A MARGINALIZED MULTILEVEL MODEL
FOR BIVARIATE LONGITUDINAL BINARY DATA

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

GÜL İNAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
STATISTICS

AUGUST 2014

Approval of the thesis:

A MARGINALIZED MULTILEVEL MODEL FOR BIVARIATE LONGITUDINAL BINARY DATA

submitted by **GÜL İNAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Statistics Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. İnci Batmaz
Head of Department, **Statistics**

Assoc. Prof. Dr. Özlem İlk Dağ
Supervisor, **Statistics Department, METU**

Examining Committee Members:

Prof. Dr. Gülser Köksal
Industrial Engineering Department, METU

Assoc. Prof. Dr. Özlem İlk Dağ
Statistics Department, METU

Prof. Dr. Ergun Karaağaoğlu
Biostatistics Department, Hacettepe University

Assist. Prof. Dr. Zeynep Kalaylıoğlu
Statistics Department, METU

Assist. Prof. Dr. Ceylan Yozgatlıgil
Statistics Department, METU

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: GÜL İNAN

Signature :

ABSTRACT

A MARGINALIZED MULTILEVEL MODEL
FOR BIVARIATE LONGITUDINAL BINARY DATA

İnan, Gül
Ph.D., Department of Statistics
Supervisor : Assoc. Prof. Dr. Özlem İlk Dağ

August 2014, 122 pages

This thesis study considers analysis of bivariate longitudinal binary data. We propose a model based on marginalized multilevel model framework. The proposed model consists of two levels such that the first level associates the marginal mean of responses with covariates through a logistic regression model and the second level includes subject/time specific random intercepts within a probit regression model. The covariance matrix of multiple correlated time-specific random intercepts for each subject is assumed to represent the within-subject association. The subject-specific random effects covariance matrix is further decomposed into its dependence and variance components through modified Cholesky decomposition method to handle possible computational and statistical problems that may be associated with its high-dimensionality. Then the unconstrained version of resulting parameters are modelled in terms of covariates with low-dimensional regression parameters, which provides better explanations related to dependence and variance parameters and a reduction in the number of parameters to be estimated in random effects covariance matrix to avoid possible identifiability problems. Marginal correlations between responses of subjects and within the responses of a subject are derived through a Taylor series-based approximation. Data cloning computational algorithm is used to compute the maximum likelihood estimates of the parameters in the proposed model and their standard errors. The validity of the proposed model is assessed through a Monte Carlo simulation study under different scenarios, and results are observed to be at acceptable level.

Lastly, the proposed model is illustrated through Mother's Stress and Children's Morbidity study data, where both population-averaged and subject-specific interpretations are drawn through Emprical Bayes estimation of random effects.

Keywords: Bivariate binary responses, multilevel models, multiple correlated random effects, covariance matrix decomposition, data cloning

ÖZ

İKİ DEĞİŞKENLİ İKİ ELEMANLI UZANLAMASINA VERİ İÇİN MARJİNALLEŞTİRİLMİŞ ÇOK SEVİYELİ BİR MODEL

İnan, Gül
Doktora, İstatistik Bölümü
Tez Yöneticisi : Doç. Dr. Özlem İlk Dağ

Ağustos 2014 , 122 sayfa

Bu tez çalışması iki elemanlı iki değişkenli uzunlamasına verinin analizini ele almaktadır. Marjinalleştirilmiş çok seviyeli model çerçevesinde, bir model önermekteyiz. Önerilen model iki seviyeden oluşmaktadır; öyle ki, ilk seviye bir lojistik regresyon modeli aracılığıyla ortalama cevap değişkenlerini açıklayıcı değişkenlerle ilişkilendirmektedir; ikinci seviye bir probit regresyon modeli içerisine kişi/zamana bağlı rassal kesişim terimleri dahil etmektedir. Her kişi için, birden çok korelasyonlu ve zamana bağlı rassal kesişim terimlerinin kovaryans matrisinin, kişi içi ilişkiyi temsil ettiği varsayılmıştır. Rassal etkiler kovaryans matrisinin yüksek boyutlu olması ile ilişkili olabilecek hesapsal ve istatistiksel sorunlarla baş edebilmek için bu matris, değiştirilmiş Cholesky ayrıştırma yöntemi ile bağımlılık ve varyans parçalarına ayrıştırılmıştır. Daha sonra, bağımlılık ve varyans parametrelerine dair daha iyi açıklamalar getirmesi ve olası tanınabilirlik sorunlarını önlemek için rassal etkiler kovaryans matrisinde hesaplanması gereken parametre sayısının azalmasını sağlamasından dolayı, ortaya çıkan parametrelerin kısıtsız halleri düşük boyutta açıklayıcı değişkenlerle modellenmiştir. Kişiler arası cevaplar ve kişinin kendi cevapları arasındaki marjinal korelasyonlar, Taylor serisi tabanlı bir yaklaşımla elde edilmiştir. Önerilen modeldeki parametrelerin, en çok olabilirlik yöntemi parametre ve standart hata tahmin edicilerini hesaplamak için veri klonlaması hesapsal algoritması kullanılmıştır. Önerilen modelin geçerliliği farklı senaryolar altında bir Monte Carlo benzetim çalışması aracılığıyla incelenmiştir, ve çıkan sonuçların kabul edilebilir seviyede olduğu gözlenmiştir. Son olarak, önerilen model, Anne'nin Stresi ve Çoçukların Hastalığı verisi ile örneklendirilmiştir; rassal etkilerin Ampirik Bayes yöntemi ile tahmin edilmesi sayesinde, hem populasyon ortalaması hem de kişi bazlı yorumlar çıkarılmıştır.

Anahtar Kelimeler: İki elemanlı iki değişkenli cevaplar, çok seviyeli modeller, birden çok korelasyonlu rassal etkiler, kovaryans matrisi ayrıştırması, veri klonlaması

To my beautiful country

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my thesis supervisor Assoc. Prof. Dr. Özlem İlk Dağ, who has supported me throughout my thesis with extreme patience. Her continuous guidance, support, and feedbacks have turned this study to an immeasurable learning experience for me. I am deeply indebted to her for her suggestions, insights and encouragements in all the process of research and writing of this thesis. This thesis would not have been possible without her persistent help and effort.

I would like to acknowledge my appreciation to my thesis committee members Prof. Dr. Gülser Köksal and Assist. Prof. Dr. Zeynep Kalaylıoğlu for guiding my thesis for more than two years. I can truly say that their valuable comments and suggestions have improved the quality of my thesis. I would also like to acknowledge my appreciation to Prof. Dr. Ergun Karaağaoğlu and Assist. Prof. Dr. Ceylan Yozgatlıgil for spending their valuable time in reviewing my thesis.

I would like to thank to Assoc. Prof. Dr. Alexander de Leon, who suggested me to use data cloning computational algorithm in my thesis during his visit to our department. It was a great opportunity for me to take advantage of his academic knowledge and experience.

I am truly indebted to Prof. Dr. İnci Batmaz, who offered me an office to continue my studies in our department. I could not have enough motivation to complete my thesis without her kind offer.

I also owe special thanks to The Scientific and Technological Research Council of Turkey (TÜBİTAK) for funding my graduate studies.

My sincere thanks go to İlkcan Keleş, who provided me an ultra efficient environment to use high performance computing facility of Department of Computer Engineering.

I am grateful to Bülent Alptekin, who offered me a lot of friendly help and provided me an efficient computer environment in our department computer lab to complete my thesis.

I would also like to thank to Dr. Peter Solymos and Assoc. Prof. Dr. Mahmoud Torabi, who kindly replied my questions via e-mail. While Dr. Peter Solymos helped me about usage of dclone package in R, Assoc. Prof. Dr. Mahmoud Torabi helped me to figure out prediction of random effects through dclone package in R.

I am very thankful to Serdar Neslihanoğlu for sending me the papers, which I could not have an access to, in a few minutes although he was in England.

I would like to thank to Emin Mustaçoğlu, who was interested in the progress of my thesis, offered me a lot of help, and encouraged me every time since I started to work as a research assistant in our department.

Interdisciplinary nature of Statistics introduced many friends to me. In this sense, I would like to convey my heartfelt thanks to my friends Tuğba Terzi Kalyoncuoğlu, Umut Okkan, and Farid Rajabli for their gentle friendship. Although they have different specializations, I have been always happy to spend time with them and to talk about science and Statistics.

I want to thank to all my friends in our department for their kind friendship and support over the past years. My very special thanks go to Tuğba Erdem, Sibel Balci, Ceyda Yazıcı, and Didem Egemen. Thesis research and writing was both a painful and enjoyable journey. They were always there to listen and to give me advice. I greatly appreciate their gentle friendship, everlasting patience, and loving kindness.

Last, but not least, I want to thank to my mummy for her unconditional support and encouragement to pursue my academic career, to my brother Başar for his words of encouragement, and to my sister Gonca for cheering me up and stood by me through the good and bad times. I could not finish my thesis if they were not with me. I owe my all achievements to them.

TABLE OF CONTENTS

CHAPTERS

APPENDICES

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

AGQ	Adaptive Gaussian quadrature
AIC	Akaike information criteria
AUROC	Area under the ROC curve
ASE	Average of standard errors
CI	Confidence Interval
CP	Coverage Probability
DC	Data Cloning
DIC	Deviance information criterion
EB	Empirical Bayes
F-S	Fisher-Scoring
GARP	Generalized autoregressive parameter
GQ	Gaussian quadrature
GEE	Generalized estimating equations
GLM	General linear model
GLMM	Generalized linear mixed model
IFT	Implicit function theorem
IV	Innovation variance
INLA	Integrated nested Laplace approximation
LMM	Linear mixed model
LRT	Likelihood ratio test
MAR	Missing at random
MCMC	Markov Chain Monte Carlo
MC	Monte Carlo
MLE	Maximum likelihood estimation
MMM	Marginalized multilevel models
MMM	Multivariate marginal models
MREM	Marginalized random effects model
MSCM	Mother’s Stress and Children’s Morbidity

MSE	Mean squared error
MTM	Marginalized transition model
MTREM	Marginalized transition random effects model
NLMM	Nonlinear mixed model
N-R	Newton-Raphson
OR	Odds-ratio
PNMTREM	Probit normal marginalized transition random effects model
PCC	Percentage of correct classification
SD	Standard deviation
TN	True negative
TP	True positive
QMC	Quasi-Monte Carlo
UMM	Univariate marginal models

CHAPTER 1

INTRODUCTION

Longitudinal studies arise from repeated measurements on a given response for the same subjects over time. Many longitudinal studies can be considered from different fields such as clinical trials, epidemiology, behavioral sciences, econometrics and so on. Multiple responses are sometimes collected repeatedly over time to better understand the underlying phenomenon. As an example, one can consider Mother's Stress and Children's Morbidity (MSCM) study (Diggle et al., 2002), where a total of 167 mothers and their preschool children were enrolled and mothers' stress status (0 = absence, 1 = presence) and their children's illness status (0 = absence, 1 = presence) were collected along with some demographic and domestic variables for 28 days. MSCM study aimed to investigate whether mother's stress status and child's illness status were jointly associated with the demographic and domestic variables such as mother's education level, employment and marriage status, child's gender and race, and the household size.

Building joint regression models for longitudinal data with multiple responses may help answer multiple questions simultaneously (Gueorguieva, 2013) such as the effect of mother's employment status on mother's stress status and child's illness status, as in the case of MSCM study, it may provide information on how the association between multiple outcomes evolves over time, and it may estimate the parameters in the model with an improved efficiency due to information exchange between responses.

Analysis of multivariate longitudinal data (e.g., longitudinal data with multiple responses) requires special attention since repeated measurements of a subject over time for a given response are expected to be correlated and yield within-subject association, while multiple responses of a subject measured at a given time point are also expected to be correlated and yield multivariate response association. As a consequence, although the effects of covariates on longitudinal response means may be of primary interest in multivariate longitudinal data, within-subject association and multivariate response association should be taken into account at the model building stage for accurate inferences (Asar, 2012).

Models developed for the analysis of longitudinal data can be categorized into 2 classes: single-level and multilevel models (Asar, 2012). While the single-level models can be subcategorized as marginal and conditional models, models that have marginalized mean structure fall into the latter category.

Marginal models directly specify a regression model to assess the effects of covariates on the longitudinal mean response. The parameters related to longitudinal association are treated as if they are nuisance parameters and are specified through a working covariance matrix. The regression parameters in marginal models have population-averaged interpretations like in a traditional general linear model (GLM). They are consistent and asymptotically normally distributed, even if covariance structure for longitudinal association is misspecified (Liang and Zeger, 1986). One considerable drawback of marginal models is their lack of likelihood-based inferences since a joint distribution for longitudinal responses cannot be specified (Griswold, 2005). In the statistical literature of marginal models for multivariate longitudinal binary data, Shelton et al. (2004) analysed multivariate longitudinal binary data through marginal models for univariate longitudinal binary via SAS. Asar and Ilk (2014) implemented their approach within R and proposed `mmm` library. While regression coefficients were response-specific in this marginal model, another marginal model, which included regression coefficients shared by responses, was also proposed by Asar and Ilk (2013a) and was presented in `mmm2` library in R. In these studies, while the regression parameters were estimated via generalized estimating equations (GEEs), the association parameters were estimated via method of moments.

Conditional models generally model the longitudinal mean response conditioned on either random effects or previous history of responses (Griswold, 2005). Random effects or previous history of responses are included into the model to account for the dependence across the responses of a subject (Griswold, 2005). The regression parameters in conditional models have interpretation conditional on either random effects or past responses. For that reason, the estimates of the regression parameters are sensitive to the longitudinal association specifications (Heagerty and Zeger, 2000), as can be seen in simulation study example of Lee et al. (2012). Beyond the interpretation of random effect terms and previous history of responses in the model, they also enable to construct a joint distribution for longitudinal binary responses and, in turn, enable likelihood-based inference capabilities. While a joint distribution for longitudinal binary responses can be constructed through conditioning on distribution imposed on the random effect terms in random effects models (Zhang et al., 2011), a joint distribution for those can be specified through conditioning on the past responses in the transition models. In the statistical literature of conditional models for multivariate longitudinal binary data, Chan et al. (1997) introduced a bivariate autoregressive model with random effects and used maximum likelihood estimation (MLE) method for inference. Zeng and Cook (2007) proposed a joint transitional model for multivariate longitudinal binary data using GEE2 (Zhao and Prentice, 1990; Liang et al., 1992).

Marginalized multilevel models (MMMs), introduced in the seminal papers of Heagerty (1999), Heagerty and Zeger (2000), and Heagerty (2002) combine the strengths of marginal and conditional models. In this sense, they build different regression models for the marginal mean, i.e., marginal mean model, which formulates the association between the longitudinal responses and covariates, and for longitudinal association(s), i.e., conditional mean model, which restructures the mean of longitudinal responses conditional on either random effects and/or a Markov structure. Hence, while MMMs take the interpretation and robustness of marginal regression parameters from marginal models, they take likelihood-based inference capabilities and flexible longitudinal association specifications from conditional models (Griswold and Zeger, 2004; Griswold, 2005). In the statistical literature of MMMs for multivariate longitudinal binary data, Ilk and Daniels (2007) proposed marginalized transition random effects models (MTREMs), which consisted of three different levels, where each level was a different logistic regression model to specify marginal mean, within-subject association, and multivariate response association, respectively. They used

Bayesian inference for the parameter estimation of their model. Lee et al. (2009) proposed a two-level marginalized model, which was based on MLE for inference. While the first level of their model defined the marginal mean, the last level of their model represented within-subject and multivariate response associations by the covariance matrix of correlated subject/time/response specific random intercepts. Asar (2012) extended the model of Ilk and Daniels (2007) by replacing the logit link functions in their model with probit link functions and offered a MLE based inference. This new model, which was named as probit normal marginalized transition random effects models (PNMTREMs), was presented in an R package `pnmtrem` (Asar and Ilk, 2013b). Iddi and Molenberghs (2012) developed a two-level marginalized model by relaxing the assumption of common link function for all levels, which was based on MLE for inference. In their model, while the first level defined the marginal mean through a logistic regression model, the last level of their model represented association between the multiple outcomes through the covariance matrix of correlated subject/response specific random effects within a probit regression model.

Finally, two comprehensive reviews on the analysis of multivariate longitudinal data and one comprehensive review on the joint modelling of multiple responses were given by Bandyopadhyay et al. (2011) and Verbeke et al. (2012), and Rizopoulos and Lesaffre (2014), respectively.

Objective of This Thesis

In this thesis study, we propose a model, which is based on marginalized multilevel model framework, to analyse bivariate longitudinal binary data. In the proposed model, the first level associates marginal mean of responses with covariates through a logistic regression model to keep odds-ratio interpretation of the marginal regression parameters. The second level includes subject/time specific random intercepts, i.e., time-specific random intercepts for each subject, within a probit regression model to take the random-variation within the responses of the same subject and as well as between the responses of different subjects into account. For each subject, time-specific random intercepts are further assumed to be correlated, which results in a marginalized multilevel model with multiple correlated random effects. Multivariate normal distribution is imposed on the vector of random intercepts for each subject, so that the covariance matrix, which may be high-dimensional, is assumed to represent the within-subject association.

To handle possible computational and statistical problems that may be associated with high-dimensionality of the random effects covariance matrix, we further decompose the random effects covariance matrix into its dependence and variance components through modified Cholesky decomposition method. Then we model the unconstrained version of resulting parameters (e.g., natural logarithm of variance parameters) in terms of covariates with low-dimensional regression parameters. Hence, we mainly get better explanations related to the dependence and variance parameters and a reduction in the number of parameters to be estimated in random effects covariance matrix to avoid from possible identifiability problems. In this sense, to the best of our knowledge, this is the first time the modified Cholesky decomposition method is used within multivariate longitudinal binary data and as well as MMMs framework to deal with computational and statistical problems associated with covariance matrix of multiple random effects.

Like any random effects model for binary responses, the parameter estimation for the proposed model is based on the numerical optimization of marginal likelihood with respect to unknown parameters, which includes the computation of high-dimensional covariance matrix inversion and differentiation, following the numerical evaluation of integration of conditional distribution of responses over the distribution of high-dimensional random effects.

In this thesis study, to avoid numerical evaluation of the high-dimensional integral, computation of high-dimensional covariance matrix inversion and differentiation of the covariance matrix with respect to unknown parameters, and numerical optimization of the marginal likelihood, we use data cloning (DC) computational algorithm (Lele et al., 2007; Lele, 2010) to compute the maximum likelihood estimates (MLEs) of the parameters in the proposed model and their standard errors. In this sense, to the best of our knowledge, this is the first time that DC computational algorithm is used for multivariate longitudinal binary data as well as within MMMs framework for parameter estimation.

Upon building the structure of the model, we carry out a Monte Carlo (MC) simulation study under different scenarios to assess the performance of the model parameters through some performance measures such as average, bias, Monte Carlo standard error (SE), average of the model based standard errors (ASE), mean squared error (MSE), and coverage probability (CP).

We fit the proposed model on Mother's Stress and Children's Morbidity (MSCM) study data to illustrate the idea. We draw population-averaged interpretations as well as subject-specific interpretations through Emprical Bayes (EB) estimates of random effects.

The subsequent chapters of this thesis study are organized as follows: Chapter 2 briefly reviews marginalized multilevel models (MMMs) for longitudinal binary data, maximum likelihood estimation (MLE) method for MMMs with random effects and as well as data cloning (DC) computational algorithm, and reparametrization of random effects covariance matrix in longitudinal data analysis. Chapter 3 introduces the proposed model, describes definitions, assumptions and required derivations related to the model, talks about the estimation of the model. Chapter 4 assesses the validity of the proposed model under different simulation scenarios and then illustrates the proposed model through Mother's Stress and Children's Morbidity (MSCM) study data. Chapter 5 summarizes the main findings explored during the thesis study, presents some recommendations and guidance for models including random effects, and give some extensions of the proposed model as future works.

CHAPTER 2

HISTORICAL BACKGROUND

In this chapter, we briefly review marginalized multilevel models (MMMs) for longitudinal binary data in Section 2.1, maximum likelihood estimation (MLE) method for MMMs with random effects in Section 2.2, and reparametrization of random effects covariance matrix in longitudinal data analysis in Section 2.3, respectively.

2.1 Marginalized Multilevel Models for Longitudinal Binary Data

Marginalized multilevel models for longitudinal binary data, introduced in the seminal papers of Heagerty (1999), Heagerty and Zeger (2000), and Heagerty (2002) build separate regression models for marginal and conditional means of the longitudinal responses to combine the strengths of marginal and conditional models. Specifically, the model for the marginal mean builds the relationship between the longitudinal responses and covariates, and the model for conditional mean parametrizes the longitudinal association(s) through either random effects (marginalized random effects models (MREMs) (e.g., the model in Heagerty (1999)) and/or a Markov structure (marginalized transition models (MTMs)) (e.g., the model in Heagerty (2002)). As a consequence of this model building structure, MMMs take the population-averaged interpretation and robustness of regression parameters from marginal models, whereas they take likelihood-based inference capabilities and flexible association specifications from conditional models (Griswold and Zeger, 2004; Griswold, 2005; Griswold et al., 2013), which are also summarized in Table 2.1.

The formulation of MREMs, which is the scope of this thesis study, for univariate

Table 2.1: Comparison of strengths and weakness of marginal, conditional and marginalized multilevel models.

	Marginal Models	Conditional Models	Marginalized Multilevel Models
Population-averaged interpretation of regression parameters	✓	✗	✓
Robustness of regression parameters to longitudinal association specifications	✓	✗	✓
Likelihood-based inference	✗	✓	✓
Flexibility in specifications for longitudinal association	✗	✓	✓

longitudinal data, define a general linear model (GLM) for the marginal mean of responses, a nonlinear mixed model (NLMM) for the conditional mean of responses, and further specify a statistical distribution for the random effects, which is assumed to be the source of within-subject association, as follows:

$$\begin{aligned} \text{Marginal Mean Model:} \quad & g(\mu_{it}^{M}) = g(E(Y_{it}|\mathbf{X}_{it})) = \mathbf{X}_{it}\boldsymbol{\beta}^{M} \\ \text{Conditional Mean Model:} \quad & g(\mu_{it}^{C}) = g(E(Y_{it}|\mathbf{X}_{it}, \mathbf{b}_i)) = \Delta_{it}(\mathbf{X}_{it}) + b_{it}, \end{aligned} \tag{2.1}$$

where Y_{it} represents the measurement of ith subject at time t ($i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, n_i$), *g(.)* is a non-linear link function for both marginal and conditional mean models, $\boldsymbol{\beta}^{M}$ is a $P \times 1$ vector of marginal regression parameters associated with $\mathbf{X}_{it}$, Δ_{it} is a term that connects the conditional mean to the marginal mean of the model as shown in equation 2.2, $\mathbf{b}_i = (b_{i1}, \ldots, b_{it}, \ldots, b_{in_i})$ is a $n_i \times 1$ vector of subject-specific and time-varying intercepts and $\mathbf{b}_i \sim F_{\mathbf{b}}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a $n_i \times n_i$ covariance matrix and may be a function of time. In practice, a normally distributed time-invariant random intercept b_i with mean 0 and variance σ^2 is assumed. Given $\mathbf{b}_i$, Y_{it}'s are assumed to independently come from a distribution, which is a member of exponential family with conditional mean μ_{it}^{C}. This is known as conditional independence assumption and it also allows to construct a joint marginal distribution for Y_{it}'s for ith subject (Zhang et al., 2011), which will be discussed in Section 2.2.

In probability theory, it is known that any conditional expectation can be written in terms of marginal expectation such that $E_{b_{it}}(\mu_{it}^C) = E_{b_{it}}(E(Y_{it}|\mathbf{X}_{it}, b_{it})) = E(Y_{it}|\mathbf{X}_{it})$ $= \mu_{it}^M$. This implies that the integration of conditional mean μ_{it}^C over the distribution of random effects results in marginal mean μ_{it}^M given as follows:

$$\begin{aligned} \mu_{it}^M &= \int \mu_{it}^C f(b_{it}) db_{it} \\ \Rightarrow E(Y_{it}|\mathbf{X}_{it}) &= \int E(Y_{it}|\mathbf{X}_{it}, b_{it}) f(b_{it}) db_{it} \\ \Rightarrow g^{-1}(\mathbf{X}_{it}\boldsymbol{\beta}^M) &= \int g^{-1}(\Delta_{it}(\mathbf{X}_{it}) + b_{it}) f(b_{it}) db_{it}, \end{aligned} \tag{2.2}$$

where $g^{-1}(.)$ is the inverse of the non-linear link function, $f(b_{it})$ is a univariate distribution with 0 mean and variance $Var(b_{it})$. The name "Marginalized" naturally originates from equation 2.2 since the conditional mean is transformed into the marginal mean. Furthermore, the convolution equation 2.2 links the second level to the first level of the model and provides a solution for Δ_{it}. Hence, it is obvious that Δ_{it} in equation 2.1 plays a bridge role between the two levels and is a non-linear function of $\mathbf{X}_{it}$, $\boldsymbol{\beta}^M$, and $Var(b_{it})$.

For the binary responses, equation 2.2 implies that the integration of conditional probability $Pr(Y_{it} = 1|\mathbf{X}_{it}, b_{it})$ over the distribution of random effects results in marginal probability $Pr(Y_{it} = 1|\mathbf{X}_{it})$ given as follows:

$$\begin{aligned} \mu_{it}^M &= \int \mu_{it}^C f(b_{it}) db_{it} \\ \Rightarrow E(Y_{it}|\mathbf{X}_{it}) &= \int E(Y_{it}|\mathbf{X}_{it}, b_{it}) f(b_{it}) db_{it} \\ \Rightarrow Pr(Y_{it} = 1|\mathbf{X}_{it}) &= \int Pr(Y_{it} = 1|\mathbf{X}_{it}, b_{it}) f(b_{it}) db_{it}. \end{aligned} \tag{2.3}$$

The marginalized multilevel model in equation 2.1 can able to adapt different response types with appropriate choices for the link function $g(.)$ (e.g., logit or probit link functions for binary responses, log link function for non-gaussian continuous or count responses). Furthermore, different choices for μ_{it}^C, $\mathbf{b}_i$, and $F_{\mathbf{b}(.)}$ in equation 2.1 result in different models even for the same response type.

Griswold and Zeger (2004), Griswold (2005), and Griswold et al. (2013) provided a variety of MMMs for univariate correlated data, where extension to longitudinal data is straightforward. For binary responses, these are: a logistic-logistic-normal model, a probit-probit-normal, a logistic-probit-normal model, and a logistic-logistic-bridge model. The three words within the name of models refer to the link functions used in the first and the second level of the models and the distribution assumed for the random effects, respectively. The first two models are examples for the specification of different $g(.)$'s for binary responses, where $g(.)$ is a logit link function in the first model and a probit link function in the second model. The third model relaxes the assumption of common link function for marginal and conditional mean models and offers different link functions for each level, namely, a logit link for the first level and a probit link for the second level of the model. The latter model is a variation of the first model, where a bridge distribution is assumed for $F_{\mathbf{b}}(.)$.

As summarized in Table 2.2, only the models with logit link function in the first level provide an odds-ratio interpretation for marginal regression parameters. However, for the probit-probit-normal model, an approximate relationship exists between logit and probit regression parameter estimates such as $\beta_{logit} = c * \beta_{probit}$, where c is called as JKB constant (Griswold, 2005) and $c = (\frac{15}{16})\frac{\pi}{\sqrt{3}}$. On the other hand, only the models, whose link function in the second level is coherent with the distribution of the random effects, provide a closed-form solution for Δ_{it}. In this framework, while the logit link gets on well with the bridge distribution, the probit link is a good collaborator with normal distribution. For example, for the logistic-probit-normal model, equation 2.3 provides $\Delta_{it} = \mathbf{X}_{it}\boldsymbol{\beta}^{M}\sqrt{(1 + Var(b_{it})}$, which is apparently a function of $\mathbf{X}_{it}$, $\boldsymbol{\beta}^{M}$, and $Var(b_{it})$. Specifically, the probit link-normal distribution collaboration finds application in many cases other than MMMs due to its computational easiness (Caffo and Griswold, 2006). Examples include Kalaylioglu and Ozturk (2013) and Teixeira-Pinto and Harezlak (2013). In other cases, where a closed-form solution does not exist for Δ_{it}, estimation of it is possible through a numerical optimization method such as Newton-Raphson (N-R) along with a numerical integration method such as Gaussian quadrature (GQ) (Heagerty, 1999), which adds extra computational burden to the parameter estimation of the model. On the other hand, extension to multiple correlated random effects is possible only with the models including normally distributed random effects. For example, although Wang and Louis (2004) extended the logistic-logistic-bridge model through using covariate-dependent random effects, generalization of this model for multiple correlated random effects requires a different task such as modelling through copulas like in Parzen et al. (2011), since there may not be a multivariate bridge distribution. So far, all aforementioned models assumed a time-invariant random intercept b_i, but Caffo et al. (2007) assumed a vector of time-varying random intercepts, where $\mathbf{b}_i$ is a vector of $n_i \times 1$ and a mixture of normal distributions for $F_{\mathbf{b}}(.)$. Swihart et al. (2013) extended these models to copulas to create marginal joint distributions for clustered binary responses. Other extensions of these models for the univariate longitudinal binary case can be found in Miglioretti and Heagerty (2004) and Schildcrout and Heagerty (2007).

Table 2.2: Comparison of marginalized multilevel models for longitudinal binary data.

	Logistic-Logistic-Normal-Model-	Probit-Probit-Normal-Model-	Logistic-Probit-Normal-Model-	Logistic-Logistic-Bridge-Model-
Odds-ratio interpretation of regression parameters	✓	✗	✓	✓
A closed-form solution for Δ_{it}	✗	✓	✓	✓
Extension to multiple correlated random effects	✓	✓	✓	✗

The marginalized multilevel models for multivariate longitudinal data consist of different extensions of equation 2.1 to jointly modelling of multiple responses Y_{itj}'s, where Y_{itj} now represents tth measurement of ith subject for jth dimension, ($t = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, J$). In this sense, within the framework of multivariate longitudinal binary data, Ilk and Daniels (2007) combined MTMs and MREMs and proposed marginalized transition random effects models (MTREMs), which consisted of three different logistic regression models. In this model, the first level introduced a logistic regression model for marginal mean with regression coefficients shared by the responses, the second level built up a transition model for serial dependence between responses of a subject at different time points, and the third level included normally distributed uncorrelated subject/time specific random intercepts for the dependence between the multiple responses of a subject for a given time point, which were further scaled by a response-specific coefficient. Hence, while serial dependence between the responses was directly modelled in the second level, multivariate response dependence at a given time point was represented by random effects. In this sense, Ilk and Daniels (2007) showed that multivariate dependence between the responses at a given time point existed through a Taylor series-based approximation to the function, which measured marginal correlation between the responses. Δ_{itj}'s in the second and third levels of the model were estimated through N-R, while a 20-points GQ was used to evaluate the one-dimensional integral in the convolution equation of Δ_{itj} in the third level. The parameter estimation for MTREM was based on Bayesian inference through Fortran. An application of MTREM to financial statements of Turkey can be found in Akinc (2008) and a simulation study on the robustness of mean response model of MTREM to longitudinal association models can be seen in Yalcinoz (2008).

