INFERENCE FOR CLUSTERED MIXED OUTCOMES FROM A MULTIVARIATE GENERALIZED LINEAR MIXED MODEL

A Dissertation

by

HSIANG-CHUN CHEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Thomas E. Wehrly
Committee Members,	Jeffrey D. Hart
	Huiyan Sang
	Dominique Lord
Department Head,	Simon J. Sheather

August 2013

Major Subject: Statistics

ABSTRACT

Multivariate generalized linear mixed models (MGLMM) are used for jointly modeling the clustered mixed outcomes obtained when there are two or more responses repeatedly measured on each individual in scientific studies. The relationship among these responses is often of interest. In the clustered mixed data, the correlation could be present between repeated measurements either within the same observer or between different observers on the same subjects. This study proposes a series of indices, namely, intra, inter and total correlation coefficients, to measure the correlation under various circumstances of observations from a multivariate generalized linear model, especially for joint modeling of clustered count and continuous outcomes.

Bayesian methods are widely used techniques for analyzing MGLMM. The need for noninformative priors arises when there is insufficient prior information on the model parameters. Another aim of this study is to propose an approximate uniform shrinkage prior for the random effect variance components in the Bayesian analysis for the MGLMM. This prior is an extension of the approximate uniform shrinkage prior. This prior is easy to apply and is shown to possess several nice properties. The methods are illustrated in terms of both a simulation study and a case example.

I dedicate this dissertation to

my beloved parents, my sisters, and my husband, Yi-Chin,

for their endless love and support.

ACKNOWLEDGEMENTS

First and foremost, I want to express my deepest gratitude to my Ph.D. advisor, Dr. Thomas E. Wehrly, for his excellent guidance. He kindly encouraged me by showing his enthusiasm for research and positive attitude toward life. He will always be my role model.

I would like to sincerely thank my Ph.D. committee members, Dr. Jeffrey D. Hart, Dr. Huiyan Sang and Dr. Dominique Lord for their valuable suggestions. Special thanks goes to Dr. Michael Longnecker, who was always willing to help and give his best suggestions. I would also like to express my gratitude to all faculty and staff members of the Department of Statistics for their advice and help during my years at Texas A&M University.

I would like to thank my friends in the Department of Statistics, Taiwanese Student Association and Tzu Chi Collegiate Association at Texas A&M University for the great times that we have shared. I am particularly grateful to my best friends, Tanya Garcia, Fang-Yu Lin, and Rebecca Lai. They always encourage and support me with their best wishes. I also thank all my teachers and friends in Taiwan. This journey would have not been possible without their support and inspiration.

I will be forever thankful to my family with all my heart for their endless support and love. My parents raised me with love and care. My sisters have always been my best friends. Finally, I would like to thank my beloved husband, Yi-Chin, who always stood by me along this journey and has made these years the best of my life. I love you all dearly.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE Page

1. INTRODUCTION

Clustered mixed outcomes arise in scientific studies such as longitudinal trials when there is more than one response repeatedly measured on each individual. Methods have been proposed for modeling the clustered mixed outcomes. The multivariate generalized linear mixed model (MGLMM) is one of the most widely used models for accommodating these measurements when they are assumed to independently follow distributions in the exponential family, depending on fixed effects and subject-specific correlated random effects (An et al., 2009; Coull and Agresti, 2000; Gueorguieva, 2001; Gueorguieva and Agresti, 2001; McCulloch, 2008). Approaches such as the adaptive Gaussian-Hermite quadrature, Monte Carlo EM algorithm, generalized estimating equations approach, and penalized quasi-likelihood have been developed for maximum likelihood estimation in MGLMM.

Assessing agreement between multiple measurements taken by several observers on the same subject is of interest because it evaluates the interaction of different observers and whether one can be substituted by the other. The two observers can be interchangeable if they have perfect correlation and perfect agreement.

Nevertheless, it is very common that multivariate outcomes with discrete and continuous components are observed. Measurements from different observers or methods under this situation usually have extremely different values. For example, the number of tumors and the blood pressure of patients in a clinical study are in discrete and continuous scales, respectively. Since measurements are taken on different scales, it is not appropriate to use the intra correlation coefficient or the concordance correlation coefficient to measure the agreement between observers. The problem arises when there is a need to assess the interchangeability or relationship of observers using extremely different scales, especially in qualitative and quantitative scales, respectively. Instead of agreement, the correlation between the mixed outcomes is of particular interest. Various types of correlation could be present, including the correlation between measurements taken by the same observers on different subjects and the correlation between measurements from different observers for a given subject. These relationships may attribute to the intra-observer variability or inter-observer variability. Furthermore, the measurements might not follow a normal distribution, which violates the assumption for inference using the concordance correlation coefficient. Situations may arise in which repeated measurements are taken for each method, such as in clinical trials and longitudinal studies. One of the main goals of this study is to develop a series of correlation coefficients to investigate the correlation between clustered mixed measurements under multivariate generalized linear mixed model. Multivariate clustered mixed outcomes are considered in this study. Correlations between measurements in the multivariate clustered mixed outcomes are assumed to be present either on the same observers or on different observers within subjects when each of the different observers produce replicated measurements on each subject. Subject-specific models for discrete and continuous measurements in the exponential family are accommodated using the MGLMM. Assume each of the different observers produces replicated measurements on each subject. Based on this model, three indices for measuring the consistency between clustered mixed outcomes are proposed, including the intra correlation coefficient (intra-CC), the inter correlation coefficient (inter-CC), and the total correlation coefficient (total-CC). The intra-CC measures the correlation among multiple measurements from the same observer. The inter-CC coefficient measures the correlation among multiple measurements from different observers based on the average of multiple measurements. The total-CC measures the correlation among multiple measurements from different observers based on individual measurements.

Generalized linear mixed models can be viewed as hierarchical models containing two stages. Therefore, a Bayesian approach (Gelman, 2006; Tierney, 1994; Zeger and Karim, 1991) has been widely used in estimating the joint posterior distributions of the fixed-effect parameters and the variance components of the random effects. Several assumptions for the prior distribution on the fixed effect parameters and the variance components of random effects have been studied. The standard noninformative prior, or Jeffreys prior (Tiao and Tan, 1965), is one of the most widely used assumption in the Bayesian approach. The drawback for using the Jeffreys prior is that it may lead to an improper joint posterior distribution for fixed effects and variance component of the random effect (Ibrahim and Laud, 1991; Natarajan and McCulloch, 1995).

For univariate generalized linear mixed model, Natarajan and Kass (2000) introduced the approximate uniform shrinkage prior as an alternative prior for the variance component of the random effects. The main idea of the approximate uniform shrinkage prior is that the weight of the prior mean used in the approximate shrinkage estimate is assumed to be componentwise uniformly distributed. Then using the transformation theorem, we can find the distribution of the variance structure of random effect. This prior has been shown to possess several desirable properties.

When the clustered mixed outcomes are considered, MGLMM is applied and the random effects are assumed to follow a multivariate normal distribution. The inverse Wishart distribution is one of the most widely used prior distribution for the covariance matrix of the random effects since the inverse Wishart distribution is the conjugate prior of the multivariate normal distribution (Dunson, 2000). However, the estimation is very sensitive to the choice of the scale matrix in the prior distribution. Therefore, the need of noninformative priors arises when there is insufficient prior information on the model parameters. Another goal of this dissertation is to propose an approximate uniform shrinkage prior for the random effect variance components in the Bayesian analysis for the MGLMM.

The rest of this dissertation is organized as follows. In Chapter 2, we introduce the multivariate generalized linear mixed model and give a brief literature review. A series of indices, namely, intra, inter and total correlation coefficients, to measure the correlation under various circumstances in a multivariate generalized linear model, especially for joint modeling of clustered count and continuous outcomes, is developed in Chapter 3. We demonstrate the methodology with a simulation study. A case example is provided to illustrate the use of these proposed indices. In Chapter 4, an approximate uniform shrinkage prior for a multivariate generalized linear mixed model is introduced. The results of simulation studies are compared with those based on other widely used priors. Some concluding remarks are given in Chapter 5. Mathematical details are given in the Appendix.

2. LITERATURE REVIEW

The univariate generalized linear mixed model (GLMM) and the multivariate generalized linear mixed model (MGLMM) are briefly reviewed in Sections 2.1 and 2.2. The Bayesian methods for GLMM and MGLMM are also discussed. The literature review of previous research about the two studies in this dissertation are provided in Sections 2.3 and 2.4.

2.1 Generalized Linear Mixed Model

Consider data composed of N subjects and T_i repeated measurements within the i-th subject. Let y_{it} be the t-th univariate measurement on the i-th subject. Conditional on a subject-specific random effect b_i, $\{y_{i1}, \cdots, y_{iT_i}\}$ is assumed to independently follow a distribution with density in the exponential family

$$f(y_{it}|b_i) \propto \exp\left\{\frac{y_{it}\theta_{it} - a(\theta_{it})}{\phi}\right\}$$

where the dispersion parameter ϕ is assumed known, and θ_{it} is the canonical parameter. The conditional mean is $\mu_{it} = \mathrm{E}(y_{it}|b_i) = a'(\theta_{it})$, and the conditional variance is $\upsilon_{it} = \mathrm{Var}(y_{it}|b_i) = \phi a''(\theta_{it})$. Also, assume that the conditional mean is related to the linear form of predictors by the link function :

$$g(\mu_{it}) = x_{it}^T\beta + z_{it}^T b_i = \beta_1 x_{1,it} + \cdots + \beta_p x_{p,it} + b_{i1} z_{1,it} + \cdots + b_{iq} z_{q,it},$$

where $g(\cdot)$ is a monotonic differentiable link function, $x_{it} = (x_{1,it}, \cdots, x_{p,it})^T$ is a vector of covariates, $\beta = (\beta_1, \cdots, \beta_p)^T$ is a vector of fixed effect parameters, and $z_{it} = (z_{1,it}, \cdots, z_{q,it})^T$ is a vector of covariates corresponding to the random effect $b_i =$

$(b_{i1}, \cdots, b_{iq})^T$. The random effect b_i is shared by repeated measurements within the same subject. Assume that b_i has a multivariate normal distribution with mean 0 and covariance matrix D. The aforementioned model is known as the generalized linear mixed model, or GLMM (Breslow and Clayton, 1993; Zeger and Karim, 1991). The normal linear mixed model is a special case of GLMM when y_{it} independently follows a normal distribution conditional on the random effects b_i and the link function is the identity function.

The likelihood for the parameter β and D is

$$L(\beta, D) \propto \prod_{i=1}^{N} |D|^{-\frac{1}{2}} \int \exp\left\{ \sum_{t=1}^{T_i} \frac{y_{it}\theta_{it} - a(\theta_{it})}{\phi} - \frac{1}{2} b_i^T D^{-1} b_i \right\} \pi(\beta) db_i.$$

However, the maximum likelihood estimates cannot be simplified or evaluated in closed form. Because of the complexity of the likelihood in GLMM, several numerical integration methods such as Gauss-Hermite quadrature or the Bayesian approach have been proposed for analyzing data in GLMM.

2.2 Multivariate Generalized Linear Mixed Model

The multivariate generalized linear mixed model (MGLMM) can accommodate clustered data when measurements are repeatedly observed from two or more observers on each subject. Consider data comprising N subjects and T_i repeated measurements within the i-th subject, measured by L observers. Measurements for different subjects are assumed to be independent, and the numbers of replications differ from subject to subject. Let the measurement for the i-th subject be $Y_i = (Y_{i1}^T, \cdots, Y_{iL}^T)^T$, where $Y_{ij}^T = (Y_{ij1}, \cdots, Y_{ijT_i})^T$ are repeated measurements from the j-th observer, $j = 1, \ldots, L$. Assume that given the random effects b_i, $\{Y_{ij1}, \cdots, Y_{ijT_i}\}$ are conditionally independent given observer j and subject i, and

Y_{ijt} has density $f_j(\cdot)$ in the exponential family. Let $\mu_{ij} = (\mu_{ij1}, \cdots, \mu_{ijT_i})^T$ be the conditional mean of $Y_{ijt} = (Y_{ij1}, \cdots, Y_{ijT_i})^T$ given the random effects b_i. Covariates are $X = (X_1, \cdots, X_N)^T$, where $X_i = (x_{1,i11}, \cdots, x_{p_1,i1T_i}, \cdots, x_{1,iL1}, \cdots, x_{p_L,iLT_i})^T$ are the covariates for the i-th subject. Assume the model has correlated random effects which follow a multivariate normal distribution. The multivariate generalized linear model is defined by the following :

$$
\begin{aligned}
&Y_{i1t}|b_{i1} \text{ is from a particular distribution } F_1 \text{ in the exponential family} \\
&\qquad \text{with mean } \mu_{i1t} \text{ and density } \exp\left\{\frac{y_{i1t}\theta_{i1t} - a_1(\theta_{i1t})}{\phi_1}\right\} \\
&\qquad \vdots \\
&Y_{iLt}|b_{iL} \text{ is from a particular distribution } F_L \text{ in the exponential family} \\
&\qquad \text{with mean } \mu_{iLt} \text{ and density } \exp\left\{\frac{y_{iLt}\theta_{iLt} - a_L(\theta_{iLt})}{\phi_L}\right\} \\
&g_1(\mu_{i1t}) = x_{i1t}^T\beta_1 + z_{i1t}^T b_{i1} \\
&\qquad \vdots \\
&g_L(\mu_{iLt}) = x_{iLt}^T\beta_L + z_{iLt}^T b_{iL} \\
&\quad b_i = (b_{i1}, \cdots b_{iL}) \sim \text{ iid multivariate normal}(0, D)
\end{aligned}
$$

where for $j = 1, \ldots, L$, the dispersion parameters ϕ_j is assumed known; θ_{ijt} is the canonical parameter; $g_j(\cdot)$ is the link function; $x_{ijt} = (x_{1,ijt}, \cdots, x_{p_j,ijt})^T$ is a vector of covariates; $\beta_j = (\beta_{j1}, \cdots, \beta_{jp_j})^T$ is a vector of fixed effect parameters; $z_{ijt} = (z_{1,ijt}, \cdots, z_{q,ijt})^T$ is a vector of covariates corresponding to the random effects $b_{ij} = (b_{1,ij}, \cdots, b_{q,ij})^T$; $D = [\sigma_{ij}]_{i=1,\cdots,L;j=1,\cdots,L}$ is the covariance matrix. The conditional means are $\mu_{ijt} = \mathrm{E}(y_{ijt}|b_{ij}) = a_j'(\theta_{ijt})$, and the conditional variances are $v_{ijt} = \mathrm{Var}(y_{ijt}|b_{ij}) = \phi_j a_j''(\theta_{ijt})$.

Similar to GLMM, the maximum likelihood estimates of MGLMM cannot be solved in closed form. Various approximate methods have been developed, such as the Bayesian approach, penalized quasi-likelihood, Monte Carlo EM algorithms, and maximum likelihood estimation, for model fitting of the MGLMM (Gueorguieva, 2001).

2.3 Assessing Correlation in Generalized Linear Mixed Model

A number of measurements of agreement among multiple measurements taken by several observers or methods have been proposed. Cohen's kappa statistics (Cohen, 1960) and weighted kappa statistics (Cohen, 1968) are used for assessing agreement among observers when the measurements are categorical. The intra correlation coefficient (ICC) (Pearson et al., 1901) and concordance correlation coefficient (CCC) (Lin, 1989) are two of the most widely applied indices for assessing the agreement between observers for continuous data. The ICC has been shown to be equivalent to a particular specification of the CCC (Carrasco and Jover, 2003). The concordance correlation coefficient was used to measure the agreement between two variables, Y_1 and Y_2 and was defined as

$$\rho_c = 1 - \frac{E\{(Y_1 - Y_2)^2\}}{\sigma_{Y_1}^2 + \sigma_{Y_2}^2 + (\mu_{Y_1} - \mu_{Y_2})^2} = \frac{2\rho\sigma_{Y_1}\sigma_{Y_2}}{\sigma_{Y_1}^2 + \sigma_{Y_2}^2 + (\mu_{Y_1} - \mu_{Y_2})^2}$$

where μ_{Y_1} and μ_{Y_2} are the means for the two variables, $\sigma_{Y_1}^2$ and $\sigma_{Y_2}^2$ are the corresponding variances, and ρ is the correlation coefficient between the two variables.

The CCC was extended to application on categorical data (King and Chinchilli, 2001) as

$$\rho_g = \frac{\left\{E_{F_{Y_1}F_{Y_2}} g(Y_1 - Y_2) - E_{F_{Y_1}F_{Y_2}} g(Y_1 + Y_2)\right\} - \left\{E_{F_{Y_1 Y_2}} g(Y_1 - Y_2) - E_{F_{Y_1 Y_2}} g(Y_1 + Y_2)\right\}}{E_{F_{Y_1}F_{Y_2}} g(Y_1 - Y_2) - E_{F_{Y_1}F_{Y_2}} g(Y_1 + Y_2) + \frac{1}{2} E_{F_{Y_1 Y_2}} \{g(2Y_1) + g(2Y_2)\}}$$

where $g(\cdot)$ is a convex function of distance defined on the real line and $g(Y_1 - Y_2)$ is an integrable function with respect to $F_{Y_1 Y_2}$.

An extended concordance correlation coefficient (Barnhart et al., 2002; King and Chinchilli, 2001) was defined to assess the amount of agreement among more than two observers. The CCC and ICC were both extended for assessing the repeated measurements such as longitudinal data in a clinical study (Carrasco et al., 2009; King et al., 2007a,b).

The intra-CCC, inter-CCC and total-CCC (Barnhart et al., 2005; Lin et al., 2007) were proposed to measure the intra, inter and total agreement among replicated measurements by using several observers, respectively. Consider the model

$$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijt}$$

where y_{ijt}, $t = 1, \cdots, T$, are replicated measurements given the subject i and the observer j. They proposed a series of indices for assessing agreement, precision and accuracy when there were multiple observers each with multiple readings. Among them, the intra, inter, and total precisions were defined as

$$\begin{aligned}
\rho_j^{intra} &= \frac{\mathrm{Cov}(y_{ijt}, y_{ijt'})}{\sqrt{\mathrm{Var}(y_{ijt})\,\mathrm{Var}(y_{ijt'})}} = \frac{\sigma_\alpha^2 + \sigma_\gamma^2}{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_e^2} \\
\rho^{inter} &= \frac{\mathrm{Cov}(\bar{y}_{ij\cdot}, \bar{y}_{ij'\cdot})}{\sqrt{\mathrm{Var}(\bar{y}_{ij\cdot})\,\mathrm{Var}(\bar{y}_{ij'\cdot})}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_e^2/m} \\
\rho^{total} &= \frac{\mathrm{Cov}(y_{ijt}, y_{ij't'})}{\sqrt{\mathrm{Var}(y_{ijt})\,\mathrm{Var}(y_{ij't'})}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_e^2}
\end{aligned}$$

However, inference based on the above measures of correlation assume that both random effects and the residuals are normally distributed, which is sometimes not appropriate. The CCC was extended (Carrasco, 2010; Carrasco and Jover, 2005) for assessing the agreement for data from any distribution of the exponential family in terms of the generalized linear mixed model. Let b_{ij} be the random effects consisting of variabilities such as the subject, observer, and subject by observer interaction effects. The conditional distribution of $Y = \{y_{ijt}\}$ given the covariates X and b, are assumed to follow a distribution from the exponential family. The conditional mean of Y given b is $\mathrm{E}(Y_{ijt}|b_{ij}) = \mu_{ij} = g^{-1}(\lambda_i + \alpha_i + \beta_j + \gamma_{ij})$ and the conditional variance of Y given b is $\mathrm{Var}(Y_{ijt}|b_{ij}) = \phi h(\mu_{ij})$. Also assume that the random effect b follows a multivariate normal distribution. The intra, inter and total generalized concordance correlation coefficients are defined as

$$\begin{aligned}
\rho_{GCCC,j}^{intra} &= \frac{\mathrm{Cov}(y_{ijt}, y_{ijt'})}{\mathrm{Var}(y_{ijt})} = \frac{\mathrm{Var}(\mu_{ij})}{\mathrm{Var}(\mu_{ij}) + \mathrm{E}\{\phi h(\mu_{ij})\}} \\
\rho_{GCCC}^{inter} &= \frac{\mathrm{Cov}(\bar{y}_{ij\cdot}, \bar{y}_{ij'\cdot})}{\mathrm{Var}(\bar{y}_{ij\cdot})} = \frac{\mathrm{Cov}(\mu_{ij}, \mu_{ij'})}{\mathrm{Var}(\mu_{ij}) + \mathrm{E}\{\phi h(\mu_{ij})\}/m} \\
\rho_{GCCC}^{total} &= \frac{\mathrm{Cov}(y_{ijt}, y_{ij't'})}{\mathrm{Var}(y_{ijt})} = \frac{\mathrm{Cov}(\mu_{ij}, \mu_{ij'})}{\mathrm{Var}(\mu_{ij}) + \mathrm{E}\{\phi h(\mu_{ij})\}}
\end{aligned}$$

All indices mentioned above can be used to measure agreement and evaluate whether the observers or methods are interchangeable when the observers are assumed to follow identical distributions. However, for clustered mixed outcome data, measurements may be taken on extremely different scales. As a result, indices are needed to measure the consistency among clustered mixed outcomes.

2.4 Bayesian Method for the Generalized Linear Mixed Model

The Bayesian approach is a very popular method used in the analysis of the GLMM. The GLMM can be thought as a two-stage hierarchical model. The measurements conditional on given subject-specific random effects are assumed to follow a particular distribution from the exponential family at the first stage, while the random effects are assumed to follow a multivariate normal distribution at the second stage. There is a need of the specification of the prior distribution for the fixed effect parameters β and the random effect variance components D. We assume that β and D are independent of each other in this study. When there is no subjective prior information about β, the most widely used noninformative prior assumption for the fixed effect coefficient is the improper uniform distribution, which will be used throughout this study. However, various noninformative prior distributions for the variance components of the random effects, D, have been suggested in the previous literature, including Jeffreys prior, a proper conjugate prior and the approximate uniform shrinkage prior.

The standard noninformative prior, or a Jeffreys prior, $\pi(D) \propto |D|^{-\frac{q+1}{2}}$ (Tiao and Tan, 1965; Zeger and Karim, 1991) is one of the most widely used prior assumptions in the Bayesian approach. It is obtained by applying Jeffreys rule to the second-stage random effect distribution. The posterior distribution of D corresponding to a Jeffreys prior follows an inverse Wishart distribution with scale matrix $S = \sum_{i=1}^{N} b_i b_i^T$ and degrees of freedom N, $IW(N, S)$. The advantage of choosing the Jeffreys prior is that the posterior distribution is specified and easy to implement, however the disadvantage is that it may lead to an improper joint posterior distribution for β and D (Ibrahim and Laud, 1991; Natarajan and McCulloch, 1995).

Another popular choice of the prior distribution is a proper conjugate prior. The inverse Wishart distribution with scale matrix Ψ and degrees of freedom λ, $IW(\lambda, \Psi)$, is a conjugate prior for D. Since a univariate specialization of the inverse Wishart distribution is the inverse gamma distribution, the prior reduces to an inverse gamma distribution when the dimension of D is one. The most popular choice is to set $\lambda = q$ and $\Psi = qD^0$, where D^0 is the prior guess of D (Spiegelhalter et al., 1996). The advantage of this conjugate prior is that it is computationally easy to implement, while the disadvantage is that the estimation results can be very sensitive to the choices of D^0 (Natarajan and Kass, 2000).

Natarajan and Kass (2000) introduced the approximate uniform shrinkage prior as an alternative prior for D. It is a generalization of the uniform shrinkage prior proposed by Strawderman (1971). The main idea in the approximate uniform shrinkage prior is induced by placing a componentwise uniform distribution on the weight given to the prior mean in the approximate shrinkage estimate of the random effects. Specifically, the approximate shrinkage estimate $\hat{b}_i$ has the form

$$\hat{b}_i = DZ_i^T(W_i^{-1} + Z_iDZ_i^T)^{-1}(y_i^* - X\hat{\beta}) = S_i 0_q + (1 - S_i)DZ_i^T W_i(y_i^* - X\hat{\beta}),$$

where y_i^* is a working dependent variable and W_i is the GLM weight matrix obtained by replacing b^* with 0. The weight $S_i = (D^{-1} + Z_i^T W_i Z_i)^{-1} Z_i^T W_i Z_i$ is a function of D and varies with i. Thus by replacing the individual weights with the average across the cluster, they define the overall weight matrix

$$S = \left(D^{-1} + \frac{1}{N}\sum_{i=1}^{N} Z_i^T W_i Z_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} Z_i^T W_i Z_i\right).$$

Assume S is componentwise uniformly distributed, then using the transformation theorem, we can find the distribution of D,

$$\pi_{us}(D) \propto \left| I + \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^T W_i Z_i\right) D \right|^{-q-1}$$

which is the approximate uniform shrinkage prior. The advantage of the approximate uniform shrinkage prior is that it is proper, the corresponding posterior distribution under some situations is proper, and the calculation is quite simple.

The Bayesian approach can also be used for MGLMM. The fixed effect parameters β and the random effect variance components D are assumed to be a priori independent. The prior distribution assumptions of the fixed effect parameters β and the variance of the random effects D are needed in the Bayesian approach. One of the most widely used prior distributions for β is uniform. However, there are several choices of prior distribution for D.

One of the most widely used prior distribution assumptions of the variance components of random effects is inverse Wishart distribution (Dunson, 2000). Assume the prior distribution of D is inverse Wishart(m, Ψ), then the posterior distribution of D is inverse Wishart$(m + N, \Psi + S)$, where $S = \sum_{i=1}^{N} b_i b_i^T$. Thus, it is a conjugate prior for the covariance matrix of a multivariate normal distribution. However, the estimation is very sensitive to the choice of the scale matrix Ψ in the prior inverse Wishart distribution. Therefore, the need of an objective prior arises in the multivariate case. In this dissertation, we will modify and extend the approximate uniform shrinkage prior proposed by Natarajan and Kass for the MGLMM.

3. ASSESSING CORRELATION OF CLUSTERED MIXED OUTCOMES FROM A MULTIVARIATE GENERALIZED LINEAR MIXED MODEL

3.1 Introduction

The classic concordance correlation coefficient measures the agreement between two variables. In recent studies, concordance correlation coefficients have been generalized to deal with responses from a distribution from the exponential family using the univariate generalized linear mixed model. Multivariate data arise when responses on the same unit are measured repeatedly by several observers. The relationship among these responses is often of interest. In clustered mixed data, the correlation could be present between repeated measurements either within the same observer or between different observers on the same subjects. Indices for measuring such association are needed. Therefore, we propose a series of indices, namely, intra, inter and total correlation coefficients to measure the correlation under various circumstances in a multivariate generalized linear model, especially for joint modeling of clustered count and continuous outcomes. The proposed indices are natural extensions of the concordance correlation coefficient.

This chapter is structured as follows. A series of the measurements for the multivariate generalized linear mixed model for joint modeling of clustered mixed outcomes are proposed in Section 3.2. Extensions of the proposed correlations among multiple observers are defined in Section 3.3. In Section 3.4, we show examples of bivariate and trivariate models. Simulation studies of evaluating the proposed correlation are included in Section 3.5. The relationship among the proposed indices is investigated. We illustrate the application of the proposed indices using one case example in Section 3.6. Finally the conclusions are stated in Section 3.7.

