ON MULTIVARIATE LONGITUDINAL BINARY DATA MODELS AND THEIR APPLICATIONS IN FORECASTING

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

BY

ÖZGÜR ASAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

JULY 2012

Approval of the thesis:

ON MULTIVARIATE LONGITUDINAL BINARY DATA MODELS AND THEIR APPLICATIONS IN FORECASTING

submitted by **ÖZGÜR ASAR** in partial fulfillment of the requirements for the degree of **Master of Science in Statistics Department, Middle East Technical University** by,

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

Prof. Dr. H. Öztaş Ayhan
Head of Department, **Statistics** ________________

Assist. Prof. Dr. Özlem İlk
Supervisor, **Statistics Department, METU** ________________

Examining Committee Members:

Assoc. Prof. Dr. İnci Batmaz
Statistics Department, METU ________________

Assist. Prof. Dr. Özlem İlk
Statistics Department, METU ________________

Assist. Prof. Dr. Ali Devin Sezer
Institute of Applied Mathematics, METU ________________

Assist. Prof. Dr. Berna Burçak Başbuğ Erkan
Statistics Department, METU ________________

Assist. Prof. Dr. Zeynep Işıl Kalaylıoğlu
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: ÖZGÜR ASAR

Signature :

ABSTRACT

ON MULTIVARIATE LONGITUDINAL BINARY DATA MODELS AND THEIR APPLICATIONS IN FORECASTING

Asar, Özgür
M.S., Department of Statistics
Supervisor : Assist. Prof. Dr. Özlem İlk

July 2012, 219 pages

Longitudinal data arise when subjects are followed over time. This type of data is typically dependent, due to including repeated observations and this type of dependence is termed as within-subject dependence. Often the scientific interest is on multiple longitudinal measurements which introduce two additional types of associations, between-response and cross-response temporal dependencies. Only the statistical methods which take these association structures might yield reliable and valid statistical inferences. Although the methods for univariate longitudinal data have been mostly studied, multivariate longitudinal data still needs more work. In this thesis, although we mainly focus on multivariate longitudinal binary data models, we also consider other types of response families when necessary. We extend a work on multivariate marginal models, namely multivariate marginal models with response specific parameters (MMM1), and propose multivariate marginal models with shared regression parameters (MMM2). Both of these models are generalized estimating equation (GEE) based, and are valid for several response families such as Binomial, Gaussian, Poisson, and Gamma. Two different R packages, **mmm** and **mmm2** are proposed to fit them, respectively. We further develop a marginalized multilevel model, namely probit normal marginalized transition random effects models (PNMTREM) for multivariate longitudinal binary response. By this model, *implicit function theorem* is introduced to explicitly link the levels of marginalized multilevel models with transition structures for the first time. An R package, **pnmtrem** is proposed to fit the model. PNMTREM is applied to data collected through Iowa Youth and Families Project (IYFP). Five different models, including univariate and multivariate ones, are considered to forecast multivariate longitudinal binary data. A comparative simulation study, which includes a model-independent data simulation process, is considered for this purpose. Forecasting independent variables are taken into account as well. To assess the forecasts, several accuracy measures, such as expected proportion of correct prediction (ePCP), area under the receiver operating characteristic (AUROC) curve, mean absolute scaled error (MASE) are considered. Mother's Stress and Children's Morbidity (MSCM) data are used to illustrate this comparison in real life. Results show that marginalized models yield better forecasting results compared to marginal models. Simulation results are in agreement with these results as well.

Keywords: Accuracy measures, marginalized multilevel models, maximum likelihood estimation, multiple outcomes, statistical software.

ÖZ

ÇOK DEĞİŞKENLİ İKİ DEĞERLİ UZUNLAMASINA VERİ MODELLERİ VE BU MODELLERİN ÖNGÖRÜ UYGULAMALARI

Asar, Özgür
Yüksek Lisans, İstatistik Bölümü
Tez Yöneticisi : Yrd. Doç. Dr. Özlem İlk

Temmuz 2012, 219 sayfa

Uzunlamasına veri şahıslar zaman içinde takip edildiğinde meydana gelmektedir. Bu tip veriler tekrarlı gözlemler içerdiğinden tipik olarak bağımlıdırlar ve bu tip bağımlılık, birey içi bağımlılığı olarak adlandırılmaktadır. Bilimsel ilgi, sıklıkla birden fazla uzunlamasına bağımlı değişken üzerine olabilir. Bu durumda iki ek bağımlılık tipi meydana gelmektedir. Bunlar belirli bir zamanda gözlenen, farklı bağımlı değişkenler arasındaki ilişki ve farklı zamanlara ait farklı bağımlı değişkenler arası çaprazlama ilişki olarak adlandırılmaktadır. Sadece, bu bahsedilen bağımlılık yapılarını hesaba katan istatistiksel yöntemler güvenilir ve geçerli istatistiksel çıkarımlar elde edilmesine olanak verebilir. Tek bağımlı değişken içeren uzunlamasına veri analizi için uygun olan yöntemler üzerine şimdiye kadar çokça çalışılmış olmasına rağmen, birden fazla bağımlı değişken içeren uzunlamasına veri halen daha fazla çalışma gerektirmektedir. Bu tezde esas olarak, çok değişkenli iki değerli uzunlamasına veri üzerine yoğunlaşıldığı halde, gerektiğinde diğer bağımlı değişken aileleri de dikkate alınmıştır. Çok değisşkenli marjinal model üzerine olan ve bağımlı değişkene özgü parametre varsayımı yapan çok değişkenli marjinal model (MMM1) olarak adlandırılan bir modeli genişlettik ve bütün bağımlı değişkenler için ortak parametre varsayımı yapan marjinal model (MMM2) olarak adlandırılan çok değişkenli marjinal modeli geliştirdik. Bu modeller genelleştirilmiş tahmin denklemlerini (GEE) esas almaktadır, ve çeşitli bağımlı değişken aileleri için geçerlidir, örneğin Binom, Gauss, Poisson ve Gamma aileleri. Iki farklı R paketi, **mmm** ve **mmm2**, sırasıyla bu modeller ile parametre tahmini yapabilmek için hazırlanmıştır. Ayrıca, çok değişkenli iki değerli uzunlamasına veri için, probit normal marjinalleştirilmiş geçişli rastgele etkili model (PNMTREM) adlı marjinalleştirilmiş çok değişkenli modeli geliştirdik. Bu model ile birlikte, *kapalı fonksiyonun varlığı teoremi*, geçişli yapıya sahip marjinalleştirilmiş çok seviyeli modellerin seviyelerini açıkça bağlamak için ilk kez sunulmuştur. **pnmtrem** adlı R paketi bu modelin parametre tahmini için hazırlanmıştır. PNMTREM ile Iowa Gençlik ve Aileleri Projesi (IYFP) kapsamında toplanan veri analiz edilmiştir. Tek ve çok değişkenli modellerden oluşan beş farklı model, çok değişkenli iki değerli uzunlamasına verinin ileriye dönük tahmini için ele alınmıştır. Kıyaslama amacı güden ve herhangi bir modele bağımlı olmayan veri üretimini esas alan bir simülasyon çalışması, bu amaç için gerçekleştirilmiştir. Bu çalışmada, bağımsız değişkenlerin ileriye dönük tahmini de ele alınmıştır. İleriye dönük tahminlerin değerlendirilmesi için çeşitli doğruluk ölçütleri, örneğin, beklenen doğru tahmin etme oranı (ePCP), alıcı işletim sistemi karakteristiği eğrisi altında kalan alan (AUROC) ve ortalama mutlak ayarlanmış hata (MASE). Annelerin Stresi ve Çocukların Hastalığı (MSCM) verisi bu karşılaştırmayı gerçek hayat uygulamasında örneklemek için kullanılmıştır. Elde edilen sonuçlar marjinalleştirilmiş modellerin, marjinal modellere göre daha iyi ileriye dönük tahmin yaptığını göstermektedir. Simülasyon sonuçları da bu sonuçları doğrular niteliktedir.

Anahtar Kelimeler: Doğruluk ölçüleri, marjinalleştirilmiş çok seviyeli modeller, en büyük olabilirlik tahmini, çoklu sonuçlar, istatistik programları.

to my mother and sister

ACKNOWLEDGMENTS

This thesis has a relatively long history than expected due to our broadening the scope and the content of it continuously. Nevertheless, somewhere we had to stop despite the fact that there are lots of work to do. I am very grateful to Assist. Prof. Dr. Özlem İlk for her support and invaluable advices in this long history of study. She patiently listened all my ideas and directed me when I needed her guidance. The various statistical topics that she never blocked me to work on led me to have a broad statistical perception and knowledge. I could say without hesitation that I gained unique experiences while studying with her. I am also very thankful to her for opening a course called "Panel Data Analysis". The related lecture notes often inspired and motivated me when constructing the skeleton of this thesis and writing it.

I would like to thank my committee members, Assoc. Prof. Dr. İnci Batmaz, and Assist. Prof. Drs. Ali Devin Sezer, Berna Burçak Başbuğ Erkan and Zeynep Işıl Kalaylıoğlu for their time and effort which they spent for me. I am also very grateful for their positive feedbacks which made me gain extra motivation for my studies.

I would like to acknowledge that the recommendation of Assist. Prof. Dr. Ali Devin Sezer regarding explicitly linking the levels of marginalized models has an important role in Chapter 4. I would also like to acknowledge that the guidance of Kuo Hui-Chen (from the National Chung-Hsing Library) to finding one of the important references we used in Chapter 4 is invaluable for us. They helped us without looking after any benefit. I believe that this is the essence of science. I am also thankful to my friends who provided some of the articles we benefited while preparing this thesis to which my university did not have access. I would like to acknowledge that R and LaTeX are god send and like to express my respect to the ones who ever contributed to them. I also would like to express my respect to the researchers whom we referred in this study.

I am very thankful for the financial support of Scientific and Technological Research Institution of Turkey (TUBITAK) during my master study.

I am hearty grateful to my mother, Atike, for her devotion and strong support. I would like to acknowledge that I am very lucky having a sister like you, Keziban. You should know that you are the best of my life. I am sorry for the stolen times during my studies that we should spend together. This study is dedicated to my family.

Last bu not least, I would like to state that the datasets which we used to illustrate our methods are not just numbers for us. However, we know that the people included in those studies had very hard times. We hope that our studies would be beneficial to the people who face similar hardships in life. This study is also dedicated to the people whose datasets were used in this thesis.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

Longitudinal data (LD; also known as panel data, used throughout interchangeably) comprise measurements taken repeatedly over time from same individuals/units/cases. This type of data is common in medical studies, clinical trials, economical studies, social sciences, psychiatry, educational and behavioral sciences, industry etc. LD has many advantages compared to cross-sectional (CSD) and time series data (TSD) (Diggle et al., 2002; Ilk, 2008). Since CSD is collected only at a single time point, it can not measure the change. Therefore, the conclusions/inferences drawn from such data are often typically inconsistent with the future studies. TSD usually include long and a few series and no associated covariates, hence their modeling framework is on history of the series. A key assumption in TSD analysis is that the series should be stationary (constant mean and constant variance across time). Additionally, observations are usually assumed to be collected with equispaced time points, e.g. biannually or annually collection. LD contains the features of both the CSD and TSD and extend them. Unlike CSD, they allow measuring the change, hence related conclusions/inferences are more suitable for future studies. LD usually contains short series compared to TSD and does not need the stationarity assumption. Furthermore, LD consists of multiple series and associated covariates. One outstanding discrimination between LD analysis and TSD analysis could be that while in the former the correlation within the series are mostly treated as nuisance parameters, in the latter, models are mainly built on that correlation. Another flexibility of LD over TSD is that LD does not need to be collected at equispaced time points. Lastly, while LD and TSD allow forecasting future events, CSD does not allow forecasting those events.

Following Ilk (2008), LD structures could be threefold: unconstrained, constrained and fully constrained. Unconstrained data structure is the most general and flexible data structure and contains repeated measures at uncommon and irregular time points, i.e., the number of repeated measures vary from subject to subject and they are collected at unequally spaced time points. Constraint data structure comprises common but irregular time points, i.e, while all the subjects have equal number of follow-ups, those are made at unequally spaced time points. Fully constraint data structure stresses on common and regular time points, i.e. all the subjects have same number of equally spaced follow-ups. In literature, whereas unconstrained data structure is the least common one among the aforementioned LD structures, fully constrained data structure is the most common one among them. Therefore, in our applications, we restrict our attention on fully constraint data structures. For the details of the aforementioned LD structures and examples for per structure, interested reader may refer to Ilk (2008).

LD might consist of continuous, discrete and/or categorical independent variables (covariates). They might be changing with time (time-varying/variant/dependent) or might not be changing with time (time-invariant/independent). Whereas gender of subjects might be an example for time-independent variables, employment status of them might be an example for the time-dependent one. The dependent variables (responses) in longitudinal data might be continuous, discrete or categorical as well. Whereas CD4+ cell numbers (used as a marker of HIV; Diggle et al., 2002) of subjects might be an example for longitudinal continuous response, epileptic seizure counts and stress status of the subjects might be examples of count and binary longitudinal responses, respectively. LD often might include multiple responses. Moreover, those responses might belong to different response families or might belong to the same response family. For example, income and life satisfaction could be two longitudinal responses that belong to different response families; anxiety, depression and hostility status of subjects could be three longitudinal responses all belonging to the same family. Here, we restrict our attention on multivariate longitudinal binary response data. However, we consider other response types (families) when necessary, e.g. in Chapter 3.

LD is typically not independent, since it is collected from same subjects/units. The related associations might be weak after a certain time-lag (short term dependence), e.g. only strong associations between time t and $t-1$, $t-2$. On the other hand, they might last being strong for most of the time-lags (long term dependence). Additionally, they might differ for same time-lags between different time points (unstructured), e.g. different associations between times t

and $t-1$ and times $t-1$ and $t-2$ although their time lags are equal. For the details of some of the patterns of those association structures and their formal definitions, see Section 2.3.2. Multivariate longitudinal data introduce an additional association which is the one between multivariate responses at a certain time point. Although multivariate LD introduces cross response associations additionally, e.g. association between the response k at time t and the response $k+1$ at time $t+1$, usually this type of association are suppressed due to the fact that they are naturally indicated by the two aforementioned associations. Here, we consider primarily former two associations, but we consider and discuss the latter as necessary.

Missing data (MD) is a portion of the whole dataset that are not observed but intended to be observed. For this reason, it should not to be confused with the missingness that occurs due to unconstrained longitudinal designs, since in those designs some subjects are not aimed to be followed by design. MD is very common in longitudinal studies due to several reasons such as following same subjects for a long period of time, etc. In the sense of Little and Rubin (2002) the mechanisms that create missing data could be threefold: missing completely at random (MCAR), missing at random (MAR), missing not at random (MNAR); while the former two are ignorable, latter is non-ignorable. If MD depends only on the covariates, then the mechanism is MCAR. However, if MD depends on both the covariates and observed data, then the mechanism is MAR. Nonetheless, if MD depends on both covariates, observed and missing observations, the the mechanism is MNAR. The identification of the mechanism that create missing data is more sensible in longitudinal studies, since data is available across time. For example if a subject has a missing observation at time t, then investigation of the observation (assuming it is observed) at time $t-1$ would help the identification of the related MD mechanism. Besides, the reason of why being missing at time t could be asked to the respondent at time $t+1$. However, in cross-sectional studies there is no such chance, hence related identification mostly depends on assumptions and expertise. The patterns of MD occurs in longitudinal studies could be twofold: monotone (drop-out) and non-monotone (intermittent). If a subject left the study at time t and never came back, then the pattern is monotone. However, if the subject left the study at time $t-1$ and came back at time t then the pattern is non-monotone. Both of the patterns are commonly faced in longitudinal studies. Since the former introduces a more systematic pattern, dealing with it is more easy compared to the latter. Analysis with MD could be threefold: complete case analysis, complete data analysis and modeling missing data. The first one includes ignoring the subjects with missing observations and/or ignoring the time points with missing data and doing analyses with the remaining data. The second one includes imputation of the missing data by proper imputation tools such as mean/mode/median imputation of subjects/variable/time points, conditional mean imputation, assigning baseline or last observation to the missing one and multiple imputation. The last one comprises modeling the missing data as well as the observed data such as pattern-mixture, selection and shared-parameter models alongside with sensitivity analyses. Although, modeling MD is beyond the scope of this thesis, we consider MD mechanisms and complete data analysis in our real life data analyses.

Family of generalized linear models (GLM) is a probabilistic modeling family which extends the normality assumption of linear models (McCullagh and Nelder, 1989). GLM allows modeling the mean response conditional only on the effects of the independent variables, i.e., $E(Y_i|X_i)$ with the aim of better explaining the variation of responses. Therefore, interpretation and calculation of covariate effects are direct, meaning not conditioned on any response conditions such as unobserved heterogeneity of them. GLM basically assumes that the subjects are independent of each other, hence direct implementation of GLM to longitudinal data would neglect the natural association structures of such data and behave the repeated observations are taken from separate subjects. Therefore, new methods were needed to be developed for LD modeling. Following Diggle et al. (2002), those models could be fourfold: marginal, transition, random effects and marginalized multilevel models. Marginal models could be considered as the extension of GLM to longitudinal data. They allow direct interpretation of the regression coefficients and useful for population-averaged inference such as comparing subgroups. Due to the fact that LD introduces multivariate response distribution in contrast to usual GLM in which response distribution is a univariate one, there are some challenges in terms of directly specifying the full distribution function. Whereas multivariate normal distribution might suit for longitudinal continuous data, counterparts for longitudinal discrete data is much more challenging. Luckily, multivariate quasi-likelihood method (also known as generalized estimating equations) help extending GLM to longitudinal data without fully identifying the multivariate distribution. In LD analysis literature, GLM with serial dependence (transition models) and/or random effects are widely used methodologies. Those models rely upon completing the univariate distribution function specification of the responses by response history and/or random effects, respectively instead of directly specifying the multivariate distribution and conceptualize that the repeated measures are independent given the response history and/or unobserved heterogeneity of the responses. They are called conditional models since covariate effects are modeled together with other conditions (history and/or unobserved heterogeneity of the responses). In conditional models calculation and interpretation of the regression coefficients might be difficult and they are useful for subject-specific purposes such as comparing certain individuals. Whereas transition models are mostly used for LD with short term dependence (significant dependence only between time t and $t-1$, $t-2$), random effects models are mostly used for LD with long range dependence (significant dependence between time t and $t-1, \ldots, t-p$ where p might be $t-1$). Choice between marginal and conditional models would be based on the purpose of the study. To illustrate, if the aim is looking for gender differences then marginal models would be used, but if the aim is investigating subjects one-by-one then conditional models would be used. Marginalized multilevel models acquire an approach of separating the marginal and conditional parts of the conditional models. They accommodate those parts in different levels of the models where the levels connected to each other by the expectation of the subsequent level over the distribution of the past responses and/or random effects distribution. They combine the best of two modeling frameworks for longitudinal data by multi-modeling and allow both population-averaged and subject-specific inference at the same time. One great advantage of marginalized models is that their regression coefficient estimates (the ones in the first levels) are not much affected by the misspecification of the true dependence structure. In addition to univariate LD, multivariate LD introduces more complex multivariate response distributions. However, solutions of multivariate marginal and marginalized models are valid as well by necessary modifications in estimating equations and the conditioning terms, respectively. Besides, multivariate conditional models are also valid by necessary modifications to the conditioning terms. One necessary note for appropriate model usage is that the design of the data might determine the usage of the models such as the use of marginal models via generalized estimating equations with autoregressive working correlation matrix, transition models and related marginalized multilevel counterparts and random effects models with autocorrelated structure and related multilevel counterparts are meaningful only for data collected at equally spaced time points. The related discussion will be enriched later. Here our concern is mostly on multivariate marginal models and multivariate marginalized models for multivariate LD; we consider also univariate marginal models throughout.

First order generalized estimating equations (GEE1; Liang and Zeger, 1986) is the most common method to fit the marginal models (especially for discrete response), since they do not require the full specification of multivariate distribution of the longitudinal responses. However, only the specification of first and second moments of that distribution is enough. GEE1 yields consistent regression parameter and related variance estimates under misspecification of the association structure, i.e., working correlation matrix. However, correct specification increases the efficiency of the parameter estimates. Nevertheless, the association parameter estimates under GEE1 are not reliable, since GEE1 only models the regression coefficients but estimates the association parameters by the method of moment estimation. Alternatively, second order generalized estimating equations (GEE2) were proposed to efficiently estimate the association parameters (via modeling them as well) together with the regression coefficients (see Chapter 3 of Fitzmaurice et al., 2009 and the original references therein; extension of GEE2 could be found in the alternating logistic regressions of Carey et al., 1993). Unless the association parameters are of primary interest, the use of GEE1 for marginal models might be preferred due its simplicity and abundance of available software. Even though GEE2 provides reliable association parameter estimates, it does not contribute to the marginal regression parameter estimates. To the best of our knowledge, there is only one software to fit GEE2 in R, called **orth** package (By et al., 2011). This package only handles with binary response data. However, there is wealth of software for GEE1. We discuss these software in Sections 2.3.2 and 3. Here our primary interest is only on the marginal regression parameter estimates and we consider univariate marginal models as a naive approach compared to other models such as marginalized multilevel models. Therefore, we consider only GEE1 to fit marginal models and call it GEE in short. Although GEE has simple forms, parameter estimates are not in closed forms and require numerical solutions such as Iteratively Re-weighted Least Squares or Fisher Scoring Algorithm. Maximum likelihood estimation (MLE) is mostly preferred for parameter estimation of conditional and multilevel models. The maximization of the likelihood might be mostly challenging due to the need to integrate over the random effects distribution and the absence of explicit functional forms of the related parameter estimates. Those challenges could be overcome by utilizing numerical integration such as Gauss-Hermite Quadratures or Monte Carlo Methods; numerical methods for root finding such as Newton's Method (also known as Newton-Raphson Method) or Fisher-Scoring Method (FS, also known as Quasi-Newton's Method). Multilevel models introduce more challenges in terms of the maximization of the likelihood compared to the conditional ones, since related levels are connected via constraint equations to be a valid probability model. Whereas they might require multiple usage of those numerical methods, they also might require the use of complex mathematical theorems to solve marginal constraints. For instance, in this study we proposed the use of implicit function theorem for explicit linkage of the levels of marginalized models. Bayesian methods, specifically Gibbs Sampling, are also commonly considered for parameter estimation in LD analysis. They are also full likelihood based methods and useful for complex data analysis like longitudinal data. One great advantage of those methods is that they do not require integration over the random effects distribution. Full likelihood based methods have some advantages over GEE such as they yield robust parameter estimates under missing at random data while GEE does only under missing completely at random data and while the usual model selection criterion are valid for the former, there is poverty of model selection for the latter. In this thesis, we consider GEE and MLE for parameter estimation.

Software for marginal models with GEE and mixed models are available in almost all the statistical packages. However, software for multilevel models are not available in any of the statistical packages. Nonetheless, they are in author written form and available upon request from the authors or on their web-pages. Here, we consider development of publicly available and user friendly R packages for multivariate longitudinal binary data.

Forecasting could be defined as making inferences for future events. It relies upon the time information available in the present data, hence time series and longitudinal data are convenient for forecasting studies. Although forecasting is a common study area in time series literature, there are few such studies in longitudinal data literature. All of the forecasting studies for longitudinal data are for univariate response and most of them are for continuous response. Additionally, they are common in longitudinal studies in economy, but rare in longitudinal studies in biostatistics. In this thesis, we consider forecasting multivariate longitudinal binary data.

1.2 Literature Review

We start with more general resources for longitudinal data (LD) and mainly focus on the ones for discrete response. Diggle et al. (2002), Agresti (2002), Fitzmaurice et al. (2004), Molenberghs and Verbeke (2005), Weiss (2005), Hedeker and Gibbons (2006), Ilk (2008), McCulloch et al. (2008) and Fitzmaurice et al. (2009) are some of the great books which mainly focus on longitudinal discrete data. Diggle et al. (2002) provided great resources on variety of topics such as exploratory methods, marginal, transition, random effects and multilevel models and missing data regarding the historical perspective of LD. Ilk (2008) introduced exploratory data analysis methods especially dynamic graphics via GGobi (Cook and Swayne, 2007) for multivariate LD and confirmatory methods for multivariate longitudinal binary data and considered Bayesian methods for parameter estimation. McCulloch et al. (2008) discussed the LD models in great depth with enormous focus on theoretical aspects and estimation methodologies. Fitzmaurice et al. (2009) comprised several topics and perspectives of LD in each of its chapters which were prepared by some of the head researchers of correlated data. Many LD examples and case studies could be found in those books. Also, a nice historical look at LD by Fitzmaurice and Molenberghs can be found in the Chapter 1 of Fitzmaurice et al. (2009). Although the book of Agresti (2002) did not primarily focus on LD, it provided necessarily much valuable information about categorical LD. Additionally, McCullagh and Nelder (1989) and Kutner et al. (2005) are great resources about generalized linear models (GLM) and GLM is important for LD analyses since LD models are built on GLM.

Molenberghs and Verbeke (2004) discussed meaningful model construction for repeated measures data by bringing theoretical meanings of the models into discussion such as negative variance components in linear mixed models and hypothesis testing for variance terms, illustrated those models with real life examples and compared different approaches. Gardiner et al. (2009) discussed the differences between marginal and random effects models in various perspectives and illustrated them with case studies. Molenberghs and Kenward (2010) provided philosophical discussion about marginal models with semi-parametric approaches and their unspecified parent (full) families. They brought up Bahadur model and hybrid models as the parent families of marginal models with generalized estimating equations. Horrocks and Van Den Heuvel (2009) introduced a joint model for longitudinal continuous and binary data in the concept of prediction of pregnancy and considered Bayesian methods for parameter estimation.

Here, our main focus is not on missing data in longitudinal studies. Nonetheless, since we discuss them throughly when necessary, it is beneficial to mention about a few great references. Molenberghs and Kenward (2007) and Ibrahim and Molenberghs (2009) provided great materials on missing data with longitudinal studies. While the former offers thorough resource, the latter provided a detailed review about the approaches and methods for missing data in repeated measures. Diggle et al. (2007) discussed an interesting approach to handle missing not at random (MNAR) data in longitudinal setting by considering the random effects as martingales and fit the proposed model in an available software.

Marginal models are placed in at least one chapter of almost all of the books referred above. As marginal models and generalized estimating equations (GEE) are well suit even confused, discussing literature of marginal models together with GEE would be beneficial; for alternative approaches to parameter estimation in marginal models see Molenberghs and Verbeke (2005). Liang and Zeger (1986) first proposed the GEE methodology to fit the marginal models extending the quasi-likelihood methodology of Wedderburn (1974) to correlated data. In this article, they introduced their methods and investigated it by simulation studies. In the same year, they illustrated their approach by a real life data application, called Mother's Stress and Children's Morbidity Study (MSCM; Zeger and Liang, 1986). MSCM data is considered in this thesis as well, see Sections 1.3.1 and 3.4. and Chapter 7. A great discussion of GEE alongside with its historical perspective and a little about the founders of the methodology could be found in Diggle (1997). A through work about GEE including its theory, extensions, behavior under missing data mechanisms, goodness of fit and model diagnostics and some case studies could be found in Chapter 3 of Fitzmaurice et al. (2009). Chapters 7, 8 and 12 of Diggle et al. (2002), Chapter 11 of Fitzmaurice et al. (2004) and Chapter 8 of Hedeker and Gibbons (2006) are also through texts for marginal models with GEE. Due to the outstanding and practical features of GEE and lack of literature for multivariate discrete distributions, GEE is the leading methods to fit marginal models (especially for discrete response) since its proposition (Fitzmaurice et al., 2009). Although we consider only longitudinal data perspective here, GEE deals with other correlated and clustered data. For instance, Carey et al. (1993) used an extension of GEE, namely Alternating Logistic Regressions, for a dataset for which association arose due to family members rather than following same subjects. There are two books mainly concentrating on GEE: Hardin and Hilbe (2003) and Ziegler (2011).

