

**Assessment of Agreements
in Linear and Generalized Linear Mixed Models**

BY

YUE YU

Bachelor of Science, Fudan University, Shanghai, China, 2007

Master of Science, University of Illinois at Chicago, Chicago, Illinois, 2009

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Chicago, 2012

Chicago, Illinois

Defense Committee:

Samad Hedayat, Chair and Advisor

Jie Yang

Jing Wang

Huayun Chen, School of Public Health

Lawrence Lin, Baxter Healthcare

Copyright by

Yue Yu

2012

This thesis is dedicate to my parents,
my wife Jiewei Zeng, and daughter Sophia S. Yu

ACKNOWLEDGMENTS

I would like to express my deep gratitude to my advisor Prof. Samad Hedayat, whose excellent guidance and unsurpassed support for me has not only made this dissertation possible, but also made my entire graduate study at UIC an enjoyable and rewarding experience.

My special thanks are for my co-advisor Dr. Lawrence Lin, for giving me tremendous guidance in my research. Without his continuous encouragement, enlightening instruction, and patient advice, this work could not been accomplished.

My another sincere thanks go to Prof. Jie Yang for guiding me on the dimension reduction in regression project. His great advice, support and friendship has been invaluable on both an academic and a personal level.

I also want to thank Prof. Hakan Demirtas, Prof. Douglas Bates in University of Wisconsin at Madison, and Dr. Alice Richardson in University of Canberra for their helpful discussions and comments for this dissertation.

It is my honor to have Dr. Lawrence Lin, Prof. Jie Yang, Prof. Jing Wang, and Prof. Huayun Chen serving as members of my defense committee. I am extremely grateful for their extraordinary support and assistance.

Last, but not the least, I want to thank all my family members, for their love, understanding, and support.

ACKNOWLEDGMENTS (Continued)

The research work and writing for this dissertation is supported by National Science Foundation (NSF) Grand DMS-0603761 and DMS-0904125. The contents are solely the responsibility of the author and do not necessarily represent the official view of NSF.

Y.Y.

TABLE OF CONTENTS

<u>CHAPTER</u>	<u>PAGE</u>
1 INTRODUCTION	1
2 ASSESSMENT OF AGREEMENT	5
2.1 Mean Square Deviation (MSD)	8
2.2 Total Deviation Index (TDI) and Coverage Probability (CP)	9
2.3 Precision Coefficient and Accuracy Coefficient	10
2.4 Concordance Correlation Coefficient (CCC)	11
3 AGREEMENT STATISTICS FOR LINEAR ADDITIVE MIXED MODELS	13
3.1 Linear Additive Mixed Effects Model for Repeated Measurements	14
3.2 Individual Agreement Statistics	15
3.2.1 Intra-Rater Precision	15
3.2.2 Inter-Rater Agreement	16
3.2.3 Total-Rater Agreement	17
3.3 Overall Agreement Statistics	17
3.4 Estimations	21
3.4.1 Estimating Fixed Effects	23
3.4.2 Estimating Variance Components	24
3.5 Asymptotic Variances of Variance Components	26
3.5.1 With Normality	26
3.5.2 Without Normality	26
3.6 Asymptotic Properties of Agreement Statistics	27
4 SIMULATIONS AND CASE STUDIES FOR LINEAR MIXED MODELS	30
4.1 Simulations	31
4.1.1 Normal Cases	31
4.1.1.1 Without Covariates	31
4.1.1.2 With Covariates and 10% Missing Values	32
4.1.2 Non-Normal Cases	33
4.1.2.1 Discrete and Skewed Cases	33
4.1.2.2 Non-Linear Case	34
4.2 Case Studies	40
4.2.1 Cardiac Function Measurements	40
4.2.2 Antihypertensive Patch Dataset	44

TABLE OF CONTENTS (Continued)

<u>CHAPTER</u>		<u>PAGE</u>
5	AGREEMENT STATISTICS FOR GENERALIZED LINEAR MIXED MODELS	49
5.1	Generalized Linear Mixed Models (GLMM)	50
5.2	Agreement Statistics for Generalized Linear Mixed Models . .	53
5.2.1	Intra-Rater Precision	53
5.2.2	Inter-Rater Agreement	55
5.2.3	Total-Rater Agreement	56
5.2.4	Restricted Agreement Statistics	57
5.2.5	Overall Agreement Statistics	59
5.3	Estimation for Different Response Models	61
5.3.1	Dispersion Parameters and Variance Functions	61
5.3.2	Link Functions	61
5.3.3	Confidence Intervals	64
6	SIMULATIONS AND CASE STUDY FOR GENERALIZED LINEAR MIXED MODELS	66
6.1	Simulations	67
6.1.1	Log-Gamma Distribution	67
6.1.2	Log-Negative Binomial Distribution	69
6.1.3	Logistic-Binary Distribution	69
6.2	Antihypertensive Patch Dataset	74
7	CONCLUSION AND DISCUSSIONS	78
	APPENDIX	82
	CITED LITERATURE	86
	VITA	90

LIST OF TABLES

<u>TABLE</u>		<u>PAGE</u>
I	Simulation Results for Normal Data without Covariates and Missing Values	36
II	Simulation Results for Normal Data with Covariates and 10% Missing Values	36
III	Simulation Results for Binomial Data	37
IV	Simulation Results for Gamma Data	37
V	Simulation Results for Binomial Data with Covariates and Missing Values	38
VI	Simulation Results for Gamma Data with Covariates and Missing Values	38
VII	Simulation Results for Negative Binomial Data	39
VIII	Simulation Results for Negative Binomial Data with Covariates and Missing Values	39
IX	Cardiac Outputs by RV and IC for 12 Patients	41
X	Estimated Agreement Statistics for Cardiac Outputs	43
XI	Summary of the Crossover Design	44
XII	Estimated Restricted Agreement Statistics for $\log(\text{AUC})$	48
XIII	List of Dispersion Parameters and Variance Functions for Selected Distributions	62
XIV	List of the Canonical Links for Selected Distributions	62
XV	Simulation Results for Log-Gamma Distribution with 10% Missing Values	71
XVI	Simulation Results for Log-Gamma Distribution with Covariates .	71
XVII	Simulation Results for Log-Negative Binomial Distribution with 10% Missing Values	72
XVIII	Simulation Results for Log-Negative Binomial Distribution with Covariates	72
XIX	Simulation Results for Logistic-Binary Distribution with 10% Missing Values	73
XX	Simulation Results for Logistic-Binary Distribution with Covariates	73
XXI	Estimated Agreement Statistics for AUC	75

LIST OF FIGURES

<u>FIGURE</u>		<u>PAGE</u>
1	Scatter Plots for Several Cases	7
2	Means of the Cardiac Outputs for each Subject with Error Bars . .	42
3	Agreement Plots for $\log(\text{AUC})$	47
4	Histogram of $\log(\text{AUC})$	76
5	Histogram of Selected Estimates for 1000 Bootstrap Simulations . .	77

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
AUC	Area Under the Curve
CCC	Concordance Correlation Coefficient
CP	Coverage Probability
GEE	Generalized Estimating Equations
GLMM	Generalized Linear Mixed Models
IC	Impedance Cardiography
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimates
MSD	Mean Square Deviation
RCCC	Restricted Concordance Correlation Coefficient
REML	Restricted Maximum Likelihood
RV	Radionuclide Ventriculography
TDI	Total Deviation Index
WLS	Weighted Least Square

SUMMARY

Study of measuring agreement is intend to evaluate whether the readings from one rater/measurement agree with those from other raters/measurements. In this dissertation, we are going to present a general method to assess agreement for a large variety of data with repeated measurements using linear and generalized linear mixed models. In the first place, a set of agreement statistics, including mean square deviation, concordance correlation coefficient, precision and accuracy coefficients, is presented for evaluating the intra-, inter-, and total-rater agreement in the multiple-rater and multiple-replications cases. Secondly, likelihood-based approaches are developed to estimate all the agreement statistics. Asymptotic properties of these estimates are also discussed for different data structures. Furthermore, our method has the merit of handling missing values and covariates naturally, and a new set of restricted agreement statistics is proposed in order to capture the true random variations and between-instrument effects adjusted for the covariate effects. Simulations for both linear and generalized linear mixed models are conducted to show the accuracy and effectiveness of our approaches. In the end, two industry datasets are evaluated using our approach. One is the cardiac function measurements used to determine the agreement between impedance cardiography and radionuclide ventriculography estimates, and the other one is an antihypertensive patch dataset given by FDA for assessing individual bioequivalence.

CHAPTER 1

INTRODUCTION

Assessment of agreement has been a popular research topic in applied statistics during the past decades. It is mainly aimed to answer one crucial question, whether the readings from one rater/measurement agree with those from other raters/measurements. For example, we may want to compare several laboratory results collected in various labs, evaluate the performance between two medical devices, or find out if a newly developed method is interchangeable with the existing one. It has been widely used in bioinformatics, chemical engineering, pharmaceutical studies, psychology, and others.

Starting from the Pearson correlation coefficient, paired t -test, there are many methodologies and statistics developed in this area. Cohen (1), (2), Fleiss, Cohen, and Everitt (3), Fleiss (4), and Fleiss and Cuzick (5) studied the kappa coefficient to measure agreement for categorical data. Bland and Altman (6) stated agreement using graphical techniques. Lin (7), (8) introduced the famous concordance correlation coefficient (CCC) for a bivariate data, following with a lot of literatures in this area, for instance, Chinchilli, Martel, Kumanyika, and Lloyd (9), Robieson (10), Lin, Hedayat, Sinha, and Yang (11), Carrasco and Jover (12), and Lin, Hedayat, and Wu (13). Additionally, Barnhart and Williamson (14), Barnhart, Haber, and Song (15), and Barnhart, Song, and Haber (16) constructed the CCC for multiple raters and readings using generalized estimating equations (GEE).

In this dissertation, we are going to assess the agreement for both continuous and categorical data with repeated measurements using likelihood based approach in mixed models, since the maximum likelihood principle is well developed, and it is known to have consistency

and asymptotic normality properties, which leads to the desirable asymptotic behavior of our agreement estimates.

To begin with, a linear additive mixed models is discussed. All the fixed effects and variance components for random effects are estimated via the restricted maximum likelihood (REML) approach proposed by Thompson (17). Normality is not required in our model. For balanced data, REML estimates are exactly the moment-based ANOVA estimates, which are unbiased and have minimum variances in the class of all quadratic unbiased estimators of the variance components (Graybill (18), and Graybill and Hultquist (19)), no matter whether normality is assumed. For unbalanced data, REML is also preferred over other moment-based estimates by Corbeil and Searle (20), Khuri and Sahai (21), and Searle, Casella, and McCulloch (22). The asymptotic behavior of the agreement estimates is studied using the second derivatives of the log likelihood function for normal data, or the asymptotic variances estimates for non-normal data given by Richardson and Welsh (23).

After studying the linear mixed models, it is natural to ask the question, “what if the data do not fit in the linear mixed model setup?” Log-Gamma distributed data, and binary data with logistic regression are two examples. Thus, the agreement statistics should be able to generate into the generalized linear mixed model framework, in order to cover a wider range of distribution families. Carrasco (24) has studied the case of overdispersed count data using generalized concordance correlation coefficient. Based on his work, a complete set of multi-rater agreement estimates for arbitrary distribution families and link functions are proposed in this dissertation, with parametric bootstrap percentile intervals for characterizing uncertainty.

Furthermore, if covariates are involved, the measurements are more likely affected by those covariates. For example, readings from a sophisticated medical device may depend on the room temperature and humidity. These readings can not present the true differences among devices, even the same device can produce different readings under different environments. Hence, in our study, we eliminate the covariates effect estimated from the linear mixed models, and only consider the between and within instruments/raters differences, and random measuring errors.

The remaining chapters are organized as follows. Chapter 2 gives a short introduction for the idea of measuring agreement with some agreement statistics for a pair of observations. Chapter 3 proposes the intra-, inter-, and total-rater agreement statistics for repeated measurement using linear additive mixed models. Estimating procedures by restricted maximum likelihood approach are also given, along with the asymptotic properties of all agreement statistics. Simulation studies with covariates and missing values are showed, and two applications on real industry datasets are presented in Chapter 4. Chapter 5 extends the intra-, inter-, and total-rater agreement statistics into generalized linear mixed model framework for modeling a large variety of cases. Chapter 6 contains results of simulations and an interesting application. And the last Chapter is the conclusion and discussion.

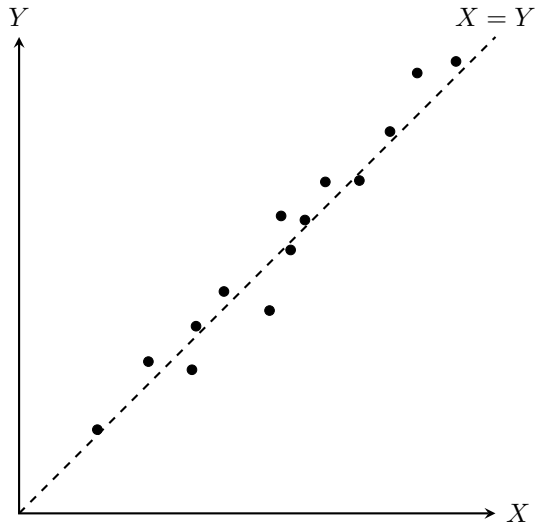
CHAPTER 2

ASSESSMENT OF AGREEMENT

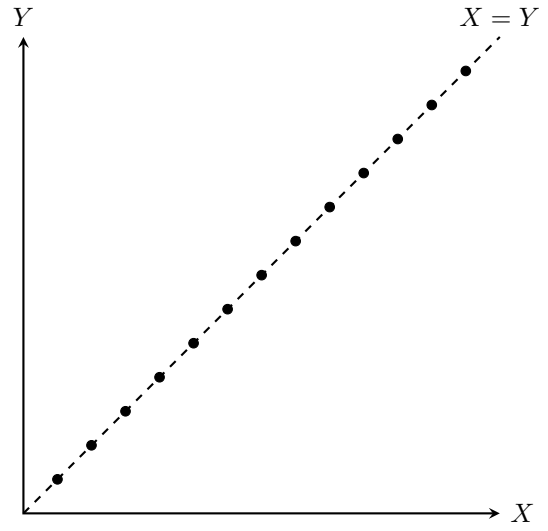
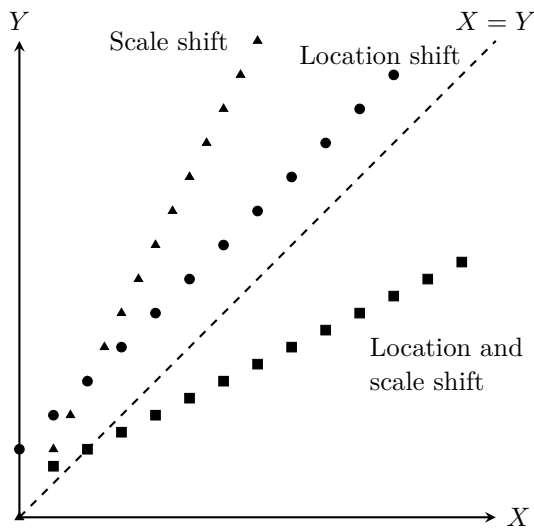
Consider a paired vector of observations X and Y , Figure 1a shows a scatter plot of an arbitrary paired vector X and Y , all the paired data are fluctuating around the 45° concordance line $X = Y$. If each reading in X is identical to the corresponding one in Y , then we say X and Y are in perfect agreement. In this case, all the points of the paired observations X and Y should be located exactly on the concordance line, which is presented in Figure 1b.

There are several widely used statistics or tests designed for measuring the agreement. The Pearson correlation coefficient, and the paired t -test statistic are two such statistics. However, both Pearson correlation coefficient and paired t -test are not good enough to characterize the agreement between X and Y . The Pearson correlation coefficient reflects the linear relationship but fails to detect the location and scale shift, while paired t -test evaluates the means rather than individual pairs. They can be misleading in some cases (Lin, (7)), as shown in Figure 1c and Figure 1d.

After the introduction of concordance correlation coefficient by Lin (7), there are some new statistics proposed for measuring agreement. To begin with, the most basic model is used to illustrate the idea of assessment of agreement, assuming X and Y have a bivariate distribution with means μ_x and μ_y , variances σ_x^2 and σ_y^2 , and covariance $\sigma_{xy} = \rho\sigma_x\sigma_y$.



(a) Scatter Plot of an Arbitrary Data

(b) Plot of Data if X and Y are Perfectly Agreed

(c) Cases when Person Correlation Coefficient Failed

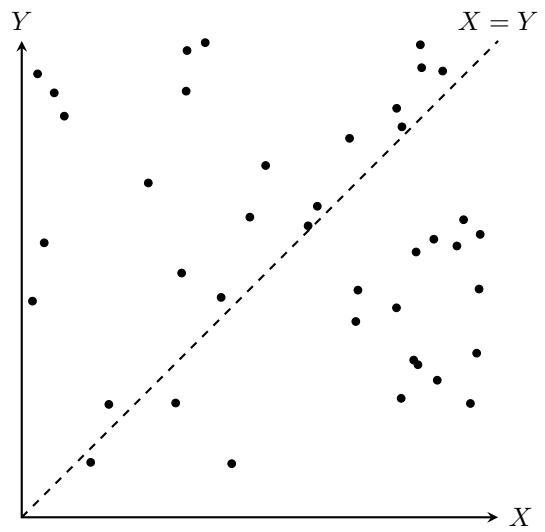
(d) Cases when t -test Failed

Figure 1: Scatter Plots for Several Cases

2.1 Mean Square Deviation (MSD)

Mean square deviation is a frequently-used measure for studying the differences between two vectors of observations. It evaluates an aggregated deviation from the identity line (25).

