 3
Simulation Study

3.1 Simulation Set-up

We conducted a simulation study to investigate the performance of MCEM algorithm on estimation of parameters and the associated standard errors of the two-part model with AR(1) covariance structure random effects and the impact of model violations on the parameter estimates in this type of models. For the linear mixed model, a proper random effects structure is essential for standard error estimates. And several researchers explored misspecification of random effects structure in generalized linear mixed models. Heagerty and Kurland [17] suggest that for generalized linear mixed models, the selection of the random effects structure may be required both for valid point estimates and for correct standard errors. More specifically, simulation by Heagerty and Kurland have shown that for generalized linear mixed models that have autoregressive random effects $b_{it}$ with

$$\text{Cov}(b_{it}, b_{it'}) = \frac{\phi |t' - t| \sigma_i^2}{1 - \phi^2},$$

misspecification of the random effects structure lead to substantial negative bias in the estimated fixed effects.

In our simulation study, data were simulated from two-part logistic normal models with AR(1) covariance structure using a few different sets of parameter values. Here we first briefly describe the set-up of our simulation study.

To simplify the simulation study, we assume that design matrices $X_1$ and $X_2$ are identical. Their first columns are two vectors consist of 1 while the second columns denote different time point with numbers going from 1 to $T$ repeatedly; the third columns include 1’s and 0’s which are designed to mimic the indicator of restatement; and the last column is the product of the second and third column, which can be viewed as the interaction between
time and restatement. The data were generated from the following model:

\[ R_{it} = \begin{cases} 
0 & \text{if } Y_{it} = 0 \\
1 & \text{if } Y_{it} > 0 
\end{cases} \]

\[ P(R_{it} = r_{it} | \theta_1) = \begin{cases} 
1 - p_{it}(\theta_1), & \text{if } r_{it} = 0 \\
p_{it}(\theta_1), & \text{if } r_{it} = 1 
\end{cases} . \]

\[ \logit(p_{it}(\theta_1)) = \log\left(\frac{p_{it}(\theta_1)}{1 - p_{it}(\theta_1)}\right) = \eta_{it}(\theta_1). \]

\[ \begin{cases} 
\eta_{it} = X_{1i}\beta_1 + b_{it} \\
V_{it} = X_{2i}\beta_2 + b_{it} + \epsilon_{it} 
\end{cases} \]

where \( i = 1, \ldots, N \) and \( t = 1, \ldots, T \), and random effect \( b_i \) has an AR(1) covariance structure:

\[ \text{Var}(b_i) = \frac{\sigma_1^2}{(1 - \phi_1^2)} \begin{pmatrix} 1 & \phi_1 & \phi_1^2 \\ \phi_1 & 1 & \phi_1 \\ \phi_1^2 & \phi_1 & 1 \end{pmatrix} \]

### 3.2 Simulation Results

#### 3.2.1 Parameter Estimates and Standard Errors

In this section, we compare parameter estimates and standard errors calculated using the MCEM method and the adaptive Gaussian quadrature method in SAS when simulated data are generated from a two-part model with autoregressive random effects. The adaptive Gaussian quadrature (AGQ) method is a numerical integration method for calculating the marginal maximum likelihood. The method was described in Pinheiro and Bates (1995) [29]. It approximates a given integral by using a weighted sum over the random effects. The AGQ is the default approximation method used in SAS procedure “nlmixed”. There is no written procedure available in SAS that fits two-part mixed distribution models. The two-part model with correlated random effects proposed by Tooze et. al. [35] was fitted using a SAS macro written by the authors. And even though proc nlmixed allows users to write their own likelihood functions and specify distribution of random effects, it is very difficult to incorporate autoregressive covariance structure into the model. The program we used to fit the two-part model with AR(1) covariance structure random effect was written in Matlab.
To demonstrate how both methods perform in a simplified situation, we first generated data from a two-part mixed distribution model without AR(1) structure, in other words, the autocorrelation coefficient $\phi_1 = 0$. We generated 800 subjects with 6 time points for each subject. Note that even though we set $\phi_1 = 0$, there is a slight difference between our model and the two-part model with correlated random effects (Tooze) shown in section 2.2.2. The random effect $b_i$ in our model is a random vector that follows a multivariate distribution, which means we have different realization of the random effect at each time point even though all $b_{it}$’s are independent and identically distributed when $\phi_1 = 0$. However, the random effects in the two-part model with correlated random effects introduced by Tooze et. al. remain constant within each subject. Therefore, we expect to see the standard errors estimated by the MCEM and adaptive Gaussian quadrature methods to be similar but not identical in this example. Table 3.1 shows parameter estimates and their standard errors for this dataset. The first column of the table shows parameters that are being estimated; the second column is the true values we assigned to the parameters; the third and fourth columns report parameter estimated we obtained from MCEM method and the last two columns report results from SAS. From Table 3.1 we see that in this particular example, the MCEM estimates for the first part of the model were closer to the true values than estimates from SAS. Moreover, the adaptive Gaussian method does not seem to decompose the variances correctly as we see that the estimate of residual variance departs from its true value. Other than that, the estimates of standard errors for all parameters are very similar but not identical as we expected.

Next, we generated another dataset using the same set of parameter values except a different value of autocorrelation parameter $\phi_1$. Instead of no autocorrelation, we now assume that $\phi_1 = 0.5$ and examine the performance of both methods. Other than parameters estimates and standard errors, we are also interested in finding out if there is any difference between the significance levels. Simulation setting was the same as in the last example: we generated 800 subjects and each subject has 6 observations. Results are shown in Table 3.2. Similarly, most parameter estimates from both methods appear to be unbiased but the adaptive Gaussian method does not decompose the variances correctly. In this example, we further compared corresponding p-values for each parameter in the model. Now that the autocorrelation is no longer zero, it turns out that the difference in standard errors between the two methods is greater even though the significance levels suggested by the p-values are
Table 3.1: Dataset 1: the Two-part mixed model with $\phi_1=0$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>MCEM Estimate</th>
<th>Standard Error</th>
<th>MCEM Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{10}$</td>
<td>1</td>
<td>0.65</td>
<td>0.43</td>
<td>0.49</td>
<td>0.42</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>2</td>
<td>2.08</td>
<td>0.34</td>
<td>2.02</td>
<td>0.33</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>-1</td>
<td>-0.81</td>
<td>0.45</td>
<td>-0.67</td>
<td>0.44</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>-2</td>
<td>-2.03</td>
<td>0.34</td>
<td>-1.97</td>
<td>0.33</td>
</tr>
<tr>
<td>$\beta_{20}$</td>
<td>6</td>
<td>6.25</td>
<td>0.07</td>
<td>6.26</td>
<td>0.09</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>5</td>
<td>4.94</td>
<td>0.02</td>
<td>4.94</td>
<td>0.02</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>-3</td>
<td>-3.26</td>
<td>0.19</td>
<td>-3.14</td>
<td>0.21</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>1</td>
<td>1.10</td>
<td>0.05</td>
<td>1.12</td>
<td>0.05</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>1</td>
<td>0.74</td>
<td>0.08</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0</td>
<td>0.00</td>
<td>0.08</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>2</td>
<td>2.12</td>
<td>0.09</td>
<td>2.23</td>
<td>0.11</td>
</tr>
</tbody>
</table>

identical in this example.

After comparing standard errors using data generated from our two-part model without and with serial correlation, we generated data from Tooze’s model and examined standard errors estimated by each method. The random effects in this example remain constant within each subject and differ from subject to subject. Results are summarized in Table 3.3. In this example, both methods seem to yield unbiased parameter estimates and generally speaking, the MCEM method yields more accurate estimates than the adaptive Gaussian method. As for the standard errors, both methods have similar values of standard errors; furthermore, the associated p-values suggest identical significance levels.

Simulation Results: Short Time Period

Now that we have tested the performance of the MCEM and the adaptive Gaussian methods in terms of estimating parameters and standard errors through a few examples, we are going to increase the number of simulations and compare their performance on average.

Since simulation sample size necessary to make MCEM algorithm perform very accurately is usually not feasible, the estimates will “converge” close to the correct answer and then vary in the neighborhood of the correct answer [23]. For each dataset, we ran 10 iterations since we found that the estimates jump to a close neighborhood of the true value within 3 to 4 iterations. And we allowed a burn-in period of 5000 generations of samples from the
Table 3.2: Simulation Results for the Two-part mixed model with $\phi_1=0.5$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>MCEM Value</th>
<th>MCEM Standard Error</th>
<th>MCEM p-value</th>
<th>SAS Value</th>
<th>SAS Standard Error</th>
<th>SAS p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{10}$</td>
<td>1</td>
<td>1.18</td>
<td>0.28</td>
<td>&lt;0.001</td>
<td>1.06</td>
<td>0.34</td>
<td>0.002</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>2</td>
<td>1.63</td>
<td>0.24</td>
<td>&lt;0.001</td>
<td>1.58</td>
<td>0.23</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>-1</td>
<td>-1.19</td>
<td>0.28</td>
<td>&lt;0.001</td>
<td>-1.07</td>
<td>0.36</td>
<td>0.003</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>-2</td>
<td>-1.62</td>
<td>0.24</td>
<td>&lt;0.001</td>
<td>-1.58</td>
<td>0.24</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_{20}$</td>
<td>6</td>
<td>6.14</td>
<td>0.05</td>
<td>&lt;0.001</td>
<td>6.15</td>
<td>0.09</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>5</td>
<td>4.98</td>
<td>0.02</td>
<td>&lt;0.001</td>
<td>4.97</td>
<td>0.02</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>-3</td>
<td>-2.88</td>
<td>0.10</td>
<td>&lt;0.001</td>
<td>-2.80</td>
<td>0.22</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>1</td>
<td>1.03</td>
<td>0.05</td>
<td>&lt;0.001</td>
<td>1.06</td>
<td>0.06</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>1</td>
<td>0.84</td>
<td>0.07</td>
<td>&lt;0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.5</td>
<td>0.33</td>
<td>0.03</td>
<td>&lt;0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>2</td>
<td>2.12</td>
<td>0.10</td>
<td>&lt;0.001</td>
<td>2.25</td>
<td>0.12</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

posterior so that we obtained stable Monte Carlo samples. The simulation set-up is similar as before except for each set of parameters, we generated 20 different datasets. The parameter estimates and standard errors were calculated by taking the averages of all 20 estimates and standard errors, respectively. The simulation results are shown below in tables 3.4- 3.6. In these tables we show parameter true values, their estimates and the associated standard errors.

Keep in mind that due to high computational cost, we used only 200 subject for these five simulation experiments. Yet, the estimates from the MCEM method appear to be unbiased. Compared to previous single dataset simulation examples, we see that by increasing the simulation size, we get more accurate estimates. When comparing results obtained using MCEM and SAS, we notice that on average, MCEM algorithm performs better than the adaptive Gaussian quadrature method in terms of estimation accuracy. Moreover, as we expected, the precision of estimate of variance of residuals obtained from Gaussian quadrature is poorer than that of the MCEM estimates. For these four sets of parameter values, we found that standard error estimates of coefficients in the model are very similar between the MCEM method and the Gaussian adaptive method and the Gaussian method over-estimates the standard errors of residual variances.
Table 3.3: Simulation Results for the Two-part mixed model (Tooze) with correlated random effects

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>MCEM Value</th>
<th>Standard Error</th>
<th>p-value</th>
<th>SAS Value</th>
<th>Standard Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{10}$</td>
<td>1</td>
<td>1.3</td>
<td>0.34</td>
<td>&lt;0.001</td>
<td>1.47</td>
<td>0.41</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>2</td>
<td>1.80</td>
<td>0.30</td>
<td>&lt;0.001</td>
<td>1.74</td>
<td>0.30</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>-1</td>
<td>-1.28</td>
<td>0.34</td>
<td>&lt;0.001</td>
<td>-1.48</td>
<td>0.44</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>-2</td>
<td>-1.86</td>
<td>0.30</td>
<td>&lt;0.001</td>
<td>-1.80</td>
<td>0.30</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_{20}$</td>
<td>6</td>
<td>6.05</td>
<td>0.04</td>
<td>&lt;0.001</td>
<td>6.02</td>
<td>0.09</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>5</td>
<td>4.99</td>
<td>0.02</td>
<td>&lt;0.001</td>
<td>5.00</td>
<td>0.02</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>-3</td>
<td>-2.84</td>
<td>0.10</td>
<td>&lt;0.001</td>
<td>-3.07</td>
<td>0.21</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>1</td>
<td>0.95</td>
<td>0.05</td>
<td>&lt;0.001</td>
<td>0.99</td>
<td>0.06</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>1</td>
<td>0.97</td>
<td>0.08</td>
<td>&lt;0.001</td>
<td>1.04</td>
<td>0.09</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>2</td>
<td>2.03</td>
<td>0.08</td>
<td>&lt;0.001</td>
<td>1.98</td>
<td>0.09</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Table 3.4: Simulation Results: Set One

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{10}$</td>
<td>1</td>
<td>0.92</td>
<td>0.2678</td>
<td>0.93</td>
<td>0.2725</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>-0.5</td>
<td>-0.48</td>
<td>0.1241</td>
<td>-0.49</td>
<td>0.1260</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>2</td>
<td>2.21</td>
<td>0.5872</td>
<td>2.24</td>
<td>0.5937</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>-0.8</td>
<td>-0.87</td>
<td>0.2624</td>
<td>-0.89</td>
<td>0.2646</td>
</tr>
<tr>
<td>$\beta_{20}$</td>
<td>6</td>
<td>6.01</td>
<td>0.0328</td>
<td>6.02</td>
<td>0.0495</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>-1</td>
<td>-1</td>
<td>0.0157</td>
<td>-1</td>
<td>0.0240</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>2</td>
<td>2.007</td>
<td>0.0678</td>
<td>2.001</td>
<td>0.0840</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>1</td>
<td>0.998</td>
<td>0.0355</td>
<td>1.001</td>
<td>0.0436</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.2</td>
<td>0.15</td>
<td>0.0059</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.5</td>
<td>0.38</td>
<td>0.056</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>0.2</td>
<td>0.236</td>
<td>0.0022</td>
<td>0.249</td>
<td>0.011</td>
</tr>
</tbody>
</table>

**Simulation Results: Increased Time Period**

In order to find out if the MCEM performs better than the Gaussian adaptive when we have longer time period and AR(1) correlation structure present, we increased the number of time points and repeated the simulation process 100 times. Table 3.8 shows the summary of our simulation results.