Shelton et al. (2004) proposed a methodology to fit multivariate marginal models in the univariate marginal modeling framework considering multivariate longitudinal binary response data and proposed a publicly available SAS macro for these models. Asar and Ilk (2012) carried their methodology into R and considered other response families rather than binomial. They proposed an R package called **mmm**. For the details of software for marginal models, see Section 2.3 and Chapter 3 and for multivariate marginal models, see Chapter 3. Ilk (2011b) is a great introduction to R book and is one of the early R books written in Turkish. The tutorial prepared by Leisch (2008) provided very user friendly steps to create R packages with simple examples.

Early works of generalized linear mixed models (GLMM) for binary response include Stiratelli et al. (1984) and Breslow and Clayton (1993). While the former discuss original setting, the latter discussed better estimation of regression parameters with approximate methods. Diggle et al. (2002) and McCulloch et al. (2008) are also great sources for such models. For transition models and methodologies to fit them, see Chapter 10 of Diggle et al. (2002). Since our main focus is not on those conditional models, here we briefly discuss some of the key works. Interested reader may refer the references related to conditional models cited in Diggle et al. (2002) and McCulloch et al. (2008).

Multilevel models for longitudinal data arose due to the deficiencies of marginal and mixed models. Heagerty (1999) proposed a two-level logistic regression models called Marginally Specified Logistic Regression Models (MSLNM) extending Azzalini's correlated logistic regressions (Azzalini, 1994, cited in Heagerty, 1999). The related framework has the form of random effects models and marginal and conditional components were separated into different levels. In that article, maximum likelihood estimation (MLE) and quadratic estimating equations were discussed for parameter estimation of multilevel models. Heagerty and Zeger (2000) provided a through work by completing and generalizing some aspects of the MSLNM with a philosophical perspective. For instance, they considered the behavior of marginal model parameters under the misspecification of the dependence structure and interpretation of marginal model parameters. Heagerty and Kurland (2001) compared the behaviors of GLMM and their marginalized structured counterparts on the marginal regression parameters under the misspecification of the random effects structures. They reported that GLMM is more sensitive to misspecification of dependence (random effects) structures compared to multilevel models. With the motivation of that marginalized models are also somehow sensitive to misspecification of random effects distributions, Mills et al. (2002) robustified the MSLNM of Heagerty (1999). They considered the same modeling framework but with Huber's least favorable distribution as the random effects distribution and doubly-weighted the likelihood function to protect it first to the outlying covariate effects (leverage effects) and outlying response effects. They also changed the numerical methods by adding flags in the process. They considered contamination in the data such as mistyping in covariates (e.g., 0.01 instead of 0.1) and reverse typing (by mistake) in responses (e.g., 1 instead of 0) and reported their model did well compared to MSLNM. Heagerty (2002b) proposed marginalized counterparts of the transition models by considering first and second orders and reported that the model is well working with long series of repeated observations. Wang and Louis (2004) considered a multilevel model with a heuristic approach of special bridge distribution for the random effects. Schildcrout and Heagerty (2007) combined the random effects and transition models at the same time and separated them in the second level of the multilevel model with the aim of well capturing the long term dependency which might occur with the long series. Those aforementioned multilevel models are all for univariate and binary longitudinal response. Additionally, all of those works used MLE approach for parameter estimation. Ilk and Daniels (2007) considered modeling multivariate longitudinal binary response with multilevel models. They proposed a three level model capturing the serial and multivariate response dependencies at different levels at second and third levels, respectively by extending the models of Heagerty (1999, 2002b). Unlike the former ones, they considered Bayesian methods, specifically Gibbs sampling with Hybrid steps (Neal, 1996), for parameter estimation. Lee et al. (2009) also considered modeling multivariate longitudinal binary data by extending the MSLNM of Heagerty (1999). They proposed a two-level model and captured serial and multivariate response dependencies in the second level of the models via the random effects. Bartolucci and Farcomeni (2009) considered single level models for dichotomous and ordinal categorical data. Their models accommodate longitudinal and multivariate response dependencies in a single level model alongside with the covariate effects as well.

For common methodologies used for parameter estimation in LD, interested reader may refer to Agresti (2002), McCulloch et al. (2008) and Demidenko (2004). For numerical methodologies to solve the likelihood function or estimating equations, one may refer to Conte and de Boor (1980), Kennedy and Gentle (1980), Gentle et al. (2004) and Gentle (2009). For the methodologies of numerical integration, one may refer to Abramowitz and Stegun (1972), Caflish (1998) and McCulloch et al. (2008).

Baltagi (2008) stated that forecasting is common in time series data, but not common in panel data. The works about forecasting with longitudinal data mostly placed in econometric longitudinal data literature and are rare in biostatistics literature. Although both of them have the common longitudinal data structures, the former datasets include relatively longer series (e.g., 40 time points in Frees and Miller, 2004) compared to the latter datasets. Therefore, in the analysis of econometric panel data some features of time series might be inherited such as stationary assumption, seasonality of the series etc. Some of the forecasting studies in the longitudinal data literature could be exemplified as follows. Frees and Miller (2004) forecasted the lottery sales by longitudinal mixed models considering a data set of the Wisconsin State Lottery. Baadsgaard et al. (2004) forecasted herd health by comparing a naive approach which is a simple time series model using previous situation as predictors and Bayesian state space models. Aslan (2010) considered the forecasting features of 21 different methods including too naive methods such as moving averages and relatively complex methods such as multilevel models (Heagerty, 2002b) to forecast longitudinal binary data.

Before closing this section, we should note that the literature review provided here aims providing a general review. Nonetheless, it will be enhanced in detail throughout the thesis in the related chapters.

1.3 Key Examples

In this sub-section, we introduce two multivariate longitudinal binary datasets, namely Mother's Stress and Children's Morbidity Study (MSCM) and Iowa Youth and Families Project (IYFP) to illustrate the properties of such data and depict the association structures of them. While the first dataset includes bivariate responses, the second one consists of trivariate responses.

1.3.1 Mother's Stress and Children's Morbidity Study

Mother's Stress and Children's Morbidity Study (MSCM; Alexander and Markowitz, 1986) aimed to investigate the effect of the mother's employment status on the use of pediatric care. 167 mothers and their preschool children (age of 18 months - 5 years) with no chronic disease were enrolled in this study. The demographic and family information of them including the marriage status, education level, employment status of mothers, children's race and gender, the health status of both the mothers and children at baseline and the size of the household were collected in a preliminary interview. All the collected variables are categorical variables; full variable list and the related explanations could be found in Table 1.1.

Table 1.1: MSCM Study - Demographic and Family Information List

Variable	Explanation
married	marriage status of the mother: 0=other, 1=married
education	mother's education level: 0=high school or less, 1=high school graduate
employed	mother's employment status: 0=unemployed, 1=employed
chlth	child's health status at baseline: 0=very poor/poor, 1=fair, 2=good, 3=very good
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
csex	child's gender: 0=male, 1=female
housize	size of the household: 0=2-3 people, 1=more than 3 people

Mothers were requested to keep a health diary about their stress and their children's morbidity status after the baseline interview for 28 days. For both variables, while 0 indicates absence of stress/illness, 1 indicates presence of stress/illness (Table 1.2).

Table 1.2: MSCM Study - Health Diary Variables

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

Almost 67% of the mothers were employed and 48% of them were married. While 14% of them had poor or fair level of health at baseline, 86% had at least good level of health. Additionally 55% of them were at least high school graduate. Furthermore, only 9 (5%) of the children were at low health level (poor or fair); 22 (13%) of them were at good level of health and 136 (almost 82%) of them were at very good or excellent level of health at baseline. Besides, 45% of the children were white. Lastly, while 67% of the related families contained at least 3 people, 33% of them contain less than 3 people in the household. A portion of MSCM data is placed in Table 1.3.

Some possible questions of interest that might be answered by the MSCM study would be:

- Do the children of employed mothers tend to be more likely to have illness compared to the ones of the unemployed mothers?
- Are less educated mothers more likely to be stressed compared to the ones with higher education levels?

Table 1.3: A Portion of MSCM data.

ID	Day	Stress	Illness	Married	Education	Employed	Chlth	Mhlth	Race	Csex	Housize
1021	1	0	0	0	1	0	2	2	1	1	0
1021	2	0	0	0	1	0	2	2	1	1	0
1021	3	0	0	0	1	0	2	2	1	1	0
1021	4	0	0	0	1	0	2	2	1	1	0
1021	5	0	0	0	1	0	2	2	1	1	0
1021	6	0	0	0	1	0	2	2	1	1	0
1021	7	0	NA	0	1	0	2	2	1	1	0
1021	8	0	0	0	1	0	2	2	1	1	0
1021	9	0	0	0	1	0	2	2	1	1	0
1021	10	0	0	0	1	0	2	2	1	1	0
1021	11	0	NA	0	1	0	2	2	1	1	0
1021	12	1	0	0	1	0	2	2	1	1	0
1021	13	0	0	0	1	0	2	2	1	1	0
1021	14	0	NA	0	1	0	2	2	1	1	0
1021	15	0	0	0	1	0	2	2	1	1	0
1021	16	0	0	0	1	0	2	2	1	1	0
1021	17	0	0	0	1	0	2	2	1	1	0
1021	18	0	0	0	1	0	2	2	1	1	0
1021	19	0	0	0	1	0	2	2	1	1	0
1021	20	0	0	0	1	0	2	2	1	1	0
1021	21	0	0	0	1	0	2	2	1	1	0
1021	22	0	NA	0	1	0	2	2	1	1	0
1021	23	0	1	0	1	0	2	2	1	1	0
1021	24	0	0	0	1	0	2	2	1	1	0
1021	25	0	0	0	1	0	2	2	1	1	0
1021	26	0	0	0	1	0	2	2	1	1	0
1021	27	0	0	0	1	0	2	2	1	1	0
1021	28	0	0	0	1	0	2	2	1	1	0

Note: Only the data for the family with ID = 1021 are presented here. The full dataset consists of 166 more families' data in a stacked to each others' end form. NA represents missing data.

- Do the mothers of large sized houses be more likely to be stressed compare to the ones of less sized houses?
- How does the stress status of mothers change over time given her education level?

Since the data were collected from same mother-children pairs over 28 days, the study is a longitudinal study (with regular and common time points, hence a fully constraint data). Therefore, stress status of a mother at a specific day (say, t) would be associated with the past stress status of her (say, $t\text{-}1$, $t\text{-}2$ ...). Likewise, the illness status of a child at a specific day would be associated with the past illness status of him/her. These associations correspond to within-subject association or serial dependence. Additionally, stress status of a mother at time t is associated with illness status of her child at the same time point. This type of association corresponds to multivariate response dependence at a specific time point. These two associations could be depicted by Figure 1.1. Some of these associations can be extremely weak, e.g. association between stress at days 28 and 1; hence, they can be neglected.

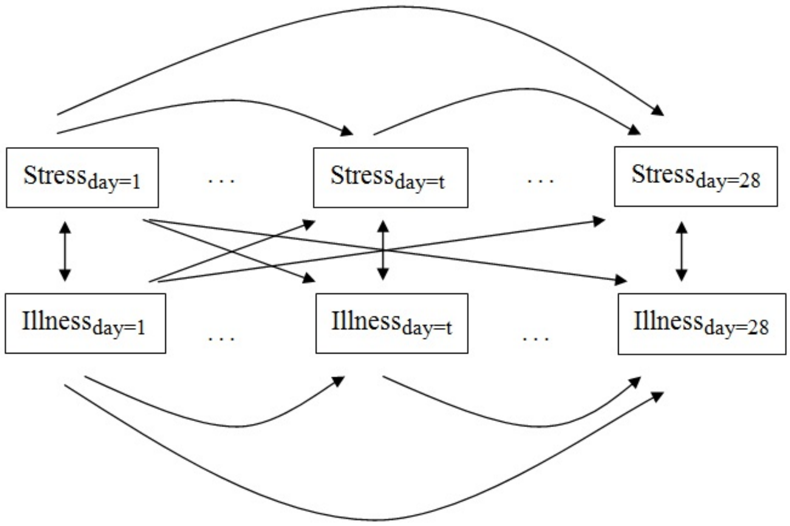


Figure 1.1: Association Structures in Mother's Stress and Children's Morbidity Data

In MSCM study, while the variables which supply demographic and family information of the respondents (Table 3.5) have no missing observation, the variables obtained by the 4-week health diary (Table 1.2) have very low percent of missing values, 0.97% and 1.42% for stress and illness, respectively. Interested reader may refer to Alexander and Markowitz (1986), Zeger and Liang (1986) and Chapter 12 of Diggle et al. (2002) for the full details of the MSCM study. The data could be reached from Heagerty (2002a).

1.3.2 Iowa Youth and Families Project

Iowa Youth and Families Project (IYFP) aimed to investigate the long term effects of the farm crisis, which began in 1980's in America, on family life of the people living in rural parts of America. 451 families from eight rural parts of north central Iowa were selected. These families had the characteristics of including 7th graders with two alive and biological parents and with a sibling aged up to 4 years. It was started in 1989 and conducted yearly until 1992. It was continued at 1994, 1995, 1997 and 1999. At each year, both the parents and the children were surveyed. In the beginning of the study, those young people were at the average age of 12.7 years and 48% (215/451) of them were male (Ilk, 2008). Young people were followed during their adolescent period as well by this 11-year follow-up. Instead of marginally investigating the effects of the aforementioned farm crisis on the emotional status of young people, the study focused on effects of it on family life and indirectly on the children such as harsh parenting due to having economic hardship.

The emotional status of young people were measured by three main distress measurements which are anxiety, hostility and depression. These variables were collected by a symptom check list and dichotomized with respect to having at least one of the distress symptoms. These symptoms include nervousness, shakiness, an urge to break things and feeling low in energy etc. Additionally, some variables thought to be determinative on these emotional variables such as gender of the children, degree of negative life events (such as having a close friend left permanently) and financial cutbacks (such as moving to a cheaper residence) and facing with any negative economical events (such as being had to replace the jobs for a worse one) in the last 12 months (Table 1.4). Moreover, a portion of IYFP data is presented in Table 1.5.

Table 1.4: IYFP Study - A portion of the collected information

Variable	Explanation
anxiety	whether the young person had symptoms: 0=absence, 1=presence
hostility	whether the young person had symptoms: 0=absence, 1=presence
depression	whether the young person had symptoms: 0=absence, 1=presence
gender	gender of the young person: 0=male, 1=female
NLE	whether the young person had any negative life event: 0=none, 1=some, 2=lots of
NEE	whether the household had any negative economical event: 0=no, 1=yes
cut	whether the household had any financial cutbacks: 0=none, 1= between 1 and 5, 2= more than 5

Some possible questions of interest that might be answered by the IYFP study would be:

- Are females more likely to report distress compared to males?
- Do negative life and economic experiences increase the distress level of young people?
- How does the depression status of young people change over time given their negative life event experiences?

IYFP is a longitudinal study (with irregular and common time points, hence a constraint data), since same families were followed across years (with unequal time spaces, i.e. from 1989-1990 and 1992-1995). In our analyses, we considered the portion of this dataset for the period

Table 1.5: A portion of IYFP dataset.

Id	Year	Anxiety	Hostility	Depression	Gender	NLE	NEE	Cut
1	1989	0	0	1	1	1	1	2
1	1990	NA	NA	NA	1	NA	NA	NA
1	1991	NA	NA	NA	1	NA	NA	NA
1	1992	1	1	1	1	1	0	1
1	1994	1	1	1	1	2	1	2
1	1995	1	1	1	1	0	0	1
1	1997	1	1	1	1	1	0	1
1	1999	0	1	1	1	1	1	1
2	1989	1	1	1	1	2	1	2
2	1990	1	1	1	1	2	1	2
2	1991	1	1	1	1	2	1	2
2	1992	1	1	1	1	2	1	2
2	1994	1	1	1	1	0	1	2
2	1995	1	1	1	1	2	1	2
2	1997	1	1	1	1	2	1	1
2	1999	1	0	1	1	1	1	1

Note: Only the families with ID=1 & 2 are presented here. The full dataset consist of 449 more families' data in a stacked to each others' end form. NA represents missing data.

of 1989-1992. This 4-year yearly follow-up allowed us studying with a fully constraint data. Due to being a longitudinal study and including multivariate responses (anxiety, hostility and depression here), it introduces associations both within and between response measurements. It has a more complex form of associations compared to MSCM study, since it has trivariate responses. These associations could be depicted in Figure 1.2. Note that cross response dependencies (dependency between Y_{itj} and $Y_{it'j'}$ for $j = 1, 2, 3$ and $t = 1989, \ldots, 1999$) are not shown in Figure 1.2 for the sake of simplicity. For the illustration of such dependencies, see Figure 1.2.

IYFP dataset has no missing observation at baseline (year=1989). However, it has differing amount of missing observations (both in responses and covariates) in later time points (year $\geq$ 1990). We will discuss the nature of the mechanisms that create those missing data and the methods to handle them in Chapter 4.

We considered a part of IYFP study here in the concept of multivariate longitudinal data, although it has a broader scope. We mainly followed Ilk (2008). For details of it, interested reader may refer to Ilk (2008) and the IYFP related references therein. The data could be reached at Ilk (2011a).

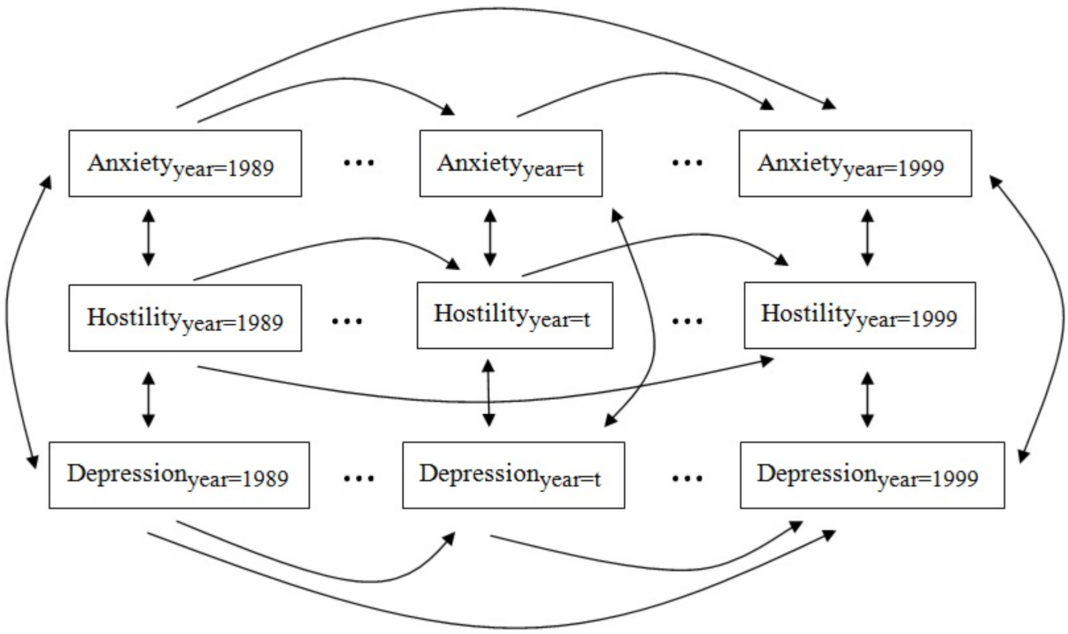


Figure 1.2: Association Structures in Iowa Youth and Families Project Data

1.4 Motivation and Objectives of Our Study

As stated earlier, most of the forecasting studies on longitudinal data are for univariate response and most of them are for continuous response. Also most of them are placed in econometric longitudinal data literature. There is no work for forecasting multivariate longitudinal binary data, hence it is an open area to work on. We take our primary motivation from these facts and consider mainly the biostatistics perspective.

Similar to forecasting studies, there is a lack of literature for modeling multivariate longitudinal binary data. Ilk and Daniels (2007) and Lee et al. (2009) proposed multilevel structured models for such data; Marginalized Transition Random Effects Models (MTREM) and Marginalized Multivariate Random Effects Models (MMREM), respectively. These models are designed to make population-averaged interpretations together with subject-specific interpretations via different levels. In short, they have complex modeling frameworks and use complex algorithms to estimate the parameters, hence they are computationally cumbersome and take more time relative to naive methods. Marginal models with generalized estimating equations (GEE; Liang and Zeger, 1986) have very simple framework and straightforward methodology to fit the parameters. However, they only permit population-averaged interpretation of the regression coefficients. Moreover, they ignore the multivariate response dependency. Fitting multivariate marginal models via GEE were first introduced by Shelton et al. (2004) by a SAS macro. They also only permit population-averaged interpretation but takes into account the multivariate response dependency in addition to the serial dependence. Therefore, they are more valid for multivariate data compared to the univariate counterparts.

In literature, data are mostly generated from the proposed model and based on that data, results of several models are compared. We think that model based data simulation would favor the data generation model. Moreover, in forecasting literature, the covariates are assumed to be known for the forecasting time periods. However, in real life situations, the covariates, similar to the responses, are unknown as well.

One of the key properties of comparative simulation studies to fairly compare different methods is that same simulated datasets should be analyzed with all the methods. Also, in simulation studies when the number of runs is increased more representative results could be obtained due to law of large numbers. We designed the data generation in R, since most of the model codes were written in R except that the code of MTREM were written in FORTRAN. Only the code of multivariate marginal models were available in SAS and parallel connection between R and SAS had to be constructed for the simulation study. However, there is poverty of information to connect those two software. Additionally, although including MTREM were possible in the simulation study, parameter fitting of this model takes a long time which naturally yields a decrease in simulation runs. Moreover, no compensation to any mistake might be done in the simulation setup, such as forgetting to save some of the statistics or a computer crash, since starting from a scratch would be too time consuming and restrictive in terms of our aims. On the other hand, in multivariate modeling literature, the common approach followed while formulating the models is that the covariate effects on multiple responses are different. This approach stresses one to fit different regression parameters for each of the covariate effects. Nonetheless, these effects might be very similar on different responses. In this case, one needs model building flexibilities to get rid the estimation of the redundant parameters. In the light of those facts, our study evolved from just concentrating on forecasting to modeling and software preparation of multivariate longitudinal binary data as well.

In this thesis, the following contributions are done to longitudinal data literature in the modeling, software and forecasting topics:

- An extensive, up to date and simplified literature and discussion were prepared for univariate marginal models, GEE and MMREM. We derived Empirical Bayesian estimator of the random effects coefficients for MMREM and prepared the related R code, since in the original article of the model (Lee et al., 2009), these procedures were not considered.

- The multivariate marginal models (with response specific regression parameters) and the related model fitting algorithm of Shelton et al. (2004) were implemented in R. Although they only considered binary data, we generalized their methods to other response types and illustrated this validity by different datasets. An R package **mmm** was prepared for these multivariate marginal models and investigated in details. This work is under review for a journal at the time of this thesis written. **mmm** will be submitted to Comprehensive R Archive Network (CRAN), i.e. will be publicly available.

- More flexible multivariate marginal models (with shared regression parameters) and the related model fitting algorithms were proposed in this study. An R package **mmm2** (Asar and Ilk, 2012b) was prepared for these multivariate marginal models and investigated in details. We showed that these models are valid for several response families as well. **mmm2** is available from the CRAN at http://CRAN.R-project.org/package=mmm2.

- Inheriting the original modeling framework of MTREM, a different version of it in terms of the link functions and parameter estimation procedure was proposed. Mainly, this new model uses probit links and MLE. The use of implicit function theorem in marginalized models was introduced as well. It has been shown that the new version of the model is superior compared to the original one in terms of computation time consumed for model fitting. An R package **pnmtrem** was prepared for the proposed model and illustrated in details. **pnmtrem** will be submitted to CRAN.

- Forecasting multivariate longitudinal binary data were introduced by a simulation study and real life data considering naive and complex models. In the simulation study, a model independent multivariate longitudinal binary data were considered. Moreover, in the forecasting studies forecasting of the covariates were considered as well.

1.5 Organization

Each chapter of this thesis could be read separately with their own motivation, literature review and discussion sections. In Chapter 2, we introduce the univariate and multivariate longitudinal data structures, provide extensive discussion and literature review for univariate marginal models, generalized estimating equations (GEE) and Marginalized Multivariate Random Effects Models (MMREM) and propose empirical Bayesian estimation of the random effects coefficients for MMREM. In Chapter 3, we introduce two different multivariate marginal models and the related software, the R packages **mmm** and **mmm2**. Additionally, these models and software are illustrated by three different datasets in Chapter 3. Chapter 4 introduces Probit Normal Marginalized Transition Random Effects Models (PNMTREM) together with simulation studies and Iowa Youth and Families Project (IYFP) data application. In Chapter 4, we also introduce the R package **pnmtrem** prepared for MTREM. In Chapter 5, we discuss about forecasting methodologies and accuracy measures for binary data. While Chapter 6 provides the comparison of forecasting abilities of the models via application of Mother's Stress and Children's Morbidity (MSCM) data, Chapter 7 provides a comparative simulation study on forecasting multivariate longitudinal binary data. We close this thesis by Chapter 8, mainly by providing discussion and conclusion regarding the whole study. A list of abbreviations used in this thesis is presented in Appendix A.

CHAPTER 2

LONGITUDINAL DATA STRUCTURES AND AVAILABLE MODELS

In this chapter, we discuss longitudinal data structures and two of the models we consider in this thesis. The chapter is divided into four parts. In Section 2.1, we start by motivating our study. In Section 2.2, we discuss univariate and multivariate longitudinal data structures which are beneficial to understand the methodologies introduced throughout the thesis. In Section 2.3, we continue by univariate marginal models (UMM) together with the quasi-likelihood theory and generalized estimating equations (GEE). Then, we introduce one of the complex models we consider in this thesis, namely Marginalized Multivariate Random Effects Models (MMREM) and propose empirical Bayes estimation of random effects coefficients for that model.

2.1 Motivation

One of the main objectives of this study is to forecast multivariate longitudinal binary response data. Models for such data are the natural tools for forecasting. Modeling multivariate longitudinal binary data has many challenges in addition to the ones that arise in univariate longitudinal data. Since the form of probability distribution function of multivariate longitudinal binary responses could not be written explicitly, it is completed by conditional structures such as transition and/or random effects together with the usual generalized linear modeling structures. A fantastic approach is to marginalize these complements, which means putting those different structures in different levels of the models, i.e., multilevel models. Although multilevel models are superior in terms of inference by permitting both population and subject specific inferences, they have complicated computational structures which naturally means that they require more effort and patience and take more time to fit the parameters. Additionally, forecasting procedures with these models are complicated as well, since they require extra forecasting of the time specific parameters and use of numerical methods. A natural question would be whether it is necessary to use complex models for forecasting. A good way to find satisfactory answers to this question is to compare complex and simple methods in terms of forecasting abilities with suitable accuracy measures. Univariate marginal models with generalized estimating equations (GEE) are robust, simple, flexible and easy to implement methods. Two drawbacks of those models when applied to multivariate longitudinal binary data is that (1) they ignore the multivariate dependence structure, and (2) they permit only population-averaged inference. A relatively complex but still simple method alternative to these models is multivariate marginal models with GEE. In addition to the univariate ones, they accommodate the multivariate response dependency, but still permit only population-averaged inference.