ε^2 is used to denote the MSD, which is defined as

$$\varepsilon^2 = E(X - Y)^2. \quad (2.1)$$

Based on the model assumption, MSD can be rewritten as

$$\varepsilon^2 = (\mu_x - \mu_y)^2 + \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}. \quad (2.2)$$

From Equation (2.2), we can see that MSD is always non-negative, and it equals zero if and only if $\mu_x = \mu_y$, $\sigma_x^2 = \sigma_y^2$ and $\rho = 1$. MSD increases when there is a large difference between means, or large variances. However, the drawback of MSD is that it is not intuitively meaningful and therefore there is no easy way to setup a common allowance of MSD for evaluating the agreement.

Lin (26) showed the asymptotic normality for MSD using log transformation.

$$\left(\ln(\hat{\varepsilon}^2) - \ln(\varepsilon^2) \right) \sim N \left(0, \frac{2}{n-2} \left[1 - \frac{(\mu_x - \mu_y)^4}{\varepsilon^4} \right] \right).$$

2.2 Total Deviation Index (TDI) and Coverage Probability (CP)

Lin, Hedayat, Sinha, and Yang (11) proposed a method to capture a proportion of data within a boundary from target value. Such proportion is called coverage probability and denoted as π , and such boundary is called total deviation index and denoted as δ .

The testing of TDI and CP are interchangeable. For example, we may want to evaluate that at least 90% of observations are located within 10% relative deviation of their target values. In this case, we can set the TDI at 10% and test whether the CP exceeds 90% or not, or conversely, set the CP at 90% and test whether the TDI exceeds 10% or not.

Assume that the difference between two variables is $D = X - Y$, which has a normal distribution with mean $\mu_d = \mu_x - \mu_y$ and variance $\sigma_d^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}$.

For a given TDI (δ_0), CP can be calculated by

$$\pi(\delta_0) = P(D^2 < \delta_0^2) = \chi^2\left(\delta_0^2, 1, \frac{\mu_d^2}{\sigma_d^2}\right), \quad (2.3)$$

where $\chi^2(\cdot)$ is the cumulative non-central chi-square distribution up to δ_0^2 , with one degree of freedom and non-centrality parameter of μ_d^2/σ_d^2 .

Similarly, for a given CP (π_0), TDI can be calculated by

$$\delta(\pi_0) = \sqrt{(\chi^2)^{-1}\left(\pi_0, 1, \frac{\mu_d^2}{\sigma_d^2}\right)}. \quad (2.4)$$

Lin (26) and Lin, Hedayat, Sinha, and Yang (11) suggested a simple version of TDI and CP using normal approximation.

$$\delta(\pi_0) \approx \Phi^{-1}\left(1 - \frac{1 - \pi_0}{2}\right)|\varepsilon^2|, \quad (2.5)$$

$$\pi(\delta_0) \approx \chi^2\left(\frac{\delta_0^2}{\varepsilon^2}, 1\right). \quad (2.6)$$

From the above Equations (2.5) and (2.6), we can see that both TDI and CP are functions of MSD, and a large MSD leads to a large TDI and a small CP, and they are determined once MSD is calculated.

2.3 Precision Coefficient and Accuracy Coefficient

The precision coefficient ρ is defined as the Pearson correlation coefficient of X and Y , which indicates the agreement of observations with their best fit linear line based on within-sample variations.

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}. \quad (2.7)$$

The accuracy coefficient indicates the agreement of the two marginal distributions. It can be decomposed into two parts, one is the location shift defined as $v = \frac{\mu_x - \mu_y}{\sqrt{\sigma_y \sigma_x}}$, the other one is the scale shift defined as $\varpi = \frac{\sigma_y}{\sigma_x}$ or $\frac{\sigma_x}{\sigma_y}$. Hence, the accuracy coefficient is

$$\chi_a = \frac{2}{\varpi + 1/\varpi + v^2}. \quad (2.8)$$

If $\mu_x = \mu_y$ and $\sigma_x^2 = \sigma_y^2$, the accuracy coefficient equals one, and if one of the variances approach infinity or there is a huge difference in means, the accuracy coefficient degenerates to zero.

2.4 Concordance Correlation Coefficient (CCC)

Since precision coefficient captures the linear relationship while accuracy coefficient captures the location and scale shift, Lin (7) proposed a compounded statistic called concordance correlation coefficient for measuring the agreement between two variables. CCC is denoted by ρ_c and has the form

$$\begin{aligned} \rho_c &= 1 - \frac{\text{Expected squared perpendicular deviation from } 45^\circ \text{line}}{\text{Expected squared perpendicular deviation from } 45^\circ \text{line} \\ &\quad \text{when X and Y are uncorrelated}} \\ &= 1 - \frac{E(X - Y)^2}{E(X - Y)^2|_{\rho=0}} \end{aligned} \quad (2.9)$$

$$= \frac{2\sigma_{xy}}{(\mu_x - \mu_y)^2 + \sigma_x^2 + \sigma_y^2} \quad (2.10)$$

$$= \rho\chi_a. \quad (2.11)$$

Here, $E(X - Y)^2$ gives the mean square for within sample total deviation, and $E(X - Y)^2|_{\rho=0}$ gives the mean square for total deviation. Thus, CCC is a standardized version of MSD, which measures whether these two vectors of observations are agreed along the concordance line or not. It is easy to see that CCC takes the range from -1 to 1 , where 1 means perfect agreement, -1 means perfect reverse agreement, and 0 means totally uncorrelated ($\sigma_{xy} = 0$).

Lin (7) also showed the asymptotic normality for CCC using Fisher's Z -transformation.

$$\left(\tanh^{-1}(\hat{\rho}_c) - \tanh^{-1}(\rho_c) \right) \sim N \left(0, \frac{1}{n} \left[\frac{(1 - \rho^2)\rho_c^2}{(1 - \rho_c^2)\rho^2} + \frac{2\rho_c^3(1 - \rho_c)v^2}{\rho(1 - \rho_c^2)^2} - \frac{\rho_c^4 v^4}{2\rho^2(1 - \rho_c^2)^2} \right] \right).$$

CHAPTER 3

AGREEMENT STATISTICS FOR LINEAR ADDITIVE MIXED MODELS

3.1 Linear Additive Mixed Effects Model for Repeated Measurements

In this chapter, we will study the agreement statistics for multiple raters with multiple replicates. The linear additive mixed effects model we use is assumed to follows,

$$y_{ijl} = \mu + \mathbf{x}_{ij}^c \boldsymbol{\theta} + \beta_j + \alpha_i + \gamma_{ij} + \epsilon_{ijl}. \quad (3.1)$$

The subscript i in (3.1) is the index of subjects, j is the index of raters, and l is the index of replications for each subject and rater. If the total number of subjects is n , the total number of raters is k , and the number of replications for subject i and rater j is m_{ij} , we have $i = 1, 2, \dots, n$, $j = 1, 2, \dots, k$, and $l = 1, 2, \dots, m_{ij}$.

In Model (3.1), μ is the overall mean and \mathbf{x}_{ij}^c is the $p \times 1$ design vector for $p \times 1$ covariate coefficients $\boldsymbol{\theta}$, β_j is the fixed inter-rater effect, α_i is the random subject effect with mean 0 and variance σ_α^2 , γ_{ij} is the random subject-rater interaction effect with mean 0 and variance σ_γ^2 , and ϵ_{ijl} is the random error with mean 0 and variance σ_ϵ^2 . In addition, ϵ_{ijl} is assumed to be uncorrelated with α_i and γ_{ij} .

Generally, the fixed inter-rater effect β_j can be included in the covariates $\boldsymbol{\theta}$. Therefore, Model (3.1) can be rewritten as

$$y_{ijl} = \mu + \mathbf{x}_{ij}' \boldsymbol{\vartheta} + \alpha_i + \gamma_{ij} + \epsilon_{ijl}, \quad (3.2)$$

where $\boldsymbol{\vartheta}' = (\beta_1, \beta_2, \dots, \beta_k, \boldsymbol{\theta}')$ and \mathbf{x}_{ij} is the corresponding design matrix.

3.2 Individual Agreement Statistics

3.2.1 Intra-Rater Precision

Intra-rater precision is used for measuring the agreement among multiple replications performed by a single rater. Here, replicates within a rater is assumed to be interchangeable. For given i and j , the intra-rater mean square deviation is defined as the mean of square difference between any two replications l and l' , $l, l' = 1, 2, \dots, m_{ij}$.

$$\begin{aligned}
 \varepsilon_{\text{intra}}^2|i, j &= E(y_{ijl} - y_{ijl'})^2 \\
 &= (\mathbf{x}'_{ij}\boldsymbol{\vartheta} - \mathbf{x}'_{ij}\boldsymbol{\vartheta})^2 + 2(\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2) - 2(\sigma_\alpha^2 + \sigma_\gamma^2) \\
 &= 2\sigma_\epsilon^2.
 \end{aligned} \tag{3.3}$$

From (2.9), for given i and j , the intra-rater CCC is

$$\rho_{c,\text{intra}}|i, j = \frac{\sigma_\alpha^2 + \sigma_\gamma^2}{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2}. \tag{3.4}$$

And the precision and accuracy coefficients become

$$\rho_{\text{intra}}|i, j = \frac{\sigma_\alpha^2 + \sigma_\gamma^2}{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2}, \tag{3.5}$$

$$\chi_{\alpha,\text{intra}}|i, j = 1. \tag{3.6}$$

3.2.2 Inter-Rater Agreement

Inter-rater agreement is used for measuring the agreement among the mean of raters. For given subject i , and any two raters j and j' , $j, j' = 1, 2, \dots, k, j \neq j'$, the inter-rater mean square deviation can be defined as follows

$$\begin{aligned}\varepsilon_{\text{inter}}^2|i, j, j' &= E(\bar{y}_{ij\cdot} - \bar{y}_{ijj'\cdot})^2 \\ &= (\mathbf{x}'_{ij}\boldsymbol{\vartheta} - \mathbf{x}'_{ijj'}\boldsymbol{\vartheta})^2 + 2\left(\sigma_\gamma^2 + \frac{\sigma_\epsilon^2}{m_{ijj'}}\right),\end{aligned}\quad (3.7)$$

where $\bar{y}_{ij\cdot}$ is the mean of m_{ij} replications from subject i and rater j , and $m_{ijj'} = 2m_{ij}m_{ij'}/(m_{ij} + m_{ij'})$.

Let $\mathbf{A}_{\mathbf{x}ijj'} = (\mathbf{x}_{ij} - \mathbf{x}_{ij'})(\mathbf{x}_{ij} - \mathbf{x}_{ij'})'/2$, then the inter-rater MSD (3.7) can be expressed as

$$\varepsilon_{\text{inter}}^2|i, j, j' = 2\left(\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}ijj'}\boldsymbol{\vartheta} + \sigma_\gamma^2 + \frac{\sigma_\epsilon^2}{m_{ijj'}}\right). \quad (3.8)$$

The inter-rater CCC, precision and accuracy coefficients are

$$\rho_{c,\text{inter}}|i, j, j' = \frac{\sigma_\alpha^2}{\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}ijj'}\boldsymbol{\vartheta} + \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2/m_{ijj'}}, \quad (3.9)$$

$$\rho_{\text{inter}}|i, j, j' = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2/m_{ijj'}}, \quad (3.10)$$

$$\chi_{\alpha,\text{inter}}|i, j, j' = \frac{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2/m_{ijj'}}{\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}ijj'}\boldsymbol{\vartheta} + \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2/m_{ijj'}}. \quad (3.11)$$

3.2.3 Total-Rater Agreement

Total-rater agreement is used for measuring the agreement based on any individual observation among all raters. For given subject i , and any two raters j and j' , $j, j' = 1, 2, \dots, k, j \neq j'$, the total-rater mean square deviation can be defined as follows

$$\begin{aligned}
 \varepsilon_{\text{total}}^2|i, j, j' &= E(y_{ijl} - y_{ij'l'})^2 \\
 &= (\mathbf{x}'_{ij}\boldsymbol{\vartheta} - \mathbf{x}'_{ij'}\boldsymbol{\vartheta})^2 + 2(\sigma_\gamma^2 + \sigma_\epsilon^2) \\
 &= 2(\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}ijj'}\boldsymbol{\vartheta} + \sigma_\gamma^2 + \sigma_\epsilon^2).
 \end{aligned} \tag{3.12}$$

Similarly, the total-rater CCC, precision and accuracy coefficients can be calculated as

$$\rho_{c,\text{total}}|i, j, j' = \frac{\sigma_\alpha^2}{\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}ijj'}\boldsymbol{\vartheta} + \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2}, \tag{3.13}$$

$$\rho_{\text{total}}|i, j, j' = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2}, \tag{3.14}$$

$$\chi_{\alpha,\text{total}}|i, j, j' = \frac{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2}{\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}ijj'}\boldsymbol{\vartheta} + \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2}. \tag{3.15}$$

3.3 Overall Agreement Statistics

The above intra-, inter-, and total-rater agreement statistics are calculated for individual subject and paired raters. They should be able to extended to multi-rater cases when a large number of raters are involved. Notice the intra- and total-rater precision coefficients ρ and all the intra-rater agreement statistics do not depend on the index i and j , and the MSDs are

linear functions of $\mathbf{A}_{\mathbf{x}ijj'}$. Hence, we can directly use their sample means to estimate the overall agreements for multiple subjects and raters.

$$\begin{aligned}\varepsilon_{\text{intra}}^2 &= \frac{\sum_{i=1}^n \sum_{j=1}^k \varepsilon_{\text{intra}}^2 |i, j|}{nk} \\ &= 2\sigma_\epsilon^2,\end{aligned}\tag{3.16}$$

$$\begin{aligned}\varepsilon_{\text{inter}}^2 &= \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \varepsilon_{\text{inter}}^2 |i, j, j'|}{nk(k-1)/2} \\ &= 2\left(\boldsymbol{\vartheta}' \mathbf{A}_{\mathbf{x}} \boldsymbol{\vartheta} + \sigma_\gamma^2 + \frac{\sigma_\epsilon^2}{m}\right),\end{aligned}\tag{3.17}$$

$$\begin{aligned}\varepsilon_{\text{total}}^2 &= \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \varepsilon_{\text{total}}^2 |i, j, j'|}{nk(k-1)/2} \\ &= 2\left(\boldsymbol{\vartheta}' \mathbf{A}_{\mathbf{x}} \boldsymbol{\vartheta} + \sigma_\gamma^2 + \sigma_\epsilon^2\right),\end{aligned}\tag{3.18}$$

$$\rho_{c,\text{intra}} = \frac{\sigma_\alpha^2 + \sigma_\gamma^2}{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2},\tag{3.19}$$

$$\rho_{\text{intra}} = \frac{\sigma_\alpha^2 + \sigma_\gamma^2}{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2},\tag{3.20}$$

$$\chi_{\alpha,\text{intra}} = 1,\tag{3.21}$$

$$\rho_{\text{total}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2},\tag{3.22}$$

where $\mathbf{A}_{\mathbf{x}} = \sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k (\mathbf{x}_{ij} - \mathbf{x}_{ij'}) (\mathbf{x}_{ij} - \mathbf{x}_{ij'})' / nk(k-1)$, and m is the harmonic mean of all $m_{ijj'}$'s.