This tables shows that when we have a longer time period and the AR(1) correlation
Table 3.5: Simulation Results: Set Two

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimates (MCEM)</th>
<th>StErr.</th>
<th>Estimates (SAS)</th>
<th>StErr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{10}$</td>
<td>1</td>
<td>0.89</td>
<td>0.4509</td>
<td>0.86</td>
<td>0.4290</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>1</td>
<td>0.98</td>
<td>0.2737</td>
<td>0.93</td>
<td>0.2497</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>-1</td>
<td>-0.62</td>
<td>0.6756</td>
<td>-0.57</td>
<td>0.6442</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>-2</td>
<td>-2.04</td>
<td>0.3843</td>
<td>-1.96</td>
<td>0.3741</td>
</tr>
<tr>
<td>$\beta_{20}$</td>
<td>6</td>
<td>6.09</td>
<td>0.2300</td>
<td>6.21</td>
<td>0.2904</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>2</td>
<td>1.95</td>
<td>0.1092</td>
<td>1.91</td>
<td>0.1268</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>3.5</td>
<td>3.51</td>
<td>0.8050</td>
<td>3.58</td>
<td>0.8820</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>-3</td>
<td>-2.85</td>
<td>0.3843</td>
<td>-2.70</td>
<td>0.5010</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>1</td>
<td>0.85</td>
<td>0.2086</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.6</td>
<td>0.44</td>
<td>0.0359</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>2</td>
<td>2.07</td>
<td>0.1915</td>
<td>2.14</td>
<td>0.3740</td>
</tr>
</tbody>
</table>

Table 3.6: Simulation Results: Set Three

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimates (MCEM)</th>
<th>StErr.</th>
<th>Estimates (SAS)</th>
<th>StErr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{10}$</td>
<td>0.5</td>
<td>0.53</td>
<td>0.3185</td>
<td>0.46</td>
<td>0.2671</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>-0.5</td>
<td>-0.55</td>
<td>0.1517</td>
<td>-0.43</td>
<td>0.1258</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>4</td>
<td>4.23</td>
<td>1.2272</td>
<td>3.91</td>
<td>1.1617</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>-0.2</td>
<td>-0.22</td>
<td>0.5053</td>
<td>-0.39</td>
<td>0.4658</td>
</tr>
<tr>
<td>$\beta_{20}$</td>
<td>2</td>
<td>1.82</td>
<td>0.0560</td>
<td>2.45</td>
<td>0.2044</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>1</td>
<td>0.97</td>
<td>0.0308</td>
<td>1.02</td>
<td>0.1031</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>1</td>
<td>0.99</td>
<td>0.0824</td>
<td>0.82</td>
<td>0.2994</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>-0.5</td>
<td>-0.39</td>
<td>0.0399</td>
<td>-0.45</td>
<td>0.1453</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>1</td>
<td>1.19</td>
<td>0.1098</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>-0.2</td>
<td>-0.12</td>
<td>0.0843</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>0.5</td>
<td>0.29</td>
<td>0.0072</td>
<td>1.03</td>
<td>0.0958</td>
</tr>
</tbody>
</table>

A coefficient that is significantly different from zero, the MCEM method yields more accurate parameter estimates than Gaussian Quadrature method. And we note that when the overall sample size increases, the MCEM algorithm converges much faster than it does when the sample size is small. We also noticed when we increase the sample size, the MCEM results improve significantly. Similarly as before, we calculated standard errors for each dataset used in the simulations and we took the average of these values to be the estimates of standard errors. Recall that we’ve seen in the first four simulation that the Gaussian adaptive
Table 3.7: Simulation Results: Set Four

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimates (MCEM)</th>
<th>StErr.</th>
<th>Estimates (SAS)</th>
<th>StErr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{10}$</td>
<td>1</td>
<td>0.71</td>
<td>0.4110</td>
<td>0.74</td>
<td>0.4203</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>1</td>
<td>1.02</td>
<td>0.2423</td>
<td>0.99</td>
<td>0.2426</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>-1</td>
<td>-0.82</td>
<td>0.6413</td>
<td>-0.89</td>
<td>0.6608</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>-2</td>
<td>-1.92</td>
<td>0.3588</td>
<td>-1.86</td>
<td>0.3673</td>
</tr>
<tr>
<td>$\beta_{20}$</td>
<td>6</td>
<td>6.14</td>
<td>0.2505</td>
<td>6.22</td>
<td>0.2680</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>2</td>
<td>1.95</td>
<td>0.1209</td>
<td>1.91</td>
<td>0.1223</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>3.5</td>
<td>3.57</td>
<td>0.9095</td>
<td>3.59</td>
<td>0.9449</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>-3</td>
<td>-2.95</td>
<td>0.5076</td>
<td>-2.80</td>
<td>0.5295</td>
</tr>
<tr>
<td>$\sigma_1^*$</td>
<td>1</td>
<td>0.56</td>
<td>0.1315</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.2</td>
<td>0.1</td>
<td>0.6300</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>2</td>
<td>2.09</td>
<td>0.2652</td>
<td>2.12</td>
<td>0.3112</td>
</tr>
</tbody>
</table>

Table 3.8: Simulation Results: Set Five

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimates (MCEM)</th>
<th>StErr.</th>
<th>Estimates (SAS)</th>
<th>StErr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{10}$</td>
<td>1</td>
<td>0.86</td>
<td>0.7941</td>
<td>0.35</td>
<td>0.5911</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>2</td>
<td>2.06</td>
<td>0.4880</td>
<td>1.57</td>
<td>0.3959</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>-1</td>
<td>-0.74</td>
<td>0.9572</td>
<td>-0.27</td>
<td>0.6750</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>-2</td>
<td>-2.08</td>
<td>0.5063</td>
<td>-1.58</td>
<td>0.4040</td>
</tr>
<tr>
<td>$\beta_{20}$</td>
<td>6</td>
<td>6.03</td>
<td>0.4501</td>
<td>6.38</td>
<td>0.2809</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>5</td>
<td>4.98</td>
<td>0.0993</td>
<td>4.92</td>
<td>0.0658</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>-3</td>
<td>-2.92</td>
<td>0.7924</td>
<td>-2.07</td>
<td>0.6435</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>1</td>
<td>1.02</td>
<td>0.1869</td>
<td>1.08</td>
<td>0.1579</td>
</tr>
<tr>
<td>$\sigma_1^*$</td>
<td>2</td>
<td>2.00</td>
<td>0.5235</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.5</td>
<td>0.51</td>
<td>0.0405</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>2</td>
<td>1.96</td>
<td>0.1510</td>
<td>2.65</td>
<td>0.4272</td>
</tr>
</tbody>
</table>

method mixes variances of the random effects and variance of residuals and overestimates the standard errors for the variances of residuals. For simulation 5, the standard error estimates from the two methods are farther apart than the ones in simulation 1 to 4. In summary, we have shown that in the two-part model with serial correlation with different parameter values and different length of time, both the MCEM and Gaussian adaptive method yield unbiased parameter estimates and similar standard error estimates.
CHAPTER 4

Applications

4.1 APPLICATION TO FINANCE

4.1.1 Description of the Restatement and Executive Compensation Data

Data Sources

For the application of our model to financial data, we first obtained a list of restatements from the General Accounting Office (GAO) studies (2002 and 2006)[25][26]. The list includes financial restatements due to accounting irregularities announced from January 1997 through June 2006. These restatements were identified by the General Accounting Office through a Lexus-Nexus database key word search. These companies restated one or more previously issued financial statements. However, the GAO reports do not detailed information on years misreported. To get the detailed misreported periods corresponding to the restatement announcements, we searched the Audit Analytic Database (Audit Analytic dataset includes restatements from 2000 through 2007.) and press release when data are not available in the Audit Analytic Database. In addition to the companies in the GAO studies, We identified more restatements resulting from fraud by using a fraud identifier in the Audit Analytic data. By including firms that restated their financial statements due to accounting irregularities, the sample we used in this research attempts to capture restatements resulting from purposeful, aggressive accounting choices[6].

We then matched these companies to Compustat ExecuComp database (ExecuComp has executive compensation information) and Compustat Company Annual Report database (Annual Report has company information). For the purpose of our study, we keep companies that can be found in both databases from year 1997 through 2007.
Table 4.1: Frequency distribution of restatement announcement by year

<table>
<thead>
<tr>
<th>Announcement Year</th>
<th>Number of Announcements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>8</td>
</tr>
<tr>
<td>1998</td>
<td>7</td>
</tr>
<tr>
<td>1999</td>
<td>33</td>
</tr>
<tr>
<td>2000</td>
<td>47</td>
</tr>
<tr>
<td>2001</td>
<td>72</td>
</tr>
<tr>
<td>2002</td>
<td>89</td>
</tr>
<tr>
<td>2003</td>
<td>94</td>
</tr>
<tr>
<td>2004</td>
<td>105</td>
</tr>
<tr>
<td>2005</td>
<td>201</td>
</tr>
<tr>
<td>2006</td>
<td>101</td>
</tr>
<tr>
<td>2007</td>
<td>24</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>781</strong></td>
</tr>
</tbody>
</table>

**Data Structure and the Sample**

The GAO database includes restating company names, ticker symbols, announcement dates, and entity that prompted the restatement and reasons for restatements. Following Burns and Kedia’s[6] approach, we excluded all the financial companies with Standard Industry Classification Codes (SIC) 6000 - 6999.¹ because their corporate governance is substantially different from those of other companies due to regulatory oversight[14]. We initially began with a sample of restating companies that consists of 458 companies with 781 restatements throughout the study period. However the number of companies used in each specific test may vary depending on data availability because the analysis was done mainly in SAS 9.2 and Matlab, which automatically skip over records when missing values are present.

Note that a lot of recent empirical studies on financial restatements include only the years before 2002[6][7][14] but our sample extends the study period to 2006. Table 4.1 shows the distribution of restatement announcement across years from which we can see that, number of restatements increases rapidly since year 1999 and this increasing trend did not stop until year 2005. For example, the number of restatements announced in 2005 was 201, almost twice as many as that in 2004. The number of companies announced restatement is listed in

¹Standard Industry Classification Codes are four-digit codes assigned by the U.S. government to business establishments to identify the primary business of the establishment.
Table 4.2: Distribution of restating firms by year

<table>
<thead>
<tr>
<th>Announcement Year</th>
<th>Number of Restating-Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>8</td>
</tr>
<tr>
<td>1998</td>
<td>7</td>
</tr>
<tr>
<td>1999</td>
<td>32</td>
</tr>
<tr>
<td>2000</td>
<td>43</td>
</tr>
<tr>
<td>2001</td>
<td>72</td>
</tr>
<tr>
<td>2002</td>
<td>77</td>
</tr>
<tr>
<td>2003</td>
<td>87</td>
</tr>
<tr>
<td>2004</td>
<td>94</td>
</tr>
<tr>
<td>2005</td>
<td>176</td>
</tr>
<tr>
<td>2006</td>
<td>94</td>
</tr>
<tr>
<td>2007</td>
<td>18</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>715</strong></td>
</tr>
</tbody>
</table>

Table 4.3: Frequency of restatement announcements

<table>
<thead>
<tr>
<th>Announcement Number</th>
<th>Frequency</th>
<th>Percent of sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>269</td>
<td>58.6</td>
</tr>
<tr>
<td>2</td>
<td>110</td>
<td>23.9</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>4.13</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>1.74</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>1.09</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.22</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Table 4.2, which also peaked in 2005. Both total numbers in Tables 4.1 and 4.2 are greater than the number of restating companies (458) since some companies restated more than once during the study period. Table 4.3 presents the distribution of the number of restatements during the study period: about 58.6% of the restating companies restated only once while nearly 38% of them restated between 2 to 4 times, and about 3% restated more than 4 times.