3.2 Method

The scenario of bivariate clustered mixed data from two different observers for N subjects is considered. An interesting case is where the measurement from one observer is discrete and the measurement from the other observer is continuous. A multivariate generalized linear model can be used to fit bivariate clustered mixed data with the assumption of joint multivariate random effects. Conditional on random effects, y_{i1t} and y_{i2t} are assumed independent. The multivariate generalized linear model is defined as follows :

$$
\begin{aligned}
& Y_{i1t}|b_{i1} \text{ is from a particular distribution } F_1 \text{ in the exponential family} \\
& \qquad \text{with mean } \mu_{i1t} \text{ and density } \exp\left\{\frac{y_{i1t}\theta_{i1t} - a_1(\theta_{i1t})}{\phi_1}\right\} \\
& Y_{i2t}|b_{i2} \text{ is from a particular distribution } F_2 \text{ in the exponential family} \\
& \qquad \text{with mean } \mu_{i2t} \text{ and density } \exp\left\{\frac{y_{i2t}\theta_{i2t} - a_2(\theta_{i2t})}{\phi_2}\right\} \\
& g_1(\mu_{i1t}) = x_{i1t}^T\beta_1 + z_{i1t}^T b_{i1} \\
& g_2(\mu_{i2t}) = x_{i2t}^T\beta_2 + z_{i2t}^T b_{i2} \\
& \quad b_i = (b_{i1}, b_{i2}) \sim \text{ iid multivariate normal}(0, D)
\end{aligned}
$$

where the dispersion parameters ϕ_1 and ϕ_2 are assumed known, θ_{i1t} and θ_{i2t} are the canonical parameters, $g_1(\cdot)$ and $g_2(\cdot)$ are link functions, $x_{i1t} = (x_{1,i1t}, \cdots, x_{p_1,i1t})^T$ and $x_{i2t} = (x_{1,i2t}, \cdots, x_{p_2,i2t})^T$ are vectors of covariates, $\beta_1 = (\beta_{11}, \cdots, \beta_{1p_1})^T$ and $\beta_2 = (\beta_{21}, \cdots, \beta_{2p_2})^T$ are vectors of fixed effect parameters, and $z_{i1t} = (z_{1,i1t}, \cdots, z_{q,i1t})^T$ and $z_{i2t} = (z_{1,i2t}, \cdots, z_{q,i2t})^T$ are vectors of covariates corresponding to the random effects $b_{i1} = (b_{1,i1}, \cdots, b_{q,i1})^T$ and $b_{i2} = (b_{1,i2}, \cdots, b_{q,i2})^T$.

The bivariate generalized linear mixed model is equivalent to two separate univariate GLMMs when the correlations between the random effects are zero. The multivariate Rasch model and the multivariate binomial-logit normal model are both special cases of the above model (Gueorguieva and Agresti, 2001). Based on this model, the marginal means and variances of the measurement Y_{ijt} are found to be

$$
\begin{aligned}
\mathrm{E}(Y_{ijt}) &= \mathrm{E}\{\mathrm{E}(Y_{ijt}|b_i)\} = \mathrm{E}(\mu_{ijt}) \\
\mathrm{Var}(Y_{ijt}) &= \mathrm{E}\{\mathrm{Var}(Y_{ijt}|b_i)\} + \mathrm{Var}\{\mathrm{E}(Y_{ijt}|b_i)\} = \mathrm{E}\{\phi_j h_j(\mu_{ijt})\} + \mathrm{Var}(\mu_{ijt})
\end{aligned}
$$

where ϕ_j is the dispersion parameter and $h_j(\cdot)$ is the corresponding variance function for $F_j(\cdot)$. Based on this model, a series of correlation coefficients are defined in the next few subsections.

3.2.1 Intra correlation coefficient (Intra-CC)

The intra correlation coefficient (abbreviated intra-CC) measures the linear relationship among multiple measurements from a given observer on a subject. In other words, it assesses the intra-observer correlation. For a given observer j, the intra-CC is defined as the correlation between any two replications t and t' measured by the same observer j on a subject. The intra-CC of the j-th observer and i-th subject is written as

$$
\rho^{intra,i,j} = \frac{\mathrm{Cov}(Y_{ijt}, Y_{ijt'})}{\sqrt{\mathrm{Var}(Y_{ijt})\,\mathrm{Var}(Y_{ijt'})}}.
$$

In the proposed model, the covariance of different replicates from j-th observer on i-th subject is

$$
\begin{aligned}
&\mathrm{Cov}(Y_{ijt}, Y_{ijt'}) \\
&= \mathrm{E}\{\mathrm{Cov}(Y_{ijt}, Y_{ijt'}|b_i)\} + \mathrm{Cov}\{\mathrm{E}(Y_{ijt}|b_i), \mathrm{E}(Y_{ijt'}|b_i)\} \\
&= \mathrm{Cov}(\mu_{ijt}, \mu_{ijt'})
\end{aligned}
$$

since for each observer j, Y_{ijt} and $Y_{ijt'}$ are conditionally independent given the random effects b_i. As a result, the intra-CC for the j-th observer can be expressed using the marginal variance and covariance as

$$\rho^{intra,i,j} = \frac{\mathrm{Cov}(\mu_{ijt}, \mu_{ijt'})}{\sqrt{[\mathrm{E}\{\phi_j h_j(\mu_{ijt})\} + \mathrm{Var}(\mu_{ijt})]\,[\mathrm{E}\{\phi_j h_j(\mu_{ijt'})\} + \mathrm{Var}(\mu_{ijt'})]}}.$$

Note that both $\mathrm{Cov}(\mu_{ijt}, \mu_{ijt'})$ and $\mathrm{E}\{\phi_j h_j(\mu_{ijt})\} + \mathrm{Var}(\mu_{ijt})$ can be expressed in terms of $\mathrm{E}(\mu_{ijt})$ and $\mathrm{E}(\mu_{ijt'})$, hence $\rho^{intra,i,j}$ defined above can be expressed as a ratio of two functions: K_N^{intra} and K_D^{intra}, or,

$$\rho^{intra,i,j} = \frac{K_N^{intra}\{\mathrm{E}(\mu_{ijt}), \mathrm{E}(\mu_{ijt'})\}}{K_D^{intra}\{\mathrm{E}(\mu_{ijt}), \mathrm{E}(\mu_{ijt'})\}}.$$

Since $\mathrm{E}(\mu_{ijt})$ depends on the covariates X, $\mathrm{E}(\mu_{ijt})$ varies not only from subject to subject but also from replicate to replicate, and so does the intra-CC. Therefore, an overall intra-CC is obtained by replacing $\mathrm{E}(\mu_{ijt})$ and $\mathrm{E}(\mu_{ijt'})$ with their marginal expectation over X, $\mu_j^* = E_X\{E(\mu_{ijt})\}$. That is,

$$\rho^{intra,j} = \frac{K_N^{intra}\left(\mu_j^*, \mu_j^*\right)}{K_D^{intra}\left(\mu_j^*, \mu_j^*\right)}.$$

Based on this model, the marginal expectations over X can be shown equal to

$$\mu_j^* = \mathrm{E}_X\{\mathrm{E}(\mu_{ijt})\} = \mathrm{E}_X\left\{g_j^{-1}(x_{ijt}^T\beta_j + z_{ijt}^T b_{ij})\right\}$$

If the log links are used, only random intercept is considered ($q = 1$), and the covariates X are independently and identically distributed having standard normal distributions, then $\mu_j^* = \exp(\beta_{j1}^2/2 + \cdots + \beta_{jp_j}^2/2 + \sigma_{b_j}^2/2)$. If a covariate is a categorical variable, then the correlations should be calculated separately over the levels of the variable.

3.2.2 Inter correlation coefficient (Inter-CC)

The inter correlation coefficient (abbreviated inter-CC) measures the linear relationship among different observers based on the average of replicated measurements when more than one measurement are observed from a subject. The inter-CC is defined as the correlation between the averages of multiple replicated measurements from each observer on the same subject and is used to measure the inter-observer correlation. Let $\overline{Y}_{ij\cdot}$ denote the arithmetic mean of replicated measurements from the i-th subject given by the j-th observer. Then the inter-CC of the ith subject is defined as

$$\rho^{inter,i} = \frac{\mathrm{Cov}(\overline{Y}_{i1\cdot}, \overline{Y}_{i2\cdot})}{\sqrt{\mathrm{Var}(\overline{Y}_{i1\cdot})\,\mathrm{Var}(\overline{Y}_{i2\cdot})}}$$

where $\mathrm{Cov}(\overline{Y}_{i1\cdot}, \overline{Y}_{i2\cdot})$ is the marginal covariance of averages of replicated measurements taken from two different observers on the same subject, and $\mathrm{Var}(\overline{Y}_{i1\cdot})$ and $\mathrm{Var}(\overline{Y}_{i2\cdot})$ are the marginal variances of the average of replicated measurements taken from first and second observer on the i-th subject, respectively. The inter-CC depends on the number of replications since it is a measurement in terms of the averages of replicated measurements taken by each observer.

Based on the proposed model, we know that the marginal covariance of the averages is

$$\begin{aligned}
\mathrm{Cov}(\overline{Y}_{i1\cdot}, \overline{Y}_{i2\cdot}) &= \frac{1}{T_i^2}\sum_{t=1}^{T_i}\sum_{t'=1}^{T_i}\mathrm{Cov}(Y_{i1t}, Y_{i2t'}) \\
&= \frac{1}{T_i^2}\sum_{t=1}^{T_i}\sum_{t'=1}^{T_i}\left[\mathrm{E}\left\{\mathrm{Cov}(Y_{i1t}, Y_{i2t'}|b_i)\right\} + \mathrm{Cov}\left\{\mathrm{E}(Y_{i1t}|b_i), \mathrm{E}(Y_{i2t'}|b_i)\right\}\right] \\
&= \frac{1}{T_i^2}\sum_{t=1}^{T_i}\sum_{t'=1}^{T_i}\mathrm{Cov}(\mu_{i1t}, \mu_{i2t'})
\end{aligned}$$

since Y_{i1t} and $Y_{i2t'}$ are assumed to be conditionally independent given the random effects. The marginal variance of the average is

$$\begin{aligned}
\mathrm{Var}(\overline{Y}_{ij\cdot}) &= \mathrm{E}\{\mathrm{Var}(\overline{Y}_{ij\cdot}|b_i)\} + \mathrm{Var}\{\mathrm{E}(\overline{Y}_{ij\cdot}|b_i)\} \\
&= \frac{\sum_{t=1}^{T_i} \mathrm{E}\{\phi_j h_j(\mu_{ijt})\}}{T_i^2} + \frac{\mathrm{Var}\left(\sum_{t=1}^{T_i} \mu_{ijt}\right)}{T_i^2}
\end{aligned}$$

where $j = 1, 2$. Combining the above equations gives

$$\rho^{inter,i} = \frac{\sum_{t=1}^{T_i}\sum_{t'=1}^{T_i} \mathrm{Cov}(\mu_{i1t}, \mu_{i2t'})}{\sqrt{\left[\sum_{t=1}^{T_i} \mathrm{E}\{\phi_1 h_1(\mu_{i1t})\} + \mathrm{Var}\left(\sum_{t=1}^{T_i} \mu_{i1t}\right)\right]\left[\sum_{t'=1}^{T_i} \mathrm{E}\{\phi_2 h_2(\mu_{i2t'})\} + \mathrm{Var}\left(\sum_{t'=1}^{T_i} \mu_{i2t'}\right)\right]}}.$$

It can be shown that the inter-CC is a ratio of functions of $\sum_{t=1}^{T_i} \mathrm{E}(\mu_{i1t})$ and $\sum_{t=1}^{T_i} \mathrm{E}(\mu_{i2t})$. That is,

$$\rho^{inter,i} = \frac{K_{N,1,2}^{inter}\left\{\sum_{t=1}^{T_i} \mathrm{E}(\mu_{i1t}), \sum_{t=1}^{T_i} \mathrm{E}(\mu_{i2t})\right\}}{K_{D,1,2}^{inter}\left\{\sum_{t=1}^{T_i} \mathrm{E}(\mu_{i1t}), \sum_{t=1}^{T_i} \mathrm{E}(\mu_{i2t})\right\}}$$

which depends on the covariates X as well as the numbers of replicates for each subject.

An overall inter-CC is obtained by replacing $\sum_{t=1}^{T_i} \mathrm{E}(\mu_{i1t})$ and $\sum_{t=1}^{T_i} \mathrm{E}(\mu_{i2t})$ with $T^*\mu_1^*$ and $T^*\mu_2^*$, where $T^* = \frac{\sum_{i=1}^{N} T_i}{N}$ and $\mu_j^* = \mathrm{E}_X\{\mathrm{E}(\mu_{ijt})\}$ is the marginal expectation over X defined in the previous subsection, $j = 1, 2$. In other words, the overall inter-CC is

$$\rho^{inter} = \frac{K^{inter}_{N,1,2}\left(T^*\mu_1^*, T^*\mu_2^*\right)}{K^{inter}_{D,1,2}\left(T^*\mu_1^*, T^*\mu_2^*\right)}.$$

If all subjects have the same number of replicates, T, for all i, then $T^* = T_i = T$.

3.2.3 Total correlation coefficient (Total-CC)

The total correlation coefficient (abbreviated total-CC) measures the linear relationship among different observers based on individual measurements. It can be viewed as an intraclass correlation between any measurements from each observer on the same subjects, which is equal to the proportion of subject variability over the total variability. The total-CC of the ith subject is defined as

$$\rho^{total,i} = \frac{\mathrm{Cov}(Y_{i1t}, Y_{i2t'})}{\sqrt{\mathrm{Var}(Y_{i1t})\,\mathrm{Var}(Y_{i2t'})}}$$

where $\mathrm{Cov}(Y_{i1t}, Y_{i2t'})$ is the marginal covariance of measurements taken from two different observers on the same subjects, and $\mathrm{Var}(Y_{i1t})$ and $\mathrm{Var}(Y_{i2t'})$ are the marginal variances of individual measurements. The value of the total-CC is independent of the number of replications.

Based on the proposed model, since Y_{i1t} and $Y_{i2t'}$ are assumed to be conditionally independent given the random effects, the covariance of any measurement from different observer on the same subjects can be expressed as

$$\mathrm{Cov}(Y_{i1t}, Y_{i2t'}) = \mathrm{E}\left\{\mathrm{Cov}(Y_{i1t}, Y_{i2t'}|b_i)\right\} + \mathrm{Cov}\left\{\mathrm{E}(Y_{i1t}|b_i), \mathrm{E}(Y_{i2t'}|b_i)\right\} = \mathrm{Cov}(\mu_{i1t}, \mu_{i2t'}).$$

The total-CC can be written using the expressions of the marginal variance and covariance as

$$\rho^{total} = \frac{\mathrm{Cov}(\mu_{i1t}, \mu_{i2t'})}{\sqrt{[\mathrm{E}\{\phi_1 h_1(\mu_{i1t})\} + \mathrm{Var}(\mu_{i1t})]\,[\mathrm{E}\{\phi_2 h_2(\mu_{i2t'})\} + \mathrm{Var}(\mu_{i2t'})]}}.$$

The total-CC defined above can be shown to be a ratio of functions of $\mathrm{E}(\mu_{i1t})$ and $\mathrm{E}(\mu_{i2t'})$, which vary across different covariates X. That is,

$$\rho^{total} = \frac{K^{total}_{N,1,2}\left\{\mathrm{E}(\mu_{i1t}), \mathrm{E}(\mu_{i2t'})\right\}}{K^{total}_{D,1,2}\left\{\mathrm{E}(\mu_{i1t}), \mathrm{E}(\mu_{i2t'})\right\}}.$$

An overall total-CC is obtained by replacing $\mathrm{E}(\mu_{i1t})$ and $\mathrm{E}(\mu_{i2t'})$ with their marginal expectations over X, $\mu_1^* = \mathrm{E}_X\left\{\mathrm{E}(\mu_{i1t})\right\}$ and $\mu_2^* = \mathrm{E}_X\left\{\mathrm{E}(\mu_{i2t})\right\}$, which are shown in the previous subsections. In other words, the overall total-CC is

$$\rho^{total} = \frac{K^{total}_{N,1,2}\left(\mu_1^*, \mu_2^*\right)}{K^{total}_{D,1,2}\left(\mu_1^*, \mu_2^*\right)}.$$

3.2.4 Properties

The intra-CC, inter-CC and total-CC possess several notable properties. First, the values of intra-CC, inter-CC and total-CC are always scaled between -1 and 1, and increase as the within-subject variability increases. Secondly, according to the definition of inter-CC and total-CC, the value of inter-CC is always greater than or equal to the value of total-CC, and the inter-CC reduces to the total-CC when there is only one replicate from the same observer on each subject. Thirdly, when the covariances between the conditional means are zero, then the inter-CC and total-CC are both equal to zero. In other words, when the subject-specific random effects are uncorrelated with each other, there is no correlation between different observers. However, the inter-CC and total-CC are not necessarily equal to ± 1 when the correlation between the conditional means are ± 1. Moreover, only the inter-CC depends on the number of replications, while the intra-CC and total-CC are independent of the number of replications.

Several previous proposed indices in measuring agreement are special cases of the intra-CC, inter-CC and total-CC proposed in this study. If the observers are identically distributed, then the intra-CC, inter-CC and total-CC reduce to the generalized intra CCC, inter CCC and total CCC proposed by Carrasco (2010). Since the measurements are from particular distributions in the exponential family in GLMM, the normal linear mixed model with identity link is a special case of GLMM. When the observers are independent identically normally distributed with identity link functions, the intra-CC, inter-CC and total-CC turn out to be equivalent to the intra, inter and total precision index proposed by Lin et al. (2007). Under this circumstance, the total-CC further reduces to the CCC proposed by Lin (1989) when there is only one replicate of each subject. In summary, these former proposed indices for assessing agreement are special cases of the indices proposed in this study.

3.2.5 Estimation and inference

Based on the definition of intra-CC, inter-CC and total-CC, these indices are functions of the marginal expectation of μ_{ijt}, $\mu_j^* = \mathrm{E}_X\left\{\mathrm{E}(\mu_{ijt})\right\}$. Let θ denote the set of the fixed-effect parameters, the variance components of the random effects in the generalized linear mixed model, and possible additional model parameters in the fitted distribution. Then $\mu_j^* = \mathrm{E}_X\left\{\mathrm{E}(\mu_{ijt})\right\}$ can be expressed as functions of θ and thus intra-CC, inter-CC and total-CC can also be expressed as functions of θ. Therefore, the sample estimates of θ can be substituted for θ to construct estimators of the intra-CC, inter-CC and total-CC. To obtain these sample estimates, the model is fitted by maximum likelihood methods. Since the conditional distributions are nonnormal and the link functions are non-linear, the maximum likelihood estimation does not have a closed form. Numerical methods such as the adaptive Gaussian-Hermite quadrature or Monte Carlo EM algorithm (Gueorguieva, 2001) can be used to approximate the maximum likelihood estimate. Since the intra-CC, inter-CC and total-CC can all be expressed as ratios of smooth functions of θ, the estimators of the intra-CC, inter-CC and total-CC can be expressed as ratios of smooth functions of the MLE of θ. Specifically, the intra-CC, inter-CC and total-CC are estimated by

$$\begin{aligned}
\hat{\rho}^{intra,j} &= \frac{K_N^{intra,j}\left(\hat{\mu}_1^*, \hat{\mu}_2^*\right)}{K_D^{intra,j}\left(\hat{\mu}_1^*, \hat{\mu}_2^*\right)} \\
\hat{\rho}^{inter} &= \frac{K_{N,1,2}^{inter}\left(T^*\hat{\mu}_1^*, T^*\hat{\mu}_2^*\right)}{K_{D,1,2}^{inter}\left(T^*\hat{\mu}_1^*, T^*\hat{\mu}_2^*\right)} \\
\hat{\rho}^{total} &= \frac{K_{N,1,2}^{total}\left(\hat{\mu}_1^*, \hat{\mu}_2^*\right)}{K_{D,1,2}^{total}\left(\hat{\mu}_1^*, \hat{\mu}_2^*\right)}
\end{aligned}$$

where $\hat{\mu}_j^* = \mathrm{E}_X\left\{\mathrm{E}(\hat{\mu}_{ijt})\right\}$ and $T^* = \frac{\sum_{i=1}^N T_i}{N}$.

Since the estimator of intra-CC, inter-CC or total-CC, say $\hat{\rho}$, is a function of $\hat{\theta}$, the standard error of $\hat{\rho}$ is therefore a function of the standard error of $\hat{\theta}$ and so is the variance of $\hat{\rho}$. Hence, the variance of estimated intra-CC, inter-CC and total-CC can be approximated by applying the delta method. Thus, the variance of $\hat{\rho}$ is $\mathrm{Var}(\hat{\rho}) = d'\Sigma d$, where d is the vector of derivatives of the index ρ with respect to each element of θ, and Σ is the covariance matrix of $\hat{\theta}$. The covariance matrix of $\hat{\theta}$ can be approximated using the inverse of the Fisher's information matrix.

The maximum likelihood estimate $\hat{\theta}$ can be assumed to have a normal distribution asymptotically. Therefore, using the transformation theory of functions of an asymptotically normal vector (Serfling, 1980), the estimator $\hat{\rho}$ is a consistent estimator of ρ and has an asymptotic normal distribution with mean $\mathrm{E}(\hat{\rho}) = \rho$ and variance

$\text{Var}(\hat{\rho}) = d'\Sigma d$. That is,

$$\hat{\rho} \sim AN(\rho, d'\Sigma d).$$

Statistical inference concerning ρ can further be obtained. By assuming asymptotic normality, the $(1-\alpha)$ confidence interval for each index ρ is estimated as

$$[\hat{\rho}_L, \hat{\rho}_U] = \left[\hat{\rho} - z_{\alpha/2}\sqrt{d'\Sigma d},\ \hat{\rho} + z_{\alpha/2}\sqrt{d'\Sigma d}\ \right].$$

These confidence intervals do not bound the values within the open interval $(-1, 1)$.

The normal approximation may be improved by using Fisher's Z-transformation,

$$\hat{Z} = \tan^{-1}(\hat{\rho}) = \frac{1}{2}\log\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right).$$

The Z-transformation performs well for the ordinary Pearson's correlation coefficients. Since $\hat{Z}$ is a function of $\hat{\rho}$, $\hat{Z}$ can be shown to be asymptotically normally distributed. Thus

$$\hat{Z} \sim AN\left(\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right), \frac{\text{Var}\left(\hat{\rho}\right)}{\left(1-\hat{\rho}^2\right)^2}\right).$$

Moreover, by applying the Z-transformation, the $(1-\alpha)$ confidence interval for each index ρ is estimated as

$$[\hat{\rho}_{Z,L}, \hat{\rho}_{Z,U}] = \left[\frac{\exp(2\hat{Z}_L)-1}{\exp(2\hat{Z}_L)+1}, \frac{\exp(2\hat{Z}_U)-1}{\exp(2\hat{Z}_U)+1}\right]$$

where $\hat{Z}_L$ and $\hat{Z}_U$ are lower bound and upper bound of the $(1-\alpha)$ confidence interval for $\hat{Z}$, i.e., $\left(\hat{Z}_L, \hat{Z}_U\right) = \hat{Z} \pm z_{\alpha/2}\sqrt{\text{Var}(\hat{Z})}$ and $z_{\alpha/2}$ is the $(1-\alpha/2)$ percentile of a standard normal distribution. In this study, we will also investigate whether the Z-transformation improves the normality of these CC estimates.

3.3 Extension to Correlations Among Multiple Observers

In a scientific study, it is frequently important to assess the relationship among outcomes observed from more than two different observers. The correlation coefficients defined in the previous sections can be extended to evaluate such relationships. According to the definition, the intra-CC measures the linear relationship among multiple measurements from each observer. As a result, it applies when there are more than two observers. Nonetheless, adjustments to the inter-CC and total-CC are needed when there are more than two observers since the inter-CC and total-CC measure the linear relationship among measurements from two different observers.

Assume that outcomes $Y_i = (Y_{i1}^T, Y_{i2}^T, \cdots, Y_{iL}^T)^T$ are measurements from L observers on the i-th subject and $L > 2$. These measurements can be continuous or discrete responses. One way to assess the relationship among L observers is to use a matrix of pairwise coefficients, as defined in the previous section. Additionally, we define the extended correlation coefficients in this section.

Let $j = 1, \ldots, L-1$ and $k = 2, \ldots, L$ index the pairwise combinations of the L observers. The extended inter correlation coefficient and extended total correlation coefficient are defined as follows :

$$\rho_E^{inter,i} = \frac{\sum_{\substack{j,k \\ j<k}}^{L} \operatorname{Cov}(\overline{Y}_{ij\cdot}, \overline{Y}_{ik\cdot})}{\sum_{\substack{j,k \\ j<k}}^{L} \sqrt{\operatorname{Var}(\overline{Y}_{ij\cdot}) \operatorname{Var}(\overline{Y}_{ik\cdot})}}$$

and

$$\rho_E^{total,i} = \frac{\sum_{\substack{j,k \\ j<k}}^{L} \operatorname{Cov}(Y_{ijt}, Y_{ikt'})}{\sum_{\substack{j,k \\ j<k}}^{L} \sqrt{\operatorname{Var}(Y_{ijt}) \operatorname{Var}(Y_{ikt'})}}.$$

An overall extended inter correlation coefficient and extended total correlation coefficient are further defined as follows :

$$\rho_E^{inter} = \frac{\sum_{\substack{j,k \\ j<k}}^{L} K_{N,j,k}^{inter}\left(T^* \mu_j^*, T^* \mu_k^*\right)}{\sum_{\substack{j,k \\ j<k}}^{L} K_{D,j,k}^{inter}\left(T^* \mu_j^*, T^* \mu_k^*\right)}$$

and

$$\rho_E^{total} = \frac{\sum_{\substack{j,k \\ j<k}}^{L} K_{N,j,k}^{total}\left(\mu_j^*, \mu_k^*\right)}{\sum_{\substack{j,k \\ j<k}}^{L} K_{D,j,k}^{total}\left(\mu_j^*, \mu_k^*\right)}$$

where $K_{N,j,k}^{inter}(\cdot)$, $K_{D,j,k}^{inter}(\cdot)$, $K_{N,j,k}^{total}(\cdot)$ and $K_{D,j,k}^{toal}(\cdot)$ are defined in the previous section. Both of the extended inter-CC and extended total-CC are weighted averages of all pairwise inter-CCs and total-CCs. Hence, the extended inter-CC and extended total-CC are natural extensions of the inter-CC and total-CC. When $L = 2$, $\rho_E^{inter} = \rho^{inter}$ and $\rho_E^{total} = \rho^{total}$.

As a result, the estimates of the extended inter-CC and extended total-CC are obtained by replacing μ_j^* with its sample estimate $\hat{\mu}_j^*$, which can be expressed as

$$\hat{\rho}_E^{inter} = \frac{\sum_{\substack{j,k \\ j<k}}^{L} K_{N,j,k}^{inter}\left(T^*\hat{\mu}_j^*, T^*\hat{\mu}_k^*\right)}{\sum_{\substack{j,k \\ j<k}}^{L} K_{D,j,k}^{inter}\left(T^*\hat{\mu}_j^*, T^*\hat{\mu}_k^*\right)}$$

and

$$\hat{\rho}_E^{total} = \frac{\sum_{\substack{j,k \\ j<k}}^{L} K_{N,j,k}^{total}\left(\hat{\mu}_j^*, \hat{\mu}_k^*\right)}{\sum_{\substack{j,k \\ j<k}}^{L} K_{D,j,k}^{total}\left(\hat{\mu}_j^*, \hat{\mu}_k^*\right)}.$$

Let θ denote the set of the fixed-effect parameters, the variance components of the random effects in the generalized linear mixed model, and possible additional model parameters in the fitted distribution. As shown before, statistical inference can be made using the result that $\hat{\theta}$ has a normal distribution asymptotically. Thus

$$\hat{\rho}_E^{inter} \sim AN\left(\rho_E^{inter}, d^{inter\prime}\hat{\Sigma}d^{inter}\right)$$

and

$$\hat{\rho}_E^{total} \sim AN\left(\rho_E^{total}, d^{total\prime}\hat{\Sigma}d^{total}\right)$$

where d^{inter} is a vector of derivatives of ρ_E^{inter} with respect to each element of θ, d^{total} is a vector of derivatives of the index ρ_E^{total} with respect to each element of θ, and Σ is the covariance matrix of $\hat{\theta}$.