Among others, UMM together with GEE and MMREM are discussed in this chapter due to the fact that they are “available” in literature and ready to use. Nonetheless, here we shall note that empirical Bayes estimation of random effects coefficients for MMREM was not available in the original article (Lee et al., 2009). In this study, we propose Empirical Bayes estimation of random effects coefficients for MMREM. Other models are introduced in later chapters.

2.2 Longitudinal Data Structures

Before formulating the models, let’s illustrate the longitudinal data structures briefly. First we consider the structures for univariate longitudinal data, then we continue with their multivariate counterparts.

Suppose that there are N subjects for whom data is collected repeatedly over time. Then, Y_i corresponds to an $n_i \times 1$ response vector of the i^{th} subject where Y_i can be illustrated by

$$Y_i = \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \\ \vdots \\ Y_{i,n_i} \end{pmatrix} \quad , \; i{=}1, \ldots, N$$

Note that the numbers of repeated measures for subject i is defined as n_i, since it is assumed that they can vary from subject to subject. Also, it is assumed that the time at which measurements are taken might be uncommon.
Additionally, X_i is an associated $n_i \times p$ covariate matrix where p is the number of covariates and X_i can be illustrated by

$$X_i = \begin{pmatrix} X_{i11} & X_{i12} & \cdots & X_{i1p} \\ X_{i21} & X_{i22} & \cdots & X_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{in_i1} & X_{in_i2} & \cdots & X_{in_ip} \end{pmatrix} \quad , \ i\text{=}1, \ldots , N$$

where X_{itp} corresponds to the p^{th} covariate for subject i at time t. Note that the elements of X_i might be changing with time or not changing with time.

Multivariate counterparts are an extension of these data structures. Note that we continue assuming the existence of N subjects. This time Y_i corresponds to a $n_i \times k$ response matrix of the i^{th} subject where Y_i can be illustrated by

$$Y_i = \begin{pmatrix} Y_{i,1,1} & \cdots & Y_{i,1,k} \\ Y_{i,2,1} & \cdots & Y_{i,2,k} \\ \vdots & \ddots & \vdots \\ Y_{i,n_i,1} & \cdots & Y_{i,n_i,k} \end{pmatrix} \quad , \ i\text{=}1, \ldots , N$$

The associated covariate matrix for the multiple responses are similar to the one of the univariate case; X_i is an associated $n_i \times p$ covariate matrix where p is the number of covariates and X_i can be illustrated by

$$X_i = \begin{pmatrix} X_{i11} & X_{i12} & \cdots & X_{i1p} \\ X_{i21} & X_{i22} & \cdots & X_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{in_i1} & X_{in_i2} & \cdots & X_{in_ip} \end{pmatrix} \quad , \ i\text{=}1, \ldots , N$$

where X_{itp} corresponds to the p^{th} covariate for subject i at time t. Note that the elements of X_i can be changing across time or not.

The assumption of varying number of repeated measures for subjects, n_i, is valid for the models we consider in this thesis except for probit normal marginalized transition random effects models (PNMTREM). However, while both marginalized multivariate random effects models (MMREM) and PNMTREM assume the data are collected at equispaced time intervals, multivariate marginal models (MMM) do not have such an assumption.

MMREM and MMM consider same set of covariates for each multiple responses and estimate different regression parameters; these will be discussed and illustrated in Chapters 2 and 3, respectively. Besides, MTREM consider a pooled covariate matrix and estimate shared regression parameters across responses. Nonetheless, transition to the case of separate regression parameters for every multiple responses just like the ones in MMREM and MMM is possible by adding response types as covariates (indicator variables) and interaction(s) of them with other covariates as well; this will be discussed and illustrated in Chapter 4, in details. Note that the specification of n_i indicates that the number of repeated measures are not same for each subject and is a general representation. This might be due to the design of the study (unconstrained design) and/or missing data. These two reasons should not be confused since while the former is intentional, the latter is unintentional and such a confusion might create serious bias, inefficiency and inconsistency.

2.3 Univariate Marginal Models

In this section, marginal models for univariate response is discussed. It is divided into two subsections. In the first part, the modeling framework and the features of marginal models are discussed. In the second part, the quasi-likelihood approach, which forms the basis of parameter estimation, and generalized estimating equations (GEE) are discussed.

2.3.1 Marginal Models

Marginal models are the extension of generalized linear models (GLM) to longitudinal data analysis. They permit regressing the effect of the covariates on the mean response directly, i.e. covariate effects are not conditioned on random effects and/or previous history of the responses. Therefore, the interpretation of the regression parameters does not depend on the specification of within-person association. Hence, the models are termed as marginal models (Fitzmaurice et al., 2004). In regression analysis of longitudinal data, the association between the repeated outcomes is usually of secondary interest. Nevertheless, to make proper inferences, it should be taken into account. In marginal models, the effect of the covariates and within-subject association are modeled separately. They are suitable when the scientific interest is on population-averaged response rather than on individual effects and permit comparing sub-groups of the covariates (Diggle et al., 2002). Marginal models are widely used in epidemiological and health studies in which the interest is to compare sub-groups in terms of the population-averaged response rather than individual level (Fitzmaurice et al., 2004; Diggle et al., 2002).

The formulation of marginal models is given by

$$g(\mu_{it}) = X_{it}\beta \tag{2.1}$$

where μ_{it}=$E(Y_{it}|X_{i1}, \ldots, X_{it})$ is the mean response depending on the covariates for subject i at time t. Specifically, it is assumed that $E(Y_{it}|X_{i1}, \ldots, X_{it})$=$E(Y_{it}|X_{it})$ which means that on the mean outcome at time t the effect of covariate at time t is considered; $g(.)$ is a link function which linearizes the effect of the covariates on the mean response and β is a $p \times 1$ vector of regression parameters.

Modeling diagram of marginal models can be depicted by Figure 2.1. The within-subject association parameters, α will be more transparent when the parameter estimation procedure is introduced (see Section 2.3.2).

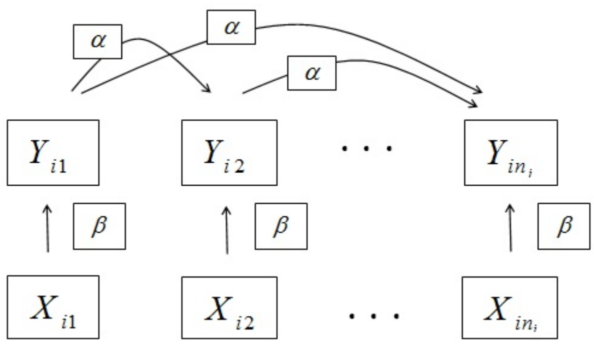


Figure 2.1: Modeling Schemes of Marginal Models

To illustrate the marginal models, let's consider a real life data called Mother's Stress and Children's Morbidity Study (MSCM) which was introduced in Chapter 1. For simplicity, let's consider only one response and one covariate which are maternal stress (0=absence, 1=presence) and employment status of mothers (0=unemployed, 1=employed), respectively. A possible scientific research question that might be answered by marginal models is whether employed mothers more likely to be stressed compared to unemployed ones. The corresponding marginal model is given by

$$\text{logit}P(\text{Stress}_{it} = 1|\text{Employment Status}_{it}) = \beta_0 + \beta_1 * \text{Employment Status}_{it} \tag{2.2}$$

where *logit* is defined as *logarithm* of odds of being stressed; $exp(\beta_0)$ corresponds to the ratio of mothers who are stressed to the mothers who are not for the sub-group of unemployed mothers; $exp(\beta_1)$ corresponds to the odds ratio of being stressed which compares employed to unemployed mothers in terms of being stressed. Note that the meaning and interpretation of the regression parameters of marginal models for the given example is very similar to those of ordinary logistic regression. Besides, this is valid for other types of longitudinal data. Also note that the regression parameters represent the subgroups (employed or unemployed) rather than individuals.

Marginal modeling of longitudinal data has challenges about accommodating the within-subject dependence. For continuous response, multivariate normal distribution assumption about the joint distribution of repeated measures might hold that dependence by specifying the variance-covariance matrix. Since the full distribution of the repeated measures is defined, maximum likelihood approach for parameter estimation is available. However, for discrete data, specification of the joint distribution of them is not that straightforward, since there is no counterpart of multivariate normal distribution for discrete data (Liang and Zeger, 1986; Fitzmaurice and Laird, 1993). An alternative approach is *generalized estimating equations* (GEE) in which only few assumptions about the distribution of the outcome variables are done and within-subject dependence is captured in the estimating equations by introducing working correlation structure within the variance-covariance matrix of longitudinal responses. GEE is valid both for continuous and discrete responses. Note that GEE is an approach for parameter estimation in marginal models rather than a model (Molenberghs and Kenward, 2010). It alternates maximum likelihood estimation (MLE). For detailed discussion about approaches for parameter estimation in marginal models for discrete response, see Diggle et al. (2002) and Molenberghs and Verbeke (2005). Throughout, marginal models recognized in this section are referred as Univariate Marginal Models (UMM) to distinguish them from the multivariate ones.

2.3.2 Quasi-Likelihood Method and Generalized Estimating Equations

Liang and Zeger (1986) and Zeger and Liang (1986) proposed a class of Generalized Estimating Equations to model both univariate longitudinal continuous and discrete outcomes by extending the quasi-likelihood method of Wedderburn (1974) to correlated data. In other words, it can be said that they enabled extending Generalized Linear Models (GLM) to longitudinal data (Hedeker and Gibbons, 2006). Over the past 25 years, GEE is the leading approach and has the most widely use to model univariate longitudinal data, especially for discrete response (e.g. binary or count) (Fitzmaurice et al., 2009; Molenberghs, 2010). Aforementioned approach facilitates taking into account the longitudinal association while modeling the effect of the covariates on the mean response directly. The within-subject association is captured by the *working correlation matrix*. Under mild regularity conditions of the serial dependence structure, GEE approach yields consistent regression parameter estimates and their variances. It is well known that GEE estimates are only robust under Missing Completely at Random (MCAR) response data, because it is not likelihood based. However, extensions are possible to other missing data mechanisms (see Robins et al., 1995; Rotnitzky et al., 1998).

Before explaining the GEE procedure, it is beneficial to mention about GLM with quasi-likelihood theory briefly. (For details of GLM, see Kutner et al., 2005; Agresti, 2002; McCulloch et al., 2008; McCullagh and Nelder, 1989 and for the details of quasi-likelihood theory, see Wedderburn, 1974; McCullagh and Nelder, 1989). Quasi-likelihood method is an approach to obtain the estimates of regression parameters in GLM. Unlike maximum likelihood estimation, the full distribution of the response is not needed to be specified. However, two assumptions about it are enough which are given below.

- The specification of the relationship between mean of the response variable and the covariates,
- The specification of the relationship between mean and variance of the response variable.

For example, for binary response, the relationship between the response mean and the covariates are specified by equating *logit* of the response mean to covariates multiplied with regression parameters, and the relationship between the mean and variance of the response is specified by setting the mean as μ_i and the variance as $\mu_i(1\text{-}\mu_i)$. Therefore, the method is called as a semi-parametric approach. One great advantage of it is that it permits flexibility about working with various types of data easily, e.g. Poisson, Gamma etc. However, it has major disadvantages such as poverty about model comparison and hypothesis tests. When it was first proposed, foundation was based on cross-sectional data, i.e. single observation was considered for each subject. The aforementioned approach is briefly explained below.

Let μ_i be the mean of the response conditional on the covariates, $E(Y_i|X_i)$, and expressed by

$$g(\mu_i) = X_i\beta \tag{2.3}$$

where X_i is a $1 \times$ p vector of explanatory variables, β is a $p \times 1$ vector of regression parameters and g is the link function. For example, for binary response, the link function is logit.

Moreover, let v_i be the variance of the response conditional on the covariates, $V(Y_i|X_i)$, and expressed by

$$v_i = h(\mu_i)/\phi \tag{2.4}$$

where h is a known variance function. For example, h is μ_i $(1 - \mu_i)$ for binary response. ϕ is a scale parameter, and because the primary interest is on regression parameters it is considered as a nuisance parameter which can be either estimated or considered as to be known. After all, the two assumptions about the distribution of the outcome variable is identified by (2.3) and (2.4), respectively.

The quasi-likelihood regression parameter estimates could be obtained by solving the following system utilizing iteratively re-weighted least squares method (Gentle et al., 2004).

$$U(\beta) = \sum_{i=1}^{N} \frac{\partial \mu_i}{\partial \beta} v_i^{-1} (Y_i - \mu_i) = 0 \tag{2.5}$$

where N is the number of subjects and $\partial\mu_i/\partial\beta$ is the first derivative of the response mean with respect to the regression parameters.

In GEE approach, the relationship between the outcome mean and the covariates is specified by

$$g(\mu_{it}) = X_{it}\beta \tag{2.6}$$

The relationship between the outcome mean and outcome variance is specified by

$$v_{it} = h(\mu_{it})/\phi \tag{2.7}$$

where $v_{it}=Var(Y_{it}|X_{it})$.

In addition to quasi-likelihood method, the covariance structure of the longitudinal responses, $Y_i=(Y_1, \ldots, Y_{n_i})$, should be specified. Let V_i be the *working* covariance matrix of Y_i given by

$$V_i = A_i^{1/2} R_i(\alpha) A_i^{1/2} / \phi \tag{2.8}$$

where A_i is a $n_i \times n_i$ diagonal matrix with $h(\mu_{it})$ as the t^{th} diagonal element. $R_i(\alpha)$ is a $n_i \times n_i$ *working* correlation matrix or *weighting* matrix (Carey et al., 1993) which indeed captures the association between the repeated measures. Note that $R_i(\alpha)$ might vary from subject to subject, and is to be fully specified by s $\times$ 1 vector of parameter, α, which is common for all subjects. $R_i(\alpha)$ is named as working correlation matrix because it does not need to be correctly specified. Even under incorrect specification, the parameter estimates and their variances are consistent. Note that because of $R_i(\alpha)$ and to distinguish from the true covariance matrix of the repeated measures, the covariance matrix of Y_i is called as *working* covariance matrix.

As emphasized earlier, $R_i(\alpha)$ is not expected to be correctly specified. Although GEE approach yields consistent parameter and variance estimates, correct specification of $R_i(\alpha)$ increases the efficiency of them. Empirical investigation of the association structure of repeated measures might help to choose the best *working* correlation structure to analyze the longitudinal data at hand (for details, see Hin and Wang, 2009; Shults et al., 2009; Sabo and Chaganty, 2010; Carey and Wang, 2011). Although there are several choices, only common structures are illustrated here (for other possible choices, see Zeger and Liang, 1986; Ziegler, 2011).

First of them is *independence* correlation structure which is the simplest case. It assumes that the repeated measures are uncorrelated, hence it does not seem to be reasonable. $R_i(\alpha)$ is set to be an $n_i \times n_i$ identity matrix which can be illustrated by

$$cor(Y_{it}, Y_{it'}) = \begin{cases} 1, & \text{if t = t} \\ 0, & \text{if t} \neq t' \end{cases} \tag{2.9}$$

In the setting of independence working correlation structure, no correlation parameter is estimated, i.e. s=0. Additionally, in this setting, it is assumed that the responses measured at different time points have the same variance, i.e. $\sigma_t^2=\sigma^2$ for all t=1, ..., n_i.

Second working correlation structure is *unstructured* correlation structure which is the most complex case. It assumes different correlations for all possible time lags and can be illustrated by

$$cor(Y_{it}, Y_{it'}) = \begin{cases} 1, & \text{if t} = t' \\ \alpha_{tt'}, & \text{if t} \neq t' \end{cases} \tag{2.10}$$

A cautionary note is that this correlation structure can be used if the number of repeated measures of each subjects are same; i.e. $n_i = n$, hence $R_i(\alpha)$=$R(\alpha)$. It has the most computational burden, since n(n-1)/2 correlation parameters (s=n(n-1)/2) should be estimated. Zeger and Liang (1986) reported that estimates obtained by assuming unstructured working correlation structure has the minimum variance compared to the estimates of other working correlation structures. Lastly, in the setting of unstructured working correlation structure it is assumed that the responses measured at different time points have different variances, i.e. $\sigma_t^2 \neq \sigma_{t'}^2$ for $t \neq t'$ $(t, t'$=$1, \ldots, n)$.

Third working correlation structure is *exchangeable (compound symmetry)* working correlation structure which assumes equal correlation between the time points regardless of lag and can be depicted by

$$cor(Y_{it}, Y_{it'}) = \begin{cases} 1, & \text{if t} = t' \\ \alpha, & \text{if t} \neq t' \end{cases} \tag{2.11}$$

Note that in random effects models this correlation structure is assumed. Only 1 correlation parameter (s=1) is estimated. Furthermore, alongside the choice of this structure it is assumed that the variances of responses at different time points are equal, i.e. σ_t^2=σ^2 for all t=1, ..., n_i.

The last one is *autoregressive working* correlation structure which is usually denoted by AR or AR(1) in the case of lag-1 structure. It assumes that as the lag increases the correlation decreases with an exponential decay. The structure can be illustrated by

$$cor(Y_{it}, Y_{it'}) = \begin{cases} 1, & \text{if t} = t' \\ e^{-\alpha|t-t'|}, & \text{if t} \neq t' \end{cases} \tag{2.12}$$

Note that assuming AR (1) correlation structure seems reasonable if the observations are collected at equally-spaced time points. As in the case of exchangeable structure, in AR(1) structure only 1 correlation parameter (s=1) is estimated. Additionally, likewise the independence and exchangeable working correlation structures, in the setting of this structure equality of variances for the responses measured at different time points is assumed.

The parameter estimates of GEE, $\hat{\beta}$, $\hat{\alpha}$, $\hat{\phi}$, could be obtained by solving the following equation which is indeed an extension of (2.5).

$$\sum_{i=1}^{N} D_i^T V_i^{-1}(Y_i - \mu_i) = 0 \tag{2.13}$$

Here, D_i=$\partial\mu_i/\partial\beta$, μ_i'=$(\mu_{i1}, \ldots, \mu_{in_i})$. Note that for simplicity, the subscript t is suppressed.

Although there are other ways to solve (2.13) (see Fitzmaurice et al., 2009), here the method proposed by Liang and Zeger (1986) is outlined. To compute the regression parameters of GEE, an iterative algorithm with two-stage estimation is needed. In every iteration, given the estimates of α and ϕ, the estimates of β are improved by Modified Fisher Scoring Algorithm given by

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} - \left[\sum_{i=1}^{N} \left(D_i^{(m)}\right)^T \left(V_i^{(m)}\right)^{-1} D_i^{(m)}\right]^{-1} \left[\sum_{i=1}^{N} \left(D_i^{(m)}\right)^T \left(V_i^{(m)}\right)^{-1} (Y_i - \hat{\mu_i}^{(m)})\right] \tag{2.14}$$

where $D_i^{(m)} = D_i\left(\hat{\beta}^{(m)}\right)$, $V_i^{(m)} = V_i\left(\hat{\beta}^{(m)}, \hat{\alpha}^{(m)}, \hat{\phi}^{(m)}\right)$, $\hat{\mu_i}^{(m)}$=$\mu_i\left(\hat{\beta}^{(m)}\right)$.

Then given the estimates of β, the estimates of α and ϕ are calculated by the Method of Moment Estimation (MME) by using the Standardized Pearson Residuals given by

$$e_{it'} = \frac{Y_{it'} - \hat{\mu}_{it'}}{\sqrt{h\left(\hat{\mu}_{it'}\right)}} \tag{2.15}$$

After all, estimate of ϕ can be calculated by

$$\hat{\phi} = \frac{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{n_i} e_{it}^2}{\sum\limits_{i=1}^{N} n_i} \tag{2.16}$$

and estimate of α can be calculated by

$$\hat{\alpha}_{tt'} = \left(\frac{1}{\hat{\phi}N}\right)\sum_{i=1}^{N} e_{it}e_{it'} \tag{2.17}$$

for unstructured correlation structure which is the most general case for estimation of α components. For estimation of correlation parameters for the other structures see Liang and Zeger (1986).

This two-stage estimation procedure is iterated until convergence. Note that obtaining the initial values of β from usual generalized linear model would decrease the number of iterations.

At convergence, $\hat{\beta}$ is the consistent estimator of the true regression parameters, β, regardless of the choice of the within-subject association structure. The sampling distribution of $\hat{\beta}$ is asymptotically multivariate normal with mean β and

$$cov(\hat{\beta}) = B^{-1}MB^{-1} \tag{2.18}$$

where

$$B = \frac{1}{N}\sum_{i=1}^{N} D_i^T V_i^{-1} D_i \tag{2.19}$$

and

$$M = \frac{1}{N}\sum_{i=1}^{N} D_i^T V_i^{-1} cov(Y_i) V_i^{-1} D_i \tag{2.20}$$

The estimate of the $cov(\hat{\beta})$, $\widehat{cov}(\hat{\beta})$, is obtained by substituting the estimates of β, α, ϕ and $cov(Y_i)$ where $\hat{\beta}$, $\hat{\alpha}$, $\hat{\phi}$ and $(Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)^T$ are the estimates, respectively. This setting of $cov(\hat{\beta})$ is called as *sandwich* estimator where B corresponds to *bread* and M corresponds to *meat* (Fitzmaurice et al., 2004). One great advantage of sandwich estimator of $cov(\hat{\beta})$ is that the results of it are valid under misspecification of the serial dependence structure. Hence, it is a robust estimator of covariance of the regression parameters. Alternatively, $cov(\hat{\beta})$ could be estimated by naive methods given by

$$cov(\hat{\beta}) = B^{-1} \tag{2.21}$$

However, the results of the naive estimators of $cov(\hat{\beta})$ are not robust to that misspecification. A kind of an intuitive way to be sure that the correct within-subject association structure is specified is checking how close the two covariance estimates are.

There is a poverty in model comparison and goodness-of-fit test for marginal models with GEE, albeit appealing features of them. Pan (2001) proposed an information criterion for model selection with GEE called Quasi-Information Criterion (QIC) by extending Akaike Information Criterion (AIC; Akaike, 1973). Nevertheless, the author reported that the QIC performs well only under independence working correlation structure. Eventually, he offered an obscure methodology for model selection which includes firstly obtaining the parameter estimates by any desired working correlation structure and then calculating the information criterion by using these estimates under the independence working correlation structure. On the other hand, Pan (2002) proposed two goodness-of-fit statistics which are extensions of Pearson chi-square statistic and unweighted sum of residual squares originally investigated by Hosmer et al. (1997) for ordinary logistic regression. The author came to a conclusion that these two goodness-of-fit statistics are useful under independence working correlation structure. Since the repeated measures are naturally dependent, QIC and aforementioned goodness-of-fit statistics do not seem to well suit to correlated data.

To sum up, GEE is a crucial parameter estimation method to marginal models for longitudinal data. It can be easily applied to various types of responses such as continuous, binary, count etc. Without specifying the full joint distribution of the repeated measures it allows capturing the association between them. However, it needs the correct specifications of the outcome mean and outcome variance conditional on the covariates with known link functions (for the details of link functions for different types of responses, see Ballinger, 2004). It permits obtaining consistent regression parameter and their variance estimates even under misspecification of the serial dependence structure. Although it is known that full likelihood methods are more efficient compared to semi-parametric ones, in many longitudinal studies GEE approach is capable to compete with them (Fitzmaurice et al., 2004). For years, it has been applied extensively in various fields such as medical care (Diggle et al., 2002), clinical trials (Martus et al., 2004), political sciences (Zorn, 2001), educational and behavioral sciences (Ghisletta and Spini, 2004) and public health (Fitzmaurice et al., 2004) etc. The implementation of GEE is straightforward and available in many statistical softwares such as R (R Core Development Team, 2011) (packages: **gee** by Carrey, 2011a; **geepack** by Højsgaard et al., 2005; **yags** by Carrey, 2011b), SAS (**proc genmod** by SAS Institute Inc, 2009), STATA (StataCorp LP, 2011), MATLAB (**GEEQBOX** by Ratcliffe and Shults, 2008) and etc.

Note that throughout, in the sense of Fitzmaurice et al. (2004) the term *association* is preferred rather than *correlation* while referring serial dependence of the repeated measures. For detailed discussion of GEE, see the great books of Hardin and Hilbe (2003) and Ziegler (2011).

2.4 Marginalized Multivariate Random Effects Models

In this section, a marginalized model for multivariate longitudinal binary data is discussed. The discussion is separated into two parts. In the first part, the model and its features are presented. Additionally, the parameter estimation of the model via maximum likelihood estimation is discussed in the second part.

2.4.1 Model

Lee et al. (2009) proposed a marginalized model structure for multivariate longitudinal binary data which is a framework of two logistic regression models. They extended Marginally Specified Logistic Normal Models of Heagerty (1999) and Ordinal Marginalized Random Effects Models of Lee and Daniels (2008) to multivariate longitudinal binary data by introducing a new variance-covariance structure for the random effects distribution with a Kronecker way of decomposition. Both within-subject association and multivariate response dependence at a given time point are captured by the random effects in the second level. Parameter estimation was held by maximum likelihood estimation which utilizes quasi-Newton Algorithm and quasi-Monte Carlo Methods to numerically solve the likelihood equations and the integrals over the random effects distribution, respectively.

The model formulation of the proposed model, Marginalized Multivariate Random Effects Models (MMREM), is given below.

$$\text{logit}P(Y_{itj} = 1|X_{it}) = X_{it}^T\beta_j \tag{2.22}$$

$$\text{logit}P(Y_{itj} = 1|b_{itj}, X_{it}) = \Delta_{itj} + b_{itj} \tag{2.23}$$

In (2.22)-(2.23), Y_{itj} denotes the j^{th} binary response (j=1, ..., k) for i^{th} person (i=1, ..., N) at time t (t=1, ..., n_i). β_j denotes response specific marginal mean parameters which represents the population-averaged effect of covariates on the mean response. Δ_{itj} are subject/time/response specific intercepts which account for the non-linear relationship between the 1^{st} (2.22) and 2^{nd} (2.23) levels of the model (Ilk and Daniels, 2007). b_{itj} are subject/time/response specific random effects parameters which account for individual level correlations, i.e serial dependence and multivariate response dependence. X_{it} is the corresponding set of covariates which is assumed to be common for the multivariate responses.

It is assumed that Y_{itj} are conditionally independent given $b_{i.j}$ where $b_{i.j}$=(b_{i1j}, ..., b_{in_ij}) which indicates the independence of the repeated measures. Similarly, it is assumed that Y_{itj} and $Y_{itj'}$ are conditionally independent for all j=1, ..., k given $b_{i.j}$ and $b_{i.j'}$ which indicates the independence of multivariate responses. Also it is assumed that responses for different subjects are independent. In addition, it is posited that b_i=(b_{i11}, ..., b_{in_i1}, ..., b_{i1k}, $b_{in_ik})^T$ ~ i.i.d.N(0,Σ), where 0 is a zero vector with length of $n_i \times k$ and Σ is an $(n_i \times k) \times (n_i \times k)$ matrix.

The variance-covariance matrix of the random effects, Σ, is defined as Σ=$\Sigma_1 \otimes \Sigma_2$ with

$$\Sigma_1 = \begin{pmatrix} 1 & e^{-\alpha} & \cdots & e^{-\alpha(n_i-1)} \\ e^{-\alpha} & 1 & \cdots & e^{-\alpha(n_i-2)} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\alpha(n_i-1)} & e^{-\alpha(n_i-2)} & \cdots & 1 \end{pmatrix} \quad , \quad \Sigma_2 = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1k} & \sigma_{2k} & \cdots & \sigma_{kk} \end{pmatrix}$$

where $\otimes$ corresponds to the Kronecker product. Σ_1 is a $n_i \times n_i$ covariance matrix which accounts for the serial dependence of the repeated measures having AR(1) correlation structure, and α is the transition parameter which determines the structure of Σ_1. Σ_2 is a $k \times k$ covariance matrix for the multivariate responses at a given time. Note that for multivariate response dependence, constant covariance structure is assumed for all time points. Consequently, the random effects parameters, b_{itj}, accommodate the two types of associations which arise while working with the multivariate longitudinal data.