Of all the sample of 781 restatements we collected, 17.26% did not report any reason for restatement; about 63% of the restatements were due to single accounting issue and about 19.79% restatements were made for multiple reasons. The two largest categories of
restatements are revenue recognition and expense and cost. The distribution of reasons for restatements is tabulated in Table 4.4. Table 4.5 presents distribution of restating companies across different industries using the 2-digit SIC code, which shows that the sampled restating companies are from a broad range of industries. Among all industries classified by the two-digit SIC code, Business Services, Electric Power and Computer Equipment are the three categories with the largest numbers of restating companies.

As the comparison group, non-restating companies are defined to be all companies that have data available in the Compustat Execucomp database and did not restate between 1992 and 2006. Initially we had 1901 non-restating companies.

**Data Cleanup Procedure**

Before moving on to the next section, let us briefly discuss the data cleanup procedure we adopted.

First, in order to control for industry and time effect, we excluded industries in which there were no restating companies as well as companies that had only data before 1997. We also excluded companies that had fewer than 5 records from the sample since highly unbalanced data may cause model fitting problems. After these two steps, there are 1718 non-restating companies and 451 restating companies left in the sample.

Second, in the process of data cleaning, we encountered problems when trying to determine the CEOs in certain firm-specific fiscal years. One problem is that in some fiscal years CEOs who left the position and their replacements were all documented in the dataset. Therefore, there exist multiple records for the same company in one year. Another problem is

<table>
<thead>
<tr>
<th>Reason</th>
<th>Number of Restatements</th>
<th>Percent of Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expense and cost</td>
<td>277</td>
<td>35.15</td>
</tr>
<tr>
<td>Revenue Recognition</td>
<td>237</td>
<td>30.08</td>
</tr>
<tr>
<td>Restructuring, assets or inventory</td>
<td>140</td>
<td>17.77</td>
</tr>
<tr>
<td>Merger</td>
<td>46</td>
<td>5.84</td>
</tr>
<tr>
<td>Reclassification</td>
<td>76</td>
<td>9.64</td>
</tr>
<tr>
<td>Security Related</td>
<td>65</td>
<td>8.25</td>
</tr>
<tr>
<td>Other</td>
<td>136</td>
<td>17.26</td>
</tr>
</tbody>
</table>
that although there is only one record in a particular fiscal year, an old CEO of a company left the position and next year an interim CEO (might be someone who worked for the company in the past) stepped in. In either case, CEOs were not fully compensated because they did not work the full year. Moreover, new CEOs and those who left usually get signing bonus or severance payment, respectively. These may cause problems since they might bring noise to the data. We marked these years where there are two or more CEOs in the company “transition years”.

One possible approach to deal with these “transition years” is to remove all the observations in these years, which include transition years as well as first and last years of each CEO. However, as appealing as it may seem, this approach removes a large amount of data points and will cause the dataset to be more unbalanced as well as causing unnecessary missingness. A second possible approach would be to replace values in the transition years with weighted average of previous fiscal year and the next fiscal year. In the simplest example, if two CEOs served a company in the same fiscal year, the old CEO was in the company for 4 months and the new one worked 8 months, the weighted average will be: \( \frac{4}{12} \times \) old CEO’s amount of last fiscal year + \( \frac{8}{12} \times \) new CEO’s amount of next year. Another more complicated example is: one old CEO and one interim CEO worked in the same fiscal year, which means, not only this fiscal year but also the next one are transition years. If we were to adopt the first approach, we would have to lose a fair amount of information. In this case, we can replace first transition year with numbers from previous fiscal year and the next one with numbers from the year after it. The last approach, which is also the one adopted by most researchers, adopts the built-in identifier *Annual Title* in execucomp to identify CEOs. Execucomp identifies the one person who worked longer in the year to be the CEO of that fiscal year. And we should note that there are also drawbacks in this identifier: missing values and incorrect values exist. There are no perfect solution to the problem we described above. It’s not possible to get accurate data. In the next section, we are going to present some descriptive results using the third approach.
## Table 4.5: Distribution of restating firms across industry

<table>
<thead>
<tr>
<th>Industry Description</th>
<th>SIC code</th>
<th>Number</th>
<th>Industry Description</th>
<th>SIC code</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural Production-crops</td>
<td>1</td>
<td>1</td>
<td>Railroad Transportation</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>Metal Mining</td>
<td>10</td>
<td>5</td>
<td>Transit and Passenger Transit</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td>Coal Mining</td>
<td>12</td>
<td>1</td>
<td>Motor Freight</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td>Oil and Gas Extraction</td>
<td>13</td>
<td>18</td>
<td>Air Transportation</td>
<td>45</td>
<td>6</td>
</tr>
<tr>
<td>Mining, Quarrying Of</td>
<td>14</td>
<td>1</td>
<td>Transportation Services</td>
<td>47</td>
<td>3</td>
</tr>
<tr>
<td>Nonmetallic Minerals, Except Fuels</td>
<td></td>
<td>48</td>
<td>Electric and Gas</td>
<td>48</td>
<td>15</td>
</tr>
<tr>
<td>Operative Builders</td>
<td>15</td>
<td>1</td>
<td>Electric Power</td>
<td>49</td>
<td>38</td>
</tr>
<tr>
<td>Heavy Construction</td>
<td>16</td>
<td>1</td>
<td>Nondurable Goods</td>
<td>51</td>
<td>6</td>
</tr>
<tr>
<td>Food Products</td>
<td>20</td>
<td>13</td>
<td>Building Materials, Hardware, Garden Supply</td>
<td>52</td>
<td>2</td>
</tr>
<tr>
<td>Apparel</td>
<td>23</td>
<td>5</td>
<td>General Merchandise</td>
<td>53</td>
<td>11</td>
</tr>
<tr>
<td>Lumber, Wood Products, Except Furniture</td>
<td>24</td>
<td>1</td>
<td>Automotive Dealers</td>
<td>55</td>
<td>4</td>
</tr>
<tr>
<td>Furniture Fixtures</td>
<td>25</td>
<td>3</td>
<td>Accessory Stores</td>
<td>56</td>
<td>22</td>
</tr>
<tr>
<td>Paper Products</td>
<td>26</td>
<td>9</td>
<td>Home Furniture</td>
<td>57</td>
<td>6</td>
</tr>
<tr>
<td>Printing</td>
<td>27</td>
<td>5</td>
<td>Eating and Drinking Places</td>
<td>58</td>
<td>19</td>
</tr>
<tr>
<td>Chemicals</td>
<td>28</td>
<td>26</td>
<td>Retail Stores</td>
<td>59</td>
<td>11</td>
</tr>
<tr>
<td>Petroleum Refining and Related Industries</td>
<td>29</td>
<td>2</td>
<td>Personal Services</td>
<td>72</td>
<td>5</td>
</tr>
<tr>
<td>Rubber and Plastics</td>
<td>30</td>
<td>3</td>
<td>Business Services</td>
<td>73</td>
<td>58</td>
</tr>
<tr>
<td>Leather And Leather Products</td>
<td>31</td>
<td>1</td>
<td>Motion pictures</td>
<td>78</td>
<td>2</td>
</tr>
<tr>
<td>Primary Metals</td>
<td>33</td>
<td>6</td>
<td>Amusement And Recreation Services</td>
<td>79</td>
<td>1</td>
</tr>
<tr>
<td>Fabric Metal</td>
<td>34</td>
<td>8</td>
<td>Health Services</td>
<td>80</td>
<td>6</td>
</tr>
<tr>
<td>Computer Equipment</td>
<td>35</td>
<td>35</td>
<td>Educational Services</td>
<td>82</td>
<td>2</td>
</tr>
<tr>
<td>Electrical Equipment</td>
<td>36</td>
<td>31</td>
<td>Engineering, Accounting, Research, Management</td>
<td>87</td>
<td>8</td>
</tr>
<tr>
<td>Automotive</td>
<td>37</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurements and instruments</td>
<td>38</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miscellaneous Manufacturing Industries</td>
<td>39</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonclassifiable Establishments</td>
<td>99</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.2 Analysis

4.2.1 Preliminary Tests

The primary interest in this section is to test whether there is a difference in CEO’s compensation between restating companies and non-restating companies, as well as to model the time dependency and restatement effect on CEOs’ compensation, especially performance based compensation. In this section, we are going to show some preliminary test results that will help us formulate reasonable marginal models and covariance structures for our data.

In order to compare restating and non-restating firms with different sizes and in different industries, we matched each restating firms with two control firms selected from non-restating companies using two-digit SIC code and firm size. The matched companies should be in the same two-digit SIC industry with comparable firm size\(^2\). To avoid having different control firms for the same restating company in different fiscal years, we took the average of total asset of each firm and take log of the average total asset instead of comparing firm size every year. Furthermore, we followed Efendi’s approach\(^14\) that allows the difference in firm size between the two companies to range between \(-30\%\) and \(30\%\). Therefore, a company that has firm size within \(30\%\) of the restating firm in comparison and belong in the same industry is considered as a potential match for the restating firm. Based on this matching criteria, we do not have enough non-restating companies in the control group in some industries. Among these 451 restating companies, one company does not have any match. The restating companies are first sorted by potential number of matches. Then two matching companies are selected at random sequentially as follows: companies that have fewer matches are matched with control firms first. After the matching process, 8 restating companies have only one match while another 25 are left out of the sample due to lack of data. In summary, we have 417 restating firms and 842 matched control firms in this sample. This case-control sample is called sample2. By using a “case-control” sample, we are able to evaluate confounding and interaction effects more precisely. Moreover, it makes the two comparison groups more comparable than before as we are controlling for industry and firm size. However, it is possible that there is potential bias in this sample when we start selecting subsamples from the whole sample. And sample size of sample2 is much smaller than sample1, which means there is

\(^2\)Company revenue, sales and natural logarithm of total assets are the most common proxies for firm size in literature\(^8\)\[^24\]. In our study, we used log(total assets) as a measure of firm size.
potential information loss during the matching process. Therefore, sample\textsuperscript{2} will be used for descriptive study and exploratory tests only.

We will now present results of some exploratory analyses. The mean and median for the major components of CEO compensation as well as their matched control companies are presented in Table 4.6. For descriptive purpose, we use matched samples because the comparison will be more straightforward and easier to interpret. We report dollar amounts for misreported years and non-misreported years separately.

The average salary paid by non-restating firms during misreported years is 590,460 dollars, and we can see from Table 4.6 that the average amount paid during normal years do not differ much. Meanwhile, non-restating firms seem to pay a bit higher salary than restating firms, with an average of 669,330 dollars in misreported period and 630,670 dollars in normal firm years (medians are also very similar). Amount of compensation in the form of bonus is similar between restating firms and their matched controls. We see a huge jump in average restricted stock grants from 456,580 for regular years to 1,044,080 for misreported years of restating firms. However, all the medians of restricted stock grants of both restating and non-restating firms are zeros, which suggest that companies do not grant restricted stock in many firm years and there may be some outliers in the data. Both the mean and median of option grant show that restating companies grant more options in misreported years, and they also grant more options than their matches. In terms of total compensation, the amounts paid in misreported years were also larger than those in normal years for both restating and non-restating firms. In comparison to non-restating companies, the ones that restated have much higher total compensation mainly because they grant more options. This may lead people to think that more options may cause more frauds. However, if we compare the percentage of options in total compensation, we see that there is not much of a difference between the two. On average, 34\% and 37\% of CEOs’ pay in restating firms comes from options in normal years and misreported years, respectively. For non-restating firms, the percentages are 31\% in normal years and 38\% in misreported years. The average percentage of CEO compensation paid in terms of cash (salary+bonus) does not differ much between restating and non-restating firms either.

Among many empirical studies in finance, one-sample t-test is the most common one used to examine the difference. This simple yet effective approach is valid only when all the observations are independent. However, correlations among subjects are commonly present.
in longitudinal data. We always need to take correlations and time effect into account when analyzing longitudinal data. Here we examine several variables of interest to see the general trend in time. We can see from Figure 4.1a that percentage of total compensation paid in options (per_option) was increasing until year 2001; after 2001, it went downward. And before year 2001, restating firms had higher average option to total compensation ratios than matched non-restating firms; after 2001 the average of restating firms option level dropped below control companies (Restating firms and all non-restating firms yield the same trend). It seems that there is a quadratic time effect on per_option. Figure 4.1b shows the general trend of percentage of total compensation paid in cash (per_cash). Not surprisingly, we see a decreasing trend in time before 2001 and a slight increasing trend between 2001 and 2005. The level of per_cash again drops after 2005. It clearly shows a nonlinear time trend. Option grant and cash compensation are the two main components of compensation. And we see from the figures that on average trend of option grants and compensation in cash move in the same direction for both types of firms.

![Options as a Percentage of Total Compensation](image1)

From Figure 4.1, we see that the two plots basically show similar trend: level of in-the-money unexercised exercisable options drop sharply in 2002 and it climbs up gradually after that. However, before 2002, average of restating firms in-the-money options changes dramatically, which may be due to outliers.