The other agreement statistics depend on the index i and j . King and Chinchilli (27) and Barnhart, Haber, and Song (15) proposed the multi-rater CCC in two different ways. King and Chinchilli (27) defined the generalized concordance correlation coefficient using a convex function of distance, while Barnhart, Haber, and Song (15) defined the overall concordance correlation coefficient for multiple raters as following,

$$\text{Overall CCC} = 1 - \frac{E \sum_{i=1}^{k-1} \sum_{j=i+1}^k (Y_i - Y_j)^2}{E \left[\sum_{i=1}^{k-1} \sum_{j=i+1}^k (Y_i - Y_j)^2 | Y_1, \dots, Y_k \text{ are uncorrelated} \right]} \quad (3.23)$$

These two methods turn out to be identical when a squared distance function is used in King and Chinchilli's generalized concordance correlation coefficient. It can be shown that the Barnhart et al.'s overall CCC (3.23) is actually a weighted mean of pairwise CCCs, and the weights are designed to penalize the pairs of observations which have higher variances and larger mean differences (15). Following the weighting scheme by Barnhart et al., the other overall agreement statistics are defined as follows,

$$\rho_{c,\text{inter}} = \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \rho_{c,\text{inter}}|i, j, j' \times \xi_{\text{inter}}|i, j, j'}{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \xi_{\text{inter}}|i, j, j'}, \quad (3.24)$$

$$\rho_{c,\text{total}} = \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \rho_{c,\text{total}}|i, j, j' \times \xi_{\text{total}}|i, j, j'}{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \xi_{\text{total}}|i, j, j'}, \quad (3.25)$$

$$\rho_{\text{inter}} = \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \rho_{\text{inter}}|i, j, j' \times \xi_{\text{inter}}|i, j, j'}{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \xi_{\text{inter}}|i, j, j'}, \quad (3.26)$$

$$\chi_{a,\text{inter}} = \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \chi_{a,\text{inter}}|i, j, j' \times \xi_{\text{inter}}|i, j, j'}{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \xi_{\text{inter}}|i, j, j'}, \quad (3.27)$$

$$\chi_{a,\text{total}} = \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \chi_{a,\text{total}}|i, j, j' \times \xi_{\text{total}}|i, j, j'}{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \xi_{\text{total}}|i, j, j'}. \quad (3.28)$$

Where the weight ξ for a pair of observations X and Y is defined as $E(X - Y)^2|_{\rho=0}$. Thus, based on the definition of inter-, and total-rater agreements, the weights $\xi_{\text{inter}}|i, j, j'$ and $\xi_{\text{total}}|i, j, j'$ can be computed as follows.

$$\xi_{\text{inter}}|i, j, j' = 2(\boldsymbol{\vartheta}' \mathbf{A}_{\mathbf{x}ijj'} \boldsymbol{\vartheta} + \sigma_{\alpha}^2 + \sigma_{\gamma}^2 + \sigma_{\epsilon}^2/m_{ijj'}),$$

$$\xi_{\text{total}}|i, j, j' = 2(\boldsymbol{\vartheta}' \mathbf{A}_{\mathbf{x}ijj'} \boldsymbol{\vartheta} + \sigma_{\alpha}^2 + \sigma_{\gamma}^2 + \sigma_{\epsilon}^2).$$

After simple algebra, we have

$$\rho_{c,\text{inter}} = \frac{\sigma_\alpha^2}{\boldsymbol{\vartheta}' \mathbf{A}_x \boldsymbol{\vartheta} + \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2/m}, \quad (3.29)$$

$$\rho_{c,\text{total}} = \frac{\sigma_\alpha^2}{\boldsymbol{\vartheta}' \mathbf{A}_x \boldsymbol{\vartheta} + \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2}, \quad (3.30)$$

$$\begin{aligned} \rho_{\text{inter}} &= \frac{\sigma_\alpha^2}{\boldsymbol{\vartheta}' \mathbf{A}_x \boldsymbol{\vartheta} + \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2/m} \\ &+ \frac{2\sigma_\alpha^2}{nk(k-1)} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \boldsymbol{\vartheta}' \mathbf{A}_{xijj'} \boldsymbol{\vartheta} / (\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2/m_{ijj'})}{\boldsymbol{\vartheta}' \mathbf{A}_x \boldsymbol{\vartheta} + \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2/m}, \end{aligned} \quad (3.31)$$

$$\chi_{a,\text{inter}} = \frac{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2/m}{\boldsymbol{\vartheta}' \mathbf{A}_x \boldsymbol{\vartheta} + \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2/m}, \quad (3.32)$$

$$\chi_{a,\text{total}} = \frac{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2}{\boldsymbol{\vartheta}' \mathbf{A}_x \boldsymbol{\vartheta} + \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2}. \quad (3.33)$$

3.4 Estimations

Model (3.2) can be rewritten in the vector form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\boldsymbol{\alpha} + \mathbf{Z}_2\boldsymbol{\gamma} + \mathbf{Z}_0\boldsymbol{\epsilon}, \quad (3.34)$$

with

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} \gamma_{11} \\ \gamma_{12} \\ \vdots \\ \gamma_{1k} \\ \gamma_{21} \\ \vdots \\ \gamma_{2k} \\ \vdots \\ \gamma_{i1} \\ \vdots \\ \gamma_{nk} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \boldsymbol{\vartheta} \end{pmatrix},$$

where \mathbf{X} and \mathbf{Z}_r , $r = 0, 1, 2$, are the corresponding design matrices, \mathbf{y} and $\boldsymbol{\epsilon}$ vectors contain elements of observations and residuals of $i = 1, 2, \dots, n$, $j = 1, 2, \dots, k$, and $l = 1, 2, \dots, m_{ij}$, respectively.

The variance of \mathbf{y} is

$$\mathbf{V} = \mathbf{Z}_1 \mathbf{Z}_1' \sigma_\alpha^2 + \mathbf{Z}_2 \mathbf{Z}_2' \sigma_\gamma^2 + \mathbf{Z}_0 \mathbf{Z}_0' \sigma_\epsilon^2. \quad (3.35)$$

For example, if the data is balanced, we have $m = m_{ijj'} = m_{ij}, i = 1, \dots, n, j, j' = 1, \dots, k$, and

$$\begin{aligned} \mathbf{X} &= \tilde{\mathbf{X}} \otimes \mathbf{1}_m, & \tilde{\mathbf{X}} &= \begin{pmatrix} 1 & \mathbf{x}'_{11} \\ \vdots & \vdots \\ 1 & \mathbf{x}'_{ij} \\ \vdots & \vdots \\ 1 & \mathbf{x}'_{nk} \end{pmatrix}, \\ \mathbf{Z}_0 &= \mathbf{I}_{n \times k \times m}, & \mathbf{Z}_0 \mathbf{Z}_0' &= \mathbf{I}_{n \times k \times m}, \\ \mathbf{Z}_1 &= \mathbf{I}_n \otimes \mathbf{1}_k \otimes \mathbf{1}_m, & \mathbf{Z}_1 \mathbf{Z}_1' &= \mathbf{I}_n \otimes \mathbf{J}_k \otimes \mathbf{J}_m, \\ \mathbf{Z}_2 &= \mathbf{I}_n \otimes \mathbf{I}_k \otimes \mathbf{1}_m, & \mathbf{Z}_2 \mathbf{Z}_2' &= \mathbf{I}_n \otimes \mathbf{I}_k \otimes \mathbf{J}_m. \end{aligned}$$

where \mathbf{I}_{dim} is the identity matrix with dimension $dim \times dim$, $\mathbf{1}_{dim}$ is the vector of all ones with dimension $dim \times 1$, and \mathbf{J}_{dim} is the square matrix of all ones with dimension $dim \times dim$.

3.4.1 Estimating Fixed Effects

The weighted least square (WLS) estimator is used to estimate the fixed effects β , which is identical to the maximum likelihood estimates (MLE) when the errors are assumed to be independently normally distributed.

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \quad (3.36)$$

And its asymptotic variance matrix is,

$$\mathbf{V}_{\hat{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}. \quad (3.37)$$

Note that the overall mean μ and fixed rater effects β_1, \dots, β_k in Model (3.1) can not be unbiased estimated individually, we can assume $\beta_1 = 0$, or add the constraint $\sum_{i=1}^k \beta_i = 0$.

3.4.2 Estimating Variance Components

Assume all the random effects are independently normally distributed, it can be shown that,

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}). \quad (3.38)$$

Restricted maximum likelihood (REML) approach is used to find the estimators of the variances components.

Let $N = \sum_{i=1}^n \sum_{j=1}^k m_{ij}$, and \mathbf{K} be an $N \times (N - p + 1)$ matrix with

$$\text{rank}(\mathbf{K}) = N - p + 1, \quad \mathbf{K}'\mathbf{X} = 0.$$

From Searle, Casella, and McCulloch (22), the REML estimate equations for variance components are,

$$\begin{pmatrix} \text{tr}(\mathbf{Z}'_0 \mathbf{P} \mathbf{Z}_0 \mathbf{Z}'_0 \mathbf{P} \mathbf{Z}_0) & \text{tr}(\mathbf{Z}'_0 \mathbf{P} \mathbf{Z}_1 \mathbf{Z}'_1 \mathbf{P} \mathbf{Z}_0) & \text{tr}(\mathbf{Z}'_0 \mathbf{P} \mathbf{Z}_2 \mathbf{Z}'_2 \mathbf{P} \mathbf{Z}_0) \\ \text{tr}(\mathbf{Z}'_1 \mathbf{P} \mathbf{Z}_0 \mathbf{Z}'_0 \mathbf{P} \mathbf{Z}_1) & \text{tr}(\mathbf{Z}'_1 \mathbf{P} \mathbf{Z}_1 \mathbf{Z}'_1 \mathbf{P} \mathbf{Z}_1) & \text{tr}(\mathbf{Z}'_1 \mathbf{P} \mathbf{Z}_2 \mathbf{Z}'_2 \mathbf{P} \mathbf{Z}_1) \\ \text{tr}(\mathbf{Z}'_2 \mathbf{P} \mathbf{Z}_0 \mathbf{Z}'_0 \mathbf{P} \mathbf{Z}_2) & \text{tr}(\mathbf{Z}'_2 \mathbf{P} \mathbf{Z}_1 \mathbf{Z}'_1 \mathbf{P} \mathbf{Z}_2) & \text{tr}(\mathbf{Z}'_2 \mathbf{P} \mathbf{Z}_2 \mathbf{Z}'_2 \mathbf{P} \mathbf{Z}_2) \end{pmatrix} \begin{pmatrix} \sigma_\epsilon^2 \\ \sigma_\alpha^2 \\ \sigma_\gamma^2 \end{pmatrix} = \begin{pmatrix} \mathbf{y}' \mathbf{P} \mathbf{Z}_0 \mathbf{Z}'_0 \mathbf{P} \mathbf{y} \\ \mathbf{y}' \mathbf{P} \mathbf{Z}_1 \mathbf{Z}'_1 \mathbf{P} \mathbf{y} \\ \mathbf{y}' \mathbf{P} \mathbf{Z}_2 \mathbf{Z}'_2 \mathbf{P} \mathbf{y} \end{pmatrix} \quad (3.39)$$

where

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} = \mathbf{K} (\mathbf{K}' \mathbf{V} \mathbf{K})^{-1} \mathbf{K}'.$$

Patterson and Thompson (28) showed that for the balanced data of mixed model, with or without assuming normality, the ANOVA estimates for variance components are the same as the solutions of REML estimating equations (3.39). For unbalanced data, maximum likelihood approach is preferred due to its well established asymptotic sampling dispersion matrix of the estimates. Therefore, we use the REML estimating equations (3.39) for both normal and non-normal data.

3.5 Asymptotic Variances of Variance Components

3.5.1 With Normality

The estimated variance of variance components with normal assumption can be calculated from the second derivatives of the log likelihood function (22), let $\hat{\boldsymbol{\varsigma}} = (\hat{\sigma}_\epsilon^2, \hat{\sigma}_\alpha^2, \hat{\sigma}_\gamma^2)'$,

$$\mathbf{V}_{\hat{\boldsymbol{\varsigma}}} = \text{var} \begin{pmatrix} \hat{\sigma}_\epsilon^2 \\ \hat{\sigma}_\alpha^2 \\ \hat{\sigma}_\gamma^2 \end{pmatrix} = 2 \begin{pmatrix} \text{tr}(\mathbf{P}\mathbf{Z}_0\mathbf{Z}_0'\mathbf{P}\mathbf{Z}_0\mathbf{Z}_0') & \text{tr}(\mathbf{P}\mathbf{Z}_0\mathbf{Z}_0'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1') & \text{tr}(\mathbf{P}\mathbf{Z}_0\mathbf{Z}_0'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2') \\ \text{tr}(\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_0\mathbf{Z}_0') & \text{tr}(\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1') & \text{tr}(\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2') \\ \text{tr}(\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_0\mathbf{Z}_0') & \text{tr}(\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1') & \text{tr}(\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2') \end{pmatrix}^{-1}. \quad (3.40)$$

3.5.2 Without Normality

Richardson and Welsh (23) discussed the asymptotic behavior of the REML estimates without normality assumption. They were using hierarchical linear mixed models with certain mild conditions, which are easy to be verified in our model. Their theorem is restated here using our own notations.

Theorem 1. (*Richardson and Welsh (23)*)

For hierarchical models, observations \mathbf{y} can be partitioned into g vectors \mathbf{y}_i so that \mathbf{V} is block diagonal with g blocks \mathbf{V}_j . Each $\mathbf{Z}_i\mathbf{Z}_i'$ is also block diagonal, we denote the j th diagonal block as $[\mathbf{Z}_i\mathbf{Z}_i']_{jj}$.

Assume some mild conditions hold, in order to ensure continuity, there is a solution $\hat{\boldsymbol{\varsigma}} = (\hat{\sigma}_\epsilon^2, \hat{\sigma}_\alpha^2, \hat{\sigma}_\gamma^2)'$ of the estimating equations (3.39) satisfying $|\hat{\boldsymbol{\varsigma}} - \boldsymbol{\varsigma}| = O_p(N^{-1/2})$ as $N \rightarrow \infty$.

Moreover,

$$N^{1/2}(\hat{\boldsymbol{\varsigma}} - \boldsymbol{\varsigma}) \xrightarrow{D} \mathcal{N}(0, \mathbf{G}^{-1} \mathbf{F} \mathbf{G}^{-1}), \quad (3.41)$$

where the elements of \mathbf{G} and \mathbf{F} can be estimated by,

$$\begin{aligned} [\hat{\mathbf{G}}]_{ik} &= (2N)^{-1} \sum_{j=1}^g \text{tr}(\hat{\mathbf{V}}_j^{-1} [\mathbf{Z}_{i-1} \mathbf{Z}'_{i-1}]_j \hat{\mathbf{V}}_j^{-1} [\mathbf{Z}_{k-1} \mathbf{Z}'_{k-1}]_j), \\ [\hat{\mathbf{F}}]_{ik} &= (4N)^{-1} \sum_{j=1}^g \left\{ \left[(\mathbf{y}_j - \mathbf{X}_j \hat{\boldsymbol{\beta}})' \hat{\mathbf{V}}_j^{-1} [\mathbf{Z}_{i-1} \mathbf{Z}'_{i-1}]_j \hat{\mathbf{V}}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \hat{\boldsymbol{\beta}}) (\mathbf{y}_j - \mathbf{X}_j \hat{\boldsymbol{\beta}})' \hat{\mathbf{V}}_j^{-1} \right. \right. \\ &\quad \left. \left. \times [\mathbf{Z}_{k-1} \mathbf{Z}'_{k-1}]_j \hat{\mathbf{V}}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \hat{\boldsymbol{\beta}}) \right] - \text{tr}(\hat{\mathbf{V}}_j^{-1} [\mathbf{Z}_{i-1} \mathbf{Z}'_{i-1}]_j) \text{tr}(\hat{\mathbf{V}}_j^{-1} [\mathbf{Z}_{k-1} \mathbf{Z}'_{k-1}]_j) \right\}. \end{aligned}$$

3.6 Asymptotic Properties of Agreement Statistics

Write $\hat{\boldsymbol{\tau}} = (\hat{\boldsymbol{\vartheta}}', \hat{\boldsymbol{\varsigma}}')'$, it is well known that the REML estimates take into account the degrees of freedom associated with the fixed effects, they are invariant to the fixed effects ((17), (22)).

Let

$$\mathbf{V}_{\hat{\boldsymbol{\tau}}} = N \cdot \left(\begin{array}{c|c} \mathbf{V}_{\hat{\boldsymbol{\vartheta}}} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{V}_{\hat{\boldsymbol{\varsigma}}} \end{array} \right), \quad (3.42)$$

where $\mathbf{V}_{\hat{\boldsymbol{\vartheta}}}$ is the estimating variance for $\hat{\boldsymbol{\vartheta}}$ from (3.37), and $\mathbf{V}_{\hat{\boldsymbol{\varsigma}}}$ can be calculated either from (3.40) or (3.41), for normal or non-normal cases respectively.

From Richardson and Welsh (23), we know that

$$N^{1/2}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \xrightarrow{D} \mathcal{N}(0, \mathbf{V}_{\boldsymbol{\tau}}), \quad (3.43)$$

Furthermore, all the agreement statistics are functions of $\boldsymbol{\tau}$, say $h(\boldsymbol{\tau})$, where $h(\cdot)$ can be any expressions of MSD, CCC, precision and accuracy coefficients. Hence, delta method (29) is used to obtain the asymptotic normality for all the agreement statistics.

$$N^{1/2}(h(\hat{\boldsymbol{\tau}}) - h(\boldsymbol{\tau})) \xrightarrow{D} \mathcal{N}(0, \nabla h(\boldsymbol{\tau})' \mathbf{V}_{\boldsymbol{\tau}} \nabla h(\boldsymbol{\tau})). \quad (3.44)$$

For example, the asymptotic variance for inter-rater CCC estimate is

$$\text{var}(\hat{\rho}_{c,\text{inter}}) = \begin{pmatrix} \frac{-2\mathbf{A}_{\mathbf{x}}\boldsymbol{\vartheta}}{(\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}}\boldsymbol{\vartheta} + \sigma_{\alpha}^2 + \sigma_{\gamma}^2 + \sigma_{\epsilon}^2/m)^2} \\ \frac{-1/m}{(\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}}\boldsymbol{\vartheta} + \sigma_{\alpha}^2 + \sigma_{\gamma}^2 + \sigma_{\epsilon}^2/m)^2} \\ \frac{\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}}\boldsymbol{\vartheta} + \sigma_{\gamma}^2 + \sigma_{\epsilon}^2/m}{(\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}}\boldsymbol{\vartheta} + \sigma_{\alpha}^2 + \sigma_{\gamma}^2 + \sigma_{\epsilon}^2/m)^2} \\ \frac{-1}{(\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}}\boldsymbol{\vartheta} + \sigma_{\alpha}^2 + \sigma_{\gamma}^2 + \sigma_{\epsilon}^2/m)^2} \end{pmatrix}^T \times \frac{\mathbf{V}_{\boldsymbol{\tau}}}{N} \times \begin{pmatrix} \frac{-2\mathbf{A}_{\mathbf{x}}\boldsymbol{\vartheta}}{(\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}}\boldsymbol{\vartheta} + \sigma_{\alpha}^2 + \sigma_{\gamma}^2 + \sigma_{\epsilon}^2/m)^2} \\ \frac{-1/m}{(\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}}\boldsymbol{\vartheta} + \sigma_{\alpha}^2 + \sigma_{\gamma}^2 + \sigma_{\epsilon}^2/m)^2} \\ \frac{\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}}\boldsymbol{\vartheta} + \sigma_{\gamma}^2 + \sigma_{\epsilon}^2/m}{(\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}}\boldsymbol{\vartheta} + \sigma_{\alpha}^2 + \sigma_{\gamma}^2 + \sigma_{\epsilon}^2/m)^2} \\ \frac{-1}{(\boldsymbol{\vartheta}'\mathbf{A}_{\mathbf{x}}\boldsymbol{\vartheta} + \sigma_{\alpha}^2 + \sigma_{\gamma}^2 + \sigma_{\epsilon}^2/m)^2} \end{pmatrix}.$$

Other variances can be calculated similarly. And all these asymptotic variances can be estimated by replacing $\boldsymbol{\tau}$ and $\mathbf{V}_{\boldsymbol{\tau}}$ by their sample estimators from REML.