To illustrate how the associations are captured within a single covariance structure yielded by the Kronecker product of two covariance matrices (for details see Harville, 1997), the following equation can be considered.

$$cov(b_{itj}, b_{it'j'}) = \sigma_{jj'}\rho^{|t-t'|} \tag{2.24}$$

for all j=1, ..., k; and t=1, ..., n_i. In (2.24), when j=j' the within-subject correlation for a specific response type is explained. Additionally, when t=t' the multivariate response dependence at a specific time point is explained by the random effects parameters. Modeling scheme of MMREM can be depicted by Figure 2.2 where for the sake of simplicity, the bivariate response case is considered. Although not shown in Figure 2.2 for the sake of simplicity, MMREM accommodates the association between different responses at different time points (cross-response dependency) explicitly (via random effects), i.e. association between Y_{itj} and $Y_{it'j'}$ for $t \neq t'$.

As emphasized earlier, the levels of the model are not independent. However, 1^{st} level of the model is the marginalized version of the 2^{nd} level over the random effects distribution and the non-linear relationship of the levels are captured by the intercept parameters, Δ_{itj}. Hence, Δ_{itj} are the functions of both the marginal mean parameters and random effects variance and can be obtained by the following equation.

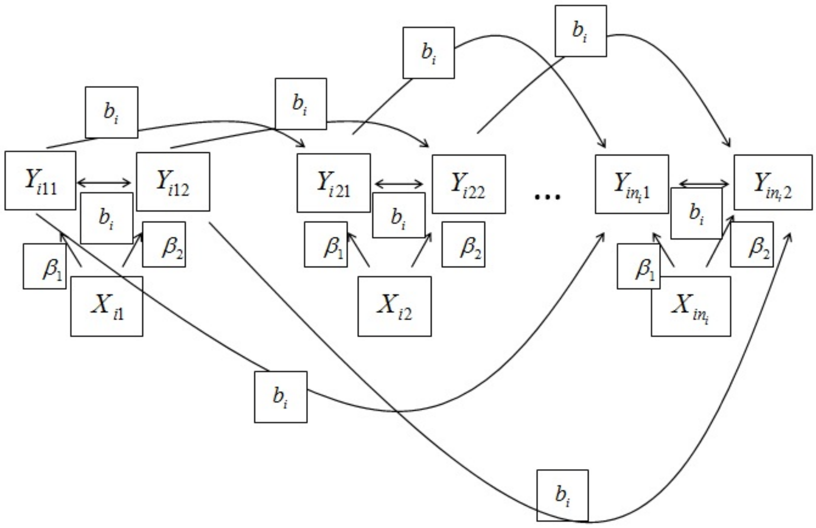


Figure 2.2: Modeling Schemes of MMREM where b_i=($b_{i11}, \ldots, b_{in_ik}$)

$$P(Y_{itj} = 1|X_{it}) = \int P(Y_{itj} = 1|b_{itj}, X_{it}) f(b_{itj})\, db_{itj} \tag{2.25}$$

where f(.) is a univariate normal distribution with mean 0 and variance var(b_{itj}). To obtain the Δ_{itj} given β_j and σ_j, (2.25) is solved by Newton-Raphson Algorithm along with 40-points Gauss-Hermite Quadrature. (For details see the Appendix of Lee et al., 2009). This procedure is repeated for all i=1, ..., N, t=1, ..., n_i and j=1, ..., k in every iteration of quasi-Newton Algorithm. Note that quasi-Newton Algorithm will be introduced in Section 2.4.2.

The orthogonalization of the random effects by setting $b_i = \Sigma_1^{1/2} \otimes \Sigma_2^{1/2} z_i$, where $\Sigma_1^{1/2}$ and $\Sigma_2^{1/2}$ are lower triangular matrices with positive diagonal elements, the Cholesky factors of Σ_1 and Σ_2, respectively, and z_i are the subject specific independent standard normal random variables. This orthogonalization yields both computational ease and time effectiveness in terms of parameter estimation. The new covariance structures takes the following forms.

$$\Sigma_1^{1/2} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ e^{-\alpha} & \sqrt{1-e^{-2\alpha}} & 0 & \cdots & 0 \\ e^{-2\alpha} & e^{-\alpha}\sqrt{1-e^{-2\alpha}} & \sqrt{1-e^{-2\alpha}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{-\alpha(n_i-1)} & e^{-\alpha(n_i-2)}\sqrt{1-e^{-2\alpha}} & e^{-\alpha(n_i-3)}\sqrt{1-e^{-2\alpha}} & \cdots & \sqrt{1-e^{-2\alpha}} \end{pmatrix}$$

and

$$\Sigma_2^{1/2} = \begin{pmatrix} s_{11} & 0 & 0 & \cdots & 0 \\ s_{12} & s_{22} & 0 & \cdots & 0 \\ s_{13} & s_{23} & s_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{1k} & s_{2k} & s_{3k} & \cdots & s_{kk} \end{pmatrix}$$

Note that Cholesky factorizations of the matrices, Σ_1 and Σ_2 are defined as Σ_1=$(\Sigma_1^{1/2})(\Sigma_1^{1/2})^T$ and Σ_2=$(\Sigma_2^{1/2})(\Sigma_2^{1/2})^T$, respectively.

Consequently, the re-parameterized association model (2^{nd} level of the model) of MMREM takes the following form:

$$\text{logit}P(Y_{itj} = 1|z_i, X_{it}) = \Delta_{itj} + c^{k(t-1)+j} z_i \tag{2.26}$$

where $c^{k(t-1)+j}$ is the $(k(t-1)+j)^{th}$ row vector of the Cholesky factor of Σ, $\Sigma^{1/2}$, which is computed as the Kronecker product of $\Sigma_1^{1/2}$ and $\Sigma_2^{1/2}$, i.e. $\Sigma^{1/2}$=$\Sigma_1^{1/2} \otimes \Sigma_2^{1/2}$; and $z_i \sim$ N(0, I). A great advantage of this re-parametrization is that it permits more accurate estimation in terms of near-zero variance parameters Lee et al. (2009).

2.4.2 Estimation

The likelihood function over the re-parametrized association level of MMREM is given by

$$L(\theta; y) = \prod_{i=1}^{N} \int \prod_{t=1}^{n_i} \prod_{j=1}^{k} (P(Y_{itj} = 1|z_i, X_{it})^{y_{itj}} (1 - P(Y_{itj} = 1|z_i, X_{it})^{1-y_{itj}} \phi(z_i)\, dz_i \tag{2.27}$$

where $\phi(.)$ is a multivariate standard normal distribution function with mean vector 0 length of $n_i \times k$ and variance-covariance matrix I of dimension $(n_i \times k) \times (n_i \times k)$ and θ=(β, s, α). The right term to the integral corresponds to the conditional independence of the repeated measurements and multivariate responses given the random effects, respectively. These two types of associations are accommodated by the integral which is taken over the random effects distribution; re-parametrized random effects, $\phi(.)$ here. The joint distribution function of the subjects are represented by the outmost product, which follows from their independence.

Maximized log-likelihood equations pertaining the parameters, θ, have the following form:

$$\sum_{i=1}^{N} \frac{\partial logL(\theta; y_i)}{\partial \theta} = \sum_{i=1}^{N} L^{-1}(\theta; y_i) \int \frac{\partial logL(\theta, z_i; y_i)}{\partial \theta} \phi(z_i)\, dz_i = 0 \tag{2.28}$$

where

$$L(\theta; y_i) = \int \prod_{t=1}^{n_i} \prod_{j=1}^{k} (P(Y_{itj} = 1|z_i, X_{it})^{y_{itj}} (1 - P(Y_{itj} = 1|z_i, X_{it})^{1-y_{itj}} \phi(z_i)\, dz_i \tag{2.29}$$

and

$$L(\theta, z_i; y_i) = \prod_{t=1}^{n_i} \prod_{j=1}^{k} (P(Y_{itj} = 1|z_i, X_{it})^{y_{itj}} (1 - P(Y_{itj} = 1|z_i, X_{it})^{1-y_{itj}} \tag{2.30}$$

For details of the first derivatives of the log-likelihood function with respect to θ including the derivatives of the Δ_{itj} parameters, see Appendix of Lee et al. (2009).

The maximized log-likelihood equations are not in the explicit forms. Therefore, numerical methods should be applied. Since deriving the second derivatives of the likelihood equations is not trivial, Newton-Raphson Algorithm cannot be used. Fortunately, quasi-Newton Algorithm does not need the second derivatives, instead works with the first derivatives and has the following form.

$$\theta^{g+1} = \theta^g + [I_e(\theta^g; y)]^{-1} \frac{\partial logL}{\partial \theta^g} \tag{2.31}$$

where θ^g corresponds to the parameter estimates at the g^{th} step which are improved in each step and $I_e(\theta)$ is the sample empirical covariance matrix calculated by

$$I_e(\theta; y) = \sum_{i=1}^{N} \frac{\partial L(\theta; y_i)}{\partial \theta} \frac{\partial L(\theta; y_i)}{\partial \theta^T} \tag{2.32}$$

Here, $I_e(\theta)$ is a consistent estimator of the information matrix at the g^{th} step. Because the parameter estimates are obtained by a numerical approach, the explicit form of the variance-covariance matrix of the parameters is not available. Luckily, the inverse of $I_e(\hat{\theta}; y)$ yields the large-sample variance-covariance matrix estimate of the parameter estimates at convergence. For the detailed use of quasi-Newton Algorithm within MMREM see the Appendix of Lee et al. (2009).

The integral over the random effects distribution are approximated by the quasi-Monte Carlo Method specifically Sobol Sequences which ensure optimal sampling from the prior distribution, multivariate standard normal distribution here. For the detailed discussion of selecting an appropriate numerical integration method for MMREM, see section 2.1 of Lee et al. (2009). Following Schildcrout and Heagerty (2007), the parameter estimation of MMREM can be summed up by the following algorithm.

1. Select initial values of θ=(β, s, α). Note that initials of β could be obtained from independent logistic regression models by using the available software to decrease the convergence burden.
2. Get the Δ_{itj} by using (2.25) by the help of Newton-Raphson Algorithm and Gauss Hermite Quadratures.
3. Maximize the log-likelihoods by using the current values of the parameters and Δ_{itj} values by the help of quasi-Monte Carlo Method.
4. Improve the parameter estimates by using quasi-Newton Algorithm ((2.31)-(2.32)).
5. Repeat steps 2-4 until convergence is satisfied, i.e. $\sqrt{(\hat{\theta}^{old} - \hat{\theta}^{new})^T(\hat{\theta}^{old} - \hat{\theta}^{new})} \leqslant$ tolerance. The tolerance is upon the choice of the user. Nonetheless, commonly it is taken as 10^{-3} (Heagerty, 1999), 10^{-4} (Heagerty, 2002b) or 10^{-5} (Lee et al., 2009).

Lee et al. (2009) showed that the parameter estimates are essentially unbiased both in cases of complete data and Missing at Random (MAR) dropout data by a simulation study. They also illustrated their proposed model by a real life data application. Furthermore, behavior of MMREM under informative dropout case were investigated. The model code was written in Fortran along with R and the authors stated that each quasi-Newton steps with 1000 Monte-Carlo Method and 40 point Gauss-Hermite Quadrature takes about 1 min on a PC 1.86 GHz. The model code is available upon request from the authors and was requested for using in this thesis.

2.4.3 Empirical Bayesian Estimation of Random Effects Coefficients

To calculate the individual probabilities such as $P(Y_{itj} = 1|z_i, X_{it}) = \frac{exp(\Delta_{itj}+c^{k(t-1)+j}z_i)}{1+exp(\Delta_{itj}+c^{k(t-1)+j}z_i)}$, we need the estimates of Δ_{itj}, $c^{k(t-1)+j}$ and z_i where z_i and $c^{k(t-1)+j}$ form the random effects b_{itj}, as $b_{itj} = c^{k(t-1)+j}z_i$. Note that Δ_{itj} are deterministic functions of $\theta = (\beta, \alpha, s)$ as given in (2.25) and $\hat{\Delta}_{itj}$ could be obtained by solving (2.25) via Gauss-Hermite Quadratures method and Newton-Raphson root finding algorithm based on the MLE of θ, $\hat{\theta} = (\hat{\beta}, \hat{\alpha}, \hat{s})$. Also, note that $c^{k(t-1)+j}$ is the $(k(t-1)+j)^{th}$ row vector of $\Sigma^{1/2}$ (the Cholesky factor of Σ) which is decomposed as $\Sigma^{1/2} = \Sigma_1^{1/2} \otimes \Sigma_2^{1/2}$ where $\Sigma_1^{1/2}$ and $\Sigma_2^{1/2}$ are formed based on the parameters α

and s. Therefore, $\hat{c}^{k(t-1)+j}$ could be directly obtained by putting the MLE of α and s ($\hat{\alpha}$ and $\hat{s}$, respectively) in $\Sigma^{1/2}$. The only unestimated parameters are the individual characteristics, z_i which together with $c^{k(t-1)+j}$ forms the random effects b_{itj}. Note that z_i's are replaced by the Gauss-Hermite Quadrature nodes, during the maximum likelihood estimation of θ. In the original article of MMREM (Lee et al., 2009), the authors did not consider the estimation of z_i, and did not include the related code. In this study, we consider the estimation of z_i in MMREM: we derive the Empirical Bayesian estimator of z_i and provide the related R code. The detailed procedure is illustrated below.

Given MLE of θ, $\hat{\theta}$, Empirical Bayes estimates of b_{itj}, $\tilde{b}_{itj}$ could be obtained by solving the posterior score equations of z_i (Heagerty, 1999). The posterior distribution of z_i is proportional to the conditional distribution of the observed data, Y_i given z_i, i.e. $[Y_i|z_i]$, times the prior distribution of z_i such that

$$
\begin{aligned}
f(z_i|-) &\propto f(Y_i|z_i)f(z_i) \\
&\propto \left\{ \prod_{t=1}^{n_i} \prod_{j=1}^{k} \left[l(\Delta_{itj} + c^{k(t-1)+j} z_i) \right]^{Y_{itj}} \left[1 - l(\Delta_{itj} + c^{k(t-1)+j} z_i) \right]^{1-Y_{itj}} \right\} e^{-z_i^2/2}
\end{aligned}
$$

where $l(\Delta_{itj} + c^{k(t-1)+j} z_i) = \frac{exp(\Delta_{itj}+c^{k(t-1)+j} z_i)}{1+exp(\Delta_{itj}+c^{k(t-1)+j} z_i)}$.

The estimates of z_i, $\hat{z}_i$, could be obtained as the mode of its posterior distribution based on $\hat{\theta}$ such that equating the first derivative of natural logarithm of the posterior distribution taken with respect to z_i, to 0, and solving the related score equations for z_i. These operations are given below.

$$
log(f(z_i|-)) \propto \left\{ \sum_{t=1}^{n_i} \sum_{j=1}^{k} \left[Y_{itj} log\left(l(\Delta_{itj} + c^{k(t-1)+j} z_i) \right) + (1 - Y_{itj}) log\left(1 - l(\Delta_{itj} + c^{k(t-1)+j} z_i) \right) \right] \right\} - \frac{z_i^2}{2}
$$

$$
\frac{\partial log(f(z_i|\_))}{\partial z_i} \propto \left\{ \sum_{t=1}^{n_i} \sum_{j=1}^{k} \left[\frac{Y_{itj}}{l(\Delta_{itj} + c^{k(t-1)+j} z_i)} l'(\Delta_{itj} + c^{k(t-1)+j} z_i) + \frac{(1 - Y_{itj})}{1 - l(\Delta_{itj} + c^{k(t-1)+j} z_i)} \left(-l'(\Delta_{itj} + c^{k(t-1)+j} z_i) \right) \right] \right\} - z_i = 0
$$

After some simple calculus, we obtain

$$
\frac{\partial log(f(z_i|\_))}{\partial z_i} \propto \left\{ \sum_{t=1}^{n_i} \sum_{j=1}^{k} \left[\frac{\left(Y_{itj} - l(\Delta_{itj} + c^{k(t-1)+j} z_i) \right) l'(\Delta_{itj} + c^{k(t-1)+j} z_i)}{l(\Delta_{itj} + c^{k(t-1)+j} z_i) \left(1 - l(\Delta_{itj} + c^{k(t-1)+j} z_i) \right)} \right] \right\} - z_i = 0
$$

Note that,

$$l'(\Delta_{itj} + c^{k(t-1)+j}z_i) = \frac{c^{k(t-1)+j}l(\Delta_{itj} + c^{k(t-1)+j}z_i)}{1 + exp(\Delta_{itj} + c^{k(t-1)+j}z_i)} \quad \text{and}$$

$$1 - l(\Delta_{itj} + c^{k(t-1)+j}z_i) = \frac{1}{1 + exp(\Delta_{itj} + c^{k(t-1)+j}z_i)}$$

and replacement of them yields

$$\frac{\partial log(f(z_i|\_))}{\partial z_i} \propto \left\{ \sum_{t=1}^{n_i} \sum_{j=1}^{k} \left[\left(Y_{itj} - l(\Delta_{itj} + c^{k(t-1)+j}z_i) \right) c^{k(t-1)+j} \right] \right\} - z_i = 0 \tag{2.33}$$

The estimates of z_i, $\hat{z}_i$, could be obtained by solving the following (2.33) with respect to z_i. Since this does not permit obtaining $\hat{z}_i$ explicitly, some optimization methods such as Newton-Raphson (N-R) root finding algorithm should be employed. To use N-R, we need the second derivative of the log-posterior distribution of z_i with respect to z_i, $\frac{\partial^2 log(f(z_i|\_))}{\partial z_i^2}$ for which the calculation is rather simple compared to the first derivative of the log-posterior distribution. The related calculation is given below.

$$\begin{aligned} \frac{\partial^2 log(f(z_i|\_))}{\partial z_i^2} &= \left\{ \sum_{t=1}^{n_i} \sum_{j=1}^{k} \left[-l'(\Delta_{itj} + c^{k(t-1)+j}z_i) c^{k(t-1)+j} \right] \right\} - 1 \\ &= \left\{ \sum_{t=1}^{n_i} \sum_{j=1}^{k} \left[-\frac{\left(c^{k(t-1)+j}\right)^2 l(\Delta_{itj} + c^{k(t-1)+j}z_i)}{1 + exp(\Delta_{itj} + c^{k(t-1)+j}z_i)} \right] \right\} - 1 \end{aligned}$$

Sample R codes to obtain $\hat{z}_i$ is available from us upon request. The Newton-Raphson is achieved by the `newton` function under the R package **OOmisc** (Asar and Ilk, 2012c).

CHAPTER 3

MULTIVARIATE MARGINAL MODELS

In this chapter, we consider marginal modeling of multivariate longitudinal data and introduce two multivariate marginal models (MMM) and the related software. The chapter is divided into five sections. Since the univariate marginal models (UMM) and generalized estimating equations (GEE) have been discussed in Section 2.3 in details, we skip them here. In Section 3.1, we motivate our study alongside with discussing the need for multivariate modeling of multivariate longitudinal data, specifically in the marginal modeling concept and the need for statistical software for such models. In Section 3.2, details of two MMM and the methodologies to fit them are provided. In Section 3.3, the details of the proposed software, R packages **mmm** and **mmm2** are introduced. Section 3.4 provides the details of the use of those packages via three multivariate longitudinal datasets and discusses the parameter estimates. We close Chapter 3 by conclusions and discussions.

3.1 Motivation

As is mentioned throughout this study, longitudinal data consist of repeated measurements belonging to the same subjects/units. This type of data might arise in many research fields such as sociology, medicine, psychiatry, economy, industry etc. Since the data are collected on the same subjects/units, the repeated measurements are typically not independent. Although the within-subject associations may not be the primary interest of the study (in some studies they are; e.g., Carey et al., 1993), they must be taken into account to make accurate statistical inferences.

The models proposed for longitudinal data might be classified into two: single level and multilevel models. While marginal, random effects and/or transition models are referred as single level models, the marginalized structured models are referred as multilevel models. Marginal models permit population averaged interpretation of the effects of regression parameters on the mean response which is in the same manner with generalized linear models. However, in random effects and transition models such interpretation depends on the subject specific conditions, hence they are called conditional models (Diggle et al., 2002). Conditional models might have challenges in terms of computation and interpretation of the marginal regression parameter estimates. Two seminal articles of Heagerty (1999; 2002b) might be respected as the pioneer works of multilevel models for longitudinal data. Despite the fact that they have many advantages such as robustness of the mean parameters under misspecification of the dependence structure and the availability of information about the within-subject association structure, they have some disadvantages in terms of computational difficulties such as convergence problems and time burden, difficulty in the use of author-written model codes etc. In this study, a specific type of marginal models is considered.

The longitudinal data are often collected on multiple responses. In such a case, in addition to the within-subject association, a second association structure, multivariate response association at a specific time point occurs. The multivariate responses either could belong to the same response family or different response families. In this study, multivariate responses belonging to the same family are considered. In literature, there are two common approaches to handle multivariate longitudinal data rather than joint modeling. First one is to ignore the multivariate response association and construct univariate models (Shelton et al., 2004). The second one is to consider one of the responses as the dependent variable and the other(s) as independent one(s) together with the other independent variables and construct a univariate model (Weiss, 2005). In spite of their practicability, they do not yield valid statistical inferences. In this study, we consider the accommodation of the aforementioned two association structures.

Methods are applied on a real life dataset on the Mother's Stress and Children's Morbidity Study (MSCM; Alexander and Markowitz, 1986) in which two binary responses which are mother's stress (0=absence of stress, 1=presence of stress) and children's morbidity (0=absence of illness, 1=presence of illness) were collected for 28 days (see Section 1.2.1). The effect of mother's employment status and other demographic factors such as child's gender and race, mother's education and marriage status and household size would be examined on the mother's stress and children's illness status. Since two responses are collected over time, the aforementioned two types of associations occur for this study. If the averaged effects of the factors on these responses are of interest, then the use of multivariate marginal models is reasonable. Additionally, simulated data examples are considered to illustrate the validity of multivariate marginal models for the other types of response families rather than binomial.

Generalized estimating equations (GEE; Liang and Zeger, 1986) is an approach for parameter estimation in marginal models for univariate longitudinal data. Since it was proposed, it has been used extensively in many research fields. GEE was not restricted to any specific response family, but it can handle several response families such as binomial, gaussian, Poisson and gamma. Even under incorrect specification of the working correlation structure, it yields consistent parameter and variance estimates. It is available in many statistical software and the use of it is straightforward; **gee** (Carey, 2011a), **yags** (Carey, 2011b) and **geepack** (Højsgaard et al., 2005) packages in R (R Core Development Team, 2011) , **proc genmod** in SAS (SAS Institute Inc, 2009) and **GEEQBOX** toolbox (Ratcliffe and Shults, 2008) in MATLAB (The MathWorks Inc, 2010). However, there is lack of available software and literature for multivariate marginal models. Shelton et al. (2004) considered a specific type of multivariate marginal models for multivariate longitudinal binary data and propose a SAS macro extending the univariate GEE solver, **proc genmod** to fit the proposed models in the formulation of univariate marginal models. Asar and Ilk (2012) considered the availability of the models of Shelton et al. (2004) for other response families as well as binomial, hence generalized the multivariate marginal models and proposed an R function `mmm`, under the R package **mmm** to fit these models by extending the available package **gee**. Ilk and Daniels (2007) considered multilevel modeling of multivariate longitudinal binary data with shared intercept and covariate effects for different multiple responses while still allowing to build models with differing effects of those terms in the same model formulation. This approach permits fitting more parsimonious models, hence it might yield more efficient parameter estimates. In this study, we propose multivariate marginal models which include their modeling approach and propose another R function `mmm2` to fit them. We show that these models are not restricted to any type of response family and the available package **gee** is used to fit them as well. `mmm2` is available in the R package **mmm2** (Asar and Ilk, 2012b) which is available from the Comprehensive R Archive Network (CRAN) at http://CRAN.R-project.org/package=mmm2.

3.2 Models

In the analysis of multivariate longitudinal data, in addition to the within-subject association a second type of association, multivariate response dependence at a specific time point arises. It should be taken into account to make appropriate statistical inferences alongside with the serial dependence. Marginal models for multivariate longitudinal data permit interpretation of the regression parameters free of the dependence structure. Nevertheless, they allow capturing the two type of associations, hence they yield reliable statistical inference. When the question of interest is based on the population-averaged effects of the covariates, marginal modeling is beneficial. In addition, marginal models for multivariate data yields reliable comparison of the effects of covariates on different multivariate responses. In this study, we consider two different multivariate marginal models.

3.2.1 Multivariate Marginal Models with Response Specific Regression Parameters

The model formulation of multivariate marginal models with response specific regression parameters (MMM1; Shelton et al., 2004, Asar and Ilk, 2012) can be constructed as below.

$$g(\mu_{itj}) = X_{it}\beta_j \tag{3.1}$$

Here μ_{itj} is the mean response for the j^{th} response ($j = 1, \ldots, k$) of subject i ($i = 1, \ldots, N$) at time t ($t = 1, \ldots, n_i$) conditioned only on the covariates, i.e., $E(Y_{itj}|X_{it})$; $g(.)$ is a known link function which makes the effects of the covariates on the mean response linear; X_{it} is a $1 \times p$ vector of covariates for subject i at time t, common for different multivariate responses and β_j is a p×1 vector of response specific regression coefficients. As in the case of univariate marginal models, the covariates might be changing with time or not changing with time. The modeling diagram of multivariate marginal models can be illustrated in Figure 3.1; for the sake of simplicity the diagram is drawn for bivariate responses. Although not shown in Figure 3.1 for simplicity, multivariate marginal models explicitly accommodate the association between different responses at different time points, cross-response temporal dependence (via the working correlation matrices), i.e. association between Y_{itj} and $Y_{it'j'}$ for $t \neq t'$ and $j \neq j'$. Note that since both the multivariate response and serial associations are captured within the same working correlation matrix, in Figure 3.1 we suppressed these associations in α.

The methodology to fit MMM1 in the univariate marginal modeling framework is depicted

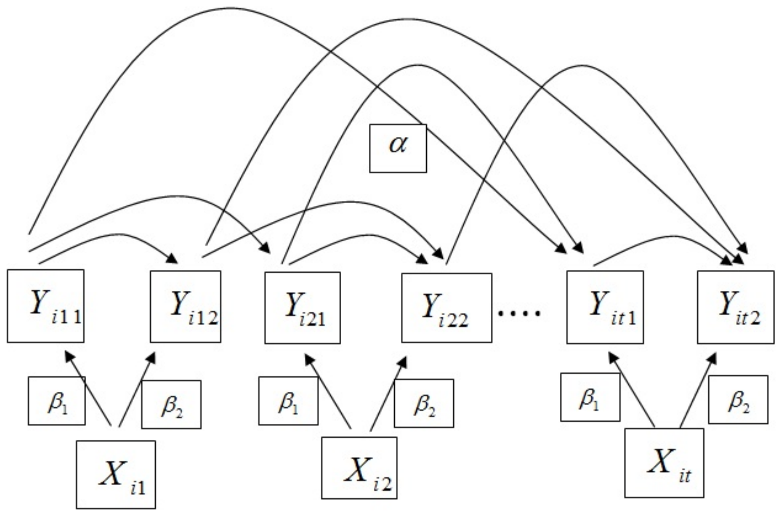


Figure 3.1: Modeling Schemes of Multivariate Marginal Models

below.