The $(1-\alpha)$ confidence interval for extended inter-CC and extended total-CC are estimated as

$$\left[\hat{\rho}_{E,L}^{inter}, \hat{\rho}_{E,U}^{inter}\right] = \left[\hat{\rho}_E^{inter} - z_{\alpha/2}\sqrt{d^{inter\prime}\hat{\Sigma}d^{inter}},\ \hat{\rho}_E^{inter} + z_{\alpha/2}\sqrt{d^{inter\prime}\hat{\Sigma}d^{inter}}\ \right]$$

and

$$\left[\hat{\rho}_{E,L}^{total}, \hat{\rho}_{E,U}^{total}\right] = \left[\hat{\rho}_E^{total} - z_{\alpha/2}\sqrt{d^{total\prime}\hat{\Sigma}d^{total}},\ \hat{\rho}_E^{total} + z_{\alpha/2}\sqrt{d^{total\prime}\hat{\Sigma}d^{total}}\ \right].$$

For $\hat{\rho}_E^{inter}$ and $\hat{\rho}_E^{total}$, the normal approximation may be improved by using Fisher's Z-transformation,

$$\hat{Z}_E^{inter} = \tan^{-1}(\hat{\rho}_E^{inter}) = \frac{1}{2}\log\left(\frac{1+\hat{\rho}_E^{inter}}{1-\hat{\rho}_E^{inter}}\right)$$

and

$$\hat{Z}_E^{total} = \tan^{-1}(\hat{\rho}_E^{total}) = \frac{1}{2}\log\left(\frac{1+\hat{\rho}_E^{total}}{1-\hat{\rho}_E^{total}}\right).$$

By applying the Z-transformation, the $(1-\alpha)$ confidence interval for ρ_E^{inter} and ρ_E^{total} can be developed as

$$\left[\hat{\rho}_{E,Z,L}^{inter}, \hat{\rho}_{E,Z,U}^{inter}\right] = \left[\frac{\exp(2\hat{Z}_{E,L}^{inter})-1}{\exp(2\hat{Z}_{E,L}^{inter})+1}, \frac{\exp(2\hat{Z}_{E,U}^{inter})-1}{\exp(2\hat{Z}_{E,U}^{inter})+1}\right]$$

and

$$\left[\hat{\rho}_{E,Z,L}^{total}, \hat{\rho}_{E,Z,U}^{total}\right] = \left[\frac{\exp(2\hat{Z}_{E,L}^{total})-1}{\exp(2\hat{Z}_{E,L}^{total})+1}, \frac{\exp(2\hat{Z}_{E,U}^{total})-1}{\exp(2\hat{Z}_{E,U}^{total})+1}\right]$$

where $\hat{Z}_{E,L}^{inter}$ and $\hat{Z}_{E,U}^{inter}$ are the lower bound and upper bound of the $(1-\alpha)$ confidence interval for Z_E^{inter}; $\hat{Z}_{E,L}^{total}$ and $\hat{Z}_{E,U}^{total}$ are the lower bound and upper bound of the $(1-\alpha)$ confidence interval for Z_E^{total}. In this study, we will also investigate whether the Z-transformation improves the normality of these extended CC estimates.

3.4 Illustrative Example

3.4.1 Joint modeling of Poisson-gamma bivariate outcomes

Let $Y_i = (Y_{i1}^T, Y_{i2}^T)^T$ represent the bivariate outcomes measured from the i-th subject, where $Y_{i1}^T = (Y_{i11}, \cdots, Y_{i1T_i})^T$ and $Y_{i2}^T = (Y_{i21}, \cdots, Y_{i2T_i})^T$ are the repeated measurements from the first and second observer, respectively. Conditional on the random effects b_i, Y_{i1t} is assumed to follow a Poisson distribution with mean μ_{i1t} and variance μ_{i1t}, and Y_{i2t} is assumed to follow a gamma distribution with mean μ_{i2t} and variance μ_{i2t}^2/ν. Y_{i1t} and Y_{i2t} are assumed to be conditionally independent given the random effects. Let $x_{1,ijt}, \ldots, x_{p_j,ijt}$ be covariates of j-th observer on the i-th subject. The link function between the linear predictors and the conditional mean of either Y_{i1t} and Y_{i2t} is the natural logarithm. The model conditional on the correlated random effects is as follows:

$$\begin{aligned}
Y_{i1t}|b_{i1} &\sim \text{Poisson distribution with mean } \mu_{i1t} \text{ and variance } \mu_{i1t} \\
Y_{i2t}|b_{i2} &\sim \text{gamma distribution with mean } \mu_{i2t} \text{ and variance } \mu_{i2t}^2/\nu \\
\log(\mu_{i1t}) &= \beta_{10} + \beta_{11}x_{1,i1t} + \cdots + \beta_{1p_1}x_{p_1,i1t} + b_{i1} \\
\log(\mu_{i2t}) &= \beta_{20} + \beta_{21}x_{1,i2t} + \cdots + \beta_{2p_2}x_{p_2,i2t} + b_{i2} \\
b_i &= (b_{i1}, b_{i2})^T \sim \text{iid multivariate normal}\,(0, D)
\end{aligned}$$

where the covariance matrix $D = \begin{pmatrix} \sigma_{b1}^2 & \rho_b\sigma_{b1}\sigma_{b2} \\ \rho_b\sigma_{b1}\sigma_{b2} & \sigma_{b2}^2 \end{pmatrix}$.

Based on this model, we can compute the marginal expectations and variances of the outcomes, and the marginal expectations and variances of the conditional means. Derivation details are included in the Appendix. Define $\mu_1^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i1t})\}$, $\mu_2^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i2t})\}$ and $T^* = \frac{\sum_{i=1}^N T_i}{N}$. Then the overall intra-CC of the first observer and second observer are

$$\rho^{intra,1} = \frac{\mu_1^*(e^{\sigma_{b_1}^2} - 1)}{1 + \mu_1^*(e^{\sigma_{b_1}^2} - 1)}$$

and

$$\rho^{intra,2} = \frac{e^{\sigma_{b_2}^2} - 1}{\left(1 + \frac{1}{\nu}\right) e^{\sigma_{b_2}^2} - 1}.$$

The overall inter-CC of the bivariate measurements is defined as

$$\rho^{inter} = \frac{e^{\rho_b \sigma_{b_1} \sigma_{b_2}} - 1}{\sqrt{\left\{\frac{1}{T^* \mu_1^*} + \left(e^{\sigma_{b_1}^2} - 1\right)\right\} \left\{\left(1 + \frac{1}{T^* \nu}\right) e^{\sigma_{b_2}^2} - 1\right\}}}$$

and the overall total-CC of the bivariate measurements is defined as

$$\rho^{total} = \frac{e^{\rho_b \sigma_{b_1} \sigma_{b_2}} - 1}{\sqrt{\left\{\frac{1}{\mu_1^*} + \left(e^{\sigma_{b_1}^2} - 1\right)\right\} \left\{\left(1 + \frac{1}{\nu}\right) e^{\sigma_{b_2}^2} - 1\right\}}}.$$

3.4.2 Joint modeling of Poisson-exponential-normal multivariate outcomes

Now consider the scenario of more than two observers measuring either discrete outcomes or continuous outcomes. For simplicity, the situation of three observers is investigated in this study. It is straightforward to extend this to models with more than three observers. To reduce the complexity of the model, a joint model of Poisson-exponential-normal multivariate outcomes is considered.

Let $Y_i = (Y_{i1}^T, Y_{i2}^T, Y_{i3}^T)^T$ represent the multivariate outcomes for the i-th subject, where $Y_{ij}^T = (Y_{ij1}, \cdots, Y_{ijT_i})^T$ is the repeated measurement from the j-th observer. Conditional on the random effects, Y_{i1t} is assumed to be count outcomes and follow a Poisson distribution, Y_{i2t} is assumed to be continuous outcomes and follow an exponential distribution, and Y_{i3t} is assumed to be continuous outcomes and follow a normal distribution with variance σ_N^2. Assume that conditional on the random effects, Y_{i1t}, Y_{i2t} and Y_{i3t} are conditionally pairwise independent. Let $x_{1,ijt}, \ldots, x_{p_j,ijt}$ be covariates of the j-th observer on the i-th subject. The link functions among the linear predictors and the conditional means of the measurements are logarithm, logarithm and identity, respectively. Specifically, the model can be expressed as

$$
\begin{aligned}
Y_{i1t}|b_{i1} &\sim \text{Poisson distribution with mean } \mu_{i1t} \text{ and variance } \mu_{i1t} \\
Y_{i2t}|b_{i2} &\sim \text{exponential distribution with mean } \mu_{i2t} \text{ and variance } \mu_{i2t}^2 \\
Y_{i3t}|b_{i3} &\sim \text{normal distribution with mean } \mu_{i3t} \text{ and variance } \sigma_N^2 \\
\log(\mu_{i1t}) &= \beta_{10} + \beta_{11}x_{1,i1t} + \cdots + \beta_{1p_1}x_{p_1,i1t} + b_{i1} \\
\log(\mu_{i2t}) &= \beta_{20} + \beta_{21}x_{1,i2t} + \cdots + \beta_{2p_2}x_{p_2,i2t} + b_{i2} \\
\mu_{i3t} &= \beta_{30} + \beta_{31}x_{1,i3t} + \cdots + \beta_{3p_3}x_{p_3,i3t} + b_{i3} \\
b_i &= (b_{i1}, b_{i2}, b_{i3})^T \sim \text{iid multivariate normal}(0, D)
\end{aligned}
$$

where the covariance matrix $D = \begin{pmatrix} \sigma_{b_1}^2 & \rho_{b_{12}}\sigma_{b_1}\sigma_{b_2} & \rho_{b_{13}}\sigma_{b_1}\sigma_{b_3} \\ \rho_{b_{12}}\sigma_{b_1}\sigma_{b_2} & \sigma_{b_2}^2 & \rho_{b_{23}}\sigma_{b_2}\sigma_{b_3} \\ \rho_{b_{13}}\sigma_{b_1}\sigma_{b_3} & \rho_{b_{23}}\sigma_{b_2}\sigma_{b_3} & \sigma_{b_3}^2 \end{pmatrix}$.

Details of the derivations are provided in the Appendix. Let $\mu_1^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i1t})\}$, $\mu_2^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i2t})\}$ and $T^* = \frac{\sum_{i=1}^N T_i}{N}$. Then the overall extended intra-CCs of the first, second and third observer are

$$
\rho_E^{intra,1} = \frac{\mu_1^*(e^{\sigma_{b_1}^2} - 1)}{1 + \mu_1^*(e^{\sigma_{b_1}^2} - 1)},
$$

$$
\rho_E^{intra,2} = \frac{e^{\sigma_{b_2}^2} - 1}{2(e^{\sigma_{b_2}^2} - 1)},
$$

and

$$
\rho_E^{intra,3} = \frac{\sigma_{b_3}^2}{\sigma_{b_N}^2 + \sigma_{b_3}^2}.
$$

An overall extended inter correlation coefficient is defined as $\rho_E^{inter} = NI^*/DI^*$, where

$$NI^* = \mu_1^*\mu_2^*(e^{\rho_{b12}\sigma_{b1}\sigma_{b2}} - 1) + \mu_1^*\rho_{b13}\sigma_{b1}\sigma_{b3}e^{\sigma_{b_1}^2/2(|\rho_{b13}|-1)} + \mu_2^*\rho_{b23}\sigma_{b2}\sigma_{b3}e^{\sigma_{b_2}^2/2(|\rho_{b23}|-1)}$$

and

$$\begin{aligned} DI^* =& \sqrt{\left\{\frac{\mu_1^*}{T^*} + \mu_1^{*^2}(e^{\sigma_{b_1}^2} - 1)\right\}\left[\mu_2^{*^2}\left\{\left(1 + \frac{1}{T^*\nu}\right)e^{\sigma_{b_2}^2} - 1\right\}\right]} \\ &+ \sqrt{\left\{\frac{\mu_1^*}{T^*} + \mu_1^{*^2}(e^{\sigma_{b_1}^2} - 1)\right\}\left(\frac{\sigma_N^2}{T^*} + \sigma_{b_3}^2\right)} \\ &+ \sqrt{\left[\mu_2^{*^2}\left\{\left(1 + \frac{1}{T^*\nu}\right)e^{\sigma_{b_2}^2} - 1\right\}\right]\left(\frac{\sigma_N^2}{T^*} + \sigma_{b_3}^2\right)}. \end{aligned}$$

Analogously, an overall extended inter correlation coefficient is defined as $\rho_E^{total} = NT^*/DT^*$, where

$$NT^* = \mu_1^*\mu_2^*(e^{\rho_{b12}\sigma_{b1}\sigma_{b2}} - 1) + \mu_1^*\rho_{b13}\sigma_{b1}\sigma_{b3}e^{\sigma_{b_1}^2/2(|\rho_{b13}|-1)} + \mu_2^*\rho_{b23}\sigma_{b2}\sigma_{b3}e^{\sigma_{b_2}^2/2(|\rho_{b23}|-1)}$$

and

$$\begin{aligned} DT^* =& \sqrt{\left\{\mu_1^* + \mu_1^{*^2}(e^{\sigma_{b_1}^2} - 1)\right\}\left[\mu_2^{*^2}\left\{\left(1 + \frac{1}{\nu}\right)e^{\sigma_{b_2}^2} - 1\right\}\right]} \\ &+ \sqrt{\left\{\mu_1^* + \mu_1^{*^2}(e^{\sigma_{b_1}^2} - 1)\right\}\left(\sigma_N^2 + \sigma_{b_3}^2\right)} \\ &+ \sqrt{\left[\mu_2^{*^2}\left\{\left(1 + \frac{1}{\nu}\right)e^{\sigma_{b_2}^2} - 1\right\}\right]\left(\sigma_N^2 + \sigma_{b_3}^2\right)}. \end{aligned}$$

3.5 Simulation Study

Simulation studies are conducted to investigate the performance of estimated correlation coefficients of clustered mixed data from a multivariate generalized linear model. Consider the scenario that each subject is repeatedly measured by each observer numerous times. Simulations are performed for three studies, illustrating how we calculate the proposed correlation coefficients when there are two or three different observers and evaluating how the estimates of each proposed correlation coefficient are affected by the parameter settings. In the first simulation study, included in Section 3.5.1, we investigate how the correlations of the subject-specific random effects affect the estimates of the proposed correlation coefficients. In the second simulation study, provided in Section 3.5.2, we investigate how the sample size and the number of repeated measurements for each observer per subject affect the estimates of the proposed correlation coefficients. In Section 3.5.3, a simulation study of clustered mixed data from three observers is assessed. The performance of these correlation coefficient estimates is evaluated in terms of the relative bias of the point estimate, the relative bias of the standard error, the mean square error of the estimates, the p-values of the Shapiro-Wilk test for normality, and the confidence interval coverage rates.

3.5.1 Correlation estimates vs. correlation between random effects

In the first simulation study, the relationship between the correlation estimates and the correlations of the random effects is of interest. Consider the bivariate mixed outcomes $Y_{i \cdot t} = (Y_{i1t}, Y_{i2t})$ observed from i-th subject and t-th repetition. Two fixed predictors, $x_{1,ijt}$ and $x_{2,ijt}$, for each observer are generated independently from the standard normal distribution for each subject. The random effects $b_i = (b_{i1}, b_{i2})^T$ are generated independently from a bivariate normal distribution with mean $\mu = 0$ and covariance matrix

$$D = \begin{pmatrix} \sigma_{b1}^2 & \rho_b \sigma_{b1} \sigma_{b2} \\ \rho_b \sigma_{b1} \sigma_{b2} & \sigma_{b2}^2 \end{pmatrix}.$$

Given the random effects b_i, the measurements Y_{i1t} and Y_{i2t} are generated independently from a Poisson distribution with mean μ_{i1t}, and a gamma distribution with mean μ_{i2t} and variance μ_{i2t}^2/ν, respectively, using the log links for those two distributions. Specifically, the model is as follow:

$$\begin{aligned}
Y_{i1t}|b_{i1} &\sim \text{Poisson distribution with mean } \mu_{i1t} \text{ and variance } \mu_{i1t} \\
Y_{i2t}|b_{i2} &\sim \text{gamma distribution with mean } \mu_{i2t} \text{ and variance } \mu_{i2t}^2/\nu \\
\log(\mu_{i1t}) &= \beta_{10} + \beta_{11} x_{1,i1t} + \beta_{11} x_{2,i1t} + b_{i1} \\
\log(\mu_{i2t}) &= \beta_{20} + \beta_{21} x_{1,i2t} + \beta_{22} x_{2,i2t} + b_{i2} \\
b_i &= (b_{i1}, b_{i2})^T \sim \text{iid multivariate normal}\,(0, D)
\end{aligned}$$

Repeated measurements samples are generated using the SAS IML procedure, and are fitted by maximum likelihood using the adaptive Gaussian quadrature through the SAS NLMIXED procedure. Simulations are performed for sample size of $N = 100$ and $T = 10$ repeated measurements for each observer on each subject with two parameter settings. In the first setting of the parameter, we assume that there are high within-observer variabilities. Specifically, the variances of the random effects are set to be $(\sigma_{b_1}^2, \sigma_{b_2}^2) = (1, 1)$. The second setting of the parameter assumes that there are low within-observer variabilities. Specifically, the variances of the random effects are set to be $(\sigma_{b_1}^2, \sigma_{b_2}^2) = (0.25, 0.25)$. The fixed effects are set to be $(\beta_{10}, \beta_{11}, \beta_{12}, \beta_{20}, \beta_{21}, \beta_{22}) = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1)$, and the shape parameter in the fitted gamma distribution is set to be $\nu = 20$ in both cases.

In each run, fixed effect parameters, variance components, addition model parameters in the conditional distributions, and then the proposed correlation coefficients are estimated. The combinations of parameters and theoretical correlations are listed in Table 3.1. It is built as the correlation between two random effects varies between strong positive correlation ($\rho_b = 0.9$), moderately positive correlation ($\rho_b = 0.5$), no correlation ($\rho_b = 0$), moderate negative correlation ($\rho_b = -0.5$) and strong negative correlation ($\rho_b = -0.9$). The first five combinations correspond to a situation of high within-observer variability, while the last five combinations correspond to a situation of low within-observer variability. The first column presents the number of combinations. The second and third columns present the number of parameter settings and the correlations of the random effects used in the simulation, respectively. The fourth, fifth, sixth, and seventh columns present theoretical intra-CC 1, theoretical intra-CC 2, theoretical inter-CC and theoretical total-CC, respectively. As seen from the table, the values of theoretical intra-CC 1, theoretical intra-CC 2 remain the same over all combinations. Conversely, the values of theoretical inter-CC and theoretical total-CC increases as the correlation of random effects ρ_b increases, and tends to the limits which are not equal to 1. Both theoretical inter-CC and theoretical total-CC are equal to zero when the random effects are independent (ρ_b=0).

No.	Parameter Setting	Correlation	Theoretical Intra-CC 1	Theoretical Intra-CC 2	Theoretical Inter-CC	Theoretical Total-CC
1	1	-0.9	0.7598	0.9267	-0.3387	-0.2898
2		-0.5	0.7598	0.9267	-0.2246	-0.1921
3		0	0.7598	0.9267	0	0
4		0.5	0.7598	0.9267	0.3703	0.3168
5		0.9	0.7598	0.9267	0.8331	0.7128
6	2	-0.9	0.2643	0.8156	-0.6204	-0.3294
7		-0.5	0.2643	0.8156	-0.3618	-0.1921
8		0	0.2643	0.8156	0	0
9		0.5	0.2643	0.8156	0.4100	0.2177
10		0.9	0.2643	0.8156	0.7770	0.4125

Table 3.1: Simulated Combinations in Simulation Study 1.

Table 3.2 summarizes the results based on a thousand simulation runs. The first and second columns stand for the number of combinations and the index, respectively. The third column presents the relative bias of the point estimate, which is calculated by taking the difference between the mean of the estimates and the theoretical value and dividing it by the theoretical value. The robustness of the estimates is evaluated by the relative bias of the point estimate. The fourth column presents the relative bias of the standard error, which is calculated by taking the difference between the mean of the standard error and the standard deviations of estimates and dividing it by the standard deviations of estimates. The precision of the estimates is evaluated by the relative bias of the standard error. The fifth column presents the mean square error (MSE) which is equal to the sum of the variance and the squared bias of the estimate. MSE measures the accuracy and precision of the CC estimates. The sixth and seventh columns shows the p-values in the Shapiro-Wilk normality test, which is used to access the normal approximation of the raw estimates and the transformed estimates. The 95% confidence intervals are built by assuming asymptotic normality and by Fisher's Z-transformation, respectively. The eighth and ninth columns in the table show the coverage, which are reported by the percentage of times that 95% confidence intervals included the true correlation coefficient. The performance of 95% confidence intervals is evaluated by the confidence interval coverages.

In the simulation considering different correlations between random effects, we find that most of the point estimation and standard error estimation tend to be smaller than expected. The relative biases of the point estimate are always less than 6%. Almost all combinations yield good point estimates, while few of them yield poor standard error estimates. Increasing the correlation between random effects generally made little difference to the relative bias of the point estimates and standard errors. These estimations are quite robust against the variability of each

No.	Index	Est. Relative Bias %	Std Err Est. Relative Bias %	MSE	S.W. P-value for Raw Est.	S.W. P-value for Trans. Est.	CI Coverage (Asym. normal)%	CI Coverage (Z-trans.)%
1	Intra-CC 1	-1.23	-3.39	0.004143	<0.0001	0.0120	93.90	93.60
	Intra-CC 2	-0.11	-1.65	0.000046	<0.0001	0.0301	95.30	94.90
	Inter-CC	1.29	-0.59	0.001720	0.2293	0.3779	92.50	92.90
	Total-CC	0.21	0.24	0.000647	<0.0001	<0.0001	92.10	91.90
2	Intra-CC 1	-2.16	-4.75	0.004742	<0.0001	0.0175	92.70	92.20
	Intra-CC 2	-0.11	-5.83	0.000050	<0.0001	0.0009	94.30	94.00
	Inter-CC	0.33	3.16	0.000834	0.6684	0.8764	96.50	96.20
	Total-CC	-0.87	1.50	0.000558	0.0007	0.0028	96.50	96.30
3	Intra-CC 1	-1.49	-1.19	0.004258	<0.0001	0.0051	95.10	94.10
	Intra-CC 2	-0.11	-0.07	0.000045	<0.0001	0.0081	95.00	94.30
	Inter-CC	NA	-0.87	0.003685	<0.0001	<0.0001	93.20	93.00
	Total-CC	NA	-0.77	0.002649	<0.0001	<0.0001	93.30	93.00
4	Intra-CC 1	-1.05	-0.64	0.004081	<0.0001	0.0013	94.20	93.80
	Intra-CC 2	-0.14	-2.94	0.000049	<0.0001	0.0106	94.40	93.20
	Inter-CC	0.75	1.48	0.004986	0.9493	0.0326	94.20	95.20
	Total-CC	0.26	2.42	0.003930	0.9672	0.1990	95.00	95.20
5	Intra-CC 1	-1.61	-0.42	0.004018	<0.0001	0.0035	94.40	93.40
	Intra-CC 2	-0.14	-0.29	0.000046	<0.0001	0.0789	95.20	94.60
	Inter-CC	-0.18	-2.35	0.001098	0.0061	0.0009	94.10	95.10
	Total-CC	-0.94	-1.99	0.002185	0.0002	0.0845	95.00	94.60
6	Intra-CC 1	-2.01	0.30	0.002110	0.0431	0.0017	92.39	91.79
	Intra-CC 2	-0.39	-1.13	0.000448	<0.0001	0.0044	94.89	94.80
	Inter-CC	-0.14	2.08	0.000586	0.0003	0.1708	95.60	95.39
	Total-CC	-1.19	0.02	0.000656	<0.0001	0.0003	94.59	94.29
7	Intra-CC 1	-1.64	-2.01	0.002306	0.0080	0.0003	92.30	91.80
	Intra-CC 2	-0.44	-1.69	0.000456	<0.0001	0.0639	95.80	94.90
	Inter-CC	-1.84	-1.81	0.004233	<0.0001	0.0006	93.70	93.40
	Total-CC	-2.48	-1.97	0.001544	0.0064	0.0334	95.00	94.60
8	Intra-CC 1	-0.22	-6.07	0.002559	0.0207	0.0003	93.10	92.60
	Intra-CC 2	-0.46	-3.18	0.000472	<0.0001	0.0949	95.20	95.00
	Inter-CC	NA	-3.66	0.008752	0.7760	0.7862	94.10	94.30
	Total-CC	NA	-3.50	0.002464	0.7874	0.7961	94.80	94.50
9	Intra-CC 1	-1.63	-0.89	0.002253	0.6565	0.2047	93.20	92.60
	Intra-CC 2	-0.30	-3.28	0.000459	<0.0001	0.3026	95.60	94.50
	Inter-CC	0.04	-4.51	0.007422	0.0411	0.6690	93.50	93.60
	Total-CC	-0.45	-3.60	0.002662	0.6463	0.3738	94.70	94.40
10	Intra-CC 1	-0.98	2.11	0.002033	0.0286	0.0013	94.48	94.18
	Intra-CC 2	-0.29	-2.74	0.000454	<0.0001	0.2306	94.38	93.88
	Inter-CC	0.27	0.76	0.001909	<0.0001	0.6512	94.08	95.59
	Total-CC	-0.16	2.94	0.001955	0.9822	0.8262	95.49	95.29

Table 3.2: The Relative Bias of the Point Estimate, the Relative Bias of the Standard Errors, the Mean Square Error of the Estimates, P-value in Shapiro-Wilk Normality Test, and the Confidence Interval Coverage Rate in Simulation Study 1.

observer. As expected, the mean square errors are very small over all combinations, again implying that correlation between random effects does not have a great effect on the CC estimates. In addition, small p-values in Shapiro-Wilk normality tests imply that the asymptotic approximation to normal distribution does not behave well, and the approximation does not improve significantly when the Fisher's Z-transformation is applied. The coverage rates of the confidence intervals based on the raw estimates and transformed estimates yield almost identical results and are close to nominal coverage rate of confidence intervals. The coverage rates of the confidence intervals based on the raw estimates are accurate compared to the nominal coverage rates in most combinations, probably due to the slight underestimation of standard error. The Z-transformed estimates do not give more accurate coverage rates. In summary, using the Z-transformation does not appear to be beneficial. The main cause of this behavior is that the distributions for some CC estimates are skewed.

3.5.2 Correlation estimates vs. sample size and number of replicates

In the second simulation study, we are interested in the relationship between the correlation estimates and the sample size, as well as the number of repeated measurements for each observer on each subject. Simulations are performed for small ($N = 10$), moderate ($N = 50$), large ($N = 100$) and extra large sample sizes ($N = 200$), considering scenarios with five ($T = 5$), ten ($T = 10$), and twenty ($T = 20$) repeated measurements for each observer per subject. Assume that there is a strong positive correlation among random effects of each subject, a large within-observer variability, and a small mean change among the measurements of each observer. Specifically, the fixed effect parameters are set to be $(\beta_{10}, \beta_{11}, \beta_{12}, \beta_{20}, \beta_{21}, \beta_{22}) = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1)$, the shape parameter in the fitted gamma distribution is set to be $\nu = 20$, and the covariance matrix of the random effects is set to be

$$D = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}.$$

The combinations generated are shown in Table 3.3. The first column presents the number of combinations. The second and third columns present the sample size and the number of replicates used in the simulation. The fourth, fifth, sixth and seventh rows present the theoretical intra-CC 1, theoretical intra-CC 2, theoretical inter-CC and theoretical total-CC, respectively. Notice that the values of theoretical intra-CC 1, theoretical intra-CC 2 and theoretical total-CC are same over all combinations, while the values of theoretical inter-CC vary as the number of replicates.

No.	Number of Subjects	Number of Replicates	Theoretical Intra-CC 1	Theoretical Intra-CC 2	Theoretical Inter-CC	Theoretical Total-CC
1	10	5	0.7598	0.9267	0.8174	0.7128
2		10	0.7598	0.9267	0.8330	0.7128
3		20	0.7598	0.9267	0.8412	0.7128
4	50	5	0.7598	0.9267	0.8174	0.7128
5		10	0.7598	0.9267	0.8330	0.7128
6		20	0.7598	0.9267	0.8412	0.7128
7	100	5	0.7598	0.9267	0.8174	0.7128
8		10	0.7598	0.9267	0.8330	0.7128
9		20	0.7598	0.9267	0.8412	0.7128
10	200	5	0.7598	0.9267	0.8174	0.7128
11		10	0.7598	0.9267	0.8330	0.7128
12		20	0.7598	0.9267	0.8412	0.7128

Table 3.3: Simulated Combinations in Simulation Study 2.

The simulation results based on a thousand runs are reported in Table 3.4. Similar to Table 3.2, the third column and fourth column present the relative biases of the point estimates and standard errors, which evaluate the robustness and the precision of the estimates, respectively. The relative biases of the estimates of the inter-CC and total-CC in combinations 3 and 8 are not defined because the theoretical correlations are zero in these cases. The fifth column shows the mean square error and the sixth column shows the p-values in the Shapiro-Wilk test for normality. The coverage rates of 95% confidence intervals built by assuming asymptotic normality and by Fisher's Z-transformation are shown in the seventh and eighth columns.