It is well known that the Fisher's Z -transformation can markedly improve the normal approximation for the precision coefficient. Lin (7), Robieson (10), and Lin, Hedayat, Sinha, and Yang (11) showed that the asymptotic normality of MSD, CCC, and accuracy coefficient can be enhanced by log, Z , and logit transformation respectively. Hence, the 95% confidence intervals

for estimated agreement statistics are recommended to be calculated based on the transformed values, and corresponding inverse transformations, such as exponential, inverse Z or logistic functions, are used to transform confidence intervals back to the original scale.

CHAPTER 4

SIMULATIONS AND CASE STUDIES FOR LINEAR MIXED MODELS

4.1 Simulations

Simulation studies were conducted for both normal and non-normal data in several situations to demonstrate the performance of our approach.

4.1.1 Normal Cases

4.1.1.1 Without Covariates

We began with the simplest model, balanced normally distributed data without covariates. The data were generated using the following equation

$$y_{ijl} = \mu + \beta_j + \alpha_i + \gamma_{ij} + \epsilon_{ijl}, \quad i = 1, 2, \dots, 50, \quad j = 1, 2, 3, \quad l = 1, 2, \quad (4.1)$$

where $\mu = 10$, $\beta_1 = 0$, $\beta_2 = -1$, $\beta_3 = -2$, α_i , γ_{ij} and ϵ_{ijl} were independently normally distributed with mean zeros and variances $\sigma_\alpha^2 = 1$, $\sigma_\gamma^2 = 0.5^2$ and $\sigma_\epsilon^2 = 0.1^2$.

Table I presents the simulation results for all intra-, inter- and total-rater agreement statistics in 1000 runs. The column **True** gives the theoretical values calculated from the simulation parameters, **Mean(Est)** gives the means of all the estimates, **Std(Est)** gives standard deviations of all these estimates in 1000 runs, **Mean(Std)** gives the means of the estimated standard deviations for each run using delta method, and the last two columns **SIG(%)** and **SIG*(%)** gives the proportions of estimates whose 95% confidence interval did not contain the mean of estimates, with and without normalizing transformations, respectively.

The intra-rater precision and accuracy coefficients are not presented in the table, because there was no location shift within raters, the accuracy coefficient was always zero and the

precision coefficient equaled the intra-rater CCC. TDI and CP are also not reported, because they can be calculated directly from MSD using (2.5) and (2.6).

From Table I, we can see that the results were satisfactory, all the estimates were nearly equal to the true values. The relative differences between the true values and the means of estimates were less than 1%. In addition, all the $\text{Std}(\text{Est})$ were very close to $\text{Mean}(\text{Est})$, and the $\text{SIG}^*(\%)$ were all around 5%, which justifies the advantages of those normalizing transformations.

4.1.1.2 With Covariates and 10% Missing Values

For the covariates case, we used the following model to generate the data.

$$y_{ijl} = \mu + x_{ij} + \beta_j + \alpha_i + \gamma_{ij} + \epsilon_{ijl}, \quad i = 1, 2, \dots, 50, \quad j = 1, 2, 3, \quad l = 1, 2, \quad (4.2)$$

where $x_{ij} = \ln(i \times j)$, and the other parameters were the same as those we used in (4.2).

Because the method of REML can handle missing values by simply removing the rows in (3.34), 10% randomly missing values were added into this simulation.

After removing the missing values, the data became unbalanced, the replicates m_{ij} could be 0, 1, or 2. We removed the subject i if $m_{ij} = 0$, because it indicated there was no readings for subject i on rater j . The other subjects with replicates 1 and 2 were kept in the data.

The covariate effect was removed by assuming $\theta = 0$ in the ϑ of (3.17), (3.18), (3.29), (3.30), (3.31), (3.32), and (3.33). So the true value of all agreement statistics should be the same as those in our previous simulation without covariates.

Table II presents the simulation results in 1000 runs. Comparing Table I and Table II, we can tell that the results with covariates and missing values were also very closed to the true values, but had larger variances, and the asymptotic normalities were truly improved by transformations.

4.1.2 Non-Normal Cases

4.1.2.1 Discrete and Skewed Cases

Binomial distribution was used for the discrete case, data were generated using the following model

$$y_{ijl} = \mu + \beta_j + \alpha_i + \gamma_{ij} + \epsilon_{ijl}, \quad i = 1, 2, \dots, 50, \quad j = 1, 2, \quad l = 1, 2, \quad (4.3)$$

where $\mu = 10$, $\beta_1 = 0$, $\beta_2 = -1$, α_i , γ_{ij} and ϵ_{ijl} were independently binomially distributed,

$$\alpha_i \sim \text{Binom}(n = 10, p = 0.5), \quad \gamma_{ij} \sim \text{Binom}(n = 3, p = 0.2), \quad \epsilon_{ijl} \sim \text{Binom}(n = 1, p = 0.5).$$

Gamma distribution with a small shape parameter was used for the skewed case. The data generating model was

$$y_{ijl} = \mu + \beta_j + \alpha_i + \gamma_{ij} + \epsilon_{ijl}, \quad i = 1, 2, \dots, 50, \quad j = 1, 2, \quad l = 1, 2, \quad (4.4)$$

where $\mu = 10$, $\beta_1 = 0$, $\beta_2 = -1$, ϵ_{ijl} was normally distributed with mean zero and variance $\sigma_\epsilon^2 = 0.5^2$, α_i and γ_{ij} were independently gamma distributed,

$$\alpha_i \sim \text{Gamma}(\text{shape} = 2, \text{scale} = 2/\sqrt{2}), \quad \gamma_{ij} \sim \text{Gamma}(\text{shape} = 2, \text{scale} = 1/\sqrt{2}).$$

Table III and Table IV present the simulation results for binomial and Gamma data in 1000 runs. Similar to the previous normal cases, all the estimates were very accurate. The SIGs were a little bit away from 5%, due to the discreteness or skewness. However, the results in $\text{SIG}^*(\%)$ demonstrate that all the results were approximately normally distributed after normalizing transformations. Furthermore, simulations with sample size $n = 200$, which are not reported here, confirmed the asymptotic normalities for all the agreement estimates.

Table V and Table VI present the simulation results for binomial and Gamma data in 1000 runs, with a covariate $x_{ij} = \ln(i \times j)$ and 10% missing values. Similar to the normal case, the covariate effect was removed by assuming $\boldsymbol{\theta} = 0$. Only fixed effect β_j , and random effects α_i , γ_{ij} and ϵ_{ijl} were considered for computing the agreement estimates. The results were very close to the ones without covariates and missing values, but the estimated standard deviations were a little bit larger, due to the existence of covariate and missing values.

4.1.2.2 Non-Linear Case

All the data in previous cases were simulated based on the linear additive model (3.1). One may question the performance of our approach on the non-linear data.

For the illustration purpose, correlated negative binomial data were generated using Gamma-Poisson mixture with parameters $r = 10$ and $p = 0.91$. The algorithm of Gamma-

Poisson random number generator is described in the appendix. The dimensions were set as $n = 50$, $k = 2$, $m = 2$, and Pearson correlation coefficient matrix for the four data series was

$$\begin{pmatrix} 1 & 0.9 & 0.8 & 0.8 \\ 0.9 & 1 & 0.8 & 0.8 \\ 0.8 & 0.8 & 1 & 0.9 \\ 0.8 & 0.8 & 0.9 & 1 \end{pmatrix}.$$

After generating the multivariate negative binomial series, an additional inter-rater effect $\beta_2 = 20$ was added to the series of the second rater, in order to make a between rater location shift.

Table VII presents the simulation results in 1000 runs, and Table VIII presents the simulation results in 1000 runs, with a covariate $x_{ij} = \ln(i \times j)$ and 10% missing values. Notice that we could not compute all the true values since it was not a linear additive model, the random effects varied in each simulation. But we did know that the intra-rater CCC and precision were 0.9 from the Pearson correlation coefficient. The other values in the **True** column were computed using a large sample simulation ($n \sim 10,000$).

Table VII and Table VIII justify that even when the true model is non-linear and unclear, our method still can provide desirable results.

TABLE I: Simulation Results for Normal Data without Covariates and Missing Values

		True	Mean(Est)	Std(Est)	Mean(Std)	SIG(%)	SIG*(%)
Intra	MSD	0.020	0.020	0.0023	0.0023	6.5	5.9
	CCC	0.992	0.992	0.0018	0.0016	8.2	6.2
Inter	MSD	2.510	2.518	0.212	0.214	4.5	4.8
	CCC	0.443	0.440	0.059	0.058	5.2	5.0
	Precision	0.797	0.792	0.045	0.045	6.1	5.3
	Accuracy	0.557	0.552	0.051	0.049	6.9	5.5
Total	MSD	2.520	2.527	0.212	0.214	4.6	4.9
	CCC	0.442	0.438	0.059	0.058	5.2	5.0
	Precision	0.794	0.791	0.045	0.046	5.9	5.2
	Accuracy	0.558	0.553	0.050	0.047	6.9	5.5

TABLE II: Simulation Results for Normal Data with Covariates and 10% Missing Values

		True	Mean(Est)	Std(Est)	Mean(Std)	SIG(%)	SIG*(%)
Intra	MSD	0.020	0.020	0.0026	0.0026	6.2	4.5
	CCC	0.992	0.992	0.0019	0.0017	8.3	5.4
Inter	MSD	2.510	2.523	0.439	0.430	6.3	5.3
	CCC	0.443	0.440	0.069	0.070	5.7	5.2
	Precision	0.797	0.803	0.059	0.057	6.9	5.3
	Accuracy	0.557	0.555	0.067	0.063	8.2	5.6
Total	MSD	2.520	2.532	0.439	0.430	6.3	5.3
	CCC	0.442	0.439	0.068	0.070	5.8	5.2
	Precision	0.794	0.787	0.046	0.047	6.6	5.0
	Accuracy	0.558	0.556	0.067	0.063	8.2	5.7

TABLE III: Simulation Results for Binomial Data

		True	Mean(Est)	Std(Est)	Mean(Std)	SIG(%)	SIG*(%)
Intra	MSD	0.500	0.500	0.050	0.057	2.7	3.2
	CCC	0.923	0.920	0.016	0.018	3.3	3.6
Inter	MSD	2.210	2.232	0.392	0.427	3.5	4.9
	CCC	0.693	0.686	0.063	0.066	6.0	4.8
	Precision	0.805	0.801	0.050	0.052	7.9	6.1
	Accuracy	0.861	0.856	0.043	0.038	8.2	6.9
Total	MSD	2.460	2.482	0.391	0.426	3.7	4.7
	CCC	0.670	0.663	0.063	0.068	5.5	4.3
	Precision	0.774	0.769	0.052	0.055	6.7	4.8
	Accuracy	0.866	0.861	0.042	0.038	8.4	6.4

TABLE IV: Simulation Results for Gamma Data

		True	Mean(Est)	Std(Est)	Mean(Std)	SIG(%)	SIG*(%)
Intra	MSD	0.500	0.496	0.071	0.068	7.5	5.9
	CCC	0.952	0.950	0.013	0.015	5.2	4.6
Inter	MSD	3.250	3.287	0.718	0.680	7.7	6.3
	CCC	0.711	0.701	0.082	0.081	6.9	6.3
	Precision	0.780	0.773	0.074	0.070	7.5	6.5
	Accuracy	0.911	0.905	0.041	0.038	9.8	7.8
Total	MSD	3.500	3.535	0.718	0.681	7.7	6.0
	CCC	0.696	0.691	0.082	0.084	5.7	5.6
	Precision	0.762	0.755	0.074	0.070	7.2	6.6
	Accuracy	0.913	0.907	0.040	0.037	9.3	7.5

TABLE V: Simulation Results for Binomial Data with Covariates and Missing Values

		True	Mean(Est)	Std(Est)	Mean(Std)	SIG(%)	SIG*(%)
Intra	MSD	0.500	0.501	0.065	0.071	4.0	4.3
	CCC	0.923	0.919	0.017	0.018	3.8	4.3
Inter	MSD	2.210	2.320	0.555	0.569	3.9	5.1
	CCC	0.693	0.685	0.070	0.068	7.1	6.9
	Precision	0.805	0.796	0.051	0.054	4.2	5.2
	Accuracy	0.861	0.855	0.052	0.053	4.6	4.5
Total	MSD	2.460	2.484	0.566	0.578	4.0	5.1
	CCC	0.670	0.659	0.069	0.068	6.4	6.2
	Precision	0.774	0.766	0.056	0.057	3.1	4.2
	Accuracy	0.866	0.861	0.050	0.051	5.7	5.2

TABLE VI: Simulation Results for Gamma Data with Covariates and Missing Values

		True	Mean(Est)	Std(Est)	Mean(Std)	SIG(%)	SIG*(%)
Intra	MSD	0.500	0.504	0.089	0.095	3.5	3.9
	CCC	0.952	0.949	0.015	0.015	6.2	4.9
Inter	MSD	3.250	3.304	0.850	0.885	2.9	3.3
	CCC	0.711	0.699	0.091	0.083	8.6	7.6
	Precision	0.780	0.770	0.080	0.076	9.7	9.6
	Accuracy	0.911	0.902	0.054	0.047	7.2	5.7
Total	MSD	3.500	3.550	0.851	0.886	3.9	4.0
	CCC	0.696	0.690	0.091	0.086	6.7	6.0
	Precision	0.762	0.754	0.081	0.078	8.7	6.3
	Accuracy	0.913	0.904	0.053	0.046	9.6	7.5

TABLE VII: Simulation Results for Negative Binomial Data

		True	Mean(Est)	Std(Est)	Mean(Std)	SIG(%)	SIG*(%)
Intra	MSD	224.788	223.807	33.057	33.350	3.1	3.2
	CCC	0.900	0.897	0.025	0.027	3.0	3.3
Inter	MSD	735.476	741.954	127.160	123.183	5.1	5.0
	CCC	0.709	0.701	0.064	0.067	3.3	3.7
	Precision	0.842	0.839	0.046	0.047	3.7	3.8
	Accuracy	0.842	0.834	0.046	0.044	6.3	5.5
Total	MSD	847.870	850.858	128.879	124.753	2.2	2.1
	CCC	0.679	0.671	0.065	0.067	2.6	3.1
	Precision	0.800	0.796	0.050	0.052	3.4	3.7
	Accuracy	0.849	0.842	0.043	0.041	7.4	6.4

TABLE VIII: Simulation Results for Negative Binomial Data with Covariates and Missing Values

		True	Mean(Est)	Std(Est)	Mean(Std)	SIG(%)	SIG*(%)
Intra	MSD	224.788	222.452	37.603	42.017	3.0	3.1
	CCC	0.900	0.897	0.026	0.025	8.0	7.2
Inter	MSD	735.476	742.531	198.931	203.390	6.9	5.2
	CCC	0.709	0.697	0.077	0.069	8.2	5.7
	Precision	0.842	0.834	0.077	0.073	8.7	6.2
	Accuracy	0.842	0.839	0.067	0.063	6.3	6.1
Total	MSD	847.870	856.438	200.106	205.268	6.5	5.0
	CCC	0.679	0.670	0.076	0.068	9.6	7.1
	Precision	0.800	0.792	0.052	0.049	7.7	6.7
	Accuracy	0.849	0.844	0.055	0.051	7.2	6.3

4.2 Case Studies

4.2.1 Cardiac Function Measurements

Consider an example from Bowling, Sageman, O'Connor, Cole, and Amundson (30), which was used to determine the limits of agreement between left ventricular ejection fraction estimated by impedance cardiography (IC) and radionuclide ventriculography (RV). Sixty measurements of cardiac ejection fraction by these two methods were made on 12 patients, with unbalanced number of replicates per patient. The cardiac outputs are listed in Table IX.

Figure 2 presents the plots for the means of the outputs by two methods on 12 patients. The vertical bars give the mean ± 1 standard deviation of RV outputs on each patient, and the horizontal bars give the mean ± 1 standard deviation of IC outputs on each patient.

Normalizing transformations were used for computing 95% confidence limits for all the estimates. The allowance for CP was 0.9, and the allowance for TDI was 1. This means that we required 90% of paired data must not deviate more than 1 unit from each other.

Table X presents the agreement statistics estimates, and 95% confidence limits for cardiac outputs, only lower limits for CCC, precision and accuracy coefficients and CP, and upper limits for TDI are reported. Intra-rater accuracy is not reported since it always equals to 1. The intra-rater CCC estimate is 0.932, indicating good agreement within each method, while the inter-rater CCC estimate is 0.642, indicating worse agreement between these two methods. These low inter-rater agreement statistics values reveal that these two methods impedance cardiography and radionuclide ventriculography are not interchangeable.