Let's assume a longitudinal dataset in which k multivariate longitudinal responses and $p-1$ covariates are available. Also assume that the longitudinal data is collected on N subjects for t time points. The (long) form of the data would be like the one illustrated in Table 3.1. Then, $(N{*}t){\times}k$ multivariate response matrix is manipulated to construct a new response matrix which has the dimension of $(N*t*k)\times 1$. This manipulation involves combining the multivariate responses of each subjects into one column, i.e., $Resp_{new}$=$(Resp_{1,new}, \ldots, Resp_{N,new})^T$ where $Resp_{i,new}$=$(Y_{i.1}, \ldots, Y_{i.k})^T$ and $Y_{i.j}$=$(Y_{i1j}, \ldots, Y_{itj})$. The matrix of covariates is extended from an $(N{*}t){\times}(p{-}1)$ matrix to an $(N{*}t{*}k){\times}((p{-}1){*}k)$ matrix with a Kronecker product operator (Harville, 1997). This manipulation aims to extend the matrix of covariates for multivariate responses, i.e., X_{new}=$(X_{1,new}, \ldots, X_{N,new})^T$ where $X_{i,new}$=$X_i \otimes I_k$ and X_i=$(X_{i.1}, \ldots, X_{i.(p-1)})$, $X_{i.s}$=$(X_{i1s}, \ldots, X_{its})^T$, I_k is an k ⊗ k identity matrix and ⊗ corresponds to a right Kronecker product. The illustration of the form of reconstructed multivariate longitudinal data can be depicted in Table 3.2. The corresponding regression coefficients β_{new} is a $(p*k)\times 1$ vector such as $\beta_{new} = (\beta_1, \ldots, \beta_k)^T$ where β_j=$(\beta_{j0}, \ldots, \beta_{j(p-1)})$. Response specific intercepts are succeeded by placing an $(N{*}t){\times}1$ vector for which all elements are 1 left to the first covariate (X_1) in the original multivariate data form (Table 3.1). Although not shown in Tables 3.1 and 3.2, it is inherited to the appropriate places during the data manipulation process. All the aforementioned data manipulations are achieved within `mmm` automatically. Eventually, the parameter estimations are done just like in the univariate marginal modeling case.

Table 3.1: An illustration of multivariate longitudinal data.

Subject	$Resp_1$	$Resp_2$	$\cdots$	$Resp_k$	Cov_1	Cov_2	$\cdots$	Cov_{p-1}
1	Y_{111}	Y_{112}	$\cdots$	Y_{11k}	X_{111}	X_{112}	$\cdots$	$X_{11(p-1)}$
1	Y_{121}	Y_{122}	$\cdots$	Y_{12k}	X_{121}	X_{122}	$\cdots$	$X_{12(p-1)}$
.	.	.	$\cdots$	.	.	.	$\cdots$	.
1	Y_{1t1}	Y_{1t2}	$\cdots$	Y_{1tk}	X_{1t1}	X_{1t2}	$\cdots$	$X_{1t(p-1)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	Y_{N11}	Y_{N12}	$\cdots$	Y_{N1k}	X_{N11}	X_{N12}	$\cdots$	$X_{N1(p-1)}$
N	Y_{N21}	Y_{N22}	$\cdots$	Y_{N2k}	X_{N21}	X_{N22}	$\cdots$	$X_{N2(p-1)}$
.	.	.	$\cdots$	.	.	.	$\cdots$	.
N	Y_{Nt1}	Y_{Nt2}	$\cdots$	Y_{Ntk}	X_{Nt1}	X_{Nt2}	$\cdots$	$X_{Nt(p-1)}$

Table 3.2: An illustration of reconstructed multivariate longitudinal data for MMM1.

Subject	$Resp_{new}$	Cov_{11}	$\cdots$	$Cov_{1(p-1)}$	$\cdots$	Cov_{k1}	$\cdots$	$Cov_{k(p-1)}$
1	Y_{111}	X_{111}	$\cdots$	$X_{11(p-1)}$	$\cdots$	0	$\cdots$	0
.	.	.	.	.	.	.	.	.
1	Y_{1t1}	X_{1t1}	$\cdots$	$X_{1t(p-1)}$	$\cdots$	0	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	Y_{11k}	0	$\cdots$	0	$\cdots$	X_{111}	$\cdots$	$X_{11(p-1)}$
.	.	.	.	.	.	.	.	.
1	Y_{1tk}	0	$\cdots$	0	$\cdots$	X_{1t1}	$\cdots$	$X_{1t(p-1)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	Y_{N11}	X_{N11}	$\cdots$	$X_{N1(p-1)}$	$\cdots$	0	$\cdots$	0
.	.	.	.	.	.	.	.	.
N	Y_{Nt1}	X_{Nt1}	$\cdots$	$X_{Nt(p-1)}$	$\cdots$	0	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	Y_{N1k}	0	$\cdots$	0	$\cdots$	X_{N11}	$\cdots$	$X_{N1(p-1)}$
.	.	.	.	.	.	.	.	.
N	Y_{Ntk}	0	$\cdots$	0	$\cdots$	X_{Nt1}	$\cdots$	$X_{Nt(p-1)}$

3.2.2 Multivariate Marginal Models with Shared Regression Parameters

The model formulation of multivariate marginal models with shared regression parameters (MMM2) can be constructed as below.

$$g(\mu_{itj}) = X_{itj}\beta \tag{3.2}$$

Here μ_{itj} is the mean response for the j^{th} response ($j = 1, \ldots, k$) of subject i ($i = 1, \ldots, N$) at time t ($t = 1, \ldots, n_i$) conditioned only on the covariates, i.e., $E(Y_{itj}|X_{itj})$; $g(.)$ is a known link function which makes the effects of the covariates on the mean response linear; X_{itj} is a $1 \times p$ vector of covariates for the j^{th} response of subject i at time t and β is a $p \times 1$ vector of regression coefficients which are assumed to be shared across multiple responses. As in the case of univariate marginal models, the covariates might be changing with time or not changing with time. We might allow multiple responses having their own intercepts by including the response types as indicator variables in the design matrix. Additionally, we might allow multiple responses having their own slopes by including the interactions of the responses types and the covariates in the design matrix. They will be clear when illustrated by examples (Section 3.4). The modeling diagram of MMM2 can be illustrated in Figure 3.1 as well. Note that for the sake of simplicity the diagram is drawn for bivariate responses. Likewise MMM1, multivariate marginal models explicitly accommodate the cross-response temporal dependence (via the working correlation matrices), i.e. association between Y_{itj} and $Y_{it'j'}$ for $t \neq t'$ and $j \neq j'$, although not shown in Figure 3.1 for simplicity. Note that by MMM1 formulation we assume that $\beta_1 \neq \beta_2$. On the other hand, by MMM2 formulation there are two options: 1) $\beta_1 = \beta_2$, 2) $\beta_1 \neq \beta_2$.

The methodology to fit MMM2 in the univariate marginal modeling framework is depicted below.

Let's assume a longitudinal dataset in which k multivariate longitudinal responses and $p-1$ covariates are available. Also assume that the longitudinal data is collected on N subjects for t time points. The (long) form of the data would be like the one illustrated in Table 3.1. Then, $(N{*}t){\times}k$ multivariate response matrix is manipulated to construct a new response matrix which has the dimension of $(N * t * k) \times 1$. This manipulation involves combining the multivariate responses of each subjects into one column, i.e., $Resp_{new}$=$(Resp_{1,new}, \ldots, Resp_{N,new})^T$ where $Resp_{i,new}$=$(Y_{i1.}, \ldots, Y_{in_i.})^T$ and $Y_{it.}$=$(Y_{it1}, \ldots, Y_{itk})$. The matrix of covariates is extended from an $(N * t) \times (p - 1)$ matrix to an $(N * t * k) \times (p - 1)$ matrix. This manipulation aims to extend the matrix of covariates for multivariate responses, i.e., X_{new}=$(X_{1,new}, \ldots, X_{N,new})^T$ where $X_{i,new}$ = (replicate($X_{i1.}$), ..., replicate($X_{in_i.}$))^T. The replicate function creates $k \times (p-1)$ matrices for which the rows are identical and equal to $X_{it.}$ where $X_{it.} = (X_{it1}, \ldots, X_{it(p-1)})$. The illustration of the form of reconstructed multivariate longitudinal data can be depicted in Table 3.3. Unlike MMM1, MMM2 assumes a common intercept for multiple responses and it is succeeded by placing an $(N * t * k) \times 1$ matrix for which all elements are 1 left to the first covariate (X_1) in the manipulated data form (Table 3.3). Response types (indicator variables) could be added as new covariates to the data form presented in Table 3.3 to build a model with differing intercepts for multiple responses. Similarly, the interactions of response types and some certain covariates could be added as covariates to the data form presented in Table 3.3 as well to build a model with differing slopes for multiple responses. All the aforementioned data manipulations including adding response types and related interactions are achieved within `mmm2` automatically. Eventually, the parameter estimations are done just like in the univariate marginal modeling case.

Table 3.3: An illustration of reconstructed multivariate longitudinal data for MMM2.

Subject	$Resp_{new}$	Cov_1	Cov_2	$\cdots$	$Cov_{(p-1)}$
1	Y_{111}	X_{111}	X_{112}	$\cdots$	$X_{11(p-1)}$
.	.	.	.	.	
1	Y_{11k}	X_{111}	X_{112}	$\cdots$	$X_{11(p-1)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
1	Y_{1t1}	X_{1t1}	X_{1t2}	$\cdots$	$X_{1t(p-1)}$
.	.	.	.	.	
1	Y_{1tk}	X_{1t1}	X_{1t2}	$\cdots$	$X_{1t(p-1)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
N	Y_{N11}	X_{N11}	X_{N12}	$\cdots$	$X_{N1(p-1)}$
.	.	.	.	.	
N	Y_{N1k}	X_{N11}	X_{N12}	$\cdots$	$X_{N1(p-1)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
N	Y_{Nt1}	X_{Nt1}	X_{Nt2}	$\cdots$	$X_{Nt(p-1)}$
.	.	.	.	.	
N	Y_{Ntk}	X_{Nt1}	X_{Nt2}	$\cdots$	$X_{Nt(p-1)}$

3.3 Details of the R Functions

In this section, we provide the details of the proposed functions, `mmm` and `mmm2` which fit the aforementioned two multivariate marginal models, MMM1 and MMM2, respectively.

3.3.1 Details of mmm

Asar and Ilk (2012) proposed an R function `mmm`, under the **mmm** package to fit multivariate marginal models with response specific regression parameters, MMM1. It largely utilizes the features of the `gee` function available under the R package **gee** (Carey, 2011a) and extends them to multivariate longitudinal data. Therefore, `mmm` requires the installation of that package and automatically loads it. Usage of `mmm` is very practical and results could be obtained in very short time. For the cases which contain so many subject (1000 or more) with large sized (15 or more) clusters and so many multivariate responses (5 or more) and covariates (15 or more), its termination might take a few seconds. The default function of `mmm` has the form of

```
R> mmm(data, nresp, family = "gaussian", corstr = "independence",
+ coefnames = NULL, tol = 0.001, maxiter = 25, Mv = 1, silent = TRUE)
```

The argument `data` is a data frame or matrix which consist of the multivariate responses and covariates to be included in the multivariate marginal modeling. All the information of subject i should be placed in subsequent rows and in time order and different subjects should be placed subsequently just as the form **gee** accepts. The first column of `data` should contain the identity information of the subjects in the study. This information is mandatory, since the data reconstructions within `mmm` and parameter estimations via **gee** rely on its availability. The first longitudinal response should be placed in the second column of the data frame. Similarly, k^{th} longitudinal response should be placed in the $(k+1)^{th}$ column of the data frame. The first associated covariate follow the k^{th} longitudinal response in column order, i.e., should be placed in $(k+2)^{th}$ column of the data frame. Similarly, the last associated covariate is to be placed in the last column of the data frame. One should be careful about that all the responses and covariates contained in the data frame entered into the function `mmm` via the `data` argument are automatically included in both of the multivariate and univariate marginal models.

The `nresp` argument is an integer which defines the number of multivariate longitudinal responses to be included in the marginal models. The correct entry of this argument is essential, since `mmm` distinguishes between the identity of the subjects, multivariate responses and the associated covariates by the help of the number of multivariate responses.

The `family` argument expects a character string which defines the class of response family. The available family choices of `mmm` are same with the ones of **gee**, despite only four common family options are mentioned here. It is set to `"gaussian"` by default which is valid for the continuous (gaussian type) data. `"binomial"` family should be chosen for binary response. Additionally, while `"poisson"` should be selected for count response, `"gamma"` should be selected for continuous (gamma type) response.

The `corstr` argument expects a character string to determine the structures of the serial and multivariate response dependences. It is set to `"independence"` by default which indicates the use of independence working correlation structure for the aforementioned dependence structures. It could be set to `"exchangeable"` for the use of exchangeable working correlation structure. Similarly, `"unstructured"` could be set for the unstructured working correlation structure. `"AR-M"` together with `"MV=1"` could be chosen for AR(1) working correlation structure. In addition to the four common working correlation structures, `"stat_M_dep"` and `"non_stat_M_dep"` could be chosen for stationary and non-stationary m-dependence structures, respectively alongside the specification of `"MV"` argument which defines the significant time lag value (Ziegler, 2011). For the details of the aforementioned four working correlation structures, see Section 2.3.2.

`coefnames` expects a vector of character strings in the same order of multivariate responses which is useful to easily identify the regression parameter estimates for different responses. Additionally, it yields a prettier output.

`tol` is the measure of tolerance defining the convergence of the fitting algorithm. The default is set to 0.001. However, decreasing `tol` would yield more sensitive parameter estimates. Nevertheless, there could be convergence problems.

`maxiter` is the maximum number of iterations which the fitting algorithm is supposed to consume during the parameter estimation. The default is set to 25. Nonetheless, in the cases for which the convergence of the algorithm fails, `maxiter` could be increased.

`silent` is a logical variable which determines the print of the details of each iteration. The default is set to be `TRUE` which means the details of the iterations not to be printed. In the cases for which those details are needed, setting `silent` to `FALSE` would report the details of the iterations.

`mmm` returns an object which includes both the multivariate and univariate marginal modeling results. The manipulation of the output of `mmm` is illustrated in Section 3.4.

3.3.2 Details of mmm2

We proposed an R function `mmm2` under the **mmm2** package to fit multivariate marginal models with shared regression parameters, MMM2. Likewise `mmm`, `mmm2` largely utilizes the features of `gee` and extends them to multivariate longitudinal data. Therefore, `mmm2` requires the installation of the R package **gee** and automatically loads it, as well. Usage of `mmm2` is also very practical and results could be obtained in very short time as well. The default function of `mmm2` has the form of

```
R> mmm2(data, nresp, rtype=TRUE, interaction=NULL, coefnames=NULL,
+ family="gaussian", tol=0.001, maxiter=25, corstr="independence",
+ Mv=1, silent=TRUE)
```

The argument `data` is a data frame or matrix which consists of the multivariate responses and covariates to be included in the multivariate marginal modeling. All the information of subject i should be placed in subsequent rows and in time order and different subjects should be placed subsequently just as the form **gee** accepts. The first column of `data` should contain the identity information of the subjects in the study. This information is mandatory, since the data reconstructions within `mmm2` and parameter estimations via **gee** rely on its availability. The first longitudinal response should be placed in the second column of the data frame. Similarly, k^{th} longitudinal response should be placed in the $(k+1)^{th}$ column of the data frame. The first associated covariate follow the k^{th} longitudinal response in column order, i.e., should be placed in $(k+2)^{th}$ column of the data frame. Similarly, the last associated covariate is to be placed in the last column of the data frame. One should be careful about that all the responses and covariates contained in the data frame entered into the function `mmm2` via the `data` argument are automatically included in both of the multivariate and univariate marginal models.

The `nresp` argument is an integer which defines the number of multivariate longitudinal responses to be included in the marginal models. The correct entry of this argument is essential, since `mmm2` distinguishes between the identity of the subjects, multivariate responses and the associated covariates by the help of the number of multivariate responses.

The `rtype` argument is a logical statement which determines the inclusion of response types as new covariates in the indicator variable format (variables taking only 0 or 1). The default is set to TRUE which means inclusion of response types by placing them right after the $(p-1)^{th}$ covariate in the design matrix. For k multiple responses, $k-1$ indicator variables are to be created and `mmm2` has a systematic way of creating these variables: The first response takes 0 for all the $k-1$ indicator variables and j^{th} response ($j = 2, \ldots, k$) takes 1 only for the $(k-j+1)^{th}$ indicator variable and takes 0 otherwise. To illustrate, let's consider a simple example where only 5 multiple responses are available. For these multiple responses, 4 indicator variables are to be created and the ones which would be created by `mmm2` systematically are presented in Table 3.4. If differing intercepts are not of interest, `rtype` could be set to FALSE.

Table 3.4: An illustration of the systematic way of mmm2 in terms of creating indicator variables for 5 multiple responses.

	$rtype_1$	$rtype_2$	$rtype_3$	$rtype_4$
$Response_1$	0	0	0	0
$Response_2$	0	0	0	1
$Response_3$	0	0	1	0
$Response_4$	0	1	0	0
$Response_5$	1	0	0	0

Note: Rtype denotes the indicator variables.

`interaction` expects a vector of integers which includes the column number of the covariates (by considering only the covariate matrix) which are to be interacted with the response types. If `rtype` is set to FALSE, `mmm2` ignores `interaction` even if it is set to a vector of column numbers of some covariates.

`coefnames` expects a vector of character strings which includes the coefficient names to have a prettier output.

The `family` argument expects a character string which defines the class of response family. The available family choices of `mmm2` are same with the ones of **gee**, despite only four common family options are mentioned here. It is set to `"gaussian"` by default which is valid for the continuous (gaussian type) data. `"binomial"` family should be chosen for binary response. Additionally, while `"poisson"` should be selected for count response, `"gamma"` should be selected for continuous (gamma type) response.

`tol` is the measure of tolerance defining the convergence of the fitting algorithm. The default is set to 0.001. However, decreasing `tol` would yield more sensitive parameter estimates. Nevertheless, there could be convergence problems.

`maxiter` is the maximum number of iterations which the fitting algorithm is supposed to consume during the parameter estimation. The default is set to 25. Nonetheless, in the cases for which the convergence of the algorithm fails, `maxiter` could be increased.

The `corstr` argument expects a character string to determine the structures of the serial and multivariate response dependences. It is set to `"independence"` by default which indicates the use of independence working correlation structure for the aforementioned dependence structures. It could be set to `"exchangeable"` for the use of exchangeable working correlation structure. Similarly, `"unstructured"` could be set for the unstructured working correlation structure. `"AR-M"` together with `"MV=1"` could be chosen for AR(1) working correlation structure. In addition to the four common working correlation structures, `"stat_M_dep"` and `"non_stat_M_dep"` could be chosen for stationary and non-stationary m-dependence structures, respectively alongside the specification of `"MV"` argument which defines the significant time lag value. For the details of the aforementioned four working correlation structures, see Section 2.3.2.

`silent` is a logical variable which determines the print of the details of each iteration. The default is set to be `TRUE` which means the details of the iterations not to be printed. In the cases for which those details are needed, setting `silent` to `FALSE` would report the details of the iterations.

`mmm2` returns a list of output for MMM2 which is very similar to the one **gee** produces. The manipulation of the output will be illustrated in Section 3.4.

3.4 Multivariate Longitudinal Data Applications

In this sub-section, we illustrate the features of MMM1 and MMM2 together with the proposed R functions, `mmm` and `mmm2`, respectively. We show that those models are not specific to any response family, but they can handle variety of them. Here, we consider Binomial, Poisson and Gaussian response families. Whereas we consider a real life dataset to illustrate the former, the latter two are illustrated via artificial datasets. We also show that equivalent models could be fitted by MMM1 and MMM2. Here, we shall note that the following examples are not carried with data analysis purposes, but they are designed to illustrate the features of the models and the proposed software.

3.4.1 Multivariate Longitudinal Binary Data Applications

The multivariate longitudinal binary data used to illustrate the aforementioned multivariate marginal models (MMM1 and MMM2) comes from the study called Mother's Stress and Children's Morbidity Study (MSCM; Alexander and Markowitz, 1986) which was introduced in Section 1.2.1 in detail. In MSCM study, 167 mothers and their preschool children were enrolled for 28 days. The bivariate binary responses are mother's stress status (0=absence, 1=presence) and children's illness status (0=absence, 1=presence). All the variables can be found in Table 3.5. Investigation of the serial dependence structures of the two longitudinal responses (mother's stress and child's illness) suggested a weak correlation structure for the period of days 1 to 16. Therefore, only the period of days 17 to 28 is considered here. To catch the specific characteristics of the mothers and children, the average response values of the ignored time period (days 1 to 16) were added as new covariates which are baseline stress and baseline illness (bstress and billness in Table 3.5), respectively. Additionally, the standardized time information is considered as a new variable (week in Table 3.5).

Only the responses had missing values; 0.97% and 1.42% of the stress and illness observation are missing, respectively. Since data analysis under missing data is beyond the scope of this study, the missing values were simply imputed by occasion mode imputation.

Table 3.5: List of the variables that appear in the MSCM study and the related explanations.

Variable	Explanation
stress	mother's stress status: 0=absence, 1=presence
illness	child's illness status: 0=absence, 1=presence
married	marriage status of the mother: 0=other, 1=married
education	mother's education level: 0=high school or less, 1=high school graduate
employed	mother's employment status: 0=unemployed, 1=employed
chlth	child's health status at baseline: 0=very poor/poor, 1=fair, 2=good, 3=very good
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
csex	child's gender: 0=male, 1=female
housize	size of the household: 0=2-3 people, 1=more than 3 people
bstress	baseline stress: average value of the mother's stress status for the first 16 days
billness	baseline illness: average value of the child's illness status for the first 16 days
week	a time variable: calculated as (day-22)/7

3.4.1.1 Application of Multivariate Marginal Models with Response Specific Regression Parameters

A multivariate marginal models with response specific parameters (MMM1) constructed for MSCM data would be like the one given in the following equation.

$$logit(P(Y_{itj} = 1|X_{it})) = \beta_{0j} + \beta_{1j} * married_{it} + \cdots + \beta_{11j} * week_t \tag{3.3}$$

The setting of $j = 1$ in (3.3) corresponds to response=stress, and $j = 2$ corresponds to response=illness.

The related model fitting of MSCM by `mmm` could be obtained by

```
R> install.packages("mmm")
R> library("mmm")
R> data("mscm.mmm")
R> coefnames1<-c("str.intercept","str.married","str.education",
+ "str.employed","str.chlth","str.mhlth","str.race","str.csex",
+ "str.housize","str.bstress","str.billness","str.week",
+ "ill.intercept","ill.married","ill.education","ill.employed",
```

```
+ "ill.chlth","ill.mhlth","ill.race","ill.csex","ill.housize",
+ "ill.bstress","ill.billness","ill.week")
R> mmm1<-mmm(data=mscm.mmm,nresp=2,family=binomial(link=logit),
+ corstr='exchangeable',coefnames=coefnames1)
```

`mmm` returns a list of output for both the univariate and multivariate marginal models which is very similar to the ones produced by **gee**. The output of multivariate models could be obtained by

```
R> mmm1$multiv$multivout
```

The multivariate marginal modeling regression parameter estimates and their related statistics (naive and robust variance and Z estimates) could be obtained by

```
R> mmm1$multiv$multivout$coef
```

Additionally, the univariate marginal model fitting output for the first response (response=stress) could be obtained by

```
R> mmm1$univ$univout[[1]]
```

Also the univariate marginal modeling regression parameter estimates of the first response and their related statistics could be obtained by

```
R> mmm1$univ$univout[[1]]$coef
```

Similar to the first response, the univariate marginal modeling output of the second response (response=illness) could be obtained by

```
R> mmm1$univ$univout[[2]]
```

Lastly, the univariate marginal modeling regression parameter estimates of the second response and their related statistics could be obtained by

```
R> mmm1$univ$univout[[2]]$coef
```

Although we directed the users how to obtain the multivariate and univariate marginal modeling results from the output of `mmm`, they should not memorize these steps. Instead, they can use the `names` function to see the possible options. We illustrated the use of this function in `mmm` below.

One can look at the main parts of a `mmm` output by

```
R> names(mmm1)
[1] "multiv" "univ"
```

Here, `"multiv"` corresponds to the outputs of multivariate marginal models, and `"univ"` corresponds to the outputs of univariate marginal models.

Let's illustrate `"multiv"` and `"univ"` one by one. The components of `"multiv"` can be viewed by

```
R> names(mmm1$multiv)
[1] ""          "multivout"
```

Here, `""` corresponds to the title of the `multiv` section. When one selected `""` , s/he would have the following.

```
R> mmm1$multiv$""
[1] "Multivariate Marginal Model"
```

Note that when one types `R>names(mmm1$multiv$"")`, s/he would see `NULL` due to the fact that the related content is a title. Additionally, `"multivout"` corresponds to the output of the MMM1, and one would print the related whole output by `R>mmm1$multiv$multivout`. On the other hand, the related content could be obtained by

```
R> names(mmm1$multiv$multivout)
 [1] "call"         "version"      "nobs"             "residual.summary"
 [5] "model"        "title"        "coefficients"     "working.correlation"
 [9] "scale"        "error"        "iterations"
```

Here, `"coefficients"`corresponds to the regression parameter estimates and the related naive and robust variance and Z estimates. One interesting feature of R is that one can use either the full name of a content or its abbreviated name. To illustrate, to obtain the regression parameter estimates and related statistics, one can either use

```
R> mmm1$multiv$multivout$coefficients
```

or

```
R> mmm1$multiv$multivout$coef
```

Similarly, the outputs of univariate marginal models could be obtained. One can look at the contents of `univ` by

```
R> names(mmm1$univ)
[1] ""        "univout"
```

As for the `"multiv"`, `""` corresponds to the title of the `"univ"`component. When one looks at the content of the `"univout"`, s/he would have `NULL`, since `"univout"`comprises of univariate marginal modeling outputs for each responses and the number of multiple responses varies for different datasets. Yet, the related outputs could be extracted by `[[]]` operation. To illustrate, the univariate marginal modeling output of the first response could be obtained by

```
R> mmm1$univ$univout[[1]]
```

On the other hand, the components of the output could be printed by

```
R> names(mmm1$univ$univout[[1]])
 [1] "call"         "version"        "nobs"              "residual.summary"
 [5] "model"        "title"          "coefficients"      "working.correlation"
 [9] "scale"        "error"          "iterations"
```

The regression parameter estimates and related naive and robuts Z values could be extracted as for the multivariate case. The related operations for other responses are straightforward, i.e. for the j^{th} response one should use `[[j]]`.

The estimates obtained by `mmm` for MSCM data could be found in Table 3.6. The multivariate marginal model estimates of `mmm` were compared with the ones of the SAS macro of Shelton et al. (2004) and found to be identical; hence only results of mmm were reported here. The multivariate and univariate marginal modeling regression parameter estimates of the MSCM data seemed to be in agreement with each other in terms of magnitude and direction, standard errors and statistical significance. However, while multivariate results indicated significant effect at the boundary for the child health on the presence of mother's stress (Z value=-1.96), univariate results indicated slightly more significant effect of it on the presence of mother's stress (Z value=-2.01). Furthermore, multivariate results indicated negative effect of child's race on the presence of mother's stress (est.=-0.02), univariate results indicated positive effect of it on the presence of mother's stress (est.=0.04). Nevertheless, both of the estimates were statistically insignificant. Despite the fact that the results of the multivariate and univariate marginal models for the MSCM data seem to be close to each other, this might not be the case for other datasets. These close results are natural for the datasets with weak multivariate response dependence, e.g. for MSCM dataset the mean Spearman correlation between the mother's stress and children's illness for the period of days 17 to 28 was found to be 0.13. Nevertheless, even this weak association yielded different statistical inferences. Therefore, while doing data analysis, multivariate response associations should be taken into account.