Asar (2012) extended MTREM and proposed probit normal marginalized transition random effects models (PNMTREMs), which consisted of three different probit regression models. Since Asar (2012) used the probit link function in the first level, he used the JKB constant to get odds-ratio interpretation of the regression parameters in the first level. As a novel approach, Asar (2012) estimated Δ_{itj} in the second level of this 3-level model by an iterative version of implicit function theorem (IFT) (Krantz and Parks, 2003; Stewart, 2008), whereas Δ_{itj} in the third level of the model had a closed-form solution due to the computational advantage of probit link-normal distribution collaboration. Details regarding to when and how to use iterative version of IFT were given in Asar (2012, pg 90-92). The parameter estimation of PNMTREM was based on MLE method including the numerical optimization Fisher-Scoring (F-S) algorithm as well as a numerical integration with a 20-points GQ. Furthermore, this new model was implemented in `pnmtrem` package in R (Asar and Ilk, 2013b).

Lee et al. (2009) proposed a MREM, which consisted of two different logistic regression models. They defined serial dependence and multivariate response dependence by introducing normally distributed and correlated subject/time/response specific random intercepts, where the covariance matrix of random effects was constructed by the help of Kronecker product. Unlike Ilk and Daniels (2007) and Asar (2012), the regression coefficients in the first level of the model were response-specific and the covariance matrix of random effects in Lee et al. (2009) represented the source of two longitudinal associations: within-subject association and multivariate response association. The parameter estimation in the model of Lee et al. (2009) was based on MLE with F-S algorithm as well as a 1000-points Quasi-Monte Carlo (QMC) integration for high-dimensional integration, which also included a N-R step with one-dimensional integration with 20-points GQ for the estimation of Δ_{itj} in the second level. The implementation of the model was done through R along with Fortran due to the high-dimensional integration.

Lastly, Iddi and Molenberghs (2012) proposed a logistic-probit-normal model for bivariate longitudinal binary data, which consisted of a logistic regression model in the first level with response-specific regression coefficients and a probit regression model with subject/response specific random effects (random intercept as well as random slopes) in the second level, where the covariance structure of random effects was similar to Gueorguieva (2001). The model of Iddi and Molenberghs (2012) did not take within-subject association into account through neither random effects, nor other terms. Δ_{itj} in the second level had a closed-form solution due to the probit link-normal distribution collaboration. They implemented their model with built-in numerical optimizers (e.g., N-R algorithm through 10-points GQ) within Proc NLMIXED in SAS.

Although the first level of the model of Lee et al. (2009) and that of Iddi and Molenberghs (2012) were the same, the difference between two models lied in the specification for the second level of their models. While Lee et al. (2009) used a logit link and included correlated subject/time/response specific random intercepts, which take serial dependence and multivariate response dependence into account, Iddi and Molenberghs (2012) used a probit link and included correlated subject/response specific random effects, which took only multivariate response dependence into account. Naturally, the dimension of covariance matrix of random effects in the model of Lee et al. (2009) was larger compared to that in the model of Iddi and Molenberghs (2012). However, Lee et al. (2009) kept the number of parameters to be estimated related to the covariance matrix at minimum by assuming a restrictive structure to avoid high-dimensional covariance matrix estimation and possible identifiability problems. At this point, it is worthy to say that in aforementioned models, only the model of Ilk and Daniels (2007) and, naturally, that of Asar (2012) used the observed information in the data to estimate within-subject association parameters. The multivariate response dependence parameters in both models and all dependence parameters in Lee et al. (2009) and Iddi and Molenberghs (2012) originated from the random effects, which were unobserved. Finally, Table 2.3 summarizes features of the aforementioned four models for multivariate longitudinal binary data.

Based on the literature review, in this thesis study, in Chapter 3, we propose a new logistic-probit-normal marginalized multilevel model with multiple correlated random effects since logistic-probit-normal marginalized multilevel models (e.g., the models with logit link in the first level and probit link in the second level) provide odds-ratio interpretation for the marginal mean parameters through logit link and computational advantages due to the probit link-normal distribution collaboration. The covariance matrix of multiple correlated random effects is further decomposed through modified Cholesky decomposition method to avoid potential statistical and computational problems that may be associated with its high-dimensionality and to provide better explanations related to the elements of it. The estimation of unknown parameters in the proposed model is based on a frequentist approach. For that reason, Section 2.2 continues with the principle idea behind MLE for random effects models for longitudinal binary data.

Table 2.3: Characteristics of marginalized multilevel models for multivariate longitudinal binary data.

	Ilk and Daniels (2007)	Lee et al. (2009)	Asar (2012) PNMTREM	Iddi and Molenberghs (2012b)
Type of regression parameters	Shared by response	Response specific	Shared by response	Response specific
Odds-ratio interpretation of regression parameters	✓	✓	✗	✓
A closed-form solution for Δ_{itj}	✗	✗	✓	✓
Type of random effects	Uncorrelated subject/time specific random intercepts	Correlated subject/time response/specific random intercepts	Uncorrelated subject/time specific random intercepts	Correlated subject/response specific random slopes
Dimension of random effects	One-dimensional	High-dimensional	One-dimensional	High-dimensional
Source of within-subject association	Through history of responses	Through random effects	Through history of responses	✗
Source of multivariate-response association	Through random effects	Through random effects	Through random effects	Through random effects
Estimation method	Bayesian Inference	MLE	MLE	MLE
Implementation	Fortran	R with Fortran	R package `pnmtrem`	Sas Proc NLMIXED

2.2 Maximum Likelihood Estimation for Random Effects Models for Longitudinal Binary Data

In this thesis study, upon building a logistic-probit-normal marginalized multilevel model with multiple correlated random effects, we estimate unknown parameters in the proposed model, which includes marginal regression parameters and random effects covariance matrix parameters, through maximum likelihood estimation (MLE) method. It is well-known that parameter estimation relies on defining marginal likelihood of data, which requires a formulation for joint marginal distribution of longitudinal responses. In a general framework of joint modelling in longitudinal studies, the construction of a joint marginal distribution for multiple responses can be possible through i) assuming a multivariate distribution, ii) including past responses into the model and then factorizing the joint distribution of responses conditional on the distribution of past responses in time order, or iii) including random effects into the model, factorizing the joint distribution of responses and random effects as conditional distribution of responses given the random effects and marginal distribution of random effects and then integrating out the distribution of random effects (Rizopoulos and Lesaffre, 2014). While longitudinal associations are imposed on the marginal covariance matrix of multiple responses in i), those are represented through the past responses in ii), and random effects in iii), respectively. Specifically, for ii) and iii), while the past responses/random effects account for the dependence between the responses of a subject in univariate longitudinal binary data, they are further responsible for the dependence between different responses of a subject in multivariate longitudinal data.

Within the marginalized multilevel models (MMMs) for clustered/longitudinal binary data, Swihart et al. (2013) provided a novel example for i). Since constructing a multivariate distribution for correlated binary responses is not available in the literature, they used copulas to construct a joint marginal distribution for correlated binary responses. While Heagerty (2002) is the first example for ii), Ilk and Daniels (2007) and Asar (2012) proposed a model, which is a combination of ii) and iii). The remaining models, which are given in Section 2.1, are examples for iii).

As mentioned in Section 2.1, MMMs with random effects, which is the focus of this thesis study, take the likelihood-based inference capabilities from generalized linear mixed models (GLMMs). For that reason, they inherit the same characteristics, which GLMMs experience during the parameter estimation.

Like any GLMMs, in MMMs for multivariate longitudinal binary data with random effects, the marginal likelihood of data $L(\boldsymbol{\theta}|\mathbf{y})$ involves the integration of conditional probabilities of responses over the distribution of random effects, as follows:

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^{N} \int \prod_{t=1}^{n_i} \prod_{j=1}^{J} f(y_{itj}|\mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i, \tag{2.4}$$

where y_{itj} represents the measurement of ith subject at time t for jth response type, ($i = 1, 2, \ldots, N, t = 1, 2, \ldots, n_i$, and $j = 1, 2, \ldots, J$), $\mathbf{b}_i$ is a $n_i \times 1$ vector of random effects, which comes from a multivariate normal distribution with $\mathbf{0}$ mean and $\boldsymbol{\Sigma}$ covariance matrix, $f(y_{itj}|\mathbf{b}_i)$ is the conditional distribution of y_{itj}, which is the last level of the marginalized multilevel model. The dimension of the integral in equation 2.4 is equal to the dimension of $\mathbf{b}_i$ (Karl et al., 2014), and $\boldsymbol{\theta}$ includes all the unknown parameters (e.g., marginal regression parameters and random effects covariance matrix parameters) in the model. The underlying assumption in equation 2.4 is that given $\mathbf{b}_i$, y_{itj} is assumed to be independent of $y_{it'j'}$, $\forall\, t \neq t'$ and $j \neq j'$, which is well-known as conditional independence assumption.

Note that under the conditional independence assumption in random effects models, it is straightforward to handle different type of responses (Rizopoulos and Lesaffre, 2014). Recent statistical literature has provided examples for joint modelling of longitudinal continuous and binary outcomes (Iddi and Molenberghs, 2012), that of longitudinal time-to-event data (Efendi et al., 2013), that of any type longitudinal response (e.g., continuous, binary or poisson) with time-to-event data (Njagi et al., 2013), that of longitudinal continuous, ordinal outcomes, and time-to-event data (Baghfalaki et al., 2014), that of non-Gaussian bounded covariates with a primary outcome (Din et al., 2014), that of censored longitudinal covariates with a time-to-event endpoint (Pike and Weissfel, 2014), and that of non-ignorable missing longitudinal covariates with a cross-sectional binary outcome (Zhang et al., 2014).

There are two mathematical problems associated with equation 2.4. First of all, in case of the non-conjugacy of $f(\mathbf{b}_i)$ with $f(y_{itj}|\mathbf{b}_i)$, the integral in equation 2.4 is not tractable and, in turn, it does not provide a closed-form solution for $L(\boldsymbol{\theta}|\mathbf{y})$. For that reason, the integral in equation 2.4 should be either evaluated numerically or approximated. There exist several numerical evaluation methods for integration in random effects models for binary data such as Gaussian quadrature (GQ), adaptive Gaussian quadrature (AGQ), Monte Carlo (MC) or Quasi-Monte Carlo (QMC), which all have different i) degree of accuracy, ii) complexity in implementation, and iii) computational time (Capanu et al., 2013). Within MMMs for multivariate longitudinal binary data, when $f(\mathbf{b}_i)$ is a univariate normal distribution, the one-dimensional integral in equation 2.4 can be evaluated numerically through GQ or AGQ, like Asar (2012) and Iddi and Molenberghs (2012). However, when $f(\mathbf{b}_i)$ is a multivariate normal distribution, the high-dimensional integral in equation 2.4 can be evaluated numerically through MC or QMC integration, like Lee et al. (2009). In this sense, it is worthy to say that in case of high-dimensional $\mathbf{b}_i$, the numerical evaluation of the integral may be further complicated. Note also that one can increase the accuracy of estimates by increasing the number of points to evaluate the integral. However, whichever method s/he uses, one should keep in mind that the increase in the number of points may result in a computational time problem. This computational problem dramatically increases in models with high-dimensional $\mathbf{b}_i$ (Pan and Thompson, 2007). For that reason, one should be careful when making a choice between accuracy and computational time. A comparative analysis between several numerical integration methods for random effects models for univariate clustered/longitudinal binary data can be done by critically analysing the studies of and Zhang et al. (2011), Capanu et al. (2013), Dey and Lim (2013), and Kim et al. (2013). However, among these studies only Kim et al. (2013) focused on random effects models with multiple correlated random effects.

The second problem associated with equation 2.4 is that due to the non-linear link function (e.g., logit or probit) used in $f(y_{itj}|\mathbf{b}_i)$ to associate conditional probabilities of y_{itj}'s with covariates on the continuous scale, the optimization of the marginal likelihood $L(\boldsymbol{\theta}|\mathbf{y})$ (or, naturally, the optimization of $logL(\boldsymbol{\theta}|\mathbf{y})$) does not provide a closed-form solution for $\hat{\boldsymbol{\theta}}$, which is MLE of $\boldsymbol{\theta}$, even $L(\boldsymbol{\theta}|\mathbf{y})$ is approximated or numerically evaluated. For that reason, $logL(\boldsymbol{\theta}|\mathbf{y})$ should be either approximated or optimized numerically through iterative algorithms such as Newton-Raphson (N-R) or Fisher-Scoring (F-S), in which the former method requires the computation of second-order derivatives of $logL(\boldsymbol{\theta}|\mathbf{y})$ with respect to $\boldsymbol{\theta}$, i.e., Hessian matrix for $logL(\boldsymbol{\theta}|\mathbf{y})$, the latter method does not. Similarly, in case of high-dimensional $\mathbf{b}_i$, the computation of second-order partial derivatives of $logL(\boldsymbol{\theta}|\mathbf{y})$ for N-R algorithm becomes cumbersome, since the optimization of $logL(\boldsymbol{\theta}|\mathbf{y})$ includes taking high-dimensional covariance matrix inversion and high-dimensional covariance matrix differentiation with respect to $\boldsymbol{\theta}$, which may also result in a computational burden. Within MMMs for multivariate longitudinal binary data, while Iddi and Molenberghs (2012) used N-R for parameter estimation, Lee et al. (2009) and Asar (2012) preferred F-S, where the iterative algorithm is illustrated below:

$$\boldsymbol{\theta}^{(s+1)} = \boldsymbol{\theta}^{(s)} + I(\boldsymbol{\theta}^{(s)})^{-1}\frac{\partial logL(\boldsymbol{\theta}^{(s)}|\mathbf{y}_i)}{\partial\boldsymbol{\theta}^{(s)}}, \tag{2.5}$$

where s is iteration number,

$$\begin{aligned}\frac{\partial logL(\boldsymbol{\theta}|\mathbf{y})}{\partial\boldsymbol{\theta}} &= \sum_{i=1}^{N}\frac{\partial logL(\boldsymbol{\theta}|\mathbf{y}_i)}{\partial\boldsymbol{\theta}} \\ &= \sum_{i=1}^{N} L(\boldsymbol{\theta}|\mathbf{y}_i)^{-1}\int\frac{\partial L(\boldsymbol{\theta}|\mathbf{y}_i,\mathbf{b}_i)}{\partial\boldsymbol{\theta}}f(\mathbf{b}_i)d\mathbf{b}_i\end{aligned} \tag{2.6}$$

with

$$\begin{aligned}L(\boldsymbol{\theta}|\mathbf{y}_i) &= \int\prod_{t=1}^{n_i}\prod_{j=1}^{J}f(y_{itj}|\mathbf{b}_i)f(\mathbf{b}_i)d\mathbf{b}_i \quad \text{and} \\ L(\boldsymbol{\theta}|\mathbf{y}_i,\mathbf{b}_i) &= \prod_{t=1}^{n_i}\prod_{j=1}^{J}f(y_{itj}|\mathbf{b}_i)f(\mathbf{b}_i)d\mathbf{b}_i,\end{aligned} \tag{2.7}$$

and information matrix is

$$I(\boldsymbol{\theta}) = \sum_{i=1}^{N}\left(\frac{\partial logL(\boldsymbol{\theta}|\mathbf{y}_i)}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial logL(\boldsymbol{\theta}|\mathbf{y}_i)}{\partial\boldsymbol{\theta}}\right)^{\top}. \tag{2.8}$$

For example, equation 2.7 for the model of Asar (2012) required an integration over the distribution of the one-dimensional subject/time specific random effect b_{it}, which was numerically evaluated as follows:

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{y}_i) &= \int \prod_{t=1}^{n_i} \prod_{j=1}^{J} f(y_{itj}|b_{it}) f(b_{it}) db_{it} \\ &\approx \sum_{q=1}^{20} \prod_{t=1}^{n_i} \prod_{j=1}^{J} w_q f(y_{itj}|\sqrt{2} z_q), \end{aligned} \tag{2.9}$$

where z_q and w_q ($q = 1, 2, \ldots, 20$) are GQ points and corresponding weights, respectively, and are available in `fastGHQuad` package in R (Blocker, 2011). The integral in equation 2.6 can be evaluated in a similar way.

On the other hand, equation 2.7 for the model of Lee et al. (2009) required the integration over the distribution of the high-dimensional subject/time specific random effects $\mathbf{b}_i = (b_{i1}, \ldots, b_{it}, \ldots, b_{in_i})$, which was numerically evaluated as follows:

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{y}_i) &= \int \prod_{t=1}^{n_i} \prod_{j=1}^{J} f(y_{itj}|\mathbf{b}_i) f(\mathbf{b}_i) \mathbf{b}_i \\ &\approx \frac{\sum_{m=1}^{1000} \prod_{t=1}^{n_i} \prod_{j=1}^{J} f(y_{itj}|b_{itj}^{(m)})}{M}, \end{aligned} \tag{2.10}$$

where $b_{itj}^{(m)}$ is mth value of QMC point. QMC points for standard normal distribution are available in `rnorm.sobol` function in `fOptions` library in R (Wuertz, 2013). For non-standard normal distributions, QMC points can be obtained through Cholesky transformations of QMC points in `rnorm.sobol` function. The integral in equation 2.6 can be evaluated in a similar fashion.

To avoid estimation problems for the parameters, which have restriction in parameter space such as variance components of random effects, while Lee et al. (2009) used the Cholesky root elements of the random effects covariance matrix $\boldsymbol{\Sigma}$, Asar (2012) used the logarithm of the variance of b_{it}.

It is also worthy to say that for models with high-dimensional b_i, since the dimension of $\boldsymbol{\Sigma}$ also increases, this will result in an inflation in the number of parameters to be estimated in $\boldsymbol{\Sigma}$, which may also cause discrepancy in parameter estimates (Kim et al., 2013) and even identifiability problems in the data.

Although the estimation of the model parameters $\boldsymbol{\theta}$ in MMMs have a major priority, MMMs with random effects include two additional estimation problems. These are i) the estimation of Δ_{itj} terms, which connect the levels to each other, along with the estimation algorithm for $\boldsymbol{\theta}$ and ii) estimation of b_i, which is unobservable, based on the estimated value of $\boldsymbol{\theta}$.

As mentioned in Section 2.1, unless the link function in the level including Δ_{itj} is coherent with the distribution of the random effects, a closed-form expression cannot be provided for Δ_{itj}. In this sense, as summarized in Table 2.3, Δ_{itj}'s in the models of Ilk and Daniels (2007) and Lee et al. (2009) did not have closed-form solutions and they used N-R algorithm along with 20-points GQ to estimate Δ_{itj}'s. For illustration, Δ_{itj}'s in the model of Lee et al. (2009) can be obtained by the following N-R algorithm:

$$\Delta_{itj}^{l+1} = \Delta_{itj}^{l} - \left(\frac{\partial h(\Delta_{itj}^{l})}{\partial \Delta_{itj}^{l}} \right)^{-1} h(\Delta_{itj}^{l}), \tag{2.11}$$

where l is iteration number and

$$h(\Delta_{itj}) = \int Pr(Y_{it} = 1|\mathbf{X}_{it}, b_{it}) f(b_{it}) db_{it} - Pr(Y_{it} = 1|\mathbf{X}_{it}). \tag{2.12}$$

The integral in equation 2.12 can be evaluated by GQ. Upon convergence of the N-R algorithm in equation 2.11, an estimate for Δ_{itj} can be obtained. It is obvious that equation 2.11 adds extra computational burden to equation 2.5. In fact, Lee et al. (2009) used Fortran 77 subroutines (.dll files) to cope with this computational problem.

Asar (2012) further estimated Empirical Bayes (EB) estimates of the random effects to compute individual conditional probabilities. Following Heagerty (1999), Asar (2012) firstly derived the posterior distribution of random effects and then solved posterior score equations with respect to random effects through N-R algorithm, which is known as finding the mode of the posterior distribution. Detailed information related to EB estimates of random effects is also available in Hedeker and Gibbons (2006) and Paule et al. (2011).

The implementation of the models with only one random effect (e.g., a random intercept term), which includes only one variance component to be estimated, is possible through either author-written codes within a well-known statistical software, like Ilk and Daniels (2007) and Asar (2012) or built-in functions such as PROC NLMIXED in SAS, like Iddi and Molenberghs (2012). The main advantage of built-in functions over author-written codes is that the model developers do not need to have a good knowledge of numerical analysis (e.g., numerical evaluation of integrals or optimization of the marginal likelihood). They do not even need to supply initial values for the parameters of the random effects covariance matrix to start the estimation algorithm, which are very hard to guess. However, for the implementation of the models with multiple correlated random effects, the built-in functions may not be sufficient to implement the model due to the computational burden (e.g., numerical evaluation of the high-dimensional integral, taking the inverse of high-dimensional random effects covariance matrix) (Iddi and Molenberghs, 2012). This may lead model developers to write their own functions to implement their model with multiple correlated random effects within a standard statistical software such as Lee et al. (2009). Lastly, Table 2.4 summarizes features related to the maximum likelihood estimation of the aforementioned models for multivariate longitudinal binary data.

At this point, it is worthy to say that due to the computational problems (e.g., high-dimensional matrix inversion and differentiation) in random effects models mentioned above, only a random intercept term was assumed in most of the models in the literature like Gueorguieva (2001) and Iddi and Molenberghs (2012). On the other hand, in the models with multiple correlated random effects, which is very few in number, either a very simple covariance structure was assumed like Lee et al. (2009) or matrix decomposition methods such as modified Cholesky decomposition was used for Σ like Lee et al. (2012) to deal with high-dimensional random effects covariance matrix.

So far, the reason for unable to add as many random effects as possible to the model has been shown to be due to computational complexity. However, in multiple correlated random effects models, inflation in the number of parameters to be estimated in Σ may result in identifiability problems for some of the parameters. In this sense, little attention is devoted in the literature to the evaluation and comparison of the model complexity with the information stored in the data. In this context, we use data cloning (DC) computational algorithm (Lele et al., 2007; Lele et al., 2010) to compute MLEs of the parameters and their standard errors for the proposed marginalized multilevel model with multiple random effects in Chapter 3.

Table 2.4: Features related to the maximum likelihood estimation of marginalized multilevel models for multivariate longitudinal binary data.

	Lee et al. (2009)	Asar (2012)	Iddi and Molenberghs (2012b)
Estimation method	MLE	MLE	MLE
Implementation	R with Fortran	R package `pnmtrem`	Sas Proc NLMIXED
Dimension of random effects	High-dimensional	One-dimensional	One-dimensional
Numerical evaluation type	QMC with 1000 points	GQ with 20 points	GQ with 10 points
Numerical optimization algorithm for marginal likelihood	N-R	F-S	F-S
Emprical Bayes estimation of random effects	✗	✓	✗

Data Cloning Approach for Parameter Estimation

Data cloning (DC), which was proposed by Lele et al. (2007) and Lele et al. (2010), is a computational algorithm to compute the maximum likelihood estimates (MLEs) of the parameters and their standard errors for complex models and to check the identifiability of these parameters for the given data. To compute MLEs, as the name suggests, DC algorithm clones the data multiple times and then uses the Bayesian formulation to obtain posterior distribution of the parameters through Markov Chain Monte Carlo (MCMC) computational algorithm (Lele et al., 2007; Lele et al., 2010).

Like Bayesian inference, DC avoids the evaluation of the high-dimensional integration and the numerical optimization of the likelihood function, and hence avoids the computational burden in models with high-dimensional random effects. However, unlike the Bayesian methodology, the choice of prior distribution has no effect on the MLEs of the parameters.

The DC approach can be illustrated through the following imaginary scenario: consider a statistical experiment, which is simultaneously and independently repeated by K times, and all K experiments yield the same result by happenstance. Then cloned data and likelihood for cloned data, which is based on K independent experiments, can be represented by $\mathbf{Y}^{(K)} = \mathbf{y}^{(K)} = (\mathbf{y}, \ldots, \mathbf{y})$ and $L(\boldsymbol{\theta}|\mathbf{y}^{(K)}) = L(\boldsymbol{\theta})^{(K)}$, respectively. Hence, the posterior distribution of $\boldsymbol{\theta}$ given $\mathbf{y}^{(K)}$ can be given as follows:

$$\begin{aligned} \prod\left(\boldsymbol{\theta}|\mathbf{y}^{(K)}\right) &= \frac{1}{C(\mathbf{y}^{(K)})} L(\boldsymbol{\theta})^{(K)} \prod(\boldsymbol{\theta}) \\ &\propto L(\boldsymbol{\theta})^{(K)} \prod(\boldsymbol{\theta}), \end{aligned} \tag{2.13}$$

where $\prod\left(\boldsymbol{\theta}|\mathbf{y}^{(K)}\right)$ is the posterior distribution for $\boldsymbol{\theta}$, $\prod(\boldsymbol{\theta})$ is the prior distribution for $\boldsymbol{\theta}$, and $C(\mathbf{y}^{(K)}) = \int L(\boldsymbol{\theta})^{(K)} \prod(\boldsymbol{\theta})\, d\boldsymbol{\theta}$ is the normalizing constant. The maximum of the cloned likelihood $L(\boldsymbol{\theta})^{(K)}$ is exactly the maximum of $L(\boldsymbol{\theta}|\mathbf{y})$ and naturally it is $\hat{\boldsymbol{\theta}}$, the MLE for $\boldsymbol{\theta}$, based on the equation 2.4. Furthermore, Fisher information matrix based on the cloned likelihood $I(\hat{\boldsymbol{\theta}})^{(K)}$ is K times of that based on original likelihood $I(\hat{\boldsymbol{\theta}})$ since the cloned data contains K times more information than the original data (Torabi, 2012). Furthermore, Lele et al. (2007) and Lele et al. (2010) showed that, for sufficiently large K, $\prod\left(\boldsymbol{\theta}|\mathbf{y}^{(K)}\right)$ converges to a multivariate Gaussian distribution with mean $\hat{\boldsymbol{\theta}}$ and covariance matrix $\frac{I(\hat{\boldsymbol{\theta}})^{-1}}{K}$. Hence, $\hat{\boldsymbol{\theta}}$ is the mean of samples drawn from the posterior distribution in equation 2.13 and the asymptotic covariance matrix for $\hat{\boldsymbol{\theta}}$ is K times the sample covariance matrix of samples drawn from the posterior distribution in equation 2.13.

Determining the number of clones, namely, K is possible through plotting the standardized largest eigenvalue of the posterior covariance matrix versus K and comparing it with K^{-1}. If the standardized largest eigenvalue and K^{-1} converge to zero at the same rate, that K can be considered as the optimum value for K. The same plot can be used to check identifiability of the parameters in the model as well. Alternatively, the plot of posterior variance of each parameter against K also shows the identifiability of that parameter. In fact, a plot does not converging to zero as K increases implies an identifiability problem for that parameter. In this sense, DC allows statistical model developers to study the identifiability of the parameters in their complex models through diagnostic measures as well as diagnostic plots, which is an advantage over other estimation or computational algorithms. Hence, DC enables model developers to build reasonable complex models since it alerts when the complexity of the model exceeds information stored in the data (Lele et al., 2010). This issue is also important in the case of models with high-dimensional random effects since apart from adding extra computational complexity, unconsciously adding too many random effects to the model may result in unidentifiable parameters and the developer may not be aware of it. In short, DC algorithm helps the model developer to balance the model complexity with information available in the data.

Although DC is a good alternative computational algorithm to draw statistical inference for models including random effects, it has some limitations. First of all, as Ponciano et al. (2009) and Baghishani and Mohammadzadeh (2011) stated, DC provides only Wald-type confidence intervals, which yield inaccurate results in small samples and are generally less than nominal coverage rate. In this sense, Ponciano et al. (2009) suggested using profile likelihood based confidence intervals, which are more likely to be at nominal coverage rate, and improving statistical accuracy of estimators in small sized samples via bootstrap methods, which requires further attention. The second limitation associated with DC is that it cannot provide maximized value of the marginal likelihood, which is required to construct likelihood ratio tests (LRTs) and, in turn, to do model selection. Although Ponciano et al. (2009) proposed a DC based computational algorithm to calculate the ratio of likelihoods for each pair of models, this algorithm did not provide any specific Akaike information criterion (AIC) value for each model separately. Overcoming this limitation of DC still needs further research. Third, although the selection of K is not associated with sample size N, K should be large enough in small N to obtain reliable estimates for standard errors. However, increase in K results in computational burden, which may lose the appeal of the method.

Data cloning computational algorithm is available in `dclone` package in R (Sólymos, 2010). Any model developer can implement his/her model through either `jags.fit` or `dc.fit` functions and can check identifiability of parameters in the model through other functions available. Since drawing samples from the posterior distribution through MCMC is based on JAGS software (Plummer, 2003) for both `jags.fit` and `dc.fit` functions, model statement part requires knowledge of a syntax, which is very similar to WinBUGS. For more computationally time consuming models, DC through parallel computing is also available via `jags.parfit` or `dc.parfit` functions within the same R package. Furthermore, to the best of our knowledge, a `stan.parfit` function is an on-going work by Dr. Peter Sólymos from University of Alberta to implement DC through Stan (Stan Development Team, 2014). Please visit `http://dcr.r-forge.r-project.org/extras/stan.fit.R`. On the other hand, Baghishani et al. (2012) utilized DC with an integrated nested Laplace approximation (INLA) as a hybrid version of DC. Further information on DC algorithm can be found in Lele et al. (2010), Sólymos (2010), Torabi (2012), Torabi (2013), and Withanage (2013).

The literature review revealed that DC computational algorithm has not been either adressed in multivariate longitudinal binary data or used to estimate the parameters of a marginalized multilevel model. In this sense, to the best of our knowledge, this is the first time that DC computational algorithm is used for multivariate longitudinal binary data as well as within MMMs framework for parameter estimation.

2.3 Reparametrization of Random Effects Covariance Matrix

Section 2.2 revealed that including normally distributed random effects into a model plays an important role in constructing a joint marginal distribution for longitudinal binary responses as well as capturing the longitudinal association within measurements of a subject for a given response and between the measurements of different responses of a subject. Depending on the model specification, random effects may be subject/time specific (Ilk and Daniels, 2007; Asar, 2012), subject/response specific (Iddi and Molenberghs, 2012) or subject/time/response specific (Lee et al., 2009), which may be further assumed to be correlated or not.