TABLE IX: Cardiac Outputs by RV and IC for 12 Patients

Patients	Method	Outputs					
1	IC	6.57	5.62	6.90	6.57	6.35	
	RV	7.83	7.42	7.89	7.12	7.88	
2	IC	4.06	4.29	4.26	4.09		
	RV	6.16	7.26	6.71	6.54		
3	IC	4.71	5.50	5.08	5.02	6.01	5.67
	RV	4.75	5.24	4.86	4.78	6.05	5.42
4	IC	4.14	4.20	4.61	4.68	5.04	
	RV	4.21	3.61	3.72	3.87	3.92	
5	IC	3.03	2.86	2.77	2.46	2.32	2.43
	RV	3.13	2.98	2.85	3.17	3.09	3.12
6	IC	5.90	5.81	5.70	5.76		
	RV	5.92	6.42	5.92	6.27		
7	IC	5.09	4.63	4.61	5.09		
	RV	7.13	6.62	6.58	6.93		
8	IC	4.72	4.61	4.36	4.20	4.36	4.20
	RV	4.54	4.81	5.11	5.29	5.39	5.57
9	IC	3.17	3.12	2.96			
	RV	4.48	4.92	3.97			
10	IC	4.35	4.62	3.16	3.53	3.53	
	RV	4.22	4.65	4.74	4.44	4.50	
11	IC	7.20	6.09	7.00	7.10	7.40	6.80
	RV	6.78	6.07	6.52	6.42	6.41	5.76
12	IC	4.50	4.20	3.80	3.80	4.20	4.50
	RV	5.06	4.72	4.90	4.80	4.90	5.10

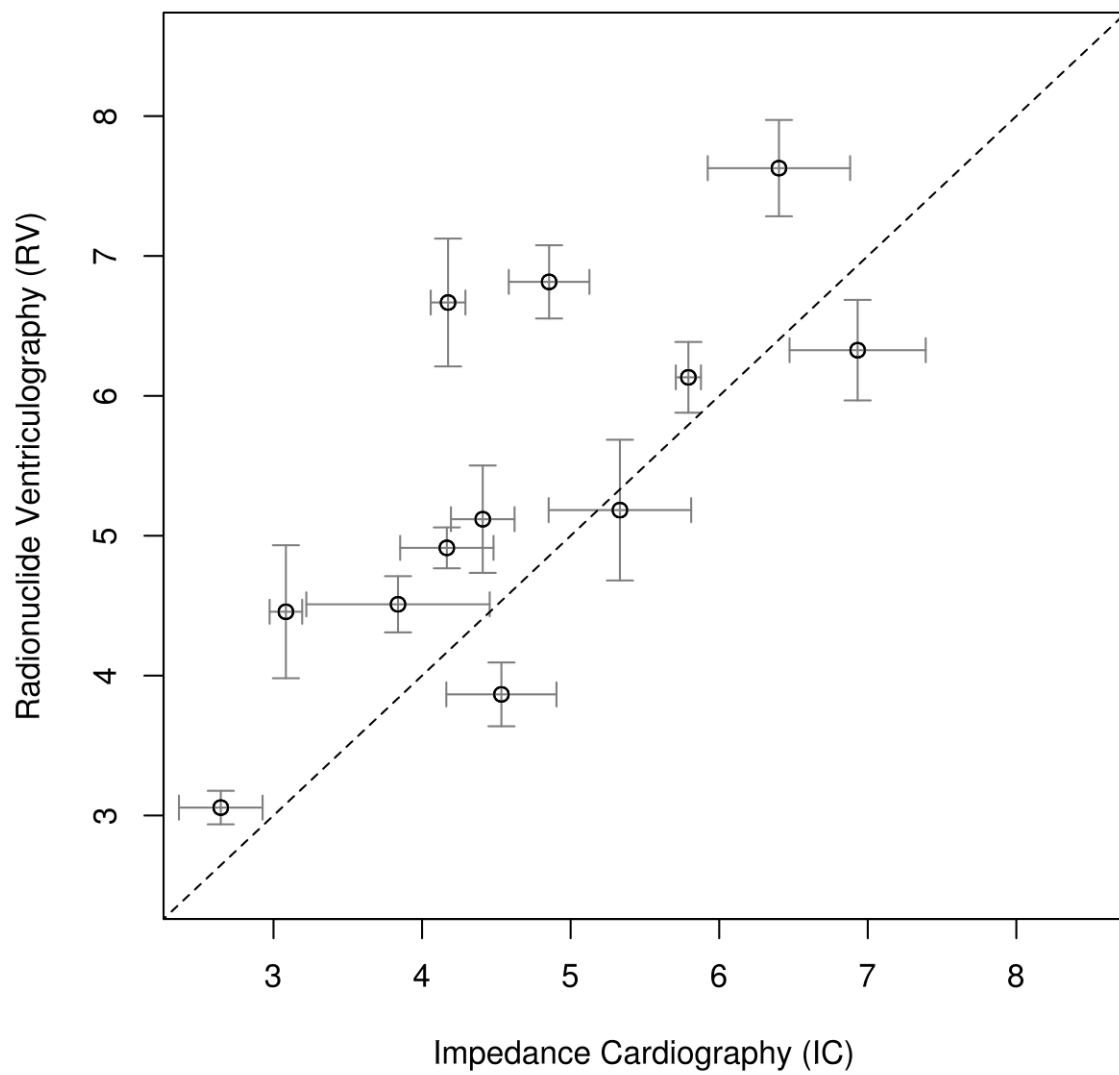


Figure 2: Means of the Cardiac Outputs for each Subject with Error Bars

TABLE X: Estimated Agreement Statistics for Cardiac Outputs

	Statistics	CCC	Precision	Accuracy	TDI _{0.9}	CP ₁
Intra	Estimate	0.932	0.932	.	0.403	0.957
	95% Conf. Limit	0.889	0.889	.	0.568	0.911
Inter	Estimate	0.642	0.716	0.874	2.303	0.602
	95% Conf. Limit	0.363	0.462	0.719	4.422	0.458
Total	Estimate	0.612	0.695	0.880	2.622	0.572
	95% Conf. Limit	0.344	0.429	0.740	4.639	0.448

4.2.2 Antihypertensive Patch Dataset

This antihypertensive patch dataset is an example given by the FDA (<http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Biostatistics/ucm081434.htm>) and had been used in (31) for the population and individual bioequivalence.

We shall study the AUC values from that dataset, which was collected by a four-period, two-sequence crossover trial to a total of 37 subjects. Table XI gives the summary of the crossover design for reference (R) and test (T) formulations. The data were considered to have large subject-by-formulation interaction, and the logarithm of AUC values were assumed to be normally distributed.

TABLE XI: Summary of the Crossover Design

Sequence	Number of Subjects	Period
1	18	TRRT
2	19	RTTR

We assume the following linear mixed effects model,

$$\log(\text{AUC})_{ijl} = \mu + \beta_j + s_i + p_l + \alpha_i + \gamma_{ij} + \epsilon_{ijl}. \quad (4.5)$$

The subscript i in (4.5) is the index of subjects, j is the index of formulations (T/R), l is the index of periods, μ is the overall mean, β_j is the fixed formulation effects, s_i is the fixed sequence effects, which depends on the index of subjects i , and p_l is the fixed period effects. In this case, there are two sequences with four periods TRTR/RTRT, $s_i = s_1, s_2$, and $p_l = p_1, \dots, p_4$. The random subject effect α_i is nested in sequences h with mean 0 and variance σ_α^2 , γ_{ij} is the random subject-formulation interaction effect with mean 0 and variance σ_γ^2 , ϵ_{ijl} is the random error with mean 0 and variance σ_ϵ^2 , and all random effects are assumed to be mutually independent.

Likelihood ratio tests showed that the period effects were not statistically significant. The model (4.5) can be reduced to

$$\log(\text{AUC})_{ijl} = \mu + \beta_j + s_i + \alpha_i + \gamma_{ij} + \epsilon_{ijl}. \quad (4.6)$$

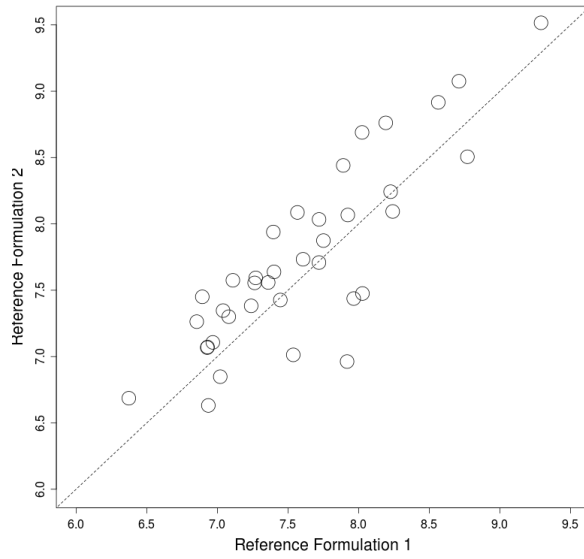
Because the period effects were negligible, we can list the data in the format of two replications for two formulations. Our linear agreement approach was applied to the model (4.6), and the sequence effect s_i was treated as a between subject covariate.

Figure 3 presents the plots for the $\log(\text{AUC})$ between two formulations on 37 subjects. Figure 3a and Figure 3b show the agreement plots within each formulation. We can see that within formulation readings of reference formulation were more precise than those of test formulation. Figure 3c presents the agreement plot of the first test formulation readings and the second reference formulation readings, reflecting total agreement among individual readings. Figure

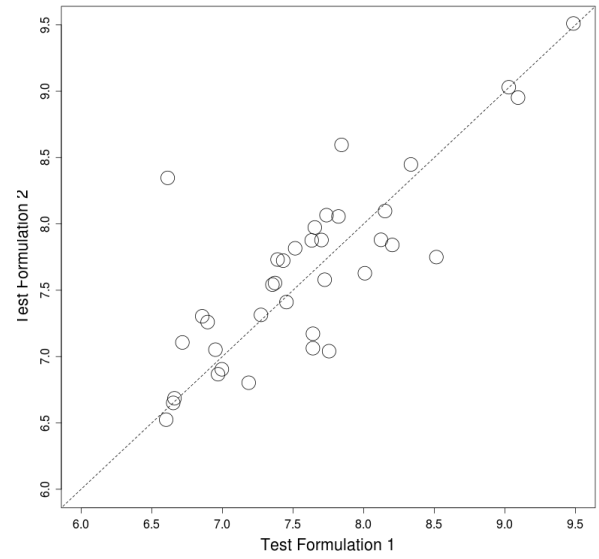
3d presents the agreement plot of the test and reference formulations based on their means, reflecting inter-formulation agreement.

Table XII presents the agreement statistics estimates by removing the fixed sequence effects. The allowance for CP was 0.9, and the allowance for TDI was 1. This means that we required 90% of paired data must not deviate more than 1 unit from each other.

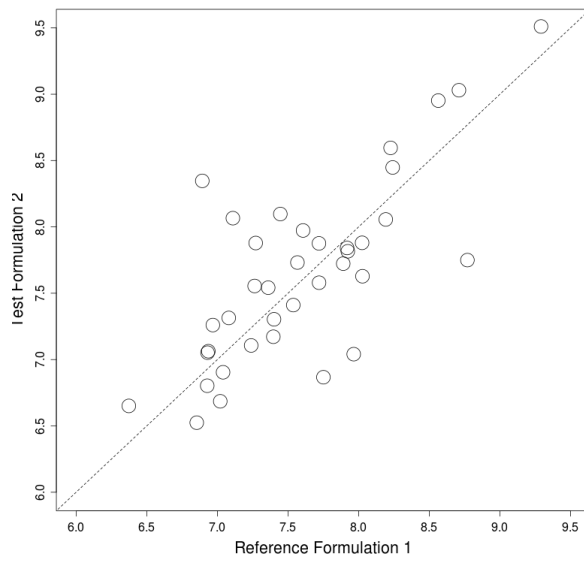
From Table XII, all the accuracy coefficients are very close to 1, which indicate no major location shift for the marginal distributions of reference and test formulations. The 95% upper limit of inter formulation $TDI_{0.9}$ is 0.323, meaning that we are 95% confident that 90% of the mean of the readings from two formulations do not deviate more than 0.323. In contrast, the 95% upper limit of total formulation $TDI_{0.9}$ is 0.470, meaning that we are 95% confident that 90% of the individual readings from two formulations do not deviate more than 0.470, which is larger than the inter formulation $TDI_{0.9}$, as seen in the figures. The inter-rater CCC is 0.841, and its 95% confidence interval is (0.646, 1.000), which covers the perfect agreement case. Therefore, the agreement between the reference and test formulations are satisfactory, they can be considered interchangeable.



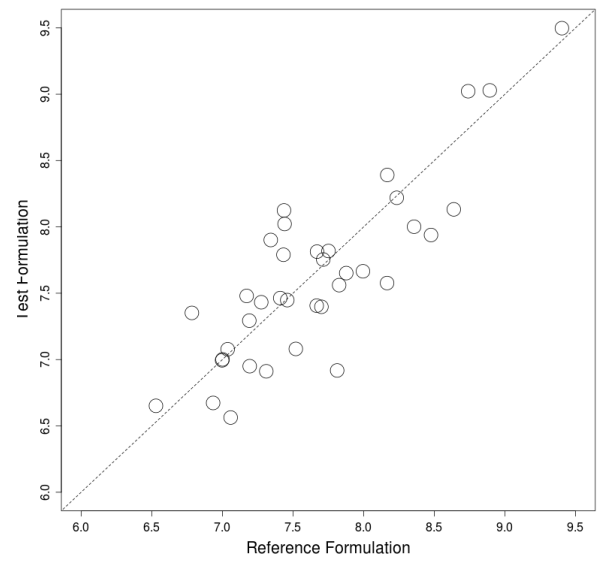
(a) Between the Reference Formulation 1 and 2



(b) Between the Test Formulation 1 and 2



(c) Between the Test 1 and the Reference 2



(d) Between the Test and Reference Means

Figure 3: Agreement Plots for $\log(\text{AUC})$

TABLE XII: Estimated Restricted Agreement Statistics for $\log(\text{AUC})$

	Statistics	CCC	Precision	Accuracy	TDI_{0.9}	CP₁
Intra	Estimate	0.820	0.820	.	0.275	0.986
	95% Conf. Limit	0.676	0.676	.	0.364	0.967
Inter	Estimate	0.841	0.843	0.998	0.221	0.994
	95% Conf. Limit	0.646	0.653	0.970	0.323	0.976
Total	Estimate	0.766	0.767	0.998	0.359	0.968
	95% Conf. Limit	0.577	0.558	0.975	0.470	0.939

CHAPTER 5

AGREEMENT STATISTICS FOR GENERALIZED LINEAR MIXED MODELS

5.1 Generalized Linear Mixed Models (GLMM)

Agreement statistics for linear mixed models are studied in the previous chapters, and we have developed a set of agreement estimates not only for normally distributed data, but also for non-normally distributed data. However, there still exist some cases, which are not able to fit into the linear mixed model framework. For example, log-Gamma distributed data, binary data using logistic regression, over-dispersion Poisson data, zero-inflating negative binomial data, etc. Generalized linear mixed model is a more appropriate approach for these cases, because it allows an arbitrary link function rather than the identical link between the observed mean and linear predictors, in order to provide more flexibility for modeling data.

Consider the simplest case, balanced data without replications, i.e., $m_{ij} = 1$, the generalized linear mixed model can be described as follows.

$$f(y_{ij}|\alpha_i, \gamma_{ij}) = \exp\left(\frac{y_{ij}\tau_{ij} - b(\tau_{ij})}{a(\phi)} + c(y_{ij}, \phi)\right), \quad (5.1)$$

$$E(y_{ij}|\alpha_i, \gamma_{ij}) = \mu_{ij}, \quad (5.2)$$

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\boldsymbol{\alpha} + \mathbf{Z}_2\boldsymbol{\gamma}, \quad (5.3)$$

where

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1k} \\ y_{21} \\ \vdots \\ y_{ij} \\ \vdots \\ y_{nk} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1k} \\ \mu_{21} \\ \vdots \\ \mu_{ij} \\ \vdots \\ \mu_{nk} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} & & & \mathbf{x}'_{11} \\ & & & \vdots \\ & & & \mathbf{x}'_{1k} \\ \mathbf{1}_{n \times k} & \mathbf{I}_k \otimes \mathbf{1}_n & & \vdots \\ & & & \mathbf{x}'_{ij} \\ & & & \vdots \\ & & & \mathbf{x}'_{nk} \end{pmatrix},$$

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} \gamma_{11} \\ \gamma_{12} \\ \vdots \\ \gamma_{1k} \\ \gamma_{21} \\ \vdots \\ \gamma_{ij} \\ \vdots \\ \gamma_{nk} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \\ \boldsymbol{\theta} \end{pmatrix}.$$

In Equation (5.1) – (5.3), τ_{ij} is the canonical parameter, ϕ is the dispersion parameter and $g(\cdot)$ is the link function. Similar to the linear mixed model, the subscript i is the index of subjects, j is the index of raters, $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, k$. In addition, with a little abuse of notations, μ is the overall mean, \mathbf{x}_{ij} is the $p \times 1$ design vector for $p \times 1$ covariate coefficients $\boldsymbol{\theta}$, β_j is the fixed inter-rater effect, α_i is the random subject effect with mean 0 and variance σ_α^2 , γ_{ij} is the random subject-rater interaction effect with mean 0 and variance σ_γ^2 , and $\mathbf{Z}_1 = \mathbf{I}_n \otimes \mathbf{1}_k$, $\mathbf{Z}_2 = \mathbf{I}_n \otimes \mathbf{I}_k$.

It can be easily extended to unbalanced replicated data by adding rows for m_{ij} replications of y_{ij} respectively. Furthermore, for missing values, the model can be derived by removing the rows corresponding to the missing values from the design matrices \mathbf{X} , \mathbf{Z}_1 , and \mathbf{Z}_2 .

It is well known that

$$b'(\tau_{ij}) = E(y_{ij}|\alpha_i, \gamma_{ij}) = \mu_{ij}, \quad (5.4)$$

$$\text{var}(y_{ij}|\alpha_i, \gamma_{ij}) = a(\phi)v(\mu_{ij}), \quad (5.5)$$

where $v(\cdot)$ is the variance function for μ .