3.4.1.2 Application of Multivariate Marginal Models with Shared Regression Parameters

A typical multivariate marginal models with shared regression parameters (MMM2) for MSCM data can be illustrated in (3.4).

$$logit(P(Y_{itj} = 1|X_{itj})) = \beta_0 + \beta_1 * married_{itj} + \cdots + \beta_{11} * week_{tj} \tag{3.4}$$

The model given in (3.4) assumes that the covariates have the same effects on different responses. For example, it indicates that time (week) has the same effect on the log odds of presence of a mother's stress and her child's illness, β_{11}. The related model fitting via mmm2 under exchangeable working correlation structure could be obtained by the following R script.

```
R> install.packages("mmm2")
R> library("mmm2")
R> data("mscm")
```

Table 3.6: Results of the multivariate and univariate marginal model analyses of MSCM.

	multivariate results			univariate results		
	Est.	(SE)	Z	Est.	(SE)	Z
Response=Stress						
Intercept	-2.14	(0.42)	-5.15	-2.19	(0.41)	-5.38
married	-0.01	(0.24)	-0.02	-0.01	(0.23)	-0.03
education	0.36	(0.23)	1.61	0.38	(0.22)	1.67
employed	-0.65	(0.25)	-2.59	-0.66	(0.24)	-2.70
chlth	-0.26	(0.13)	-1.96	-0.26	(0.13)	-2.01
mhlth	-0.17	(0.12)	-1.39	-0.17	(0.12)	-1.41
race	-0.02	(0.24)	-0.06	0.04	(0.24)	0.15
csex	-0.04	(0.22)	-0.20	-0.003	(0.22)	-0.01
housize	0.06	(0.24)	0.26	0.06	(0.24)	0.26
bstress	3.89	(0.71)	5.48	3.94	(0.71)	5.56
billness	0.86	(0.71)	1.21	0.81	(0.71)	1.14
week	-0.43	(0.16)	-2.65	-0.43	(0.16)	-2.64
Response=Illness						
Intercept	-1.58	(0.49)	-3.30	-1.67	(0.49)	-3.43
married	0.50	(0.27)	1.88	0.52	(0.27)	1.91
education	-0.06	(0.29)	-0.19	-0.01	(0.29)	-0.04
employed	-0.22	(0.33)	-0.66	-0.23	(0.33)	-0.69
chlth	-0.40	(0.16)	-2.56	-0.38	(0.16)	-2.43
mhlth	0.03	(0.17)	0.15	0.01	(0.17)	0.05
race	0.02	(0.24)	0.08	0.12	(0.25)	0.47
csex	0.02	(0.25)	0.08	0.05	(0.25)	0.18
housize	-0.56	(0.26)	-2.18	-0.60	(0.26)	-2.32
bstress	0.06	(0.98)	0.06	0.20	(0.97)	0.20
billness	2.18	(0.75)	2.89	2.04	(0.75)	2.74
week	-0.20	(0.22)	-0.90	-0.20	(0.22)	-0.90

Note: Only robust standard error and Z estimates are reported. Both of the models were fitted under exchangeable working correlation structure.

```
R> coefnames12<-c("intercept","married","education","employed","chlth",
+ "mhlth","race","csex","housize","bstress","billness","week")
R> mmm12<-mmm2(data=mscm,nresp=2,rtype=FALSE,coefnames=coefnames12,
+ family=binomial(link=logit),corstr="exchangeable")
```

Here, we shall note that the datasets “mscm.mmm” under the **mmm** package and “mscm” under the **mmm2** package are exactly same, but they have been assigned different names. `mmm2` returns a list of output for multivariate marginal models which is very similar to the one produced by **gee**. The related content could be displayed by

```
R> names(mmm12)
 [1] "title"               "version"            "model"            "call"
 [5] "terms"               "nobs"               "iterations"       "coefficients"
 [9] "nas"                 "linear.predictors"  "fitted.values"    "residuals"
[13] "family"              "y"                  "id"               "max.id"
[17] "working.correlation" "scale"              "robust.variance"  "naive.variance"
[21] "xnames"              "error"
```

`summary(mmm12)` would print the whole output of MMM2, and parameter estimates and the related statistics (naive and robust variance and Z estimates) could be obtained by

```
R> summary(mmm12)$coef
```

The results are placed in Table 3.7 under the column of Model 1. Model 1 produced estimates averaged over the bivariate responses. For example, MMM1 for MSCM data (Table 3.6) estimated the intercept for response=stress as -2.14 and the intercept for response=illness as -1.58. On the other hand, MMM2 (Model 1) estimated a shared intercept as -1.89. By the formulation of the MMM2 (3.4), we are able to estimate 12 less regression parameters compared to the MMM1 (3.3) for the same data.

Both MMM1 and MMM2 formulations require the estimation of $(k * t) * (k * t - 1)/2$ different correlation parameters under the unstructured working correlation structure. For example, for the MSCM data, it requires the estimation of 276 different correlation parameters in addition to the 24 regression parameters (3.3). Therefore, sometimes the modeling algorithms could not convergence. For example, we could not fit an MMM1 for MSCM data (3.3) under the unstructured working correlation structure. Instead, we fitted an MMM1 under exchangeable working correlation structure which only estimates 1 correlation parameter. By the MMM2 formulation (3.4), we could achieve fitting a model under unstructured working correlation structure. By this modeling formulation, we still had to estimate $(k*t)*(k*t-1)/2$ correlation parameters. Nonetheless, since MMM2 permits fitting more parsimonious models via shared regression parameters, we were able to fit multivariate marginal models under unstructured working correlation structure. Such model fitting and parameter estimates by mmm2 could be obtained by

```
R> library("mmm2")
R> data("mscm")
R> coefnames13<-c("intercept","married","education","employed","chlth",
+ "mhlth","race","csex","housize","bstress","billness","week")
R> mmm13<-mmm2(data=mscm,nresp=2,rtype=FALSE,coefnames=coefnames13,
+ family=binomial(link=logit),corstr="unstructured")
R> summary(mmm13)$coef
```

The related results are displayed in Table 3.7 under the column of Model 2. While Model 1 (with the exchangeable working correlation structure) indicates statistically significant effect of employment status of a mother (at 95% confidence level) on the presence of her stress and her child's illness, Model 2 (with the unstructured working correlation structure) indicates such effect is statistically insignificant; robust Z values are -2.01 and -1.50, respectively. As expected, GEE yielded consistent parameter and variance estimates under different working correlation assumptions.

We can fit equivalent models by MMM1 and MMM2 formulations. For example, a model by MMM2 indicating an equivalent one to MMM1, which is given in (3.3), can be constructed as below.

$$
\begin{aligned}
logit(P(Y_{itj} = 1|X_{itj})) = \beta_0 + \beta_1 * married_{itj} + \cdots + \beta_{11} * week_{tj} + \beta_{12} * rtype_j + \\
\beta_{13} * (married_{itj} * rtype_j) + \cdots + \beta_{23} * (week_{tj} * rtype_j) \qquad (3.5)
\end{aligned}
$$

Let's assume rtype=0 for response=stress and rtype=1 for response=illness by following mmm2 (see rtype argument). Then the MMM2 (3.5) indicates two different models for these bivariate responses. For response=stress, it indicates

$$logit(P(Y_{it1} = 1|X_{it1})) = \beta_0 + \beta_1 * married_{it1} + \cdots + \beta_{11} * week_{t1} \tag{3.6}$$

On the other hand, for response=illness, it indicates

$$logit(P(Y_{it2} = 1|X_{it2})) = (\beta_0 + \beta_{12}) + (\beta_1 + \beta_{13}) * married_{it2} + \cdots + (\beta_{11} + \beta_{23}) * week_{t2} \tag{3.7}$$

The related model fitting and parameter estimates via mmm2 could be obtained by

```
R> library("mmm2")
R> data("mscm")
R> coefnames14<-c("intercept","married","education","employed","chlth",
+ "mhlth","race","csex","housize","bstress","billness","week","resptype",
+ "married*resptype","education*resptype","employed*resptype",
+ "chlth*resptype","mhlth*resptype","race*resptype","csex*resptype",
+ "housize*resptype","bstress*resptype","billness*resptype","week*resptype")
R> mmm14<-mmm2(data=mscm,nresp=2,interaction=seq(1:11),
+ coefnames=coefnames14,family=binomial(link=logit),corstr="exchangeable")
R> summary(mmm14)$coef
```

The parameter estimates are placed in Table 3.7 under the column of Model 3. Since mmm2 assigns response type indicator variable (rtype) to 0 for response=stress, the related results are identical with the results for response=stress obtained by MMM1 (Table 3.6). We can obtain the estimates of response=illness by setting rtype=1. For example, the estimate of baseline illness (billness) for response=illness can be calculated by 0.86 + 1.32 = 2.18 which is equivalent to the one presented in Table 3.6. Moreover, Model 3 indicates that some of the interactions of response types and covariates are statistically insignificant. This might be interpreted as the effects of the related covariates could be considered shared across responses, hence these insignificant interactions could be omitted from the model. To illustrate this approach, we only considered the significant or barely significant interaction terms. Specifically, we considered the interactions of response types and marriage status of mothers, household size, baseline stress and baseline illness and fit Model 4 (Table 3.7). Note that Model 4 still follows the formulation given in (3.5) with less interactions. The related model fit could be obtained by

```
R> library("mmm2")
R> data("mscm")
R> coefnames15<-c("intercept","married","education","employed","chlth",
+ "mhlth","race","csex","housize","bstress","billness","week","resptype",
+ "married*resptype","housize*resptype","bstress*resptype","billness*resptype")
R> mmm15<-mmm2(data=mscm,nresp=2,interaction=c(1,8,9,10),coefnames=coefnames15,
+ family=binomial(link=logit),corstr="exchangeable")
R> summary(mmm15)$coef
```

While Model 3 indicates slightly significant effect of the child health at baseline (chlth) on the log odds of the presence of a mother's stress and her child's illness, Model 4 indicates more significant effect; related robust Z values are -1.96 and -2.75, respectively. The conclusion drawn by Model 4 might be regarded as more sensible, since the health status of a child at baseline is expected to effect his/her health status and the stress status of his/her mother at later days.

`mmm2` includes more data reconstruction compared to `mmm`, hence it requires more computational time. The required computational times for the analyses of MSCM are presented in Table 3.8. The analyses were done on a PC with 4.00 GB RAM and 2.53 GHz processor. While `mmm` requires very short time (1.14 seconds) for model fitting, `mmm2` requires relatively more time.

3.4.2 Multivariate Longitudinal Count Data Applications

In addition to multivariate longitudinal binary data, MMM1 and MMM2 can handle the multivariate marginal modeling of multivariate longitudinal count data. Since no such real life data is available at hand, one was simulated by utilizing the R package **corcounts** (Erhardt, 2009). For simplicity, bivariate longitudinal count responses measured at four time points and an age-at-baseline type (time independent) covariate were considered. The bivariate responses were assumed to be coming from Poisson distribution with different mean values, 5 and 8. Furthermore, their means were assumed to be constant across time. Additionally, the serial correlation structures of both the bivariate responses were assumed to have an AR(1) type correlation structure which is common in longitudinal studies. The assumed correlation matrix is illustrated in Table 3.9. Note that since the simulated covariate is time independent, we generated it only for the first time point.

Table 3.7: Results of the MMM2 analyses of MSCM.

	Model 1			Model 2			Model 3			Model 4		
	Est.	(SE)	Z	Est.	(SE)	Z	Est.	(SE)	Z	Est.	(SE)	Z
Intercept	-1.89	(0.36)	-5.32	-2.19	(0.33)	-6.59	-2.14	(0.42)	-5.15	-2.09	(0.37)	-5.70
married	0.26	(0.19)	1.40	0.20	(0.18)	1.09	-0.01	(0.24)	-0.02	0.002	(0.24)	0.01
education	0.18	(0.20)	0.91	0.26	(0.20)	1.29	0.36	(0.23)	1.62	0.17	(0.20)	0.84
employed	-0.44	(0.22)	-2.02	-0.32	(0.21)	-1.50	-0.65	(0.25)	-2.59	-0.45	(0.22)	-2.00
chlth	-0.32	(0.12)	-2.76	-0.27	(0.11)	-2.39	-0.26	(0.13)	-1.96	-0.33	(0.12)	-2.75
mhlth	-0.08	(0.11)	-0.73	-0.15	(0.10)	-1.46	-0.17	(0.12)	-1.39	-0.10	(0.12)	-0.82
race	0.06	(0.18)	0.30	0.20	(0.19)	1.07	-0.02	(0.24)	-0.06	-0.02	(0.18)	-0.10
csex	0.02	(0.18)	0.13	0.08	(0.17)	0.46	-0.04	(0.22)	-0.20	0.02	(0.18)	0.13
housize	-0.26	(0.19)	-1.37	-0.16	(0.19)	-0.85	0.06	(0.24)	0.26	0.06	(0.24)	0.26
bstress	2.13	(0.61)	3.48	1.97	(0.58)	3.40	3.89	(0.71)	5.48	3.90	(0.70)	5.55
billness	1.40	(0.58)	2.40	1.61	(0.55)	2.91	0.86	(0.71)	1.21	0.86	(0.69)	1.24
week	-0.31	(0.14)	-2.20	-0.35	(0.13)	-2.62	-0.43	(0.16)	-2.65	-0.31	(0.14)	-2.20
rtype							0.56	(0.54)	1.04	0.57	(0.35)	1.63
married*rtype							0.50	(0.32)	1.57	0.49	(0.32)	1.53
education*rtype							-0.42	(0.31)	-1.35			
employed*rtype							0.43	(0.38)	1.13			
chlth*rtype							-0.14	(0.17)	-0.82			
mhlth*rtype							0.20	(0.18)	1.12			
race*rtype							0.04	(0.32)	0.11			
csex*rtype							0.06	(0.29)	0.21			
housize*rtype							-0.63	(0.32)	-1.95	-0.66	(0.32)	-2.07
bstress*rtype							-3.83	(1.10)	-3.50	-4.04	(1.00)	-4.05
billness*rtype							1.32	(0.88)	1.50	1.38	(0.87)	1.58
week*rtype							0.24	(0.26)	0.91			

Note: Only robust standard error and Z estimates are reported. While Models 1, 3 & 4 were fitted under exchangeable working correlation structure, only Model 2 was fitted under unstructured working correlation assumption.

Table 3.8: Computational times consumed for the analyses of MSCM dataset.

Model	Time (in seconds)
MMM1	1.14
MMM2 - Model 1	16.11
MMM2 - Model 2	16.40
MMM2 - Model 3	16.15
MMM2 - Model 4	16.15

Table 3.9: Assumed correlation matrix for the multivariate longitudinal count data.

	Y_{i11}	Y_{i12}	X_{i11}	Y_{i21}	Y_{i22}	Y_{i31}	Y_{i32}	Y_{i41}	Y_{i42}
Y_{i11}	1.00	0.40	0.60	0.90	0.37	0.80	0.34	0.70	0.31
Y_{i12}	0.40	1.00	0.60	0.37	0.90	0.34	0.80	0.31	0.70
X_{i11}	0.60	0.60	1.00	0.60	0.60	0.60	0.60	0.60	0.60
Y_{i21}	0.90	0.37	0.60	1.00	0.40	0.90	0.37	0.80	0.34
Y_{i22}	0.37	0.90	0.60	0.40	1.00	0.37	0.90	0.34	0.80
Y_{i31}	0.80	0.34	0.60	0.90	0.37	1.00	0.40	0.90	0.37
Y_{i32}	0.34	0.80	0.60	0.37	0.90	0.40	1.00	0.37	0.90
Y_{i41}	0.70	0.31	0.60	0.80	0.34	0.90	0.37	1.00	0.40
Y_{i42}	0.31	0.70	0.60	0.34	0.80	0.37	0.90	0.40	1.00

The time independent covariate was assumed to be coming from Poisson distribution with mean 20 and its simulation was achieved together with the simulation of responses. The data simulation via **corcounts** could be obtained by

```
R> library("corcounts")
R> set.seed(12)
R> n1 <- 500
R> margins <- c("Poi","Poi","Poi","Poi","Poi","Poi","Poi","Poi","Poi")
R> mu <- c(5, 8, 20, 5, 8, 5, 8, 5, 8)
R> corstr <- "unstr"
R> corpar<-matrix(c(1,0.4,0.6,0.9,0.37,0.8,0.34,0.7,0.31,
+ 0.4,1,0.6,0.37,0.9,0.34,0.8,0.31,0.7,
+ 0.6,0.6,1,0.6,0.6,0.6,0.6,0.6,0.6,
+ 0.9,0.37,0.6,1,0.4,0.9,0.37,0.8,0.34,
+ 0.37,0.9,0.6,0.4,1,0.37,0.9,0.34,0.8,
+ 0.8,0.34,0.6,0.9,0.37,1,0.4,0.9,0.37,
+ 0.34,0.8,0.6,0.37,0.9,0.4,1,0.37,0.9,
+ 0.7,0.31,0.6,0.8,0.34,0.9,0.37,1,0.4,
+ 0.31,0.7,0.6,0.34,0.8,0.37,0.9,0.4,1),ncol=9,byrow=T)
R> data1 <- rcounts(N=n1,margins=margins,mu=mu,corstr=corstr,
+ corpar=corpar)
```

The simulated data should be reconstructed to be analyzed by `mmm1` and `mmm2`. The data reconstruction could be achieved by

```
R> time11<-data1[,1:3]
R> time12<-cbind(data1[,4:5],data1[,3])
R> time13<-cbind(data1[,6:7],data1[,3])
R> time14<-cbind(data1[,8:9],data1[,3])
R> data12<-rbind(time11,time12,time13,time14)
R> time1<-matrix(rep(seq(1:4),each=n1))
R> id1<-matrix(rep(seq(1:n1),4))
R> data13<-cbind(id1,data12,time1)
R> # sorting the data for subjects
R> data14<-NULL
R> for (i in 1:n1){
R> data14<-rbind(data14,data13[data13[,1]==i,])
R> }
```

In addition to the simulated covariate, time was added as a covariate as well. To observe the changes in the regression coefficients easier, the natural logarithm of the simulated covariate was taken and time was standardized. Moreover, their interaction was added into the model. These data manipulations could be achieved by

```
R> data14[,4]<-log(data14[,4])
R> data14[,5]<-scale(data14[,5])
R> data14<-cbind(data14,data14[,4]*data14[,5])
```

Note that we call the simulated multivariate longitudinal count data, `data14` as MLCD throughout. Also note that since we simulated MLCD by fixing the seed number to `seed=12`, it is reproducible.

3.4.2.1 Application of Multivariate Marginal Models with Response Specific Regression Parameters

A multivariate marginal models with response specific regression parameters (MMM1) constructed for MLCD is given below.

$$log(E(Y_{itj}|X_{it})) = \beta_{0j} + \beta_{1j} * X_{it} + \beta_{2j} * time_t + \beta_{3j} * (X_{it} * time_t) \tag{3.8}$$

Setting j=1 in (3.8) corresponds to the first longitudinal count response and j=2 corresponds to the second longitudinal count response. The model fittings of MLCD via `mmm` could be done by

```
R> library("mmm")
R> coefnames2<-c("Y1.int","Y1.beta1","Y1.beta2","Y1.beta3",
+ "Y2.int","Y2.beta1","Y2.beta2","Y2.beta3")
R> mmm2<-mmm(data=data14,nresp=2,family="poisson",
R> coefnames=coefnames2,corstr='unstructured')
```

The multivariate and univariate marginal modeling results produced by `mmm` could be obtained by

Table 3.10: Multivariate and univariate marginal model analyses of multivariate longitudinal count data.

	multivariate results			univariate results		
	Est.	(SE)	Z	Est.	(SE)	Z
Response 1						
Intercept	-1.98	(0.17)	-11.76	-1.97	(0.17)	-11.76
X	1.20	(0.06)	21.78	1.20	(0.06)	21.83
time	-0.02	(0.08)	-0.31	-0.02	(0.08)	-0.28
X*time	0.01	(0.03)	0.39	0.01	(0.03)	0.36
Response 2						
Intercept	-0.72	(0.15)	-4.96	-0.71	(0.15)	-4.92
X	0.94	(0.05)	19.75	0.94	(0.05)	19.72
time	-0.04	(0.07)	-0.66	-0.02	(0.06)	-0.65
X*time	0.02	(0.02)	0.77	0.02	(0.02)	0.77

Note: Only robust standard error and Z estimates are reported. The models were fitted under unstructured working correlation assumption.

```
R> mmm2$multiv$multivout$coef
R> mmm2$univ$univout[[1]]$coef
R> mmm2$univ$univout[[2]]$coef
```

respectively. The related results are placed in Table 3.10. As in the case of MSCM, the multivariate and univariate marginal modeling regression parameter estimates for MLCD seemed to be in agreement with each other in terms of magnitude and direction, standard errors and statistical significance.

3.4.2.2 Application of Multivariate Marginal Models with Shared Regression Parameters

A typical multivariate marginal models with shared regression parameters (MMM2) constructed for MLCD is illustrated below.

$$log(E(Y_{itj}|X_{itj})) = \beta_0 + \beta_1 * X_{itj} + \beta_2 * time_{tj} + \beta_3 * (X_{itj} * time_{tj}) \tag{3.9}$$

The model given in (3.9) assumes that the effects of the covariates on bivariate responses are same. For example, the covariate X has same effects on the first and second count responses, β_1. The related model fitting and the parameter estimates via `mmm2` could be obtained by

```
R> library("mmm2")
R> coefnames22<-c("intercept","X","time","X*time")
R> mmm22<-mmm2(data=data14,nresp=2,rtype=FALSE,coefnames=coefnames22,
+ family="poisson",corstr="exchangeable")
R> summary(mmm22)$coef
```

The results are displayed in Table 3.11 under Model 1 column. Since MMM2 assumes that the effects of covariates on multiple responses are same, Model 1 produced averaged parameter estimates over the responses. For example, whereas MMM1 estimated the time effect as -0.02 and -0.04 for the first and second responses, respectively; MMM2 estimated the time effect as -0.03. Sometimes, the fitting algorithm for MMM2 might fail to converge under the unstructured working correlation structure due to the fact that fitting a model which assumes that all the covariate effects are shared across responses (a too parsimonious model) might hide the necessary and significant information of the covariates required for model building. For example, while Model 2 which will be introduced next (Table 3.11) shows that the intercepts and the covariate effects (X) are significantly different for the first and second responses, Model 1 assumes that both of these effects are same for the multiple responses. Model 1 failed to converge under unstructured correlation structure most probably due to this reason and it was built under exchangeable working correlation structure.

As for the multivariate longitudinal binary data, we may fit equivalent models under MMM1 and MMM2 formulations for the count counterparts. An MMM2 for MLCD which indicates an equivalent model constructed in the MMM1 formulation given in (3.8) is given below.

$$\begin{aligned} log(E(Y_{itj}|X_{itj})) = \beta_0 + \beta_1 * X_{itj} + \beta_2 * time_{tj} + \beta_3 * (X_{itj} * time_{tj}) + \beta_4 * rtype_j + \\ \beta_5 * (X_{itj} * rtype_j) + \beta_6 * (time_{tj} * rtype_j) + \beta_7 * (X_{itj} * time_{tj} * rtype_j) \end{aligned} \tag{3.10}$$

Following the `rtype` argument of `mmm2`, let's assign the response type=0 (rtype) for the first count response, and response type=1 for the second one. The MMM2 given in (3.10) actually indicates two different models which are distinguished by the response type variable. For example, for the first response it indicates

$$log(E(Y_{it1}|X_{it1})) = \beta_0 + \beta_1 * X_{it1} + \beta_2 * time_t + \beta_3 * (X_{it1} * time_t) \tag{3.11}$$

On the other hand, for the second response it indicates

$$log(E(Y_{it2}|X_{it2})) = (\beta_0 + \beta_4) + (\beta_1 + \beta_5) * X_{it2} + (\beta_2 + \beta_6) * time_t + (\beta_3 + \beta_7) * (X_{it2} * time_t) \tag{3.12}$$

The related model fitting and parameter estimates could be obtained via `mmm2` by

```
R> library("mmm2")
R> coefnames23<-c("intercept","X","time","X*time","resptype",
+ "X*resptype","time*resptype","X*time*resptype")
R> mmm23<-mmm2(data=data14,nresp=2,interaction=seq(1:3),
+ coefnames=coefnames23,family="poisson",corstr="unstructured")
R> summary(mmm23)$coef
```

The results are displayed in Table 3.11 under Model 2 column. As mentioned above, in contrast to Model 1, Model 2 indicated statistically significant difference in intercept and covariate effect (X) on the first and second responses; corresponding robust Z values are 5.38 and -3.41, respectively. It further indicated that the time and interaction of time and X could be shared by these bivariate responses; the corresponding robust Z values are -0.22 and 0.23, respectively. Following the results of Model 2, we allow the time effect and interaction between time and X to be shared by multiple responses by fitting Model 3. This model fit and related results could be obtained by

```
R> library("mmm2")
R> coefnames24<-c("intercept","X","time","X*time","resptype","X*resptype")
R> mmm24<-mmm2(data=data14,nresp=2,interaction=c(1),coefnames=coefnames24,
+ family="poisson",corstr="unstructured")
R> summary(mmm24)$
```

The results are displayed in Table 3.11 under Model 3 column. This MMM2 formulation permit fitting 2 less regression parameters compared to the MMM1 given in (3.8), while indicating very similar parameter estimates. For example, while MMM2 estimated the intercept for the first response as -1.98, MMM1 estimated it as -1.97 (Table 3.10).

Table 3.11: Multivariate and univariate marginal model analyses of multivariate longitudinal count data.

	Model 1			Model 2			Model 3		
	Est.	(SE)	Z	Est.	(SE)	Z	Est.	(SE)	Z
Intercept	-1.21	(0.11)	-11.12	-1.98	(0.17)	-11.76	-1.98	(0.17)	-11.76
X	1.03	(0.04)	29.31	1.20	(0.06)	21.78	1.20	(0.06)	21.78
time	-0.03	(0.05)	-0.60	-0.02	(0.08)	-0.31	-0.04	(0.05)	-0.67
X*time	0.01	(0.02)	0.73	0.01	(0.03)	0.39	0.01	(0.02)	0.80
rtype				1.26	(0.23)	5.39	1.26	(0.23)	5.38
X*rtype				-0.26	(0.08)	-3.41	-0.26	(0.08)	-3.40
time*rtype				-0.02	(0.09)	-0.22			
X*time*rtype				0.01	(0.03)	0.23			

Note: Only robust standard error and Z estimates are reported. While Models 2 & 3 were fitted under unstructured working correlation structure, Model 1 was fitted under exchangeable working correlation assumption.

Table 3.12: Computational times consumed for the analyses of MLCD.

Model	Time (in seconds)
MMM1	1.13
MMM2 - Model 1	8.72
MMM2 - Model 2	9.25
MMM2 - Model 3	8.87

The computational times required for model fitting of MLCD are placed in Table 3.12. The analyses were done on a PC with 4.00 GB RAM and 2.53 GHz processor. While `mmm` (MMM1) took very short time (1.13 seconds), `mmm2` (MMM2) took relatively more time.