These results indicate that most of CC estimates tend to underestimate. It can also be found that the biases of CC estimates decrease as the number of subjects increases. In other words, the CC estimates tend to be unbiased with larger sample. It is worth notice that for larger sample sizes, almost all CC estimates relative biases are lower than 5%, which indicates that CC estimates are very close to the theoretical values. Though the CC estimates are sensitive to the sample size, they are not sensitive to the number of replicates. Differing the number of replicates does not have noteworthy impact on the CC estimation. On the other hand, appreciable differences are observed between the standard error and the standard deviation of the estimates when the sample size is small. In most combinations, the standard error is underestimated by the standard deviation of estimates. Similar to the relative bias of the point estimate, the relative bias of the standard error decreases as the sample size increases. The estimated mean standard errors are very close to the empirical standard deviations when the sample size is greater than 50. However, the number of replicates of each subject has no influence on the standard error estimates. As seen in the table, MSE decreases as the sample size increases. MSE does not change as the number of replicates increases. MSE has the same trend as the relative biases of the point estimate and standard error since it is the combination of these measures. The normal approximation of these estimates is evaluated in terms of the Shapiro-Wilk normality test for both raw estimates and Z-transformed estimates. For all raw estimates, the p-values in the Shapiro-Wilk normality test are very small, indicating the

No.	Index	Est. Relative Bias %	Std Err Est. Relative Bias %	MSE	S.W. P-value for Raw Est.	S.W. P-value for Trans. Est.	CI Coverage (Asym. normal)%	CI Coverage (Z-trans.)%
1	Intra-CC 1	-10.5	-2.08	0.045412	<0.0001	<0.0001	86.0	83.4
	Intra-CC 2	-1.31	-15.95	0.001357	<0.0001	<0.0001	91.2	87.3
	Inter-CC	-3.20	-3.14	0.020257	<0.0001	0.0012	89.1	89.9
	Total-CC	-7.00	-5.99	0.028688	<0.0001	0.1188	88.7	86.3
2	Intra-CC 1	-12.41	-15.25	0.051842	<0.0001	<0.0001	85.9	79.9
	Intra-CC 2	-1.5	-20.75	0.001291	<0.0001	<0.0001	94.1	88.0
	Inter-CC	-1.85	-13.81	0.015465	<0.0001	<0.0001	88.6	91.2
	Total-CC	-7.68	-15.01	0.028289	<0.0001	0.3715	89.8	86.5
3	Intra-CC 1	-12.25	-6.61	0.043640	<0.0001	<0.0001	87.4	83.5
	Intra-CC 2	-1.68	-21.53	0.001287	<0.0001	<0.0001	96.4	90.6
	Inter-CC	-0.79	-13.48	0.010901	<0.0001	<0.0001	86.3	92.8
	Total-CC	-7.39	-9.96	0.024037	<0.0001	0.0331	91.0	86.5
4	Intra-CC 1	-3.13	-5.63	0.010075	<0.0001	0.0069	92.3	91.7
	Intra-CC 2	-0.16	-4.09	0.000130	<0.0001	0.0041	96.3	94.5
	Inter-CC	-0.25	-2.42	0.003215	<0.0001	<0.0001	93.0	94.6
	Total-CC	-1.46	-3.17	0.005179	<0.0001	0.0019	93.2	93.3
5	Intra-CC 1	-3.11	-8.36	0.009485	<0.0001	0.0009	91.0	89.9
	Intra-CC 2	-0.28	-3.30	0.000106	<0.0001	<0.0001	96.8	95.9
	Inter-CC	-0.18	-1.48	0.002147	<0.0001	<0.0001	93.5	95.5
	Total-CC	-1.67	-5.51	0.004688	<0.0001	0.2998	93.9	92.9
6	Intra-CC 1	-3.18	-5.10	0.008403	<0.0001	<0.0001	92.1	90.9
	Intra-CC 2	-0.29	-5.43	0.000098	<0.0001	<0.0001	95.0	92.8
	Inter-CC	-0.06	-0.76	0.001602	<0.0001	0.1332	94.1	95.9
	Total-CC	-1.73	-2.39	0.004047	<0.0001	0.3236	94.7	93.2
7	Intra-CC 1	-1.62	-2.54	0.004640	<0.0001	0.2406	94.4	94.0
	Intra-CC 2	-0.10	-2.24	0.000061	<0.0001	0.0031	95.3	94.4
	Inter-CC	0.14	0.11	0.001500	<0.0001	0.0143	94.3	95.2
	Total-CC	-0.52	1.80	0.002314	<0.0001	0.0737	94.6	95.4
8	Intra-CC 1	-1.50	-0.51	0.004003	<0.0001	0.0038	93.9	94.0
	Intra-CC 2	-0.13	-2.10	0.000047	<0.0001	<0.0001	95.1	94.3
	Inter-CC	-0.10	-2.43	0.001091	<0.0001	0.0767	93.5	94.9
	Total-CC	-0.81	-0.92	0.002123	<0.0001	0.7810	94.9	93.8
9	Intra-CC 1	-1.13	-2.77	0.003804	<0.0001	0.0259	94.0	94.2
	Intra-CC 2	-0.14	-3.31	0.000043	<0.0001	<0.0001	95.6	94.5
	Inter-CC	-0.10	-3.19	0.000849	<0.0001	0.0524	94.5	94.5
	Total-CC	-0.71	-3.15	0.001999	<0.0001	0.8074	93.9	94.2
10	Intra-CC 1	-0.87	-3.84	0.002367	<0.0001	0.3336	94.0	93.8
	Intra-CC 2	-0.04	0.07	0.000028	0.0002	0.4347	95.5	95.1
	Inter-CC	-0.10	-6.45	0.000875	0.0236	0.0083	92.5	93.1
	Total-CC	-0.43	-6.52	0.001387	<0.0001	0.8045	93.7	93.6
11	Intra-CC 1	-0.76	0.74	0.001930	0.0232	0.0020	94.6	94.4
	Intra-CC 2	-0.05	-1.16	0.000022	0.001	0.5379	95.1	94.9
	Inter-CC	-0.02	2.98	0.000488	0.0056	0.0092	95.2	95.9
	Inter-CC	-0.37	1.30	0.001006	0.1236	0.1076	95.6	95.8
12	Intra-CC 1	-0.75	2.79	0.001734	0.0141	0.0010	95.6	96.3
	Intra-CC 2	-0.07	-0.26	0.000019	<0.0001	0.0912	95.4	95.2
	Inter-CC	0.01	0.92	0.000388	0.0384	0.0205	94.5	95.7
	Total-CC	-0.38	1.46	0.000914	0.0973	0.5107	95.4	95.2

Table 3.4: The Relative Bias of the Point Estimate, the Relative Bias of the Standard Errors, the Mean Square Error of the Estimates, P-value in Shapiro-Wilk Normality Test, and the Confidence Interval Coverage Rate in Simulation Study 2.

rejection of the normality assumption. However, for the Z-transformed estimates, p-values increases as the sample size increases. The normal approximation is less likely to be rejected with larger sample size. It appears that the Z-transformation significantly improves the normal approximation in most combinations. Nevertheless, it is well known that the normality test will detect even trivial deviations from normality and usually give a significant result when the sample size is very large. Q-Q plots and histograms are also built to analyze whether the normal assumption holds true. They indicate that most of these CC estimates have symmetric and bell-shaped distributions, suggesting that the asymptotic approximation to normal is appropriate. The coverage rates for both raw data and transformed data give very similar results. The Z-transformation that attempts to improve the normal approximation does not provide greater coverage rates. Both the raw estimates and Z-transformed estimates underestimate the coverage rates in most combinations. This may be due to the fact that the standard errors underestimate the standard deviations. The coverage rates in many combinations do not reach 95 percent coverage and get worse as sample size decreases. Nonetheless, the coverage rates in the combinations with larger samples are above 90 percent.

It is noteworthy that when fitting the MGLMM, the optimization is not guaranteed to achieve convergence. With small sample and few replicates of each subject, the model converge ratio among a thousand simulation runs is not as high as expected, but still has a high level of convergence. With the sample size $N = 10$ and the number of replicates $T = 5$, the model converge ratio is around 73%. With the sample size $N = 10$ and the number of replicates $T = 10$, the model converge ratio is around 87%. Nevertheless, with larger sample size and more replicates of each subject, the converge ratio rises to 100%.

3.5.3 Extended correlation coefficients among multiple measurements

A simulation is conducted to assess the performance of the extended intra-CC, inter-CC and total-CC. Measurements are from three different observers. The study is carried out for different sample sizes ($N = 50, 100$) and different numbers of replicated measurements ($T = 10, 25$) taken from each observer on each subject. For each subject, the covariates $x_{1,ijt}$ and $x_{2,ijt}$ are generated independently from the standard normal distribution, $j = 1, 2, 3$. The random effects $b_i = (b_{i1}, b_{i2}, b_{i3})^T$ are generated from a trivariate normal distribution with mean zero and covariance matrix D for each subject. Data is generated according to the correlated random effects model as follow:

$$
\begin{aligned}
Y_{i1t}|b_{i1} &\sim \text{Poisson distribution with mean } \mu_{i1t} \text{ and variance } \mu_{i1t} \\
Y_{i2t}|b_{i2} &\sim \text{exponential distribution with mean } \mu_{i2t} \text{ and variance } \mu_{i2t}^2 \\
Y_{i3t}|b_{i3} &\sim \text{normal distribution with mean } \mu_{i3t} \text{ and variance } \sigma_N^2 \\
\log(\mu_{i1t}) &= \beta_{10} + \beta_{11}x_{1,i1t} + \beta_{12}x_{2,i1t} + b_{i1} \\
\log(\mu_{i2t}) &= \beta_{20} + \beta_{21}x_{1,i2t} + \beta_{22}x_{2,i2t} + b_{i2} \\
\mu_{i3t} &= \beta_{30} + \beta_{31}x_{1,i3t} + \beta_{32}x_{2,i3t} + b_{i3} \\
b_i &= (b_{i1}, b_{i2}, b_{i3})^T \sim \text{ iid multivariate normal}\,(0, D)
\end{aligned}
$$

where $D = \begin{pmatrix} \sigma_{b_1}^2 & \rho_{b_{12}}\sigma_{b_1}\sigma_{b_2} & \rho_{b_{13}}\sigma_{b_1}\sigma_{b_3} \\ \rho_{b_{12}}\sigma_{b_1}\sigma_{b_2} & \sigma_{b_2}^2 & \rho_{b_{23}}\sigma_{b_2}\sigma_{b_3} \\ \rho_{b_{13}}\sigma_{b_1}\sigma_{b_3} & \rho_{b_{23}}\sigma_{b_2}\sigma_{b_3} & \sigma_{b_3}^2 \end{pmatrix}$.

For simplicity, the scale parameter in the normal distribution, σ_N, is set to be 1. Simulation results are based on five hundred simulated datasets under each scenario with the following parameter specifications are used: $(\beta_{10}, \beta_{11}, \beta_{12}, \beta_{20}, \beta_{21}, \beta_{22}, \beta_{30},$

No.	No. of Sub.	No. of Rep.	Theoretical Intra-CC 1	Theoretical Intra-CC 2	Theoretical Intra-CC 3	Theoretical Inter-CC	Theoretical Total-CC
1	50	10	0.7598	0.3873	0.5	0.6946	0.4022
2		25	0.7598	0.3873	0.5	0.7325	0.4022
3	100	10	0.7598	0.3873	0.5	0.6946	0.4022
4		25	0.7598	0.3873	0.5	0.7325	0.4022

Table 3.5: Simulated Combinations in Simulation Study 3.

$\beta_{31}, \beta_{32}) = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)$, and the covariance matrix of the random effects

$$D = \begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{pmatrix}$$

Table 3.5 summarizes the simulation combinations and the corresponding true values. Only the extended inter-CCs are sensitive to the number of replicates. All others would remain the same for all scenarios.

The point estimates and standard error of the extend intra-CC, inter-CC and total-CC are calculated in each simulation run. The corresponding relative bias of the point estimate, the relative bias of the standard error, the mean square error, the p-values in the normality test and the coverage rate of 95% confidence intervals built by assuming asymptotic normality and by Fishers Z-transformation based on five hundred runs are reported in Table 3.6. There results demonstrate that there are very slight bias in the point estimates of the extend intra-CCs, inter-CC and total-CC. It tends underestimate more seriously when the sample size is smaller. The corresponding relative biases of the standard error are not too big under all scenarios. MSEs in all combination are very small, indicating the estimations are good in both accuracy and precision. The p-values of normality test show that the normal approximations is more appropriate when the sample size is larger. The normal

No.	Index	Est. Relative Bias %	Std Err Est. Relative Bias %	MSE	S.W. P-value for Raw Est.	S.W. P-value for Trans. Est.	CI Coverage (Asym. normal)%	CI Coverage (Z-trans.)%
1	Intra-CC 1	-2.72	-4.86	0.008669	<0.0001	0.0063	92.5	90.6
	Intra-CC 2	-2.00	-3.89	0.001089	0.0001	0.0013	93.7	92.3
	Intra-CC 3	-1.50	-0.03	0.003042	0.0418	0.5015	93.7	94.1
	Inter-CC	-0.07	-3.79	0.001645	0.0121	0.9955	93.7	93.7
	Total-CC	-1.57	-5.33	0.002107	0.0340	0.3547	93.3	92.1
2	Intra-CC 1	-3.41	-4.60	0.008284	<0.0001	0.0113	91.8	91.6
	Intra-CC 2	-2.59	-1.43	0.000978	0.0026	0.0257	94.6	93.2
	Intra-CC 3	-3.10	-8.26	0.003389	0.0109	0.1960	91.0	90.2
	Inter-CC	0.14	0.24	0.000872	0.0002	0.1186	94.8	94.8
	Total-CC	-2.48	-4.73	0.001939	0.1091	0.5832	93.2	92.4
3	Intra-CC 1	-0.76	3.39	0.003577	0.0005	0.0081	96.6	95.6
	Intra-CC 2	-0.55	5.62	0.000422	0.0501	0.1121	95.6	95.0
	Intra-CC 3	-0.66	4.79	0.001380	0.0053	0.0938	95.2	95.0
	Inter-CC	0.70	0.30	0.000730	0.0297	0.5195	94.4	94.8
	Total-CC	0.27	5.27	0.000847	0.6893	0.7224	96.2	96.6
4	Intra-CC 1	-0.62	-0.12	0.003461	0.0020	0.0819	94.8	95.0
	Intra-CC 2	-0.81	-2.44	0.000446	0.0003	0.0018	94.2	94.0
	Intra-CC 3	-1.02	-3.26	0.001458	0.0037	0.0975	94.2	92.4
	Inter-CC	0.19	-0.79	0.000428	0.0159	0.4556	94.8	94.8
	Total-CC	-0.40	-1.24	0.000863	0.0243	0.1804	94.8	95.2

Table 3.6: The Relative Bias of the Point Estimate, the Relative Bias of the Standard Errors, the Mean Square Error of the Estimates, P-value in Shapiro-Wilk Normality Test, and the Confidence Interval Coverage Rate in Simulation Study 3.

approximation is improved when the Z-transformation is applied. The confidence interval coverage rates for raw estimates and Z-transformed estimate are very close to the nominal coverage. In summary, the extend intra-CC, inter-CC and total-CC perform very well in multivariate generalized linear mixed model when the sample size is moderately large, which is consistent with the result in Section 3.5.2.

3.6 Case Example : Data From The Osteoarthritis Initiative

The data used to illustrate the methods introduced in the previous sections come from the Osteoarthritis Initiative (OAI) database, which is available for public access at http://www.oai.ucsf.edu/ and described in detail by McCulloch (2008). OAI is a cohort study of the causations of knee osteoarthritis for more than four thousand people aged 45 and above. Briefly, persons at high risk for developing knee osteoarthritis are observed at baseline, 12 months, 24 months, 36 months and 48 months, resulting in five measurements per individual. The outcomes investigated here are the Western Ontario and McMaster Universities (WOMAC) disability scores and the number of workdays missed. WOMAC is a numeric score used to rate patients' pain, stiffness, and physical function with hip and/or knee osteoarthritis, while the number of days of missed work due to knee pain, aching or stiffness in the past 3 months is a count variable. In this study, we use the average of the WOMAC scores for the left and right knee as the final WOMAC score. We restrict our study to the complete data, which reduces our data to 1499 individuals. The primary objective of the study is to investigate the relationship between the WOMAC score and the number of days of missed work in the past 3 months. The scatter plot of these two outcomes is given in Figure 3.1, indicating that they seem to be uncorrelated.

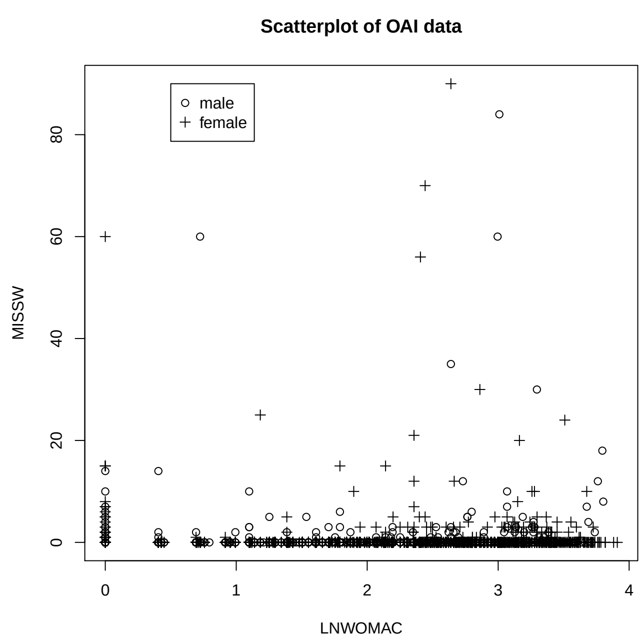

Figure 3.1: Scatter Plot of OAI Data.

	Male	Female
intraclass CC 1	0.3646	0.3720
intraclass CC 2	0.0060	0.0265
sample inter CC	0.1590	0.1305
sample total CC	0.0857	0.0697

Table 3.7: Sample Correlation for Male and Female in OAI Data.

The intraclass correlation coefficients, the sample correlation coefficients based on the average of replicated measurements, and the sample correlation coefficients based on individual measurements for male and female are presented in Table 3.7, showing that these two outcomes are not highly correlated.

Three predictors under consideration are the age, sex and body mass index (BMI). Age and BMI are continuous variables, while sex is a categorical variable. We assume that these covariates are independent, and the covariates age and BMI are normally distributed. Same as McCulloch, we jointly model the log transformation of the WOMAC scores plus one and the number of days of missed work in the past 3 months. However, to consider overdispersion in the count data, negative binomial distribution is used. Thus, the multivariate generalized linear mixed model of normal-negative binomial distributions is fitted with age, sex and BMI as fixed effects, and subject as random effect. More specifically, the model is given by

$$\begin{aligned}
WOMAC_{i1t}|b_i &\sim \text{normal with mean } \mu_{i1t} \text{ and variance } \sigma_N^2 \\
MISSW_{i2t}|b_i &\sim \text{negative binomial with mean } \mu_{i2t} \text{ and variance } \mu_{i2t}\left(1+\frac{1}{\delta_N}\right) \\
\mu_{i1t} &= \beta_{10} + \beta_{11}AGE_{it} + \beta_{12}SEX_i + \beta_{13}BMI_{it} + b_{i1} \\
\log(\mu_{i2t}) &= \beta_{20} + \beta_{21}AGE_{it} + \beta_{22}SEX_i + \beta_{23}BMI_{it} + b_{i2} \\
b_i &= (b_{i1}, b_{i2})^T \sim \text{ iid multivariate normal}(0, D)
\end{aligned}$$

Effects	WOMAC Para.	WOMAC Est.	WOMAC Std Err	MISSW Para.	MISSW Est.	MISSW Std Err
Fixed effect parameters						
Intercept	β_{10}	-0.4511	0.1845	β_{20}	-10.9993	1.8395
AGE	β_{11}	-0.0024	0.0026	β_{21}	-0.0261	0.0247
SEX	β_{12}	0.2100	0.0390	β_{22}	0.2211	0.3532
BMI	β_{13}	0.0513	0.0038	β_{23}	0.2284	0.0361
Covariance structure parameters						
	σ_{b1}	0.6365	0.0165	σ_{b2}	3.3212	0.2925
	ρ_b	0.5103	0.0526			
	σ_N	0.8079	0.0148	δ	0.0581	0.0090

Table 3.8: Parameter Estimates and Standard Errors of MGLMM for OAI Data.

where the covariance matrix $D = \begin{pmatrix} \sigma_{b1}^2 & \rho_b\sigma_{b1}\sigma_{b2} \\ \rho_b\sigma_{b1}\sigma_{b2} & \sigma_{b2}^2 \end{pmatrix}$.

Using the NLMIXED procedure in SAS, the parameter estimates and standard errors are given in Table 3.8. The estimate of ρ_b is 0.5103, indicating that there is a moderate positive correlation between the subject-specific random effects.

Since sex is a categorical variable, the marginal expectation of the conditional mean over the covariates should be calculated in different genders separately. Thus the intra-CC, inter-CC and total-CC are estimated in different genders. The estimated CCs, standard errors, and corresponding 95% confidence intervals are reported in Table 3.9. Since sex effect is not significant for MISSW, the CC estimates for different genders based on the normal-negative binomial model are almost identical. For both males and females, the estimated inter-CC and total-CC are very close to zero, which means that the two outcomes are not strongly correlated though there is a moderate positive correlation between the subject-specific random effects. These results are consistent with what we expected from the scatter plot. The values of the estimated inter-CC and total-CC are lower than the sample inter-CC and total-

	Male			Female		
Index	Estimate	Std Error	95% CI	Estimate	Std Error	95% CI
Intra-CC 1	0.3340	0.0127	(0.3090 , 0.3590)	0.3340	0.0127	(0.3090 , 0.3590)
Intra-CC 2	0.0549	0.0080	(0.0391 , 0.0706)	0.0549	0.0080	(0.0391 , 0.0706)
Inter-CC	0.0002	0.0003	(-0.0003, 0.0007)	0.0002	0.0003	(-0.0003, 0.0007)
Total-CC	0.0001	0.0001	(-0.0001, 0.0002)	0.0001	0.0001	(-0.0001, 0.0002)

Table 3.9: Estimated CC for Male and Female in OAI Data.

CC. In conclusion, there is no strong association between the Western Ontario and McMaster Universities (WOMAC) disability scores and the numbers of workdays missed in the past 3 months.

3.7 Conclusion

In this study, three different types of correlation coefficients which measure various linear relationship between replicated measurements from different observers are proposed. The intra-CC measures the within-observer correlation, the inter-CC measures the between-observer correlation, and the total-CC measures the overall correlation. These indices are very useful for measuring correlation in clustered mixed data. The intra-CC, inter-CC and total-CC proposed in this study give more flexibility since they allow negative correlations, and can be extended when there exist more than two observers.

4. APPROXIMATE UNIFORM SHRINKAGE PRIOR FOR A MULTIVARIATE GENERALIZED LINEAR MIXED MODEL

4.1 Introduction

The multivariate generalized linear mixed models (MGLMM) are used for jointly modeling the clustered mixed outcomes obtained when there is more than one response repeatedly measured on each individual in scientific studies. The Bayesian methods are widely used techniques for analyzing MGLMM. The need of noninformative priors arises when there is insufficient prior information on the model parameters. The main purpose of this study is to propose an approximate uniform shrinkage prior for the random effect variance components in the Bayesian analysis for the MGLMM. This prior is an extension of the approximate uniform shrinkage prior proposed by Natarajan and Kass (2000).

The rest of this chapter is organized as follows. In Section 4.2, the approximate uniform shrinkage prior for multivariate generalized linear mixed model is derived. Illustrative examples are also provided in this section. Section 4.3 presents properties of the approximate uniform shrinkage prior distribution and its corresponding posterior distribution. Section 4.4 explains how the posterior simulation is performed. In Section 4.5, the performance of the approximate uniform shrinkage prior is evaluated by a simulation study. A case example is provided to illustrate the application of the approximate uniform shrinkage prior in Section 4.6. Lastly, Section 4.7 concludes with implications for future research.

4.2 Approximate Uniform Shrinkage Prior for MGLMM

4.2.1 Model

The MGLMM model in the Chapter 2 can be re-expressed as:

$$
\begin{aligned}
y^*|b^* &\sim \prod_{i=1}^{N}\prod_{t=1}^{T_i}\exp\left\{\sum_{j=1}^{L}\frac{y_{ijt}\theta_{ijt}-a_j(\theta_{ijt})}{\phi_j}\right\}\\
g^*(\mu^*) &= X^*\beta^* + Z^*b^*\\
b^* &\sim N(0, D^*).
\end{aligned}
$$

Here the multivariate measurements are expressed as $y^* = (y_{111}, \cdots, y_{N1T_N}, \cdots, y_{1L1}, \cdots, y_{NLT_N})^T$; the conditional mean can be expressed as $\mu^* = (\mu_{111}, \cdots, \mu_{N1T_N}, \cdots, \mu_{1L1}, \cdots, \mu_{NLT_N})^T$; the link function g^* can be expressed as $g^*(t) = (g_1(t_{111}), \cdots, g_1(t_{N1T_N}), \cdots, g_L(t_{1L1}), \cdots, g_L(t_{NLT_N}))^T$; the fixed effect parameter can be expressed as $\beta^* = (\beta_{11}, \cdots, \beta_{1p_1}, \cdots, \beta_{Lp_L})^T$; the covariate matrix can be rewritten as $X^* = \bigoplus_{j=1}^{L} X_j^* = \mathrm{diag}(X_1^*, \cdots, X_L^*)$, where X_j^* is the matrix of covariates for the j-th observer and $\bigoplus$ is the direct sum; the known random effects design matrix can be expressed as $Z^* = \bigoplus_{j=1}^{L}\bigoplus_{i=1}^{N} Z_{ij}^*$, where Z_{ij}^* is the design matrix; the random effects can be expressed as $b^* = (b_{1,11}, \cdots, b_{q,11}, \cdots, b_{q,N1}, \cdots, b_{1,1L}, \cdots, b_{q,1L}, \cdots, b_{q,NL})^T$; the covariance matrix of the random effect is $D^* = [D_{ij}^*]_{i=1,\cdots,L;j=1,\cdots,L}$, where $D_{ij}^* = \bigoplus_{k=1}^{N}\sigma_{ij}$. We assume the prior distribution of β is uniform distribution in this study.

4.2.2 Motivation

We will later show that the weight matrix for the prior mean of the random effect in the approximate uniform shrinkage estimate of the random effect is equal to

$$S = \left(D^{-1} + \frac{1}{N} \sum_{i=1}^{N} Z_i^T W Z_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} Z_i^T W Z_i \right)$$

where D is the covariance matrix of the random effect, W is the GLM weight matrix (McCullagh and Nelder, 1989), and Z_i is the design matrix for the random effects of subject i. S defined above is a function of D. To obtain the approximate uniform shrinkage prior, we assume that $\pi_S(s)$ is componentwise uniformly distributed. Then using the transformation theorem, we find that D has probability density function,

$$\pi_D(D) \propto \left| I_{Lq} + \left(\frac{1}{N} \sum_{i=1}^{N} Z_i^T W Z_i \right) D \right|^{-2q-1}.$$

This is defined as the approximate uniform shrinkage prior for D. Since only positive-semidefinite matrix can be a covariance matrix, we define the approximate uniform shrinkage prior distribution on real-valued positive-definite matrices.

4.2.3 Derivation of weight matrix

The weight matrix S is obtained by finding the approximate uniform shrinkage estimate $\hat{b^*}$ (Breslow and Clayton, 1993), which can be derived from the likelihood function for the parameters β and D :

$$L(\beta, D) \propto |D|^{-\frac{N}{2}} \int \exp \left[- \sum_{j=1}^{L} \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \{ y_{ijt}\theta_{ijt} - a_j(\theta_{ijt}) \}}{2\phi_j} - \frac{1}{2} \sum_{i=1}^{N} b_i^T D^{-1} b_i \right] db.$$

Since the quasi-likelihood method generates efficient estimators without making precise distribution assumptions, it is considered here to estimate the parameters in the above model. The integrated quasi-likelihood function is

$$L(\beta, D) \propto |D^*|^{-\frac{1}{2}} \int \exp\left\{-\sum_{j=1}^{L} \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i} d_j(y_{ijt};\mu_{ijt})}{2\phi_j} - \frac{1}{2} b^{*T} D^{-1} b^*\right\} db$$

where $d_j(y;\mu) = -2\int_y^{\mu} \frac{y-\mu}{a_j''(u)} du$ is the quasi-deviance function (McCullagh and Nelder, 1989). Then the log quasi-likelihood function is

$$ql(\beta, D) \approx -\frac{1}{2}\log|D^*| + \log\int e^{-\kappa(b^*)} db^*$$

where $\kappa(b^*) = \sum_{j=1}^{L} \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i} d_j(y_{ijt};\mu_{ijt})}{2\phi_j} + \frac{1}{2} b^{*T} D^{-1} b^*$.