Assuming the linear predictor $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\boldsymbol{\alpha} + \mathbf{Z}_2\boldsymbol{\gamma}$, and the inverse link function $h(\cdot) = g^{-1}(\cdot)$, we have

$$E(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\gamma}) = h(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\boldsymbol{\alpha} + \mathbf{Z}_2\boldsymbol{\gamma}). \quad (5.6)$$

There are some developed methods to estimate the fixed and random effects via maximum likelihood (ML) or restricted maximum likelihood (REML) approach. For example, penalized quasi-likelihood (Breslow and Clayton (32)), Laplace approximation (Lin and Breslow (33)), pseudo likelihood (Wolfinger (34)), and Gauss-Hermite quadrature. In this dissertation, Laplace approximation is used due to computational simplicity, and the results were acceptable.

5.2 Agreement Statistics for Generalized Linear Mixed Models

In this section, we shall develop MSD, CCC, precision and accuracy coefficients for intra-, inter- and total-rater for any distribution families and link functions using generalized linear mixed model. TDI and CP are no longer available because they require the normal assumption. In addition, an overall version of these agreement statistics for multiple raters is also provided.

To begin with, all the parameters, involving the dispersion parameter, overall mean, fixed effects, and variance components are estimated via Laplace approximation and restricted maximum likelihood approach. Then, μ_{ij} can be computed using Equation (5.3) for arbitrary link functions.

5.2.1 Intra-Rater Precision

From McCulloch and Searle (35) and Carrasco (24), we can show

$$\begin{aligned} \text{var}(y_{ijl}) &= \text{var}(E(y_{ijl}|\mu_{ij})) + E(\text{var}(y_{ijl}|\mu_{ij})) \\ &= \text{var}(\mu_{ij}) + E(a(\phi)v(\mu_{ij})), \end{aligned}$$

and

$$\text{cov}(y_{ijl}, y_{ijl'}) = \text{var}(\mu_{ij}).$$

Thus, for given i and j , the intra-rater MSD between any two replications l and l' is

$$\begin{aligned} \varepsilon_{\text{intra}}^2|i, j &= E(y_{ijl} - y_{ijl'})^2 \\ &= [E(y_{ijl}) - E(y_{ijl'})]^2 + \text{var}(y_{ijl}) + \text{var}(y_{ijl'}) - 2\text{cov}(y_{ijl}, y_{ijl'}) \\ &= 2E(a(\phi)v(\mu_{ij})). \end{aligned} \tag{5.7}$$

The intra-rater CCC for given i and j is

$$\begin{aligned} \rho_{c, \text{intra}}|i, j &= \frac{\text{cov}(y_{ijl}, y_{ijl'})}{\text{var}(y_{ijl})} \\ &= \frac{\text{var}(\mu_{ij})}{\text{var}(\mu_{ij}) + E(a(\phi)v(\mu_{ij}))}. \end{aligned} \tag{5.8}$$

The intra-rater precision coefficient for given i and j is identical to the intra-rater CCC, and the corresponding intra-rater accuracy coefficient is one.

$$\rho_{\text{intra}}|i, j = \frac{\text{var}(\mu_{ij})}{\text{var}(\mu_{ij}) + E(a(\phi)v(\mu_{ij}))}, \tag{5.9}$$

$$\chi_{a, \text{intra}}|i, j = 1. \tag{5.10}$$

5.2.2 Inter-Rater Agreement

Similarly, we have

$$\begin{aligned}\text{var}(\bar{y}_{ij\cdot}) &= \text{var}(E(\bar{y}_{ij\cdot}|\mu_{ij})) + E(\text{var}(\bar{y}_{ij\cdot}|\mu_{ij})) \\ &= \text{var}(\mu_{ij}) + \frac{E(a(\phi)v(\mu_{ij}))}{m_{ij}},\end{aligned}$$

and

$$\text{cov}(\bar{y}_{ij\cdot}, \bar{y}_{ij'\cdot}) = \text{cov}(\mu_{ij}, \mu_{ij'}).$$

Thus, for given i, j and $j', j \neq j'$, the inter-rater MSD between any two raters j and j' is

$$\begin{aligned}\varepsilon_{\text{inter}}^2|i, j, j' &= E(\bar{y}_{ij\cdot} - \bar{y}_{ij'\cdot})^2 \\ &= [E(\bar{y}_{ij\cdot}) - E(\bar{y}_{ij'\cdot})]^2 + \text{var}(\bar{y}_{ij\cdot}) + \text{var}(\bar{y}_{ij'\cdot}) - 2\text{cov}(\bar{y}_{ij\cdot}, \bar{y}_{ij'\cdot}) \\ &= [E(\mu_{ij}) - E(\mu_{ij'})]^2 + \text{var}(\mu_{ij}) + \frac{E(a(\phi)v(\mu_{ij}))}{m_{ij}} \\ &\quad + \text{var}(\mu_{ij'}) + \frac{E(a(\phi)v(\mu_{ij'}))}{m_{ij'}} - 2\text{cov}(\mu_{ij}, \mu_{ij'}). \tag{5.11}\end{aligned}$$

The inter-rater CCC is

$$\begin{aligned}\rho_{c,\text{inter}}|i, j, j' &= \frac{2\text{cov}(\mu_{ij}, \mu_{ij'})}{\sqrt{[E(\mu_{ij}) - E(\mu_{ij'})]^2 + \text{var}(\mu_{ij}) + \frac{1}{m_{ij}}E(a(\phi)v(\mu_{ij})) + \text{var}(\mu_{ij'}) + \frac{1}{m_{ij'}}E(a(\phi)v(\mu_{ij'}))}}.\end{aligned} \tag{5.12}$$

The inter-rater precision coefficient is

$$\rho_{\text{inter}}|i, j, j' = \frac{\text{cov}(\mu_{ij}, \mu_{ij'})}{\sqrt{\text{var}(\mu_{ij}) + \frac{1}{m_{ij}}E(a(\phi)v(\mu_{ij}))} \sqrt{\text{var}(\mu_{ij'}) + \frac{1}{m_{ij'}}E(a(\phi)v(\mu_{ij'}))}}. \quad (5.13)$$

And the inter-rater accuracy coefficient can be calculated as

$$\chi_{a,\text{inter}}|i, j, j' = \frac{\rho_{c,\text{inter}}|i, j, j'}{\rho_{\text{inter}}|i, j, j'}. \quad (5.14)$$

5.2.3 Total-Rater Agreement

It is straightforward to show that

$$\text{cov}(y_{ijl}, y_{ij'l'}) = \text{cov}(\mu_{ij}, \mu_{ij'}).$$

Thus, for given i, j and $j', j \neq j'$, the total-rater MSD is

$$\begin{aligned} \varepsilon_{\text{total}}^2|i, j, j' &= E(y_{ijl} - y_{ij'l'})^2 \\ &= [E(y_{ijl}) - E(y_{ij'l'})]^2 + \text{var}(y_{ijl}) + \text{var}(y_{ij'l'}) - 2\text{cov}(y_{ijl}, y_{ij'l'}) \\ &= [E(\mu_{ij}) - E(\mu_{ij'})]^2 + \text{var}(\mu_{ij}) + E(a(\phi)v(\mu_{ij})) \\ &\quad + \text{var}(\mu_{ij'}) + E(a(\phi)v(\mu_{ij'})) - 2\text{cov}(\mu_{ij}, \mu_{ij'}). \end{aligned} \quad (5.15)$$

The total-rater CCC is

$$\rho_{c,\text{total}}|i, j, j' = \frac{2\text{cov}(\mu_{ij}, \mu_{ij'})}{[E(\mu_{ij}) - E(\mu_{ij'})]^2 + \text{var}(\mu_{ij}) + E(a(\phi)v(\mu_{ij})) + \text{var}(\mu_{ij'}) + E(a(\phi)v(\mu_{ij'}))}. \quad (5.16)$$

The total-rater precision coefficient is

$$\rho_{\text{total}}|i, j, j' = \frac{\text{cov}(\mu_{ij}, \mu_{ij'})}{\sqrt{\text{var}(\mu_{ij}) + E(a(\phi)v(\mu_{ij}))} \sqrt{\text{var}(\mu_{ij'}) + E(a(\phi)v(\mu_{ij'}))}}. \quad (5.17)$$

And the total-rater accuracy coefficient can be calculated as

$$\chi_{a,\text{total}}|i, j, j' = \frac{\rho_{c,\text{total}}|i, j, j'}{\rho_{\text{total}}|i, j, j'}. \quad (5.18)$$

5.2.4 Restricted Agreement Statistics

In many cases, data may come with several covariates, or other fixed effects. For instance, readings from a sophisticated medical device may depend on the room temperature and humidity. If the readings of one medical device are collected under room temperature 80°F and humidity 80%, while the readings of another device are collected under room temperature 50°F and humidity 50%. It is obvious that we can not obtain good agreement estimates, like CCC, precision and accuracy coefficients. The agreement will be reduced due to the difference of these covariates. That is to say, large differences of the covariates between raters will explode

the inter-rater MSD and make the inter-rater CCC decreasing to zero. However, we may not be interested in these covariates, but the inter- and total-rater effects or variabilities. Thus, it is more appropriate to evaluate agreement when all the readings are collected under the same condition.

The restricted agreement estimates is proposed for this kind of situations, by eliminating covariates effects. For example, the restricted concordance correlation coefficient (RCCC) for two raters X and Y is defined as follows,

$$\tilde{\rho}_c = 1 - \frac{E(X - Y)^2|_{\mathbf{x}_X = \mathbf{x}_Y = \mathbf{x}_0}}{E(X - Y)^2|_{\mathbf{x}_X = \mathbf{x}_Y = \mathbf{x}_0, \rho = 0}}, \quad (5.19)$$

where \mathbf{x}_X and \mathbf{x}_Y is the covariates for X and Y respectively, and both of them are set as a constant value \mathbf{x}_0 .

This RCCC (5.19) indicates the concordance correlation coefficient if all observations have the same covariates. It measures the agreement under the same condition. Moreover, it can be shown that RCCC possesses all the characteristics of CCC. Using this idea, for given i, j and $j', j \neq j'$, we can conclude the intra-, inter- and total-rater individual RCCC as follows,

$$\tilde{\rho}_{c,\text{intra}}|i, j = \frac{\text{var}(\hat{\mu}_{ij})}{\text{var}(\hat{\mu}_{ij}) + E(a(\phi)v(\hat{\mu}_{ij}))} \Big|_{\mathbf{x}_{ij}=\mathbf{x}_0},$$

$$\tilde{\rho}_{c,\text{inter}}|i, j, j' =$$

$$\frac{2\text{cov}(\hat{\mu}_{ij}, \hat{\mu}_{ij'})}{[E(\hat{\mu}_{ij}) - E(\hat{\mu}_{ij'})]^2 + \text{var}(\hat{\mu}_{ij}) + \frac{1}{m_{ij}}E(a(\phi)v(\hat{\mu}_{ij})) + \text{var}(\hat{\mu}_{ij'}) + \frac{1}{m_{ij'}}E(a(\phi)v(\hat{\mu}_{ij'}))} \Big|_{\mathbf{x}_{ij}=\mathbf{x}_{ij'}=\mathbf{x}_0},$$

$$\tilde{\rho}_{c,\text{total}}|i, j, j' =$$

$$\frac{2\text{cov}(\hat{\mu}_{ij}, \hat{\mu}_{ij'})}{[E(\hat{\mu}_{ij}) - E(\hat{\mu}_{ij'})]^2 + \text{var}(\hat{\mu}_{ij}) + E(a(\phi)v(\hat{\mu}_{ij})) + \text{var}(\hat{\mu}_{ij'}) + E(a(\phi)v(\hat{\mu}_{ij'}))} \Big|_{\mathbf{x}_{ij}=\mathbf{x}_{ij'}=\mathbf{x}_0},$$

where $\hat{\mu}_{ij}, \hat{\mu}_{ij'}$ are the fitted values when $\mathbf{x}_{ij} = \mathbf{x}_{ij'} = \mathbf{x}_0$.

Other restricted agreement statistics can be defined similarly, by using $\hat{\mu}_{ij}, \hat{\mu}_{ij'}$ at $\mathbf{x}_{ij} = \mathbf{x}_{ij'} = \mathbf{x}_0$.

5.2.5 Overall Agreement Statistics

Just like the linear mixed model case, weighted averages are used to calculate the overall agreement statistics for multiple raters. The overall CCC, precision and accuracy coefficient can be defined as following.

$$\rho_{c,\text{intra}} = \rho_{\text{intra}} = \frac{\sum_{i=1}^n \sum_{j=1}^k \rho_{c,\text{intra}}|i, j \times \xi_{\text{intra}}|i, j}{\sum_{i=1}^n \sum_{j=1}^k \xi_{\text{intra}}|i, j}, \quad (5.20)$$

$$\rho_{c,\text{inter}} = \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \rho_{c,\text{inter}}|i, j, j' \times \xi_{\text{inter}}|i, j, j'}{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \xi_{\text{inter}}|i, j, j'}, \quad (5.21)$$

$$\rho_{c,\text{total}} = \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \rho_{c,\text{total}}|i, j, j' \times \xi_{\text{total}}|i, j, j'}{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \xi_{\text{total}}|i, j, j'}, \quad (5.22)$$

$$\rho_{\text{inter}} = \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \rho_{\text{inter}}|i, j, j' \times \xi_{\text{inter}}|i, j, j'}{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \xi_{\text{inter}}|i, j, j'}, \quad (5.23)$$

$$\rho_{\text{total}} = \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \rho_{\text{total}}|i, j, j' \times \xi_{\text{total}}|i, j, j'}{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \xi_{\text{total}}|i, j, j'}, \quad (5.24)$$

$$\chi_{a,\text{inter}} = \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \chi_{a,\text{inter}}|i, j, j' \times \xi_{\text{inter}}|i, j, j'}{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \xi_{\text{inter}}|i, j, j'}, \quad (5.25)$$

$$\chi_{a,\text{total}} = \frac{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \chi_{a,\text{total}}|i, j, j' \times \xi_{\text{total}}|i, j, j'}{\sum_{i=1}^n \sum_{j=1}^{k-1} \sum_{j'=j+1}^k \xi_{\text{total}}|i, j, j'}, \quad (5.26)$$

where the weights are

$$\begin{aligned} \xi_{\text{intra}}|i, j &= \text{var}(\mu_{ij}) + E(a(\phi)v(\mu_{ij})), \\ \xi_{\text{inter}}|i, j, j' &= [E(\mu_{ij}) - E(\mu_{ij'})]^2 + \text{var}(\mu_{ij}) + \frac{1}{m_{ij}}E(a(\phi)v(\mu_{ij})) \\ &\quad + \text{var}(\mu_{ij'}) + \frac{1}{m_{ij'}}E(a(\phi)v(\mu_{ij'})), \\ \xi_{\text{total}}|i, j, j' &= [E(\mu_{ij}) - E(\mu_{ij'})]^2 + \text{var}(\mu_{ij}) + E(a(\phi)v(\mu_{ij})) + \text{var}(\mu_{ij'}) + E(a(\phi)v(\mu_{ij'})). \end{aligned}$$

The overall restricted agreement statistics can also be defined using the same weighted averages as (5.20) – (5.26).

5.3 Estimation for Different Response Models

5.3.1 Dispersion Parameters and Variance Functions

For different types of data, the conditional distribution of y in (5.1) can take several different distribution families. For example, Gamma distribution can be used for continuous responses in which all the data are positive, binary distribution can be used for 0/1 responses, and Poisson and negative binomial distribution can be used for count data. Table XIII lists dispersion parameters and variance functions for some commonly used distribution families.

5.3.2 Link Functions

For a given distribution family, the frequently used link function is the canonical link function, which makes the canonical parameter τ equal to the linear predictor η . Table XIV gives a list of canonical links for several selected distribution families.

The negative binomial model is often served as an advanced model for overdispersed Poisson regression model (Hardin and Hilbe, (36)). Therefore, it is rarely used in canonical link, and a natural log link is used more often.

Next, we shall develop the expressions for $E(\mu_{ij})$, $\text{var}(\mu_{ij})$, and $\text{cov}(\mu_{ij}, \mu_{ij'})$, which will be used to calculate the agreement statistics for different link functions.

TABLE XIII: List of Dispersion Parameters and Variance Functions for Selected Distributions

Distribution	$a(\phi)$	$v(\mu)$
Gaussian(μ, σ_ϵ^2)	σ_ϵ^2	1
Gamma(μ, ϕ)	ϕ	μ^2
Binomial(p, N)	1	$\mu(1 - \mu)/N$
Poisson(μ)	1	μ
Negative Binomial(r, p)	1	$\mu + \mu^2 r$

TABLE XIV: List of the Canonical Links for Selected Distributions

Distribution	Canonical Links	Inverse of Canonical Link
Gaussian(μ, σ_ϵ^2)	μ	η
Gamma(μ, ϕ)	$1/\mu$	$1/\eta$
Binomial(p, N)	$\ln(\mu/(1 - \mu))$	$\exp(\eta)/(1 + \exp(\eta))$
Poisson(μ)	$\ln(\mu)$	$\exp(\eta)$
Negative Binomial(r, p)	$-\ln(r/\mu + 1)$	$r/(\exp(-\eta) - 1)$

For natural log link, the expressions for $E(\mu_{ij})$, $\text{var}(\mu_{ij})$, and $\text{cov}(\mu_{ij}, \mu_{ij'})$ are easy to find, because

$$\begin{aligned} E(\mu_{ij}) &= E(\exp(X_{ij}\boldsymbol{\beta} + Z_{1ij}\boldsymbol{\alpha} + Z_{2ij}\boldsymbol{\gamma})) \\ &= \exp(X_{ij}\boldsymbol{\beta})E(\exp(Z_{1ij}\boldsymbol{\alpha}))E(\exp(Z_{2ij}\boldsymbol{\gamma})) \\ &= \exp(X_{ij}\boldsymbol{\beta})M_{\boldsymbol{\alpha}}(Z_{1ij})M_{\boldsymbol{\gamma}}(Z_{2ij}), \end{aligned}$$

where $M_{\dagger}(\cdot)$ is the moment generating function for variable \dagger at \cdot . X_{ij} is the fixed effects including means, Z_{1ij} and Z_{2ij} stand for the rows corresponding to the response y_{ij} in \mathbf{Z}_1 and \mathbf{Z}_2 .