3.4.3 Multivariate Longitudinal Continuous Data Applications

The validity of MMM1 and MMM2 for multivariate longitudinal continuous data was investigated as well. As in the case of multivariate longitudinal count data, a simulated dataset was considered. The data simulation was done by the help of an R library called **mvtnorm** (Genz et al., 2011). For simplicity, bivariate longitudinal continuous responses measured at four time points and a covariate measured only at baseline were considered. The bivariate responses were assumed to have different variances which are 0.97 and 1.1 for the first and second responses, respectively. Additionally, the variances of the responses were assumed to be constant across time and their serial correlation structures were assumed to have an AR(1) type correlation structure. The time independent covariate was assumed to have a variance

4, and its simulation was succeeded together with the simulation of the responses. These assumptions regarding data simulation are arbitrary. Besides, we think that they are realistic. The data simulation via **mvtnorm** could be achieved by

```
R> library("mvtnorm")
R> n2<-500
R> set.seed(12)
R> cormat<-matrix(c(1,0.4,0.6,0.9,0.37,0.8,0.34,0.7,0.31,
+ 0.4,1,0.6,0.37,0.9,0.34,0.8,0.31,0.7,
+ 0.6,0.6,1,0.6,0.6,0.6,0.6,0.6,0.6,
+ 0.9,0.37,0.6,1,0.4,0.9,0.37,0.8,0.34,
+ 0.37,0.9,0.6,0.4,1,0.37,0.9,0.34,0.8,
+ 0.8,0.34,0.6,0.9,0.37,1,0.4,0.9,0.37,
+ 0.34,0.8,0.6,0.37,0.9,0.4,1,0.37,0.9,
+ 0.7,0.31,0.6,0.8,0.34,0.9,0.37,1,0.4,
+ 0.31,0.7,0.6,0.34,0.8,0.37,0.9,0.4,1),ncol=9,byrow=T)
R> variance<-c(0.97,1.1,4,0.97,1.1,0.97,1.1,0.97,1.1)
R> std<-diag(sqrt(variance),9)
R> sigma<-std%*%cormat%*%std
R> data2<-rmvnorm(n2,mean = rep(0,nrow(sigma)),sigma=sigma,method="svd")
```

The dataset should be reconstructed to be analyzed by `mmm1` and `mmm2`. The data reconstruction could be done by

```
R> time21<-data2[,1:3]
R> time22<-cbind(data2[,4:5],data2[,3])
R> time23<-cbind(data2[,6:7],data2[,3])
R> time24<-cbind(data2[,8:9],data2[,3])
R> data22<-rbind(time21,time22,time23,time24)
R> time2<-matrix(rep(seq(1:4),each=n2))
R> id2<-matrix(rep(seq(1:n2),4))
R> data23<-cbind(id2,data22,time2)
R> data24<-NULL
R> for (i in 1:n2){
R> data24<-rbind(data24,data23[data23[,1]==i,])
R> }
```

In addition to the invariant covariate, standardized time and the interaction of the time invariant covariate and time were added into the model. The data manipulations could be done by

```
R> data24[,5]<-scale(data24[,5])
R> data24<-cbind(data24,data24[,4]*data24[,5])
```

Note that we call the simulated multivariate longitudinal continuous (gaussian) data, `data24` as MLGD throughout. Also note that since we simulated MLGD by fixing `seed=12`, it is reproducible as well.

3.4.3.1 Application of Multivariate Marginal Models with Response Specific Regression Parameters

A multivariate marginal models with response specific regression parameters (MMM1) for MLGD is illustrated below.

$$E(Y_{itj}|X_{it}) = \beta_{0j} + \beta_{1j} * X_{it} + \beta_{2j} * time_t + \beta_{3j} * (X_{it} * time_t) \tag{3.13}$$

As in the previous applications, setting j=1 in (3.13) corresponds to first longitudinal continuous response and j=2 corresponds to the second longitudinal continuous response. The model fit via `mmm` could be obtained by

```
R> library("mmm")
R> coefnames3<-c("Y1.int","Y1.beta1","Y1.beta2","Y1.beta3",
+ "Y2.int","Y2.beta1","Y2.beta2","Y2.beta3")
R> mmm3<-mmm(data=data24,nresp=2,family="gaussian",
+ coefnames=coefnames3,corstr='unstructured')
```

The multivariate and univariate marginal modeling results produced by `mmm` could be obtained by

```
R> mmm3$multiv$multivout$coef
R> mmm3$univ$univout[[1]]$coef
R> mmm3$univ$univout[[2]]$coef
```

Table 3.13: Multivariate and univariate marginal model analyses of multivariate longitudinal continuous data.

	multivariate results			univariate results		
	Est.	(SE)	Z	Est.	(SE)	Z
Response 1						
Intercept	0.02	(0.03)	0.65	0.03	(0.03)	0.98
X	0.28	(0.01)	19.57	0.28	(0.01)	20.61
time	0.02	(0.01)	1.65	0.02	(0.01)	1.63
X*time	0.004	(0.01)	0.64	0.01	(0.01)	1.00
Response 2						
Intercept	0.0003	(0.04)	0.08	-0.02	(0.03)	-0.71
X	0.28	(0.02)	14.37	0.29	(0.02)	18.89
time	0.03	(0.02)	1.32	0.02	(0.01)	1.22
X*time	0.001	(0.01)	0.13	0.01	(0.01)	0.99

Note: Only robust standard error and Z estimates are reported. The models were fitted under unstructured working correlation assumption.

The related results are placed in Table 3.13. As in the previous applications, the multivariate and univariate marginal modeling regression parameter estimates seemed to be in agreement with each other in terms of magnitude and direction, standard errors and statistical significance. Nevertheless, multivariate and univariate models produced intercepts with reverse directions for the second response which are 0.0003 and -0.02, respectively. Moreover, multivariate and univariate models produced parameter estimates having different amount of effects (almost 10 times the other) of time independent covariate and time on the second response which are 0.001 and 0.01, respectively. Nonetheless, both the intercepts having reverse directions and interactions having different amount of effects were statistically insignificant.

3.4.3.2 Application of Multivariate Marginal Models with Shared Regression Parameters

A typical multivariate marginal models with shared regression parameters (MMM2) for MLGD could be formulated by

$$E(Y_{itj}|X_{itj}) = \beta_0 + \beta_1 * X_{itj} + \beta_2 * time_{tj} + \beta_3 * (X_{itj} * time_{tj}) \tag{3.14}$$

The MMM2 given in (3.14) assumes that the intercept and the covariate effects are same for the first and second responses. The related model fit and the estimates could be obtained by

```
R> library("mmm2")
R> coefnames32<-c("intercept","X","time","X*time")
R> mmm32<-mmm2(data=data24,nresp=2,rtype=FALSE,coefnames=coefnames32,
+ family="gaussian",corstr="unstructured")
summary(mmm32)$coef
```

The results are displayed in Table 3.14 under Model 1 column. As expected, it yielded estimates averaged over the bivariate responses. While MMM1 (3.14) estimated the intercepts as 0.02 and 0.0003 for first and second responses, respectively, the MMM2 (3.14) estimated an average intercept, 0.01=(0.02+0.0003)/2.

As in the previous examples (Sections 3.4.1 and 3.4.2), equivalent models could be fitted by MMM1 and MMM2 formulations by introducing response types and the interactions of covariates and response types as covariates. An MMM2 which is equivalent to the MMM1 (3.14) can be formulated as below.

$$
\begin{aligned}
E(Y_{itj}|X_{itj}) = \beta_0 + \beta_1 * X_{itj} + \beta_2 * time_{tj} + \beta_3 * (X_{itj} * time_{tj}) + \beta_4 * rtype_j + \\
\beta_5 * (X_{itj} * rtype_j) + \beta_6 * (time_{tj} * rtype_j) + \beta_7 * (X_{itj} * time_{tj} * rtype_j)
\end{aligned}
\tag{3.15}
$$

Following the `rtype` argument of `mmm2`, let's assign the response type=0 for the first continuous response, and response type=1 for the second one. The MMM2 given in (3.15) actually indicates two different models which are distinguished by the response type variable. For example, for the first response it indicates

$$
E(Y_{it1}|X_{it1}) = \beta_0 + \beta_1 * X_{it1} + \beta_2 * time_{tj} + \beta_3 * (X_{it1} * time_{tj}) \tag{3.16}
$$

On the other hand, for the second response it indicates

$$
E(Y_{it2}|X_{it2}) = (\beta_0 + \beta_4) + (\beta_1 + \beta_5) * X_{it2} + (\beta_2 + \beta_6) * time_{tj} + (\beta_3 + \beta_7) * (X_{it2} * time_{tj}) \tag{3.17}
$$

The related model fit and parameter estimates via `mmm2` could be obtained by

Table 3.14: Multivariate and univariate marginal model analyses of multivariate longitudinal continuous data.

	Model 1			Model 2		
	Est.	(SE)	Z	Est.	(SE)	Z
Intercept	0.01	(0.02)	0.47	0.02	(0.03)	0.65
X	0.28	(0.01)	25.22	0.28	(0.01)	19.57
time	0.03	(0.02)	1.49	0.02	(0.01)	1.65
X*time	0.002	(0.01)	0.24	0.004	(0.01)	0.64
rtype				-0.02	(0.06)	-0.37
X*rtype				-0.001	(0.03)	-0.05
time*rtype				0.01	(0.02)	0.36
X*time*rtype				-0.003	(0.01)	-0.28

Note: Only robust standard error and Z estimates are reported. Models 1 and 2 were fitted under unstructured working correlation assumption.

Table 3.15: Computational times consumed for the analyses of MLGD.

Model	Time (in seconds)
MMM1	1.01
MMM2 - Model 1	8.69
MMM2 - Model 2	9.24

```
R> library("mmm2")
R> coefnames33<-c("intercept","X","time","X*time","resptype","X*resptype",
+ "time*resptype","X*time*resptype")
R> mmm33<-mmm2(data=data24,nresp=2,interaction=seq(1:3),coefnames=coefnames33,
+ family="gaussian",corstr="unstructured")
R> summary(mmm33)$coef
```

The results are displayed in Table 3.14 under Model 2. The MMM1 (Table 3.13) estimated the coefficient of the interaction of the covariate (X) and time for the second response as 0.001. It can be calculated by the equivalent MMM2 by 0.004+(-0.003)=0.001. Moreover, the MMM2 (3.15) indicates that all the covariates have same effects on the bivariate responses. By the MMM2 formulation (Model 1), we are able to estimate 4 less parameters compared to the MMM1 formulation.

The computational times required for model fitting of the multivariate longitudinal continuous dataset are placed in Table 3.15. While `mmm` (MMM1) took very small time (1.01 seconds), `mmm2` (MMM2) took relatively more time.

3.5 Discussion and Conclusion

In this study, we discussed multivariate marginal models with response specific regression parameters (MMM1). Additionally, we introduced multivariate marginal models with shared regression parameters (MMM2). We proposed two R packages **mmm** and **mmm2** to fit these models, respectively. We showed that MMM1 and MMM2 are not restricted to any type of response families, but they handle several response families. We presented the applications of them with multivariate longitudinal continuous, binary and count data. Furthermore, we provided the details and the usage of the R packages, **mmm** and **mmm2**. In addition to the multivariate marginal models, **mmm** fits univariate marginal models for the multivariate responses one by one. Naive variance and Z estimates are produced together with the robust (sandwich) estimates in both **mmm** and **mmm2**. The usages of these packages are straightforward and parsimonious in terms of computational time. We displayed the computational time required by MMM1 and MMM2 via `mmm` and `mmm2`, respectively for datasets with different amount of time points, covariates and response families. While `mmm` takes very short time for model fitting `mmm2` takes relatively longer time. Although we consider same number of repeated measures, t for each subjects, both `mmm` and `mmm2` can handle varying number of repeated measures, n_i.

The statistical inference made by MMM1 and MMM2 could be reliable only under missing completely at random (MCAR) data, since they use conventional generalized estimating equations (GEE) for parameter estimation. One drawback of these models is that in some cases they might fail in the estimation of the regression parameters with the use of unstructured working correlation matrix; especially when the cluster sizes are large. This failure is natural since the number of correlation parameters to be estimated by GEE in the framework of multivariate marginal modeling increases rapidly compared to a univariate one. While the number of correlation parameters to be estimated by a univariate marginal model is $t*(t-1)/2$, it is $(k*t)*(k*t-1)/2$ for a multivariate marginal model.

Multivariate marginal models presented in this chapter could be extended to handle the missing data mechanisms other than MCAR. New working correlation structures which estimates less association parameters while reflecting both the multivariate and serial dependence structures compared to unstructured working correlation structure could be developed. These are on-going works.

CHAPTER 4

PROBIT NORMAL MARGINALIZED TRANSITION RANDOM EFFECTS MODELS

In this chapter, we introduce a marginalized model for multivariate longitudinal binary data called Probit Normal Marginalized Transition Random Effects Models (PNMTREM) which is an extension of Marginalized Transition Random Effects Models (MTREM) by Ilk and Daniels (2007). The extensions we proposed are in terms of link functions and parameter estimation procedure. Additionally, we introduce an R package **pnmtrem** prepared for that model. The chapter is divided into seven sections. We motivate our study in Section 4.1. In Section 4.2, we provide details of the model. Section 4.3 discusses the parameter estimation procedure. In Section 4.4, we validate the correctness of the written code and the derivations done for parameter estimation by conducting a simulation study. Section 4.5 introduces a real life data application of our model and discuss parameter estimations and interpretations. In Section 4.6, we introduce the R package **pnmtrem** and illustrate the usage of it. We close Chapter 4 by discussions and conclusions regarding PNMTREM and **pnmtrem**.

4.1 Motivation

As is indicated throughout the thesis, longitudinal data arise when same subjects/units are followed repeatedly over time and this type of data are mostly not independent. Also, longitudinal data often include multiple responses. Mother's Stress and Children's Morbidity Study (MSCM; Alexander and Markowitz, 1986) and Iowa Youth and Families Project (IYFP; Ilk, 2008) are two such examples which include multiple longitudinal responses (for details of these datasets, see Chapter 1). In MSCM Study, 167 mothers and their children were followed for 28 days mainly to investigate the effect of mother's employment status on the use of pediatric care. The bivariate longitudinal binary responses are mother's stress status (0=absence, 1=presence) and children's morbidity status (0=absence, 1=presence). In IYFP, 7th graders with two alive and biological parents and a sib within 4 years of age from eight counties of Iowa were followed for 11 years by 8 follow-up visits. The main aim of the project is to investigate the long term effects of the farm crisis occurred in 1980's in America on family life, especially on well being of young people living in rural regions. The trivariate longitudinal binary responses are anxiety (0=absence, 1=presence), hostility (0=absence, 1=presence) and depression (0=absence, 1=presence) statuses of young people. In addition to the serial dependence introduced by univariate longitudinal data structures, multivariate longitudinal data introduces a second type of association which is multivariate response dependence at a specific time point. For illustration of these two types of associations with MSCM and IYFP datasets, see Figures 1.1 and 1.2. Despite the fact that these associations are not of initial interest, they should be taken into account to make proper statistical inferences.

Generalized linear models with random effects and/or serial dependence are used widely to analyze longitudinal data. Nevertheless, they are challenging in terms of computation and interpretation of the marginal regression parameter estimates. Heagarty proposed two marginalized models for longitudinal binary data with the characteristics of random effects (Heagerty, 1999) and serial dependence (Heagerty, 2002b) structures. Both of those models are likelihood-based and have two-level modeling framework. In both of them, while the first level accommodates the marginal effects of the independent variables on longitudinal responses, second levels capture the longitudinal association between the repeated measurements. One outstanding feature of those models is that since the effects of the independent variables on the responses are modeled free of the association structures, they yield marginal interpretation of those effects just like in generalized linear models case while still permitting subject-specific estimation, e.g. by the random effects structure. Moreover, the estimates of those effects are robust even under misspecification of the association structures (Heagerty and Kurland, 2001). Heagerty and Zeger (2000), Diggle et al. (2002), Mills et al. (2002), Wang and Louis (2004), Schildcrout and Heagerty (2005), Ilk and Daniels (2007) and Lee et al. (2009) are other works which considered marginalized modeling of longitudinal binary data; latter two deals with multivariate data.

As stated earlier, marginal models are the extension of generalized linear models for longitudinal data. Also, since there is lack of literature for multivariate discrete distributions, full likelihood based approaches for parameter estimation are not possible or computationally feasible. Some exceptions could be found in Molenberghs and Verbeke (2005) such as the Bahadur model. Although this model has a closed form formulation, it has limitations in terms of computation and number of repeated measures. This is due to the fact that the model has a complicated and constraint parameter space. For instance, Molenberghs and Kenward (2010) reported that the second order Bahadur model (here second order model corresponds to the one which considers all pairwise correlations) with 12 repeated measures for each subject introduces pairwise correlations which are bounded from above by a number in the interval of [0.09,0.18]. Generalized estimating equations (GEE; Liang and Zeger, 1986 and Zeger and Liang, 1986) are alternative to full likelihood based methods and are able to compete with them in many cases (Fitzmaurice et al., 2004). One great advantage of them is that they produce consistent parameter and their variance estimates even under incorrect choice of the association structure of the repeated measures, so called working correlation matrix. However, they are robust only under missing completely at random data (MCAR). Inverse Probability weighting methods can handle other types of missing data mechanisms (Robins et al., 1995; Rotnitzky et al., 1998). Also there is lack of literature for model selection with this semi-parametric approach and classical methods do not hold for GEE since they do not contain a genuine likelihood function.

It is well known that full likelihood methods are robust even under missing at random data (MAR; Molenberghs and Kenward, 2007). Additionally, likelihood ratio test (LRT) could be used for model selection among nested models. Luckily, Akaike's Information Criterion (AIC; Akaike, 1973), Schwarz's (Bayesian) Information Criterion (BIC; Schwarz, 1978) and Deviance Information Criterion (DIC; Spiegelhalter et al., 2002) are available for model selection among non-nested models. One additional facility of likelihood based methods is that they allow observing uncertainty regarding the parameter estimates, namely profile likelihoods. Maximum Likelihood Estimation and Bayesian Approach could be considered as the full likelihood based methods. Each of these methods are widely used in longitudinal data analysis and has their own properties. Maximum likelihood estimation assumes fix parameters and searches for the parameter estimates that maximizes the likelihood function, and asymptotic properties of these estimators are well established but they rely on large samples. However, Bayesian approach assumes varying parameters meaning parameters have their own distributions and samples from their posterior distributions, e.g. by utilizing Markov Chain Monte Carlo methods (MCMC), and does not necessarily rely on large samples. While the former yields confidence intervals with boundaries having equal distance to the center (point estimate), the latter permits such intervals with probably unequal distant boundaries to the center. Whereas maximum likelihood estimation requires integration over the random effects distribution (in models with random effects structures) with numerical integration methods, yielding approximate solutions of the integrals, Bayesian approach does not require such approximations. Nonetheless, the Bayesian approach largely relies upon subjective expertise, such as in the choice of prior distributions, selection of the hyper-parameters (parameters of the prior distributions), and interaction with tuning parameters. Moreover, the related methods are computationally intensive.

In this thesis, we propose a marginalized model via probit link which was originally proposed via logit links (Ilk and Daniels, 2007). *logit* and *probit* are two widely used link functions to linearize the effects of the independent variables on the dependent ones in the concept of regression analyses for categorical data. While the former is defined as the inverse of the cumulative distribution function (CDF) of the standard logistic distribution and at the same time log odds of success, the latter is defined as the inverse of the CDF of the standard normal distribution. They reflect similar behavior of placing probabilities and are almost indistinguishable except for very low and very high probabilities; logit lets more probability to the tails (Hedeker and Gibbons, 2006; McCullagh and Nelder, 1989). Often, large amounts of high quality data are needed to detect substantial differences between the conclusions drawn from the regression models with those link functions (Doksum and Gakso, 1990, cited in Hedeker and Gibbons, 2006, pp. 153). Regression analyses with logit link produce parameter estimates with a large scale compared to probit counterparts, since while standard normal distribution has a variance 1, standard logistic distribution has a variance of $\pi^2/3$ (Agresti, 2002). While logit yields direct interpretation of the regression coefficients, e.g. change in the natural logarithm of odds ratios, this job with probit is more challenging. Additionally, approximate transition between logit and probit regression parameter estimates are possible (Agresti, 2002; Griswold, 2005). For example Johnson et al. (1995, pp. 113-163 cited in Griswold, 2005 pp. 85-96) proposed a constant (JKB constant) between logit and probit regression parameter estimates such as $\beta_{logit} = c * \beta_{probit}$ where $c = (15/16)(\pi/\sqrt{3})$. For a comparative application and discussion of other possible approximations between logit and probit estimates, see Griswold (2005). Probit link is widely used in econometric, medical and genetic studies. Some of the works utilizing probit link in the concept of longitudinal data mixed modeling include Gibbons and Hedeker (1994), Gibbons et al. (1994), Gibbons and Hedeker (1997), Hedeker and Gibbons (1994), Gueorguieva and Agresti (2001). For other possible choices of link functions for binary data such as complementary log-log and log-log link functions, see McCullagh and Nelder (1989).

There is lack of literature for multivariate longitudinal binary data. Lee et al. (2009) proposed a marginalized model for multivariate longitudinal binary data (Multivariate Marginalized Random Effects Models, MMREM) extending the work Heagerty (1999). MMREM is a likelihood based model and has a framework of two logistic regression models. While the first level directly models covariate effects, the second level accommodates both the serial and multivariate response dependencies via random effects. They considered maximum likelihood estimation (MLE) alongside with quasi-Monte Carlo method for parameter estimation. Bartolucci and Farcomeni (2009) proposed a mixed model for multivariate longitudinal binary and ordinal data. Their model is a single level model which accommodates both the covariate effects and serial and multivariate response dependencies at the same level. They considered MLE via Expectation-Maximization (EM) Algorithm for parameter estimation. Asar and Ilk (2012) proposed an R package **mmm** for multivariate marginal modeling of multivariate longitudinal data by extending the methodology of Shelton et al. (2004); see Chapter 3. **mmm** utilizes generalized estimating equations (GEE) for parameter estimation and accommodates both the serial and multivariate response dependencies within the *working correlation matrix*.

Ilk and Daniels (2007) proposed a marginalized model (Marginalized Transition Random Effects Models, MTREM) for multivariate longitudinal binary data which extends the aforementioned two works of Heagerty (1999, 2002b). The model has a structure of three level logistic regression models (i.e., uses logit link). The first level accommodates the direct effects of the independent variables on the mean longitudinal responses. While the second level captures the effects of the response history on the current ones by transition parameters, the third level captures the multivariate response dependency by random effects. While the first level permits population-average interpretation, the latter ones permit subject-level interpretation. Bayesian approach specifically Gibbs sampling with Hybrid Monte Carlo (MC) steps (Neal, 1996), was used for parameter estimation. FORTRAN codes of that model for lags 1 and 2 could be reached at Ilk (2011a). This version of MTREM is computationally cumbersome meaning it requires long time for model fitting and one needs to interact for tuning parameters. This causes limitations in terms of the model usage.

In this study, we consider an extension of MTREM in terms of link functions and parameter estimation procedure. The link functions are changed from logit to probit. Probit link permits explicitly linking second and third levels of the model. Also it eases linkage of the first and second levels of the model. Maximum likelihood estimation (MLE) with Gauss-Hermite quadratures are utilized for parameter estimation. Although MLE requires numerical integrations over the random effects distributions, it still reduces the computational burden considerably. A user-friendly R package **pnmtrem** were prepared for this new version of MTREM.

4.2 Model

Multivariate longitudinal data introduce two types of associations. While the first one corresponds to the within-subject association for each longitudinal response, the second one corresponds to the multivariate response association at a specific time point. To accommodate these two types of associations together with the marginal mean modeling, Ilk and Daniels (2007) proposed a marginalized model (MTREM) which is a framework of three logistic regressions. The model considered here inherits the framework and the properties of their model except the link functions and parameter estimation methodology. For the MTREM with logit link, interested reader may refer to Ilk and Daniels (2007), Ilk (2008), Akinc (2008) and Yalcinoz (2008).

4.2.1 General Probit Normal Marginalized Transition Random Effects Models, PN-MTREM(p)

Let Y_{itj} be the j^{th} ($j = 1, \ldots, k$) response for the i^{th} ($i = 1, \ldots, n$) subject at time t ($t = 1, \ldots, T$) and X_{itj} be the associated set of covariates. X_{itj} might include time-variant and/or time-invariant covariates. Also let $\Phi(.)$ be the cumulative distribution function of the standard normal distribution. Use of inverse *probit* link yields the following representation of the model.

$$P^m_{itj} \equiv P(Y_{itj} = 1|X_{itj}) = \Phi(X_{itj}\beta) \tag{4.1}$$

$$P^t_{itj} \equiv P(Y_{itj} = 1|y_{i,t-1,j}, .., y_{i,t-p,j}, X_{itj}) = \Phi(\Delta_{itj} + \sum_{m=1}^{p} \gamma_{itj,m} y_{i,t-m,j}) \tag{4.2}$$

$$P^r_{itj} \equiv P(Y_{itj} = 1|y_{i,t-1,j}, ..., y_{i,t-p,j}, X_{itj}, b_{it}) = \Phi(\Delta^*_{itj} + \lambda_j b_{it}) \tag{4.3}$$

In the first level of the model given in the (4.1), β are the marginal regression coefficients that directly accounts the covariate effects on the mean responses, i.e. the covariate effects are not conditioned on either the response history or the random effects. They allow comparing sub-groups of the covariates such as females vs. males and notice that related interpretations are free of the serial/multivariate response dependencies. Although it is assumed that the intercept and the slopes (covariate effects) are shared by different responses (same β for different responses), the inclusion of response types as covariates (indicator variables) allow different responses to have different intercepts. Similarly, the inclusion of the interactions of response types and covariates allow different responses have different slopes. This construction of the model provides model flexibility such as fitting a more parsimonious model when the covariate effects on multiple responses do not differ. Therefore, it might yield parameter estimates with low variances. Typical setup of the model assumes that only the covariates at time t have significant effects on the responses at that time point, i.e., $P(Y_{itj} = 1|X_{i1j}, \ldots, X_{itj})$ $= P(Y_{itj} = 1|X_{itj})$. Nevertheless, lagged covariates might be included in the model.

In the second level of the model given in the (4.2), $\gamma_{itj,m} = \alpha_{t,m} Z_{itj,m} = \alpha_{t1,m} Z_{itj1,m} + ... + \alpha_{tl,m} Z_{itjl,m}$ for $m = 1, \ldots, p$ where $\alpha_{t,m}$ are the time/covariate/order specific transition parameters which accommodates the effect of the past responses on the current ones by taking into account the interactions between the past responses and a subset of covariates as well; p is the order of the transition model and Z_{itj} are a subset of covariates with l variables. Note that Z_{itj} has the form of a design matrix, i.e. includes 1's on the first column. Choices of Z_{itj} permit various association structures between the current and past responses. For example, if the effects of the lag-1 responses on the current ones are different for males and females, then gender could be included in $Z_{itj,1}$. Similar to the first level, although the transition parameters, $\alpha_{t,m}$ are shared across multiple responses (common parameters for different responses), the inclusion of interaction(s) between the response types and the response history allows these parameters to differ for multiple responses.

In the third level of the model given in the (4.3), b_{it} are subject/time specific random effects and it is assumed that $b_{it} \sim \mathrm{N}(0, \sigma_t^2)$. b_{it} can be re-written as $b_{it}=\sigma_t\, z_i$ where z_i are the standard normal random variables; this version of b_{it} is useful in numerical integration which will be introduced later. b_{it} measures the unobserved heterogeneity between the subjects at time t. λ_j are response specific parameters that scale the random effects with respect to response j and accommodates the multivariate response dependence together with them. For identifiability, λ_1 is set to 1. Note that by allowing the random effects changing over time, i.e. having index t in b_{it}, the model accommodates differing multivariate response dependencies at different time points.