Inclusion of multiple correlated random effects into a model requires defining a structure for covariance matrix for random effects, which characterizes the normal distribution, and results in a certain number of parameters to be estimated related to this covariance matrix. However, the positive definiteness constraint imposed on the covariance matrix and high-dimensionality problem of the covariance matrix are two major problems encountered while imposing a structure for the covariance matrix and estimating the parameters in the covariance matrix (Pourahmadi, 1999). For example, although Lee et al. (2009) assumed a $2n \times 2n$ $(n_i = n, \forall i)$ random effects covariance matrix $\mathbf{\Sigma}$ for correlated subject/time/response specific random effects, where the dimension of the matrix increased as the number of time points in the longitudinal data increased, to avoid inflation in the number of parameters in high-dimensional covariance matrix $\mathbf{\Sigma}$, they wrote $\mathbf{\Sigma}$ as a Kronecker product of $\mathbf{\Sigma}_1$, which is a $n \times n$ within-subject association matrix, and $\mathbf{\Sigma}_2$, which is a 2×2 multivariate association matrix. They further assumed $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ had a structure of an AR(1) and unstructured (UN), respectively. Hence, they decreased the number of parameters in $\mathbf{\Sigma}$ from $\frac{(2n\times(2n+1))}{2}$ to 4, which is free from n. On the other hand, in the models, which assumes correlated subject/response specific random effects following Gueorguieva (2001), the dimension of $\mathbf{\Sigma}$ depends on the number of responses $\times$ the number of random intercepts and slopes. However, to reduce the number of covariance matrix components, a general strategy is to assume only response-specific random intercepts (ignoring random slopes) in the model. In this sense, Iddi and Molenberghs (2012) analysed bivariate longitudinal binary data with both assuming response-specific random intercepts and a random intercept shared by the responses, where the former case included 2 additional parameters compared to the later case. Similarly, Rodrigues-Motta et al. (2013) analysed multivariate correlated count data with six responses by assuming response-specific random intercepts. But, they decreased the number of components to be estimated in 6×6 random effects covariance matrix $\mathbf{\Sigma}$ from 21 to 4 through assuming a common correlation parameter across the responses and three different variance terms for six responses.

In longitudinal data analysis, a more common approach to manage with the positive definiteness constraint assumption of the covariance matrix and its high-dimensionality problem is to decompose the covariance matrix into its variance and the dependence components through decomposition methods and then to model unconstrained versions of the resulting parameters in terms of covariates with low-dimensional regression parameters. This approach allows i) flexibility for the structure of the covariance matrix, ii) better explanations related to the elements of covariance matrix, iii) the covariance matrix to differ by subjects, and iv) a reduction in the number of parameters to be estimated, which is an advantage for the models that cannot be implemented or are hardly implemented due to the high-dimensionality of the covariance matrix. As Section 2.2 also revealed, modelling covariance matrix of random effects parsimoniously is important in terms of both estimation and as well as balancing the model complexity to avoid parameter identifiability problems.

There are two well-known covariance matrix decomposition methods in longitudinal data, which are variance-correlation and modified Cholesky decomposition methods (Pourahmadi, 2011). The former method decomposes covariance matrix as $\mathbf{\Sigma} = \mathbf{DRD}$, where $\mathbf{D}$ is a diagonal matrix with standard deviations on the diagonals and $\mathbf{R}$ is the positive-definite correlation matrix with 1's on the diagonals and correlations on the off-diagonals. While natural logarithm of diagonal elements of $\mathbf{D}$ are unconstrained, the constraint on the off-diagonal elements of $\mathbf{R}$, which is lying within the interval [-1,1], can be removed by using Fisher's Z-transformation. Then the unconstrained parameters in $\mathbf{D}$ and $\mathbf{R}$ can be modelled through some measured covariates. As an example within marginalized multilevel models framework, Lee et al. (2013) proposed a marginalized random effects model (MREM) for bivariate longitudinal ordinal data and assumed that the covariance matrix of random effects can be written as a Kronecker product of the correlation matrix of responses of a subject over time and the covariance matrix of different responses of a subject at a given time, as in Lee et al. (2009). Following Daniels and Pourahmadi (2009), they further reparametrized the correlation matrix in terms of lag-1 correlations and partial autocorrelations and then they modelled Fisher's Z-transformation of these parameters in terms of covariates. On the other hand, this statistically motivated decomposition method is also attractive for Bayesian estimation of covariance matrix when Wishart distribution is not an adequate prior distribution for covariance matrix (Barnard et al., 2000).

The modified Cholesky decomposition method decomposes covariance matrix $\mathbf{\Sigma} = \mathbf{T}^{-1}\mathbf{D}(\mathbf{T}^{\top})^{-1}$, where $\mathbf{T}$ is a unit-lower triangular matrix with 1's on the diagonals and serial dependence parameters on the off-diagonals and $\mathbf{D}$ is a diagonal matrix with variances on the diagonals. The serial dependence parameters in $\mathbf{T}$ and the natural logarithm of variances in $\mathbf{D}$ are unconstrained, have statistical interpretations, and can be modelled in terms of covariates. In the statistical literature for longitudinal data, first of all, Pourahmadi (1999) and Pourahmadi (2000) used modified Cholesky decomposition to model marginal covariance matrix of univariate longitudinal continuous responses, then estimated resulting parameters through maximum likelihood estimation (MLE) method. They showed that 11×11 marginal covariance matrix $\mathbf{\Sigma}$ can be modelled with 8 parameters, instead of 66 parameters. Since then, there have been existed numerous examples using modified Cholesky decomposition within marginal models for continuous responses such as Pan et al. (2013), but among them, Kim and Zimmerman (2012) were the first, who extended the same approach to multivariate longitudinal data with continuous responses by proposing modified Cholesky Block decomposition method. Daniels and Pourahmadi (2002) and Pourahmadi and Daniels (2002) extended modified Cholesky decomposition to dynamic conditionally linear mixed models (LMMs) within a Bayesian context. They modelled the parameters of the decomposed marginal covariance matrix in terms of baseline covariates and assumed a constant structure for the random effects covariance matrix. On the other hand, Pan and Mackenzie (2007) and Li and Pourahmadi (2013) issued the sample problem to decompose the conditional covariance matrix of LMMs.

The first work on modelling the random effects covariance matrix within LMMs was proposed by Daniels and Zhao (2003). Daniels and Zhao (2003) assumed a vector of random effects for each subject i in the study, namely, $\mathbf{b}_i = (b_{i1}, \ldots, b_{ik}, \ldots, b_{iq})$, where b_{i1} was random intercept for ith subject and the remaining ones were random slopes, which were associated with the covariates related to time variable in the design matrix and $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{\Sigma}_i)$, where $\mathbf{\Sigma}_i$ was a $q \times q$ subject-specific covariance matrix of random effects. Then $\mathbf{\Sigma}_i$ was decomposed into a unique unit-lower triangular matrix $\mathbf{T}_i$ and a unique diagonal matrix $\mathbf{D}_i$ by modified Cholesky decomposition. The resulting parameters were modelled parsimoniously through some covariates. While the parameter estimation for the model was based on Bayesian inference, naturally, the selection of models with different random effects covariance structures was achieved through the deviance information criterion (DIC). On the other hand, Das et al. (2013) and Das and Daniels (2014) extended this approach to analyse bivariate longitudinal continuous response data, which were measured at irregularly time points, through a novel random effects model, where Bayesian inference was used for parameter estimation.

Lee et al. (2012) were the first, who extended the idea of associating the elements of random effects covariance matrix with covariates within generalized linear mixed models (GLMMs) for univariate longitudinal binary responses. Following Heagerty (1999), in the linear predictor of the GLMM, Lee et al. (2012) assumed a vector of time-varying and correlated intercepts for each subject i in the study, namely, $\mathbf{b}_i = (b_{i1}, \ldots, b_{it}, \ldots, b_{in_i})$, where b_{it} was the random intercept at time t for ith subject and $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{\Sigma}_i)$, where $\mathbf{\Sigma}_i$ was a $n_i \times n_i$ subject-specific covariance matrix of random effects. Then $\mathbf{\Sigma}_i$ was decomposed into a unique unit-lower triangular matrix $\mathbf{T}_i$ and a unique diagonal matrix $\mathbf{D}_i$ by modified Cholesky decomposition. The resulting parameters were modelled parsimoniously through some covariates. While the parameter estimation for the model was based on MLE, naturally, the selection of models with different random effects covariance structures was achieved through the Akaike information criterion (AIC). Lee (2013) further illustrated the same idea within a Bayesian framework.

The literature review revealed that the literature on modified Cholesky decomposition in longitudinal data mostly relies on univariate longitudinal continuous data with two exceptions for multivariate continuous data and one exception for binary data and has not been addressed either for multivariate longitudinal binary data or within marginalized multilevel models (MMMs) framework so far. Following the idea of Lee et al. (2012), we adapt this approach to model the elements of the covariance matrix of random effects in the proposed marginalized multilevel model with multiple correlated random effects in Chapter 3. Since we focus on the study of Lee et al. (2012), it would be useful to briefly review the modified Cholesky decomposition.

The Modified Cholesky Decomposition Theorem: If $\mathbf{\Sigma}$ is a $n \times n$ symmetric positive definite matrix, then a unique unit-lower triangular matrix $\mathbf{T}$ with unconstrained elements and a unique diagonal matrix $\mathbf{D}$ exist with positive diagonal elements such that

$$\mathbf{T}\boldsymbol{\Sigma}\mathbf{T}^\top = \mathbf{D}. \tag{2.14}$$

Corollary 1: The unit-lower triangular matrix **T** in modified Cholesky decomposition theorem is non-singular and its inverse is also unit-lower triangular.

Based on this theorem, Lee et al. (2012) considered GLMMs, where $g(\mu_{it}) = \mathbf{X}_{it}\boldsymbol{\beta} + b_{it}$ and $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$. When $\boldsymbol{\Sigma}_i$ is decomposed into $\mathbf{T}_i$ and $\mathbf{D}_i$, the lower diagonal entries of $\mathbf{T}_i$ represent the negative coefficients when b_{it} is regressed on its predecessors, $b_{i1}, \ldots, b_{it-1}$ such as $b_{i1} = e_{i1}$ and $b_{it} = \sum_{l=1}^{t-1} \phi_{i,tl} b_{il} + e_{it}$ for $t = 2, \ldots, n_i$ and $1 \leq l \leq t-1$. The diagonal entries of $\mathbf{D}_i = (\sigma^2_{i1}, \ldots, \sigma^2_{it}, \ldots, \sigma^2_{in_i})$ are the error variances such that $\mathbf{e}_i \sim N(0, \mathbf{D}_i)$.

This expression can be written in terms of matrix as follows:

$$\mathbf{T}_i \mathbf{b}_i = \mathbf{e}_i, \tag{2.15}$$

where $\mathbf{T}_i$ is a unit-lower triangular matrix with $-\phi_{i,tl}$ at its (t, l)th position for $1 \leq l < t \leq n_i$. Specifically, $\mathbf{T}_i$ is given as:

$$\mathbf{T}_i = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{i,21} & 1 & 0 & \cdots & 0 \\ -\phi_{i,31} & -\phi_{i,32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ -\phi_{i,n_i1} & -\phi_{i,n_i2} & \cdots & -\phi_{i,n_in_i-1} & 1 \end{pmatrix}$$

Then $Cov(\mathbf{T}_i, \mathbf{b}_i) = \mathbf{T}_i \boldsymbol{\Sigma}_i \mathbf{T}_i^\top = Cov(\mathbf{e}_i) = \mathbf{D}_i$, which is given below, completes the modified Cholesky decomposition theorem.

$$\mathbf{D}_i = \begin{pmatrix} \sigma^2_{i1} & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2_{i2} & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2_{i3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \cdots & \sigma^2_{in_i} \end{pmatrix}$$

The parameters in $\mathbf{T}_i$ and $\mathbf{D}_i$ measure dependence and variance, respectively. The positive definiteness constraint on σ_{it}^2's can be removed by using a natural logarithm transformation, namely, $log(\sigma_{it}^2)$'s. The modified Cholesky decomposition ensures that an unstructured covariance matrix $\mathbf{\Sigma}_i$ with $\frac{(n_i \times (n_i+1))}{2}$ parameters turns into a covariance matrix $\mathbf{\Sigma}_i$ with the same number of unconstrained parameters $\phi_{i,tl}$'s and $log(\sigma_{it}^2)$'s (Daniels and Pourahmadi, 2009). Following this, the number of parameters to be estimated in $\mathbf{\Sigma}_i$ can be reduced by modelling $\phi_{i,tl}$'s and $log(\sigma_{it}^2)$'s by using covariates with low-dimensional regression parameters, which are assumed to be shared by all $\mathbf{\Sigma}_i$'s (Pourahmadi et al., 2007), such that

$$\begin{aligned} \phi_{i,tl} &= \mathbf{w}_{i,tl}\boldsymbol{\gamma} \quad \text{and} \\ log(\sigma_{it}^2) &= \mathbf{h}_{i,tl}\boldsymbol{\lambda}, \end{aligned} \tag{2.16}$$

where $\mathbf{w}_{i,tl}$ and $\mathbf{h}_{i,tl}$ are $1 \times q$ and $1 \times r$ vectors of covariates, respectively. $\mathbf{w}_{i,tl}$ and $\mathbf{h}_{i,tl}$ are assumed to be a subset of $\mathbf{X}_{it}$. $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are the corresponding $q \times 1$ and $r \times 1$ vectors of regression parameters, respectively.

CHAPTER 3

PROPOSED MODEL

We organize this chapter into four sections. In Section 3.1, we introduce our model and give definitions, assumptions related to the proposed model in detail. We give derivations related to the connection between the levels of the proposed model and to the marginal correlation of responses in Sections 3.2 and 3.3, respectively. We present maximum likelihood estimation (MLE) of the parameters and Empirical Bayes (EB) estimation of random effects in the proposed model in Sections 3.4 and 3.5, respectively.

3.1 Model Definition and Assumptions

In this thesis study, we propose a model based on marginalized multilevel model framework for bivariate longitudinal binary data, which consists of two different levels, where each level is a different regression model. In the proposed model, the first level associates marginal mean of responses with covariates through a logistic regression model and the second level restructures the mean of responses conditional on subject/time specific random intercepts through a probit regression model to represent the longitudinal association(s).

The proposed logistic-probit-normal marginalized multilevel model for bivariate longitudinal binary data is given as follows:

$$Pr(Y_{itj} = 1|\mathbf{X}_{it}) = \frac{exp(\mathbf{X}_{it}\boldsymbol{\beta}_j)}{1 + exp(\mathbf{X}_{it}\boldsymbol{\beta}_j)},$$

$$\begin{aligned} Pr(Y_{itj} = 1|\mathbf{X}_{it}, b_{it}) = \Phi(\Delta_{itj}(\mathbf{X}_{it}) + b_{it}), \quad \text{and} \\ \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{\Sigma}_i), \end{aligned} \tag{3.1}$$

where Y_{itj} denotes the response of ith subject (i = 1,2,. . . ,N) at time t (t = 1,2,. . . ,n_i) for jth dimension (j = 1,2), $\mathbf{X}_{it}$ is a $1 \times P$ vector of covariates, which may be time-variant (e.g., the subject's weight) and/or time-invariant (e.g., the subject's gender) and is assumed to be common for each response type. $\boldsymbol{\beta}_j$ is the corresponding $P \times 1$ vector of response-specific marginal regression parameter, which estimates the association between the mean of longitudinal responses and covariates through odds-ratio formula due to the logit link function (note that the first level of the model in equation 3.1 can also be written as $logit(Pr(Y_{itj} = 1|\mathbf{X}_{it})) = \mathbf{X}_{it}\boldsymbol{\beta}_j$). Furthermore, $\boldsymbol{\beta}_j$'s have population-averaged interpretations and are robust to model specification for the second level of the model since the marginal mean model is specified separately from the conditional mean model, as a general feature of MMMs (see Table 2.1). Δ_{itj} is subject/time/response specific term, which connects the second level to the first level of the model, as will be shown in equation 3.6. b_{it} is subject-specific and time-varying intercept, which is shared by different binary response types and represents the random-variation within the measurements of ith subject as well as between the measurements of different subjects. Given b_{it}, Y_{itj} is assumed to be independent of all $Y_{it'j}$ for $t = 1, \ldots, n_i$ & $t' \neq t$, which is known as conditional independence assumption and, in turn, enables to construct a marginal joint distribution for longitudinal binary responses (e.g., $f(Y_{i11}, Y_{i21}, \ldots, Y_{in_i1}, Y_{i12}, Y_{i22}, \ldots, Y_{in_i2})$). Furthermore, $\mathbf{b}_i = (b_{i1}, \ldots, b_{it}, \ldots, b_{in_i})$ is a $n_i \times 1$ vector of random intercepts for ith subject. $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{\Sigma}_i)$, where **0** is a $n_i \times 1$ vector of zeros and $\mathbf{\Sigma}_i$ is a $n_i \times n_i$ covariance matrix, which is assumed to represent the dependence between the successive measurements of ith subject over time for a given response type. Since the dimension of $\mathbf{\Sigma}_i$ depends on total number of time points n_i, not on the number of response types, extension of the proposed model to multivariate case with more than bivariate responses is straightforward.

Inflation in the number of parameters to be estimated in $\mathbf{\Sigma}_i$ may result in identifiability problems and high-dimensional matrix inversion and differentiation of $\mathbf{\Sigma}_i$ may result in computational burden during the parameter estimation. To avoid the positive definiteness constraint on $\mathbf{\Sigma}_i$ and potential statistical and computational problems that may be associated with its high-dimensionality, $\mathbf{\Sigma}_i$ is further decomposed into its dependence and variance components through modified Cholesky decomposition method. Then unconstrained parameters of the resulting dependence and variance matrices are modelled in terms of covariates with low-dimensional regression parameters as follows:

$$\mathbf{\Sigma}_i = \mathbf{T}_i^{-1}\mathbf{D}_i(\mathbf{T}_i^{\top})^{-1}, \tag{3.2}$$

where $\mathbf{T}_i$ is a $n_i \times n_i$ unique unit-lower triangular matrix with 1's on the main diagonal and unconstrained elements $\phi_{i,tl}$'s at its (l,t)th position for $l < t$ such that

$$\mathbf{T}_i = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{i,21} & 1 & 0 & \cdots & 0 \\ -\phi_{i,31} & -\phi_{i,32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ -\phi_{i,n_i1} & -\phi_{i,n_i2} & \cdots & -\phi_{i,n_in_i-1} & 1 \end{pmatrix} \quad \text{and}$$

$\mathbf{D}_i$ is a $n_i \times n_i$ unique-diagonal matrix with positive diagonal elements σ_{it}^2's such that

$$\mathbf{D}_i = \begin{pmatrix} \sigma_{i1}^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_{i2}^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_{i3}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \cdots & \sigma_{in_i}^2 \end{pmatrix}.$$

$\phi_{i,tl}$'s and σ_{it}^2's are named as the generalized autoregressive parameters (GARPs) and the innovation variances (IVs), respectively and they have sensible statistical interpretations such as while $\phi_{i,tl}$ measures serial dependence between two random effects b_{it} and b_{il}, where $l < t$, σ_{it}^2 is prediction variance for the random effect b_{it}.

Since $\phi_{i,tl}$'s and $log(\sigma_{it}^2)$'s are unconstrained in parameter space, it is possible to associate these parameters with covariates through low-dimensional regression parameters, as follows:

$$\begin{aligned} \phi_{i,tl} &= \mathbf{w}_{i,tl}\boldsymbol{\gamma} \quad \text{and} \\ log(\sigma^2_{it}) &= \mathbf{h}_{i,tl}\boldsymbol{\lambda}, \end{aligned} \tag{3.3}$$

where $\mathbf{w}_{i,tl}$ and $\mathbf{h}_{i,tl}$ are $1 \times q$ and $1 \times r$ vectors of covariates, respectively. $\mathbf{w}_{i,tl}$ and $\mathbf{h}_{i,tl}$ are assumed to be a subset of $\mathbf{X}_{it}$. $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are corresponding $q \times 1$ and $r \times 1$ vectors of regression parameters. Consequently, as summarized in Section 2.3, this approach allows i) flexibility for the structure of $\boldsymbol{\Sigma}_i$, ii) better explanations related to the elements of $\boldsymbol{\Sigma}_i$, and iii) a reduction in the number of parameters to be estimated in $\boldsymbol{\Sigma}_i$ (e.g., a reduction from $\frac{(n_i(n_i+1))}{2}$ number of parameters to $q + r$ number of parameters), which is an advantage for the models that cannot be implemented or are hardly implemented due to the high-dimensionality of the covariance matrix and may avoid possible identifiability problems.

3.2 Connection between the Levels of the Proposed Model and Calculation of Δ_{itj}

As mentioned in Section 2.1, the fact that any conditional expectation can be written in terms of marginal expectation implies that integration of conditional probability $Pr(Y_{itj} = 1|\mathbf{X}_{it}, b_{it})$ over the distribution of random effects results in marginal probability $Pr(Y_{itj} = 1|\mathbf{X}_{it})$ in random effects models for longitudinal binary data. That is,

$$\begin{aligned} E(Y_{itj}|\mathbf{X}_{it}) &= \int E(Y_{itj}|\mathbf{X}_{it}, b_{it}) f(b_{it}) db_{it} \\ \Rightarrow Pr(Y_{itj} = 1|\mathbf{X}_{it}) &= \int Pr(Y_{it} = 1|\mathbf{X}_{it}, b_{it}) f(b_{it}) db_{it}, \end{aligned} \tag{3.4}$$

where $f(b_{it})$ is a univariate normal distribution with mean 0 and variance $var(b_{it})$. Then substituting marginal and conditional probabilities in equation 3.1, which are the first and second levels of the proposed model, respectively, into equation 3.4 gives the following expression:

$$\frac{exp(\mathbf{X}_{it}\boldsymbol{\beta}_j)}{1+exp(\mathbf{X}_{it}\boldsymbol{\beta}_j)} = \int_{-\infty}^{+\infty} \Phi(\Delta_{itj}(\mathbf{X}_{it}) + b_{it}) f(b_{it}) db_{it}. \tag{3.5}$$

Following Griswold (2005), due to good collaboration of probit link function used in the second level of the model and normal distribution assumption for random effects, it is shown that a closed-form solution for Δ_{itj} exists such that

$$\Delta_{itj} = \Phi^{-1}\left(\frac{exp(\mathbf{X}_{it}\boldsymbol{\beta}_j)}{1+exp(\mathbf{X}_{it}\boldsymbol{\beta}_j)}\right)\sqrt{1+Var(b_{it})}. \tag{3.6}$$

Details of this derivation are presented in Appendix A. It is obvious that Δ_{itj} in equation 3.1 is an explicit function of $\mathbf{X}_{it}$, the marginal regression parameters, and variance of random effects and plays a bridge role between the first and the second levels of the model.

3.3 Marginal Correlation of Y_{itj}'s

The random effects covariance matrix $\boldsymbol{\Sigma}_i$ of the proposed model in equation 3.1 represents the association within the repeated measurements of ith subject over time for a given response. To illustrate, $Corr\left(Y_{itj}, Y_{it'j}\right)$ is represented by $Corr\left(b_{it}, b_{it'}\right)$ and it is estimated through using the relevant covariance and variance components in $\boldsymbol{\Sigma}_i$. On the other hand, the association between measurements of two different responses of ith subject for a given time point is not represented by $\boldsymbol{\Sigma}_i$ in the proposed model. At this point, Das et al. (2013) pointed out that repeated measurements of a subject might be internally correlated and it is not necessary to represent them through random effects. In this sense, following Goldstein and Rasbash (1996), Vangeneugden et al. (2010), and Vangeneugden et al. (2011), it is shown that marginal correlation $Corr\left(Y_{itj}, Y_{itj'}\right)$ still exists in the model through a Taylor series-based approximation as follows:

$$Corr\left(Y_{itj}, Y_{itj'}\right) \simeq \frac{\phi(\Delta_{itj}(\mathbf{X}_{it}))Var\left(b_{it}\right)\phi(\Delta_{itj'}(\mathbf{X}_{it}))}{\sqrt{v_{itj} + (\phi(\Delta_{itj}(\mathbf{X}_{it})))^2 Var(b_{it})}\sqrt{v_{itj'} + (\phi(\Delta_{itj'}(\mathbf{X}_{it})))^2 Var(b_{it})}}, \tag{3.7}$$

where $v_{itj} = \Phi(\Delta_{itj}(\mathbf{X}_{it}))(1-\Phi(\Delta_{itj}(\mathbf{X}_{it})))$, $v_{itj'} = \Phi(\Delta_{itj'}(\mathbf{X}_{it}))(1-\Phi(\Delta_{itj'}(\mathbf{X}_{it})))$ and details are presented in Appendix B. Note that reason for failing to derive an exact expression for marginal correlations is due to the non-linear probit link function in the second level of the proposed model.

3.4 Marginal Likelihood Definition of the Proposed Model

Like any generalized linear mixed models (GLMMs), in MMMs for multivariate longitudinal binary data with random effects, the marginal likelihood of data $L(\boldsymbol{\theta}|\mathbf{y})$ involves integration of conditional probabilities over the distribution of random effects, as follows:

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^{N} \int \prod_{t=1}^{n_i} \prod_{j=1}^{J} f(y_{itj}|\mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i, \tag{3.8}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ is the vector of unknown parameters in the model (see equations 3.1 and 3.3), y_{itj} represents the measurement of ith subject at time t for jth response type ($i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, n_i$, and $j = 1, 2, \ldots, J$), $f(y_{itj}|\mathbf{b}_i)$ is the conditional distribution of y_{itj}, which is the second level of the model given in equation 3.1. In other words, $f(y_{itj}|\mathbf{b}_i) = Pr(Y_{itj} = 1|\mathbf{X}_{it}, b_{it}) = \Phi(\Delta_{itj}(\mathbf{X}_{it}) + b_{it})$, where $\Delta_{itj} = \Phi^{-1}\left(\frac{exp(\mathbf{X}_{it}\boldsymbol{\beta}_j)}{1+exp(\mathbf{X}_{it}\boldsymbol{\beta}_j)}\right)\sqrt{1 + Var(b_{it})}$ is given in equation 3.6. The $\mathbf{b}_i = (b_{i1}, \ldots, b_{it}, \ldots, b_{in_i})$ is a $n_i \times 1$ vector of random effects, which comes from a multivariate normal distribution $f(\mathbf{b}_i)$ with $\mathbf{0}$ mean and $\boldsymbol{\Sigma}_i$ covariance matrix, where $\boldsymbol{\Sigma}_i = \mathbf{T}_i^{-1}\mathbf{D}_i(\mathbf{T}_i^{\top})^{-1}$ is given in equation 3.2. The underlying assumption in equation 3.8 is that given $\mathbf{b}_i$, y_{itj} is assumed to be independent of $y_{it'j'}$, $\forall\ t \neq t'$ and $j \neq j'$, which is well-known as conditional independence assumption.

For the proposed model in equation 3.1, the likelihood function in equation 3.8 involves multiple correlated random effects $\mathbf{b}_i$ and due to the non-conjugacy of $f(\mathbf{b}_i)$ with $f(y_{itj}|\mathbf{b}_i)$, the integral in equation 3.8 is not tractable and, in turn, it does not provide a closed-form solution for $L(\boldsymbol{\theta}|\mathbf{y})$. Furthermore, since the dimension of the integral in equation 3.8 is equal to the dimension of $\mathbf{b}_i$ (Karl et al., 2014) (e.g., n_i for each i), increase in the dimension of the integral further complicates the numerical evaluation of the integral. For that reason, the integral in the equation 3.8 needs to be evaluated through numerical methods for high-dimensional integration. At this point, it is worthy to say that due to the aforementioned reasons, most of the models in the literature are restricted to only a random-intercept model and multiple correlated random effects are usually avoided (Kim et al., 2013), where several examples are available in Section 2.2. On the other hand, even $L(\boldsymbol{\theta}|\mathbf{y})$ is numerically evaluated, due to the probit link function, a non-linear link function, used in $f(y_{itj}|\mathbf{b}_i)$ to associate conditional probabilities of y_{itj}'s with covariates on a continuous scale, the optimization of the marginal likelihood $L(\boldsymbol{\theta}|\mathbf{y})$ (or, naturally, the optimization of $logL(\boldsymbol{\theta}|\mathbf{y})$) does not provide a closed-form solution for $\hat{\boldsymbol{\theta}}$, which is the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$. For that reason, $logL(\boldsymbol{\theta}|\mathbf{y})$ should be either approximated or evaluated numerically through iterative algorithms such as Newton-Raphson (N-R) or Fisher-Scoring (F-S), which requires the computation of second-order and first order derivatives of $logL(\boldsymbol{\theta}|\mathbf{y})$ with respect to $\boldsymbol{\theta}$, respectively. However, due to high-dimensional $\mathbf{b}_i$, the optimization of $logL(\boldsymbol{\theta}|\mathbf{y})$ includes taking the inverse of high-dimensional covariance matrices and the differentiation of these matrices with respect to $\boldsymbol{\theta}$, which may also result in a computational burden.

To avoid numerical evaluation of high-dimensional integral, the computation of high-dimensional matrix inversion and differentiation, and numerical optimization of the marginal likelihood, in this thesis study, we use data cloning (DC) computational algorithm (Lele et al., 2007; Lele, 2010) to compute the MLEs of the parameters in the proposed model in equation 3.1 and their standard errors. The details related to DC computational algorithm is available in Section 2.2.

3.5 Empirical Bayes Estimation of Random Effects

Both population-averaged and subject-specific interpretations are possible through MMMs, since they combine the strengths of marginal and conditional models. While the first level of MMMs enables to do population-averaged interpretations based on the MLEs for $\boldsymbol{\beta}$, the second level of the MMMs allows to do subject-specific inferences through the MLEs of $\boldsymbol{\beta}$, γ, $\boldsymbol{\lambda}$, and the prediction of random effects. Based on the MLEs for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \gamma, \boldsymbol{\lambda})$, the prediction of vector of random effects for ith subject conditional on the observed data is possible through the following posterior density via MCMC sampling:

$$\prod(\mathbf{b}_i|\mathbf{y}_i) = \frac{f(\mathbf{y}_i|\mathbf{b}_i, \hat{\boldsymbol{\theta}}) f(\mathbf{b}_i|\hat{\gamma}, \hat{\boldsymbol{\lambda}})}{C(\mathbf{y}_i)}, \tag{3.9}$$

where $f(\mathbf{y}_i|\mathbf{b}_i, \hat{\boldsymbol{\theta}}) = \prod_{t=1}^{n_i} \prod_{j=1}^{J} f(y_{itj}|\mathbf{b}_i)$ is joint distribution of responses for ith subject evaluated at $\hat{\boldsymbol{\theta}}$, $f(\mathbf{b}_i|\hat{\gamma}, \hat{\boldsymbol{\lambda}})$ is the prior distribution of random effect vector for ith subject evaluated at $\hat{\gamma}$ and $\hat{\boldsymbol{\lambda}}$, and $C(\mathbf{y}_i)$ is the normalizing constant.

CHAPTER 4

APPLICATION

In this chapter, we firstly assess the validity of the proposed model under different simulation scenarios through some statistical performance criteria such as average, bias, Monte Carlo standard error (SE), average of the model based standard errors (ASE), mean squared error (MSE), and coverage probability (CP) and then discuss the results of the simulation scenarios in Section 4.1. Then, we illustrate the proposed model through Mother's Stress and Children's Morbidity (MSCM) study data in Section 4.2.

4.1 Numerical Assessment of the Proposed Model

A simulation study is conducted to justify the validity of the proposed model in equation 3.1 under different scenarios through some performance measures. Data are generated in R and the model is fitted using `jags.fit` function under `dclone` package in R (Sólymos, 2010).

4.1.1 Simulation Design

Simulation Scenario 1: Equally spaced $\&$ equal number of time points

Under Scenario 1, the number of time points is assumed to be the same for all subjects and it is $n_i = n = 4$. The spacing between two consecutive time points is also assumed to be the same (e.g., the measurements are taken each week within a month).