We took a random sample of 20 companies from the sample, 2 restating companies from
Figure 4.1: Estimated Value of in-the-money unexercised exercisable options

each industry group defined by Fama and French\(^3\). And the other 12 are from these restating firms’ matches. Figure 4.2 shows empirical growth plot of percentage of option grant in total compensation versus year. Due to data limitation, some companies have fewer observations than others and they have different starting points (We should note that there are certain missing values in the dataset). But the plot gives us some sense of individual change over time. From the plot, we see that for some companies, the trajectories change more smoothly (id 006376, 027957) and for some other companies, the trajectories change more dramatically. Furthermore, it is clear that each firm has its own pattern of change over time, which suggest there is \textit{between-subject variation} in the sample. Figure 4.2b plots the individual change of proportion of cash compensation over time for the same companies in Figure 4.2a. It is not surprising that the change of \textit{per\_cash} is opposite of \textit{per\_option}.

Table 4.6: Components of CEO compensations in misreported years and normal years

<table>
<thead>
<tr>
<th>Normal</th>
<th>Salary</th>
<th>Bonus</th>
<th>Restricted Stock</th>
<th>Option Grant</th>
<th>Total</th>
<th>Value Realized on Option</th>
<th>In-the-Money Estimate</th>
<th>Option-to-Total</th>
<th>Cash-to-Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>restating(1)</td>
<td>630.67</td>
<td>610.52</td>
<td>456.58</td>
<td>2840.09</td>
<td>5056.19</td>
<td>2032.47</td>
<td>10427.73</td>
<td>0.34</td>
<td>0.49</td>
</tr>
<tr>
<td>restating(2)</td>
<td>572</td>
<td>285</td>
<td>0</td>
<td>557.4</td>
<td>2207.49</td>
<td>0</td>
<td>1164.49</td>
<td>0.31</td>
<td>0.45</td>
</tr>
<tr>
<td>control(1)</td>
<td>565.67</td>
<td>536.67</td>
<td>301.61</td>
<td>1808.80</td>
<td>3696.79</td>
<td>1797.67</td>
<td>6724.78</td>
<td>0.31</td>
<td>0.54</td>
</tr>
<tr>
<td>control(2)</td>
<td>513.00</td>
<td>247.80</td>
<td>0</td>
<td>410.25</td>
<td>1747.97</td>
<td>0</td>
<td>840.84</td>
<td>0.27</td>
<td>0.52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Restated</th>
<th>Salary</th>
<th>Bonus</th>
<th>Restricted Stock</th>
<th>Option Grant</th>
<th>Total</th>
<th>Value Realized on Option</th>
<th>In-the-Money Estimate</th>
<th>Option-to-Total</th>
<th>Cash-to-Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>restating(1)</td>
<td>669.33</td>
<td>646.30</td>
<td>1044.08</td>
<td>3517.77</td>
<td>6317.01</td>
<td>2640.91</td>
<td>11834.17</td>
<td>0.37</td>
<td>0.49</td>
</tr>
<tr>
<td>restating(2)</td>
<td>633.1</td>
<td>337.56</td>
<td>0</td>
<td>855.94</td>
<td>2590.72</td>
<td>0</td>
<td>1325.86</td>
<td>0.36</td>
<td>0.43</td>
</tr>
<tr>
<td>control(1)</td>
<td>590.46</td>
<td>576.26</td>
<td>479.45</td>
<td>2331.91</td>
<td>4338.16</td>
<td>1985.74</td>
<td>8711.09</td>
<td>0.38</td>
<td>0.49</td>
</tr>
<tr>
<td>control(2)</td>
<td>540.00</td>
<td>294.5</td>
<td>0</td>
<td>751.78</td>
<td>2310.55</td>
<td>0</td>
<td>1216.93</td>
<td>0.38</td>
<td>0.45</td>
</tr>
</tbody>
</table>

a(1) is Mean.
b(2) is Median.
cAll numbers are in thousand dollars.
dSalary: amount of CEOs' base salary.
eBonus: amount of CEOs' bonus.
fRestricted Stock: value of restricted stock granted to CEOs during the fiscal year.
gOption Grant: total Black-Scholes value of stock options granted to CEOs during the fiscal year.
hTotal: CEOs’ total compensation.
iValue realized on option: value obtained from exercising options.
jIn the money estimate: estimated value CEOs would have realized from exercising all vested options and unvested options.
kOption-to-Total: Ratio of Black-Scholes value of stock options to CEOs’ total compensation.
lCash-to-Total: Ratio of bonus plus base salary to CEOs’ total compensation.
The firm characteristics of restating firms and matched control firms are also compared, the descriptive results are presented in Table 4.7. We calculate means and medians for these two types of firms in misreported years and non-misreported years separately. Note that in the table we list actual value of total assets but not logarithm of it, but we use log(assets) as the proxy of firm size. Both restating and their matches have higher sales in the violation years. And on average, restating firms have larger sales than their matches and the median of sales for restating firms in misreported years is even larger than the median of the controls. The medians of total dividends paid by both types of firms are similar, however, on average restating companies pay more dividends than control companies. This is evidence that supports our hypothesis that restating firms use dividend as one way to signal to the market they are doing well while in fact they are not.

We also examined market value and book value at fiscal year end. Market value is number of shares outstanding times price at fiscal year end. Market capitalization of the firm is often used as another proxy of firm size[6]. In our sample, both market value and book values have similar patterns to sales. It is suggested by other studies that companies that are close to violating debt covenants are more likely to engage in aggressive accounting practices to avoid the penalties [6][30]. They test the hypothesis by examining leverage level as a proxy for closeness to debt covenant violations or costs of financial distress. In Burns
and Kedia’s study, mean and median levels of leverage of “restating firm-years” are found to be significantly higher than “nonrestating firm-years”. However, the approach they adopt defines only misreported years to be restating-years; all other years of both restating and non-restating firms are defined as nonrestating years. Therefore, time effect is easily overlooked in this approach. In another approach by Efendi, Debt-to-Asset ratio is examined among matched pairs in all years. No significant difference is found between restating companies and nonrestating companies in their study. We did not find that restating firms have higher leverage level, on the contrary, we find that in both restated and normal firm-years, mean and median of leverage levels in restating firm are lower.

The market-to-book (MTB) ratio is a measure used to compare a company’s book value to its market value. Book value is company’s total tangible assets less its liabilities. A higher market-to-book ratio means investors expect the firm to generate more value from a given set of assets with all else equal, because investors are paying more than what they actually worth. In some studies, market-to-book ratio is used as proxy for growth opportunity. Researchers suggest that high-growth firms facing reduced growth opportunities are likely to engage in aggressive accounting practices. By comparing means and medians of market-to-book ratios in both types of firms, we see that on average, they are very close. and even though the mean of misreported years for restating firm is a bit higher (2.17) than the one of normal years (2.12), the medians are the other way around (1.5 and 1.53, respectively). The control group also has higher market-to-book ratio than restating firms. Figure 4.3 shows the mean trajectory of market-to-book ratios over the studied period. Both restating and non-restating companies have fast growing MTBs until 2000. The averages of the two groups are very close throughout the entire period except for years between 1997 and 1999, 2002 and 2003. In these periods, the restating companies have mean higher MTBs than their matched controls.

While the market-to-book ratio tells us what investors expect, earnings-per-share (EPS) is an important metric to determine a company’s profitability. The higher EPS is with all else equal, the higher each share should worth. We can see from Table 4.7 that in misreported years and normal years, control group has higher EPS than restating firms.
Figure 4.3: Market to Book Ratio by year
Table 4.7: Descriptive Summary of Firm Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Assets</th>
<th>Sale</th>
<th>Total Dividends</th>
<th>Equity Book Value</th>
<th>Equity Market Value</th>
<th>Market-to-Book Ratio</th>
<th>Leverage</th>
<th>Earning Per Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>restating(1)</td>
<td>5152.66</td>
<td>4089.72</td>
<td>73.57</td>
<td>1534.59</td>
<td>5302.94</td>
<td>2.12</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>restating(2)</td>
<td>895.62</td>
<td>970.44</td>
<td>2.14</td>
<td>371.11</td>
<td>1009.04</td>
<td>1.53</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>control(1)</td>
<td>3307.95</td>
<td>2636.65</td>
<td>48.79</td>
<td>1354.80</td>
<td>3384.87</td>
<td>2.12</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>control(2)</td>
<td>778.81</td>
<td>807.87</td>
<td>1.67</td>
<td>326.68</td>
<td>845.81</td>
<td>1.57</td>
<td>0.22</td>
</tr>
<tr>
<td>Restated</td>
<td>restating(1)</td>
<td>9308.79</td>
<td>5961.73</td>
<td>95.42</td>
<td>2546.95</td>
<td>6988.49</td>
<td>2.17</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>restating(2)</td>
<td>1218.45</td>
<td>1304.48</td>
<td>0</td>
<td>540.99</td>
<td>1186.05</td>
<td>1.50</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>control(1)</td>
<td>4234.04</td>
<td>3382.70</td>
<td>46.15</td>
<td>1734.40</td>
<td>4203.35</td>
<td>2.25</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>control(2)</td>
<td>907.47</td>
<td>966.65</td>
<td>0</td>
<td>389.69</td>
<td>1031.31</td>
<td>1.59</td>
<td>0.21</td>
</tr>
</tbody>
</table>

*a(1) is Mean.  
*b(2) is Median.  
*cAll numbers are in thousand dollars.  
*dAssets: total assets at fiscal year end.  
*eSale: net sales at fiscal year end.  
*fTotal dividends: total dividends paid.  
*gEquity Book Value: book value of company’s equity at fiscal year end.  
*hEquity Market Value: market value of company’s equity at fiscal year end.  
*iMarket-to-Book Ratio: (Total Asset- Book Value of Equity + Market Value of Equity)/Total Asset  
*jLeverage: (Long Term Debt + Short Term Debt)/Total Assets  
*kEarning per Share: diluted earning per share before extraordinary items is calculated as follows: (Net Income - Preferred Stock Dividends)/(Weighted Average of Common Stock Shares Outstanding )
4.2.2 Two-Part Models

One important thing to remember when we conduct longitudinal analysis is to take time
effect and interindividual differences into account. A lot of research done using longitudinal
data overlooked time effect and dependency, which may lead to unreliable results. As we have
seen from the exploratory tests above that not all companies response variables (those we are
interested in) change in the same direction or at the same rate. Therefore, the trajectories
of the the sample mean only display average change of different groups, not anything about
shapes of individual trajectories. The analyses in the previous section are helpful in the
sense that we now know how the average of CEO compensation components go. However,
those techniques suffer from several drawbacks: first, we are not using the whole sample but
matched sample. In the process of pairing up the firms, we lose a certain number of firms
as there are not matches for them. In other words, we do not use all of the relevant data
and information available. Second, simple summaries cause people to overlook the richness
inherent in longitudinal data. It is suggested that we should not omit any subject when it
is not really necessary. Therefore, in this section, we are going to analyze the entire sample
we have.

While researchers traditionally make assumptions of independence when performing
linear regression analysis, mixed effect models are more appropriate for presenting correlated/
cluster collected data because we bring in random effects into the model and therefore
allow heterogeneity among individuals and correlation within each subject. In our context,
companies are subjects. However, the main goal in most statistical analyses is to use a
flexible model that represents the data as adequately as possible. One major assumption
that underlies most models for continuous responses is the normality assumption. The reason
is the unrivaled mathematical properties of the normal distribution. But, as we mentioned
before, there are cases where making this assumption will not lead to satisfactory results.
One example will be executive compensation data we present.

Logistic-Lognormal Mixed Model

For the executive compensation data, we first plotted the histogram of option grant and
proportion of option in total compensation. From Figure 4.4 we see that after log-
transformation, the continuous part of option grant has a nice symmetric bell-shaped
distribution. And Figure 4.5 shows distribution of proportion of option in total compensation after arcsine transformation. Note that in Figure 4.5 we purposely exclude all zeros values to get a better vision of continuous values (If zeros included, the distribution looks rather flat). The second response variable does not have a perfectly symmetric distribution and both transformed responses have positive values. Unlike many other studies that adopt mixed distribution or similar models, in this study, all the zeros are true values that indicate there is no option granted in a certain firm-year. While about 25% of our samples are restating companies, zeros and continuous values in our sample also present similar distributions between the two: 24.07% of zeros come from restating firms and 75.93% from non-restating firms.