Assume $\boldsymbol{\alpha} \sim N(0, \sigma_{\alpha}^2 \cdot \mathbf{I}_n)$, $\boldsymbol{\gamma} \sim N(0, \sigma_{\gamma}^2 \cdot \mathbf{I}_{n \times k})$, and the rows Z_{1ij} and Z_{2ij} have all zeros but only one 1. From McCulloch and Searle (35) and Carrasco (24), we have

$$E(\mu_{ij}) = \exp(X_{ij}\boldsymbol{\beta}) \exp\left(\frac{\sigma_{\alpha}^2 + \sigma_{\gamma}^2}{2}\right), \quad (5.27)$$

$$\text{var}(\mu_{ij}) = \exp(2X_{ij}\boldsymbol{\beta} + 2\sigma_{\alpha}^2 + 2\sigma_{\gamma}^2) - \exp(2X_{ij}\boldsymbol{\beta} + \sigma_{\alpha}^2 + \sigma_{\gamma}^2), \quad (5.28)$$

$$\text{cov}(\mu_{ij}, \mu_{ij'}) = \exp(X_{ij}\boldsymbol{\beta} + X_{ij'}\boldsymbol{\beta} + 2\sigma_{\alpha}^2 + \sigma_{\gamma}^2) - \exp(X_{ij}\boldsymbol{\beta} + X_{ij'}\boldsymbol{\beta} + \sigma_{\alpha}^2 + \sigma_{\gamma}^2). \quad (5.29)$$

For other links, like logit, square root or inverse links, we may use Taylor expansion to find the first and the second moments, but the radius of convergence for such Taylor series is hard to determine. Hence, adaptive numerical integration algorithm (Berntsen, Espelid, and

Genz (37)) is used to calculate the $E(\mu_{ij})$, $E(\mu_{ij}^2)$, and $E(\mu_{ij}\mu_{ij'})$ numerically for all the link functions, assuming both α and γ are normally distributed.

5.3.3 Confidence Intervals

Unfortunately, unlike the well developed asymptotic normality of restricted maximum likelihood estimates in linear mixed models, the asymptotic behavior of the variance components estimates from GLMM is not clear. It is known that the sampling distribution of variance estimates from likelihood-based approach is in general strongly asymmetric (38). The estimating variances may not be a good measure to characterize the uncertainty.

Therefore, instead of using standard normal confidence interval calculated from the variances estimates, a 95% percentiles interval is used via parametric bootstrap, which is more suitable for a asymmetric distribution.

The estimating algorithm can be described as following.

1. Estimate the fixed effects, dispersion parameters, and variance components via restricted maximum likelihood and Laplace approximation.
2. Compute the agreement statistics based on the chosen link function and parameter estimates.
3. Draw new samples randomly from the fitted model, using the same sample size as the original data.
4. Proceed to Step 1 and Step 2 again for the new samples, and repeat this procedure one thousand times.

5. Evaluate the agreement statistics on each bootstrap sample, and report the 95% percentile interval of these one thousand replications.

CHAPTER 6

SIMULATIONS AND CASE STUDY FOR GENERALIZED LINEAR MIXED MODELS

6.1 Simulations

For evaluating the performance of our agreement estimates, several simulations using different distribution families and link functions were conducted.

The data generating procedure was based on the generalized linear mixed model (5.1). The main steps can be described as follows.

1. Generate the random subject and interaction effect α_i and γ_{ij} using independent normal distribution with mean zero, and pre-determined standard deviations σ_α and σ_γ .
2. Calculate the individual mean μ_{ij} as $\mu_{ij} = g^{-1}(\mu + \beta_j + \alpha_i + \gamma_{ij})$, where $g(\cdot)$ is the link function and μ is the overall mean.
3. For each element in the dataset, generate one sample from the desired distribution with mean μ_{ij} and some other pre-set parameters.

6.1.1 Log-Gamma Distribution

We began with the Gamma distribution with natural log link, which is widely used as an alternative to the log-normal data with large skewness.

Fifty samples were generated at each run for three raters with two replications, which is to say, $n = 50$, $k = 3$ and $m = 2$. The standard deviations for generating linear predictors were chosen at $\sigma_\alpha = 1$ and $\sigma_\gamma = 0.5$, the scale parameter for random Gamma generator was set at 2, the overall mean was chosen at $\mu = 0.8$, and fixed effects was set at $\beta_1 = 0$, $\beta_2 = 0.1$, $\beta_3 = 0.2$. In addition, 10% randomly missing values were added for demonstrating the capability of handling unbalanced data in our approach.

Table XV presents the simulation results for all intra-, inter- and total-rater agreement statistics in 1000 simulation runs. MSD is no longer reported due to less importance. The column **True** gives the theoretical values calculated from the simulation parameters, **Mean(Est)** gives the means of all the estimates, **Std(Est)** gives standard deviations of all these estimates in 1000 runs, and **95% P.I.** gives the means of lower (2.5%) and upper (97.5%) boundary for the 95% percentiles interval from parametric bootstrap for each run.

Table XVI presents the simulation results in 1000 simulation runs involving a covariate $x_{ij} = \ln(i \times j)$. All the true values and agreement estimates are reported as restricted agreement estimates at all $x_{ij} = 0$. No missing values were conducted in this simulation.

Table XV and Table XVI show that our approach had very good estimates for log-Gamma distribution, even when missing values or covariates were involved. All the means of estimates were very close to the true value, differences only appeared after the second digit after the decimal point, and all the standard deviations were within an acceptable range.

Comparing Table XV and Table XVI, we may notice that they had the same true values, because the covariates effects were eliminated by using the fitted value at $x_{ij} = 0$ during the regression of generalized linear mixed models. The standard deviations of estimates with covariates were a little larger than those without covariates, and the corresponding 95% percentiles intervals were a little wider than those without covariates. Such differences were caused by the introduction of covariates.

6.1.2 Log-Negative Binomial Distribution

For the negative binomial case with natural log link function, fifty samples were generated at each run for two raters with two replications, which is to say, $n = 50$, $k = 2$ and $m = 2$. The standard deviations for generating linear predictors were chosen at $\sigma_\alpha = 1$ and $\sigma_\gamma = 0.5$, the shape parameter for random negative binomial generator was set at 10, the overall mean was chosen at $\mu = 1.8$, and fixed effects was set at $\beta_1 = 0$, $\beta_2 = 0.2$.

Table XVII presents the simulation results for all intra-, inter- and total-rater agreement statistics in 1000 simulation runs with 10% missing values, while Table XVIII presents the restricted agreement estimates in 1000 simulation runs involving a covariate $x_{ij} = \ln(i \times j)$ but no missing values.

Table XVII and Table XVIII show that all the means of estimates were very close to the true value, and all the standard deviations were within an acceptable range. This demonstrates that our approach also had excellent estimates for log-negative binomial distribution, no matter there were missing values or covariates.

6.1.3 Logistic-Binary Distribution

Binary data with logistic link function was also conducted. Two hundred samples were generated at each run for two raters with two replications, which is to say, $n = 200$, $k = 2$ and $m = 2$. The standard deviations for generating linear predictors were chosen at $\sigma_\alpha = 2$ and $\sigma_\gamma = 1.5$, the overall mean was chosen at $\mu = -2$, and fixed effects was set at $\beta_1 = 0$, $\beta_2 = 1$.

Like the previous two cases, Table XIX presents the simulation results for all intra-, inter- and total-rater agreement statistics in 1000 simulation runs with 10% missing values, and Ta-

ble XX presents the restricted agreement estimates in 1000 simulation runs involving a covariate $x_{ij} = \ln(i \times j)$ but no missing values.

Table XIX and Table XX also demonstrate that our approach worked well for this logistic-binary distribution. Although the differences between the true values and estimates were somewhat bigger than those in the log-Gamma or log-negative binomial case, the relative differences were still no larger than 5%, and all the standard deviations were acceptable.

TABLE XV: Simulation Results for Log-Gamma Distribution with 10% Missing Values

		95% P.I.				
		True	Mean(Est)	Std(Est)	Lower	Upper
Intra	CCC	0.588	0.594	0.042	0.547	0.672
Inter	CCC	0.504	0.496	0.063	0.367	0.610
	Precision	0.511	0.509	0.064	0.379	0.622
	Accuracy	0.987	0.981	0.022	0.931	0.999
Total	CCC	0.401	0.403	0.053	0.301	0.506
	Precision	0.406	0.413	0.054	0.310	0.516
	Accuracy	0.988	0.983	0.021	0.935	0.999

TABLE XVI: Simulation Results for Log-Gamma Distribution with Covariates

		95% P.I.				
		True	Mean(Est)	Std(Est)	Lower	Upper
Intra	CCC	0.588	0.590	0.043	0.510	0.663
Inter	CCC	0.504	0.493	0.068	0.367	0.630
	Precision	0.511	0.502	0.069	0.370	0.633
Total	Accuracy	0.987	0.981	0.039	0.871	0.999
	CCC	0.401	0.392	0.056	0.286	0.497
	Precision	0.406	0.407	0.056	0.300	0.516
	Accuracy	0.988	0.982	0.037	0.876	0.999

TABLE XVII: Simulation Results for Log-Negative Binomial Distribution with 10% Missing Values

		95% P.I.				
		True	Mean(Est)	Std(Est)	Lower	Upper
Intra	CCC	0.853	0.851	0.031	0.784	0.901
Inter	CCC	0.618	0.614	0.084	0.447	0.764
	Precision	0.635	0.630	0.085	0.462	0.787
Total	Accuracy	0.974	0.969	0.033	0.887	1.000
	CCC	0.573	0.566	0.080	0.409	0.713
	Precision	0.588	0.583	0.081	0.422	0.732
	Accuracy	0.974	0.968	0.032	0.890	1.000

TABLE XVIII: Simulation Results for Log-Negative Binomial Distribution with Covariates

		95% P.I.				
		True	Mean(Est)	Std(Est)	Lower	Upper
Intra	CCC	0.853	0.846	0.036	0.763	0.898
Inter	CCC	0.618	0.613	0.092	0.450	0.778
	Precision	0.635	0.630	0.089	0.441	0.817
Total	Accuracy	0.974	0.969	0.046	0.851	1.000
	CCC	0.573	0.568	0.088	0.404	0.725
	Precision	0.588	0.582	0.087	0.414	0.740
	Accuracy	0.974	0.970	0.045	0.883	1.000

TABLE XIX: Simulation Results for Logistic-Binary Distribution with 10% Missing Values

		95% P.I.				
		True	Mean(Est)	Std(Est)	Lower	Upper
Intra	CCC	0.459	0.472	0.057	0.293	0.571
Inter	CCC	0.347	0.340	0.059	0.230	0.502
	Precision	0.364	0.356	0.061	0.249	0.474
	Accuracy	0.953	0.948	0.025	0.894	0.986
Total	CCC	0.256	0.266	0.052	0.172	0.366
	Precision	0.265	0.270	0.054	0.171	0.376
	Accuracy	0.965	0.957	0.021	0.910	0.989

TABLE XX: Simulation Results for Logistic-Binary Distribution with Covariates

		95% P.I.				
		True	Mean(Est)	Std(Est)	Lower	Upper
Intra	CCC	0.459	0.477	0.060	0.291	0.586
Inter	CCC	0.347	0.335	0.081	0.135	0.468
	Precision	0.364	0.352	0.087	0.140	0.509
Total	Accuracy	0.953	0.946	0.031	0.877	0.992
	CCC	0.256	0.268	0.076	0.083	0.399
	Precision	0.265	0.276	0.081	0.763	0.419
	Accuracy	0.965	0.951	0.028	0.888	0.993

6.2 Antihypertensive Patch Dataset

Let's revisit the antihypertensive patch dataset in Chapter 4. The logarithm of AUC values were assumed to be normally distributed in (31) and Chapter 4. However, the histogram plot of $\log(\text{AUC})$ in Figure 4 reveals that the distribution of logarithm of AUC values is actually skewed. Hence, the log-Gamma model is more appropriate for this dataset.

Applying the GLMM model (5.1) – (5.3) with Gamma distribution family and natural log link function. Assume the linear predictor,

$$\eta_{ij} = \mu + \beta_j + s_i + \alpha_i + \gamma_{ij}, \quad (6.1)$$

where the subscript i is the index of subjects, j is the index of formulations (T/R), μ is the overall mean, β_j is the fixed formulation effects, s_i is the fixed sequence effects, α_i is the random subject effect nested in sequences, and γ_{ij} is the random subject-formulation interaction effect.

Table XXI presents the intra-, inter- and total-rater agreement estimates with 95% percentiles intervals (95% P.I.) from parametric bootstrap. All the accuracy coefficients are very close to 1, which indicate no major location shift for the marginal distributions of reference and test formulations. If an allowance of 0.950 is used for CCC, we can see that the inter-CCC is 0.954, which exceeds that allowance. Therefore, the agreement between the reference and test formulations are satisfactory, and thus can be considered interchangeable.

Figure 5 presents the histograms of selected agreement estimates for 1000 bootstrap simulations, these histograms reveal the asymmetric distributions of agreement estimates. Thus,

the parametric bootstrap percentile interval is more appropriate than the normal confidence interval for characterizing the uncertainty.

TABLE XXI: Estimated Agreement Statistics for AUC

		Est	95% P.I.
Intra	CCC	0.914	(0.861, 0.949)
Inter	CCC	0.954	(0.925, 0.973)
	Precision	0.955	(0.925, 0.974)
	Accuracy	0.999	(0.998, 1.000)
Total	CCC	0.913	(0.859, 0.948)
	Precision	0.914	(0.859, 0.949)
	Accuracy	0.999	(0.998, 1.000)