Then, the bivariate longitudinal binary data is simulated based on the proposed model in equation 3.1 with two different total number of subjects ($N = 150, 300$) and two different number of covariates ($P = 2, 4$), which results in $2^2 = 4$ different sub-scenarios.

In case of $P = 2$, the marginal probabilities are specified as follows:

$$\begin{aligned} Pr(Y_{it1} = 1|\mathbf{x}_{it}) &= \frac{exp(\beta_{10} + \beta_{11}x_{it1} + \beta_{12}x_{it2})}{1 + exp(\beta_{10} + \beta_{11}x_{it1} + \beta_{12}x_{it2})} \\ Pr(Y_{it2} = 1|\mathbf{x}_{it}) &= \frac{exp(\beta_{20} + \beta_{21}x_{it1} + \beta_{22}x_{it2})}{1 + exp(\beta_{20} + \beta_{21}x_{it1} + \beta_{22}x_{it2})}, \end{aligned} \tag{4.1}$$

where $i = 1, 2, \ldots, 150$ or $i = 1, 2, \ldots, 300$, $t = 1, 2, 3, 4$, x_{it1} is the centered time at tth measurement, $x_{it2} = x_{i2}$ is the centered time-invariant binary covariate generated from Bernoulli distribution with equal probability, $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \beta_{12}) = (0.9, 0.6, -0.6)$, and $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}) = (1.1, 0.7, -0.9)$. The true value for each β is chosen such that exponential of the true value, which is odds-ratio, is far from 1 (e.g., $exp(\beta_{11}) = 1.82$).

The conditional probabilities are specified as follows:

$$\begin{aligned} Pr(Y_{it1} = 1|\mathbf{x}_{it}, b_{it}) &= \Phi(\Delta_{it1}(\mathbf{X}_{it}) + b_{it}) \\ Pr(Y_{it2} = 1|\mathbf{x}_{it}, b_{it}) &= \Phi(\Delta_{it2}(\mathbf{X}_{it}) + b_{it}), \end{aligned} \tag{4.2}$$

where Δ_{itj} ($j = 1, 2$) is given such that:

$$\begin{aligned} \Delta_{it1} &= \Phi^{-1}\left(\frac{exp(\beta_{10} + \beta_{11}x_{it1} + \beta_{12}x_{it2})}{1 + exp(\beta_{10} + \beta_{11}x_{it1} + \beta_{12}x_{it2})}\right)\sqrt{1 + Var(b_{it})} \\ \Delta_{it2} &= \Phi^{-1}\left(\frac{exp(\beta_{20} + \beta_{21}x_{it1} + \beta_{22}x_{it2})}{1 + exp(\beta_{20} + \beta_{21}x_{it1} + \beta_{22}x_{it2})}\right)\sqrt{1 + Var(b_{it})}, \end{aligned} \tag{4.3}$$

b_{it} is tth element of $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3}, b_{i4})$ and $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ with $\boldsymbol{\Sigma_i} = \mathbf{T}_i^{-1}\mathbf{D}_i(\mathbf{T}_i^{\top})^{-1}$ (see equation 3.2 for the general formulation). $Var(b_{it})$ is tth diagonal component of $\boldsymbol{\Sigma}_i$. Furthermore, only lag-1 serial dependencies (dependence between two consecutive time points, e.g., $t = 2$ & $l = 1$; $t = 3$ & $l = 2$; and $t = 4$ & $l = 3$) are assumed to be significant and other lags are equated to zero for simplicity. Hence, $\mathbf{T}_i$ and $\mathbf{D}_i$ can be represented as follows:

$$\mathbf{T}_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\phi_{i,21} & 1 & 0 & 0 \\ 0 & -\phi_{i,32} & 1 & 0 \\ 0 & 0 & -\phi_{i,43} & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{D}_i = \begin{pmatrix} \sigma_{i1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{i2}^2 & 0 & 0 \\ 0 & 0 & \sigma_{i3}^2 & 0 \\ 0 & 0 & 0 & \sigma_{i4}^2 \end{pmatrix}.$$

Then the parameters in $\mathbf{T}_i$ and $\mathbf{D}_i$ are associated with x_{it2} as follows:

$$\begin{aligned} \phi_{i,tl} &= \gamma_0 I_{(|t-l|=1)} + \gamma_1 x_{it2} I_{(|t-l|=1)} \quad \text{and} \\ log(\sigma_{it}^2) &= \lambda_0 + \lambda_1 x_{it2}, \end{aligned} \tag{4.4}$$

where $I_{(|t-l|=1)}$ denotes that only lag-1 serial dependencies are significant, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)$ $= (-0.3, -1.3)$, and $\boldsymbol{\lambda} = (\lambda_0, \lambda_1) = (0.5, 1.5)$. Note that increasing number of significant lags, which are different from zero, do not inflate the number of parameters to be estimated in $\boldsymbol{\Sigma}_i$ since $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are free from t and l ($1 \leq l < t \leq n_i$). On the other hand, only one covariate is associated with the random effects parameters not to increase the model complexity.

Under this sub-scenario, the association between two longitudinal binary responses for each time point (e.g., $t = 1, 2, 3, 4$) is also investigated through odds-ratio formula (Chen et al., 2012), given as follows:

$$OR_t = \frac{Pr(Y_{t1} = 0, Y_{t2} = 0) Pr(Y_{t1} = 1, Y_{t2} = 1)}{Pr(Y_{t1} = 1, Y_{t2} = 0) Pr(Y_{t1} = 0, Y_{t2} = 1)}, \tag{4.5}$$

where $Pr(Y_{t1}, Y_{t2})$ denotes joint probability of bivariate binary responses at time t ($t = 1, 2, 3, 4$). For a couple of simulated data sets, $\hat{OR}_t$ for $t = 1, 2, 3, 4$ is estimated around $7, 13, 17$ and 28, respectively, which indicates that association between two responses is high and increases as time increases.

In case of $P = 4$, the marginal probabilities are specified as follows:

$$\begin{aligned} Pr(Y_{it1}=1|\mathbf{x}_{it}) &= \frac{exp(\beta_{10}+\beta_{11}x_{it1}+\beta_{12}x_{it2}+\beta_{13}x_{it3}+\beta_{14}x_{it4})}{1+exp(\beta_{10}+\beta_{11}x_{it1}+\beta_{12}x_{it2}++\beta_{13}x_{it3}+\beta_{14}x_{it4})} \\ Pr(Y_{it2}=1|\mathbf{x}_{it}) &= \frac{exp(\beta_{20}+\beta_{21}x_{it1}+\beta_{22}x_{it2}+\beta_{23}x_{it3}+\beta_{24}x_{it4})}{1+exp(\beta_{20}+\beta_{21}x_{it1}+\beta_{22}x_{it2}+\beta_{23}x_{it3}+\beta_{24}x_{it4})}, \end{aligned} \tag{4.6}$$

where $i = 1, 2, \ldots, 150$ or $i = 1, 2, \ldots, 300$, $t = 1, 2, 3, 4$, x_{it1} is the centered time at tth measurement, $x_{it2} = x_{i2}$ is the centered time-invariant binary covariate generated from Bernoulli distribution with equal probability, $x_{it3} = x_{i3}$ is the centered time-invariant continuous covariate generated from Uniform distribution within $(-1, 1)$, x_{it4} is the centered time-varying binary covariate generated from Bernoulli distribution with equal probability, $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}) = (0.9, 0.6, -0.6, 0.5, -0.3)$, and $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}) = (1.1, 0.7, -0.9, 0.4, -0.4)$.

The conditional probabilities are specified as follows:

$$\begin{aligned} Pr(Y_{it1}=1|\mathbf{x}_{it}, b_{it}) &= \Phi(\Delta_{it1}(\mathbf{X}_{it}) + b_{it}) \\ Pr(Y_{it2}=1|\mathbf{x}_{it}, b_{it}) &= \Phi(\Delta_{it2}(\mathbf{X}_{it}) + b_{it}), \end{aligned} \tag{4.7}$$

where Δ_{itj} $(j = 1, 2)$ is given such that:

$$\begin{aligned} \Delta_{it1} &= \Phi^{-1}\left(\frac{exp(\beta_{10}+\beta_{11}x_{it1}+\beta_{12}x_{it2}+\beta_{13}x_{it3}+\beta_{14}x_{it4})}{1+exp(\beta_{10}+\beta_{11}x_{it1}+\beta_{12}x_{it2}+\beta_{13}x_{it3}+\beta_{14}x_{it4})}\right)\sqrt{1+Var(b_{it})} \\ \Delta_{it2} &= \Phi^{-1}\left(\frac{exp(\beta_{20}+\beta_{21}x_{it1}+\beta_{22}x_{it2}+\beta_{23}x_{it3}+\beta_{24}x_{it4})}{1+exp(\beta_{20}+\beta_{21}x_{it1}+\beta_{22}x_{it2}++\beta_{23}x_{it3}+\beta_{24}x_{it4})}\right)\sqrt{1+Var(b_{it})}, \end{aligned} \tag{4.8}$$

b_{it} is tth element of $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3}, b_{i4})$ and $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ with $\boldsymbol{\Sigma_i} = \mathbf{T}_i^{-1}\mathbf{D}_i(\mathbf{T}_i^{\top})^{-1}$ (see equation 3.2 for the general formulation). $Var(b_{it})$ is tth diagonal component of $\boldsymbol{\Sigma}_i$. Furthermore, the same setting in equation 4.4 is assumed for $\boldsymbol{\Sigma}_i$.

Simulation Scenario 2: Unequally spaced $\&$ equal number of time points

Under Scenario 2, similar to Scenario 1, the number of time points is assumed to be the same for all subjects and it is $n_i = n = 4$. However, the spacing between two consecutive time points is assumed to be unequal (e.g., while one measurement is taken at the first week, the other one is taken at the second week, but the next one is taken at a longer time point than one-week). Then, the bivariate longitudinal binary data is simulated based on the proposed model in equation 3.1 with a total number of subjects $N = 150$ and number of covariates $P = 2$.

The marginal probabilities are specified as follows:

$$\begin{aligned} Pr(Y_{it1} = 1|\mathbf{x}_{it}) &= \frac{exp(\beta_{10} + \beta_{11}x_{it1} + \beta_{12}x_{it2})}{1 + exp(\beta_{10} + \beta_{11}x_{it1} + \beta_{12}x_{it2})} \\ Pr(Y_{it2} = 1|\mathbf{x}_{it}) &= \frac{exp(\beta_{20} + \beta_{21}x_{it1} + \beta_{22}x_{it2})}{1 + exp(\beta_{20} + \beta_{21}x_{it1} + \beta_{22}x_{it2})}, \end{aligned} \tag{4.9}$$

where $i = 1, 2, \dots, 150$, $t = 1, 2, 4, 7$, x_{it1} is the centered time at tth measurement, $x_{it2} = x_{i2}$ is the centered time-invariant binary covariate generated from Bernoulli distribution with equal probability, $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \beta_{12}) = (0.9, 0.6, -0.6)$, and $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}) = (1.1, 0.7, -0.9)$. The conditional probabilities are specified as in equations 4.2-4.4. $I_{(|t-l|=1)}$ in equation 4.4 still denotes the consecutive serial dependencies (i.e., dependence between two consecutive time points, e.g., $t = 2$ & $l = 1$; $t = 4$ & $l = 2$; and $t = 7$ & $l = 4$) are significant, and other lags are equated to zero for simplicity.

Simulation Scenario 3: Equally spaced & unequal number of time points

Under Scenario 3, three different number of time points is assumed across the subjects (e.g., $n_i = 3, 4$, or 5), whereas the spacing between two consecutive time points is assumed to be equal. Then, the bivariate longitudinal binary data is simulated based on the proposed model in equation 3.1 with a total number of subjects $N = 150$ and number of covariates $P = 2$.

The marginal probabilities are specified as follows:

$$\begin{aligned} Pr(Y_{it1} = 1|\mathbf{x}_{it}) &= \frac{exp(\beta_{10} + \beta_{11}x_{it1} + \beta_{12}x_{it2})}{1 + exp(\beta_{10} + \beta_{11}x_{it1} + \beta_{12}x_{it2})} \\ Pr(Y_{it2} = 1|\mathbf{x}_{it}) &= \frac{exp(\beta_{20} + \beta_{21}x_{it1} + \beta_{22}x_{it2})}{1 + exp(\beta_{20} + \beta_{21}x_{it1} + \beta_{22}x_{it2})}, \end{aligned} \tag{4.10}$$

where $t = 1, 2, 3$ for $i = 1, 2, \ldots, 50$; $t = 1, 2, 3, 4$ for $i = 51, 52, \ldots, 100$; and $t = 1, 2, 3, 4, 5$ for $i = 101, 102, \ldots, 150$, x_{it1} is the centered time at tth measurement, $x_{it2} = x_{i2}$ is the centered time-invariant binary covariate generated from Bernoulli distribution with equal probability, $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \beta_{12}) = (0.9, 0.6, -0.6)$, and $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}) = (1.1, 0.7, -0.9)$. The conditional probabilities are specified as in equations 4.2-4.4. The expression for $\mathbf{T}_i$ and $\mathbf{D}_i$ when $n_i = 3, 4$, or 5 is given respectively, as follows:

$$\mathbf{T}_i = \begin{pmatrix} 1 & 0 & 0 \\ -\phi_{i,21} & 1 & 0 \\ 0 & -\phi_{i,32} & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{D}_i = \begin{pmatrix} \sigma_{i1}^2 & 0 & 0 \\ 0 & \sigma_{i2}^2 & 0 \\ 0 & 0 & \sigma_{i3}^2 \end{pmatrix},$$

$$\mathbf{T}_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\phi_{i,21} & 1 & 0 & 0 \\ 0 & -\phi_{i,32} & 1 & 0 \\ 0 & 0 & -\phi_{i,43} & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{D}_i = \begin{pmatrix} \sigma_{i1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{i2}^2 & 0 & 0 \\ 0 & 0 & \sigma_{i3}^2 & 0 \\ 0 & 0 & 0 & \sigma_{i4}^2 \end{pmatrix},$$

$$\mathbf{T}_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -\phi_{i,21} & 1 & 0 & 0 & 0 \\ 0 & -\phi_{i,32} & 1 & 0 & 0 \\ 0 & 0 & -\phi_{i,43} & 1 & 0 \\ 0 & 0 & 0 & -\phi_{i,54} & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{D}_i = \begin{pmatrix} \sigma_{i1}^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{i2}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{i3}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{i4}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{i5}^2 \end{pmatrix}.$$

Simulation Scenario 4: Misspecification of Time Lag Order in Random Effects Covariance Matrix

Under Scenario 4, like Scenario 1, while the number of time points is $n_i = n = 4$ and the spacing between two consecutive time points is assumed to be the same, the total number of subjects is $N = 150$ and number of covariates is $P = 2$.

The marginal probabilities are specified as follows:

$$\begin{aligned} Pr(Y_{it1}=1|\mathbf{x}_{it}) &= \frac{exp(\beta_{10}+\beta_{11}x_{it1}+\beta_{12}x_{it2})}{1+exp(\beta_{10}+\beta_{11}x_{it1}+\beta_{12}x_{it2})} \\ Pr(Y_{it2}=1|\mathbf{x}_{it}) &= \frac{exp(\beta_{20}+\beta_{21}x_{it1}+\beta_{22}x_{it2})}{1+exp(\beta_{20}+\beta_{21}x_{it1}+\beta_{22}x_{it2})}, \end{aligned} \tag{4.11}$$

where $i = 1, 2, \ldots, 150$, $t = 1, 2, 3, 4$, x_{it1} is the centered time at tth measurement, $x_{it2} = x_{i2}$ is the centered time-invariant binary covariate generated from Bernoulli distribution with equal probability, $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \beta_{12}) = (0.9, 0.6, -0.6)$, and $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}) = (1.1, 0.7, -0.9)$.

The conditional probabilities are specified as in equations 4.2 and 4.3. $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{\Sigma}_i)$ with $\mathbf{\Sigma}_{\boldsymbol{i}} = \mathbf{T}_i^{-1}\mathbf{D}_i(\mathbf{T}_i^{\top})^{-1}$. Furthermore, all lags, i.e., up to 3rd order, are assumed to be significant. Hence, $\mathbf{T}_i$ and $\mathbf{D}_i$ can be represented as follows:

$$\mathbf{T}_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\phi_{i,21} & 1 & 0 & 0 \\ -\phi_{i,31} & -\phi_{i,32} & 1 & 0 \\ -\phi_{i,41} & -\phi_{i,42} & -\phi_{i,43} & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{D}_i = \begin{pmatrix} \sigma_{i1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{i2}^2 & 0 & 0 \\ 0 & 0 & \sigma_{i3}^2 & 0 \\ 0 & 0 & 0 & \sigma_{i4}^2 \end{pmatrix}.$$

Then the parameters in $\mathbf{T}_i$ and $\mathbf{D}_i$ are associated with x_{it2} as follows:

$$\begin{aligned} \phi_{i,tl} &= \gamma_0 I_{(|t-l|\leq 3)} + \gamma_1 x_{it2} I_{(|t-l|\leq 3)} \quad \text{and} \\ log(\sigma_{it}^2) &= \lambda_0 + \lambda_1 x_{it2}, \end{aligned} \tag{4.12}$$

where $I_{(|t-l|\leq 3)}$ denotes that lags up to 3rd order are significant, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1) = (-0.3, -1.3)$, and $\boldsymbol{\lambda} = (\lambda_0, \lambda_1) = (0.5, 1.5)$. However, during the model fitting, the conditional model is assumed to be specified through equation 4.4:

$$\begin{aligned} \phi_{i,tl} &= \gamma_0 I_{(|t-l|=1)} + \gamma_1 x_{it2} I_{(|t-l|=1)} \quad \text{and} \\ log(\sigma_{it}^2) &= \lambda_0 + \lambda_1 x_{it2}, \end{aligned}$$

where $I_{(|t-l|=1)}$ denotes that lags up to 1st order are significant. By this scenario, we aim to study the sensitivity of model parameters to the misspecification of significant lags in random effects covariance matrix.

Simulation Scenario 5: Misspecification of Simple Regression Structure in Random Effects Covariance Matrix

Under Scenario 5, like Scenario 4, while the number of time points is $n_i = n = 4$ and the spacing between two consecutive time points is assumed to be the same, the total number of subjects is $N = 150$ and number of covariates is $P = 2$.

The marginal probabilities are specified through the equation 4.11 and the conditional probabilities are specified as in equations 4.2 and 4.3. $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{\Sigma}_i)$ with $\mathbf{\Sigma_{\boldsymbol{i}}} = \mathbf{T}_i^{-1}\mathbf{D}_i(\mathbf{T}_i^{\top})^{-1}$. The parameters in $\mathbf{T}_i$ and $\mathbf{D}_i$ are assumed to have a simple structure as follows:

$$\begin{aligned} \phi_{i,tl} &= \gamma_0 I_{(|t-l|=1)} \quad \text{and} \\ log(\sigma^2_{it}) &= \lambda_0, \end{aligned} \tag{4.13}$$

where $I_{(|t-l|=1)}$ denotes that lags up to 1st order are significant, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1) = (-0.3, 0)$, and $\boldsymbol{\lambda} = (\lambda_0, \lambda_1) = (0.5, 0)$. However, during the model fitting, the conditional model is assumed to be specified through equation 4.4:

$$\begin{aligned} \phi_{i,tl} &= \gamma_0 I_{(|t-l|=1)} + \gamma_1 x_{it2} I_{(|t-l|=1)} \quad \text{and} \\ log(\sigma^2_{it}) &= \lambda_0 + \lambda_1 x_{it2}. \end{aligned}$$

where $I_{(|t-l|=1)}$ denotes that lags up to 1st order are significant.

Simulation Scenario 6: Misspecification of Complex Regression Structure in Random Effects Covariance Matrix

Under Scenario 6, like Scenarios 4 and 5, while the number of time points is $n_i = n = 4$ and the spacing between two consecutive time points is assumed to be the same, the total number of subjects is $N = 150$ and number of covariates is $P = 2$.

The marginal probabilities are specified through the equation 4.11 and the conditional probabilities are specified as in equations 4.2 and 4.3. $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{\Sigma}_i)$ with $\mathbf{\Sigma_{\boldsymbol{i}}} =$

$\mathbf{T}_i^{-1}\mathbf{D}_i(\mathbf{T}_i^{\top})^{-1}$. The parameters in $\mathbf{T}_i$ and $\mathbf{D}_i$ are assumed to have a simple structure as follows:

$$\begin{aligned} \phi_{i,tl} &= \gamma_0 I_{(|t-l|=1)} + \gamma_1 x_{it2} I_{(|t-l|=1)} \quad \text{and} \\ log(\sigma_{it}^2) &= \lambda_0 + \lambda_1 x_{it2}, \end{aligned} \tag{4.14}$$

where $I_{(|t-l|=1)}$ denotes that lags up to 1st order are significant, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1) = (-0.3, -1.3)$, and $\boldsymbol{\lambda} = (\lambda_0, \lambda_1) = (0.5, 1.5)$. However, during the model fitting, the conditional model is assumed to be specified through equation as follows:

$$\begin{aligned} \phi_{i,tl} &= \gamma_0 I_{(|t-l|=1)} \quad \text{and} \\ log(\sigma_{it}^2) &= \lambda_0. \end{aligned}$$

Simulation Scenario 7: Revisit of equally spaced $\&$ equal number of time points with n=8

The first sub-scenario in Scenario 1 is revisited under Scenario 7 such that while the number of time points is $n_i = n = 8$ and the spacing between two consecutive time points is assumed to be the same, the total number of subjects is $N = 150$ and number of covariates is $P = 2$.

The marginal probabilities are specified as follows:

$$\begin{aligned} Pr(Y_{it1} = 1|\mathbf{x}_{it}) &= \frac{exp(\beta_{10} + \beta_{11}x_{it1} + \beta_{12}x_{it2})}{1 + exp(\beta_{10} + \beta_{11}x_{it1} + \beta_{12}x_{it2})} \\ Pr(Y_{it2} = 1|\mathbf{x}_{it}) &= \frac{exp(\beta_{20} + \beta_{21}x_{it1} + \beta_{22}x_{it2})}{1 + exp(\beta_{20} + \beta_{21}x_{it1} + \beta_{22}x_{it2})}, \end{aligned} \tag{4.15}$$

where $i = 1, 2, \ldots, 150$, $t = 1, 2, \ldots, 8$, x_{it1} is the centered time at tth measurement, $x_{it2} = x_{i2}$ is the centered time-invariant binary covariate generated from Bernoulli distribution with equal probability, $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \beta_{12}) = (0.9, 0.6, -0.6)$, and $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}) = (1.1, 0.7, -0.9)$. The conditional probabilities are specified as in equations 4.2-4.4.

4.1.2 Simulation Settings

The initial values for $\boldsymbol{\beta}$ are drawn from treating the bivariate longitudinal binary data as if it consisted of two different univariate longitudinal binary data, and then each univariate longitudinal binary data is fitted through `gee` package in R (Carey, 2012). The initial values for $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are drawn randomly from Uniform distribution within a narrow interval. The convergence of the code is also tested through using different initial values. The prior distributions for each $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\lambda}$ are chosen as normal distribution with 0 mean and variance $1,000$ (Daniels and Zhao, 2003; Congdon, 2010; Lee, 2013). Based on several simulated data sets, the standardized largest eigenvalue of posterior covariance matrix λ^K is plotted against K^{-1}. The result for one simulated data is given in Figure 4.1 and the number of clones K is determined as 10. The number of iterations for convergence is fixed at $30,000$. After convergence, $5,000$ iterations are drawn from 2 different chains with a thinning value of 25. The simulation study is repeated $M = 100$ times. For illustration, the R code for model fitting and the prediction of random effects for Scenario 1 is given in Appendix C.

4.1.3 Performance Measures

Under Scenarios $1-3$ and 7, for each parameter in $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$, measures such as average, bias, Monte Carlo standard error (SE), average of the model based standard errors (ASE), mean squared error (MSE), and coverage probability (CP) are computed based on the formulas given in Burton et al. (2006) and Crespi et al. (2009). On the other hand, since asymptotic covariance matrix of parameters is affected by misspecification, it does not provide good variance estimators for parameter estimators. For that reason, under Scenarios $4-6$, for each parameter in $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$, measures such as average, bias, SE, and MSE are computed, but ASE and CP are avoided.

The average and bias of the parameter estimates are given by,

$$\begin{aligned} \text{Average} &= \bar{\hat{\theta}}_v = \sum_{m=1}^{M} \frac{\hat{\theta}_{m,v}}{M} \quad \text{and} \\ \text{Bias} &= \bar{\hat{\theta}}_v - \theta_v, \end{aligned} \tag{4.16}$$

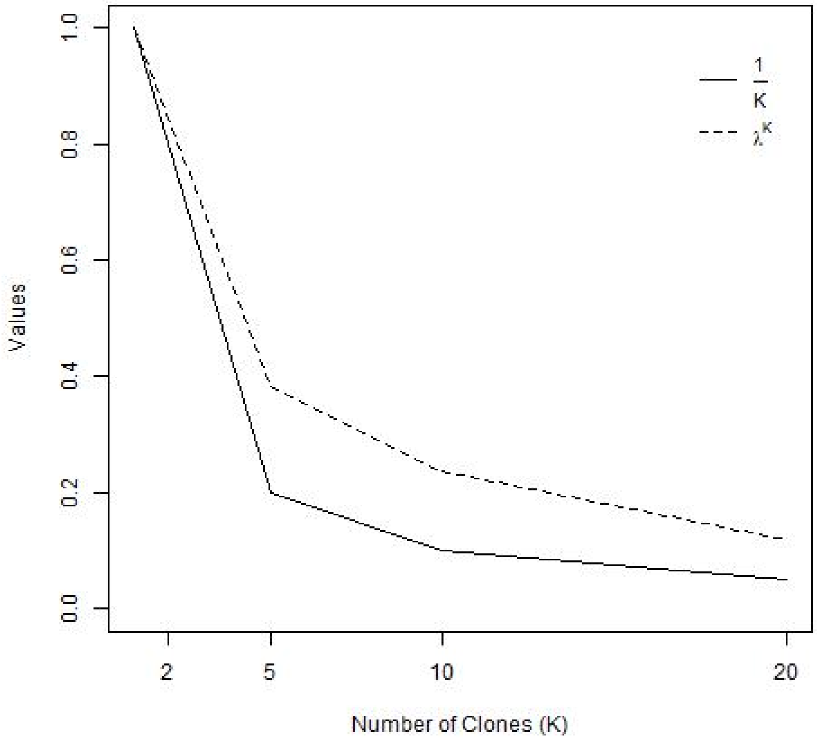

Figure 4.1: Determination of number of clones.

where $\hat{\theta}_{m,v}$ is the estimate of vth parameter in $\boldsymbol{\theta}$ at the mth simulation replication ($m = 1, 2, \ldots, M$).

Moreover, SE, ASE, and MSE of the parameter estimates are defined as follows:

$$\text{SE} = \sqrt{\sum_{m=1}^{M} \frac{(\hat{\theta}_{m,v} - \bar{\hat{\theta}}_v)^2}{M-1}},$$

$$\text{ASE} = \sum_{m=1}^{M} \frac{s.e.(\hat{\theta}_{m,v})}{M}, \quad \text{and}$$

$$\text{MSE} \;\; = (\bar{\hat{\theta}}_v - \theta_v)^2 + \textstyle\sum_{m=1}^{M} \frac{(\hat{\theta}_{m,v} - \bar{\hat{\theta}}_v)^2}{M-1}, \tag{4.17}$$

where $\bar{\hat{\theta}}_v$ is the average in equation 4.16, $s.e.(\hat{\theta}_{m,v})$ is the model based standard error of $\hat{\theta}_v$ at the mth simulation replication ($m = 1, 2, \ldots, M$) based on K times the sample covariance matrix of samples drawn from the posterior distribution in equation 2.13 in Section 2.2 on page 24.

At each m, a 95% confidence interval (CI) is computed as follows:

$$\hat{\theta}_{m,v} \pm 1.96 \times s.e.(\hat{\theta}_{m,v}), \tag{4.18}$$

where $s.e.(\hat{\theta}_{m,v})$ is the standard error of $\hat{\theta}_v$ at the mth simulation replication ($m = 1, 2, \ldots, M$) based on the Fisher information matrix and 1.96 is critical value corresponding to the $(1 - 0.05/2)$th quantile of standard normal distribution. Then a 95% coverage probability (CP) for θ_v is computed as the number of replications in which CI in equation 4.18 contains θ_v divided by M as given:

$$\text{CP} = \frac{\#((\hat{\theta}_{m,v} - 1.96 \times s.e.(\hat{\theta}_{m,v})) \leq \theta_v \leq (\hat{\theta}_{m,v} + 1.96 \times s.e.(\hat{\theta}_{m,v})))}{M}. \tag{4.19}$$

4.1.4 Simulation Results

Simulation results for Scenarios 1-7 are presented in Tables 4.1-4.10, respectively. Simulation results for the proposed model in equation 3.1 across the scenarios are promising.

Under Scenario 1, the estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\lambda}$ have generally quite small biases and moderate SEs, ASEs, and MSEs. For example, for the sub-scenario $N = 150$, $n = 4$, and $P = 2$, results in Table 4.1 are at an acceptable level for bias, SE, ASE, and MSE values. The bias values for $\hat{\boldsymbol{\beta}}$ are very low (e.g., ranging from 0.01 to 0.03). On the other hand, although the bias in $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\lambda}}$ are slightly larger than those in $\hat{\boldsymbol{\beta}}$, the bias values for $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\lambda}}$ are still low (e.g., ranging from -0.04 to -0.01). For the same setting, increasing N from 150 to 300 generally decreases the amount of bias in $\hat{\boldsymbol{\beta}}$ (except $\hat{\beta}_{22}$), $\hat{\boldsymbol{\gamma}}$, and $\hat{\boldsymbol{\lambda}}$, (except $\hat{\gamma}_1$) as can be seen in Table 4.2.

When $N = 150$, $n = 4$, and $P = 2$, the SEs and ASEs are also at an acceptable level for $\hat{\boldsymbol{\beta}}$ (e.g., SE and ASE are both ranging from 0.08 to 0.19). The SEs and ASEs for $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\lambda}}$ are larger compared to those for $\hat{\boldsymbol{\beta}}$, however, they are still at an acceptable level (e.g., SE and ASE are both ranging from 0.11 to 0.56). For all parameters, ASE values are very close to SE values, as ASE/SE ratio's are around 1. The SE and ASE values for all parameters decrease as N increases from 150 to 300. For example, while SE and ASE values for $\hat{\beta}_{12}$ are 0.19 and 0.18, respectively, for $N = 150$, those values are 0.13 and 0.13, respectively, for $N = 300$. On the other hand, ASE values of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\lambda}}$ tend to slightly underestimate SE values of those, as N increases. This situation reflects into the CP's, which are less than 95% nominal level. According to Tang et al. (2005), a CP falling within approximately two SEs of the nominal coverage probability(p) $SE(p) = \sqrt{\frac{(p(1-p))}{M}}$ is considered as acceptable. In this sense, for a 95% confidence interval, which is based on $M = 100$ simulations, a CP lying within 0.91 and 0.99 is considered as acceptable. On the other hand, the decrease in MSEs for all parameters can also be seen. For instance, while MSE value for $\hat{\beta}_{12}$ is 0.007 for $N = 150$, it is 0.004 for $N = 300$. Further comparison is available through Tables 4.1 and 4.2.