Laplace's method can be used for approximation of the higher dimensional integral in the likelihood based on second-order Taylor series expansion, which gives

$$ql(\beta, D) \approx -\frac{1}{2}\log|I_{NLq} + Z^{*T} W^* Z^* D^*| - \sum_{j=1}^{L} \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i} d_1(y_{ijt};\mu_{ijt})}{2\phi_j} - \frac{1}{2}\tilde{b^*}^T D^{-1}\tilde{b^*}$$

where $\tilde{b^*} = \tilde{b}(\alpha, \theta)$ is chosen such that $\kappa'(\tilde{b^*}) = 0$ and
$W^* = \text{diag}\left(\left[\phi_1 a_1''(\mu_{111})\{g_1'(\mu_{111})\}^2\right]^{-1}, \cdots, \left[\phi_L a_L''(\mu_{N1T_N})\{g_1'(\mu_{NLT_N})\}^2\right]^{-1}\right)$ is the diagonal block GLM weight matrix.

Since $ql(\beta, D)$ may not result in a closed form solution and cannot be used to estimate the variance-covariance structure, the penalized quasi-likelihood (PQL) (Breslow and Clayton, 1993; Green, 1987) is developed. Assuming that the GLM iterative weights vary very slowly as a function of the mean, the penalized quasi-likelihood is defined by adding a penalty function to the quasi-likelihood of the form $-\frac{1}{2}b^{*T}D^{-1}b^*$.

Therefore, the penalized quasi-likelihood is equal to

$$PQL = -\sum_{j=1}^{L} \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} d_j(y_{ijt}; \mu_{ijt})}{2\phi_j} - \frac{1}{2} b^{*T} D^{-1} b^{*}.$$

The maximum penalized quasi-likelihood equations are implemented by differentiating PQL with respect to β^* and b^*. Using Fisher scoring algorithm, these score equations are modified to an iterative weighted least squares problem (Green, 1987; Harville, 1977) :

$$\begin{bmatrix} X^{*T}W^*X^* & X^{*T}W^*Z^*D^* \\ Z^{*T}W^*X^* & I + Z^{*T}W^*Z^*D^* \end{bmatrix} \begin{bmatrix} \beta^* \\ \nu \end{bmatrix} = \begin{bmatrix} X^{*T}W^*Y^{0*} \\ Z^{*T}W^*Y^{0*} \end{bmatrix}$$

where $b^* = D^*\nu$, Y^{0*} is the working vector and W^* is the diagonal block GLM weight matrix obtained by replacing b^* with 0. Solving this equation, we can get $\hat{\beta^*} = \left(X^{*T}V^{*-1}X^*\right)^{-1} X^{*T}V^{*-1}Y^{0*}$ and $\hat{b^*} = D^*Z^{*T}V^{*-1}(Y^{0*} - X^*\hat{\beta^*})$, where $V^* = W^{*-1} + Z^*D^*Z^{*T}$. Under this model, the prior mean of b^* is a vector of zeros, 0, and a frequentist estimate of b^* is $D^*Z^{*T}W^*(Y^{0*} - X^*\hat{\beta^*})$. Thus the approximate shrinkage estimate b^* can be expressed as a weighted average of its prior mean and frequentist estimate and has the form

$$\begin{aligned} \hat{b^*} &= D^*Z^{*T}\left(W^{*-1} + Z^*D^*Z^{*T}\right)^{-1}(Y^{0*} - X^*\hat{\beta^*}) \\ &= S^* \cdot 0 + (I_{NLq} - S^*) \cdot D^*Z^{*T}W^*(Y^{0*} - X^*\hat{\beta^*}) \end{aligned}$$

where $S^* = \left(D^{*-1} + Z^{*T}W^*Z^*\right)^{-1} Z^{*T}W^*Z^*$ is the weight given to the approximate shrinkage estimate and the k-th row in S^* denotes the weights on the prior mean of the k-th element of b^*. S^* amounts to a shrinkage factor, and the approximate shrinkage estimate shrinks the frequentist estimate toward the prior mean.

Notice that the weight matrix S^* is a $NLq \times NLq$ block matrix comprising weights for every subject and its dimensionality depends on the number of subjects, which may result in high-dimensional problems. Therefore, we define an overall weight matrix

$$S = \left(D^{-1} + \frac{1}{N}\sum_{i=1}^{N} Z_i^T W Z_i\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^T W Z_i\right).$$

4.2.4 Illustrative examples: bivariate and trivariate cases

To illustrate the approximate uniform shrinkage prior, a bivariate clustered mixed model with random intercept, a bivariate clustered mixed model with both random intercept and random slope, and a trivariate clustered mixed model with random intercept are considered. The detailed derivations of their approximate uniform shrinkage priors are provided in the Appendix. The extension to higher dimensional models or models with more than one random slope is quite straightforward.

Example 1: a bivariate clustered mixed model with random intercept

For simplicity, a bivariate clustered mixed model with random intercept is taken into account first. A bivariate clustered mixed model, where the responses are assumed to be conditionally independent from a Poisson distribution and a gamma distribution, is given below :

$$\begin{aligned}
Y_{i1t}|b_{i1} &\sim \text{Poisson distribution with mean } \mu_{i1t} \text{ and variance } \mu_{i1t} \\
Y_{i2t}|b_{i2} &\sim \text{gamma distribution with mean } \mu_{i2t} \text{ and variance } \mu_{i2t}^2/\nu \\
\log(\mu_{i1t}) &= \beta_{10} + \beta_{11}x_{1,i1t} + \cdots + \beta_{1p_1}x_{p_1,i1t} + b_{i1} \\
\log(\mu_{i2t}) &= \beta_{20} + \beta_{21}x_{1,i2t} + \cdots + \beta_{2p_2}x_{p_2,i2t} + b_{i2} \\
b_i = (b_{i1}, b_{i2})^T &\sim \text{ iid multivariate normal}\,(0, D)
\end{aligned}$$

where $D = [\sigma_{ij}]_{i=1,2;\ j=1,2}$, $i = 1, \cdots, N$ and $t = 1, \cdots, T_i$. Therefore, the approximate uniform shrinkage prior is

$$\pi_D(D) \propto \left\{ \left[1 + \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{t=1}^{T_i} \mu_{i1t} \right) \sigma_{11} \right] \left[1 + \frac{1}{N} \sum_{i=1}^{N} T_i \nu \cdot \sigma_{22} \right] \right.$$
$$\left. - \left[\frac{1}{N} \sum_{i=1}^{N} \left(\sum_{t=1}^{T_i} \mu_{i1t} \right) \cdot \frac{1}{N} \sum_{i=1}^{N} T_i \nu \cdot \sigma_{12}^2 \right] \right\}^{-3}.$$

Example 2: a bivariate clustered mixed model with both random intercept and random slope

Consider the bivariate clustered mixed model with both random intercept and random slope as follows :

$$\begin{aligned}
Y_{i1t}|b_{i1} &\sim \text{Poisson distribution with mean } \mu_{i1t} \text{ and variance } \mu_{i1t} \\
Y_{i2t}|b_{i2} &\sim \text{gamma distribution with mean } \mu_{i2t} \text{ and variance } \mu_{i2t}^2/\nu \\
\log(\mu_{i1t}) &= \beta_{10} + \beta_{11} x_{1,i1t} + \cdots + \beta_{1p_1} x_{p_1,i2t} + b_{i10} + b_{i11} z_{it} \\
\log(\mu_{i2t}) &= \beta_{20} + \beta_{21} x_{1,i2t} + \cdots + \beta_{2p_2} x_{p_2,i2t} + b_{i20} + b_{i21} z_{it} \\
b_i &= (b_{i10}, b_{i11}, b_{i20}, b_{i21})^T \sim \text{ iid multivariate normal} \, (0, D)
\end{aligned}$$

where $D = [\sigma_{ij}]_{i=1,\cdots,4;\ j=1,\cdots,4}$, $i = 1, \cdots, N$ and $t = 1, \cdots, T_i$. In this case, the approximate uniform shrinkage prior can be shown to be

$$\pi_D(D) \propto \left| \begin{bmatrix}
1 + \frac{1}{N}\sum_{i=1}^N S_1(i)\sigma_{11} & \frac{1}{N}\sum_{i=1}^N S_1(i)\sigma_{12} & \frac{1}{N}\sum_{i=1}^N S_1(i)\sigma_{13} & \frac{1}{N}\sum_{i=1}^N S_1(i)\sigma_{14} \\
\frac{1}{N}\sum_{i=1}^N S_2(i)\sigma_{21} & 1 + \frac{1}{N}\sum_{i=1}^N S_2(i)\sigma_{22} & \frac{1}{N}\sum_{i=1}^N S_2(i)\sigma_{23} & \frac{1}{N}\sum_{i=1}^N S_2(i)\sigma_{24} \\
\frac{1}{N}\sum_{i=1}^N T_i\nu\sigma_{31} & \frac{1}{N}\sum_{i=1}^N T_i\nu\sigma_{32} & 1 + \frac{1}{N}\sum_{i=1}^N T_i\nu\sigma_{33} & \frac{1}{N}\sum_{i=1}^N T_i\nu\sigma_{34} \\
\frac{1}{N}\sum_{i=1}^N T_i z_{i1t}^2\nu\sigma_{41} & \frac{1}{N}\sum_{i=1}^N T_i z_{i1t}^2\nu\sigma_{42} & \frac{1}{N}\sum_{i=1}^N T_i z_{i1t}^2\nu\sigma_{43} & 1 + \frac{1}{N}\sum_{i=1}^N T_i z_{i1t}^2\nu\sigma_{44}
\end{bmatrix} \right|^{-5}$$

where $S_1(i) = \sum_{t=1}^{T_i} \mu_{i1t}$ and $S_2(i) = \sum_{t=1}^{T_i} z_{i1t}^2 \mu_{i1t}$.

Example 3: a trivariate clustered mixed model with random intercept
Assume measurements are repeatedly taken from three different observers, and assume measurements from the observers follow a Poisson, gamma and normal distribution, respectively. The model is shown below :

$$
\begin{aligned}
Y_{i1t}|b_{i1} &\sim \text{Poisson distribution with mean } \mu_{i1t} \text{ and variance } \mu_{i1t} \\
Y_{i2t}|b_{i2} &\sim \text{gamma distribution with mean } \mu_{i2t} \text{ and variance } \mu_{i2t}^2/\nu \\
Y_{i3t}|b_{i3} &\sim \text{normal distribution with mean } \mu_{i3t} \text{ and variance } \sigma_N^2 \\
\log(\mu_{i1t}) &= \beta_{10} + \beta_{11}x_{1,it} + \cdots + \beta_{1p_2}x_{p_1,it} + b_{i1} \\
\log(\mu_{i2t}) &= \beta_{20} + \beta_{21}x_{1,it} + \cdots + \beta_{2p_2}x_{p_2,it} + b_{i2} \\
\mu_{i3t} &= \beta_{30} + \beta_{31}x_{1,it} + \cdots + \beta_{3p_3}x_{p_3,it} + b_{i3} \\
b_i &= (b_{i1}, b_{i2}, b_{i3})^T \sim \text{ iid multivariate normal}\,(0, D)
\end{aligned}
$$

where $D = [\sigma_{ij}]_{i=1,\cdots,3;\ j=1,\cdots,3}$, $i = 1, \cdots, N$ and $t = 1, \cdots, T_i$. Therefore, the approximate uniform shrinkage prior is

$$
\pi_D(D) \propto \left| \begin{bmatrix} 1 + \frac{1}{N}\sum_{i=1}^N S_1(i)\sigma_{11} & \frac{1}{N}\sum_{i=1}^N S_1(i)\sigma_{12} & \frac{1}{N}\sum_{i=1}^N S_1(i)\sigma_{13} \\ \frac{1}{N}\sum_{i=1}^N T_i\nu\sigma_{21} & 1 + \frac{1}{N}\sum_{i=1}^N T_i\nu\sigma_{22} & \frac{1}{N}\sum_{i=1}^N T_i\nu\sigma_{23} \\ \frac{1}{N}\sum_{i=1}^N T_i\frac{1}{\sigma_N^2}\sigma_{31} & \frac{1}{N}\sum_{i=1}^N T_i\frac{1}{\sigma_N^2}\sigma_{32} & 1 + \frac{1}{N}\sum_{i=1}^N T_i\frac{1}{\sigma_N^2}\sigma_{33} \end{bmatrix} \right|^{-3}
$$

where $S_1(i) = \sum_{t=1}^{T_i} \mu_{i1t}$.

4.3 Properties of the Approximate Uniform Shrinkage Prior

Several properties of the approximate uniform shrinkage prior will be shown in this section. First, we prove that the approximate uniform shrinkage prior distribution proposed in Section 4.2 is a probability density function. Furthermore, Natarajan and Kass (2000) have shown that in the univariate GLMM, the approximate uniform shrinkage prior is proper and leads to a proper posterior under some circumstances. In this section, we will show the extended approximate uniform shrinkage in the multivariate GLMM is prior. We will also define sufficient conditions where the corresponding posterior is proper. Some of the following proofs adjust the proofs in the appendix in Natarajan and Kass.

Theorem 1. *The approximate uniform shrinkage prior in the MGLMM is a probability density function.*

Proof. It has been shown in Section 4.2 that the approximate uniform shrinkage prior is

$$\pi_D(D) \propto \left| I_{Lq} + \left(\frac{1}{N} \sum_{i=1}^{N} Z_i^T W Z_i \right) D \right|^{-2q-1}$$

where each matrix D is a positive definite matrix.

Denote $\pi_D(D) = C \left| I_{Lq} + AD \right|^{-2q-1}$, where C is a constant and $A = \frac{1}{N} \sum_{i=1}^{N} Z_i^T W Z_i$. Since A is a positive diagonal matrix, there exists a diagonal matrix $A^{1/2}$ such that $A = (A^{1/2})^2$ and $(A^{1/2})^T = A^{1/2}$. For any column vector x of Lq real numbers and $y = A^{1/2}x$, $x^T(A^{1/2})^T D A^{1/2} x = y^T D y > 0$ since D is positive definite. Thus $A^{1/2} D A^{1/2}$ is positive definite and thus $I_{Lq} + A^{1/2} D A^{1/2}$ is positive definite, i.e., $\left| I_{Lq} + A^{1/2} D A^{1/2} \right| > 0$. According to Sylvester's determinant theorem, $\left| I_{Lq} + AD \right| = \left| I_{Lq} + A^{1/2} D A^{1/2} \right| > 0$. Hence, $\left| I_{Lq} + AD \right|^{-2q-1} > 0$.

Let $C = (\int \left| I_{Lq} + AD \right|^{-2q-1} dD)^{-1}$, a positive constant. C is finite since $\int \pi_D(D) dD < \infty$, which will be proved in Theorem 2. Therefore, $\int \pi_D(D) dD = 1$. In addition, it can be seen that $\pi_D(D) = C \left| I_{Lq} + AD \right|^{-2q-1} > 0$. Therefore $\pi_D(D)$ is a probability density function. □

Theorem 2. *The approximate uniform shrinkage prior in the MGLMM is proper.*

Proof. Consider the weight matrix S given to the prior mean in the shrinkage estimate, which is defined in the previous section. Let R={ S: all principal minors are less than one} and R_1= {S: all first-order and second-order principal minors are positive and less than one}, then R is a subset of R_1. Hence,

$$\begin{aligned}
\int_R dS &< \int_{R_1} dS \\
&= \int_0^1 \cdots \int_0^1 \prod_{i<j} \int I\left(s_{i,j}^2 \le s_{i,i}s_{j,j}\right) ds_{i,j} ds_{1,1} \cdots ds_{2N,2N} \\
&= 2^{\frac{2N(N-1)}{2}} \prod_{i=1}^{2N} \int s_{i,i}^{\frac{2N-1}{2}} ds_{i,i} < \infty
\end{aligned}$$

Thus, a uniform prior for S is integrable. Then the approximate uniform shrinkage prior $\pi_D(D)$ is integrable. That is, $\int \pi_D(D)dD < \infty$. Therefore, we can conclude that the approximate uniform shrinkage prior in the MGLMM is proper. □

Theorem 3. *Assume the data follows a MGLMM. Also assume that β and D are a priori independent, and the prior distribution of β is uniform. If there exists p full rank vectors x_k^T, where $k = 1, \cdots, p$, such that*

$$L = \int\int \prod_k \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_k|r_{1,k}, r_{2,k}, b) dr_{1,k} dr_{2,k} f(b|D) db \pi(D) dD < \infty,$$

where $r_{1,k} = x_k^T\beta_1$ and $r_{2,k} = x_k^T\beta_2$, then the posterior distribution corresponding to the approximate uniform shrinkage prior for D is proper.

Proof. First we can show that the marginal probability density function of the data $m(y)$ is finite. Assume there exist $r_{1,k} = x_k^T\beta_1$ and $r_{2,k} = x_k^T\beta_2$ for any p full rank design vectors, then the Jacobian of the transformation is $|J| = (\det(X^*))^{-1}$, where

X^* is a $p \times p$ full rank matrix with rows x_k^T. Then,

$$
\begin{aligned}
m(y) &= \int \int \int \prod_{i=1}^{N} \prod_{j=1}^{2} \prod_{t=1}^{T_i} f(y_{ijt}|\beta, b_i) f(b_i|D) db_i d\beta \pi_D(D) dD \\
&\propto |J| \int \int \int \int f(y_k|r_k, b) dr_{1,k} dr_{2,k} f(b_i|D) db_i \pi_D(D) dD \\
&\propto L < \infty
\end{aligned}
$$

Notice that the second equation holds since individual components in the likelihood are bounded and thus can be ignored. Hence, $m(y)$ is bounded above. Since the joint posterior distribution $\pi(\beta, D)$ is proper if and only if $m(y)$ is finite, the posterior distribution corresponding to the approximate uniform shrinkage prior for D is proper when the prior distribution for β is uniform. □

Corollary 4. *The posterior distribution corresponding to the uniform prior for β and the approximate uniform shrinkage prior for D is proper if measurements of each observer are assumed to follow, but not limited to, any of the following distributional families: Poisson distribution with canonical link when y_k corresponding to the full rank x_k^T are nonzero, gamma distribution with canonical link or log link, and Gaussian distribution and inverse Gaussian distribution with canonical link.*

Proof. First, assume the measurements are from a joint model of Poisson distribution with log link and gamma distribution with log link, then $\log(\mu_{1,k}) = x_k^T\beta_1 + z_{1,k}^T b_{1,k}$ and $\log(\mu_{2,k}) = x_k^T\beta_2 + z_{2,k}^T b_{2,k}$. Let $r_{1,k} = x_k^T\beta_1$ and $r_{2,k} = x_k^T\beta_2$, then $f(y_k|r_{1,k}, r_{2,k}, b_k) \propto \exp[-\exp(r_{1,k} + z_{1,k}^T b_{1,k}) + y_{1,k}(r_{1,k} + z_{1,k}^T b_{1,k}) - \nu y_{2,k} \exp\{-(r_{2,k} + z_{2,k}^T b_{2,k})\} - \nu(r_{2,k} + z_{2,k}^T b_{2,k})]$, where ν is the shape parameter in the gamma distribution. Thus,

$$
\begin{aligned}
I &= \int\int\prod_k \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_k|r_k,b)dr_{1,k}dr_{2,k}f(b|D)db\pi(D)dD \\
&\propto \int\int\prod_k \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp[-\exp(r_{1,k}+z_{1,k}^T b_{1,k}) + y_{1,k}(r_{1,k}+z_{1,k}^T b_{1,k}) \\
&\quad -\nu y_{2,k}\exp\{-(r_{2,k}+z_{2,k}^T b_{2,k})\} - \nu(r_{2,k}+z_{2,k}^T b_{2,k})]dr_{1,k}dr_{2,k}f(b|D)db\pi(D)dD \\
&\propto \int\int\prod_k \int_{0}^{\infty}\int_{0}^{\infty} \exp\{-s_{1,k}+y_{1,k}\log(s_{1,k})\}\exp\{-\nu y_{2,k}s_{2,k}+\nu\log(s_{2,k})\}s_{1,k}s_{2,k} \\
&\quad ds_{1,k}ds_{2,k}f(b|D)db\pi(D)dD \\
&= \int\int\prod_k \int_{0}^{\infty} e^{-s_{1,k}}s_{1,k}^{y_{1,k}+1}ds_{1,k}\int_{0}^{\infty} e^{-\nu y_{2,k}s_{2,k}}s_{2,k}^{\nu+1}ds_{2,k}f(b|D)db\pi(D)dD
\end{aligned}
$$

where the transformation $s_{1,k} = \exp(r_{1,k}+z_{1,k}^T b_{1,k})$ and $s_{2,k} = \exp\{-(r_{2,k}+z_{2,k}^T b_{2,k})\}$ are made in the last two equations. Then I is finite when y_{i2t} are all nonzero. Thus, the corresponding posterior distribution is proper.

The posterior distributions for measurements from a joint model of above mentioned distributions can be shown to be proper analogously. □

4.4 Posterior Distributions Simulation

In this section the Markov chain Monte Carlo (MCMC) algorithm is outlined for estimating the joint posterior distribution of fixed effect parameters and the variance components of random effects. MCMC for univariate GLMM has been discussed in several studies. For a detailed illustration of the approach, we refer to Zeger and Karim (1991).

Suppose that all observations in the data set are independent. Assume that β and D are a priori independent. Since MGLMM can be viewed as a hierarchical Bayesian model, the Bayesian inference is obtained by estimating the full distribution of each variable conditioned on all other variables. That is, we sample from $f(\beta|y,b)$, $f(D|b)$

and $f(b_i|\beta, D, y)$, respectively. The posterior distributions of β, D, and b_i are given as follows :

$$f(\beta|y,b) \propto \left\{\prod_{i=1}^{N}\prod_{j=1}^{L}\prod_{t=1}^{T_i} f(y_{ijt}|\beta, b_i)\right\}\pi(\beta)$$

$$f(D|b) \propto \left\{\prod_{i=1}^{N}\prod_{j=1}^{L} f(b_{ij}|D)\right\}\pi(D)$$

$$f(b_i|y,\beta,D) \propto \left\{\prod_{j=1}^{L}\prod_{t=1}^{T_i} f(y_{ijt}|\beta, b_i)\right\}\exp\left(-\frac{1}{2}b_i^T D^{-1} b_i\right)$$

Gibbs sampling can be used in estimating these desired posterior distributions. Given the full conditional distributions, samples are iteratively generated and collected after convergence to gain the empirical distribution and compute the posterior summaries of interest.

Rejection sampling with normal proposal distribution based on maximum likelihood estimation can be used to sample from the posterior of the fixed effects coefficients βs and the random effects b_i. However, the main computational difficulty is that the posterior is no longer inverse Wishart distribution when the approximate uniform shrinkage prior is used. Methods such as the Metropolis-Hastings algorithm can be adopted here to generate realizations from $f(D|b)$ and the inverse Wishart distribution can be chosen as the proposal distribution.

4.5 Simulation Study

A simulation study is conducted in this section to evaluate the performance of the approximate uniform shrinkage prior. Independent bivariate mixed outcomes are considered. First for each subject, the fixed covariates x_{i1} and x_{i2} are generated independently and identically from a standard normal distribution, and the random effects $b_i = (b_{i1}, b_{i2})^T$ are generated independently from a bivariate normal distribution with mean 0 and covariance matrix D. Then given the random effect b_i, the measurements Y_{i1t} and Y_{i2t} are generated independently from a Poisson distribution with mean μ_{i1t} and a gamma distribution with mean μ_{i2t} and variance μ_{i2t}^2/ν, respectively. For simplicity, the shape parameter in the fitted gamma distribution is set to be $\nu = 1$, which produces an exponential distribution. The multivariate generalized linear model with the natural logarithm as the link functions is considered. That is, we consider the following model :

$$\begin{aligned}
Y_{i1t}|b_{i1} &\sim \text{Poisson distribution with mean } \mu_{i1t} \\
Y_{i2t}|b_{i2} &\sim \text{exponential distribution with mean } \mu_{i2t} \\
\log(\mu_{i1t}) &= \beta_{10} + \beta_{11}x_{1,i1t} + \beta_{12}x_{2,i1t} + b_{i1} \\
\log(\mu_{i2t}) &= \beta_{20} + \beta_{21}x_{1,i2t} + \beta_{22}x_{2,i2t} + b_{i2} \\
b_i = (b_{i1}, b_{i2})^T &\sim \text{iid multivariate normal}\,(0, D)
\end{aligned}$$

where $D = [\sigma_{ij}]_{i=1,2;\ j=1,2}$, $i = 1, \cdots, N$ and $t = 1, \cdots, T_i$. The situation may arise when two measurements are observed from each subject repeatedly in a longitudinal study. In this study, we consider that each dataset consists of $N = 50$ or $N = 100$ clusters of size $T_i = 1$ or $T_i = 7$, respectively. The true values of the fixed effect parameters are set to be $\beta = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{20}, \beta_{21}, \beta_{22}) = (0.5, 0.3, 0.7, 0.5, 0.3, 0.7)$. The covariance matrix of the random effects is set to be

$$D = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix},$$

which implies the strongly positive correlated random effects.

Assume an improper uniform prior distribution is placed on the fixed effect parameters β. The variance components of the random effects, D, is assumed to have an approximate uniform shrinkage prior. In addition, the common Bayesian conjugate priors are also considered for comparison. The priors for the variance of the random effect used here are inverse Wishart $(2, 2I)$ and inverse Wishart $(2, 2D)$. The simulation studies are implemented using the MCMC procedure in SAS software. For each situation, 500 MCMC runs are performed, each run consisting of 10000 iterations after a burn-in of 3000 iterations. Moreover, only every 20th sample is collected. Bayesian inferences are based on the 500 samples generated from the full posterior distribution.

For each situation, the posterior summaries for each fixed effect parameters β and variance components D are shown in Table 4.1, Table 4.2, Table 4.3 and Table 4.4, respectively. In each table, the posterior mean is the average of 500 posterior means; the posterior standard deviation is the average of 500 posterior standard deviations; the standard error of posterior means is the standard error of 500 posterior means; the relative bias is the ratio of the difference between the posterior mean and the true value to the true value; HPD interval width is the average of 95% highest posterior density interval widths in 500 simulations; the coverage rate is the percentage of times that 95% HPD interval includes the true value.