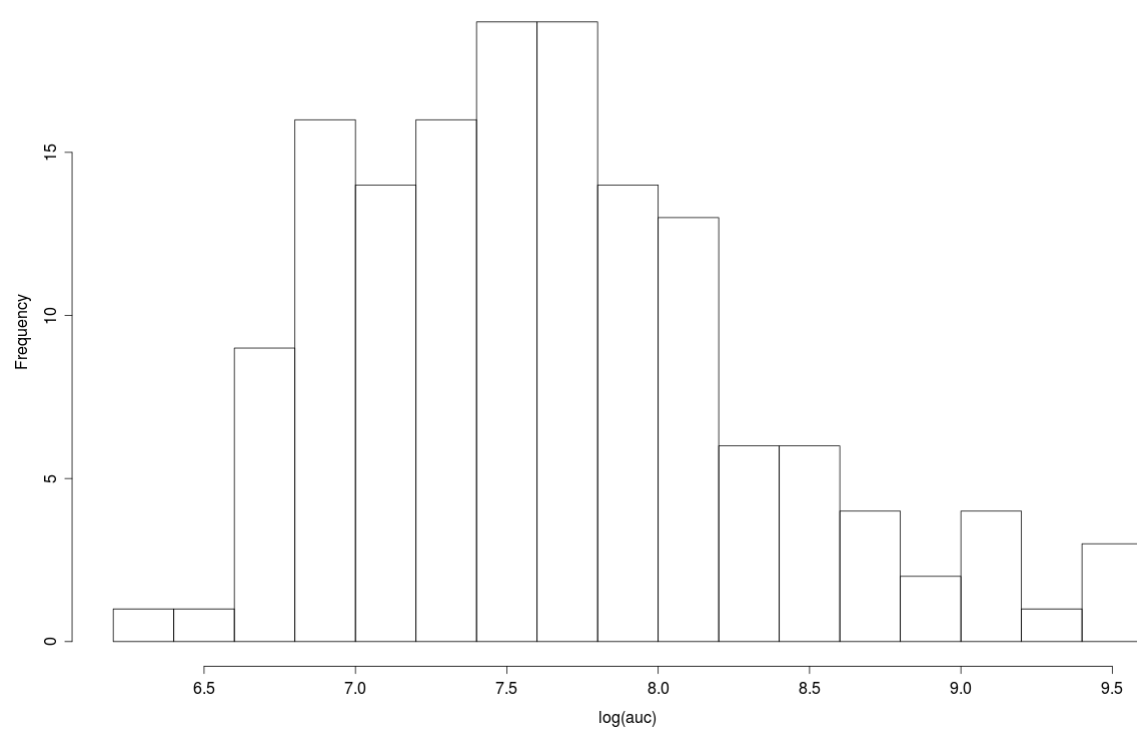


Figure 4: Histogram of $\log(\text{AUC})$

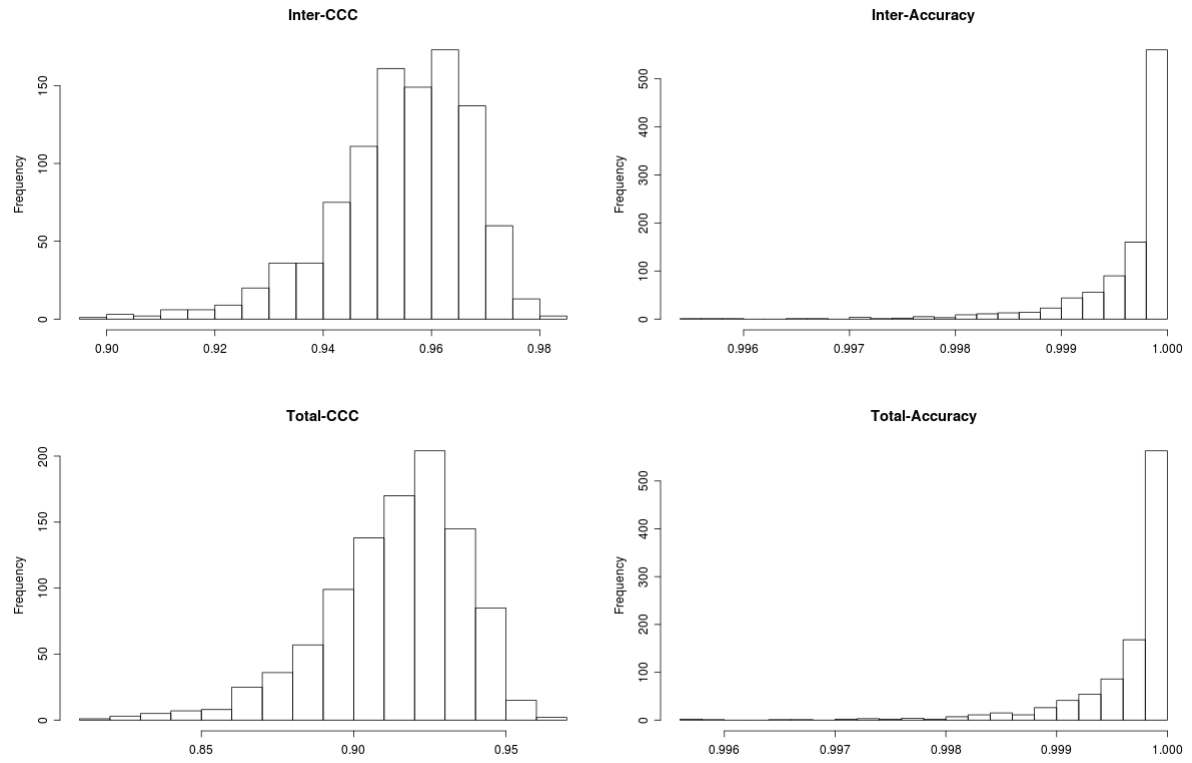


Figure 5: Histogram of Selected Estimates for 1000 Bootstrap Simulations

CHAPTER 7

CONCLUSION AND DISCUSSIONS

Assessment of agreement has been studied in many literatures. One of the most cited papers is Lin's concordance correlation coefficient paper (7), which defined a brilliant CCC statistic for evaluating reproducibility. In this dissertation, we follow Lin's idea of CCC, and extend it for measuring agreement for multiple raters with multiple replicates. Furthermore, our approach is designed not only for balanced data, but also for unbalanced data with covariates, and it can be applied to a large variety of distribution families using either linear mixed models or generalized linear mixed models.

For the linear mixed models, the well developed restricted maximum likelihood approach is used, and the desired asymptotic normality for the agreement estimates, even when the original data is not normally distributed, can be concluded. For the balanced data, with or without the normality assumption, the REML estimates are exactly the same as the ANOVA estimates, and our agreement estimates are identical to those estimated by Lin's unified approach (13). One of the advantages of using REML rather than the ANOVA approach is that the likelihood-based approach has the merit of handling unbalanced data, missing values and covariates naturally. This increases the application scope of our approach dramatically.

Simulations justify the excellent performance of our approach using linear mixed models for not only normally distributed data, but also discrete and skewed data. Our approach also works well even when the true model of the data is unclear. In addition, if the sample size is not large, say less than 100, the normalizing transformations are recommended for computing the confidence intervals for all the agreement estimates.

However, the linear mixed model does not perform well for the binary data, because the binary responses can not be decomposed into the sum of fixed and random effects. One of the widely used methods for modeling binary data is logistic regression. Hence, it is necessary to extend our results in linear mixed models to generalized linear mixed models for modeling a larger diversities of data.

For generalized linear mixed models, the agreement estimates are totally different due to the existence of various distribution families and link functions. It is not easy to come out with a unified expression for all the estimates. Thus, numerical algorithms are introduced for estimating the model parameters and computing the agreement statistics. Several exiting methods, such as Wald confidence interval, profile likelihood confidence interval via likelihood ratio test, and parametric bootstrap percentiles interval, can be used to evaluate the uncertainty of the estimates. However, Wald confidence interval is least preferred since it is only appropriate if the log-likelihood is well approximated by a quadratic function (39). There are no known substantial advantages between parametric bootstrap and profile likelihood; although some literatures point out that the performances of the likelihood based approaches depended on the model assumptions and data structures ((40), (41)). In this dissertation, we shall use the parametric bootstrap percentiles intervals rather than the likelihood based ones. The reason being the parametric bootstrap method requires no distribution assumptions and works relatively well for small sample sizes, while profile likelihood requires computing the log-likelihood many times, which is unfortunately error-prone in GLMM.

Antihypertensive patch dataset is a perfect example to illustrate the superiority of our GLMM approach. Generally, $\log(\text{AUC})$ has been regarded as normally distributed, for the purpose of being statistically analyzed in linear mixed models assume normality (42), (43). Although some studies reveal the distribution of $\log(\text{AUC})$, especially in a small sample size, is non-normal and skewed (31), (44). The skewness is confirmed in the antihypertensive patch dataset by the histogram plot. Therefore, GLMM using Gamma distribution family with natural log link function is a better way to characterize this dataset, and it can yield more appropriate agreement estimates.

The idea of assessment of agreement may not be restricted to linear and generalized mixed models alone. It can be generalized to some other models. Liu, Du, Teresi, and Hasin (45) proposed a bivariate CCC for survival data, and Barnhart, Song, and Lyles (46) studied the assay validation for left-censored data using maximum likelihood and generalized estimating equations approaches. Further research are needed for assessing agreement for survival data if multiple rater, covariates, and missing values are involved. Additionally, non-linear mixed models are commonly used in pharmaceutical industries for modeling pharmacokinetics and pharmacodynamics. Further research can also be conducted for evaluating agreement in non-linear mixed models.

APPENDIX

GENERATING MULTIVARIATE NEGATIVE BINOMIAL SERIES USING GAMMA – POISSON MIXTURE

APPENDIX (Continued)

In this appendix, we shall generate correlated multivariate negative binomial (NB) series with given parameters.

For example, we may want to generate four correlated negative binomial series for $r = 10$, $p = 0.1$, with Pearson correlation coefficient matrix

$$\begin{pmatrix} 1 & 0.9 & 0.8 & 0.8 \\ 0.9 & 1 & 0.8 & 0.8 \\ 0.8 & 0.8 & 1 & 0.9 \\ 0.8 & 0.8 & 0.9 & 1 \end{pmatrix}.$$

To begin with, consider a bivariate case, assuming

$$y_1 | \lambda_1 \sim \text{Poisson}(\lambda_1),$$

$$y_2 | \lambda_2 \sim \text{Poisson}(\lambda_2),$$

and these two Poisson distributions are independent for given λ_1 and λ_2 .

In addition, let λ_1, λ_2 follow the same Gamma distribution with shape α and scale θ (or rate $\beta = \theta^{-1}$), and $\text{Cor}(\lambda_1, \lambda_2) = \rho$.

With a little algebra, we can show that

$$y_1 \sim NB(r = \alpha, p = \frac{\theta}{1 + \theta}),$$

$$y_2 \sim NB(r = \alpha, p = \frac{\theta}{1 + \theta}).$$

APPENDIX (Continued)

The variance-covariance matrix for y_1 and y_2 can be calculated as follows,

$$\begin{aligned}
 \text{Var} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} &= E_{\lambda} \left(\text{Var} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \middle| \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \right) + \text{Var}_{\lambda} \left(E \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \middle| \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \right) \\
 &= E_{\lambda} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} + \text{Var}_{\lambda} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \\
 &= \begin{pmatrix} \alpha\theta & \\ & \alpha\theta \end{pmatrix} + \begin{pmatrix} \alpha\theta^2 & \rho\alpha\theta^2 \\ \rho\alpha\theta^2 & \alpha\theta^2 \end{pmatrix} \\
 &= \begin{pmatrix} \alpha\theta + \alpha\theta^2 & \rho\alpha\theta^2 \\ \rho\alpha\theta^2 & \alpha\theta + \alpha\theta^2 \end{pmatrix}.
 \end{aligned}$$

Therefore, the correlation matrix for y_1 and y_2 is

$$\begin{pmatrix} 1 & \frac{\theta}{1+\theta}\rho \\ \frac{\theta}{1+\theta}\rho & 1 \end{pmatrix}.$$

This can be easily extended to multivariate case with different parameters for negative binomial, which is not shown here.

APPENDIX (Continued)

In conclusion, if we want to generate four correlated negative binomial series with the same parameter set r , p and correlation matrix

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ & 1 & \rho_{23} & \rho_{24} \\ & & 1 & \rho_{34} \\ & & & 1 \end{pmatrix},$$

the steps can be described as follows.

1. Calculating $\alpha = r$, $\theta = \frac{p}{1-p}$, and correlation matrix for Gamma distribution

$$\Sigma_{Gamma} = \begin{pmatrix} 1 & \frac{1+\theta}{\theta}\rho_{12} & \frac{1+\theta}{\theta}\rho_{13} & \frac{1+\theta}{\theta}\rho_{14} \\ & 1 & \frac{1+\theta}{\theta}\rho_{23} & \frac{1+\theta}{\theta}\rho_{24} \\ & & 1 & \frac{1+\theta}{\theta}\rho_{34} \\ & & & 1 \end{pmatrix},$$

2. Generating four correlated Gamma series using multivariate Gamma generator, with correlation Σ_{Gamma} .
3. For each sample in the Gamma series, taking one sample from independent Poisson distribution, assuming the Poisson parameter equals the value of that Gamma sample.
4. The samples from Poisson distribution are the negative binomial series we want.

CITED LITERATURE

1. Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1):37–46, 1960.
2. Cohen, J.: Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin, 70(4):213, 1968.
3. Fleiss, J., Cohen, J., and Everitt, B.: Large sample standard errors of kappa and weighted kappa. Psychological Bulletin, 72(5):323, 1969.
4. Fleiss, J.: Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5):378, 1971.
5. Fleiss, J. and Cuzick, J.: The reliability of dichotomous judgments: unequal numbers of judges per subject. Applied Psychological Measurement, 3(4):537, 1979.
6. Bland, J. and Altman, D.: Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet, 327(8476):307–310, 1986.
7. Lin, L.: A concordance correlation coefficient to evaluate reproducibility. Biometrics, 45(1):255, 1989.
8. Lin, L.: Assay validation using the concordance correlation coefficient. Biometrics, 48(2):599–604, 1992.
9. Chinchilli, V., Martel, J., Kumanyika, S., and Lloyd, T.: A weighted concordance correlation coefficient for repeated measurement designs. Biometrics, 341–353, 1996.
10. Robieson, W.: On weighted kappa and concordance correlation coefficient. Doctoral dissertation, University of Illinois at Chicago, 1999.
11. Lin, L., Hedayat, A., Sinha, B., and Yang, M.: Statistical methods in assessing agreement. Journal of the American Statistical Association, 97(457):257–270, 2002.
12. Carrasco, J. and Jover, L.: Concordance correlation coefficient applied to discrete data. Statistics in Medicine, 24(24):4021–4034, 2005.

13. Lin, L., Hedayat, A., and Wu, W.: A unified approach for assessing agreement for continuous and categorical data. Journal of Biopharmaceutical Statistics, 17(4):629, 2007.
14. Barnhart, H. and Williamson, J.: Modeling concordance correlation via gee to evaluate reproducibility. Biometrics, 57(3):931–940, 2001.
15. Barnhart, H., Haber, M., and Song, J.: Overall concordance correlation coefficient for evaluating agreement among multiple observers. Biometrics, 58(4):1020–1027, 2002.
16. Barnhart, H., Song, J., and Haber, M.: Assessing intra, inter and total agreement with replicated readings. Statistics in Medicine, 24(9):1371–1384, 2005.
17. Thompson Jr, W.: The problem of negative estimates of variance components. The Annals of Mathematical Statistics, 273–289, 1962.
18. Graybill, F.: On quadratic estimates of variance components. The Annals of Mathematical Statistics, 367–372, 1954.
19. Graybill, F. and Hultquist, R.: Theorems concerning eisenhart’s model ii. The Annals of Mathematical Statistics, 261–269, 1961.
20. Corbeil, R. and Searle, S.: A comparison of variance component estimators. Biometrics, 779–791, 1976.
21. Khuri, A. and Sahai, H.: Variance components analysis: a selective literature survey. International Statistical Review/Revue Internationale de Statistique, 279–300, 1985.
22. Searle, S., Casella, G., and McCulloch, C.: Variance Components. Wiley Online Library, 1992.
23. Richardson, A. and Welsh, A.: Asymptotic properties of restricted maximum likelihood (reml) estimates for hierarchical mixed linear models. Australian & New Zealand Journal of Statistics, 36(1):31–43, 1994.
24. Carrasco, J.: A generalized concordance correlation coefficient based on the variance components generalized linear mixed models for overdispersed count data. Biometrics, 66(3):897–904, 2010.

25. Lin, L., Hedayat, A., and Wu, W.: Statistical Tools for Measuring Agreement. Springer, 2012.
26. Lin, L.: Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence. Statistics in Medicine, 19(2):255–270, 2000.
27. King, T. and Chinchilli, V.: A generalized concordance correlation coefficient for continuous and categorical data. Statistics in Medicine, 20(14):2131–2147, 2001.
28. Patterson, H. and Thompson, R.: Maximum likelihood estimation of components of variance. Proceedings of the 8th International Biometric Conference, 197:207, 1974.
29. Cramer, H.: Mathematical models of statistics. Princeton Math. Series, 1946.
30. Bowling, L., Sageman, W., O'Connor, S., Cole, R., and Amundson, D.: Lack of agreement between measurement of ejection fraction by impedance cardiography versus radionuclide ventriculography. Critical Care Medicine, 21(10):1523, 1993.
31. Hauschke, D., Steinijans, V., and Pigeot, I.: Bioequivalence Studies in Drug Development: Methods and Applications, volume 47. John Wiley & Sons Inc, 2007.
32. Breslow, N. and Clayton, D.: Approximate inference in generalized linear mixed models. Journal of the American Statistical Association, 9–25, 1993.
33. Lin, X. and Breslow, N.: Bias correction in generalized linear mixed models with multiple components of dispersion. Journal of the American Statistical Association, 1007–1016, 1996.
34. Wolfinger, R. and O'Connell, M.: Generalized linear mixed models a pseudo-likelihood approach. Journal of Statistical Computation and Simulation, 48(3):233–243, 1993.
35. McCulloch, C. and Neuhaus, J.: Generalized Linear Mixed Models. Wiley Online Library, 2001.
36. Hardin, J. and Hilbe, J.: Generalized Linear Models and Extensions. Stata Corp, 2007.
37. Berntsen, J., Espelid, T., and Genz, A.: An adaptive algorithm for the approximate calculation of multiple integrals. ACM Transactions on Mathematical Software (TOMS), 17(4):437–451, 1991.

38. Bates, D.: lme4: Mixed-Effects Modeling with R. Unpublished, 2010.
39. Fears, T., Benichou, J., and Gail, M.: A reminder of the fallibility of the wald statistic. American Statistician, 50(3):226–227, 1996.
40. Turner, R., Omar, R., Yang, M., Goldstein, H., and Thompson, S.: A multilevel model framework for meta-analysis of clinical trials with binary outcomes. Statistics in Medicine, 19(24):3417–3432, 2000.
41. Moerbeek, M., Piersma, A., and Slob, W.: A comparison of three methods for calculating confidence intervals for the benchmark dose. Risk Analysis, 24(1):31–40, 2004.
42. FDA Guidance: Statistical approaches to establishing bioequivalence. Rockville, MD: US Food and Drug Administration Center for Drug Evaluation and Research, 2001.
43. Midha, K., Ormsby, E., Hubbard, J., McKay, G., Hawes, E., Gavalas, L., and McGilveray, I.: Logarithmic transformation in bioequivalence: application with two formulations of perphenazine. Journal of Pharmaceutical Sciences, 82(2):138–144, 1993.
44. Ahn, B. and Yim, D.: Comparison of parametric and bootstrap method in bioequivalence test. The Korean Journal of Physiology & Pharmacology: Official Journal of the Korean Physiological Society and the Korean Society of Pharmacology, 13(5):367, 2009.
45. Liu, X., Du, Y., Teresi, J., and Hasin, D.: Concordance correlation in the measurements of time to event. Statistics in Medicine, 24(9):1409–1420, 2005.
46. Barnhart, H., Song, J., and Lyles, R.: Assay validation for left-censored data. Statistics in Medicine, 24(21):3347–3360, 2005.

VITA

Education

- Ph.D. in Statistics, University of Illinois at Chicago, 2012.
- M.S. in Statistics, University of Illinois at Chicago, 2009.
- B.S. in Mathematics and Applied Mathematics, Fudan University, 2007.

Working Experience

- Intern/Research Assistant, Baxter Healthcare Corp., Round Lake, Illinois, 2009–2010.

Publication

- Cluster-based Regularized Sliced Inverse Regression for Forecasting Macroeconomic Variables, *Submitted for publication, with Zhihong Chen and Jie Yang.*
- Partial Sliced Inverse Regression for Quality-Relevant Multivariate Statistical Process Monitoring, *Accepted, with Zhijie Sun.*

Membership

- American Statistical Association (ASA)
- Institute of Mathematical Statistics (IMS)
- Society for Industrial and Applied Mathematics (SIAM)