Increasing the number of covariates P from 2 to 4 does not cause any considerable change in the results. The estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\lambda}$ have still quite small biases and moderate SEs, ASEs, and MSEs as given in Table 4.3 for the sub-scenario $N = 150$, $n = 4$, and $P = 4$ and in Table 4.4 for the sub-scenario $N = 300$, $n = 4$, and $P = 4$. In fact, the results of the sub-scenarios $N = 150$, $n = 4$, and $P = 4$ and $N = 300$, $n = 4$, and $P = 4$ are very similar with those of the sub-scenarios $N = 150$, $n = 4$, and $P = 2$ and $N = 300$, $n = 4$, and $P = 2$ for the common parameters, (i.e., $\beta_{10}, \beta_{11}, \beta_{12}$, $\beta_{20}, \beta_{21}, \beta_{22}$, λ_0, $\lambda_{1,}$, γ_0, and γ_1), respectively. The only eye-catching point is that small differences in SEs result in noticeable differences in MSEs since MSEs are dominated by SEs due to small biases. To illustrate, while in sub-scenario $N = 150$, $n = 4$, and $P = 2$, $\hat{\gamma}_1$ is estimated as -1.31 with a downward bias of -0.01, a standard error of 0.20, and a MSE of 0.040, in sub-scenario $N = 150$, $n = 4$, and $P = 4$, it is estimated as -1.32 with a bias of -0.02, a standard error of 0.23, and a MSE of 0.053. Another example is that while in sub-scenario $N = 300$, $n = 4$, and $P = 2$, $\hat{\beta}_{22}$ is estimated as -0.93 with a downward bias of -0.03, a standard error of 0.15, and a MSE of 0.023, in sub-scenario $N = 300$, $n = 4$, and $P = 4$, it is estimated as -0.90 with zero bias, a standard error of 0.13, and a MSE of 0.017.

Our simulation results are also consistent with Pan and Mackenzie (2007) and Li and Pourahmadi (2013), who used the modified Cholesky decomposition method within univariate longitudinal continuous data. The $\boldsymbol{\gamma}$, which are the regression coefficients corresponding to the dependence term $\phi_{i,tl}$ are estimated with small MSEs compared to $\boldsymbol{\lambda}$, which are the regression coefficients corresponding to the variance term $log(\sigma_{it}^2)$. For instance, for sub-scenario $N = 150$, $n = 4$, and $P = 2$ under Scenario 1 (Table 4.1), MSE values of $\hat{\gamma}$ range from 0.014 to 0.04, whereas those values for $\hat{\boldsymbol{\lambda}}$ range from 0.079 to 0.314. This may be resulted from the restriction in the parameter space of variance terms compared to that of covariance parameters.

In the proposed model (equation 3.1), the number of measurements per b_{it} is 2 since information for b_{it} comes from two different responses Y_{it1} and Y_{it2}. In our opinion, the performance of the parameter estimates of $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ can also be improved through increasing the number of measurements per b_{it}, which is possible through extending the proposed model to trivariate or higher order longitudinal data. This argument may also be justified from a different point of view. For example, the model of Iddi and Molenberghs (2012) included a time-invariant random intercept b_i and the number of measurements per b_i was naturally $2 \times n$, since information for b_i came from 2 different responses, where each response was measured at n different time points. As the number of time points increased, which means the number of measurements per b_i increased, the biases and relative precision estimates related to the variance of b_i decreased, which can be seen in the simulation results. Similar phenomenon, which had parallel results to Iddi and Molenberghs (2012), can also be seen in their simulation results of Austin (2010) and Capanu et al. (2013).

Under Scenario 2, a similar case of sub-scenario under Scenario 1 is discussed. Scenario 2 also assumes $N = 150$, $n = 4$, and $P = 2$, but it further assumes that spacing between two consecutive time points are unequal. The estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\lambda}$ have generally quite small biases and moderate SEs, ASEs, and MSEs as presented in Table 4.5. The bias values for $\hat{\boldsymbol{\beta}}$, $\hat{\gamma}$, and $\hat{\boldsymbol{\lambda}}$ range from -0.03 to 0.03, -0.03 to 0.03, and -0.05 to -0.02, respectively. The SEs, ASEs, and MSEs are also at an acceptable level for $\hat{\boldsymbol{\beta}}$, $\hat{\gamma}$, and $\hat{\boldsymbol{\lambda}}$ (except $\hat{\lambda}_1$). On the other hand, since the time variable enters into the model as covariate x_{it2}, the results of Scenario 2 and those of sub-scenario under Scenario 1 cannot be comparable. Nonetheless, the results are in accordance with each other.

Under Scenario 3, N is 150, but it assumes that there are three different number of time points across the subjects, which are $n_i = 3$ for 50 subjects, $n_i = 4$ for other 50 subjects, and $n_i = 5$ for remaining 50 subjects. The number of covariates P is 2 and spacing between two consecutive time points are equal. The estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\lambda}$ have generally quite small biases and moderate SEs, ASEs, and MSEs as presented in Table 4.6. The bias values for $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$, and $\hat{\boldsymbol{\lambda}}$ range from 0.00 to 0.01, -0.03 to 0.00, and 0.00 to 0.05, respectively. The SEs, ASEs, and MSEs are also at an acceptable level for $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$, and $\hat{\boldsymbol{\lambda}}$ (except $\hat{\lambda}_1$). Note that assuming unequal number of time points in the data (e.g., an unbalanced longitudinal data) does not increase the number of parameters to be estimated in the random effects covariance matrix since $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ parameters are not indexed by n_i. Accommodation of the proposed model with unequal number of time points can be considered as an advantage over marginal models since marginal models cannot adapt an unstructured type for covariance matrix in case of unbalanced longitudinal data, where the number of parameters to be estimated in the covariance matrix changes with n_i (Hardin and Hilbe, 2003).

Scenario 4 is very similar to Scenario 1 in terms of settings. Unlike Scenario 1, under Scenario 4, data generated is based on a random effects covariance matrix, where lags up to 3rd order are assumed to be significant. However, during the model fitting, it is assumed that lags up to 1st order are significant. The estimates of $\boldsymbol{\beta}$ have generally quite small biases, moderate SEs, and MSEs as presented in Table 4.7, which indicates that they are robust to the misspecification of time lag order in random effects covariance matrix. On the other hand, the estimates of $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ have considerable biases, SEs, and MSEs as presented in Table 4.7.

Scenario 5 is also very similar to Scenario 1 in terms of settings. Unlike Scenario 1, under Scenario 5, data generated is based on a simple regression structure for random effects covariance matrix. However, during the model fitting, it is assumed that random effects covariance matrix has a complex regression structure. The estimates of $\boldsymbol{\beta}$ have generally quite small biases, moderate SEs, and MSEs as presented in Table 4.8, which implies that that they are robust to misspecification of simple regression structure in random effects covariance matrix. On the other hand, the estimates of $\boldsymbol{\gamma}$ and

$\boldsymbol{\lambda}$ have small biases, and moderate SEs and MSEs as presented in Table 4.8. Specifically, γ_1 and λ_1 are estimated with values, which are close to 0 (e.g., $\hat{\gamma_1} = 0.01$ and $\hat{\lambda_1} = 0.02$).

Scenario 6 is the opposite of Scenario 5 such that under Scenario 6, data generated is based on a complex regression structure for random effects covariance matrix. However, during the model fitting, it is assumed that random effects covariance matrix has a simple regression structure. The estimates of $\boldsymbol{\beta}$ have generally quite small biases, moderate SEs, and MSEs as presented in Table 4.9, except $\hat{\beta}_{12}$ and $\hat{\beta}_{22}$. Although $\hat{\beta}_{12}$ and $\hat{\beta}_{22}$ are estimated with a bias of 0.16 and 0.20, respectively, their SEs are very close to the ones in other scenarios. For that reason, their MSEs are still at an acceptable level, which are 0.051 and 0.080, for $\hat{\beta}_{12}$ and $\hat{\beta}_{22}$, respectively. Note that the biases are observed in the regression coefficients of the covariate, which is associated with random effects covariance matrix, but is ignored in the model fitting. On the other hand, the estimates of γ_0 and λ_0 have considerable biases. In fact, bias in γ_0 and λ_0 are -0.20 and 0.21, respectively. Note that, while γ_0 and λ_0 are estimated with considerable biases, they are estimated with moderate SEs, which are 0.12 and 0.23, respectively. This results in moderate MSE values for γ_0 and λ_0. Scenario 6 also indicates that if one of the covariates is associated with the elements of random effects covariance matrix, omitting it from regression models for the parameters of random effects covariance matrix will result in considerable biases in the regression coefficient of that covariate in the model for mean response.

Under Scenario 7, the performance of the proposed model is assessed with $N = 150$, $n = 8$, and $N = 300$. However, in this scenario, convergence problems are experienced related to $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ in more than 50% of $M = 100$ simulation runs, while $\boldsymbol{\beta}$'s converge in all $M = 100$ simulated data. Since the variation between the measurements of a subject depends on the time through b_{it}, increase in the number of time points per subject results in a considerable heterogeneity between the measurements of the subject, and in turn, this might cause poor estimation problems related to $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$. Since the first level of the model is marginalized from the second level, the convergence of $\boldsymbol{\beta}$ is not affected from the non-convergence of $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$. In this sense, the simulation is repeated until 100 converged data set is obtained, where the results are presented in Table 4.10. The results of Scenario 7 are parallel to sub-scenario of

Scenario 1, where$N = 300$, $n = 4$, and $P = 2$, except $\hat{\lambda}_1$. Specifically, λ_1 is estimated with a slightly larger SE (naturally MSE) and ASE under Scenario 7 than the sub-scenario of Scenario 1, where$N = 300$, $n = 4$, and $P = 2$.

In our opinion, the convergence problem for the case of $n = 8$ can be avoided with assuming a simple structure for 4.4 such that

$$\begin{aligned} \phi_{i,tl} &= \gamma_0 \\ log(\sigma^2_{it}) &= \lambda_0, \end{aligned} \tag{4.20}$$

which estimates two parameters γ_0 and λ_0 in $\mathbf{\Sigma}_i$, instead of $\frac{(8\times 9)}{2} = 36$ number of parameters. Hence, it is obvious that the modified Cholesky decomposition still provides an advantage in the number of parameters to be estimated related to $\mathbf{\Sigma}_i$ with a loss in interpretation. On the other hand, the simulation results of other marginalized multilevel models with time-varying random intercepts are also investigated. Within univariate longitudinal data concept, where only one observation was assumed per b_{it}, during the simulations, Lee and Daniels (2008) assumed $N = 300$ and $n = 6$ with three parameters in $\mathbf{\Sigma}_i$, Lee et al. (2011b) assumed $N = 200, 300$ and $n = 8$ with four parameters in $\mathbf{\Sigma}_i$, and Lee et al. (2011a) assumed $N = 400$ and $n = 6$ with three parameters in $\mathbf{\Sigma}_i$. Within multivariate longitudinal data concept, where only one observation was assumed per b_{itj}, during the simulations, Lee et al. (2009) and Lee et al. (2013) assumed $N = 300$ and $n = 6$ with four and five parameters in $\mathbf{\Sigma}_i$, respectively. However, this review reveals that they did not investigate the effect of different N and as well as different n values during the simulation studies. Overall simulation results (except the ones in Lee et al. (2011a)) demonstrate that large biases can occur in the marginal mean parameters when the covariance matrix structure for random effects is misspecified in the presence of missing at random (MAR) missingness. However, they did not report any performance measure results such as bias, MSE, and CP related to the covariance matrix parameters, which leads us to be in gray zone.

Table 4.1: Simulation result for Scenario 1 when N=150, n=4, and P=2.

Parameters	True value	Average	Bias	SE*	ASE**	ASE/SE	MSE	CP
β_{10}	0.90	0.91	0.01	0.10	0.09	0.90	0.010	0.90
β_{11}	0.60	0.61	0.01	0.08	0.08	1.00	0.007	0.95
β_{12}	-0.60	-0.57	0.03	0.19	0.18	0.95	0.037	0.91
β_{20}	1.10	1.11	0.01	0.10	0.10	1.00	0.010	0.95
β_{21}	0.70	0.71	0.01	0.08	0.09	1.12	0.007	0.97
β_{22}	-0.90	-0.89	0.01	0.17	0.19	1.12	0.029	0.95
γ_0	-0.30	-0.34	-0.04	0.11	0.11	1.00	0.014	0.93
γ_1	-1.30	-1.31	-0.01	0.20	0.23	1.15	0.040	0.97
λ_0	0.50	0.47	-0.03	0.28	0.27	0.96	0.079	0.95
λ_1	1.50	1.48	-0.02	0.56	0.56	1.00	0.314	0.95

* Monte Carlo standard error, ** Average of the model based standard errors

Table 4.2: Simulation result for Scenario 1 when N=300, n=4, and P=2.

Parameters	True value	Average	Bias	SE*	ASE**	ASE/SE	MSE	CP
β_{10}	0.90	0.90	0.00	0.06	0.07	1.17	0.004	0.97
β_{11}	0.60	0.59	-0.01	0.06	0.06	1.00	0.004	0.91
β_{12}	-0.60	-0.61	-0.01	0.13	0.13	1.00	0.017	0.96
β_{20}	1.10	1.11	0.01	0.07	0.07	1.00	0.005	0.94
β_{21}	0.70	0.70	0.00	0.06	0.06	1.00	0.004	0.95
β_{22}	-0.90	-0.93	-0.03	0.15	0.13	0.87	0.023	0.91
γ_0	-0.30	-0.31	-0.01	0.09	0.08	0.89	0.008	0.94
γ_1	-1.30	-1.33	-0.03	0.17	0.16	0.94	0.030	0.94
λ_0	0.50	0.47	-0.03	0.22	0.19	0.86	0.049	0.93
λ_1	1.50	1.48	-0.02	0.46	0.38	0.83	0.212	0.89

* Monte Carlo standard error, ** Average of the model based standard errors

Table 4.3: Simulation result for Scenario 1 when N=150, n=4, and P=4.

Parameters	True value	Average	Bias	SE*	ASE**	ASE/SE	MSE	CP
β_{10}	0.90	0.90	0.00	0.10	0.09	0.90	0.010	0.97
β_{11}	0.60	0.62	0.02	0.08	0.08	1.00	0.007	0.96
β_{12}	-0.60	-0.61	-0.01	0.16	0.18	1.12	0.026	1.00
β_{13}	0.50	0.49	-0.01	0.16	0.15	0.94	0.026	0.95
β_{14}	-0.30	-0.29	0.01	0.16	0.17	1.06	0.026	0.95
β_{20}	1.10	1.12	0.02	0.10	0.10	1.00	0.010	0.96
β_{21}	0.70	0.71	0.01	0.08	0.09	1.12	0.007	0.97
β_{22}	-0.90	-0.90	0.00	0.17	0.19	1.12	0.029	0.98
β_{23}	0.40	0.40	0.00	0.13	0.15	1.15	0.017	0.95
β_{24}	-0.40	-0.39	0.01	0.19	0.18	0.95	0.036	0.93
γ_0	-0.30	-0.30	0.00	0.11	0.11	1.00	0.012	0.97
γ_1	-1.30	-1.32	-0.02	0.23	0.23	1.00	0.053	0.95
λ_0	0.50	0.48	-0.02	0.31	0.28	0.90	0.096	0.92
λ_1	1.50	1.47	-0.03	0.56	0.57	1.02	0.315	0.93

* Monte Carlo standard error, ** Average of the model based standard errors

Table 4.4: Simulation result for Scenario 1 when N=300, n=4, and P=4.

Parameters	True value	Average	Bias	SE*	ASE**	ASE/SE	MSE	CP
β_{10}	0.90	0.91	0.01	0.08	0.07	0.88	0.007	0.92
β_{11}	0.60	0.61	0.01	0.06	0.06	1.00	0.004	0.93
β_{12}	-0.60	-0.61	-0.01	0.14	0.13	0.93	0.020	0.93
β_{13}	0.50	0.52	0.02	0.11	0.10	0.91	0.012	0.93
β_{14}	-0.30	-0.31	-0.01	0.13	0.12	0.92	0.017	0.92
β_{20}	1.10	1.11	0.01	0.07	0.07	1.00	0.005	0.94
β_{21}	0.70	0.70	0.00	0.05	0.06	1.20	0.003	0.96
β_{22}	-0.90	-0.90	0.00	0.13	0.14	1.08	0.017	0.96
β_{23}	0.40	0.40	0.00	0.11	0.11	1.00	0.012	0.97
β_{24}	-0.40	-0.43	-0.03	0.12	0.12	1.00	0.015	0.93
γ_0	-0.30	-0.30	0.00	0.09	0.08	0.89	0.008	0.93
γ_1	-1.30	-1.30	0.00	0.17	0.16	0.94	0.029	0.94
λ_0	0.50	0.51	0.01	0.20	0.20	1.00	0.040	0.93
λ_1	1.50	1.51	0.01	0.44	0.40	0.91	0.194	0.92

* Monte Carlo standard error, ** Average of the model based standard errors

Table 4.5: Simulation result for Scenario 2.

Parameters	True value	Average	Bias	SE*	ASE**	ASE/SE	MSE	CP
β_{10}	0.90	0.92	0.02	0.09	0.11	1.22	0.008	0.95
β_{11}	0.60	0.61	0.01	0.06	0.06	1.00	0.004	0.96
β_{12}	-0.60	-0.62	-0.02	0.20	0.18	0.90	0.040	0.93
β_{20}	1.10	1.13	0.03	0.10	0.12	1.20	0.011	0.99
β_{21}	0.70	0.70	0.00	0.07	0.06	0.86	0.005	0.94
β_{22}	-0.90	-0.93	-0.03	0.23	0.19	0.83	0.054	0.91
γ_0	-0.30	-0.33	-0.03	0.09	0.12	1.33	0.009	1.00
γ_1	-1.30	-1.27	0.03	0.22	0.25	1.14	0.049	0.94
λ_0	0.50	0.45	-0.05	0.25	0.28	1.12	0.065	0.99
λ_1	1.50	1.48	-0.02	0.66	0.57	0.86	0.436	0.88

* Monte Carlo standard error, ** Average of the model based standard errors

Table 4.6: Simulation result for Scenario 3.

Parameters	True value	Average	Bias	SE*	ASE**	ASE/SE	MSE	CP
β_{10}	0.90	0.91	0.01	0.10	0.09	0.90	0.010	0.94
β_{11}	0.60	0.60	0.00	0.07	0.08	1.14	0.005	0.95
β_{12}	-0.60	-0.60	0.00	0.19	0.18	0.95	0.036	0.94
β_{20}	1.10	1.11	0.01	0.11	0.10	0.91	0.012	0.93
β_{21}	0.70	0.71	0.01	0.09	0.08	0.89	0.008	0.94
β_{22}	-0.90	-0.89	0.01	0.17	0.19	1.12	0.029	0.98
γ_0	-0.30	-0.30	0.00	0.10	0.11	1.10	0.010	0.98
γ_1	-1.30	-1.33	-0.03	0.24	0.23	0.96	0.058	0.93
λ_0	0.50	0.50	0.00	0.30	0.28	0.93	0.090	0.95
λ_1	1.50	1.55	0.05	0.56	0.55	0.98	0.316	0.96

* Monte Carlo standard error, ** Average of the model based standard errors

Table 4.7: Simulation result for Scenario 4.

Parameters	True value	Average	Bias	SE*	MSE
β_{10}	0.90	0.91	0.01	0.11	0.012
β_{11}	0.60	0.60	-0.00	0.08	0.006
β_{12}	-0.60	-0.63	-0.03	0.17	0.030
β_{20}	1.10	1.09	-0.01	0.08	0.007
β_{21}	0.70	0.71	0.01	0.07	0.005
β_{22}	-0.90	-0.91	-0.01	0.20	0.040
γ_0	-0.30	0.03	0.33	0.16	0.134
γ_1	-1.30	-1.07	0.23	0.34	0.169
λ_0	0.50	0.63	0.13	0.35	0.139
λ_1	1.50	1.97	0.47	0.70	0.711

* Monte Carlo standard error

Table 4.8: Simulation result for Scenario 5.

Parameters	True value	Average	Bias	SE*	MSE
β_{10}	0.90	0.92	0.02	0.09	0.008
β_{11}	0.60	0.60	0.00	0.10	0.010
β_{12}	-0.60	-0.63	-0.03	0.16	0.027
β_{20}	1.10	1.12	0.02	0.09	0.008
β_{21}	0.70	0.71	0.01	0.09	0.008
β_{22}	-0.90	-0.91	-0.01	0.19	0.036
γ_0	-0.30	-0.31	-0.01	0.09	0.008
γ_1	0.00	0.01	0.01	0.21	0.044
λ_0	0.50	0.52	0.02	0.22	0.049
λ_1	0.00	0.02	0.02	0.50	0.250

* Monte Carlo standard error

Table 4.9: Simulation result for Scenario 6.

Parameters	True value	Average	Bias	SE*	MSE
β_{10}	0.90	0.88	-0.02	0.09	0.008
β_{11}	0.60	0.61	0.01	0.08	0.007
β_{12}	-0.60	-0.44	0.16	0.16	0.051
β_{20}	1.10	1.07	-0.03	0.09	0.009
β_{21}	0.70	0.72	0.02	0.10	0.010
β_{22}	-0.90	-0.70	0.20	0.20	0.080
γ_0	-0.30	-0.50	-0.20	0.12	0.054
λ_0	0.50	0.71	0.21	0.23	0.097

* Monte Carlo standard error

Table 4.10: Simulation result for Scenario 7.

Parameters	True value	Average	Bias	SE*	ASE**	ASE/SE	MSE	CP
β_{10}	0.90	0.90	0.00	0.07	0.08	1.14	0.005	0.97
β_{11}	0.60	0.60	0.00	0.04	0.04	1.00	0.002	0.93
β_{12}	-0.60	-0.60	0.00	0.14	0.14	1.00	0.020	0.96
β_{20}	1.10	1.10	0.00	0.08	0.08	1.00	0.006	0.96
β_{21}	0.70	0.70	0.00	0.04	0.04	1.00	0.002	0.97
β_{22}	-0.90	-0.89	0.01	0.16	0.15	0.94	0.026	0.93
γ_0	-0.30	-0.29	0.01	0.08	0.06	0.75	0.007	0.91
γ_1	-1.30	-1.31	-0.01	0.14	0.13	0.93	0.020	0.94
λ_0	0.50	0.54	0.04	0.23	0.21	0.91	0.054	0.90
λ_1	1.50	1.50	0.00	0.53	0.40	0.75	0.281	0.82

* Monte Carlo standard error, ** Average of the model based standard errors

4.2 Mother's Stress and Children's Morbidity (MSCM) Study Data

In Mother's Stress and Children's Morbidity (MSCM) study, Alexander and Markowitz (1986) investigated the relationship between maternal employment and paediatric health care utilization due to considerable changes in social and demographic characteristics in the US since 1950. Increasing number of working mothers leaded to mainly examine whether mother's employment status had any influence on the child's cognitive and social development. A total of 167 mothers and their preschool children (ages of between 18 months and 5 years) were enrolled in the MSCM study. At the beginning of the study, mothers were asked to provide demographic and domestic information about them such as education level, employment and marriage status, children's gender and race, maternal and child's health status at baseline and the household size, which are all categorical and time-invariant variables. Afterwards, the mothers were asked to record their maternal stress and child's illness status, whether present or not, in a health diary over a 28-day follow-up period. Information on these variables along with two binary responses, namely, mother's stress status and child's illness status, are listed in Table 4.11.

Table 4.11: Variable information list for MSCM study.

Variable	Explanation
Stress	mother's stress status at day t: 0=absence, 1=presence
Illness	child's illness status at day t: 0=absence, 1=presence
Chlth	child's health status at baseline: 0=very poor/poor, 1=fair, 2=good, 3=very good
Csex	child's gender: 0=male, 1=female
Education	mother's education level: 0=high school or less, 1=high school graduate
Employed	mother's employment status: 0=unemployed, 1=employed
Housize	size of the household: 0=2-3 people, 1=more than 3 people
Married	marriage status of the mother: 0=other,1=married
Mhlth	mother's health status at baseline: 0=very poor/poor, 1=fair, 2=good, 3=very good
Race	child's race: 0=white, 1=non-white

4.2.1 Exploratory Analysis of MSCM Study Data

Previous analysis of MSCM study data shows that two longitudinal binary responses, namely, mother's stress status and child's illness status exhibit weak serial dependencies for the first 16 days (Asar, 2012). For that reason a portion of MSCM study data for the days from 17 to 28 is presented in `mmm` package in R (Asar and Ilk, 2014).

For MSCM study data, first of all, the prevalence of maternal stress and that of child illness are computed based on the days from 17 to 28, which are displayed in Figure 4.2. It is observed that the prevalence of maternal stress at the 17th day is 14%, but then it gradually declines to 8% at the 28th day. Similarly, the prevalence of child illness is 11% at the 17th day, after an up-down pattern, it has declined to 8% and 7% in the last two days. On the other hand, it is seen that 55% of the mothers have higher education, 33% of the mothers are employed, and 48% of the mothers are married. Furthermore, at the beginning of the study, 34% and 37% of the mothers reported their health status as fair and good, respectively. 43% of the children are female and 55% of the children are non-white. Similarly, at the beginning of the study, the percentage of the children, who report their health status as good and very good, are 49% and 32%, respectively. 66% of the families live in a house consists of more than 3 people. These summary statistics are given in Table 4.12.

The association between two longitudinal binary responses, namely, mother's stress status and child's illness status at days from 17 to 28 are further investigated through odds-ratio formula (Chen et al., 2012), given as follows:

$$OR_t = \frac{Pr(Y_{t1}=0, Y_{t2}=0)Pr(Y_{t1}=1, Y_{t2}=1)}{Pr(Y_{t1}=1, Y_{t2}=0)Pr(Y_{t1}=0, Y_{t2}=1)}, \tag{4.21}$$

where $Pr(Y_{t1}, Y_{t2})$ denotes joint probability of mother's stress and child's illness outcome at time t ($t = 17, 18, \ldots, 28$). Furthermore, $(1 - 0.05/(2 \times 12))\%$ Bonferroni confidence intervals are constructed for the natural logarithm of odds-ratio at each time t ($t = 17, 18, \ldots, 28$) and results are then back transformed and estimated odds-ratios with lower and upper confidence intervals are presented in Table 4.13. It is observed that only odds-ratios at days 17, 18, and 23 are insignificant, which implies

Table 4.12: Characteristics of some MSCM study covariates.

Variable		%
Chlth		
	0 = very poor/poor	5
	1 = fair	13
	2 = good	49
	3 = very good	32
Csex		
	0 = male	57
	1 = female	43
Education		
	0 = high school or less	45
	1 = high school graduate	55
Employed		
	0 = unemployed	77
	1 = employed	33
Housize		
	0 = 2-3 people	44
	1 = more than 3 people	66
Married		
	0 = no	52
	1 = yes	48
Mhlth		
	0 = very poor/poor	14
	1 = fair	34
	2 = good	37
	3 = very good	15
Race		
	0 = white	45
	1 = non-white	55

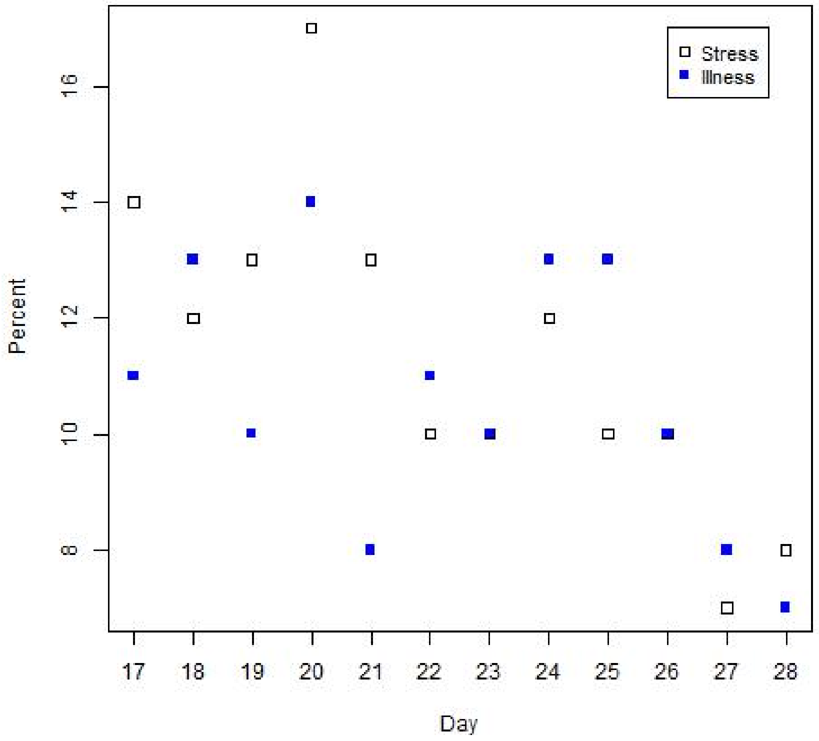

Figure 4.2: The prevalence of maternal stress and child illness in the MSCM study during the days of 17-28.

that the mother's maternal stress status is highly associated with the child's illness status from 17th day to 28th day. Hence, the MSCM study data can be considered as bivariate data and joint analysis of mother's stress status and child's illness status may help answer multiple questions simultaneously such as the effect of mother's education status or mother's employment status on both mother's stress status and child's illness status. It may also estimate the parameters with an improved efficiency due to the information exchange between responses.

Prior to analysing the MSCM study data with the proposed model in equation 3.1, marginal models are fitted to the data to investigate the relationship of mother's maternal status and child's illness status with the aforementioned covariates in Table 4.11. However, as both Figure 4.2 and Table 4.13 reveal, the prevalences of both responses are between 7% and 17%, which implies that MSCM study is a sparse data. For that reason, the data analysis is focused on the days $17, 20, 23$, and 26.

Table 4.13: The association between mother's stress status and child's illness status from days 17 to 28.