![Figure 4.4: Distribution of log-transformed option](image)

An explanatory variable in the mixed-distribution two-part model can be related to the probability of a non-zero value, magnitude of non-zero values or both. As a result of allowing for possibly same or different sets of covariates in the two parts of the model, it is more likely that the mixed-distribution two part model will provide us with more accurate results. We start by including the same list of independent variables that may affect option grants in both parts. And we will examine significance levels of each covariate. We first applied
the logistic-lognormal model proposed by Tooze[35] to our data to model the probability of companies granting options to their CEOs and the amount of options granted. Results are reported in Table 4.8. Variables we examined in exploratory tests are included in the model. The initial model is as follows:

\[ R_{ij} = \begin{cases} 0 & \text{if } Option_{ij} = 0 \\ 1 & \text{if } Option_{ij} > 0 \end{cases} \] (4.1)

where \( Option_{ij} \) is the Black-Scholes value of option grants of the \( i^{th} \) company in \( j^{th} \) year. Recall that

\[ P(R_{ij} = r_{ij} | \theta_1) = \begin{cases} 1 - p_{ij}(\theta_1), & \text{if } r_{ij} = 0 \\ p_{ij}(\theta_1), & \text{if } r_{ij} = 1 \end{cases} \] 

The first part of our model is:

\[
\text{logit}(p_{it}) = \beta_{10} + \beta_{1_{res}} \cdot \text{Restate} + \beta_{1_{year}} \cdot \text{Year} + \beta_{1_{res,year}} \cdot \text{Year} \cdot \text{Restate} + \\
\beta_{1_{leve}} \cdot \text{Leverage} + \beta_{1_{MTB}} \cdot \text{MTB} + \beta_{1_{sale}} \cdot \text{Sale} + \beta_{1_{SEQ}} \cdot \text{SEQ} + \beta_{1_{EP}} \cdot \text{EP} + \\
\beta_{1_{div}} \cdot \text{Dividend} + \beta_{1_{age}} \cdot \text{Age} + \beta_{1_{ind}} \cdot \text{Industry} + u_{1i} \] (4.2)

We assume that the second part follows a log-normal distribution and intercepts are the only
random effects in both parts:

\[
\log(\text{option}_{it}) = \beta_{20} + \beta_{2,\text{res}} \times \text{Restate} + \beta_{2,\text{year}} \times \text{Year} + \beta_{2,\text{res_year}} \times \text{Year_Restate} \\
+ \beta_{2,\text{lev}} \times \text{Leverage} + \beta_{2,\text{mtb}} \times \text{MTB} + \beta_{2,\text{sale}} \times \text{Sale} + \beta_{2,\text{SEQ}} \times \text{SEQ} \\
+ \beta_{2,\text{EP}} \times \text{EP} + \beta_{2,\text{div}} \times \text{Dividend} + \beta_{2,\text{age}} \times \text{Age} + \beta_{2,\text{ind}} \times \text{Industry} + u_{2i} + \epsilon_i 
\] (4.3)

where “Restate” is the restatement dummy variable that indicates if the company falls in the restating company category; year*restate is the interaction term between year and restatement indicator; variables such as leverage, MTB (Market-to-Book ratio), SEQ (Shareholder’s Equity), Dividend yield, industry category and sale are company characteristics that may impact CEO’s option compensation; and the variable age is CEO’s age. Both models with and without correlated random effects were fitted using the SAS macro MIXCORR developed by Tooze[35]. Parameter estimates are shown in Table 4.8. The AIC and likelihood ratio of the correlated random effects model are slightly larger than the model with uncorrelated random effects is more appropriate. Despite the difference, most parameter estimates and associated significance levels in both models are very similar. In the next step, we removed covariates that were not significant in both parts of the model backwards in order to obtain a more parsimonious model. After variables “leverage”, “Market-to-Book Ratio”, “earnings per share” and “dividend” were removed from both parts and SEQ from the logistic part, the correlation between random effects became insignificant (p-value = 0.3277). In other words, there is no longer any statistically significant evidence that shows correlation between random effects in the two parts of this model. Based on AIC and likelihood ratio test, the model with correlated random effects is better than the one with uncorrelated random effects.
Table 4.8: Parameter Estimates of the Full Model: Correlated Random Intercept

<table>
<thead>
<tr>
<th></th>
<th>Uncorrelated</th>
<th></th>
<th>Correlated</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>p-value</td>
<td>Estimate</td>
<td>p-value</td>
</tr>
<tr>
<td>Logistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.3259</td>
<td>0.0000</td>
<td>2.3424</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>restate-dummy</td>
<td>0.3994</td>
<td>0.0152</td>
<td>0.4070</td>
<td>0.0139</td>
</tr>
<tr>
<td>time</td>
<td>0.0176</td>
<td>0.0633</td>
<td>0.0111</td>
<td>0.2425</td>
</tr>
<tr>
<td>restate*time</td>
<td>-0.0564</td>
<td>0.0026</td>
<td>-0.0521</td>
<td>0.0055</td>
</tr>
<tr>
<td>industry1</td>
<td>0.0929</td>
<td>0.5520</td>
<td>0.1019</td>
<td>0.5171</td>
</tr>
<tr>
<td>industry2</td>
<td>0.3844</td>
<td>0.0113</td>
<td>0.3927</td>
<td>0.0101</td>
</tr>
<tr>
<td>industry3</td>
<td>0.3661</td>
<td>0.0195</td>
<td>0.3649</td>
<td>0.0207</td>
</tr>
<tr>
<td>industry4</td>
<td>0.7945</td>
<td>0.0001</td>
<td>0.8064</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>leverage</td>
<td>-0.0639</td>
<td>0.5657</td>
<td>-0.0691</td>
<td>0.5309</td>
</tr>
<tr>
<td>MTB</td>
<td>-0.0075</td>
<td>0.4257</td>
<td>-0.0037</td>
<td>0.7007</td>
</tr>
<tr>
<td>SALE</td>
<td>0.0000</td>
<td>0.0009</td>
<td>0.0000</td>
<td>0.0024</td>
</tr>
<tr>
<td>SEQ</td>
<td>0.0000</td>
<td>0.3750</td>
<td>0.0000</td>
<td>0.5466</td>
</tr>
<tr>
<td>EPSFX</td>
<td>-0.0015</td>
<td>0.1707</td>
<td>-0.0014</td>
<td>0.1967</td>
</tr>
<tr>
<td>DVT</td>
<td>-0.0003</td>
<td>0.1143</td>
<td>-0.0002</td>
<td>0.1682</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.0257</td>
<td>0.0000</td>
<td>-0.0255</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Var(Rndm Effect)</td>
<td>2.6439</td>
<td>0.0000</td>
<td>2.7042</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Lognormal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>6.6222</td>
<td>0.0000</td>
<td>6.6401</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>restate-dummy</td>
<td>0.2955</td>
<td>0.0001</td>
<td>0.2897</td>
<td>0.0004</td>
</tr>
<tr>
<td>time</td>
<td>0.0773</td>
<td>0.0000</td>
<td>0.0795</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>restate*time</td>
<td>-0.0297</td>
<td>0.0001</td>
<td>-0.0331</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>industry1</td>
<td>-0.1719</td>
<td>0.0460</td>
<td>-0.1802</td>
<td>0.0623</td>
</tr>
<tr>
<td>industry2</td>
<td>-0.3688</td>
<td>0.0000</td>
<td>-0.3930</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>industry3</td>
<td>0.3422</td>
<td>0.0001</td>
<td>0.3882</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>industry4</td>
<td>0.2913</td>
<td>0.0053</td>
<td>0.2856</td>
<td>0.0149</td>
</tr>
<tr>
<td>leverage</td>
<td>-0.0004</td>
<td>0.9925</td>
<td>-0.0512</td>
<td>0.2051</td>
</tr>
<tr>
<td>MTB</td>
<td>0.0346</td>
<td>0.0000</td>
<td>0.0320</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>SALE</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>SEQ</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0001</td>
</tr>
<tr>
<td>EPSFX</td>
<td>0.0008</td>
<td>0.3539</td>
<td>0.0007</td>
<td>0.4053</td>
</tr>
<tr>
<td>DVT</td>
<td>-0.0001</td>
<td>0.0393</td>
<td>-0.0001</td>
<td>0.0505</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.0051</td>
<td>0.0095</td>
<td>-0.0058</td>
<td>0.0039</td>
</tr>
<tr>
<td>Residual</td>
<td>0.7329</td>
<td>0.0000</td>
<td>0.7402</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Var(Rndm Effect)</td>
<td>0.9221</td>
<td>0.0000</td>
<td>1.2185</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Covariance</td>
<td></td>
<td></td>
<td>0.3732</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Uncorrelated</th>
<th></th>
<th>Correlated</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>175836.5</td>
<td></td>
<td>175879.6</td>
<td></td>
</tr>
<tr>
<td>-2 log-likelihood</td>
<td>175770.5</td>
<td></td>
<td>175811.6</td>
<td></td>
</tr>
</tbody>
</table>
This model setup is very similar to 4.2 and 4.3 except that we now have fewer covariates in the model.

\[
\text{logit}(p_{it}) = \beta_{10} + \beta_{1\text{res}} \times \text{Restate} + \beta_{1\text{year}} \times \text{Year} + \beta_{1\text{res,year}} \times \text{Year*Restate} + \\
\beta_{1\text{sale}} \times \log(\text{Sale}) + \beta_{1\text{age}} \times \text{Age} + \beta_{1\text{ind}} \times \text{Industry} + u_{1i} \tag{4.4}
\]

and

\[
\text{log(option}_{it}) = \beta_{20} + \beta_{2\text{res}} \times \text{Restate} + \beta_{2\text{year}} \times \text{Year} + \beta_{2\text{res,year}} \times \\
\text{Year*Restate} + \beta_{2\text{sale}} \times \log(\text{Sale}) + \beta_{2\text{SEQ}} \times \text{SEQ} + \beta_{2\text{age}} \times \text{Age} + \\
\beta_{2\text{ind}} \times \text{Industry} + u_{2i} + \epsilon_{i} \tag{4.5}
\]

The parameter estimates of this model are reported in Table 4.9. The only covariate that is in intensity model but not in the occurrence model is SEQ (Equity for Share Holders), however, after removing covariates that are not significant one by one, SEQ turns out to be insignificant as well. Therefore, all covariates that are significant are included in both occurrence and intensity models. According to this model, for the sample of companies used in this analysis, the restating firms are associated with 1.58 (e^{0.4611} = 1.58) times greater probability of having a non-zero option grant. But this probability decreases over time since interaction between restatement indicator and time is negative and significant. On average, restating companies grant more options than non-restating companies given that option grants were not zero.

Variances of both random effects in occurrence and intensity models are significant, which suggests that there are significant variabilities in the probability of granting options and the number of options granted after adjusting for the factors in the model. Some companies have higher probability of granting options than other companies; some companies grant more options to their CEO than other companies. The correlation between random effects in both parts is significantly positive in the full model, however, it becomes negative and insignificant in the model we discussed. This model suggests that based on the parsimonious model, the probability of granting options is not associated with Black-Scholes value of options granted.
Table 4.9: Parameter Estimates of the Reduced Model: Correlated Random Intercept

<table>
<thead>
<tr>
<th></th>
<th>Uncorrelated</th>
<th></th>
<th>Correlated</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>p-value</td>
<td>Estimate</td>
<td>p-value</td>
</tr>
<tr>
<td>Logistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.4511</td>
<td>0.0000</td>
<td>1.4521</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>restate-dummy</td>
<td>0.4611</td>
<td>0.0035</td>
<td>0.4596</td>
<td>0.0040</td>
</tr>
<tr>
<td>time</td>
<td>0.0010</td>
<td>0.9167</td>
<td>0.0013</td>
<td>0.8901</td>
</tr>
<tr>
<td>restate*time</td>
<td>-0.0634</td>
<td>0.0004</td>
<td>-0.0631</td>
<td>0.0004</td>
</tr>
<tr>
<td>industry1</td>
<td>0.0503</td>
<td>0.7391</td>
<td>0.0650</td>
<td>0.6706</td>
</tr>
<tr>
<td>industry2</td>
<td>0.3929</td>
<td>0.0075</td>
<td>0.3891</td>
<td>0.0089</td>
</tr>
<tr>
<td>industry3</td>
<td>0.5383</td>
<td>0.0004</td>
<td>0.5488</td>
<td>0.0003</td>
</tr>
<tr>
<td>industry4</td>
<td>1.2650</td>
<td>0.0000</td>
<td>1.2614</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>log(SALE)</td>
<td>0.2933</td>
<td>0.0000</td>
<td>0.2935</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.0446</td>
<td>0.0000</td>
<td>-0.0446</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Var(Rndm Effect)</td>
<td>2.5920</td>
<td>0.0000</td>
<td>2.6776</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Lognormal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>4.7118</td>
<td>0.0000</td>
<td>4.6526</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>restate-dummy</td>
<td>0.1967</td>
<td>0.040</td>
<td>0.1785</td>
<td>0.0091</td>
</tr>
<tr>
<td>time</td>
<td>0.0520</td>
<td>0.0000</td>
<td>0.0513</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>restate*time</td>
<td>-0.0253</td>
<td>0.0002</td>
<td>-0.0244</td>
<td>0.0003</td>
</tr>
<tr>
<td>industry1</td>
<td>-0.3368</td>
<td>0.0000</td>
<td>-0.3769</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>industry2</td>
<td>-0.5108</td>
<td>0.0000</td>
<td>-0.5536</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>industry3</td>
<td>0.6975</td>
<td>0.0000</td>
<td>0.6942</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>industry4</td>
<td>0.6901</td>
<td>0.0952</td>
<td>0.6687</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>log(SALE)</td>
<td>0.4060</td>
<td>0.00</td>
<td>0.4141</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>SEQ</td>
<td>0.0000</td>
<td>0.1034</td>
<td>0.0000</td>
<td>0.1816</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.0157</td>
<td>0.0000</td>
<td>-0.0150</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>0.7099</td>
<td>0.0000</td>
<td>0.7080</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Var(Rndm Effect)</td>
<td>0.7754</td>
<td>0.0000</td>
<td>0.7830</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Covariance</td>
<td>-0.0526</td>
<td>0.3277</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>192403.2</td>
<td></td>
<td>192000.1</td>
<td></td>
</tr>
<tr>
<td>-2 log-likelihood</td>
<td>192349.2</td>
<td></td>
<td>191944.1</td>
<td></td>
</tr>
</tbody>
</table>
Last, for the final two-part model with correlated random intercepts, we examined the goodness-of-fit of the second model by checking the normal quantile plots of both random effects and residual plots. First, Figure 4.6, and 4.7 show that the assumption of normality for random effects may need to be modified. The normal quantile-quantile plot of \( u_1 \) (random effect in the logistic/occurrence model) shows a clear departure from a straight line but the histograms of random effect \( u_2 \) (random effect in the intensity model) is approximately symmetric around 0. Second, we examined distribution of subject-specific residuals \( \log(Y_{it}) - X_{2it}\hat{\beta}_1 - \tilde{u}_{2i} \) and population-averaged residuals \( \log(Y_{it}) - X_{2it}\hat{\beta}_1 \). From Figure 4.9 we can see that some outliers are present in the data. The residuals are clustered around zero and there is no certain patterns in the population-averaged residuals. Weiss [39] suggested that the subject-specific residuals are likely to have residual correlations in them as when we estimate them, the estimation process introduces correlation into the residuals. We also plot these two types of residuals against time (not shown), no certain patterns can be seen from the graphs. There are outliers but most profile lines move closely around 0. However, Verbeke [36] argued that subject-specific residuals should not be used to check model assumptions because the random effects highly depend on the normality assumption and are themselves influenced by the assumed covariance structure.