As seen in these tables, most of the Bayesian estimates tend to estimate the true value very well, especially when there are more replicates within each subject. The posterior means are more sensitive to the number of replicates than to the sample size. Posterior estimates with approximate uniform shrinkage prior have the lowest bias among the three priors in many situations. On the other hand, there is no significant difference between the standard error of posterior means and average posterior standard deviations under all situations. Both the standard error of posterior means and the average posterior standard deviations decrease not only as the sample size increases but also as the number of replicates increases. The

Parameter	Method	True Value	Post Mean	Post Sd	SE of Post Mean	Relative Bias	HPE Interval Width	Coverage
β_{10}	AUS	0.5	0.4726	0.1014	0.1035	-5.50 %	0.3868	91.20 %
	IW$(2, 2I)$	0.5	0.4918	0.1043	0.1027	-1.60 %	0.3983	93.40 %
	IW$(2, 2D)$	0.5	0.4966	0.1032	0.1043	-0.70 %	0.3932	93.60 %
β_{11}	AUS	0.3	0.3016	0.0227	0.0227	0.50 %	0.0873	95.00 %
	IW$(2, 2I)$	0.3	0.3019	0.0228	0.0228	0.60 %	0.0876	95.00 %
	IW$(2, 2D)$	0.3	0.3019	0.0227	0.0229	0.60 %	0.0871	94.20 %
β_{12}	AUS	0.7	0.7002	0.0233	0.0240	0.00 %	0.0896	93.80 %
	IW$(2, 2I)$	0.7	0.7006	0.0234	0.0240	0.10 %	0.0899	94.60 %
	IW$(2, 2D)$	0.7	0.7005	0.0233	0.0241	0.10 %	0.0898	94.00 %
β_{20}	AUS	0.5	0.4805	0.1024	0.1066	-3.90 %	0.3928	91.80 %
	IW$(2, 2I)$	0.5	0.4958	0.1060	0.1049	-0.80 %	0.4069	94.00 %
	IW$(2, 2D)$	0.5	0.5006	0.1046	0.1070	0.10 %	0.4005	93.20 %
β_{21}	AUS	0.3	0.3006	0.0414	0.0414	0.20 %	0.1592	94.20 %
	IW$(2, 2I)$	0.3	0.3005	0.0419	0.0418	0.20 %	0.1606	94.40 %
	IW$(2, 2D)$	0.3	0.3004	0.0414	0.0412	0.10 %	0.1594	93.20 %
β_{22}	AUS	0.7	0.6974	0.0413	0.0388	-0.40 %	0.1589	95.40 %
	IW$(2, 2I)$	0.7	0.6980	0.0417	0.0391	-0.30 %	0.1605	95.80 %
	IW$(2, 2D)$	0.7	0.6977	0.0412	0.0387	-0.30 %	0.1586	96.20 %
D_{11}	AUS	1.0	1.0081	0.1626	0.1623	0.80 %	0.6122	94.40 %
	IW$(2, 2I)$	1.0	1.0515	0.1710	0.1627	5.20 %	0.6445	95.60 %
	IW$(2, 2D)$	1.0	1.0366	0.1676	0.1616	3.70 %	0.6306	95.80 %
D_{12}	AUS	0.9	0.8981	0.1465	0.1427	-0.20 %	0.5534	93.80 %
	IW$(2, 2I)$	0.9	0.8866	0.1508	0.1425	-1.50 %	0.5699	94.40 %
	IW$(2, 2D)$	0.9	0.9264	0.1515	0.1429	2.90 %	0.5727	95.60 %
D_{22}	AUS	1.0	0.9905	0.1637	0.1609	-0.90 %	0.6175	93.60 %
	IW$(2, 2I)$	1.0	1.0437	0.1721	0.1598	4.40 %	0.6496	96.20 %
	IW$(2, 2D)$	1.0	1.0211	0.1679	0.1602	2.10 %	0.6349	95.80 %

Table 4.1: Simulation Results for Fixed Effect Parameters and the Variance Components of the Random Effects When Each Dataset Consists of 100 Clusters of Size 7.

approximate uniform shrinkage prior usually has the smallest HPD interval width, resulting in the lower 95% HPD interval coverage probabilities. This may be due to that its posterior standard deviations are smaller than the other priors. It indicates that the approximate uniform shrinkage prior tends to be more conservative than the other two priors. Since almost all of the coverage probabilities are higher than

Parameter	Method	True Value	Post Mean	Post Sd	SE of Post Mean	Relative Bias	HPE Interval Width	Coverage
β_{10}	AUS	0.5	0.4495	0.1476	0.1506	-10.10 %	0.5662	93.60 %
	IW$(2, 2I)$	0.5	0.4573	0.1480	0.1459	-8.60 %	0.5677	95.00 %
	IW$(2, 2D)$	0.5	0.4744	0.1453	0.1448	-5.10 %	0.5572	95.40 %
β_{11}	AUS	0.3	0.2964	0.1295	0.1366	-1.20 %	0.4982	94.00 %
	IW$(2, 2I)$	0.3	0.3047	0.1349	0.1423	1.60 %	0.5187	93.20 %
	IW$(2, 2D)$	0.3	0.3007	0.1319	0.1383	0.20 %	0.5072	92.80 %
β_{12}	AUS	0.7	0.6905	0.1328	0.1255	-1.40 %	0.5105	96.00 %
	IW$(2, 2I)$	0.7	0.7108	0.1384	0.1315	1.50 %	0.5327	95.00 %
	IW$(2, 2D)$	0.7	0.7062	0.1360	0.1293	0.90 %	0.5223	95.80 %
β_{20}	AUS	0.5	0.4689	0.1570	0.1559	-6.20 %	0.6043	93.80 %
	IW$(2, 2I)$	0.5	0.4573	0.1595	0.1554	-8.50 %	0.6148	94.40 %
	IW$(2, 2D)$	0.5	0.4896	0.1558	0.1539	-2.10 %	0.5997	94.60 %
β_{21}	AUS	0.3	0.2885	0.1539	0.1610	-3.80 %	0.5935	93.80 %
	IW$(2, 2I)$	0.3	0.2910	0.1583	0.1631	-3.00 %	0.6104	92.80 %
	IW$(2, 2D)$	0.3	0.2927	0.1542	0.1632	-2.40 %	0.5941	92.20 %
β_{22}	AUS	0.7	0.6831	0.1537	0.1477	-2.40 %	0.5923	93.80 %
	IW$(2, 2I)$	0.7	0.6906	0.1590	0.1489	-1.30 %	0.6131	94.60 %
	IW$(2, 2D)$	0.7	0.6943	0.1547	0.1510	-0.80 %	0.5967	94.60 %
D_{11}	AUS	1.0	1.0848	0.2535	0.2720	8.50 %	0.9358	90.00 %
	IW$(2, 2I)$	1.0	1.1381	0.2704	0.2561	13.80 %	1.0028	94.00 %
	IW$(2, 2D)$	1.0	1.0923	0.2540	0.2458	9.20 %	0.9443	93.20 %
D_{12}	AUS	0.9	0.8855	0.2214	0.2273	-1.60 %	0.8341	90.20 %
	IW$(2, 2I)$	0.9	0.8051	0.2273	0.2183	-10.60 %	0.8619	89.60 %
	IW$(2, 2D)$	0.9	0.9459	0.2272	0.2143	5.10 %	0.8592	95.60 %
D_{22}	AUS	1.0	1.0781	0.3091	0.3153	7.80 %	1.1325	93.40 %
	IW$(2, 2I)$	1.0	1.1777	0.3289	0.2917	17.80 %	1.2264	96.00 %
	IW$(2, 2D)$	1.0	1.0815	0.2969	0.2641	8.20 %	1.1047	96.80 %

Table 4.2: Simulation Results for Fixed Effect Parameters and the Variance Components of the Random Effects When Each Dataset Consists of 100 Clusters of Size 1.

90%, the approximate uniform shrinkage prior still seems competitive.

Posterior inferences can also be evaluated in terms of the squared error risks. Hence, in order to assess the accuracy and the precision of the estimators of β, D and b_i, Table 4.5 reports the risks $E\left\{(\hat{\beta}-\beta)^T(\hat{\beta}-\beta)\right\}$, $E\left\{(\hat{D}_{11}-D_{11})^2\right\}$, $E\left\{(\hat{D}_{12}-D_{12})^2\right\}$,

Parameter	Method	True Value	Post Mean	Post Sd	SE of Post Mean	Relative Bias	HPE Interval Width	Coverage
β_{10}	AUS	0.5	0.4478	0.1445	0.1565	-10.40 %	0.5513	89.60 %
	$\mathrm{IW}(2, 2I)$	0.5	0.4874	0.1498	0.1583	-2.50 %	0.5736	92.40 %
	$\mathrm{IW}(2, 2D)$	0.5	0.4927	0.1489	0.1535	-1.50 %	0.5701	92.60 %
β_{11}	AUS	0.3	0.2993	0.0321	0.0328	-0.20 %	0.1235	93.60 %
	$\mathrm{IW}(2, 2I)$	0.3	0.2996	0.0324	0.0329	-0.10 %	0.1249	94.20 %
	$\mathrm{IW}(2, 2D)$	0.3	0.2995	0.0323	0.0329	-0.20 %	0.1246	95.00 %
β_{12}	AUS	0.7	0.7000	0.0333	0.0323	0.00 %	0.1281	95.00 %
	$\mathrm{IW}(2, 2I)$	0.7	0.7005	0.0334	0.0323	0.10 %	0.1288	94.60 %
	$\mathrm{IW}(2, 2D)$	0.7	0.7004	0.0333	0.0324	0.10 %	0.1283	94.80 %
β_{20}	AUS	0.5	0.4629	0.1462	0.1589	-7.40 %	0.5613	89.00 %
	$\mathrm{IW}(2, 2I)$	0.5	0.4970	0.1527	0.1601	-0.60 %	0.5869	93.00 %
	$\mathrm{IW}(2, 2D)$	0.5	0.4997	0.1515	0.1567	-0.10 %	0.5831	93.80 %
β_{21}	AUS	0.3	0.3005	0.0584	0.0600	0.20 %	0.2251	92.60 %
	$\mathrm{IW}(2, 2I)$	0.3	0.3002	0.0597	0.0600	0.10 %	0.2299	94.80 %
	$\mathrm{IW}(2, 2D)$	0.3	0.3005	0.0587	0.0600	0.20 %	0.2259	94.00 %
β_{22}	AUS	0.7	0.6971	0.0587	0.0572	-0.40 %	0.2261	93.60 %
	$\mathrm{IW}(2, 2I)$	0.7	0.6975	0.0599	0.0579	-0.40 %	0.2305	95.40 %
	$\mathrm{IW}(2, 2D)$	0.7	0.6974	0.0590	0.0570	-0.40 %	0.2263	93.40 %
D_{11}	AUS	1.0	1.0106	0.2334	0.2285	1.10 %	0.8627	93.60 %
	$\mathrm{IW}(2, 2I)$	1.0	1.0916	0.2551	0.2303	9.20 %	0.9394	96.20 %
	$\mathrm{IW}(2, 2D)$	1.0	1.0681	0.2482	0.2254	6.80 %	0.9147	95.60 %
D_{12}	AUS	0.9	0.9000	0.2102	0.2118	0.00 %	0.7793	91.60 %
	$\mathrm{IW}(2, 2I)$	0.9	0.8844	0.2211	0.2130	-1.70 %	0.8196	93.80 %
	$\mathrm{IW}(2, 2D)$	0.9	0.9529	0.2252	0.2127	5.90 %	0.8316	94.80 %
D_{22}	AUS	1.0	0.9987	0.2365	0.2402	-0.10 %	0.8742	91.20 %
	$\mathrm{IW}(2, 2I)$	1.0	1.0930	0.2571	0.2399	9.30 %	0.9489	95.40 %
	$\mathrm{IW}(2, 2D)$	1.0	1.0591	0.2507	0.2375	5.90 %	0.9242	94.60 %

Table 4.3: Simulation Results for Fixed Effect Parameters and the Variance Components of the Random Effects When Each Dataset Consists of 50 Clusters of Size 7.

$E\left\{(\hat{D}_{22} - D_{22})^2\right\}$, $\sum_{i=1}^{N} E\left\{\left(\hat{b}_{i1} - b_{i1}\right)^2\right\}$ and $\sum_{i=1}^{N} E\left\{\left(\hat{b}_{i2} - b_{i2}\right)^2\right\}$.

The approximate uniform shrinkage prior has similar risks to the inverse Wishart $(2, 2I)$ prior and inverse Wishart $(2, 2D)$ when estimating β, D and b_i. The risks of β and Ds decrease as the sample size increases or the number of replicates increases.

Parameter	Method	True Value	Post Mean	Post Sd	SE of Post Mean	Relative Bias	HPE Interval Width	Coverage
β_{10}	AUS	0.5	0.3995	0.2204	0.2444	-20.10 %	0.8460	90.20 %
	IW$(2, 2I)$	0.5	0.4333	0.2177	0.2311	-13.30 %	0.8348	91.80 %
	IW$(2, 2D)$	0.5	0.4554	0.2107	0.2207	-8.90 %	0.8078	92.00 %
β_{11}	AUS	0.3	0.2934	0.1889	0.1990	-2.20 %	0.7301	91.20 %
	IW$(2, 2I)$	0.3	0.3079	0.1995	0.2135	2.60 %	0.7695	92.00 %
	IW$(2, 2D)$	0.3	0.3064	0.1954	0.2040	2.10 %	0.7529	91.40 %
β_{12}	AUS	0.7	0.6905	0.1902	0.1911	-1.40 %	0.7316	93.40 %
	IW$(2, 2I)$	0.7	0.7300	0.2050	0.2067	4.30 %	0.7906	94.00 %
	IW$(2, 2D)$	0.7	0.7261	0.1980	0.2041	3.70 %	0.7650	94.20 %
β_{20}	AUS	0.5	0.4672	0.2257	0.2375	-6.60 %	0.8678	91.00 %
	IW$(2, 2I)$	0.5	0.4755	0.2316	0.2344	-4.90 %	0.8948	92.20 %
	IW$(2, 2D)$	0.5	0.5050	0.2242	0.2316	1.00 %	0.8635	93.00 %
β_{21}	AUS	0.3	0.2859	0.2240	0.2235	-4.70 %	0.8632	93.80 %
	IW$(2, 2I)$	0.3	0.2943	0.2328	0.2310	-1.90 %	0.8980	93.80 %
	IW$(2, 2D)$	0.3	0.2960	0.2253	0.2264	-1.30 %	0.8691	92.80 %
β_{22}	AUS	0.7	0.6756	0.2235	0.2205	-3.50 %	0.8644	94.40 %
	IW$(2, 2I)$	0.7	0.6901	0.2312	0.2249	-1.40 %	0.8918	93.60 %
	IW$(2, 2D)$	0.7	0.6984	0.2236	0.2234	-0.20 %	0.8637	93.00 %
D_{11}	AUS	1.0	1.1432	0.4039	0.3937	14.30 %	1.4477	92.60 %
	IW$(2, 2I)$	1.0	1.2143	0.4234	0.3542	21.40 %	1.5150	97.40 %
	IW$(2, 2D)$	1.0	1.1490	0.3872	0.3214	14.90 %	1.3922	97.60 %
D_{12}	AUS	0.9	0.8647	0.3262	0.3159	-3.90 %	1.2198	91.80 %
	IW$(2, 2I)$	0.9	0.7644	0.3354	0.2973	-15.10 %	1.2570	90.40 %
	IW$(2, 2D)$	0.9	0.9735	0.3370	0.2870	8.20 %	1.2486	96.40 %
D_{22}	AUS	1.0	1.1339	0.4570	0.4174	13.40 %	1.6389	93.60 %
	IW$(2, 2I)$	1.0	1.2570	0.4890	0.3876	25.70 %	1.7599	98.00 %
	IW$(2, 2D)$	1.0	1.1372	0.4301	0.3390	13.70 %	1.5455	97.60 %

Table 4.4: Simulation Results for Fixed Effect Parameters and the Variance Components of the Random Effects When Each Dataset Consists of 50 Clusters of Size 1.

But the risks of b_is decreases as the sample size decreases or the number of replicates increases. In conclusion, the approximate uniform shrinkage prior has a good overall performance in the simulation study.

Method	β	D_{11}	D_{12}	D_{22}	$b_{\cdot 1}$	$b_{\cdot 2}$
			$N = 100$ $m = 7$			
AUS	0.0287	0.0259	0.0202	0.0259	7.7441	10.8635
IW$(2, 2I)$	0.0270	0.0285	0.0203	0.0269	7.7257	10.8783
IW$(2, 2D)$	0.0275	0.0272	0.0211	0.0259	7.6974	10.8110
			$N = 100$ $m = 1$			
AUS	0.1329	0.0810	0.0518	0.1053	34.0333	39.5811
IW$(2, 2I)$	0.1354	0.0845	0.0566	0.1165	34.4263	41.3977
IW$(2, 2D)$	0.1306	0.0688	0.0479	0.0763	33.7481	38.6195
			$N = 50$ $m = 7$			
AUS	0.0627	0.0522	0.0448	0.0576	4.6146	6.1263
IW$(2, 2I)$	0.0598	0.0613	0.0455	0.0661	4.5986	6.1725
IW$(2, 2D)$	0.0571	0.0554	0.0479	0.0598	4.4571	5.9452
			$N = 50$ $m = 1$			
AUS	0.3023	0.1752	0.1008	0.1918	18.7341	21.8448
IW$(2, 2I)$	0.3062	0.1711	0.1066	0.2160	19.1468	23.2090
IW$(2, 2D)$	0.2890	0.1253	0.0876	0.1335	18.3096	20.8748

Table 4.5: Risk for β, D_{11}, D_{12}, D_{22}, $b_{\cdot 1}$, and $b_{\cdot 2}$.

4.6 Case Example : Data From The Osteoarthritis Initiative

To illustrate the methodology, we considered an osteoarthritis data from the Osteoarthritis Initiative (OAI) database, which is available for public access at http://www.oai.ucsf.edu/ and is described in detail by McCulloch (2008). Osteoarthritis Initiative is a cohort study of the determinants of knee osteoarthritis for people aged 45 and above. The data were collected from persons at high risk for developing knee osteoarthritis at baseline, 12 months, 24 months, 36 months and 48 months, resulting in five measurements per individual. We restrict our study to the complete data, which reduces our data to 1499 individuals. The outcomes of interest are the Western Ontario and McMaster Universities (WOMAC) disability scores and the numbers of workdays missed in past 3 months. WOMAC is a numeric score used to assess pain, stiffness, and physical function in patients with hip and/or knee osteoarthritis, while the number of days of missed work due to knee pain, aching or stiffness in past 3 months is a count variable. In this study, we use the average of WOMAC for left knee and right knee as the WOMAC score. The predictor variables of primary interest in this study are the age, sex and body mass index (BMI), where age and BMI are continuous variables, and sex is a categorical variable.

To accommodate such a clustered mixed outcome data, we consider a multivariate generalized linear mixed model with subject-specific random effects. Assume that conditional on the random effects, the WOMAC disability scores follow a normal distribution and the numbers of workdays missed in the past 3 months follow a negative binomial distribution since negative binomial distribution can be used to accommodate overdispersion in count data. The dispersion parameter in the negative binomial distribution is the inverse of its shape parameter, say δ_N. The negative binomial distribution approaches a Poisson distribution when the overdispersion parameter approaches infinity, i.e., when δ_N approaches zero. The normal means and Poisson means are related to the covariates via the identity link and logarithm link, respectively. More specifically, the data are accommodated by the following model :

$$\begin{aligned}
WOMAC_{i1t}|b_{i1} &\sim \text{normal}\left(\mu_{i1t}, \sigma_N^2\right) \\
MISSW_{i2t}|b_{i2} &\sim \text{negative binomial}\left(\mu_{i2t}, \delta_N\right) \\
\mu_{i1t} &= \beta_{10} + \beta_{11} AGE_{it} + \beta_{12} SEX_i + \beta_{13} BMI_{it} + b_{i1} \\
\log(\mu_{i2t}) &= \beta_{20} + \beta_{21} AGE_{it} + \beta_{22} SEX_i + \beta_{23} BMI_{it} + b_{i2} \\
b_i = (b_{i1}, b_{i2})^T &\sim \text{ iid multivariate normal}\,(0, D)
\end{aligned}$$

where $D = [\sigma_{ij}]_{i=1,2;\ j=1,2}$, $i = 1, \cdots, N$ and $t = 1, \cdots, T$. In this case, $N = 1499$ and $T = 5$.

Assume that the prior for the fixed effect coefficients βs is uniform, the prior for the variance σ_N^2 of the normal distribution is inverse gamma(α_N, β_N), and the prior for the overdispersion parameter δ_N of the negative binomial distribution is inverse gamma(α_D, β_D). We set $\alpha_N = 10$, $\beta_N = 1$, $\alpha_D = 100$, and $\beta_D = 1$ in this study. Since the inverse gamma distribution is the conjugate prior for the variance in normal distribution, then the posterior distribution of σ_N^2 is inverse gamma$\left(\frac{NT}{2} + \alpha_N, \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(y_{i1t}-\mu_{i1t})^2}{2} + \beta_N\right)$. However, unlike σ_N^2, it is difficult to directly sample from the posterior distribution of δ_N so that sampling techniques are needed.

We set the prior for the variance components of the random effects to be either the approximate uniform shrinkage prior or the inverse Wishart$(2, 2I)$. Under the circumstances, the approximate uniform shrinkage prior for D is

$$\pi_D(D) \propto \left\{ \left[1 + \frac{\sigma_{11}}{N}\sum_{i=1}^{N}\frac{n_i}{\sigma_N^2}\right]\left[1 + \frac{\sigma_{22}}{N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T_i}\frac{\delta_N\mu_{i2t}^b}{\delta_N + \mu_{i2t}^b}\right)\right] - \left[\left(\frac{\sigma_{12}}{N}\right)^2\sum_{i=1}^{N}\frac{T_i}{\sigma_N^2}\sum_{i=1}^{N}\left(\sum_{t=1}^{n_i}\frac{\delta_N\mu_{i2t}^b}{\delta_N + \mu_{i2t}^b}\right)\right]\right\}^{-3}.$$

Analysis is based on 500 samples obtained from one single chain retaining every 20th simulation iteration in 10000 iterations after a burn-in of 2000 iterations. The posterior simulation results, including the posterior means, the posterior standard deviations and the 95% highest probability density intervals, are presented in Table 4.6. For comparative purposes, data are also fitted by maximum likelihood using adaptive Gaussian quadrature. The estimates of parameters and their standard deviations and 95% confidence interval are also reported.

	approximate uniform shrinkage prior			inverse-Wishart(2,2I)			adaptive Gaussian quadrature		
Para	Post. Mean	Post. Std	95% HPD	Post. Mean	Post. Std	95% HPD	Estimate	Std. Error	95% CI
β_{10}	-0.4387	0.0745	(-0.5806 , -0.3134)	-0.4207	0.176	(-0.7608 , -0.0807)	-0.4550	0.1846	(-0.8172 , -0.0928)
β_{11}	-0.00269	0.00184	(-0.00636 , 0.00087)	-0.00289	0.00244	(-0.00767 , 0.00146)	-0.0023	0.0026	(-0.0073 , 0.0027)
β_{12}	0.2148	0.0411	(0.1394 , 0.2926)	0.2126	0.0401	(0.134 , 0.2912)	0.2077	0.0390	(0.1312 , 0.2843)
β_{13}	0.0514	0.00298	(0.0454 , 0.0567)	0.0511	0.00378	(0.0448 , 0.0594)	0.0513	0.0038	(0.0438 , 0.0587)
β_{20}	-10.7088	0.2757	(-11.2364 , -10.1901)	-8.9397	1.6514	(-12.1228 , -5.9527)	-10.0150	1.8614	(-13.6662 , -6.3637)
β_{21}	-0.0051	0.0172	(-0.0351 , 0.0305)	-0.0124	0.0233	(-0.0571 , 0.0370)	-0.0408	0.0258	(-0.0913 , 0.0097)
β_{22}	0.1594	0.2709	(-0.2877 , 0.6628)	0.1359	0.3406	(-0.5779 , 0.7207)	0.0577	0.3638	(-0.6560 ,0.7713)
β_{23}	0.2375	0.0313	(0.1738 , 0.2946)	0.2062	0.0321	(0.1493 , 0.2689)	0.2213	0.0369	(0.1490 , 0.2936)
σ_{11}	0.4083	0.0211	(0.367 , 0.4446)	0.4128	0.0221	(0.3737 , 0.4580)	0.4062	0.0210	(0.3649 , 0.4475)
σ_{12}	1.0684	0.1678	(0.7777 , 1.3909)	0.8956	0.1326	(0.6411 , 1.1484)	1.1670	0.1597	(0.8538 , 1.4802)
σ_{22}	6.7426	1.4102	(4.1907 , 9.757)	4.6397	1.3006	(2.5209 , 7.4504)	11.9830	2.2314	(7.6059 , 16.3600)
σ_N^2	0.8064	0.0138	(0.7792 , 0.8336)	0.8037	0.0149	(0.7759 , 0.8343)	0.8081	0.0148	(0.7791 , 0.8371)
δ_N^2	0.0173	0.00169	(0.0137 , 0.0201)	0.0165	0.00189	(0.0134 , 0.0204)	0.0586	0.0090	(0.0410 , 0.0762)

Table 4.6: Parameter Estimates for Osteoarthritis Study Based on Normal-Negative Binomial Model.

Most Bayesian estimates using approximate uniform shrinkage prior have similar values with the Bayesian estimates using inverse Wishart prior and maximum likelihood estimates using adaptive Gaussian quadrature, but those estimates are very different for σ_{22}. The age effect is not a significant in both WOMAC disability scores and the numbers of workdays missed in past 3 months, the sex effect is significant only for WOMAC disability scores, the BMI effect is significant in both WOMAC disability scores and the numbers of workdays missed in past 3 months. The subject-specific random effects are significant and there are moderate correlation between the two measurements.

4.7 Conclusion

A need for noninformative priors arises when there is insufficient prior information on the model parameters. In this study, we introduced an approximate uniform shrinkage prior in the multivariate generalized linear mixed model. This prior can be reduced to the approximate uniform shrinkage prior proposed by Natarajan and Kass in the univariate case.

In this study we have shown that the approximate uniform shrinkage prior is not only easy to implement, but also possess several desirable properties. This prior is proper and leads to a proper posterior distribution for numerous common distributions under MGLMM.

5. SUMMARY AND FUTURE RESEARCH

Studies of clustered data, such as repeated measurements in a longitudinal study, become more and more common in scientific research. Multiple measurements for each subject are often taken repeatedly in either a quantitative or qualitative scale by different observers. Under this circumstance, multivariate generalized linear models can accommodate such clustered mixed data from two or more observers by joint modeling the multivariate outcomes.

Investigating the relationship among measurements from different observers on the given subject and the relationship among measurements taken by the same observer on different subjects is useful and important in scientific studies. It would be useful to have indices to assess the association and consistency between clustered mixed data. In this study, three different types of correlation coefficients which measure various linear relationship between replicated measurements from different observers are proposed. The intra-CC measures the within-observer correlation, the inter-CC measures the between-observer correlation, and the total-CC measures the overall correlation. These indices are natural extensions of the intra-CC, inter-CC and total-CC proposed by Lin et al. (2007) and Carrasco (2010), and are very useful for measuring consistency in clustered mixed data. A cluster mixed data is considered in this study and is modelled by a multivariate generalized linear mixed model. The estimates of these indices are obtained from the maximum likelihood estimates for the parameters of the underlying distributions. Confidence intervals and further statistical inference are derived based on the assuming asymptotic normality of these estimates. When there are more than two observers, the extended intra-CC, inter-CC and total-CC are defined and are the weighted averages of all pairwise inter-CC and total-CC.

Since these CC estimates are developed in terms of the sample estimates of the fixed effect parameters, the variance components of the random effects, and possible additional model parameters in the conditional distributions in the MGLMM, the precision and accuracy of these CC estimates are associated to the appropriateness of the model and the approximation to the likelihood. However, maximum likelihood inference for MGLMM is very complicated due to the fact that the link function may be non-linear. Several methods are proposed to solve the estimation and inference in MGLMM, such as the adaptive Gaussian-Hermite quadrature, Monte Carlo EM algorithm, generalized estimating equations approach, and penalized quasi-likelihood. Therefore, the approximation to the likelihood also influence these CC estimates. In addition, the results from the simulation study imply that the bias grows when the number of subjects is smaller. These CC estimates are robust especially for large sample sizes and are not sensitive to the number of replicates. As a result of an increase in correlation between random effects of observers, a larger value of inter-CC and total-CC would exist.

The disadvantage of CCC in the past researches is that it cannot produce negative values since it is expressed in terms of variance components. However, the intra-CC, inter-CC and total-CC proposed in this study allow negative correlations. They give more flexibility in modeling. Furthermore, clustered mixed data is fitted by MGLMM in this study, where the distributional assumption is required. Nonparametric methods for evaluating the correlation can be investigated in future research.

In the Bayesian approach to GLMM, the choices of prior distribution may greatly influence inferences, especially when the number of subjects is small. Noninformative priors are needed when there is insufficient prior information on the model parameters. In this study, we introduced the approximate uniform shrinkage prior in the multivariate generalized linear mixed model. This prior is obtained by placing a uniform distribution on the weight given to the prior mean in the approximate shrinkage estimate of the random effects, and then transforming it to find the distribution of the variance components of the random effects. It is noteworthy to mention that when two observers are assumed to be independent and identically distributed or when there is only one observer, then the MGLMM reduces to an ordinary GLMM, thus the proposed approximate uniform shrinkage prior reduces to the approximate uniform shrinkage prior proposed by Natarajan and Kass (2000).

In this study we have shown that the approximate uniform shrinkage prior is not only easy to implement, but also possess several attractive properties. This prior is proper and leads to a proper posterior distribution for numerous common distributions under MGLMM. In addition to bivariate generalized linear mixed model with random intercept, we have shown that the approximate uniform shrinkage prior can be applied to more complicated models, such as bivariate generalized linear mixed model with both random intercept and random slope and the trivariate generalized linear mixed model with random intercept. The extension to higher dimensional models is quite straightforward. This prior is very flexible in diverse models.