		Day 17		Day 18		Day 19		Day 20		Day 21		Day 22	
		Stress		Stress		Stress		Stress		Stress		Stress	
		0	1	0	1	0	1	0	1	0	1	0	1
Illness	0	128	21	128	17	135	16	125	19	135	18	136	12
	1	16	2	19	3	11	5	14	9	10	4	14	5
$\hat{OR}$		0.76 (0.50, 1.15)		1.19 (0.84, 1.69)		3.84 (2.80, 5.26)		4.23 (3.27, 5.48)		3.00 (2.14, 4.20)		4.05 (2.95, 5.55)	

		Day 23		Day 24		Day 25		Day 26		Day 27		Day 28	
		Stress		Stress		Stress		Stress		Stress		Stress	
		0	1	0	1	0	1	0	1	0	1	0	1
Illness	0	136	14	133	12	133	12	139	12	143	10	143	12
	1	15	2	14	8	17	5	12	4	12	2	10	2
$\hat{OR}$		1.30 (0.85, 1.97)		6.33 (4.78, 8.39)		3.26 (2.39, 4.44)		3.86 (2.74, 5.43)		2.38 (1.54, 3.68)		2.38 (1.54, 3.68)	

$\hat{OR}$: Estimated odds-ratio. Numbers within brackets denote the values of lower and upper confidence intervals

4.2.2 Data Analysis through Simpler Models

As mentioned in Chapter 1, marginal models directly specify a regression model for the mean response to assess the effects of covariates and treat longitudinal association parameters, which are specified through a working covariance matrix, as if they were nuisance parameters. In this sense, first of all, MSCM study data is treated as if it consisted of two different univariate longitudinal binary data and then univariate marginal models (UMMs) are fitted to each longitudinal binary data, as follows:

$$logit(Pr(Y_{it} = 1|\mathbf{X}_{it})) = \beta_0 + \beta_1 chlth_i + \cdots + \beta_8 race_i, \tag{4.22}$$

where Y_{it} is either stress status of mother or illness status of child for ith family ($i = 1, 2, \ldots, 167$) at time t ($t = 17, 20, 23, 26$), β_p's ($p = 0, \ldots, 8$) are corresponding regression coefficients of the covariates, and logit is the logarithm of odds. In UMMs, β_p's are estimated through generalized estimating equations (GEEs) (Liang and Zeger, 1986; Zeger and Liang, 1986). The association parameters related to the association between the repeated measurements of a response are generally estimated through the method of moments using Pearson residuals. The implementation of UMMs is possible through `gee` library in R (Carey, 2012) with a large menu for within-dependence structure such as AR(1), exchangeable, and unstructured and so on.

The extension of UMMs to jointly modelling of multivariate responses with response-specific coefficients is called as multivariate marginal models (MMMs) (Asar, 2012). The MMM, which is fitted to the bivariate longitudinal binary data, is given as follows:

$$logit(Pr(Y_{itj} = 1|\mathbf{X}_{it})) = \beta_0 + \beta_{j1} chlth_i + \cdots + \beta_{j8} race_i, \tag{4.23}$$

where Y_{itj} is jth response ($j = 1$, if response type = stress status of mother, and $j = 2$, if response type = illness status of child) of ith family ($i = 1, 2, \ldots, 167$) at time t ($t = 17, 20, 23, 26$), β_{jp}'s ($p = 0, \ldots, 8$) are the corresponding response-specific regression coefficients of the covariates, and logit is the logarithm of odds. In MMMs, β_{jp}'s are also estimated through GEEs and the working correlation matrix take all of the within, between and cross response dependencies into account. The implementation of MMMs is possible through `mmm` library in R (Asar and Ilk, 2014) with a large menu for covariance structure from independent to unstructured.

The results related to model fitting of UMMs and MMM with an exchangeable correlation structure to the MSCM study data are displayed in Table 4.14. The results of MMMs are consistent with those of UMMs in terms of the magnitude, direction, and standard error of the estimate. Both univariate and multivariate modelling results indicate that at 95% confidence level, mother's stress status is highly associated with child's health status at baseline (Z value = -2.82 and -2.78, respectively) and mother's employment status (Z value = -2.25 and -2.22, respectively) and that child's illness status is highly associated with child's health status at baseline (Z value = -2.85 and -2.83, respectively), size of the house (Z value = -2.87 and -2.84, respectively) and mother's marriage status (Z value = 2.55 and 2.57, respectively).

The results related to multivariate modelling show that although there is not enough evidence to say that educated or non-educated mothers are more likely to be stressed (Z value = 1.86) at 95% confidence level, being employed make mothers less stressful (Z value = -2.22), which may be related to the economic freedom in relation with the improved quality of life. Interestingly, being married increases both the probability of mother's being stressful and child's being ill ($\hat{\beta}_{17} = 0.08$ and $\hat{\beta}_{27} = 0.81$, respectively) and children, who live in a house with more than 3 people, are less likely to be ill ($\hat{\beta}_{26} = -0.87$).

Lastly, although not presented in Table 4.14, the longitudinal association parameter related to the exchangeable correlation structure ρ is estimated as 0.14 and 0.03 in UMM, in which response is the mother's stress status and child's illness status, respectively. It is estimated as 0.05 in MMM, which is apparently between two UMM estimates and may be due to the information exchange between the responses.

Table 4.14: Results of univariate and multivariate marginal model analysis for MSCM study data.

	Univariate Marginal Model Results			Multivariate Marginal Model Results		
Response=Stress	Estimate	SE	Z	Estimate	SE	Z
Intercept	-0.82	0.43	-1.92	-0.83	0.42	-1.94
chlth	-0.45	0.16	-2.82	-0.45	0.16	-2.78
csex	0.12	0.28	0.42	0.10	0.28	0.36
education	0.54	0.28	1.91	0.52	0.28	1.86
employed	-0.77	0.34	-2.25	-0.76	0.34	-2.22
housize	-0.40	0.30	-1.34	-0.41	0.30	-1.36
married	0.08	0.29	0.28	0.08	0.29	0.29
mhlth	-0.23	0.15	-1.57	-0.22	0.15	-1.51
race	0.29	0.28	1.05	0.29	0.28	1.05
Response=Ilness						
Intercept	-1.27	0.40	-3.17	-1.27	0.40	-3.17
chlth	-0.44	0.16	-2.85	-0.44	0.16	-2.83
csex	-0.01	0.28	-0.05	-0.01	0.28	-0.04
education	0.11	0.31	0.37	0.11	0.31	0.36
employed	-0.31	0.34	-0.92	-0.31	0.34	-0.94
housize	-0.87	0.30	-2.87	-0.86	0.30	-2.84
married	0.80	0.32	2.55	0.81	0.32	2.57
mhlth	-0.07	0.18	-0.39	-0.07	0.18	-0.40
race	0.49	0.31	1.60	0.48	0.31	1.57

4.2.3 Data Analysis through the Proposed Model

Preliminary analysis reveals that while mother's stress status is highly associated with child's health status at baseline and mother's employment status, child's illness status is highly associated with child's health status at baseline, mother's marriage status, and size of the house. Although preliminary analysis does not reveal any statistically significant effect of mother's education status on both responses, this covariate is also included into the model for the sake of curiosity. Hence, six different models, whose first level includes all these five covariates (e.g., child's health status at baseline, mother's education level, mother's employment status, house size and mother's marriage status) as given in equation 4.24, but whose second level is differed by modelling the elements of covariance matrix of random effects through one of these covariates as given in equation 4.25, are fitted to the data.

For each model, the marginal probabilities are specified as follows:

$$\begin{aligned} Pr(Y_{it1} = 1|\mathbf{X}_{it}) &= \frac{exp(\mathbf{X}_{it}\boldsymbol{\beta}_1)}{1 + exp(\mathbf{X}_{it}\boldsymbol{\beta}_1)} \\ Pr(Y_{it2} = 1|\mathbf{X}_{it}) &= \frac{exp(\mathbf{X}_{it}\boldsymbol{\beta}_2)}{1 + exp(\mathbf{X}_{it}\boldsymbol{\beta}_2)}, \end{aligned} \tag{4.24}$$

where Y_{it1} is stress status of ith mother ($i = 1, 2, 3, ..., 167$) at day t ($t = 17, 20, 23, 26$ or renumbered as $t = 1, 2, 3, 4$), Y_{it2} is illness status of ith child ($i = 1, 2, 3, ..., 167$) at day t ($t = 17, 20, 23, 26$ or renumbered as $t = 1, 2, 3, 4$), $\mathbf{X}_{it} = (1, chlth_i, education_i, employed_i, housesize_i, married_i)$, $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{15})$, and $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}, \beta_{25})$.

For each model, the conditional probabilities are specified as follows:

$$\begin{aligned} Pr(Y_{it1} = 1|\mathbf{X}_{it}, b_{it}) &= \Phi(\Delta_{it1}(\mathbf{X}_{it}) + b_{it}) \\ Pr(Y_{it2} = 1|\mathbf{X}_{it}, b_{it})) &= \Phi(\Delta_{it2}(\mathbf{X}_{it}) + b_{it}), \end{aligned} \tag{4.25}$$

where

$$\begin{aligned} \Delta_{it1} &= \Phi^{-1}\left(\frac{exp(\mathbf{X}_{it}\boldsymbol{\beta}_1)}{1 + exp(\mathbf{X}_{it}\boldsymbol{\beta}_1)}\right)\sqrt{1 + Var(b_{it})} \\ \Delta_{it2} &= \Phi^{-1}\left(\frac{exp(\mathbf{X}_{it}\boldsymbol{\beta}_2)}{1 + exp(\mathbf{X}_{it}\boldsymbol{\beta}_2)}\right)\sqrt{1 + Var(b_{it})}, \end{aligned} \tag{4.26}$$

and b_{it} is tth element of $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3}, b_{i4})$ and $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$, where $\mathbf{0}$ is a 4×1 vector of zeros and $\boldsymbol{\Sigma}_i$ is a 4×4 covariance matrix such that

$$\boldsymbol{\Sigma}_i = \mathbf{T}_i^{-1}\mathbf{D}_i(\mathbf{T}_i^{\top})^{-1}, \tag{4.27}$$

where $Var(b_{it})$ is the tth diagonal element of $\boldsymbol{\Sigma}_i$. At this point, $\phi_{i,tl}$ elements in $\mathbf{T}_i$ and σ_{it}^2 elements in $\mathbf{D}_i$ are modelled differently for each model as given in Table 4.15.

To illustrate, Model 1 does not allow the elements of random effects covariance matrix to differ by covariates, whereas Models 2, 3, 4, 5, and 6 allow the elements of random effects covariance matrix with first-order structure to differ by child's health status at baseline, education status of the mother, employment status of the mother, house size, and marriage status of the mother, respectively.

Table 4.15: Models for the elements of random effects covariance matrix.

Model 1	$\phi_{i,tl} = \gamma_0$	$log(\sigma^2_{it}) = \lambda_0$
Model 2	$\phi_{i,tl} = \gamma_0 + \gamma_1 I_{(\|t-l\|=1)} chlth_i$	$log(\sigma^2_{it}) = \lambda_0 + \lambda_1 chlth_i$
Model 3	$\phi_{i,tl} = \gamma_0 + \gamma_1 I_{(\|t-l\|=1)} education_i$	$log(\sigma^2_{it}) = \lambda_0 + \lambda_1 education_i$
Model 4	$\phi_{i,tl} = \gamma_0 + \gamma_1 I_{(\|t-l\|=1)} employed_i$	$log(\sigma^2_{it}) = \lambda_0 + \lambda_1 employed_i$
Model 5	$\phi_{i,tl} = \gamma_0 + \gamma_1 I_{(\|t-l\|=1)} housesize_i$	$log(\sigma^2_{it}) = \lambda_0 + \lambda_1 housesize_i$
Model 6	$\phi_{i,tl} = \gamma_0 + \gamma_1 I_{(\|t-l\|=1)} married_i$	$log(\sigma^2_{it}) = \lambda_0 + \lambda_1 married_i$

Data cloning (DC) algorithm, which is discussed in detail in Section 2.2, is used for parameter estimation. Six models are fitted by using `jags.fit` function under `dclone` package in R (Sólymos, 2010). The prior distributions for each $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\lambda}$ are chosen as normal distribution with 0 mean and variance $1,000$ (Daniels and Zhao, 2003; Congdon, 2010; Lee, 2013). The number of clones used is $K = 10$ and the average number of iterations for convergence is $30,000$. After convergence, $5,000$ iterations are drawn from 2 chains with a thinning value of 25. Results are presented in Table 4.16.

In Table 4.16, the first eye-catching point is that the results related to the marginal part of six models are the same in terms of both estimates and their standard errors. This is expected since it is well-known that parameter estimates and their standard errors in the marginal part of the marginalized multilevel models (MMMs) are robust to the specification for the covariance matrix of random effects as mentioned in Section 2.1. Although not presented here, the results (both estimates and their standard errors) in the marginal part of Models $1-6$ are also consistent with the result of the multivariate marginal model fitted through mmm package in R (Asar and Ilk, 2014).

Table 4.16: Maximum likelihood estimates of parameters (their standard errors) for Models $1-6$.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Regression parameters: $\boldsymbol{\beta}$						
Response=Stress						
Intercept	-0.69 (0.40)	-0.70 (0.37)	-0.68 (0.38)	-0.67 (0.39)	-0.70 (0.41)	-0.67 (0.38)
chlth	$-0.57(0.16)\dagger$	$-0.56(0.15)\dagger$	$-0.57(0.15)\dagger$	$-0.57(0.16)\dagger$	$-0.56(0.16)\dagger$	$-0.57(0.15)\dagger$
education	0.45 (0.28)	0.47 (0.30)	0.43 (0.30)	0.44 (0.31)	0.45 (0.30)	0.44 (0.28)
employed	$-0.81(0.33)\dagger$	$-0.78(0.33)\dagger$	$-0.82(0.34)\dagger$	$-0.81(0.33)\dagger$	$-0.80(0.32)\dagger$	$-0.82(0.35)\dagger$
housesize	-0.40 (0.30)	-0.40 (0.27)	-0.38 (0.29)	-0.39 (0.28)	-0.40 (0.28)	-0.39 (0.29)
married	0.16 (0.29)	0.15 (0.27)	0.15 (0.28)	0.15 (0.28)	0.16 (0.29)	0.14 (0.28)
Response=Illness						
Intercept	$-0.96(0.42)\dagger$	$-0.99(0.36)\dagger$	$-0.96(0.42)\dagger$	$-0.97(0.40)\dagger$	$-0.96(0.42)\dagger$	$-0.95(0.39)\dagger$
chlth	$-0.49(0.17)\dagger$	$-0.47(0.15)\dagger$	$-0.48(0.17)\dagger$	$-0.48(0.16)\dagger$	$-0.48(0.17)\dagger$	$-0.49(0.16)\dagger$
education	0.01 (0.30)	-0.00 (0.31)	-0.03 (0.30)	-0.01 (0.31)	-0.01 (0.33)	-0.01 (0.29)
employed	-0.37 (0.32)	-0.39 (0.33)	-0.38 (0.34)	-0.38 (0.34)	-0.36 (0.34)	-0.39 (0.32)
housesize	$-0.83(0.30)\dagger$	$-0.82(0.28)\dagger$	$-0.83(0.30)\dagger$	$-0.82(0.31)\dagger$	$-0.83(0.30)\dagger$	$-0.83(0.29)\dagger$
married	$0.89(0.30)\dagger$	$0.89(0.32)\dagger$	$0.88(0.29)\dagger$	$0.87(0.30)\dagger$	$0.88(0.31)\dagger$	$0.88(0.31)\dagger$

Continued on next page

Table 4.16 – *Continued from previous page*

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Generalized autoregressive parameters: $\boldsymbol{\gamma}$						
Intercept	1.00(0.20)†	0.21(0.37)	1.22(0.32)†	1.10(0.33)†	1.05(0.34)†	1.30(0.32)†
chlth		0.51(0.16)†				
education			−0.52(0.48)			
employed				−0.33(0.37)		
housesize					−0.18(0.45)	
married						−0.54(0.54)
Innovation parameters: $\boldsymbol{\lambda}$						
Intercept	−2.08(0.43)†	−0.26(0.76)	−2.58(0.65)†	−2.41(0.71)†	−2.26(0.76)†	−2.96(0.66)†
chlth		−1.15(0.43)†				
education			1.11(0.91)			
employed				1.12(0.85)		
housesize					0.55(0.92)	
married						1.60(1.30)

† Indicates that statistically significant at 95% confidence level

On the other hand, in Table 4.16, the results related to the elements of covariance matrix of random effects reveal that the covariates in the single covariate models, Models $3-6$ are not statistically significantly associated with either dependence or variance parameters at 95% confidence level, which all reduce to the only-intercept model, Model 1. To illustrate, the slope parameters γ_1 corresponding to single covariate for dependence parameter in Models $3-6$ are estimated (with a standard error) as $-0.52(0.48)$, $-0.33(0.37)$, $-0.18(0.45)$ and $-0.54(0.54)$, respectively, which all provide Z values less than 1.96 in absolute value. Similarly, the slope parameters λ_1 corresponding to the single covariate for variance parameter in Models $3-6$ are estimated (with a standard) error as $1.11(0.91), 1.12(0.85), 0.55(0.92)$ and $1.60(1.30)$, respectively, which all provide Z values less than 1.96 in absolute value. Furthermore, the intercept parameters γ_0 corresponding to model for dependence parameter (and its standard error) are estimated as $1.22(0.32)$, $1.10(0.33)$, $1.05(0.34), 1.30(0.32)$ in Models $3-6$, respectively, which all provide Z values greater than 1.96 in absolute value. Similarly, the intercept parameters λ_0 corresponding to model for dependence parameter (and its standard error) are estimated as $-2.58(0.65), -2.41(0.71), -2.26(0.76)$, $-2.96(0.66)$ in Models $3-6$, respectively, which all provide Z values greater than 1.96 in absolute value. These results imply that Models $3-6$ reduce to Model 1. At this point, another eye-catching point is that the estimates of γ_0 and λ_0 in Model 1 are very similar to the ones in Models $3-6$ (e.g., while $\hat{\gamma_0}$ in Model 1 is 1.00, they are $1.22, 1.10, 1.05, 1.30$ in Models $3-6$, respectively), except with a slight increase in standard errors (e.g., while standard error of $\hat{\gamma_0}$ in Model 1 is 0.20, they are $0.32, 0.33, 0.34, 0.32$ in Models $3-6$, respectively). This may imply that including additional parameters related to the covariance matrix of random effects may result in increase in standard errors due to the limited information in the data.

On the other side, Model 2 reveals that child's health status at baseline is significantly associated with dependence and variance parameters ($\hat{\gamma_1} = 0.51$, $se(\hat{\gamma_1}) = 0.16$, and Z value =3.19; $\hat{\lambda_1} = -1.15$, $se(\hat{\lambda_1}) = 0.43$, and Z value =$-2.67$). It is also worthy to say that γ_0 and λ_0 are estimated as 0.21 and -0.26, respectively, which are different than the ones in Model 1 and Models $3-6$. This difference may be a result of rich information exchange between the responses, since the child's health status is the only covariate, which is highly associated with both responses (see Table 4.16).

As mentioned in Section 2.2, one of the limitations of DC algorithm is that it does not provide a maximized value of the marginal likelihood in equation 3.8, which deprives us of performing likelihood ratio (LR) tests and doing selection between models through some information criteria such as Akaike information criteria (AIC). In this sense, overall prediction accuracy of Models $1-6$ are computed by pooling the bivariate and longitudinal data together through area under the ROC curve (AUROC) via `auc` package in R (Ballings and Poel, 2013). Table 4.17 presents overall AUROC values for Models $1-6$, which are based on estimated marginal and conditional probabilities. Specifically, marginal probabilities for Models $1-6$ are computed through substituting the relevant parameter estimates for that model in Table 4.16 into the equation 4.24. On the other hand, random effects for Models $1-6$ are predicted by substituting the relevant parameter estimates for that model in Table 4.16 into the equation 3.9. Afterwards, conditional probabilities for Models $1-6$ are computed through substituting the relevant parameter estimates for that model in Table 4.16 and predicted random effects into the equation 4.25.

Table 4.17: Overall AUROC values for Models $1-6$.

Model	AUROC value based on marginal probabilities	AUROC value based on conditional probabilities
Model 1	0.666	0.875
Model 2	0.666	0.949
Model 3	0.667	0.888
Model 4	0.667	0.881
Model 5	0.668	0.887
Model 6	0.666	0.879

In Table 4.17, AUROC values for Models $1-6$ based on marginal probabilities are naturally the same (e.g., 0.666) since the estimates of regression parameters in the marginal part of six models are nearly the same (see Table 4.16). On the other hand, AUROC values for Models $1-6$ based on conditional probabilities are considerably larger than those based on marginal probabilities. For example, while AUROC value for Model 1 based on marginal probabilities is estimated as 0.666, that based on conditional probabilities is estimated as 0.875. In this sense, a general conclusion is that the conditional models have much better prediction accuracies to represent

MSCM data compared to the marginal models. Furthermore, AUROC value of Model 2 based on conditional probabilities has the largest value, which is 0.949, among six models.

In addition to AUROC values, ratio of true prediction results, namely, ratio of total number of true positives (TPs) and true negatives (TNs) to total number of observations is computed for Models $1-6$ based on conditional probabilities. This is also named as percentage of correct classification (PCC) rate, which is formulated as follows:

$$PCC = \frac{TP + TN}{N} \times 100, \tag{4.28}$$

where TP is the total number of mothers and children, who are actually stressed and ill, respectively, and are also predicted as stressed and ill by the model, respectively. TN is the total number of mothers and children, who are actually not stressed and not ill, respectively and are also predicted as not stressed and not ill by the model, respectively. In equation 4.28, $N = 167 \times 4 \times 2 = 1336$ since the data is pooled as mentioned above. Results in Table 4.18 also indicate that Model 2 has the largest PCC rate, which is 80.2%.

Table 4.18: Overall percentage of correct classification (PCC) rate values for Models $1-6$ based on conditional probabilities

Model	PCC Rate $\%$
Model 1	73.5
Model 2	80.2
Model 3	73.8
Model 4	73.0
Model 5	73.7
Model 6	73.8

Table 4.19: AUROC values of Model 2 based on the conditional probabilities for each response across different days.

Response	Day=17	Day=20	Day=23	Day=26
Stress	0.957	0.961	0.945	0.975
Illness	0.901	0.928	0.959	0.961

In this sense, based on the results in Tables 4.17 and 4.18, the prediction accuracy of Model 2 based on conditional probabilities is further investigated through for each response across the days. Results, which are presented in Table 4.19, are also promising that Model 2 is successful at predicting the binary responses for both response types across the days. Specifically, AUROC values range from 0.901 to 0.975.

Based on the Wald-tests, it has been already mentioned that Models $3-6$ reduce to Model 1. Between Model 1 and Model 2, attention for further analysis is focused on Model 2, since it provides higher prediction accuracies and it also provides explanations related to the parameters of random effects covariance matrix with two additional parameters compared to Model 1 (Model 1 is a nested model of Model 2).

Within the column named Model 2 in Table 4.16, the first two upper rows give the regression parameter estimates and their standard errors for the first level of Model 2 for each response, i.e., mother's stress status and child's illness status, respectively. The estimated first level of Model 2 in terms of *logit* scale is given for mother's stress status and child's illness status, as follows:

$$\begin{aligned} logit(Pr(Y_{it1}=1)) &= -0.70 - 0.56 chlth_i + 0.47 education_i - 0.78 employed_i \\ &\quad - 0.40 housesize_i + 0.15 married_i \\ logit(Pr(Y_{it2}=1)) &= -0.99 - 0.47 chlth_i - 0.39 employed_i - 0.82 housesize_i \\ &\quad + 0.89 married_i . \end{aligned} \tag{4.29}$$

Population-averaged results based on the first level of Model 2 for mother's stress status reveal that mother's stress status is highly associated with child's health status at baseline ($\hat{\beta}_{11} = -0.56$, $se(\hat{\beta}_{11}) = 0.15$, and Z value $=-3.73$) and mother's employment status ($\hat{\beta}_{14} = -0.78$, $se(\hat{\beta}_{14}) = 0.33$, and Z value $=-2.36$) at 95% confidence level. Results indicate that the probability of mother's being stressed decreases as child's health status at baseline moves from poor to very good and that the odds of being stressed is $exp(0.78) = 2.18$ times higher for unemployed mothers than employed mothers.

Population-averaged results based on the first level of Model 2 for child's illness status also reveal that child's illness status is highly associated with child's health status at baseline ($\hat{\beta}_{21} = -0.47$, $se(\hat{\beta}_{21}) = 0.15$, and Z value $=-3.13$), size of the house ($\hat{\beta}_{24} = -0.82$, $se(\hat{\beta}_{24}) = 0.28$, and Z value $=-2.93$), and mother's marriage status ($\hat{\beta}_{25} = 0.89$, $se(\hat{\beta}_{25}) = 0.32$, and Z value=2.78) at 95% confidence level. Results indicate that the probability of child's being ill decreases as child's health status at baseline moves from poor to very good, that the odds of child's being ill is higher for house with size of $2-3$ people than house with size of more than 3 people, and that the odds of child's being ill is $exp(0.89) = 2.44$ times higher for married mothers than unmarried mothers.

The third and last set of rows within the column named Model 2 in Table 4.16 give the estimates of the regression parameters and their standard errors for the serial dependence and variance parameters in the second level of Model 2, respectively. The estimated serial dependence and variance parameters in the second level of Model 2 are given as follows:

$$\begin{aligned} \hat{\phi}_{i,tl} &= 0.21 I_{(|t-l|=1)} + 0.51 chlth_i I_{(|t-l|=1)} \\ log(\hat{\sigma^2_{it}}) &= -0.26 - 1.15 chlth_i. \end{aligned} \tag{4.30}$$

The estimate of slope regression coefficients related to the dependence and variance parameters are statistically significant at 95% confidence level ($\hat{\gamma}_1 = 0.51, se(\hat{\gamma}_1) = 0.16$, and Z value=3.19; $\hat{\lambda}_1 = -1.15, se(\hat{\lambda}_1) = 0.43$, and Z value=$-2.67$). This implies that the serial dependence between two consecutive random effect is positive and degree of dependence increases as child's health status at baseline moves from poor to very good ($\hat{\phi}_{tl} = (0.21 + (0.51 \times 0)) = 0.21$ for poor status, $\hat{\phi}_{tl} = (0.21 + (0.51 \times 1)) = 0.72$ for fair status, $\hat{\phi}_{tl} = (0.21 + (0.51 \times 2)) = 1.23$ for good status, and $\hat{\phi}_{tl} = (0.21 + (0.51 \times 3)) = 1.74$ for very good status, respectively, for |t-l|=1 and see Table 4.11 for more information on child's health status at baseline variable, e.g., it is an ordinal variable). A positively estimated γ_1 also indicates that if the child's health status at baseline is poor or fair, the mothers are less likely to report similar answers to their stress status across days. However, when the child's health status at baseline is good or very good, the mothers are more likely to report similar answers to their stress status across days. This also have a parallel interpretation in the predicting the child's illness status. If the child's health status at baseline is poor or fair, the children are more likely to report dissimilar answers to their illness status across days. However, if the child's health status at baseline is good or very good, then they report similar answers to their illness status across days. On the other hand, the variance decreases as the health status of a child moves from poor to very good $\hat{\sigma_{it}^2} = exp(-0.26-(1.15\times0)) = 0.77$ for poor status, $\hat{\sigma_{it}^2} = exp(-0.26-(1.15\times1)) = 0.25$ for fair status, $\hat{\sigma_{it}^2} = exp(-0.26 - (1.15 \times 2)) = 0.08$ for good status, and $\hat{\sigma_{it}^2} = exp(-0.26 - (1.15 \times 3)) = 0.02$ for very good status). A negatively estimated λ_1 also indicates that heterogeneity among the mothers, whose child's health status at baseline is good or less than good is high, whereas the heterogeneity is less among the mothers, whose child's health status at baseline is very good. An overall result is that the heterogeneity within the responses of a subject and the heterogeneity between the responses of subjects are ruled by the child's health status at baseline.

These arguments can be justified through plotting the marginal and conditional probabilities of interesting families. In this sense, among the families, whose child's health status at baseline is "poor", 2 families are selected based on the criteria that other covariates are also the same. As a consequence of this selection criteria, families with ID=142 and ID=150 are drawn. In these families, child's health status at baseline is poor, mothers are educated, employed, not married, and the size of the house is less than or equal to 3 people (see Table 4.11 for variable coding, e.g, child's poor health status at baseline=0, educated mother=1, employed mother=1, unmarried mother=0, small house size=0). On the other hand, among the families, whose child's health status at baseline is "very good", 2 families are selected based on the criteria that other covariates are also the same. As a consequence of this selection criteria, families with ID=7 and ID=14 are drawn. In these families, child's health status at baseline is very good, mothers are educated, employed, not married, and the size of the house is greater than 3 people (see Table 4.11 for variable coding, e.g, child's good health status at baseline=3, educated mother=1, employed mother=1, unmarried mother=0, large house size=1). Then, for these four families, marginal probabilities based on the first level of Model 2 and conditional probabilities based on the second level of Model 2 after predicting the random effects (equation 3.9) are computed for each response across days 17, 20, 23, and 26. The marginal and conditional probabilities for families with ID=142 and ID=150 are given in Figure 4.3, while those for families with ID=7 and ID=14 are presented in 4.4. To illustrate, marginal probability for each response (e.g., mother's stress status and child's illness status) for families with ID=142 and ID=150 (e.g., $i = 142$ and $i = 150$, respectively) based on the first level of Model 2 is computed as follows:

$$\begin{aligned}
\hat{Pr}(Y_{it1} = 1) &= \frac{exp(-0.70 - (0.56 \times 0) + (0.47 \times 1) - (0.78 \times 1) - (0.40 \times 0) + (0.15 \times 0))}{1 + exp(-0.70 - (0.56 \times 0) + (0.47 \times 1) - (0.78 \times 1) - (0.40 \times 0) + (0.15 \times 0))} \\
&= 0.27 \\
\hat{Pr}(Y_{it2} = 1) &= \frac{exp(-0.99 - (0.47 \times 0) - (0.00 \times 1) - (0.39 \times 1) - (0.82 \times 0) + (0.89 \times 0))}{1 + exp(-0.99 - (0.47 \times 0) - (0.00 \times 1) - (0.39 \times 1) - (0.82 \times 1) + (0.89 \times 0))} \\
&= 0.20
\end{aligned} \tag{4.31}$$

In Figures 4.3 and 4.4, the responses of mothers and children at days 17, 20, 23, and 26 are marked with 0 if the response is the absence of stress for the mother and the absence of illness for the child, and those are marked with 1 if the response is the presence of stress for the mother and the presence of illness for the child. From these figures, it is clear that conditional probabilities are much more successful in capturing the variation between and within the responses of the subjects. Moreover, these figures illustrate the larger variability among families with "poor" child health status at baseline (Figure 4.3) compared to families with "very good" child health status (Figure 4.4). Similarly, the variability increases across time within each family is higher for the families with "poor" child health status at baseline (Figure 4.3).