![Figure 4.6: Normal qq Plot: random effects](image)

(a) Occurrence Model  
(b) Intensity Model
Mixed-Distribution Models with Correlated Random Effects and Serial Correlation

In the last section, we have shown some preliminary results that do suggest difference in CEOs’ option grants between restating and non-restating companies. However, as we discussed before, the model applied to our data in the last section does not take serial correlation into account. If serial correlation is present, which is usually the case in longitudinal data, the estimates and their significance levels may be inaccurate. In this section, we will fit the model we introduced in section 2.3, which has random effects with an AR(1) covariance structure.

Similarly as before, we first start from a “full model” and then try to reduce the dimension
by removing insignificant covariates from the model. Note that the number of observations used in each model we fit varies because different variables have different amount of missing values.

The model setup is as follows:

$$R_{it} = \begin{cases} 
0 & \text{if } \text{Option}_{it} = 0 \\
1 & \text{if } \text{Option}_{it} > 0 
\end{cases} \quad (4.6)$$

where Option$_{it}$ is the Black-Scholes value of option grants the CEO received in the $i^{th}$ company in $j^{th}$ year and we have:

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

The first part of the full model can be written in the following form:

$$\text{logit}(p_{it}) = \beta_{10} + \beta_{1,\text{res}} \ast \text{Restate} + \beta_{1,\text{year}} \ast \text{Year} + \beta_{1,\text{res,year}} \ast \text{Year} \ast \text{Restate} + \beta_{1,\text{leverage}} \ast \text{Leverage} + \beta_{1,\text{MTB}} \ast \text{MTB} + \beta_{1,\text{sale}} \ast \text{Sale} + \beta_{1,\text{SEQ}} \ast \text{SEQ} + \beta_{1,\text{div}} \ast \text{Dividend} + \beta_{1,\text{age}} \ast \text{Age} + \beta_{1,\text{ind}} \ast \text{Industry} + b_{it} \quad (4.7)$$
\[
\log(\text{option}_{it}) = \beta_{20} + \beta_{2,\text{rest}} \times \text{Restate} + \beta_{2,\text{year}} \times \text{Year} + \beta_{2,\text{rest},\text{year}} \times \text{Year} \times \text{Restate} + \beta_{2,\text{lev}} \times \text{Leverage} + \beta_{2,\text{mtb}} \times \text{MTB} + \beta_{2,\text{sale}} \times \text{Sale} + \beta_{2,\text{SEQ}} \times \text{SEQ} + \beta_{2,\text{div}} \times \text{Dividend} + \beta_{2,\text{age}} \times \text{Age} + \beta_{2,\text{ind}} \times \text{Industry} + b_{it} + \epsilon_i
\] (4.8)

where the variables are defined as same as before: “Restate” is the restatement dummy variable that indicates if the company falls in the restating company category; \(\text{year} \times \text{restate}\) is the interaction term between year and restatement indicator; variables leverage, MTB (Market-to-Book ratio), SEQ (Shareholder’s Equity), Dividend yield, industry category and sale are company characteristics control variables; and the variable age is CEO’s age.

Before fitting the two-part model with AR(1) covariance structure, we applied SAS macro MIXCORR provided by Tooze to get initial starting values of parameter estimates in order to make the MCEM algorithm converge faster to the correct values. We also transformed a few independent variables such as sale and SEQ (Shareholders’ equity) that have large values in order to make the associated coefficients smaller. We ran the MCEM algorithm until it converged when we fit the two-part AR(1) covariance structure model. The results are presented in Table 4.10.

When we compare the result from fitting the two-part AR(1) covariance structure model with the model we fitted in Table 4.8, we see that the results from the Gaussian adaptive method are mostly consistent: most parameter estimates are similar and the p-values suggest the same significance level except the sign of coefficient estimates for leverage and MTB reversed. Comparing the estimates by the MCEM with the ones from the Gaussian adaptive method, we notice that in general, the estimates are still relatively similar, however, for some variables, the significance level also changed. Based on this model, we found that restatement became less significant: in the first part (the occurrence model), the coefficient estimate for restatement is 0.4 and the p-value is around 0.4; in the second part (the intensity model), the coefficient estimate for restatement was no longer significant (p-value=0.7). Moreover, the p-value for time fell below 0.03, which suggests a stronger time effect after controlling for other factors. Other differences include SEQ and interaction between restatement and time: SEQ became more significant after the transformation and even though the interaction effect between restatement and time remained negative, our model suggests that the difference
of option grants between restating and non-restating companies over time is no longer statistically significant (p-values are 0.06 and 0.46). Other variables such as earning per share and leverage are still not significant.

We also reported the Bayesian information criterion (BIC) in the table for model comparison purpose. To compute the BIC, we performed the following approximation: recall that companies/subjects are assumed to be independent of each other and the observed data likelihood can be written in terms of the complete data likelihood in the following form:

\[ f(Y|\Theta) = \int f(Y, b|\Theta) db. \]

Therefore, the observed data log-likelihood can be written as

\[ l = \sum_{i=1}^{N} l_i = \sum_{i=1}^{N} \log \int f(y_i|b_i, \Theta)g(b_i|\Sigma)db_i. \]

The expression on the right hand side can be approximated by

\[ \hat{l} \approx \sum_{i=1}^{N} \log \left( \frac{1}{M} \sum_{j=1}^{M} f(y_i|b_i^{(j)}, \hat{\Theta}) \right) = \sum_{i=1}^{N} \log \left( \frac{1}{M} \sum_{j=1}^{M} \prod_{t=1}^{n_i} \left\{ \frac{1 - r_{it}}{1 + \exp(X_{1it}\hat{\beta}_1 + b_{it}^{(j)})} + \frac{r_{it} \exp(X_{1it}\hat{\beta}_1 + b_{it}^{(j)})}{1 + \exp(X_{1it}\hat{\beta}_1 + b_{it}^{(j)})} \right\} \Phi(X_{2it}\hat{\beta}_2 + b_{it}^{(j)}, \hat{\sigma}_e^2) \right) \]

\( b_i^{(j)} \)'s \((j = 1, \cdots, M)\) are M random samples generated from \( g(b_i|\Sigma) \) and \( \hat{\Theta} \) is the maximum likelihood estimate of all parameters in the model we obtained from the MCEM method. M is usually a large number and we specify M to be 200,000 in order to get a close approximation. Last, \( BIC = -2\hat{l} + p\log(n) \), where \( p \) is the number of parameters being estimated and \( n \) is the actual number of observations used in the estimation.

Next, we checked model goodness-of-fit by examining the normality of random effects and residual plots for both the two-part model with correlated bivariate normal random effects (we would from now on refer to this as Tootz's model) and our model that has AR(1) covariance structure random effects. From Figure 4.10, we see that the distribution of random effects in the first part of Tootz's model does not appear normal. The histogram shows that the distribution of random effect \( u_{1i} \)'s is skewed and the normal quantile-quantile plots clearly depart from a straight line. Figure 4.11 for random effect \( u_{2i} \) in the second part of Tootz's model shows that \( u_{2i} \)'s are approximately normally distributed with mean 0. Recall that for each company, we drew \( m \) Monte Carlo samples from the posterior density \( f(b|Y, \beta_1, \beta_2, \phi_1, \sigma_e, \sigma_1) \). We used the average of the \( m \) MC samples we drew in the last iteration as an estimate of random effects. The distributions of the random effects in our
Table 4.10: Results from fitting the Full Model

<table>
<thead>
<tr>
<th></th>
<th>SAS</th>
<th>MCEM</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial Value</td>
<td>Estimate</td>
<td>StErr.</td>
<td>p-value</td>
</tr>
<tr>
<td><strong>First Part</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.3460</td>
<td>1.6833</td>
<td>0.2358</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>restate-dummy</td>
<td>0.3333</td>
<td>0.2561</td>
<td>0.1299</td>
<td>0.0486</td>
</tr>
<tr>
<td>time</td>
<td>0.0143</td>
<td>0.0204</td>
<td>0.0092</td>
<td>0.0258</td>
</tr>
<tr>
<td>restate*time</td>
<td>-0.0492</td>
<td>-0.0337</td>
<td>0.0181</td>
<td>0.0628</td>
</tr>
<tr>
<td>industry1</td>
<td>0.0356</td>
<td>0.0478</td>
<td>0.0866</td>
<td>0.5810</td>
</tr>
<tr>
<td>industry2</td>
<td>0.1680</td>
<td>0.1639</td>
<td>0.0831</td>
<td>0.0487</td>
</tr>
<tr>
<td>industry3</td>
<td>0.4592</td>
<td>0.4340</td>
<td>0.0914</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>industry4</td>
<td>0.8772</td>
<td>0.8728</td>
<td>0.1214</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>log(SALE)</td>
<td>0.2158</td>
<td>0.1390</td>
<td>0.0329</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>CEO Age</td>
<td>-0.0458</td>
<td>-0.0404</td>
<td>0.0034</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Leverage</td>
<td>0.1804</td>
<td>-0.1302</td>
<td>0.1196</td>
<td>0.2764</td>
</tr>
<tr>
<td>Log(SEQ)</td>
<td>0.1193</td>
<td>0.0778</td>
<td>0.0360</td>
<td>0.0308</td>
</tr>
<tr>
<td>EPSFX</td>
<td>-0.009</td>
<td>-0.0053</td>
<td>0.0033</td>
<td>0.1067</td>
</tr>
<tr>
<td>Dividend</td>
<td>-0.0003</td>
<td>-0.0001</td>
<td>&lt; .0001</td>
<td>0.1722</td>
</tr>
<tr>
<td>MTB</td>
<td>0.005</td>
<td>-0.0323</td>
<td>0.0115</td>
<td>0.0049</td>
</tr>
<tr>
<td><strong>Second Part</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>4.1401</td>
<td>4.3611</td>
<td>0.1468</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>restate-dummy</td>
<td>0.1801</td>
<td>0.0497</td>
<td>0.1750</td>
<td>0.7763</td>
</tr>
<tr>
<td>time</td>
<td>0.0555</td>
<td>0.0547</td>
<td>0.0078</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>restate*time</td>
<td>-0.0269</td>
<td>-0.0138</td>
<td>0.0190</td>
<td>0.4676</td>
</tr>
<tr>
<td>industry1</td>
<td>-0.3267</td>
<td>-0.2953</td>
<td>0.0354</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>industry2</td>
<td>-0.5040</td>
<td>-0.5403</td>
<td>0.0485</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>industry3</td>
<td>0.5666</td>
<td>0.4791</td>
<td>0.0789</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>industry4</td>
<td>0.5634</td>
<td>0.5577</td>
<td>0.0906</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>log(SALE)</td>
<td>0.3745</td>
<td>0.1915</td>
<td>0.0252</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>CEO Age</td>
<td>-0.002</td>
<td>-0.0252</td>
<td>0.0028</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Leverage</td>
<td>0.0439</td>
<td>-0.0125</td>
<td>0.0828</td>
<td>0.8800</td>
</tr>
<tr>
<td>Log(SEQ)</td>
<td>0.0717</td>
<td>0.3288</td>
<td>0.0303</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>EPSFX</td>
<td>-0.005</td>
<td>-0.0059</td>
<td>0.0041</td>
<td>0.1456</td>
</tr>
<tr>
<td>Dividend</td>
<td>-0.000</td>
<td>-0.0002</td>
<td>&lt; .0001</td>
<td>0.0132</td>
</tr>
<tr>
<td>MTB</td>
<td>0.0079</td>
<td>0.0513</td>
<td>0.0067</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Var(Residual)</td>
<td>0.7109</td>
<td>0.8004</td>
<td>0.0077</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>-</td>
<td>0.5156</td>
<td>0.0113</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>$\phi$</td>
<td>-</td>
<td>0.8400</td>
<td>0.001</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td><strong>BIC</strong></td>
<td>35166.50</td>
<td>36582</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
model are shown in Figure 4.12 and 4.11. We see from the graphs that random effects in both part of our model appear to be approximately normal and symmetric about 0.