Simulation studies are conducted to evaluate the performance of the approximate uniform shrinkage prior. The methodology is also illustrated through an analysis of real world data in the osteoarthritis study. This prior seems to perform as well as the commonly used prior, inverse Wishart prior, and even better under some circumstances. However, the disadvantage of this prior is that the computation time is longer than the inverse Wishart prior.

In conclusion, the proposed CC estimates and the approximate uniform shrinkage prior are both very useful in the multivariate generalized linear mixed model. More complicated model structure can be considered in future research.

REFERENCES

An, L., Nkurunziza, S., Fung, K. Y., Krewski, D., and Luginaah, I. (2009), "Shrinkage estimation in general linear models," *Computational Statistics & Data Analysis*, 53, 2537–2549.

Barnhart, H. X., Haber, M., and Song, J. L. (2002), "Overall concordance correlation coefficient for evaluating agreement among multiple observers," *Biometrics*, 58, 1020–1027.

Barnhart, H. X., Song, J. L., and Haber, M. J. (2005), "Assessing intra, inter and total agreement with replicated readings," *Statistics in Medicine*, 24, 1371–1384.

Breslow, N. E. and Clayton, D. G. (1993), "Approximate inference in generalized linear mixed models," *Journal of the American Statistical Association*, 88, 9–25.

Carrasco, J. L. (2010), "A generalized concordance correlation coefficient based on the variance components generalized linear mixed models for overdispersed count data," *Biometrics*, 66, 897–904.

Carrasco, J. L. and Jover, L. (2003), "Estimating the generalized concordance correlation coefficient through variance components," *Biometrics*, 59, 849–858.

— (2005), "Concordance correlation coefficient applied to discrete data," *Statistics in Medicine*, 24, 4021–4034.

Carrasco, J. L., King, T. S., and Chinchilli, V. M. (2009), "The concordance correlation coefficient for repeated measures estimated by variance components," *Journal of Biopharmaceutical Statistics*, 19, 90–105.

Cohen, J. (1960), "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, 20, 37–46.

— (1968), "Weighted kappa - nominal scale agreement with provision for scaled disagreement or partial credit," *Psychological Bulletin*, 70, 213–220.

Coull, B. A. and Agresti, A. (2000), "Random effects modeling of multiple binomial responses using the multivariate binomial logit-normal distribution," *Biometrics*, 56, 73–80.

Dunson, D. B. (2000), "Bayesian latent variable models for clustered mixed outcomes," *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 62, 355–366.

Gelman, A. (2006), "Prior distributions for variance parameters in hierarchical models," *Bayesian Analysis*, 1, 515–533.

Green, P. J. (1987), "Penalized likelihood for general semiparametric regression-models," *International Statistical Review*, 55, 245–259.

Gueorguieva, R. (2001), "A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family," *Statistical Modelling*, 1, 177–193.

Gueorguieva, R. V. and Agresti, A. (2001), "A correlated probit model for joint modeling of clustered binary and continuous responses," *Journal of the American Statistical Association*, 96, 1102–1112.

Harville, D. A. (1977), "Maximum likelihood approaches to variance component estimation and to related problems," *Journal of the American Statistical Association*, 72, 320–338.

Ibrahim, J. G. and Laud, P. W. (1991), "On Bayesian analysis of generalized linear models using Jeffreys's prior," *Journal of the American Statistical Association*, 86, 981–986.

King, T. S. and Chinchilli, V. M. (2001), "A generalized concordance correlation coefficient for continuous and categorical data," *Statistics in Medicine*, 20, 2131–2147.

King, T. S., Chinchilli, V. M., and Carrasco, J. L. (2007a), "A repeated measures concordance correlation coefficient," *Statistics in Medicine*, 26, 3095–3113.

King, T. S., Chinchilli, V. M., Wang, K. L., and Carrasco, J. L. (2007b), "A class of repeated measures concordance correlation coefficients," *Journal of Biopharmaceutical Statistics*, 17, 653–672.

Lin, L., Hedayat, A. S., and Wu, W. (2007), "A unified approach for assessing agreement for continuous and categorical data," *Journal of Biopharmaceutical Statistics*, 17, 629–652.

Lin, L. I. (1989), "A concordance correlation-coefficient to evaluate reproducibility," *Biometrics*, 45, 255–268.

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, London: Chapman and Hall.

McCulloch, C. (2008), "Joint modelling of mixed outcome types using latent variables," *Statistical Methods in Medical Research*, 17, 53–73.

Natarajan, R. and Kass, R. E. (2000), "Reference Bayesian methods for generalized linear mixed models," *Journal of the American Statistical Association*, 95, 227–237.

Natarajan, R. and McCulloch, C. E. (1995), "A note on the existence of the posterior distribution for a class of mixed models for binomial responses," *Biometrika*, 82, 639–643.

Pearson, K., Lee, A., Warren, E., Fry, A., and Fawcett, C. D. (1901), "Mathematical contributions to the theory of evolution. IX. On the principle of homotyposis and its relation to heredity, to the variability of individual, and to that of the race. Part I. Homotyposis in the vegetable kingdom," *Philosophical Transactions of the Royal Society of London Series a-Containing Papers of a Mathematical or Physical Character*, 197, 285–379.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley series in probability and mathematical statistics, New York, N.Y.: Wiley.

Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W. R. (1996), "BUGS 0.5 . Bayesian inference using Gibbs sampling," *Cambridge: Medical Research Council Biostatistics Unit*.

Strawderman, W. E. (1971), "Proper Bayes minimax estimators of multivariate normal mean," *Annals of Mathematical Statistics*, 42, 385.

Tiao, G. C. and Tan, W. Y. (1965), "Bayesian analysis of random-effect models in analysis of variance I posterior distribution of variance components," *Biometrika*, 52, 37.

Tierney, L. (1994), "Markov chains for exploring posterior distributions," *Annals of Statistics*, 22, 1701–1728.

Zeger, S. L. and Karim, M. R. (1991), "Generalized linear-models with random effects

- a Gibbs sampling approach," *Journal of the American Statistical Association*, 86, 79–86.

APPENDIX A

ASSESSING CORRELATION OF CLUSTERED MIXED OUTCOMES FROM A MULTIVARIATE GENERALIZED LINEAR MIXED MODEL

A.1 Derivation of correlations in joint modeling of Poisson-gamma bivariate outcomes

Based on the model proposed in Section 3.4.1, we can compute the marginal mean and variance of the outcomes, and the covariate of the conditional means (McCulloch, 2008). The marginal means for Y_{i1t} and Y_{i2t} are

$$\mathrm{E}(Y_{i1t}) = \mathrm{E}\{\mathrm{E}(Y_{i1t}|b_{i1})\} = e^{\beta_{10}+\beta_{11}x_{1,i1t}+\cdots+\beta_{1p_1}x_{p_1,i1t}+\sigma^2_{b_1}/2}$$
$$\mathrm{E}(Y_{i2t}) = \mathrm{E}\{\mathrm{E}(Y_{i2t}|b_{i2})\} = e^{\beta_{20}+\beta_{21}x_{1,i2t}+\cdots+\beta_{2p_2}x_{p_2,i2t}+\sigma^2_{b_2}/2}$$

and the marginal variances of Y_{i1t} and Y_{i2t} are

$$\begin{aligned}
\mathrm{Var}(Y_{i1t}) &= \mathrm{E}\{\mathrm{Var}(Y_{i1t}|b_{i1})\} + \mathrm{Var}\{\mathrm{E}(Y_{i1t}|b_{i1})\} \\
&= e^{\beta_{10}+\beta_{11}x_{1,i1t}+\cdots+\beta_{1p_1}x_{p_1,i1t}+\sigma^2_{b_1}/2} + e^{2\left(\beta_{10}+\beta_{11}x_{1,i1t}+\cdots+\beta_{1p_1}x_{p_1,i1t}\right)}\left(e^{2\sigma_1^2} - e^{\sigma_1^2}\right) \\
\mathrm{Var}(Y_{i2t}) &= \mathrm{E}\{\mathrm{Var}(Y_{i2t}|b_{i2})\} + \mathrm{Var}\{\mathrm{E}(Y_{i2t}|b_{i2})\} \\
&= e^{2\left(\beta_{20}+\beta_{21}x_{1,i2t}+\cdots+\beta_{2p_2}x_{p_2,i2t}+\sigma^2_{b_2}\right)}/\nu + e^{2\left(\beta_{20}+\beta_{21}x_{1,i2t}+\cdots+\beta_{2p_2}x_{p_2,i2t}\right)}\left(e^{2\sigma_2^2} - e^{\sigma_2^2}\right).
\end{aligned}$$

This is due to the fact that if $Z \sim N(0,\sigma^2)$, then $\mathrm{E}\left(e^Z\right) = e^{\sigma^2/2}$ and $\mathrm{Var}\left(e^Z\right) = e^{2\sigma^2} - e^{\sigma^2}$.

Moreover, the covariances of the conditional means of the first and second observer can be developed as

$$
\begin{aligned}
\mathrm{Cov}(\mu_{i1t}, \mu_{i1t'}) &= e^{\beta_{10}+\beta_{11}x_{1,i1t}+\cdots+\beta_{1p_1}x_{p_1,i1t}} e^{\beta_{10}+\beta_{11}x_{1,i1t'}+\cdots+\beta_{1p_1}x_{p_1,i1t'}} \left(e^{2\sigma^2_{b_1}} - e^{\sigma^2_{b_1}}\right) \\
&= \mathrm{E}(\mu_{i1t})\,\mathrm{E}(\mu_{i1t'}) \left(e^{\sigma^2_{b_1}} - 1\right) \\
\mathrm{Cov}(\mu_{i2t}, \mu_{i2t'}) &= e^{\beta_{20}+\beta_{21}x_{1,i2t}+\cdots+\beta_{2p_2}x_{p_2,i2t}} e^{\beta_{20}+\beta_{21}x_{1,i2t'}+\cdots+\beta_{2p_2}x_{p_2,i2t'}} \left(e^{2\sigma^2_{b_2}} - e^{\sigma^2_{b_2}}\right) \\
&= \mathrm{E}(\mu_{i2t})\,\mathrm{E}(\mu_{i2t'}) \left(e^{\sigma^2_{b_2}} - 1\right)
\end{aligned}
$$

The covariance of the conditional means of the first and second observer is

$$
\mathrm{Cov}(\mu_{i1t}, \mu_{i2t'}) = e^{\beta_{10}+\beta_{11}x_{1,i1t}+\cdots+\beta_{1p_1}x_{p_1,i1t}+\beta_{20}+\beta_{21}x_{1,i2t}+\cdots+\beta_{2p_2}x_{p_2,i2t}}\,\mathrm{Cov}(e^{b_{i1}}, e^{b_{i2}})
$$

To calculate the covariance in the final term, we rewrite the two random effects in terms of three i.i.d. standard normal variables Z_is as

$$
\begin{aligned}
b_{i1} &= \sigma_{b_1}\left\{Z_1\sqrt{1-|\rho_b|} + Z_3\sqrt{|\rho_b|}\right\} \\
b_{i2} &= \sigma_{b_2}\left\{Z_2\sqrt{1-|\rho_b|} + Z_3\,\mathrm{sgn}(\rho_b)\sqrt{|\rho_b|}\right\}
\end{aligned}
$$

The covariance in the final term can be written as

$$
\begin{aligned}
\operatorname{Cov}(e^{b_{i1}}, e^{b_{i2}}) =& \operatorname{E}(e^{b_{i1}} e^{b_{i2}}) - \operatorname{E}(e^{b_{i1}}) \operatorname{E}(e^{b_{i2}}) \\
& - \operatorname{E}\left\{ e^{\sigma_{b_1}\left(Z_1\sqrt{1-|\rho_b|}+Z_3\sqrt{|\rho_b|}\right)} \right\} \operatorname{E}\left\{ e^{\sigma_{b_2}\left(Z_2\sqrt{1-|\rho_b|}+Z_3 sgn(\rho_b)\sqrt{|\rho_b|}\right)} \right\} \\
=& \operatorname{E}\left\{ e^{Z_1\left(\sigma_{b_1}\sqrt{1-|\rho_b|}\right)} \right\} \operatorname{E}\left\{ e^{Z_2\left(\sigma_{b_2}\sqrt{1-|\rho_b|}\right)} \right\} \\
& \times \left[\operatorname{E}\left\{ e^{Z_3\left(\sigma_{b_1}\sqrt{|\rho_b|}+\sigma_{b_2} sgn(\rho_b)\sqrt{|\rho_b|}\right)} \right\} \right. \\
& \left. - \operatorname{E}\left\{ e^{Z_3\left(\sigma_{b_1}\sqrt{|\rho_b|}\right)} \right\} \operatorname{E}\left\{ e^{Z_3\left(\sigma_{b_2} sgn(\rho_b)\sqrt{|\rho_b|}\right)} \right\} \right] \\
=& e^{\left(\sigma_{b_1}\sqrt{1-|\rho_b|}\right)^2/2} e^{\left(\sigma_{b_2}\sqrt{1-|\rho_b|}\right)^2/2} \\
& \times \left\{ e^{\left(\sigma_{b_1}\sqrt{|\rho_b|}+\sigma_{b_2} sgn(\rho_b)\sqrt{|\rho_b|}\right)^2/2} - e^{\left(\sigma_{b_1}\sqrt{|\rho_b|}\right)^2/2} e^{\left(\sigma_{b_2} sgn(\rho_b)\sqrt{|\rho_b|}\right)^2/2} \right\} \\
=& e^{\sigma_{b_1}^2/2} e^{\sigma_{b_2}^2/2} \left(e^{\rho_b\sigma_{b_1}\sigma_{b_2}} - 1\right).
\end{aligned}
$$

Thus the covariance of the conditional means is

$$
\begin{aligned}
\operatorname{Cov}(\mu_{i1t}, \mu_{i2t'}) =& \exp\left(\beta_{10} + \beta_{11}x_{1,i1t} + \cdots + \beta_{1p_1}x_{p_1,i1t} + \beta_{20} + \beta_{21}x_{1,i2t} + \cdots \right. \\
& \left. + \beta_{2p_2}x_{p_2,i2t} + \sigma_{b_1}^2/2 + \sigma_{b_2}^2/2\right) \times \left(e^{\rho_b\sigma_{b1}\sigma_{b2}} - 1\right),
\end{aligned}
$$

which can further be expressed in terms of $\operatorname{E}(\mu_{i1t})$ and $\operatorname{E}(\mu_{i2t})$ as

$$
\operatorname{Cov}(\mu_{i1t}, \mu_{i2t'}) = \operatorname{E}(\mu_{i1t}) \operatorname{E}(\mu_{i2t'}) \left(e^{\rho_b\sigma_{b_1}\sigma_{b_2}} - 1\right).
$$

Therefore, the intra-CC of the measurements from the first observer is

$$
\rho^{intra,i,1} = \frac{\operatorname{E}(\mu_{i1t}) \operatorname{E}(\mu_{i1t'}) \left(e^{\sigma_{b_1}^2} - 1\right)}{\sqrt{\left\{\operatorname{E}(\mu_{i1t}) + \operatorname{E}(\mu_{i1t})^2 \left(e^{\sigma_{b_1}^2} - 1\right)\right\} \left\{\operatorname{E}(\mu_{i1t'}) + \operatorname{E}(\mu_{i1t'})^2 \left(e^{\sigma_{b_1}^2} - 1\right)\right\}}}
$$

and the intra-CC of the measurements from the second observer is

$$\rho^{intra,i,2} = \frac{\mathrm{E}(\mu_{i2t})\,\mathrm{E}(\mu_{i2t'})\left(e^{\sigma_{b_2}^2}-1\right)}{\sqrt{\left[\mathrm{E}(\mu_{i2t})^2\left\{\left(1+\frac{1}{\nu}\right)e^{\sigma_{b_2}^2}-1\right\}\right]\left[\mathrm{E}(\mu_{i2t'})^2\left\{\left(1+\frac{1}{\nu}\right)e^{\sigma_{b_2}^2}-1\right\}\right]}}.$$

To obtain overall intra-CCs for first and second observers, we replace $\mathrm{E}(\mu_{i1t})$ and $\mathrm{E}(\mu_{i2t})$ with the marginal expectations over X, $\mu_1^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i1t})\}$ and $\mu_2^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i2t})\}$. Hence, the overall intra-CCs are

$$\rho^{intra,1} = \frac{\mu_1^*(e^{\sigma_{b_1}^2}-1)}{1+\mu_1^*(e^{\sigma_{b_1}^2}-1)}$$

and

$$\rho^{intra,2} = \frac{e^{\sigma_{b_2}^2}-1}{\left(1+\frac{1}{\nu}\right)e^{\sigma_{b_2}^2}-1}.$$

Also notice that

$$\begin{aligned}\mathrm{Var}\left(\sum_{t=1}^{T_i}\mu_{i1t}\right) &= \left(\sum_{t=1}^{T_i} e^{\beta_{10}+\beta_{11}x_{1,i1t}+\cdots+\beta_{1p_1}x_{p_1,i1t}+b_{i1}}\right)^2\left(e^{2\sigma_{b_1}^2}-e^{\sigma_{b_1}^2}\right)\\ &= \left\{\sum_{t=1}^{T_i}\mathrm{E}(\mu_{i1t})\right\}^2\left(e^{\sigma_{b_1}^2}-1\right).\end{aligned}$$

Similarly,

$$\mathrm{Var}\left(\sum_{t'=1}^{T_i}\mu_{i2t'}\right) = \left\{\sum_{t'=1}^{T_i}\mathrm{E}(\mu_{i2t'})\right\}^2\left(e^{\sigma_{b_2}^2}-1\right).$$

In addition,

$$\mathrm{E}\left(\mu_{i2t'}^2\right) = e^{2(\beta_{20}+\beta_{21}x_{1,i2t}+\cdots+\beta_{2p_2}x_{p_2,i2t})}e^{2\sigma_{b_2}^2} = \left\{\mathrm{E}(\mu_{i2t'})\right\}^2 e^{\sigma_{b_2}^2}.$$

Combining the above equations, the inter-CC of the bivariate measurements is equal to

$$\rho^{inter} = \frac{\sum_{t=1}^{T_i}\sum_{t'=1}^{T_i} \mathrm{E}(\mu_{i1t})\,\mathrm{E}(\mu_{i2t'})\,(e^{\rho_b\sigma_{b_1}\sigma_{b_2}}-1)}{\sqrt{\left[\sum_{t=1}^{T_i}\mathrm{E}(\mu_{i1t}) + \left\{\sum_{t=1}^{T_i}\mathrm{E}(\mu_{i1t})\right\}^2\left(e^{\sigma^2_{b_1}}-1\right)\right]\left[\frac{e^{\sigma^2_{b_2}}}{\nu}\sum_{t'=1}^{T_i}\mathrm{E}(\mu_{i2t'})^2 + \left\{\sum_{t'=1}^{T_i}\mathrm{E}(\mu_{i2t'})\right\}^2\left(e^{\sigma^2_{b_2}}-1\right)\right]}}.$$

We replace $\sum_{t=1}^{T_i}\mathrm{E}(\mu_{i1t})$ and $\sum_{t=1}^{T_i}\mathrm{E}(\mu_{i2t})$ with $T^*\mu_1^*$ and $T^*\mu_2^*$, and an overall inter-CC is defined as

$$\rho^{inter} = \frac{e^{\rho_b\sigma_{b_1}\sigma_{b_2}}-1}{\sqrt{\left\{\frac{1}{T^*\mu_1^*} + \left(e^{\sigma^2_{b_1}}-1\right)\right\}\left\{\left(1+\frac{1}{T^*\nu}\right)e^{\sigma^2_{b_2}}-1\right\}}}$$

where $T^* = \frac{\sum_{i=1}^{N} T_i}{N}$. If all subjects have the same number of replicates, T, for all i, then $T^* = T_i = T$.

The total-CC of the bivariate measurements is

$$\rho^{total} = \frac{\mathrm{E}(\mu_{i1t})\,\mathrm{E}(\mu_{i2t'})\,(e^{\rho_b\sigma_{b_1}\sigma_{b_2}}-1)}{\sqrt{\left\{\mathrm{E}(\mu_{i1t}) + \mathrm{E}(\mu_{i1t})^2(e^{\sigma^2_{b_1}}-1)\right\}\left[\mathrm{E}(\mu_{i2t'})^2\left\{(1+\frac{1}{\nu})e^{\sigma^2_{b_2}}-1\right\}\right]}}.$$

Analogously, an overall total-CC is given by replacing $\mathrm{E}(\mu_{i1t})$ and $\mathrm{E}(\mu_{i2t})$ with their marginal expectations over X, $\mu_1^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i1t})\}$ and $\mu_2^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i2t})\}$, and can be expressed as

$$\rho^{total} = \frac{e^{\rho_b\sigma_{b_1}\sigma_{b_2}}-1}{\sqrt{\left\{\frac{1}{\mu_1^*} + \left(e^{\sigma^2_{b_1}}-1\right)\right\}\left\{\left(1+\frac{1}{\nu}\right)e^{\sigma^2_{b_2}}-1\right\}}}.$$

A.2 Derivation of correlations in joint modeling of Poisson-exponential-normal multivariate outcomes

Based on the model proposed in Section 3.4.2, we can compute the marginal means and variances of these outcomes, and the covariances of the conditional means. The marginal means for Y_{i1t}, Y_{i2t} and Y_{i3t} are

$$
\begin{aligned}
\mathrm{E}(Y_{i1t}) &= \mathrm{E}\{\mathrm{E}(Y_{i1t}|b_{i1})\} = e^{\beta_{10}+\beta_{11}x_{1,i1t}+\cdots+\beta_{1p_1}x_{p_1,i1t}+\sigma_{b_1}^2/2} \\
\mathrm{E}(Y_{i2t}) &= \mathrm{E}\{\mathrm{E}(Y_{i2t}|b_{i2})\} = e^{\beta_{20}+\beta_{21}x_{1,i2t}+\cdots+\beta_{2p_2}x_{p_2,i2t}+\sigma_{b_2}^2/2} \\
\mathrm{E}(Y_{i3t}) &= \mathrm{E}\{\mathrm{E}(Y_{i3t}|b_{i3})\} = \beta_{30}+\beta_{31}x_{1,i3t}+\cdots+\beta_{3p_3}x_{p_3,i3t}
\end{aligned}
$$

The marginal variances of Y_{i1t}, Y_{i2t} and Y_{i3t} are

$$
\begin{aligned}
\mathrm{Var}(Y_{i1t}) &= \mathrm{E}\{\mathrm{Var}(Y_{i1t}|b_{i1})\} + \mathrm{Var}\{\mathrm{E}(Y_{i1t}|b_{i1})\} \\
&= e^{\beta_{10}+\beta_{11}x_{1,i1t}+\cdots+\beta_{1p_1}x_{p_1,i1t}+\sigma_{b_1}^2/2} + e^{2(\beta_{10}+\beta_{11}x_{1,i1t}+\cdots+\beta_{1p_1}x_{p_1,i1t})}\left(e^{2\sigma_1^2}-e^{\sigma_1^2}\right) \\
&= \mathrm{E}(\mu_{i1t}) + \mathrm{E}(\mu_{i1t})^2\left(e^{\sigma_{b_1}^2}-1\right) \\
\mathrm{Var}(Y_{i2t}) &= \mathrm{E}\{\mathrm{Var}(Y_{i2t}|b_{i2})\} + \mathrm{Var}\{\mathrm{E}(Y_{i2t}|b_{i2})\} \\
&= e^{2(\beta_{20}+\beta_{21}x_{1,i2t}+\cdots+\beta_{2p_2}x_{p_2,i2t}+\sigma_{b_2}^2)} + e^{2(\beta_{20}+\beta_{21}x_{1,i2t}+\cdots+\beta_{2p_2}x_{p_2,i2t})}\left(e^{2\sigma_2^2}-e^{\sigma_2^2}\right) \\
&= \mathrm{E}(\mu_{i2t})^2\left(2e^{\sigma_{b_2}^2}-1\right) \\
\mathrm{Var}(Y_{i3t}) &= \mathrm{E}\{\mathrm{Var}(Y_{i3t}|b_{i3})\} + \mathrm{Var}\{\mathrm{E}(Y_{i3t}|b_{i3})\} \\
&= \sigma_N^2 + \sigma_{b_3}^2
\end{aligned}
$$

Moreover, the covariances of the conditional means of the first, second and third observer can be developed as

$$
\begin{aligned}
\mathrm{Cov}(\mu_{i1t}, \mu_{i1t'}) &= e^{\beta_{10}+\beta_{11}x_{1,i1t}+\cdots+\beta_{1p_1}x_{p_1,i1t}} e^{\beta_{10}+\beta_{11}x_{1,i1t'}+\cdots+\beta_{1p_1}x_{p_1,i1t'}} \left(e^{2\sigma_{b_1}^2} - e^{\sigma_{b_1}^2}\right) \\
&= \mathrm{E}(\mu_{i1t})\,\mathrm{E}(\mu_{i1t'}) \left(e^{\sigma_{b_1}^2} - 1\right) \\
\mathrm{Cov}(\mu_{i2t}, \mu_{i2t'}) &= e^{\beta_{20}+\beta_{21}x_{1,i2t}+\cdots+\beta_{2p_2}x_{p_2,i2t}} e^{\beta_{20}+\beta_{21}x_{2,i2t'}+\cdots+\beta_{2p_2}x_{p_2,i2t'}+} \left(e^{2\sigma_{b_2}^2} - e^{\sigma_{b_2}^2}\right) \\
&= \mathrm{E}(\mu_{i2t})\,\mathrm{E}(\mu_{i2t'}) \left(e^{\sigma_{b_2}^2} - 1\right) \\
\mathrm{Cov}(\mu_{i3t}, \mu_{i3t'}) &= \mathrm{Cov}\,(b_{i3}, b_{i3}) = \sigma_{b_3}^2
\end{aligned}
$$

Similar to the result in the previous section, we can show that $\mathrm{Cov}(e^{b_{i1}}, b_{i2}) = \rho_{b_{12}}\sigma_{b_1}\sigma_{b_2}e^{\sigma_{b_1}^2|\rho_{b_{12}}|/2}$. The covariances of the conditional means of the different observers are

$$
\begin{aligned}
\mathrm{Cov}(\mu_{i1t}, \mu_{i2t'}) &= e^{\beta_{10}+\beta_{11}x_{1,i1t}+\cdots+\beta_{1p_1}x_{p_1,i1t}+\beta_{20}+\beta_{21}x_{1,i2t}+\cdots+\beta_{2p_2}x_{p_2,i2t}}\,\mathrm{Cov}(e^{b_{i1}}, e^{b_{i2}}) \\
&= \mathrm{E}(\mu_{i1t})\,\mathrm{E}(\mu_{i2t'})\,(e^{\rho_{b_{12}}\sigma_{b_1}\sigma_{b_2}} - 1) \\
\mathrm{Cov}(\mu_{i1t}, \mu_{i3t'}) &= e^{\beta_{10}+\beta_{11}x_{1,i1t}+\cdots+\beta_{1p_1}x_{p_1,i1t}}\,\mathrm{Cov}(e^{b_{i1}}, b_{i3}) \\
&= e^{\beta_{10}+\beta_{11}x_{1,i1t}+\cdots+\beta_{1p_1}x_{p_1,i1t}}\rho_{b_{13}}\sigma_{b_1}\sigma_{b_3}e^{\sigma_{b_1}^2|\rho_{b_{13}}|/2} \\
&= \mathrm{E}(\mu_{i1t})\rho_{b_{13}}\sigma_{b_1}\sigma_{b_3}e^{\sigma_{b_1}^2/2(|\rho_{b_{13}}|-1)} \\
\mathrm{Cov}(\mu_{i2t}, \mu_{i3t'}) &= e^{\beta_{20}+\beta_{21}x_{1,i2t}+\cdots+\beta_{2p_2}x_{p_2,i2t}}\rho_{b_{23}}\sigma_{b_2}\sigma_{b_3}e^{\sigma_{b_2}^2|\rho_{b_{23}}|/2} \\
&= \mathrm{E}(\mu_{i2t})\rho_{b_{23}}\sigma_{b_2}\sigma_{b_3}e^{\sigma_{b_2}^2/2(|\rho_{b_{23}}|-1)}
\end{aligned}
$$

Therefore, the extended intra-CC of the first observer is

$$
\rho_E^{intra,i,1} = \frac{\mathrm{E}(\mu_{i1t})\,\mathrm{E}(\mu_{i1t'})\left(e^{\sigma_{b_1}^2} - 1\right)}{\sqrt{\left\{\mathrm{E}(\mu_{i1t}) + \mathrm{E}(\mu_{i1t})^2\left(e^{\sigma_{b_1}^2} - 1\right)\right\}\left\{\mathrm{E}(\mu_{i1t'}) + \mathrm{E}(\mu_{i1t'})^2\left(e^{\sigma_{b_1}^2} - 1\right)\right\}}}.
$$

The extended intra-CC of the second observer is

$$\rho_E^{intra,i,2} = \frac{\mathrm{E}(\mu_{i2t})\,\mathrm{E}(\mu_{i2t'})\left(e^{\sigma_{b_2}^2}-1\right)}{\sqrt{\left\{\mathrm{E}(\mu_{i2t})^2\left(2e^{\sigma_{b_2}^2}-1\right)\right\}\left\{\mathrm{E}(\mu_{i2t'})^2\left(2e^{\sigma_{b_2}^2}-1\right)\right\}}}.$$

The extended intra-CC of the third observer is

$$\rho_E^{intra,i,3} = \frac{\sigma_{b_3}^2}{\sigma_N^2+\sigma_{b_3}^2}.$$

To obtain overall extended intra-CCs, the expectations of conditional means above are replaced by the marginal expectations over X, $\mu_1^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i1t})\}$, $\mu_2^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i2t})\}$ and $\mu_3^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i3t})\}$. Hence the overall extended intra-CCs are

$$\begin{aligned}
\rho_E^{intra,1} &= \frac{\mu_1^*(e^{\sigma_{b_1}^2}-1)}{1+\mu_1^*(e^{\sigma_{b_1}^2}-1)}\\
\rho_E^{intra,2} &= \frac{e^{\sigma_{b_2}^2}-1}{2e^{\sigma_{b_2}^2}-1}\\
\rho_E^{intra,3} &= \frac{\sigma_{b_3}^2}{\sigma_N^2+\sigma_{b_3}^2}
\end{aligned}$$

Notice that for the third observer, $\sum_{t=1}^{T_i}\mathrm{E}\{\phi_3 h_3(\mu_{i3t})\} = T_i\sigma_N^2$ and $\mathrm{Var}\left(\sum_{t=1}^{T_i}\mu_{i3t}\right) = \mathrm{Var}\left(\sum_{t=1}^{T_i} b_{i3}\right) = T_i^2\sigma_{b_3}^2$.