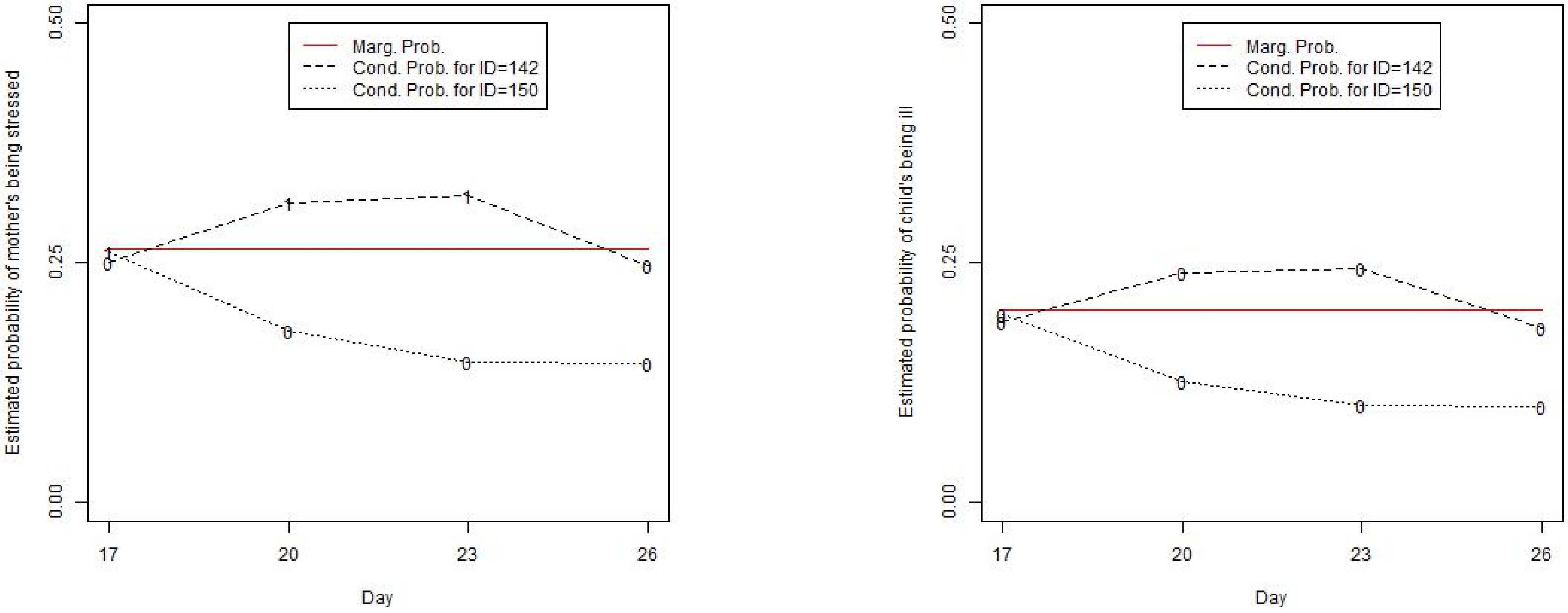

Figure 4.3: Plot of marginal and conditional probabilities of mother's being stressed and child's being ill for families with ID=142 and ID=150 at days 17, 20, 23, and 26.

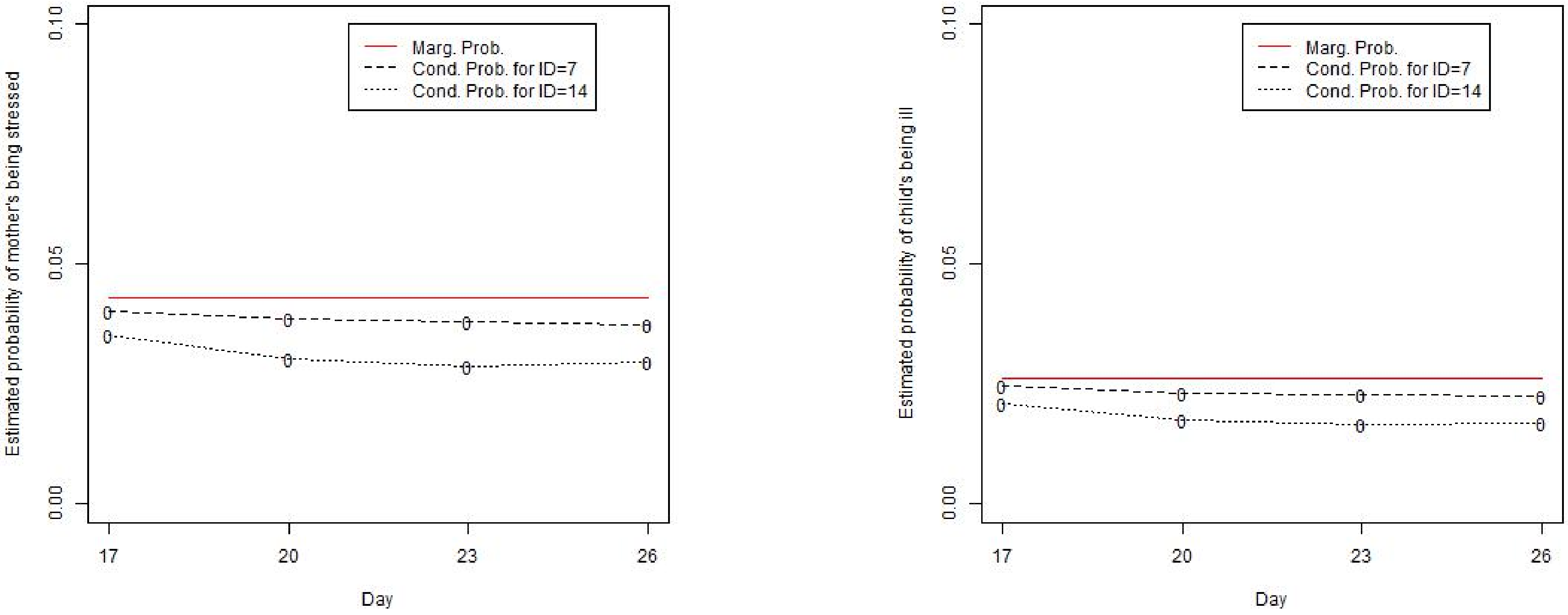

Figure 4.4: Plot of marginal and conditional probabilities of mother's being stressed and child's being ill for families with ID=7 and ID=14 at days 17, 20, 23, and 26.

CHAPTER 5

CONCLUSION AND FUTURE STUDIES

In a regression model developed to analyse multivariate longitudinal data, there exist three phenomenons to be embedded in the model structure. These building blocks are i) Marginal mean: the association between the mean of longitudinal responses and covariates, ii) Within-subject association: the association within the repeated measurements of a subject over time for a given response, and iii) Multivariate response association: the association between the multiple responses of a subject at a given time point. Drawing accurate inferences from the analysis of multivariate longitudinal data is possible when the model developed takes within-subject association as well as multivariate response association into account.

Within the regression model classes developed for the analysis of multivariate longitudinal data, marginalized multilevel models (MMMs) deserve attention since they build separate regression models for marginal mean of the longitudinal responses and for longitudinal associations to combine the strengths of marginal and conditional models, which are other two well-known regression model classes in the literature. Specifically, in MMMs, the model for the marginal mean, i.e., marginal mean model, builds the relationship between the longitudinal responses and covariates, and the model for longitudinal association(s), i.e., conditional mean model, restructures the mean of longitudinal responses conditional on either random effects and/or a Markov structure. As a consequence of this model building structure, MMMs take the population-averaged interpretation and robustness of marginal regression parameters from marginal models, while they take likelihood-based inference capabilities and flexible specifications for longitudinal associations from conditional models. Since the link function assumed for each level of MMM and the distributional assumption imposed on the random effects provide a rich class for MMMs, there exist a variety of MMMs for analysis of multivariate longitudinal binary data in the literature. In this sense, Ilk and Daniels (2007), Lee et al. (2009), Asar (2012), and Iddi and Molenberghs (2012) proposed different MMMs for multivariate longitudinal binary data, where each model had different model building structure (e.g., two-level or three-level, and logit or probit link function was used for all levels or different link functions were used for each level), represented longitudinal associations in a different point of view (e.g., either random effects or history of responses were included), used different parameter estimation algorithms (e.g., maximum likelihood estimation (MLE) or Bayesian inference was used) and implemented their models through different software (e.g., R, Fortran, or SAS). Further comparisons are available in Section 2.1.

In this thesis study, we propose a new marginalized multilevel model (MMM) for bivariate longitudinal binary data, which consists of two different levels, where each level is a different regression model. In the proposed model, the first level associates the marginal mean of responses with covariates through a logistic regression model to keep odds-ratio interpretation of the marginal regression parameters. The second level restructures the mean of responses conditional on subject/time specific random intercepts through a probit regression. For each subject, time-specific random intercepts, i.e., subject/time specific random intercepts, are included into the model to take into account the random-variation within the responses of the same subject as well as between the responses of different subjects. For each subject, time-specific random intercepts are assumed to be correlated, which results in multiple correlated random effects. Then multivariate normal distribution is imposed on each vector of random intercepts, i.e., multiple correlated random effects for each subject, so that the covariance matrix of vector of random effects are assumed to represent the within-subject association.

The reason for restructuring the conditional mean through probit link function is to utilize the good collaboration of probit link with normal distribution. This approach provides a closed-form solution for subject/time/response specific intercept terms, which connects the second level to the first level of the model. Hence, estimation of subject/time/response specific intercept term through numerical methods are avoided, which is a considerable gain in terms of computational burden. From this point of view, the proposed model can be also considered as a logistic-probit-normal marginalized multilevel model, where logistic and probit refer to the link functions used in the first and second levels of the model, respectively, and normal refers to the normal distribution assumed for random effects.

The literature review reveals that only a subject-specific random intercept is included into most of the models and multiple correlated random effects are usually avoided since multiple correlated random effects are accompanied by a high-dimensional random effects covariance matrix. MMMs with multiple correlated random effects, which take the likelihood based inference capabilities from generalized linear mixed models (GLMMs), may experience computational problems (e.g., numerical evaluation of high-dimensional integrals, numerical optimization of the marginal likelihood, which may include high-dimensional matrix inversion and differentiation computations) and statistical problems (e.g., inflation in the number of parameters to be estimated due to high-dimensionality of random effects covariance matrix increases complexity of the model, which may result in identifiability problems).

In this thesis study, to handle possible computational and statistical problems associated with high-dimensionality of the random effects covariance matrix, it is furthered decomposed into its dependence and variance components through modified Cholesky decomposition method. Then unconstrained version of resulting parameters (e.g., natural logarithm of variance parameters) are modelled in terms of covariates with low-dimensional regression parameters. Consequently, this approach allows i) flexibility for the structure of random effects covariance matrix (e.g., structure is not restricted to AR(1), exchangeable, etc.), ii) better explanations related to the elements of random effects covariance matrix, and iii) a reduction in the number of parameters to be estimated in random effects covariance matrix. The latter statement provides a considerable advantage for the implementation of the proposed model since there is a reduction in the number of parameters, which need to be jointly optimized. Furthermore, it avoids the proposed model from possible identifiability problems since the reduction in the number of parameters to be estimated are the parameters of random effects covariance matrix so that the complexity of the model does not exceed the information stored in the data. In this sense, to the best of our knowledge, this is the first time that the modified Cholesky decomposition method has been used within multivariate longitudinal binary data as well as MMMs framework to deal with computational and statistical problems associated with covariance matrix of multiple random effects.

In the proposed model, the multivariate response association are not represented through random effects. In fact, it is also shown that it internally exists in the model through a Taylor series-based approximation to the marginal correlation.

Like any GLMMs, in MMMs for multivariate longitudinal binary data with random effects, parameter estimation is based on the marginal likelihood of data, which involves the integration of conditional distribution of responses over the distribution of random effects. However, due to the non-conjugacy of conditional probability function (e.g., a probability density function for binary response) with the distribution of random effects (e.g., a multivariate normal distribution for continuous random variables), the integral is not tractable and, in turn, it does not provide a closed-form solution for the marginal likelihood. The dimension of integral, which is high, further complicates the computations. For that reason, there is a need for an approximation method or for a numerical method for high-dimensional integration. On the other hand, even if the marginal likelihood function is numerically evaluated, due to the probit link function used in conditional distribution of responses, the optimization of the marginal likelihood does not provide a closed-form solution for the parameters. For that reason, there is also a need for numerical optimization of the marginal likelihood, where the algorithm may involve computations such as taking the inverse of high-dimensional covariance matrix and differentiation of the covariance matrix with respect to unknown parameters.

In this thesis study, to avoid numerical evaluation of the high-dimensional integral, computation of high-dimensional covariance matrix inversion and differentiation of the covariance matrix with respect to unknown parameters, and numerical optimization of the marginal likelihood, we use data cloning (DC) computational algorithm (Lele et al., 2007; Lele, 2010) to compute the maximum likelihood estimates (MLEs) of the parameters in the proposed model and their standard errors. In this sense, to the best of our knowledge, this is the first time that DC computational algorithm has been used for multivariate longitudinal binary data as well as within MMMs framework for parameter estimation.

It is also shown that both population-averaged and subject-specific interpretations through estimating the random effects via Emprical Bayes (EB) estimation are possible within MMMs, which is an advantage over marginal models (e.g., provide only population-averaged interpretations) and conditional models (e.g., provide only subject-specific interpretations).

In this thesis study, simulation studies, which are carried out to assess the validity of the proposed model under different scenarios, show that the model performs well in longitudinal data with medium length series. The marginal regression parameters are estimated with ignorable bias and small mean squared errors (MSEs). The parameters related to the covariance matrix of random effects are estimated with larger biases and MSEs compared to the marginal regression parameters, but they are still at acceptable level. It is also seen that the regression parameters corresponding to the variance terms are estimated with larger MSEs compared to those corresponding to the dependence terms, which may be resulted from the restriction in the parameter space of variance terms compared to that of covariance parameters. Simulation results on misspecification of random effects covariance matrix show that while marginal regression are not generally affected by misspecification, the parameters related to the covariance matrix of random effects are estimated with considerable bias and moderate to considerable MSEs.

The extension of our proposed model to trivariate or higher order longitudinal data is very straightforward since the dimension of random effects covariance matrix depends only on the number of repeated measurements, not on the number of multiple responses. In fact, in our opinion, the performance of the parameter estimates related to the random effects covariance matrix would also be improved since information for random effects would increase as the number of responses increases.

In this thesis study, while modelling the dependence and variance parameters through covariates, we suggest parsimony in regression structures (e.g., using only one covariate). Otherwise, there would be identifiability and convergence problems, which can be identified through diagnostic measures in `dclone` package in R.

In case of convergence problems in longitudinal data with long length series, using a simple structure is suggested to model dependence and variance parameters, which still provides an advantage in the number of parameters to be estimated related to random effects covariance matrix with a loss in interpretation. At this point, we feel responsible to say that in most of studies, which have proposed new MMMs with subject/time or subject/time/response specific random intercepts, although performance measure results such as bias and MSE related to the marginal regression parameters have been presented in the simulation studies, any result related to the random effects covariance matrix parameters has not been presented, which lead us to be in gray zone.

In simulation studies, we also consider the scenarios with unequally spaced & equal number of time points, and equally spaced & unequal number of time points. However, we have not designed a simulation scenario, which included longitudinal data with unequally spaced & unequal number of time points. Assessing the performance of the proposed model under this scenario would be a good extension since the literature have not paid enough attention to this case. However, it requires substantial amount of change in the code, and hence it is left as a future work.

In simulation studies, the estimated odds-ratio between bivariate binary responses is high and the value of odds-ratio increases as time increases. Assessing the performance of the model when the odds-ratio between bivariate binary responses is 1 or around 1 would be a good future research study. Under this condition, it is expected that the estimates of marginal regression parameters would not be affected by week association of bivariate binary responses, whereas the estimates of the parameters related to the dependence terms in random effects covariance matrix would turn into insignificant, which leads estimates close to 0 in the off-diagonal entries of random effects covariance matrix.

To the best of our knowledge, the literature on reparametrization of covariance matrix and modelling the resulting variance and dependence parameters in terms of low-dimensional parameters are restricted to the covariates, which are a subset of the covariates that are associated with response means. However, this assumption can be relaxed when the data motivates that there may exist covariates, which are associated with variance and dependence parameters of the covariance matrix, but not associated with response means. Such a work would be a novel example for the literature on covariance matrix modelling.

In this thesis study, the proposed model is illustrated through Mother's Stress and Children's Morbidity (MSCM) Study data, which also provides interesting results. The data analysis shows that the estimates of marginal regression parameters are robust to the model specification for the second level, which is an advantage over conditional models. It also shows that a 4×4 random effects covariance matrix, which includes $\frac{4\times(4+1)}{2} = 10$ parameters, can also be modelled parsimoniously with 2 dependence and 2 covariance parameters. This provides better explanations related to the parameters of random effects covariance matrix (e.g., source of heterogeneity is explained in terms of child's health status at baseline) and avoided possible identifiability problems (e.g., it is highly possible that a model with 10 parameters related to the random effects covariance matrix would experience identifiability problems since the complexity of the model may exceed the information in the data). Along with population-based interpretations, subject-specific interpretations enable to identify behaviour characteristics of families. Since one of the main limitations of DC computational algorithm is that it does not provide a maximized value of the marginal likelihood, it deprives us of doing model selection through Akaike information criteria (AIC) and leads us to use accuracy measures such as area under the ROC curve (AUROC). We leave adapting an AIC type of criteria through DC computational algorithm as a future work since developing such an information criteria is an on-going work for several authors.

We hope this thesis study would be beneficial for statisticians, who work on the topics such as model developing for correlated binary data, random effects models, multiple correlated random effects, and covariance matrix decomposition. We also hope this thesis study would be beneficial for scientists in other fields such as medical doctors and sociologists, who collect longitudinal data and wish to do both population-based and subject-specific interpretations.

In the name of future studies, it is worthy to say that recent literature focuses on joint modelling of longitudinal data either with uniform responses (e.g., responses of the same type) or mixed responses through random effects, where several examples are given in Section 2.2. Within the marginalized multilevel models framework, Lee et al. (2013) proposed a marginalized multilevel model for bivariate longitudinal ordinal data with two levels with logit link. Subject/time/response specific random intercepts were included into the last level of the model to accommodate the serial correlation within responses of a subject over time and the correlation between different responses of a subject at a given time. They further reparametrized the correlation matrix in terms of lag-1 correlations and partial autocorrelations, and then they modelled Fisher's Z-transformation of these parameters in terms of covariates. Their model can be modified by assuming subject/time specific random intercepts with a probit link in the second level of their model to obtain a closed-form solution for Δ_{itj}'s and then the covariance matrix of the random effects can be further decomposed by modified Cholesky decomposition for a flexible correlation structure. Similarly, Efendi et al. (2013) proposed a marginalized multilevel model for bivariate longitudinal time-to-event data, where subject/response specific random intercepts were included into the model to induce the joint modelling of bivariate responses. During the implementation, however, they assumed only a subject-specific random intercept, which was shared by time and response. Their model also can be extended by assuming subject/time specific random intercepts, where the covariance matrix of the random effects is decomposed by modified Cholesky decomposition. On the other hand, Iddi and Molenberghs (2012) proposed a marginalized multilevel model for joint modelling of longitudinal continuous and binary outcomes, where subject/response specific random intercepts were included into the model to induce the joint modelling of responses. Following Iddi and Molenberghs (2012), Njagi et al. (2013) proposed a class of marginalized multilevel models for joint modelling of longitudinal time-to-event response with different type of longitudinal responses such as continuous, binary, and count responses. The mixed models in Iddi and Molenberghs (2012) and Njagi et al. (2013) can also be extended by assuming subject/time specific random intercepts, which may be scaled by a response-specific parameter as in Ilk and Daniels (2007) and Asar (2012), and the covariance matrix of the random effects can be decomposed by modified Cholesky decomposition. Recently, Lee et al. (2014) is the first work who has issued the non-ignorable missing longitudinal responses within the marginalized multilevel models for univariate longitudinal binary data through random effects. Hence, a natural extension of the proposed model in equation 3.1 and aforementioned models can be the adaptation of them to the missing responses and/or covariates.

REFERENCES

Akinc, D. (2008). *Statistical Modeling of Financial Statements of Turkey: A Panel Data Analysis.* M.Sc. Thesis. Department of Statistics, Middle East Technical University, Turkey.

Alexander, C.S. and Markowitz, R. (1986). "Maternal Employment and Use of Pediatric Clinic Services." *Medical Care* 24 (2), 134–147.

Asar, O. (2012). *On Multivariate Longitudinal Binary Data Models and Their Applications in Forecasting.* M.Sc. Thesis. Department of Statistics, Middle East Technical University, Turkey.

Asar, O. and Ilk, O. (2013a). *mmm2: Multivariate Marginal Models with Shared Regression Parameters. R package version 1.2.* URL: `http://cran.r-project.org/web/packages/mmm2` (visited on 04/08/2014).

— (2013b). *pnmtrem: Probit-Normal Marginalized Transition Random Effects Models. R package version 1.3.* URL: `http://cran.r-project.org/web/packages/pnmtrem` (visited on 04/08/2014).

— (2014). *mmm: Multivariate Marginal Models. R package version 1.4.* URL: `http://cran.r-project.org/web/packages/mmm` (visited on 04/08/2014).

Austin, P.C. (2010). "Estimating Multilevel Logistic Regression Models When The Number of Clusters is Low: A Comparison of Different Statistical Software Procedures." *International Journal of Biostatistics* 6 (1), Article 16.

Baghfalaki, T., Ganjali, M., and Berridge, D. (2014). "Joint Modeling of Multivariate Longitudinal Mixed Measurements and Time to Event Data Using a Bayesian Approach." *Journal of Applied Statistics* 41 (9), 1934–1955.

Baghishani, H. and Mohammadzadeh, M. (2011). "A Data Cloning Algorithm for Computing Maximum Likelihood Estimates in Spatial Generalized Linear Mixed Models." *Computational Statistics and Data Analysis* 55, 1748–1759.

Baghishani, H., Rue, H., and Mohammadzadeh, M. (2012). "On a Hybrid Data Cloning Method and Its Application in Generalized Linear Mixed Models." *Statistics and Computing* 22, 597–613.

Ballings, M. and Poel, D. Van den (2013). *AUC: Threshold independent performance measures for probabilistic classifiers. R package version 0.3.0.* URL: `http://cran.r-project.org/web/packages/AUC` (visited on 04/08/2014).

Bandyopadhyay, S., Ganguli, B., and Chatterjee, A. (2011). "A Review of Multivariate Longitudinal Data Analysis." *Statistical Methods in Medical Research* 20, 299–330.

Barnard, J., McCulloch, R., and Meng, X.L. (2000). "Modeling Covariance Matrices in terms of Standard Deviations and Correlations, with Application to Shrinkage." *Statistica Sinica* 10, 1281–1311.

Blocker, A.W. (2011). *fastGHQuad: Fast Rcpp implementation of Gauss-Hermite quadrature. R package version 0.1-1.0.* URL: `http://cran.r-project.org/web/packages/fastGHQuad` (visited on 04/08/2014).

Burton, A., Altman, D.G., Royston, P., and Holder, R.L. (2006). "The Design of Simulation Studies in Medical Statistics." *Statistics in Medicine* 25, 4279–4292.

Caffo, B. and Griswold, M. (2006). "A User-Friendly Tutorial on Link-Probit-Normal Models." *Journal of the American Statistical Association* 60, 139–145.

Caffo, B., Ming-Wen, A., and Rohde, C. (2007). "Flexible Random Intercept Models for Binary Outcomes Using Mixtures of Normals." *Computational Statistics and Data Analysis* 51, 5220–5235.

Capanu, M., Gonen, M., and Colin, B.B. (2013). "An Assessment of Estimation Methods for Generalized Linear Mixed Models with Binary Outcomes". *Statistics in Medicine* 32, 4550–4566.

Carey, V.J. (2012). *gee: Generalized Estimation Equation solver. R package version 4.13-18.* URL: `http://CRAN.R-project.org/package=gee` (visited on 04/08/2014).

Chan, J.S.K., Kuk, A.Y.C., and Bel, J. (1997). "A Likelihood Approach to Analysing Longitudinal Bivariate Binary Data." *Biometrical Journal* 39 (4), 409–421.

Chen, B., Yi, G.Y., Cook, R.J., and Zhou, X-H. (2012). "Marginal Methods for Clustered Longitudinal Binary Data with Incomplete Covariates." *Journal of Statistical Inference* 142 (10), 2819–2831.

Congdon, P. (2010). *Applied Bayesian Hierarchical Methods.* Boca Raton, FL: Chapman & Hall/CRC.

Crespi, C.M., Wong, W.K., and Mishra, S.I. (2009). "Using Second-Order Generalized Estimating Equations to Model Heterogeneous Intraclass Correlation in Cluster-Randomized Trials." *Statistics in Medicine* 28, 814–827.

Daniels, M.J. and Pourahmadi, M. (2002). "Bayesian Analysis of Covariance Matrices and Dynamic Models for Longitudinal Data." *Biometrika* 89, 553–566.

— (2009). "Modeling Covariance Matrices via Partial Autocorrelations." *Journal of Multivariate Analysis* 100 (10), 2352–2363.

Daniels, M.J. and Zhao, Y.D. (2003). "Modelling the Random Effects Covariance Matrix in Longitudinal Data." *Statistics in Medicine* 22, 1631–1647.

Das, K. and Daniels, M. (2014). "A Semiparametric Approach to Simultaneous Covariance Estimation for Bivariate Sparse Longitudinal Data." *Biometrics* 70, 33–43.

Das, K., Li, R., Sengupta, S., and Wu, R. (2013). "A Bayesian Semiparametric Model for Bivariate Sparse Longitudinal Data." *Statistics in Medicine* 32, 3899–3910.

Dey, T. and Lim, C.Y. (2013). "Comparisons of Computational Methods for Clustered Binary Data." *Journal of Statistical Computation and Simulation* 83 (11), 2030–2046.

Diggle, P.J., Heagerty, P., Liang, K.Y., and Zeger, S. (2002). *Analysis of Longitudinal Data.* Oxford University Press.

Din, S.H.M., Molas, M., Luime, J., and Lesaffre, E. (2014). "Longitudinal Profiles of Bounded Outcome Scores as Predictors for Disease Activity in Rheumatoid Arthritis Patients: A Joint Modeling Approach." *Journal of Applied Statistics* 41 (8), 1627–1644.

Efendi, A., Molenberghs, G., Njagi, E.N., and Dendale, P. (2013). "A Joint Model for Longitudinal and Time-to-Event Outcomes with Direct Marginal Interpretation." *Biometrical Journal* 55, 572–588.

Goldstein, H. and Rasbash, J. (1996). "Improved Approximations for Multilevel Models with Binary Responses." *Journal of the Royal Statistical Society Series B* 159, 505–513.

Griswold, M. (2005). *Complex Distributions, Hmmmm...Hierarchical Mixtures of Mar-*

ginalized Multilevel Models. PhD Thesis. Department of Biostatistics, Johns Hopkins University, United States of America.

Griswold, M.E. and Zeger, S.L. (2004). *On Marginalized Multilevel Models and their Computation.* Department of Biostatistics, Johns Hopkins University, United States of America, Working Papers.

Griswold, M.E., Swihart, B.J., Caffo, B.S., and Zeger, S.L. (2013). "Practical Marginalized Multilevel Models." *Stat* 2 (1), 129–142.

Gueorguieva, R. (2001). "A Multivariate Generalized Linear Mixed Model for Joint Modeling of Clustered Outcomes in the Exponential Family." *Statistical Modelling* 1, 177–193.

— (2013). "Analysis of Mixed Data. Methods & Applications." Chapman & Hall/CRC. Chap. Random Effects Models for Joint Analysis of Repeatedly Measured Discrete and Continuous Outcomes. 109–123.

Hardin, J.W. and Hilbe, J.M. (2003). *Generalized Estimating Equations.* London: Chapman & Hall/CRC Press.

Heagerty, P.J. (1999). "Marginally Specified Logistic-Normal Models for Longitudinal Binary Data." *Biometrics* 55, 688–698.

— (2002). "Marginalized Transition Models and Likelihood Inference for Longitudinal Categorical Data." *Biometrics* 58, 342–351.

Heagerty, P.J. and Zeger, S.L. (2000). "Marginalized Multilevel Models and Likelihood Inference (with discussion)." *Statistical Science* 15, 1–26.

Hedeker, D. and Gibbons, R.D. (2006). *Longitudinal Data Analysis.* New Jersey: John Wiley & Sons.

Iddi, S. and Molenberghs, G. (2012). "A Joint Marginalized Multilevel Model for Longitudinal Outcomes." *Journal of Applied Statistics* 39 (11), 2413–2430.

Ilk, O. and Daniels, M.J. (2007). "Marginalized Transition Random Effects Models for Multivariate Longitudinal Binary Data." *The Canadian Journal of Statistics* 35, 105–123.

Kalaylioglu, Z.I. and Ozturk, O. (2013). "Bayesian Semiparametric Models for Nonignorable Missing Mechanisms in Generalized Linear Models." *Journal of Applied Statistics* 40 (8), 1746–1763.

Karl, A.T., Yang, Y., and Lohr, S.L. (2014). "Computation of Maximum Likelihood Estimates for Multiresponse Generalized Linear Mixed Models with Non-Nested, Correlated Random Effects." *Computational Statistics and Data Analysis* 73, 146–162.

Kim, C. and Zimmerman, D.L. (2012). "Unconstrained Models for The Covariance Structure of Multivariate Longitudinal Data." *Journal of Multivariate Analysis* 107, 104–118.

Kim, Y., Choi, Y-K., and Emery, S. (2013). "Logistic Regression with Multiple Random Effects: A Simulation Study of Estimation Methods and Statistical Packages." *The American Statistician* 67 (3), 171–182.

Krantz, S.G. and Parks, H.R. (2003). *The Implicit Function Theorem: History, Theory and Applications.* Boston: Birkhauser.

Lee, K. (2013). "Bayesian Modeling of Random Effects Covariance Matrix for Generalized Linear Mixed Models." *Communications for Statistical Applications and Methods* 20 (3), 235–240.

Lee, K. and Daniels, M. (2008). "Marginalized Models for Longitudinal Ordinal Data with Application to Quality of Life Studies." *Statistics in Medicine* 27, 4359–4380.

Lee, K., Joo, Y., Yoo, J.K., and Lee, J.B. (2009). "Marginalized Random Effects Models for Multivariate Longitudinal Binary Data." *Statistics in Medicine* 28, 1284–1300.

Lee, K., Joo, Y., Song, J.J., and Harper, D.W.S (2011a). "Analysis of Zero-Inflated Clustered Count Data: A Marginalized Model Approach." *Computational Statistics and Data Analysis* 55, 824–837.

Lee, K., Kang, S., Liu, X., and Seo, K. (2011b). "Likelihood-Based Approach for Analysis of Longitudinal Nominal Data Using Marginalized Random Effects Models." *Journal of Applied Statistics* 38 (8), 1577–1590.

Lee, K., Lee, J., Hagan, J., and Yoo, J.K. (2012). "Modeling The Random Effects Covariance Matrix for Generalized Linear Mixed Models." *Computational Statistics and Data Analysis* 56, 1545–1551.

Lee, K., Daniels, M.J., and Joo, Y. (2013). "Flexible Marginalized Models for Bivariate Longitudinal Ordinal Data." *Biostatistics* 14 (3), 462–476.

Lee, M., Lee, K., and Lee, J. (2014). "Marginalized Transition Shared Random Effects Models for Longitudinal Binary Data with Nonignorable Dropout." *Biometrical Journal* 56 (2), 1–14.

Lele, S.R. (2010). "Model Complexity and Information in the Data: Could it be a house built on sand ?." *Ecology* 91, 3493–3496.

Lele, S.R., Dennis, B., and Lutscher, F. (2007). "Data Cloning: Easy Maximum Likelihood Estimation for Complex Ecological Models Using Bayesian Markov Chain Monte Carlo Methods." *Ecology Letters* 10, 551–563.