![Random Effects in the first part of the Model](image1)

(a)random effects in the first part (Gaussian)

![Random Effects in the second part of the Model](image2)

(b)random effects in the second part (Gaussian)

Figure 4.10: Random Effects in the first part of the Model *without* AR(1) Effect

Figure 4.11: Random Effects second part of the Model *without* AR(1) Effect

We also examined the residual plots for both models. As we discussed in section 4.2.2, there are two types of residuals we can consider, the subject-specific (conditional) residuals and the population-averaged (marginal) residuals. We observed that even though there is no obvious pattern in the population-averaged residuals for Tooze’s model, there is a systematic trend in the subject-specific residuals. The subject-specific residuals decrease when the predicted values increase. This suggests that after adjusting for fixed effects and
conditional on subject-specific random effects, the variance of the random errors $\epsilon_{it}$ is not constant. Figure 4.15 shows the residuals versus predicted values from the full model with AR(1) random effects. Based on these graphs, we see that there are outliers in the data and the clear trend we saw in the residual plots of Tooze’s model is no longer present. Therefore, we believe that our model fits the data better than Tooze’s model.
In the next step, we removed variables “leverage” and “EPSFX” because they were not significant in both parts of our model. In this reduced model, we also transformed variable “dividend” by dividing it by 100 to force it to a smaller scale. We then repeated the steps above, the results are summarized in Table 4.11. After two variables were removed from both parts of the model, the magnitude and significance levels of most of the remaining control variables did not change significantly. This model suggests that variables “Dividend” no longer has any statistically significant impact on the decision of offering option grants or the magnitude of the option granted to the CEOs. In the second part of this model, most variables are significant other than effect of restatement indicator and interaction between
restatement and time. Furthermore, BIC for this model is slightly larger than the full model (37730 compared to 36582 in the full model).

We proceed to the next model by further removing the insignificant variables from the model. In the first part of the model with AR(1) random effects, we removed variable “Dividend” while for the second part, we removed “Dividend” and the interaction of restatement and time effect. We kept the restatement indicator in the model because it is our main interest to see whether there is any difference in option grants between restating companies and non-restating companies. We show the parameter estimates and standard errors in Table 4.12. Compared with model in Table 4.11, this model has a slightly smaller BIC value and the magnitude of parameter estimates remained similar. Based on the first part of this model, restating companies are 1.23 times more likely to grant options to their CEOs than non-restating companies given that other covariates are equal; however, this effect was not statistically significant with p-value=0.1057 (In Tooze’s model restatement effect was not significant at significance level=0.1). The restatement is also found to be insignificant in the second part, which means there is no significant difference between restating and non-restating companies in the amount of options granted. Time has a positive effect on both parts, which suggests that the probability and amount of options increases over time. The interaction term between restatement and time is no longer significant in the first part of our model. As for other control variables, we found that industry was an important factor that have impacts on both the likelihood and value of option grants. The coefficient estimates of industry groups are consistent in all models we fitted and these models suggest that Industry group 3 (Business and etc.) and Industry group 4 (Healthcare related industries) are more likely to offer options than other industries while there is no significant difference among Industry 1 (Consumer), Industry 2 (Manufacturing) and other industries; interestingly, companies in Industry 1 and 2 grant fewer options and Industry 3, 4 grant more than other industries (conditional on the fact there were options granted). As we expected, company size and shareholder’s equity are positive in both parts of our model. This model also suggests that CEO’s age has negative impact on the probability and amount of options. One possible interpretation for this is the closer CEOs are to their retirement, the less willing they are to accept options as compensation. Note that the market-to-book (MTB) ratios were not significant in Tooze’s model and thus were excluded from the model we discussed in 4.9, however, the estimates from our proposed model were both significant. In the occurrence
Table 4.11: Results from fitting a Reduced Model with fewer covariates

<table>
<thead>
<tr>
<th></th>
<th>SAS Initial Value</th>
<th>SAS Estimate</th>
<th>SAS StErr.</th>
<th>SAS p-value</th>
<th>MCEM Initial Value</th>
<th>MCEM Estimate</th>
<th>MCEM StErr.</th>
<th>MCEM p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Part</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.6011</td>
<td>1.7758</td>
<td>0.2322</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>restate-dummy</td>
<td>0.4393</td>
<td>0.2968</td>
<td>0.1297</td>
<td>0.0221</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>0.0169</td>
<td>0.0239</td>
<td>0.0092</td>
<td>0.0093</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>restate*time</td>
<td>-0.0577</td>
<td>-0.0373</td>
<td>0.0183</td>
<td>0.0413</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>industry1</td>
<td>0.1717</td>
<td>0.1007</td>
<td>0.0861</td>
<td>0.2421</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>industry2</td>
<td>0.3542</td>
<td>0.2016</td>
<td>0.0828</td>
<td>0.0150</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>industry3</td>
<td>0.6114</td>
<td>0.4612</td>
<td>0.0911</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>industry4</td>
<td>1.1988</td>
<td>0.8632</td>
<td>0.1193</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(SALE)</td>
<td>0.1959</td>
<td>0.1233</td>
<td>0.0212</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEO Age</td>
<td>-0.0495</td>
<td>-0.0422</td>
<td>0.0033</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log(SEQ)</td>
<td>0.1095</td>
<td>0.0900</td>
<td>0.0137</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dividend</td>
<td>-0.0220</td>
<td>-0.0078</td>
<td>0.0079</td>
<td>0.3278</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTB</td>
<td>0.0011</td>
<td>-0.0290</td>
<td>0.0110</td>
<td>0.0086</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Second Part</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>4.1350</td>
<td>4.5167</td>
<td>0.2339</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>restate-dummy</td>
<td>0.1846</td>
<td>0.0890</td>
<td>0.1407</td>
<td>0.5272</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>0.0560</td>
<td>0.0614</td>
<td>0.0119</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>restate*time</td>
<td>-0.0272</td>
<td>-0.0176</td>
<td>0.0184</td>
<td>0.3397</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>industry1</td>
<td>-0.3313</td>
<td>-0.3574</td>
<td>0.1832</td>
<td>0.0511</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>industry2</td>
<td>-0.4848</td>
<td>-0.4578</td>
<td>0.1590</td>
<td>0.0040</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>industry3</td>
<td>0.5647</td>
<td>0.5695</td>
<td>0.1594</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>industry4</td>
<td>0.5677</td>
<td>0.6938</td>
<td>0.1456</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(SALE)</td>
<td>0.3786</td>
<td>0.3938</td>
<td>0.0200</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEO Age</td>
<td>-0.0119</td>
<td>-0.0245</td>
<td>0.0026</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log(SEQ)</td>
<td>0.0684</td>
<td>0.0568</td>
<td>0.0107</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dividend</td>
<td>-0.0052</td>
<td>-0.0070</td>
<td>0.0072</td>
<td>0.3274</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTB</td>
<td>0.0512</td>
<td>0.0535</td>
<td>0.0076</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var(Residual)</td>
<td>0.7114</td>
<td>0.8088</td>
<td>0.0077</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>-</td>
<td>0.6126</td>
<td>0.0012</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi$</td>
<td>-</td>
<td>0.7977</td>
<td>&lt; .0001</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BIC</strong></td>
<td>35956.21</td>
<td>37730</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
part, we found that higher MTB is associated with lower probability of options, in other words, the more a company is overvalued, the less likely the CEO is going to have options. Moreover, given that options were granted, MTB has a positive impact on the magnitude of options. The autocorrelation coefficient estimates from this model was around 0.8, which shows that there is a strong serial correlation in our data.

Similarly as before, we performed a model check, the histogram and QQ plots in Figure 4.16 and 4.17 show that random effect in the first part of Tooze’s model clearly depart from a normal distribution while the random effect in the second part is approximately normal.

We used standard regression diagnostics again to assess goodness of fit of the models by examining the residual plots. Both models assume normality and constant variance of residuals. Figure 4.18 show residual values versus predicted values for Tooze’s model. There is no obvious pattern shown in the population averaged (marginal) residual plot, however, the significant down trend still can be seen in the subject-specific (conditional) residual plot. Moreover, we examined the normality assumption of the residuals. From Figure 4.19 and 4.20, we see that even though these two types of residuals are symmetric about zero and have bell-shaped distributions, the normality quantile-quantile plots suggest that the assumption of normality may need to be modified.

Figure 4.21 and 4.22 are plots of random effects for the two-part model with AR(1) covariance structure. Compared to Figure 4.16 and 4.17, these estimates of random effects are more symmetric and their distributions are both approximately normal. The residual plots Figure 4.23 show both population residuals and subject-specific residuals versus predicted values from our proposed model. Similarly as the full model residuals, the plots show no obvious patterns and trend. The last two figures 4.24 and 4.25 are histograms and Q-Q plots of the residuals. We can see from the plots that both population-averaged and subject-specific residuals are symmetric about zero and bell-shaped. However, the Q-Q plot of the residuals shows violation of the normality assumption of the random errors in our model.
Table 4.12: Results from the final Model with AR(1) random effect

<table>
<thead>
<tr>
<th></th>
<th>SAS</th>
<th>MCEM</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial Value</td>
<td>Estimate</td>
<td>StErr.</td>
<td>p-value</td>
</tr>
<tr>
<td><strong>First Part</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.6797</td>
<td>1.8369</td>
<td>0.2262</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>restate-dummy</td>
<td>0.4249</td>
<td>0.2097</td>
<td>0.1296</td>
<td>0.1057</td>
</tr>
<tr>
<td>time</td>
<td>0.0168</td>
<td>0.0216</td>
<td>0.0909</td>
<td>0.0166</td>
</tr>
<tr>
<td>restate*time</td>
<td>-0.0568</td>
<td>-0.0290</td>
<td>0.0181</td>
<td>0.1080</td>
</tr>
<tr>
<td>industry1</td>
<td>0.1886</td>
<td>0.0857</td>
<td>0.0849</td>
<td>0.3130</td>
</tr>
<tr>
<td>industry2</td>
<td>0.3378</td>
<td>0.1441</td>
<td>0.0816</td>
<td>0.0775</td>
</tr>
<tr>
<td>industry3</td>
<td>0.6202</td>
<td>0.4415</td>
<td>0.0905</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>industry4</td>
<td>1.2095</td>
<td>0.8442</td>
<td>0.1187</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>log(SALE)</td>
<td>0.1851</td>
<td>0.1084</td>
<td>0.0202</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>CEO Age</td>
<td>-0.0496</td>
<td>-0.0416</td>
<td>0.0033</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Log(SEQ)</td>
<td>0.1072</td>
<td>0.0906</td>
<td>0.0135</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>MTB</td>
<td>0.0001</td>
<td>-0.0366</td>
<td>0.0113</td>
<td>0.0012</td>
</tr>
<tr>
<td><strong>Second Part</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>4.1915</td>
<td>4.6010</td>
<td>0.3015</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>restate-dummy</td>
<td>0.0052</td>
<td>-0.0604</td>
<td>0.0736</td>
<td>0.4119</td>
</tr>
<tr>
<td>time</td>
<td>0.0560</td>
<td>0.0586</td>
<td>0.0063</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>industry1</td>
<td>-0.3298</td>
<td>-0.3778</td>
<td>0.1041</td>
<td>0.0003</td>
</tr>
<tr>
<td>industry2</td>
<td>-0.4898</td>
<td>-0.5058</td>
<td>0.0738</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>industry3</td>
<td>0.5678</td>
<td>0.5553</td>
<td>0.0813</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>industry4</td>
<td>0.5576</td>
<td>0.6633</td>
<td>0.1419</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>log(SALE)</td>
<td>0.3760</td>
<td>0.3839</td>
<td>0.0241</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>CEO Age</td>
<td>-0.0118</td>
<td>-0.0245</td>
<td>0.0032</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Log(SEQ)</td>
<td>0.0691</td>
<td>0.0573</td>
<td>0.0108</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>MTB</td>
<td>0.0516</td>
<td>0.0464</td>
<td>0.0061</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Var(Residual)</td>
<td>0.7156</td>
<td>0.8151</td>
<td>0.0078</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>-</td>
<td>0.4849</td>
<td>0.0085</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>$\phi$</td>
<td>-</td>
<td>0.7977</td>
<td>0.0002</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td><strong>BIC</strong></td>
<td>36174.48</td>
<td>37725</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.16: Random Effects in the first part of the Model without AR(1) Effects

Figure 4.17: Random Effects in the second part of the Model without AR(1) Effects
Figure 4.18: Residual Plots of the Model without AR(1) Effect: further reduced model

Figure 4.19: Residual Histograms of the Model without AR(1) Effect: further reduced model
Figure 4.20: Residual QQ-Plots of the Model \textit{without} AR(1) Effect: further reduced model

Figure 4.21: Random Effects in the first part of the Model with AR(1) Effects: further reduced model
Figure 4.22: Random Effects in the second part of the Model with AR(1) Effects: further reduced model

(a) Population-Averaged Residual Plots
(b) Subject-Specific Residual Plots

Figure 4.23: Residual Plots of the Model with AR(1) Effect: further reduced model
Figure 4.24: Residual Histograms of the Model with AR(1) Effect: further reduced model

(a) Population-Averaged Residuals: histogram
(b) Population-Averaged Residuals: QQ Plot

Figure 4.25: Residual QQ-Plots of the Model with AR(1) Effect: further reduced model

(a) Subject-Specific Residuals: histogram
(b) Subject-Specific Residuals: QQ Plot
CHAPTER 5

Discussion and Future Work

Zero-inflated data are frequently encountered in many areas and two-part models have been proposed to handle zero-inflated data. Most models in literatures deal with cross-sectional data and the two separate parts are usually assumed to be independent. It has been shown by Su et al. [33] that bias can be induced for regression coefficients when random effects in different parts are correlated but misspecified as independent in a two-part mixed model. When modeling executive option grants, we encountered longitudinal data with excessive amount of zeros. Modeling zero-inflated longitudinal data is more challenging because repeated observations were obtained from the same group of companies over time, therefore, observations of the same company are very likely to be correlated. It is essential that we specify an appropriate correlation structure for the data. Following an approach proposed by Tooze et. al. [35], we applied a two-part model with correlated random effects to the executive compensation data. The correlated random effects were introduced into the model to take into account subject-specific effects. This approach can be implemented in standard statistical softwares such as SAS.