The extended inter-CC is equal to $\rho_E^{inter} = NI/DI$, where

$$\begin{aligned}
NI =& \sum_{t=1}^{T_i}\sum_{t'=1}^{T_i}\mathrm{E}(\mu_{i1t})\,\mathrm{E}(\mu_{i2t'})\left(e^{\rho_{b_{12}}\sigma_{b_1}\sigma_{b_2}}-1\right) + T_i\sum_{t=1}^{T_i}\mathrm{E}(\mu_{i1t})\rho_{b_{13}}\sigma_{b_1}\sigma_{b_3}e^{\sigma_{b_1}^2/2(|\rho_{b_{13}}|-1)}\\
&+ T_i\sum_{t=1}^{T_i}\mathrm{E}(\mu_{i2t})\rho_{b_{23}}\sigma_{b_2}\sigma_{b_3}e^{\sigma_{b_2}^2/2(|\rho_{b_{23}}|-1)}
\end{aligned}$$

and

$$
\begin{aligned}
DI = & \sqrt{\left[\sum_{t=1}^{T_i} \mathrm{E}(\mu_{i1t}) + \left\{\sum_{t=1}^{T_i} \mathrm{E}(\mu_{i1t})\right\}^2 \left(e^{\sigma_{b_1}^2} - 1\right)\right]} \\
& \times \sqrt{\left[e^{\sigma_{b_2}^2} \sum_{t'=1}^{T_i} \mathrm{E}(\mu_{i2t'})^2 + \left\{\sum_{t'=1}^{T_i} \mathrm{E}(\mu_{i2t'})\right\}^2 \left(e^{\sigma_{b_2}^2} - 1\right)\right]} \\
& + \sqrt{\left[\sum_{t=1}^{T_i} \mathrm{E}(\mu_{i1t}) + \left\{\sum_{t=1}^{T_i} \mathrm{E}(\mu_{i1t})\right\}^2 \left(e^{\sigma_{b_1}^2} - 1\right)\right] \left(T_i\sigma_N^2 + T_i^2\sigma_{b_3}^2\right)} \\
& + \sqrt{\left[e^{\sigma_{b_2}^2} \sum_{t'=1}^{T_i} \mathrm{E}(\mu_{i2t'})^2 + \left\{\sum_{t'=1}^{T_i} \mathrm{E}(\mu_{i2t'})\right\}^2 \left(e^{\sigma_{b_2}^2} - 1\right)\right] \left(T_i\sigma_N^2 + T_i^2\sigma_{b_3}^2\right)}.
\end{aligned}
$$

We replace $\sum_{t=1}^{T_i} \mathrm{E}(\mu_{i1t})$ and $\sum_{t=1}^{T_i} \mathrm{E}(\mu_{i2t})$ with $T^*\mu_1^*$ and $T^*\mu_2^*$. The overall extended inter-CC is defined as $\rho_E^{inter} = NI^*/DI^*$, where

$$
NI^* = \mu_1^*\mu_2^*\left(e^{\rho_{b_{12}}\sigma_{b_1}\sigma_{b_2}} - 1\right) + \mu_1^*\rho_{b_{13}}\sigma_{b_1}\sigma_{b_3}e^{\sigma_{b_1}^2/2(|\rho_{b_{13}}|-1)} + \mu_2^*\rho_{b_{23}}\sigma_{b_2}\sigma_{b_3}e^{\sigma_{b_2}^2/2(|\rho_{b_{23}}|-1)}
$$

and

$$
\begin{aligned}
DI^* = & \sqrt{\left\{\frac{\mu_1^*}{T^*} + \mu_1^{*^2}(e^{\sigma_{b_1}^2} - 1)\right\} \left[\mu_2^{*^2}\left\{\left(1 + \frac{1}{T^*}\right)e^{\sigma_{b_2}^2} - 1\right\}\right]} \\
& + \sqrt{\left\{\frac{\mu_1^*}{T^*} + \mu_1^{*^2}(e^{\sigma_{b_1}^2} - 1)\right\} \left(\frac{\sigma_N^2}{T^*} + \sigma_{b_3}^2\right)} \\
& + \sqrt{\left[\mu_2^{*^2}\left\{\left(1 + \frac{1}{T^*}\right)e^{\sigma_{b_2}^2} - 1\right\}\right] \left(\frac{\sigma_N^2}{T^*} + \sigma_{b_3}^2\right)}.
\end{aligned}
$$

The extended total-CC can be expressed as $\rho_E^{total} = NT/DT$, where

$$\begin{aligned} NT =& \mathrm{E}(\mu_{i1t})\,\mathrm{E}(\mu_{i2t'})\,(e^{\rho_{b_{12}}\sigma_{b_1}\sigma_{b_2}} - 1) + \mathrm{E}(\mu_{i1t})\rho_{b_{13}}\sigma_{b_1}\sigma_{b_3}e^{\sigma_{b_1}^2/2(|\rho_{b_{13}}|-1)} \\ &+ \mathrm{E}(\mu_{i2t})\rho_{b_{23}}\sigma_{b_2}\sigma_{b_3}e^{\sigma_{b_2}^2/2(|\rho_{b_{23}}|-1)} \end{aligned}$$

and

$$\begin{aligned} DT =& \sqrt{\left\{\mathrm{E}(\mu_{i1t}) + \mathrm{E}(\mu_{i1t})^2\left(e^{\sigma_{b_1}^2} - 1\right)\right\}\left\{\mathrm{E}(\mu_{i2t})^2\left(2e^{\sigma_{b_2}^2} - 1\right)\right\}} \\ &+ \sqrt{\left\{\mathrm{E}(\mu_{i1t}) + \mathrm{E}(\mu_{i1t})^2(e^{\sigma_{b_1}^2} - 1)\right\}\left(\sigma_N^2 + \sigma_{b_3}^2\right)} \\ &+ \sqrt{\left\{\mathrm{E}(\mu_{i2t})^2\left(2e^{\sigma_{b_2}^2} - 1\right)\right\}\left(\sigma_N^2 + \sigma_{b_3}^2\right)}. \end{aligned}$$

An overall total-CC is given by replacing $\mathrm{E}(\mu_{i1t})$ and $\mathrm{E}(\mu_{i2t})$ with the marginal expectation over X, $\mu_1^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i1t})\}$ and $\mu_2^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i2t})\}$, and is defined as $\rho_E^{total} = NT^*/DT^*$, where

$$NT^* = \mu_1^*\mu_2^*\,(e^{\rho_{b_{12}}\sigma_{b_1}\sigma_{b_2}} - 1) + \mu_1^*\rho_{b_{13}}\sigma_{b_1}\sigma_{b_3}e^{\sigma_{b_1}^2/2(|\rho_{b_{13}}|-1)} + \mu_2^*\rho_{b_{23}}\sigma_{b_2}\sigma_{b_3}e^{\sigma_{b_2}^2/2(|\rho_{b_{23}}|-1)}$$

and

$$\begin{aligned} DT^* =& \sqrt{\left\{\mu_1^* + \mu_1^{*2}(e^{\sigma_{b_1}^2} - 1)\right\}\left\{\mu_2^{*2}\left(2e^{\sigma_{b_2}^2} - 1\right)\right\}} \\ &+ \sqrt{\left\{\mu_1^* + \mu_1^{*2}(e^{\sigma_{b_1}^2} - 1)\right\}\left(\sigma_N^2 + \sigma_{b_3}^2\right)} \\ &+ \sqrt{\left\{\mu_2^{*2}\left(2e^{\sigma_{b_2}^2} - 1\right)\right\}\left(\sigma_N^2 + \sigma_{b_3}^2\right)}. \end{aligned}$$

A.3 Derivation of correlations in OAI example

Now consider the normal-negative binomial model. The marginal mean and variance for Y_{i1t} are shown in the previous subsections. Given the random effects b_{i2},

Y_{i2t} is assumed to be from negative binomial distribution with mean μ_{i2t} and variance $\mu_{i2t}\left(1+\frac{1}{\delta_N}\right)$, where δ_N is the shape parameter. Then the marginal mean and variance of Y_{i2t} are

$$\begin{aligned}
\mathrm{E}(Y_{i2t}) =& \mathrm{E}\{\mathrm{E}(Y_{i2t}|b_{i2})\} = e^{\beta_{20}+\beta_{21}x_{1,it}+\beta_{22}x_{2,it}+\sigma_{b_2}^2/2} \\
\mathrm{Var}(Y_{i2t}) =& \mathrm{E}\{\mathrm{Var}(Y_{i2t}|b_{i2})\} + \mathrm{Var}\{\mathrm{E}(Y_{i2t}|b_{i2})\} \\
=& e^{\beta_{20}+\beta_{21}x_{1,it}+\beta_{22}x_{2,it}+\sigma_{b_2}^2/2} + e^{2(\beta_{20}+\beta_{21}x_{1,it}+\beta_{22}x_{2,it}+\sigma_{b_2}^2)}/\delta_N \\
&+ e^{2(\beta_{20}+\beta_{21}x_{1,it}+\beta_{22}x_{2,it})}\left(e^{2\sigma_2^2}-e^{\sigma_2^2}\right) \\
=& \mathrm{E}(\mu_{i2t}) + \mathrm{E}(\mu_{i2t})^2\left\{\left(1+\frac{1}{\delta_N}\right)e^{\sigma_2^2}-1\right\}
\end{aligned}$$

Furthermore, notice that $\sum_{t=1}^{T_i}\mathrm{E}\{\phi_2 h_2(\mu_{i2t})\} = \sum_{t=1}^{T_i}\mathrm{E}(\mu_{i2t}) + \sum_{t=1}^{T_i}\frac{1}{\delta_N}\mathrm{E}(\mu_{i2t}^2)$, $\mathrm{Var}\left(\sum_{t=1}^{T_i}\mu_{i2t}\right) = \left\{\sum_{t=1}^{T_i}\mathrm{E}(\mu_{i2t})\right\}^2\left(e^{\sigma_{b_2}^2}-1\right)$, and $\mathrm{E}\left(\mu_{i2t'}^2\right) = e^{2(\beta_{20}+\beta_{21}x_{1,it}+\beta_{22}x_{2,it})}e^{2\sigma_{b_2}^2}$ $= \{\mathrm{E}(\mu_{i2t'})\}^2 e^{\sigma_{b_2}^2}$.

Therefore, the intra-CC of the first observer is

$$\rho^{intra,i,1} = \frac{\sigma_{b_3}^2}{\sigma_N^2+\sigma_{b_3}^2}$$

and the intra-CC of the second observer is

$$\rho^{intra,i,2} = \frac{\mathrm{E}(\mu_{i2t})\,\mathrm{E}(\mu_{i2t'})\left(e^{\sigma_{b_2}^2}-1\right)}{\sqrt{\left[\mathrm{E}(\mu_{i2t})+\mathrm{E}(\mu_{i2t})^2\left\{(1+\frac{1}{\delta_N})e^{\sigma_2^2}-1\right\}\right]\left[\mathrm{E}(\mu_{i2t'})+\mathrm{E}(\mu_{i2t'})^2\left\{(1+\frac{1}{\delta_N})e^{\sigma_2^2}-1\right\}\right]}}.$$

Thus, the overall intra-CC 1 and intra-CC 2 are

$$\rho^{intra,1} = \frac{\sigma_{b_3}^2}{\sigma_N^2+\sigma_{b_3}^2} \quad \text{and} \quad \rho^{intra,2} = \frac{\mu_2^*(e^{\sigma_{b_2}^2}-1)}{1+\mu_2^*\left\{\left(1+\frac{1}{\delta_N}\right)e^{\sigma_{b_2}^2}-1\right\}},$$

respectively, where $\mu_2^* = \mathrm{E}_X\{\mathrm{E}(\mu_{i2t})\}$. The inter-CC and total-CC are

$$\rho^{inter} = \frac{\sum_{t=1}^{T_i}\sum_{t'=1}^{T_i} \mathrm{E}(\mu_{i2t'})\rho_{b_{12}}\sigma_{b_1}\sigma_{b_2}e^{\sigma_{b_2}^2/2(|\rho_{b_{12}}|-1)}}{\sqrt{\left(T_i\sigma_N^2 + T_i^2\sigma_{b_1}^2\right)\left[\sum_{t=1}^{T_i}\mathrm{E}(\mu_{i2t}) + \sum_{t=1}^{T_i}\frac{1}{\delta_N}\left[\mathrm{E}(\mu_{i2t'})\right]^2 e^{\sigma_{b_2}^2} + \left\{\sum_{t=1}^{T_i}\mathrm{E}(\mu_{i2t'})\right\}^2\left(e^{\sigma_{b_2}^2}-1\right)\right]}}$$

and

$$\rho^{total} = \frac{\mathrm{E}(\mu_{i2t'})\rho_{b_{12}}\sigma_{b_1}\sigma_{b_2}e^{\sigma_{b_2}^2/2(|\rho_{b_{12}}|-1)}}{\sqrt{\left(\sigma_N^2 + \sigma_{b_1}^2\right)\left\{\mathrm{E}(\mu_{i2t'}) + \frac{1}{\delta_N}\left\{\mathrm{E}(\mu_{i2t'})\right\}^2 e^{\sigma_{b_2}^2} + \left\{\mathrm{E}(\mu_{i2t'})\right\}^2\left(e^{\sigma_{b_2}^2}-1\right)\right]}}.$$

Thus, overall inter-CC and total-CC are

$$\rho^{inter} = \frac{\rho_{b_{12}}\sigma_{b_1}\sigma_{b_2}e^{\sigma_{b_2}^2/2(|\rho_{b_{12}}|-1)}}{\sqrt{\left(\frac{\sigma_N^2}{T^*} + \sigma_{b_1}^2\right)\left\{\frac{1}{T^*\mu_2^*} + (1 + \frac{1}{T^*\delta_N})e^{\sigma_{b_2}^2} - 1\right\}}}$$

and

$$\rho^{total} = \frac{\rho_{b_{12}}\sigma_{b_1}\sigma_{b_2}e^{\sigma_{b_2}^2/2(|\rho_{b_{12}}|-1)}}{\sqrt{\left(\sigma_N^2 + \sigma_{b_1}^2\right)\left\{\frac{1}{\mu_2^*} + (1 + \frac{1}{\delta_N})e^{\sigma_{b_2}^2} - 1\right\}}}.$$

APPENDIX B

APPROXIMATE UNIFORM SHRINKAGE PRIOR FOR A MULTIVARIATE GENERALIZED LINEAR MIXED MODEL

B.1 Derivation of illustrative examples

Example 1: a bivariate clustered mixed model with random intercept To obtain the approximate uniform shrinkage prior, the GLM weighted matrix is required. Since conditional on the random effects b_{i1}, Y_{i1t} is assumed to follow a Poisson distribution, then $w_{i1t} = [\phi_1 a_1''(\mu_{i1t})\{g_1'(\mu_{i1t})\}^2]^{-1} = \left\{\mu_{i1t}\left(\frac{1}{\mu_{i1t}}\right)^2\right\}^{-1} = \mu_{i1t}$. Similarly, conditional on the random effects b_{i2}, Y_{i2t} is assumed to follow a gamma distribution, then $w_{i2t} = [\phi_2 a_2''(\mu_{i2t})\{g_2'(\mu_{i2t})\}^2]^{-1} = \left\{\frac{{\mu_{i2t}}^2}{\nu}\left(\frac{1}{\mu_{i2t}}\right)^2\right\}^{-1} = \nu$. Therefore, the GLM weight matrix for the i-th subject is $W_i^* = \text{diag}(\mu_{i11}, \cdots, \mu_{i1T_i}, \nu, \cdots, \nu)$. Furthermore, the random effects design matrix for the i-th subject is $Z_i^* = J_i \bigoplus J_i$, where $J_i = (1, \cdots, 1)^T$. Therefore, the approximate uniform shrinkage prior can be shown as

$$
\begin{aligned}
\pi_D(D) &\propto \left| I_2 + \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^T W_i Z_i\right) D \right|^{-3} \\
&= \left| \begin{bmatrix} 1 + \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T_i}\mu_{i1t}\right)\cdot\sigma_{11} & \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T_i}\mu_{i1t}\right)\cdot\sigma_{12} \\ \frac{1}{N}\sum_{i=1}^{N} T_i\nu\cdot\sigma_{12} & 1 + \frac{1}{N}\sum_{i=1}^{N} T_i\nu\cdot\sigma_{22} \end{bmatrix} \right|^{-3} \\
&= \left\{ \left[1 + \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T_i}\mu_{i1t}\right)\cdot\sigma_{11}\right]\left[1 + \frac{1}{N}\sum_{i=1}^{N} T_i\nu\cdot\sigma_{22}\right] \right. \\
&\quad \left. - \left[\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T_i}\mu_{i1t}\right)\cdot\frac{1}{N}\sum_{i=1}^{N} T_i\nu\cdot\sigma_{12}^2\right] \right\}^{-3}.
\end{aligned}
$$

Example 2: a bivariate clustered mixed model with both random intercept and random slope

In this case, the GLM weight matrix and the random effects design matrix for the i-th subject are $W_i^* = \text{diag}(\mu_{i11}, \cdots, \mu_{i1T_i}, \nu, \cdots, \nu)$ and $Z_i^* = G_{i,1} \bigoplus G_{i,2}$, where $G_{i,j} = \begin{bmatrix} 1 & \cdots & 1 \\ z_{ij1} & \cdots & z_{ijT_i} \end{bmatrix}^T$. Thus, the approximate uniform shrinkage prior is shown as

$$\pi_D(D) \propto \left| I_4 + \left(\frac{1}{N} \sum_{i=1}^{N} Z_i^T W_i Z_i \right) D \right|^{-5}$$
$$= \left| \begin{bmatrix} 1 + \frac{1}{N}\sum_{i=1}^N S_1(i)\sigma_{11} & \frac{1}{N}\sum_{i=1}^N S_1(i)\sigma_{12} & \frac{1}{N}\sum_{i=1}^N S_1(i)\sigma_{13} & \frac{1}{N}\sum_{i=1}^N S_1(i)\sigma_{14} \\ \frac{1}{N}\sum_{i=1}^N S_2(i)\sigma_{21} & 1 + \frac{1}{N}\sum_{i=1}^N S_2(i)\sigma_{22} & \frac{1}{N}\sum_{i=1}^N S_2(i)\sigma_{23} & \frac{1}{N}\sum_{i=1}^N S_2(i)\sigma_{24} \\ \frac{1}{N}\sum_{i=1}^N T_i\nu\sigma_{31} & \frac{1}{N}\sum_{i=1}^N T_i\nu\sigma_{32} & 1 + \frac{1}{N}\sum_{i=1}^N T_i\nu\sigma_{33} & \frac{1}{N}\sum_{i=1}^N T_i\nu\sigma_{34} \\ \frac{1}{N}\sum_{i=1}^N T_i z_{i1t}^2\nu\sigma_{41} & \frac{1}{N}\sum_{i=1}^N T_i z_{i1t}^2\nu\sigma_{42} & \frac{1}{N}\sum_{i=1}^N T_i z_{i1t}^2\nu\sigma_{43} & 1 + \frac{1}{N}\sum_{i=1}^N T_i z_{i1t}^2\nu\sigma_{44} \end{bmatrix} \right|^{-5}$$

where $S_1(i) = \sum_{t=1}^{T_i} \mu_{i1t}$ and $S_2(i) = \sum_{t=1}^{T_i} z_{i1t}^2 \mu_{i1t}$.

Example 3: a trivariate clustered mixed model with random intercept

Conditional on the random effects b_{i3}, the measurements Y_{i3t} from normal distribution are also considered in additional to the measurements from Poisson distribution and gamma distribution. Then $w_{i3t} = [\phi_3 a_3''(\mu_{i3t})\{g_3'(\mu_{i3t})\}^2]^{-1} = (\sigma_N^2 \cdot 1^2)^{-1} = \frac{1}{\sigma_N^2}$. Under this circumstance, the GLM weight matrix and the random effects design matrix for the i-th subject are $W_i^* = \text{diag}(\mu_{i11}, \cdots, \mu_{i1T_i}, \nu, \cdots, \nu, 1/\sigma_N^2, \cdots, 1/\sigma_N^2)$ and $Z_i^* = J_i \bigoplus J_i \bigoplus J_i$, where $J_i = (1, \cdots, 1)^T$. Therefore, the approximate uniform shrinkage prior is

$$\pi_D(D) \propto \left| I_3 + \left(\frac{1}{N} \sum_{i=1}^N Z_i^T W_i Z_i \right) D \right|^{-3}$$
$$= \left| \begin{bmatrix} 1 + \frac{1}{N}\sum_{i=1}^N \left(\sum_{t=1}^{T_i} \mu_{i1t}\right) \cdot \sigma_{11} & \frac{1}{N}\sum_{i=1}^N \left(\sum_{t=1}^{T_i} \mu_{i1t}\right) \cdot \sigma_{12} & \frac{1}{N}\sum_{i=1}^N \left(\sum_{t=1}^{T_i} \mu_{i1t}\right) \cdot \sigma_{13} \\ \frac{1}{N}\sum_{i=1}^N T_i\nu \cdot \sigma_{21} & 1 + \frac{1}{N}\sum_{i=1}^N T_i\nu \cdot \sigma_{22} & \frac{1}{N}\sum_{i=1}^N T_i\nu \cdot \sigma_{23} \\ \frac{1}{N}\sum_{i=1}^N T_i \frac{1}{\sigma_N^2} \cdot \sigma_{31} & \frac{1}{N}\sum_{i=1}^N T_i \frac{1}{\sigma_N^2} \cdot \sigma_{32} & 1 + \frac{1}{N}\sum_{i=1}^N T_i \frac{1}{\sigma_N^2} \cdot \sigma_{33} \end{bmatrix} \right|^{-3}.$$

B.2 Derivation of posterior distribution in the OAI example

In the normal-negative binomial model, the posterior distribution of σ_N^2 is

$$
\begin{aligned}
f(\sigma_N^2|b,\beta,D,y) &\propto \left\{\prod_{i=1}^{N}\prod_{t=1}^{T_i} f(y_{i1t}|b,\beta,D)\right\}\pi(\sigma_N^2) \\
&\propto (\sigma_N^2)^{-\frac{\sum_i^N T_i}{2}} \exp\left(-\frac{\sum_{i=1}^N\sum_{t=1}^{T_i}(y_{i1t}-\mu_{i1t})^2}{2\sigma_N^2}\right)\cdot(\sigma_N^2)^{-\alpha_N-1}\exp\left(-\frac{\beta_N}{\sigma_N^2}\right),
\end{aligned}
$$

which implies the posterior distribution of σ_N^2 follows an inverse gamma $\left(\frac{\sum_{i=1}^N T_i}{2}+\alpha_N, \frac{\sum_{i=1}^N\sum_{t=1}^{T_i}(y_{i1t}-\mu_{i1t})^2}{2}+\beta_N\right)$.

Similarly, the posterior distribution of δ_N is

$$
\begin{aligned}
f(\delta_N|b,\beta,D,y) &\propto \left\{\prod_{i=1}^{N}\prod_{t=1}^{T_i} f(y_{i2t}|b,\beta,D)\right\}\pi(\delta_N) \\
&\propto \left\{\prod_{i=1}^{N}\prod_{t=1}^{T_i}\frac{\Gamma(y_{i2t}+\delta_N)}{\Gamma(y_{i2t}+1)\Gamma(\delta_N)}\left(\frac{\delta_N}{\delta_{i2t}+\mu_{i2t}}\right)^{\delta_N}\left(\frac{\mu_{i2t}}{\delta_N+\mu_{i2t}}\right)^{y_{i2t}}\right\}\cdot\delta_N^{-\alpha_D-1}\exp\left(-\frac{\beta_D}{\delta_N}\right)
\end{aligned}
$$

Conditional on the random effects b_{i2}, Y_{i2t} is assumed to follow a negative binomial distribution and the diagonal element of GLM weighted matrix is $w_{i2t} = [\phi_2 a_2''(\mu_{i2t})\{g_2'(\mu_{i2t})\}^2]^{-1} = \left\{\mu_{i2t}\left(1+\frac{\mu_{i2t}}{\delta_N}\right)\left(\frac{1}{\mu_{i2t}}\right)^2\right\}^{-1} = \frac{\delta_N\mu_{i2t}}{\delta_N+\mu_{i2t}}$. Therefore, the approximate uniform shrinkage prior of D in the normal-negative binomial model is

$$
\begin{aligned}
\pi_D(D) &\propto \left|I_2+\left(\frac{1}{N}\sum_{i=1}^N Z_i^T W_i Z_i\right)D\right|^{-3} \\
&= \left|\begin{bmatrix} 1+\frac{1}{N}\sum_{i=1}^N\left(T_i\cdot\frac{1}{\sigma_N^2}\right)\cdot\sigma_{11} & \frac{1}{N}\sum_{i=1}^N\left(T_i\cdot\frac{1}{\sigma_N^2}\right)\cdot\sigma_{12} \\ \frac{1}{N}\sum_{i=1}^N\left(\sum_{t=1}^{T_i}\frac{\delta_N\mu_{i2t}}{\delta_N+\mu_{i2t}}\right)\cdot\sigma_{12} & 1+\frac{1}{N}\sum_{i=1}^N\left(\sum_{t=1}^{T_i}\frac{\delta_N\mu_{i2t}}{\delta_N+\mu_{i2t}}\right)\cdot\sigma_{22}\end{bmatrix}\right|^{-3} \\
&= \left[\left\{1+\frac{\sigma_{11}}{N}\sum_{i=1}^N\frac{T_i}{\sigma_N^2}\right\}\left\{1+\frac{\sigma_{22}}{N}\sum_{i=1}^N\left(\sum_{t=1}^{T_i}\frac{\delta_N\mu_{i2t}}{\delta_N+\mu_{i2t}}\right)\right\}\right. \\
&\quad \left.-\left\{\left(\frac{\sigma_{12}}{N}\right)^2\sum_{i=1}^N\frac{T_i}{\sigma_N^2}\sum_{i=1}^N\left(\sum_{t=1}^{T_i}\frac{\delta_N\mu_{i2t}}{\delta_N+\mu_{i2t}}\right)\right\}\right]^{-3}.
\end{aligned}
$$