Lele, S.R., Nadeem, K., and Schmuland, B. (2010). "Estimability and Likelihood Inference for Generalized Linear Mixed Models Using Data Cloning." *Journal of the American Statistical Association* 105, 1617–1625.

Li, E. and Pourahmadi, M. (2013). "An Alternative REML Estimation of Covariance Matrices in Linear Mixed Models." *Statistics & Probability Letters* 83 (4), 1071–1077.

Liang, K.Y. and Zeger, S.L. (1986). "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73, 13–22.

Liang, K.Y., Zeger, S.L., and Qaqish, B. (1992). "Multivariate Regression Analyses for Categorical Data." *Journal of Royal Statistical Society, Series B* 54, 3–40.

Miglioretti, D.L. and Heagerty, P.J. (2004). "Marginal Modeling of Multilevel Binary Data with Time-Varying Covariates." *Biostatistics* 5, 381–398.

Njagi, E.N., Molenberghs, G., Rizopolous, D., Verbeke, G., Kenward, M.G., Dendale, P., and Willekens, K. (2013). "A Flexible Joint Modeling Framework for Longitudinal and Time-to-Event Data with Overdispersion." *Statistical Methods in Medical Research* 0 (0), 1–16.

Pan, J. and Mackenzie, G. (2007). "Modelling Conditional Covariance in the Linear Mixed Model." *Statistical Modelling* 7, 49–71.

Pan, J. and Thompson, R. (2007). "Quasi-Monte Carlo Estimation in Generalized Linear Mixed Models." *Computational Statistics and Data Analysis* 51, 5765–5775.

Pan, J., Nummi, T., and Liu, K. (2013). "Modeling of Mean-Covariance Structures in Generalized Estimating Equations with Dropouts." *Statistics and Its Interface* 6, 19–26.

Parzen, M., Ghosh, S., Lipsitz, S., Sinha, D., Fitzmaurice, G.M., Mallick, B.K., and Ibrahim, J.G. (2011). "A Generalized Linear Mixed Model for Longitudinal Binary Data with a Marginal Logit Link Function." *Annals of Applied Statistics* 5 (1), 449–467.

Paule, I., Girard, P., and Tod, M. (2011). “Empirical Bayes Estimation of Random Effects of a Mixed-Effects Proportional Odds Markov Model for Ordinal Data.” *Computer Methods and Programs in Biomedicine* 104 (3), 505–513.

Pike, F. and Weissfel, L. (2014). “Joint Modeling of Censored Longitudinal and Event Time Data.” *Journal of Applied Statistics* 40 (1), 17–27.

Plummer, Martyn (2003). *JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling*. URL: `http://mcmc-jags.sourceforge.net/` (visited on 04/08/2014).

Ponciano, J.M., Taper, M.L., Dennis, B., and Lele, S.R. (2009). “Hierarchical Models in Ecology: Confidence Intervals, Hypothesis Testing, and Model Selection Using Data Cloning”. *Ecology* 90 (2), 356–362.

Pourahmadi, M. (1999). “Joint Mean-Covariance Models with Applications to Longitudinal Data: Unconstrained Parameterization.” *Biometrika* 86, 677–690.

— (2000). “Maximum Likelihood Estimation of Generalized Linear Models for Multivariate Normal Covariance Matrix.” *Biometrika* 87, 425–435.

— (2011). “Covariance Estimation: The GLM and Regularization Perspectives.” *Statistical Science* 26 (3), 369–387.

Pourahmadi, M. and Daniels, M.J. (2002). “Dynamic Conditionally Linear Mixed Models for Longitudinal Data.” *Biometrics* 58, 225–231.

Pourahmadi, M., Daniels, M.J., and Park, T. (2007). “Simultaneous Modelling of the Cholesky Decomposition of Several Covariance Matrices.” *Journal of Multivariate Analysis* 98, 568–587.

Rizopoulos, D. and Lesaffre, E. (2014). “Joint modeling techniques.” *Statistical Methods in Medical Research* 23, 3–10.

Rodrigues-Motta, M., Pinheiro, H.P., Martins, E.G., Araujo, M.S., and Reis, S.F. dos (2013). “Multivariate Models for Correlated Count Data.” *Journal of Applied Statistics* 40 (7), 1586–1596.

Rudary, M.R. (2009). *On Predictive Linear Gaussian Models.* PhD Thesis. University of Michigan, United States of America.

Schildcrout, J.S. and Heagerty, P.J. (2007). “Marginalized Models for Moderate to Long Series of Longitudinal Binary Response Data.” *Biometrics* 63, 322–331.

Shelton, B.J., Gilbert, G.H., Liu, B., and Fisher, M. (2004). “A SAS Macro for the Analysis of Multivariate Longitudinal Binary Outcomes.” *Computer Methods and Programs in Biomedicine* 76, 163–175.

Sólymos, P. (2010). “dclone: Data Cloning in R.” *The R Journal* 2, 29–37.

Stan Development Team (2014). *Stan: A C++ Library for Probability and Sampling, Version 2.2.* URL: `http://mc-stan.org/` (visited on 04/08/2014).

Stewart, J. (2008). *Calculus: Early Transcendentals.* USA: Thomson Brooks/Cole.

Swihart, B.J., Caffo, B.S., and Crainiceanu, C.M. (2013). “A Unifying Framework for Marginalized Random-Intercept Models of Correlated Binary Outcomes.” *International Statistical Review*, 1–22.

Tang, L.Q., Song, J.W., Belin, T.R., and Unutzer, J. (2005). “A Comparison of Imputation Methods in a Longitudinal Randomized Clinical Trial.” *Statistics in Medicine* 24, 2111–2128.

Teixeira-Pinto, A. and Harezlak, J. (2013). “Analysis of Mixed Data. Methods & Applications”. Chapman & Hall/CRC. Chap. Factorization and Latent Variable Models for Joint Analysis of Binary and Continuous Outcomes, 81–92.

Torabi, M. (2012). “Likelihood Inference in Generalized Linear Mixed Models with Two Components of Dispersion Using Data Cloning.” *Journal of Computational Statistics and Data Analysis* 56, 4259–4265.

— (2013). “Likelihood Inference in Generalized Linear Mixed Measurement Error Models.” *Journal of Computational Statistics and Data Analysis* 57, 549–557.

Vangeneugden, T., Molenberghs, G., Laenen, A., Geys, H., Beunckens, C., and Sotto, C. (2010). “Marginal Correlation in Longitudinal Binary Data Based on Generalized Linear Mixed Models.” *Communications in Statistics – Theory and Methods* 39 (19), 3540–3557.

Vangeneugden, T., Molenberghs, G., Verbeke, G., and Demétrio, C.G.B. (2011). “Mar ginal Correlation from an Extended Random-Effects Model for Repeated and Overdispersed Counts.” *Journal of Applied Statistics* 38 (2), 215–232.

Verbeke, G., Fieuws, S., Molenberghs, G., and Davidian, M. (2012). “The analysis of Multivariate Longitudinal Data: A Review.” *Statistical Methods in Medical Research* 0 (0), 1–18.

Wang, Z. and Louis, T.A. (2004). “Marginalized Binary Mixed-Effects Models with Covariate-Dependent Random Effects and Likelihood Inference.” *Biometrics* 60, 884–891.

Withanage, N. (2013). *Methods and Applications in the Analysis of Correlated Non-Gaussian Data.* PhD Thesis. University of Calgary, Canada.

Wuertz, D. (2013). *fOptions: Basics of Option Valuation. R package version 3010.83.* URL: `http://cran.r-project.org/web/packages/fOptions` (visited on 04/08/2014).

Yalcinoz, Z. (2008). *A Simulation Study on Marginalized Transition Random Effects Models for Multivariate Longitudinal Binary Data.* M.Sc. Thesis. Department of Statistics, Middle East Technical University, Turkey.

Zeger, S.L. and Liang, K.Y. (1986). “Longitudinal Data Analysis for Discrete and Continuous outcomes.” *Biometrics* 42, 121–130.

Zeng, L. and Cook, R.J. (2007). “Transition Models for Multivariate Longitudinal Binary Data.” *Journal of the American Statistical Association* 102, 211–223.

Zhang, H., Lu, N., Feng, C., Thurston, S.W., Xia, Y., and Tu, X.M. (2011). “On Fitting Generalized Linear Mixed-Effects Models for Binary Responses Using Different Statistical Packages.” *Statistics in Medicine* 30, 2562–2572.

Zhang, N., Chen, H., and Zou, Y. (2014). “A Joint Model of Binary and Longitudinal Data with Non-Ignorable Missingness, with Application to Marital Stress and Late-Life Major Depression in Women.” *Journal of Applied Statistics* 41 (5), 1028–1039.

Zhao, L.P. and Prentice, R.L. (1990). “Correlated Binary Regression Using a Quadratic Exponential Model.” *Biometrics* 77, 642–648.

APPENDIX A

DETAILED CALCULATIONS OF Δ_{itj}

In probability theory, it is known that any conditional expectation can be written in terms of marginal expectation. This implies that the integration of conditional probability $Pr(Y_{itj} = 1|\mathbf{X}_{it}, b_{it})$ over the distribution of random effects results in marginal probability $Pr(Y_{itj} = 1|\mathbf{X}_{it})$ for random effects models for longitudinal binary data as given follows:

$$\begin{aligned} E(Y_{itj}|\mathbf{X}_{it}) &= \int E(Y_{itj}|\mathbf{X}_{it}, b_{it})f(b_{it})db_{it} \\ \Rightarrow Pr(Y_{itj} = 1|\mathbf{X}_{it}) &= \int Pr(Y_{it} = 1|\mathbf{X}_{it}, b_{it})f(b_{it})db_{it}, \end{aligned} \tag{A.1}$$

where $f(b_{it})$ is a univariate normal distribution with mean 0 and variance $var(b_{it})$. Substituting marginal and conditional probabilities in equation 3.1, which are the first and second levels of the proposed model, respectively, into equation A.1 gives the following expression:

$$\frac{exp(\mathbf{X}_{it}\boldsymbol{\beta}_j)}{1 + exp(\mathbf{X}_{it}\boldsymbol{\beta}_j)} = \int_{-\infty}^{+\infty} \Phi(\Delta_{itj}(\mathbf{X}_{it}) + b_{it})f(b_{it})db_{it}. \tag{A.2}$$

Since $b_{it} \sim N(0, Var(b_{it}))$, it is possible to write $b_{it} = \sqrt{Var(b_{it})}z_i$, where $z_i \sim N(0, 1)$. Following Griswold (2005), define a $W_i \sim N(0, 1)$, which is $W_i \perp z_i$, where the symbol $\perp$ denotes independence. Then it is easy to show that

$$\frac{W_i}{\sqrt{Var(b_{it})}}) - z_i \sim N(0, 1 + Var(b_{it})^{-1}). \tag{A.3}$$

The right-hand side of equation A.2 can be rewritten as follows:

$$\begin{aligned}
\int_{-\infty}^{+\infty} \Phi\left(\Delta_{itj}(\mathbf{X}_{it}) + b_{it}\right) f(b_{it}) db_{it} &= \int_{-\infty}^{+\infty} \Phi\left(\Delta_{itj}(\mathbf{X}_{it}) + \sqrt{Var(b_{it})} z_i\right) \phi(z_i) dz_i \\
&= \int_{-\infty}^{+\infty} \Pr\left(W_i \leq \Delta_{itj}(\mathbf{X}_{it}) + \sqrt{Var(b_{it})} z_i\right) \phi(z_i) dz_i \\
&= \int_{-\infty}^{+\infty} \Pr\left(\frac{\frac{W_i}{\sqrt{Var(b_{it})}} - z_i}{\sqrt{1 + Var(b_{it})^{-1}}} \leq \frac{\frac{\Delta_{itj}(\mathbf{X}_{it})}{\sqrt{Var(b_{it})}}}{\sqrt{1 + Var(b_{it})^{-1}}}\right) \phi(z_i) dz_i \\
&= \int_{-\infty}^{+\infty} \Phi\left(\frac{\Delta_{itj}(\mathbf{X}_{it})}{\sqrt{1 + Var(b_{it})}}\right) \phi(z_i) dz_i \\
&= \Phi\left(\frac{\Delta_{itj}(\mathbf{X}_{it})}{\sqrt{1 + Var(b_{it})}}\right).
\end{aligned}$$

Then equation A.2 can be rewritten as follows:

$$\frac{exp(\mathbf{X}_{it}\boldsymbol{\beta}_j)}{1 + exp(\mathbf{X}_{it}\boldsymbol{\beta} j)} = \Phi\left(\frac{\Delta_{itj}(\mathbf{X}_{it})}{\sqrt{1 + Var(b_{it})}}\right). \tag{A.4}$$

Solving equation A.4 for Δ_{itj} provides a closed-form solution for it such that

$$\Delta_{itj} = \Phi^{-1}\left(\frac{exp(\mathbf{X}_{it}\boldsymbol{\beta}_j)}{1 + exp(\mathbf{X}_{it}\boldsymbol{\beta}_j)})\right) \sqrt{1 + Var(b_{it})}, \tag{A.5}$$

where Δ_{itj} is an explicit function of both marginal regression parameters and variance of random effects.

APPENDIX B

TAYLOR SERIES-BASED DERIVATION OF MARGINAL CORRELATION FUNCTION

The formula for the correlation between Y_{itj} and $Y_{it'j'}$, $Corr\left(Y_{itj}, Y_{it'j'}\right)$ is given as follows:

$$Corr\left(Y_{itj}, Y_{it'j'}\right) = \frac{Cov\left(Y_{itj}, Y_{it'j'}\right)}{\sqrt{Var(Y_{itj})}\sqrt{Var(Y_{it'j'})}} \quad \forall i, t \text{ and } j, \tag{B.1}$$

where $Cov\left(Y_{itj}, Y_{it'j'}\right)$ is covariance between Y_{itj} and $Y_{it'j'}$ and $Var(Y_{itj})$ is variance of Y_{itj}.

For any model including random effects, the formulation for $Corr\left(Y_{itj}, Y_{it'j'}\right)$ in equation B.1 relies on conditional expectation, variance, and covariance formulas. The marginal covariance $Cov\left(Y_{itj}, Y_{it'j'}\right)$ can be formulated in terms of conditional expectation and covariance as follows (Rudary, 2009):

$$Cov\left(Y_{itj}, Y_{it'j'}\right) = E\left(Cov\left(Y_{itj}, Y_{it'j'}|b_{it}, b_{it'}\right)\right) + Cov\left(E\left(Y_{itj}|b_{it}\right), E\left(Y_{it'j'}|b_{it'}\right)\right), \tag{B.2}$$

where $E\left(Cov\left(Y_{itj}, Y_{it'j'}|b_{it}, b_{it'}\right)\right) = E\left(E[(Y_{itj} - \mu_{itj}) \times (Y_{it'j'} - \mu_{it'j'})]|b_{it}, b_{it'}\right) = E\left(E\left(Y_{itj} - \mu_{itj}\right) \times E\left(Y_{it'j'} - \mu_{it'j'}\right)\right) = 0$ since $Y_{itj}|b_{it}$ and $Y_{it'j'}|b_{it'}$ are independent of each other, which is also known as the conditional independence assumption, and $Cov\left(E\left(Y_{itj}|b_{it}\right), E\left(Y_{it'j'}|b_{it'}\right)\right) = Cov\left(\mu_{itj}, \mu_{it'j'}\right)$. Then $Cov\left(Y_{itj}, Y_{it'j'}\right)$ in equation B.2 can be rewritten as follows:

$$Cov\left(Y_{itj}, Y_{it'j'}\right) = Cov\left(\mu_{itj}, \mu_{it'j'}\right). \tag{B.3}$$

The marginal variance $Var(Y_{itj})$ can be formulated in terms of conditional expectation and variance as follows:

$$Var(Y_{itj}) = E\left(Var\left(Y_{itj}|b_{it}\right)\right) + Var\left(E\left(Y_{itj}|b_{it}\right)\right), \tag{B.4}$$

where $E\left(Var\left(Y_{itj}|b_{it}\right)\right) = E\left(\mu_{itj}(1-\mu_{itj})\right)$ and $Var\left(E\left(Y_{itj}|b_{it}\right)\right) = Var\left(\mu_{itj}\right)$. Then $Var(Y_{itj})$ in equation B.4 can be rewritten as follows:

$$Var(Y_{itj}) = E\left(\mu_{itj}(1-\mu_{itj})\right) + Var\left(\mu_{itj}\right), \tag{B.5}$$

Then $Corr\left(Y_{itj}, Y_{it'j'}\right)$ in equation B.1 can be rewritten based on equations B.3 and B.5 as follows:

$$Corr\left(Y_{itj}, Y_{it'j'}\right) = \frac{Cov\left(\mu_{itj}, \mu_{it'j'}\right)}{\sqrt{E\left(\mu_{itj}(1-\mu_{itj})\right) + Var\left(\mu_{itj}\right)}\sqrt{E\left(\mu_{it'j'}(1-\mu_{it'j'})\right) + Var\left(\mu_{it'j'}\right)}}.$$

Taylor Series-Based Approximations for $E(\mu_{itj}(1-\mu_{itj})), Var(\mu_{itj})$ and $Cov(\mu_{itj}, \mu_{it'j'})$

The equation B.6 requires a closed-form solution for $E(\mu_{itj}(1-\mu_{itj}))$, $Var(\mu_{itj})$ and $Cov(\mu_{itj}, \mu_{it'j'})$. However, unless the link function in μ_{itj} is an identity function, a closed-form solution for these expressions cannot be obtained (e.g., for binary responses). In this sense, following Goldstein and Rasbash (1996), Vangeneugden et al. (2010), and Vangeneugden et al. (2011), these expressions can be approximated by a first-order Taylor series expansion around $b_{it} = 0$.

First-order Taylor series expansions of $E(\mu_{itj}(1-\mu_{itj}))$ and $Var(\mu_{itj})$ around $b_{it} = 0$ give the following approximations, respectively:

$$E(\mu_{itj}(1-\mu_{itj})) \simeq (\mu_{itj}(1-\mu_{itj})|_{b_{it}=0}) \quad \text{and} \tag{B.6}$$

$$Var(\mu_{itj}) \ \simeq \left(\tfrac{\partial\mu_{itj}}{\partial b_{it}}|_{b_{it}=0}\right)^2 Var(b_{it}). \tag{B.7}$$

In a similar fashion, a first-order Taylor series expansion for $Cov(\mu_{itj}, \mu_{it'j'})$ around $b_{it} = 0$ and $b_{it'} = 0$ gives the following approximation:

$$Cov\left(Y_{itj}, Y_{it'j'}\right) \simeq \left(\frac{\partial\mu_{itj}}{\partial b_{it}}|_{b_{it}=0}\right) Cov\left(b_{it}, b_{it'}\right) \left(\frac{\partial\mu_{it'j'}}{\partial b_{it'}}|_{b_{it'}=0}\right). \tag{B.8}$$

For the proposed model in equation 3.1, $\mu_{itj} = Pr(Y_{itj} = 1|\mathbf{X}_{it}, b_{it}) = \Phi(\Delta_{itj}(\mathbf{X}_{it})+ b_{it})$. Then equations B.6, B.7 and B.8 can be rearranged, respectively, as follows:

$$\begin{aligned} E(\mu_{itj}(1-\mu_{itj})) &\simeq (\mu_{itj}(1-\mu_{itj})|_{b_{it}=0}) \\ &= \Phi(\Delta_{itj}(\mathbf{X}_{it}))(1-\Phi(\Delta_{itj}(\mathbf{X}_{it}))), \end{aligned} \tag{B.9}$$

$$\begin{aligned} Var(\mu_{itj}) &\simeq \left(\frac{\partial\mu_{itj}}{\partial b_{it}}|_{b_{it}=0}\right)^2 \\ &= (\phi(\Delta_{itj}(\mathbf{X}_{it})))^2 Var(b_{it}), \quad \text{and} \end{aligned} \tag{B.10}$$

$$\begin{aligned} Cov\left(Y_{itj}, Y_{it'j'}\right) &\simeq \left(\frac{\partial\mu_{itj}}{\partial b_{it}}|_{b_{it}=0}\right) Cov\left(b_{it}, b_{it'}\right) \left(\frac{\partial\mu_{it'j'}}{\partial b_{it'}}|_{b_{it'}=0}\right) \\ &= \phi(\Delta_{itj}(\mathbf{X}_{it})) Cov\left(b_{it}, b_{it'}\right) \phi(\Delta_{it'j'}(\mathbf{X}_{it})). \end{aligned} \tag{B.11}$$

where $\phi(\Delta_{itj}(\mathbf{X}_{it})) = \frac{\partial\mu_{itj}}{\partial b_{it}}|_{b_{it}=0}$, $\phi(\Delta_{it'j'}(\mathbf{X}_{it})) = \frac{\partial\mu_{it'j'}}{\partial b_{it'}}|_{b_{it'}=0}$, and $Var(b_{it})$ and $Cov(b_{it}, b_{it'})$ are the corresponding variance and covariance components in Σ_i.

For the proposed model in equation 3.1, $Corr(Y_{itj}, Y_{it'j'})$ can be expressed through equations B.9-B.11 as follows:

$$\begin{gathered} Corr\left(Y_{itj}, Y_{it'j'}\right) \simeq \\ \frac{\phi(\Delta_{itj}(\mathbf{X}_{it}))Cov\left(b_{it}, b_{it'}\right)\phi(\Delta_{it'j'}(\mathbf{X}_{it}))}{\sqrt{v_{itj} + (\phi(\Delta_{itj}(\mathbf{X}_{it})))^2 Var(b_{it})}\sqrt{v_{it'j'} + (\phi(\Delta_{it'j'}(\mathbf{X}_{it})))^2 Var(b_{it'})}}, \end{gathered} \tag{B.12}$$

$\forall\, i$, t and j, where $v_{itj} = \Phi(\Delta_{itj}(\mathbf{X}_{it}))(1 - \Phi(\Delta_{itj}(\mathbf{X}_{it})))$ and $v_{it'j'} = \Phi(\Delta_{it'j'}(\mathbf{X}_{it}))$ $(1 - \Phi(\Delta_{it'j'}(\mathbf{X}_{it})))$.

APPENDIX C

R CODE FOR DATA CLONING

The R code for model fitting and prediction of random effects under Scenario 1 is as follows:

```
#############Model Fitting with Data Cloning####################
#Define the Data
#n=number of total subjects
#nt=number of total time points
#nr=number of responses
#np=number of regression parameters for each response
#(including intercept)
#resmat=a n*nt*nr by 1 response vector.
#covmat=a n*nt*nr by np covariate matrix to associate
#with responses
#Xg=a n*nt*nr by 2 covariate matrix to associate with
#dependence parameters
#Xl=a n*nt*nr by 2 covariate matrix to associate with
#variance parameters
#beta1int=a np by 1 vector of initial values for
#regression parameters
#of first response obtained from gee analysis
#beta2int=a np by 1 vector of initial values for
#regression parameters
#of second response obtained from gee analysis

data<-list("respmat"=resmat,"covariatemat"=covmat,
"Xgamma"=Xg, "Xlamda"=Xl,"betait1"=beta1int,
"betait2"=beta2int)
```

```
#Prepare the data for data cloning

#id=rep(rep(seq(1:n),each=nt),nr)  #index for subject
#time=rep(seq(1:nt),n*nr)          #index for time
#rind=rep(1:nr,each=n*nt)          #index for response type

mydata<-list(Y=dcdim(as.matrix(data[[1]])),X=data[[2]],
Xg=data[[3]],Xl=data[[4]],nt=length(unique(time)),nr=2,
n=length(unique(id)),np=dim(data[[2]])[2],
ntot=dim(data[[2]])[1],id=id,time=time,rind=rind,K=1)

#Clone the data
clmydata<-dclone(mydata, n.clones=10, multiply="K",
unchanged=c("nt","nr","n","np","ntot","id","time","rind") )

#Prepare the initials for parameters
inits<-function() {list(beta=cbind(mvrnorm(1,as.matrix(data[[5]],
np,1),diag(np)),mvrnorm(1,as.matrix(data[[6]],np,1),diag(np))),
gamma=c(runif(1,0,0.5),runif(1,0,1.5)),lambda=c(runif(1,0,0.5),
runif(1,0,1.5)))}

#Define the Model

mymodel<-function(){

#define the marginal part
for (k in 1:K)     {
for (j in 1:ntot)  {
Y[j,k] ~ dbern(cp[j,k])
cp[j,k]<-min(max(cp1[j,k],0.0000001),0.9999999)
cp1[j,k]<- phi(delta[j,k]+ b[id[j],time[j],k])
delta[j,k]<-
qnorm(mp[j,k],0,1)*sqrt(1+bigE[id[j],time[j],time[j]])
logit(mp[j,k])<-inprod(X[j,],beta[,rind[j]])
} # j
} # k
```

```
#define random effects
for (k in 1:K) {
for (i in 1:n) {
b[i,1:nt,k]~dmnorm(mu[1:nt], InvbigE[i,1:nt,1:nt])
}#i
}#k

#define the mean vector for random effects
for ( t in 1:nt) {
mu[t]<-0
}#t

#define the inverse of the subject-specific
#random effects covariance matrix
for (i in 1:n) {
InvbigE[i,1:nt,1:nt]<-
t(bigT[i,1:nt,1:nt])%*%invbigD[i,1:nt,1:nt]%*%bigT[i,1:nt,1:nt]
bigE[i,1:nt,1:nt]<-inverse(InvbigE[i,1:nt,1:nt])
}#i

#define an identity matrix
for ( t in 1:nt) {
for ( tt in 1:nt) {
Imat[t,tt]<-equals(t,tt)
}#tt
}#t

#define unit-lower-triangular matrix T
#define inverse of the diagonal matrix D
for ( i in 1:n)   {
for ( t in 1:nt)  {
for ( tt in 1:nt) {
abigT[i,t,tt]<-
equals(t, tt+1)*inprod(-Xg[t+(nt*(i-1)),],gamma[])
invbigD[i,t,tt]<-
equals(t,tt)*(1/exp(inprod(Xl[t+(nt*(i-1)),],lambda[])))
```

```
} # tt
} # t
bigT[i,1:nt,1:nt]<-abigT[i,1:nt,1:nt]+Imat[1:nt,1:nt]
} # i

#define priors for beta
for ( j in 1:nr)  {
for ( p in 1:np)  {
beta[p,j]~dnorm(0,0.001)
}#j
}#p

#define priors for gamma and lambda
for ( ii in 1:2)        {
gamma[ii]~ dnorm(0,0.001)
lambda[ii]~dnorm(0,0.001)
}#ii

}#mymodel

##jags.fit##

myfit<-jags.fit(clmydata,c("beta","gamma","lambda"),
mymodel,inits,n.chains=2,n.update=30000,n.iter=5000,thin=25)


#Prediction of random effects

pmydata<-list(Y=as.vector(data[[1]]),X=data[[2]],Xg=data[[3]],
Xl=data[[4]],nt=length(unique(time)),nr=2,n=length(unique(id)),
np=dim(data[[2]])[2],ntot=dim(data[[2]])[1],id=id,time=time,
rind=rind,param=as.vector(coef(myfit)))

#Define the Model

mymodelpred<-function(){
```

```
#define the marginal part
for (j in 1:ntot)  {
Y[j] ~ dbern(cp[j])
cp[j]<-min(max(cp1[j],0.0000001),0.9999999)
cp1[j]<- phi(delta[j]+ b[id[j],time[j]])
delta[j]<-
qnorm(mp[j],0,1)*sqrt(1 + bigE[id[j],time[j],time[j]])
logit(mp[j])<-inprod(X[j,],beta[,rind[j]])
} # j

#define random effects
for (i in 1:n) {
b[i,1:nt]~dmnorm(mu[1:nt], InvbigE[i,1:nt,1:nt])
}#i

#define the mean vector for random effects
for ( t in 1:nt) {
mu[t]<-0
}#t

#define the inverse of the subject-specific
#random effects covariance matrix
for (i in 1:n) {
InvbigE[i,1:nt,1:nt]<-
t(bigT[i,1:nt,1:nt])%*%invbigD[i,1:nt,1:nt]%*%bigT[i,1:nt,1:nt]
bigE[i,1:nt,1:nt]<-inverse(InvbigE[i,1:nt,1:nt])
}#i

#define an identity matrix
for ( t in 1:nt)  {
for ( tt in 1:nt) {
Imat[t,tt]<-equals(t,tt)
}#tt
}#t

#define unit-lower-triangular matrix T
#define inverse of the diagonal matrix D
```

```
for ( i in 1:n)   {
for ( t in 1:nt)  {
for ( tt in 1:nt) {
abigT[i,t,tt]<-
equals(t, tt+1)*inprod(-Xg[t+(nt*(i-1))],gamma)
invbigD[i,t,tt]<-
equals(t,tt)*(1/exp(inprod(Xl[t+(nt*(i-1))],lambda)))
} # tt
} # t
bigT[i,1:nt,1:nt]<-abigT[i,1:nt,1:nt]+Imat[1:nt,1:nt]
} # i


for ( j in 1:nr)  {
for ( p in 1:np)  {
beta[p,j]<-param[p+np*(j-1)]
}#j
}#p

gamma[1:2]<-param[((nr*np)+1):((nr*np)+2)]
lambda[1:2]<-param[((nr*np)+3):((nr*np)+4)]

}#mymodel

##fit##
myfit.p<-jags.fit(pmydata,c("b"),mymodelpred,
n.chains=2,n.update=30000,n.iter=5000,thin=25)

#get the random effects
summary(myfit)
```

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: İnan, Gül
Nationality: Turkish (TC)
Date and Place of Birth: 11/06/1983 & Malatya
Phone: 0 312 2105316
Fax: 0 312 2102959

EDUCATION

Degree	Institution	Year of Graduation
M.S.	Middle East Technical University (METU) Department of Statistics	2009
B.S.	Middle East Technical University (METU) Department of Statistics	2006
High School	Malatya Anatolian High School	2001

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
Sep., 2006 - Dec., 2013	Department of Statistics, METU	Research Assistant

PUBLICATIONS

1. Inan, G., Ilk Dag, O. and de Leon, A.R. (2014). "Letter to Editor Re: "Logistic regression with multiple random effects: A simulation study of estimation

methods and statistical packages" by Y. Kim, Y-K. Choi, and S. Emery." *The American Statistician* 67(3),171-182. DOI:10.1080/00031305.2014.904251.

2. Rajabli, F., Goktas, U. and Inan, G. (2013). "Application of Dempster Schafer Method in Family-Based Association Studies." *IEEE Computational Biology & Bioinformatics* 10(4),1071 - 1075.
3. Rajabli, F., Inan, G. and Ilk, O. (2013). "Power analysis of C-TDT for small sample size genome-wide association studies by the joint use of case-parent trios and pairs." *Computational and Mathematical Methods in Medicine*. DOI:10.1155/2013/235825.

International Conference Publications

1. Inan, G. and Ilk, O. A Marginalized Multilevel Model for Bivariate Longitudinal Binary Data. *S. Co. 2013-Complex Data Modeling and Computationally Intensive Methods*. 9-11 September 2013, Milan, Italy.