Although Tooze’s model adjusts for subject-specific effects, it does not account for serial correlation. In longitudinal data analysis, it is often observed that dependence between repeated measurements exist within the same subject (company in our example) and it usually decays when time separation between them increases. Heagerty and Kurlandciteref:Heagerty have proved through simulation study that ignoring variation in the within-cluster correlation can result in severe bias for regression estimates, moreover, misspecification of correlation structure of random effects when serial correlation presents in the data also leads to biased estimates. In order to adjust for serial correlation and subject effects in longitudinal data, we proposed a two-part model with AR(1) random effects. We used Monte Carlo EM (MCEM)
algorithm for model estimation and through a simulation study, we examined performance of the MCEM algorithm. We found that when the serial correlation is present in the data, our model yields more accurate estimates than those given by Tooze’s model.

Finally, we applied the two-part model with AR(1) random effects to our data. We compared the results with Tooze’s model. We found that estimates for regression coefficients from both models had similar values, however, the significance levels were not identical. By further examining model goodness-of-fit, we noticed that even though the only difference between these two models was covariance structure of the random effects, the model with AR(1) random effects fit the data better: the empirical Bayesian estimates of random effects from the model without serial correlation showed violation of the normality assumption and the corresponding subject-specific residual plots showed a clear trend.

As we mentioned before, there is no written procedure in statistical softwares that fits two-part models with serial correlation. We implemented our proposed model in Matlab. Note that, even though the MCEM algorithm has been proved to yield good and stable estimates, it is computationally expensive. In future research, we should explore alternative numerical approximation approaches that are less computational intensive to estimate the model. Another area that needs improvement is the ability to account for missing values. A critical assumption in analyses that involve autoregressive structure is that observations are sampled with the same frequency. Note that even though our proposed model allows unbalanced data, we have not explored situations where there are intermittent missing observations. There are several ways of dealing with the missing value problem, interpolation and multiple imputation are two common methods used in literatures. Missing values not only makes it difficult to implement our model, but also adds noise to data. Recall that in the two-part models with correlated random effects we fitted in section 4.2.2, we saw significant effects of restatement on the likelihood of granting options and the magnitude of options. However, when we ran the same models using companies without any intermittent missing values, effect of restatement became insignificant. This suggests that these missing values may be informative and careful measures should be taken when dealing with these missing values.

In our application example of modeling CEO’s option grants, we focused on shared random effects model, in other words, random effects in both part are assumed to have the same structure. It may be necessary to specify different structures for random effects in
our model. More importantly, when the model becomes more complicated, we will need a more sophisticated diagnostic algorithm to assess the goodness-of-fit.

For the example that motivated this study, both two-part models with and without AR(1) random effects suggest that our main interest “restatement” does have impact on the decision of granting options as part of their compensation package to CEOs: the restating companies have higher probability of granting options. However, our study did not provide enough evidence to show that there was a statistically significant difference in option grants between restating and non-restating companies given that there were option grants. Recall that our study period was from 1997 to 2006 and most company samples in previous studies were taken before year 2002. It might be interesting to see if the models give different estimates when applied to different time period given the policy change and the “IT bubble” occurred around year 2000. Cheng and Farber [7] found that CEO’s option compensation declines significantly in two years following the restatements in their study period (1997-2001) and they concluded performance of these companies improved due to the decrease of options in CEO compensation. Following this idea, we replicated their tests by running a few two-sample t-test with our sample. Surprisingly, we did not find any significant decrease in CEO’s compensation following the restatements. We also ran two-sample t-tests on different time period, results show that decrease in option grants was significant before 2002 while it became insignificant after 2002. Note that even though these exploratory tests do help us understand more about change in option grants in CEO pay package, they may not be reliable. As we have seen before, there are a large portion of zeros in CEO option grants, therefore, they are not normally distributed. The normality and independence assumption of two-sample t-test and OLS regression are clearly violated.

This study can be extended in several directions: first, other than bonus, we can examine other performance related compensation such as annual bonus and restricted stock. Second, instead of using the Black-Scholes value of newly granted options as the response variable, in future study we can also model the relative value of options to total compensation or cash compensation. Third, in the current analysis, we included a few most commonly used control variables in our model, however, they may not explain options in CEO compensation adequately, it is important that we investigate other possible factors that may have impact on CEO compensation package. Fourth, we categorized companies into two types: restating companies and non-restating companies, this specification indicates effect of restatement on
option grants in the entire study period. In future research, we can use a restatement-year indicator to indicate a restatement in a specific company year. By doing this, we will be able to model direct effect of a restatement on CEO option compensation. Moreover, it would be interesting to see if restatements have impact on option grants in the years following the restatements. Last, restatements usually result in CEO turn over. There might be difference between new CEOs’ pay package and CEOs who left their companies. In literature, researchers have attempted to capture this factor by including dummy variables that indicate change in CEOs or departing CEOs. Investing this factor may help us understand CEO compensation package better.
APPENDIX A

Results from a simplified model

Here we show some results from a simplified model we fitted. In this model, we included restatement, time, interaction between time and restatement, industry and logarithm of sale as our control variables. From results presented in table A.1, we see that all the control variables have significant impact on our responses. The restatement indicator, which was shown not significant in the models we discussed in section 4.2.2, turns out to be significant in both parts of this simplified model. Moreover, the estimate of the autocorrelation coefficient is 0.15 and it is not significant at level 0.5. This suggest some of the control variables in the more complicated models may have strong autocorrelations and the restatement indicator might be highly correlated with some of them. When we compare this simple model with the other models we showed in section 4.2.2, we see that BIC associated with this model is much higher, thus, this simplified model is not adequate for our data.
Table A.1: Parameter Estimates of the Simplified Model

<table>
<thead>
<tr>
<th></th>
<th>Initial Value</th>
<th>Estimate</th>
<th>StErr.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Part</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.0131</td>
<td>-0.5772</td>
<td>0.1049</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>restate-dummy</td>
<td>0.4363</td>
<td>0.3569</td>
<td>0.0972</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>time</td>
<td>0.0145</td>
<td>0.0145</td>
<td>0.007</td>
<td>0.0372</td>
</tr>
<tr>
<td>restate*time</td>
<td>-0.0598</td>
<td>-0.0433</td>
<td>0.0137</td>
<td>0.0016</td>
</tr>
<tr>
<td>industry1</td>
<td>0.1226</td>
<td>0.0834</td>
<td>0.0629</td>
<td>0.1852</td>
</tr>
<tr>
<td>industry2</td>
<td>0.3285</td>
<td>0.1915</td>
<td>0.0610</td>
<td>0.0017</td>
</tr>
<tr>
<td>industry3</td>
<td>0.7088</td>
<td>0.4898</td>
<td>0.0649</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>industry4</td>
<td>1.2690</td>
<td>0.8549</td>
<td>0.0881</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>log(SALE)</td>
<td>0.2702</td>
<td>0.1712</td>
<td>0.0123</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td><strong>Second Part</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>3.8309</td>
<td>3.794</td>
<td>0.0048</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>restate-dummy</td>
<td>0.1621</td>
<td>0.1834</td>
<td>0.0287</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>time</td>
<td>0.0479</td>
<td>0.0568</td>
<td>0.0023</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>restate*time</td>
<td>-0.0221</td>
<td>-0.0247</td>
<td>0.0011</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>industry1</td>
<td>-0.3864</td>
<td>-0.3470</td>
<td>0.0330</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>industry2</td>
<td>-0.5647</td>
<td>-0.4988</td>
<td>0.0151</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>industry3</td>
<td>0.6823</td>
<td>0.6825</td>
<td>0.0062</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>industry4</td>
<td>0.6583</td>
<td>0.7223</td>
<td>0.0398</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>log(SALE)</td>
<td>0.4186</td>
<td>0.4069</td>
<td>0.0012</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Residual</td>
<td>0.7484</td>
<td>1.2087</td>
<td>0.0197</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.1017</td>
<td>0.0001</td>
<td>&lt; .0001</td>
<td></td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.1509</td>
<td>0.0813</td>
<td>0.0634</td>
<td></td>
</tr>
<tr>
<td><strong>BIC</strong></td>
<td>46044</td>
<td>52268</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX B

Overall Effects of Covariates

Following Tooze et. al.’s approach [35], we will discuss briefly on how to evaluate the overall effect of a covariate on the mean of the response variable.

If a covariate is in both the occurrence model and intensity model, we will need to evaluate the overall effect of this covariate on the mean of our response. Suppose $M$ is a covariate in both parts of the model. We have the following:

$$E(Y|M = m + 1, \theta) = E(Y|M = m + 1, \theta, R = 1)P(R = 1|M = m + 1, \theta)$$

$$= E(V|M = m + 1, \theta_2)P(R = 1|M = m + 1, \theta_1)$$

then the ratio of conditional mean of $Y$ when $M = m + 1$ to $M = m$ is:

$$\frac{E(Y|M = m + 1, \theta)}{E(Y|M = m, \theta)} = \frac{P(R = 1|M = m + 1, \theta_1)}{P(R = 1|M = m, \theta_1)} \cdot \frac{E(V|M = m + 1, \theta_2)}{E(V|M = m, \theta_2)}$$ (B.1)

Recall that in this logistic-lognormal model, we have the following properties of the continuous response:

- $P(R = 1|\theta_1) = \frac{\exp(X_1 \beta_1 + \alpha_1 M + u_1)}{1 + \exp(X_1 \beta_1 + \alpha_1 M + u_1)}$

- $E(V|\theta_2) = \exp(X_2 \beta_2 + \alpha_2 M + u_2 + \frac{\sigma^2}{2})$

We can take the extreme cases of the ratio, which is when $X_1 \beta_1 + u_1$ is large and positive and when it is large and negative. The limiting ratios are $\exp(\alpha_2)$ and $\exp(\alpha_1)\exp(\alpha_2)$, respectively [35]. In our example, we used median values of continuous covariates and the limits shown above to estimate the effects of some covariates of interest on option value. Effects of restatement and industry on option value are shown in table B.1.

We calculated the estimate of means ratio using the median of firm size (log(Sale)), median of CEO’s age, median of time and 1st and 3rd quantiles of time-restatement
interaction. Having controlled for other factors, restatement is associated with increased mean B-S value of option grant ranging from 1.4 to 1.9 times. For industry effect, we used Industry 5 (Other Industry) and Industry 3 (HiTech,Business) to calculate the range of means ratio. Being in Industry group 3 is associated with higher mean option value by 2.22 to 2.32 times.

Table B.1: Effects of Restatement and Industry on Option Value

<table>
<thead>
<tr>
<th>Var</th>
<th>Restate</th>
<th>Industry</th>
<th>$e^{a1}$</th>
<th>$e^{a2}$</th>
<th>Prob Ratio</th>
<th>Estimated Ratio of mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restate</td>
<td>N/A 5</td>
<td>1.58</td>
<td>1.1954</td>
<td>0.7087</td>
<td>1.339</td>
<td></td>
</tr>
<tr>
<td>Restate</td>
<td>N/A 3</td>
<td>1.58</td>
<td>1.1954</td>
<td>0.9998</td>
<td>1.888</td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td>0 N/A 1</td>
<td>1.731</td>
<td>2</td>
<td>0.67</td>
<td>2.32</td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td>1 N/A 1</td>
<td>1.731</td>
<td>2</td>
<td>0.64</td>
<td>2.22</td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


BIOGRAPHICAL SKETCH

Anqi Tang

Anqi Tang was born in 1983 in Pingdingshan, China. In the Summer of 2006, she completed her Bachelors degree in Statistics at Sun Yat-Sen University, Guangzhou, China. Anqi enrolled in the doctoral program in statistics at FSU in the fall of 2006 and is expected to graduate in the spring of 2011 under the advisement of Prof. Xufeng Niu. Anqi’s research interests include longitudinal data analysis, two-part models for zero-inflated data, time series and applications in executive compensation.