An adaptation of the first order Taylor series approximation to the form of the pairwise correlations between the multiple responses from logit link to probit by following Ilk and Daniels (2007) yields

$$cor(y_{itj}, y_{itj'}) = \frac{\lambda_j \lambda_{j'} \sigma_t^2 \Phi(\Delta^*_{itj})(1-\Phi(\Delta^*_{itj}))\Phi(\Delta^*_{itj'})(1-\Phi(\Delta^*_{itj'}))}{\sqrt{(\lambda_j^2 \Phi^2(\Delta^*_{itj})(1-\Phi(\Delta^*_{itj}))^2\sigma_t^2 + \Phi(\Delta^*_{itj})(1-\Phi(\Delta^*_{itj})))}\sqrt{(\lambda_{j'}^2 \Phi^2(\Delta^*_{itj'})(1-\Phi(\Delta^*_{itj'}))^2\sigma_t^2 + \Phi(\Delta^*_{itj'})(1-\Phi(\Delta^*_{itj'})))}} \tag{4.4}$$

The interpretation of λ_j and σ_t^2 could be understood most easily by setting $\Phi(\Delta^*_{itj}) = \Phi(\Delta^*_{itj\prime})$ for all j and $j\prime$ and $\Phi(\Delta^*_{itj}) = \Phi(\Delta^*_{it'j})$ for all t and t'. Since λ_j are response specific parameters, they allow multivariate responses to have different pairwise correlations at a given time, i.e., $cor(y_{itj}, y_{itj'}) \neq cor(y_{itj}, y_{itj''})$ for all $j \neq j' \neq j'' (= 1, \ldots, k)$. Put another way, setting $\lambda_j = 1$ for $j = 2, \ldots, k$ (λ_1 is already set to 1 for identifiability) yields multivariate responses having equal correlation among themselves conditioning on the response history. Allowing the random effects variance changing with respect to time, i.e. having index t in σ_t^2, permits the multivariate response dependencies changing over time, i.e. $cor(y_{itj}, y_{itj'}) \neq cor(y_{itj}, y_{it'j'})$ for all $j \neq j' (= 1, \ldots, k)$ and $t \neq t' (= 1, \ldots, T)$. In other words, setting $\sigma_t^2 = \sigma^2$ for $t = 1, \ldots, T$ makes the multivariate response dependencies constant over time. Additionally, PNMTREM captures cross-response temporal dependencies implicitly (not directly accounted by the modeling framework), i.e. $cor(y_{itj}, y_{it'j'})$. This is due to the fact that accommodating serial and multivariate response dependencies at the same time permits the model naturally to induce the cross-response temporal dependencies. Ilk and Daniels (2007) investigated the behavior of MTREM with an application to IYFP dataset and reported that it captured those cross-response temporal dependencies very well.

In (4.2, 4.3), Δ_{itj} are subject/time/response specific intercepts that take into accounts the non-linear relationship between the marginal and transition probabilities (P^m_{itj} and P^t_{itj}, respectively). Similarly, Δ^*_{itj} are the subject/time/response specific intercepts that capture the non-linear relationship between the transition and random effects probabilities (P^t_{itj} and P^r_{itj}, respectively).

One of the inherited features of PNMTREM from the original setup of MTREM is that the conditional mean of the responses given all set of covariates is equal to the conditional mean of the responses given the covariate history, i.e., $E(Y_{itj}|X_{iqj}, q = 1, \ldots, T) = E(Y_{itj}|X_{isj}, s \leq t)$. This assumption is vital for the validity of the marginal constraint equation which will be introduced later while linking the levels of the model. However, the assumption is meaningful for exogenous covariates (covariates which do not depend on response history, at time t) but not meaningful for the endogenous ones (covariates which depend on response history at time t).

The three-level model specification of PNMTREM presented in this section completes the multivariate distribution of the multivariate longitudinal binary data.

4.2.2 First Order Probit Normal Marginalized Transition Random Effects Models, PNMTREM(1)

Here, we discuss first order PNMTREM, PNMTREM(1), which is a specialized form of the general PNMTREM, PNMTREM(p). PNMTREM(1) considers only the effects of lag-1 responses on the current ones in the second level of the model formulation and the related modeling framework is given by

$$P^m_{itj} \equiv P(Y_{itj} = 1|X_{itj}) = \Phi(X_{itj}\beta) \tag{4.5}$$

$$P^t_{itj} \equiv P(Y_{itj} = 1|y_{i,t-1,j}, X_{itj}) = \Phi(\Delta_{itj} + \gamma_{itj,1} y_{i,t-1,j}) \tag{4.6}$$

$$P^r_{itj} \equiv P(Y_{itj} = 1|y_{i,t-1,j}, X_{itj}, b_{it}) = \Phi(\Delta^*_{itj} + \lambda_j b_{it}) \tag{4.7}$$

where $b_{it} \sim$ N(0, σ_t^2) and b_{it}=z_i σ_t, $z_i \sim$ N(0,1); λ_1=1. Again, $\gamma_{itj,1} = \alpha_{t,1} Z_{itj,1} = \alpha_{t1,1} Z_{itj1,1} + \ldots + \alpha_{tl,1} Z_{itjl,1}$ where $Z_{itj,1}$ are a subset of covariates. Note that throughout we call this model as $t \geq 2$ model.

Since for baseline (t=1) no history data are available, second level of PNMTREM given in (4.6) is not valid anymore. Due to the fact that in longitudinal studies, commonly baseline data reflect more variability and has different covariate effects compared to later time points (t ≥ 2), a separate model is constructed for that time point. The baseline model is given below.

$$P^{m}_{i1j} \equiv P(Y_{i1j} = 1|X_{i1j}) = \Phi(X_{i1j}\beta^{*}) \tag{4.8}$$

$$P^{r}_{i1j} \equiv P(Y_{i1j} = 1|X_{i1j}, b_{i1}) = \Phi(\Delta^{*}_{i1j} + \lambda^{*}_{j} b_{i1}) \tag{4.9}$$

where $b_{i1} \sim$ N(0, σ_1^2) and b_{i1}=z_i σ_1, $z_i \sim$ N(0,1); λ^{*}_{j}=1. Note that throughout we call this model as baseline model.

The modeling diagram of PNMTREM(1) could be illustrated in Figure 4.1. For simplicity, it is assumed that there are bivariate responses and the data are collected on two time points. Similarly, possible interaction effects of the covariates and past responses are suppressed in the transition parameters, α.

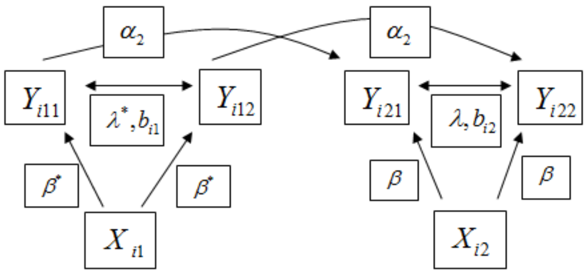


Figure 4.1: Modeling Schemes of PNMTREM(1)

4.2.2.1 Linking Levels of PNMTREM(1) for t ≥ 2 Model

To be a valid probabilistic model, the levels of PNMTREM(1) are connected to each other by constraint equations. Throughout this section we illustrate these constraints.

Linking First and Second Levels of PNMTREM(1)

Level 1 (4.5) and level 2 (4.6) of PNMTREM(1) are linked via a marginal constrained equation which is illustrated by

$$P(Y_{itj} = 1|X_{itj}) = \sum_{y_{i,t-1,j}} P(Y_{itj} = 1|y_{i,t-1,j})P(y_{i,t-1,j}) \tag{4.10}$$

which is equivalent to

$$P^m_{itj} = \sum_{y_{i,t-1,j}=0}^{1} P^t_{itj}(P^m_{i,t-1,j})^{y_{i,t-1,j}}(1 - P^m_{i,t-1,j})^{1-y_{i,t-1,j}} \tag{4.11}$$

Re-writing the marginal and transition probabilities in (4.11) yields

$$\Phi(X_{itj}\beta) = \sum_{y_{i,t-1,j}=0}^{1} \Phi(\Delta_{itj} + \gamma_{itj,1}y_{i,t-1,j})(\Phi(X_{i,t-1,j}\beta))^{y_{i,t-1,j}}(1 - \Phi(X_{i,t-1,j}\beta))^{(1-y_{i,t-1,j})} \tag{4.12}$$

where $\gamma_{itj,1} = \alpha_{t,1}Z_{itj,1}$. Finally, we get the following non-linear equation as a solution of the marginal constraint equation given in (4.10).

$$\Phi(X_{itj}\beta) = \Phi(\Delta_{itj})(1 - \Phi(X_{i,t-1,j}\beta)) + \Phi(\Delta_{itj} + \gamma_{itj})\Phi(X_{i,t-1,j}\beta) \tag{4.13}$$

Note that for t=2, β^* replaces β as the multiplier of lag-1 covariates in the marginal constraint equation and yields

$$\Phi(X_{i2j}\beta) = \Phi(\Delta_{i2j})(1 - \Phi(X_{i,1,j}\beta^*)) + \Phi(\Delta_{i2j} + \gamma_{i2j})\Phi(X_{i,1,j}\beta^*) \tag{4.14}$$

Note that we suppress (4.14) in (4.13) for the sake of simplicity and take into account the difference when necessary. The non-linear equation given in (4.13) does not permit writing Δ_{itj} in terms of β and $\alpha_{t,1}$, explicitly. Luckily, *implicit function theorem (IFT)* allows us solving (4.13) for Δ_{itj} in terms of β and $\alpha_{t,1}$, explicitly.

Brief Overview of Implicit Function Theorem (IFT) (Modified from Section 12.8 of Adams and Essex, 2010)

Let's assume that there exist three different variables x, y and z. The question which IFT addresses is

"Does the equation $F(x, y, z)$=0 permit to define z as a function of x and y, i.e. z=$z(x, y)$ near some point P_0=(x_0, y_0, z_0) which satisfies the equation?"

If $F(x, y, z)$ does so and if it has continuous first partial derivatives near P_0, then the first partial derivatives of z can be found at (x_0,y_0) by implicit differentiation of $F(x, y, z)$=0 with respect to x and y which are given by

$$\frac{\partial F(x,y,z)}{\partial x} + \frac{\partial F(x,y,z)}{\partial z}\frac{\partial z}{\partial x}=0 \quad \text{and} \quad \frac{\partial F(x,y,z)}{\partial y} + \frac{\partial F(x,y,z)}{\partial z}\frac{\partial z}{\partial y}=0$$

which are equivalent to

$$\left.\frac{\partial z}{\partial x}\right|_{(x_0,y_0)} = - \frac{\left.\frac{\partial F}{\partial x}\right|_{(x_0,y_0,z_0)}}{\left.\frac{\partial F}{\partial z}\right|_{(x_0,y_0,z_0)}} \quad \text{and} \quad \left.\frac{\partial z}{\partial y}\right|_{(x_0,y_0)} = - \frac{\left.\frac{\partial F}{\partial y}\right|_{(x_0,y_0,z_0)}}{\left.\frac{\partial F}{\partial z}\right|_{(x_0,y_0,z_0)}}$$

provided that $\left.\frac{\partial F}{\partial z}\right|_{(x_0,y_0,z_0)} \neq 0$. This condition indicates that the level surface of F through P_0 does not have a normal vector, i.e. it is not parallel to the z-axis, since $\frac{\partial F}{\partial z}$ is the z component of the gradient (first derivative vector) of F. Therefore, it is guaranteed that part of the surface near P_0 must be the graph of a function z=$z(x, y)$, i.e. z=$z(x, y)$ exists near P_0.

Based on the facts of the implicit differentiation, IFT solves z as a function of x and y near P_0 by a first order approximation as

$$z(x, y) = -\frac{\left.\frac{\partial F}{\partial x}\right|_{(x_0,y_0,z_0)}}{\left.\frac{\partial F}{\partial z}\right|_{(x_0,y_0,z_0)}}(x - x_0) - \frac{\left.\frac{\partial F}{\partial y}\right|_{(x_0,y_0,z_0)}}{\left.\frac{\partial F}{\partial z}\right|_{(x_0,y_0,z_0)}}(y - y_0)$$

Other references of IFT include Krantz and Parks (2003) and Section 14.5 of Stewart (2008).

Application of IFT to PNMTREM(1)

Let F be a function of X_{itj}, X_{it-1j}, β, Δ_{itj}, α_{t1} and $Z_{itj,1}$ such that (by re-writing (4.13))

$$F(X_{itj}, X_{it-1j}, \beta, \Delta_{itj}, \alpha_{t1}, Z_{itj,1}) = \Phi(X_{itj}\beta) - \Phi(\Delta_{itj})(1 - \Phi(X_{i,t-1,j}\beta)) - \Phi(\Delta_{itj} + \alpha_{t,1}Z_{itj,1})\Phi(X_{i,t-1,j}\beta) = 0 \tag{4.15}$$

Then, by IFT and first order approximation, the Δ_{itj} could be obtained by

$$\Delta_{itj} = -\frac{\left.\frac{\partial F}{\partial \beta}\right|_{(\beta_0,\alpha_{t,10},\Delta_{itj0})}}{\left.\frac{\partial F}{\partial \Delta_{itj}}\right|_{(\beta_0,\alpha_{t,10},\Delta_{itj0})}}(\beta - \beta_0) - \frac{\left.\frac{\partial F}{\partial \alpha_{t,1}}\right|_{(\beta_0,\alpha_{t10},\Delta_{itj0})}}{\left.\frac{\partial F}{\partial \Delta_{itj}}\right|_{(\beta_0,\alpha_{t,10},\Delta_{itj0})}}(\alpha_{t,1} - \alpha_{t,10}) \tag{4.16}$$

where,

$$\frac{\partial F}{\partial \beta} = X_{itj}\phi(X_{itj}\beta) + \Phi(\Delta_{itj})(\phi(X_{i,t-1,j}\beta))X_{i,t-1,j} - \Phi(\Delta_{itj} + \alpha_{t,1}Z_{itj,1})\phi(X_{i,t-1,j}\beta)X_{i,t-1,j}$$
$$\frac{\partial F}{\partial \Delta_{itj}} = -\phi(\Delta_{itj})(1 - \Phi(X_{i,t-1,j}\beta)) - \phi(\Delta_{itj} + \alpha_{t,1}Z_{itj,1})(\Phi(X_{i,t-1,j}\beta))$$
$$\frac{\partial F}{\partial \alpha_{t,1}} = -\phi(\Delta_{itj} + \alpha_{t,1}Z_{itj})\Phi(X_{i,t-1,j}\beta)Z_{itj,1} \quad (4.17)$$

$\phi(.)$ is the probability density function of the standard normal distribution and $\beta_0, \alpha_{t,10}$ and Δ_{itj0} are the components of P_0 around which IFT searches for solution.

For t=2, β^* replace β as the multiplier of lag-1 covariates and yields

$$F(X_{itj}, X_{it-1j}, \beta, \Delta_{itj}, \alpha_{t1}, Z_{itj,1}) = \Phi(X_{itj}\beta) - \Phi(\Delta_{itj})(1 - \Phi(X_{i,t-1,j}\beta^*)) - \Phi(\Delta_{itj} + \alpha_{t,1}Z_{itj,1})\Phi(X_{i,t-1,j}\beta^*) = 0 \quad (4.18)$$

and

$$\frac{\partial F}{\partial \beta} = X_{itj}\phi(X_{itj}\beta)$$
$$\frac{\partial F}{\partial \Delta_{itj}} = -\phi(\Delta_{itj})(1 - \Phi(X_{i,t-1,j}\beta^*)) - \phi(\Delta_{itj} + \alpha_{t,1}Z_{itj,1})(\Phi(X_{i,t-1,j}\beta^*))$$
$$\frac{\partial F}{\partial \alpha_{t,1}} = -\phi(\Delta_{itj} + \alpha_{t,1}Z_{itj,1})\Phi(X_{i,t-1,j}\beta^*)Z_{itj,1} \quad (4.19)$$

Note that we suppress the above replacements of t=2 given in (4.19) in the general derivative cases of F with respect to β, Δ_{itj} and $\alpha_{t,1}$ given in (4.17) and take those replacements into account when necessary. Re-writing (4.16) yields

$$\Delta_{itj} = A_{itj}(\beta - \beta_0) + B_{itj}(\alpha_{t,1} - \alpha_{t,10}) \quad (4.20)$$

where

$$A_{itj} = -\frac{\left.\frac{\partial F}{\partial \beta}\right|_{(\beta_0, \alpha_{t10}, \Delta_{itj0})}}{\left.\frac{\partial F}{\partial \Delta_{itj}}\right|_{(\beta_0, \alpha_{t,10}, \Delta_{itj0})}}, \quad B_{itj} = -\frac{\left.\frac{\partial F}{\partial \alpha_{t,1}}\right|_{(\beta_0, \alpha_{t10}, \Delta_{itj0})}}{\left.\frac{\partial F}{\partial \Delta_{itj}}\right|_{(\beta_0, \alpha_{t,10}, \Delta_{itj0})}} \quad (4.21)$$

From (4.16), it can be seen that the Δ_{itj} are explicit and deterministic functions of X_{itj}, X_{it-1j}, β, $\alpha_{t,1}$, and $Z_{itj,1}$, i.e. Δ_{itj}=$\Delta_{itj}(X_{itj}, X_{it-1j}, \beta, \alpha_{t,1}, Z_{itj,1})$. The β_0 and $\alpha_{t,10}$ components of P_0 for PNMTREM are taken to be 0 since the hypothesis tests about the significances of β and $\alpha_{t,1}$ place null hypotheses which assume the equality of those parameters to be 0. This yields

Δ_{itj0} for $t \geq 3$ to be equal to 0 as well (by solving (4.15) with β_0 and $\alpha_{t,10}$ are equal to 0). For $t = 2$, since β^* replaces β as the multiplier of lag-1 covariates (4.19), Δ_{i2j0} could be obtained by Newton-Raphson Algorithm (N-R). Note that this requires $n \times k$ different N-R solutions. We will assess the equality of the components of P_0 to 0 by simulation studies and discuss other conditions in Section 4.4.

Linking Second and Third Levels of PNMTREM(1)

Level 2 (4.6) and level 3 (4.7) of PNMTREM(1) are linked via a convolution equation given by

$$P(Y_{itj} = 1 | y_{i,t-1,j}, X_{itj}) = \int P(Y_{itj} = 1 | y_{i,t-1,j}, X_{itj}, b_{it}) dF(b_{it}) \tag{4.22}$$

which is equivalent to

$$P^t_{itj} = \int P^r_{itj} f(b_{it}) db_{it} \tag{4.23}$$

Re-writing the transition and random effects probabilities in (4.23) yields

$$\Phi(\Delta_{itj} + \alpha_{t,1} Z_{itj,1} y_{i,t-1,j}) = \int \Phi(\Delta^*_{itj} + \lambda_j b_{it}) f(b_{it}) db_{it} \tag{4.24}$$

Following Griswold (2005),

$$\begin{aligned}
\Delta_{itj} + \alpha_{t,1} Z_{itj,1} y_{i,t-1,j} &= \Phi^{-1} \int \Phi(\Delta^*_{itj} + \lambda_j b_{it}) f(b_{it}) db_{it} \\
\Delta_{itj} + \alpha_{t,1} Z_{itj,1} y_{i,t-1,j} &= \Phi^{-1} \Phi \left(\frac{\Delta^*_{itj}}{\sqrt{1 + \lambda_j^2 \sigma_t^2}} \right) \\
\Delta_{itj} + \alpha_{t,1} Z_{itj,1} y_{i,t-1,j} &= \frac{\Delta^*_{itj}}{\sqrt{1 + \lambda_j^2 \sigma_t^2}}
\end{aligned}$$

Then we can obtain Δ^*_{itj} in terms of $\Delta_{itj}, \alpha_{t,1}, Z_{itj,1}, y_{i,t-1,j}, \lambda_j$ and σ_t such that

$$\Delta^*_{itj} = \sqrt{1 + \lambda_j^2 \sigma_t^2} \, (\Delta_{itj} + \alpha_{t,1} Z_{itj,1} y_{i,t-1,j}) \tag{4.25}$$

From (4.25) it can be seen that the Δ^*_{itj} are explicit and deterministic functions of Δ_{itj} (hence Δ^*_{itj} are functions of X_{itj} and β), $\alpha_{t,1}$ $Z_{itj,1}$, y_{it-1j}, λ_j and b_{it},
i.e. $\Delta^*_{itj} = \Delta^*_{itj}(X_{itj}, \beta, \Delta_{itj}, \alpha_{t,1}, Z_{itj,1}, y_{it-1j}, \lambda_j, b_{it})$. The detailed proof could be found in Appendix B.

Putting Δ_{itj} (4.16) in (4.25) yields

$$\Delta^*_{itj} = \sqrt{1+\lambda_j^2\sigma_t^2}\ (A_{itj}(\beta-\beta_0) + B_{itj}(\alpha_{t,1} - \alpha_{t,10}) + Z_{itj,1}\alpha_{t,1}y_{i,t-1,j}) \tag{4.26}$$

Then, the Level 1 (4.5) and level 2 (4.6) of PNMTREM(1) could be explicitly indicated by the level 3 (4.7) as given below.

$$P^r_{itj} \equiv P(Y_{itj} = 1|y_{i,t-1,j}, X_{itj}, b_{it}) = \Phi\left(\sqrt{1+\lambda_j^2\sigma_t^2}(A_{itj}(\beta-\beta_0) + B_{itj}(\alpha_{t,1} - \alpha_{t,10}) + Z_{itj,1}\alpha_{t,1}y_{i,t-1,j}) + \lambda_j b_{it}\right) \tag{4.27}$$

4.2.2.2 Linking Levels of PNMTREM(1) for Baseline Model

Recall that we assume a different model for baseline ($t = 1$) due to reasons such as lack of history data, expectation of more variability and different covariate effects compared to later time points, i.e. $t \geq 2$. This model is still a marginalized multivariate model and related framework are given in (4.8, 4.9). As for the model of $t \geq 2$, to be a valid probabilistic model, the levels of the baseline model are linked to each other via a constraint. The related convolution equation is given below.

$$P(Y_{i1j} = 1|X_{i1j}) = \int P(Y_{i1j} = 1|, X_{i1j}, b_{i1})dF(b_{i1}) \tag{4.28}$$

which is equivalent to

$$P^m_{i1j} = \int P^r_{i1j} f(b_{i1})db_{i1} \tag{4.29}$$

Re-writing the transition and random effects probabilities in (4.29) yields

$$\Phi(X_{i1j}\beta^*) = \int \Phi(\Delta^*_{i1j} + \lambda^*_j b_{i1}) f(b_{i1}) db_{i1} \tag{4.30}$$

Again following Griswold (2005),

$$X_{i1j}\beta^* = \Phi^{-1} \int \Phi(\Delta^*_{i1j} + \lambda^*_j b_{i1}) f(b_{i1}) db_{i1}$$

$$X_{i1j}\beta^* = \Phi^{-1}\Phi\left(\frac{\Delta^*_{i1j}}{\sqrt{1 + \lambda^{*2}_j \sigma^2_1}} \right)$$

Thus, Δ^*_{i1j} could be written in terms of X_{i1j}, β^*, λ^*_j and b_{i1}, explicitly such that

$$\Delta^*_{i1j} = \sqrt{1 + \lambda^{*2}_j \sigma^2_1}\, X_{i1j}\beta^* \tag{4.31}$$

The detailed proof could be found in Appendix B.

Then, level 1 of the baseline model (4.8) could be explicitly indicated by the level 2 of the model (4.9) which is given by

$$P^r_{i1j} \equiv P(Y_{i1j} = 1 | X_{i1j}, b_{i1}) = \Phi\left(\sqrt{1 + \lambda^{*2}_j \sigma^2_1}\, X_{i1j}\beta^* + \lambda^*_j b_{i1} \right) \tag{4.32}$$

Use of *probit* link in each levels of MTREM allows us explicitly writing these levels in terms of each other. On the other hand, the use of *logit* link does not permit such explicit equations between the levels of the model. These ways of explicitly linking the levels of PNMTREM(1) (for both baseline and $t \geq 2$ models) allows us maximizing the likelihood and obtaining the maximum likelihood estimates (MLE) of the parameters without taking the derivatives of Δ_{itj}, Δ^*_{itj} and Δ^*_{i1j} with respect to model parameters. This eases the related MLE derivations and decreases computational time.

4.3 Estimation

We assume two different models for a given multivariate longitudinal binary dataset in the PNMTREM framework. These models are baseline ($t = 1$) and $t \geq 2$ models. The related likelihood function of PNMTREM(1) is the product of two likelihood functions belonging to these models. The likelihood function (also known as marginal likelihood) of PNMTREM(1) is given by

$$L(\theta|y) = \left[\prod_{i=1}^{N} \int \prod_{j=1}^{k} \left(P^r_{i1j}\right)^{y_{i1j}} \left(1 - P^r_{i1j}\right)^{1-y_{i1j}} \phi(b_{i1}) db_{i1}\right] \left[\prod_{i=1}^{N} \prod_{t=2}^{T} \int \prod_{j=1}^{k} \left(P^r_{itj}\right)^{y_{itj}} \left(1 - P^r_{itj}\right)^{1-y_{itj}} \phi(b_{it}) db_{it}\right] \tag{4.33}$$

where $\theta = (\beta^*, \lambda^*_j, \sigma^2_1, \beta, \alpha_{t1}, \lambda_j, \sigma^2_t)$; $P^r_{i1j} = P(Y_{i1j}|X_{i1j}, b_{i1})$ and $P^r_{itj} = P(Y_{itj}|X_{itj}, b_{it})$ are the random effects probabilities for baseline and $t \geq 2$ models, respectively and $\phi(b_{i1})$ and $\phi(b_{it})$ are the random effects distributions which are assumed to be univariate normal densities with mean 0 and standard deviations σ_1 and σ_t, respectively. Note that the construction of the likelihood relies on the common assumptions that the longitudinal responses are independent given the past responses and multivariate responses are independent given the random effects. By re-writing the random effects b_{i1} and b_{it} as $b_{i1} = \sigma_1 z_i$ and $b_{it} = \sigma_t z_i$ in (4.33), and by some simple transformations, we get the likelihood function as

$$L(\theta|y) = \left[\prod_{i=1}^{N} \int \prod_{j=1}^{k} \left(P^r_{i1j}\right)^{y_{i1j}} \left(1 - P^r_{i1j}\right)^{1-y_{i1j}} \phi(z_i) dz_i\right] \left[\prod_{i=1}^{N} \prod_{t=2}^{T} \int \prod_{j=1}^{k} \left(P^r_{itj}\right)^{y_{itj}} \left(1 - P^r_{itj}\right)^{1-y_{itj}} \phi(z_i) dz_i\right] \tag{4.34}$$

where $\phi(z_i)$ is the standard normal density function. For simplicity, the likelihood function $L(\theta|y)$ can be written as

$$L(\theta|y) = L_1(\theta_1|y_1) L_2(\theta_2|y_2) \tag{4.35}$$

where

$$L_1(\theta_1|y_1) = \prod_{i=1}^{N} \int \prod_{j=1}^{k} \left(P^r_{i1j}\right)^{y_{i1j}} \left(1 - P^r_{i1j}\right)^{1-y_{i1j}} \phi(z_i) dz_i \tag{4.36}$$

$$L_2(\theta_2|y_2) = \prod_{i=1}^{N} \prod_{t=2}^{T} \int \prod_{j=1}^{k} \left(P^r_{itj}\right)^{y_{itj}} \left(1 - P^r_{itj}\right)^{1-y_{itj}} \phi(z_i) dz_i \tag{4.37}$$

Here ($\theta = (\theta_1, \theta_2)$), and specifically $\theta_1 = (\beta^*, \lambda_j^*, \sigma_1^2)$ and $\theta_2 = (\beta, \alpha_{t1}, \lambda_j, \sigma_t^2)$ are parameter vectors for baseline and $t \geq 2$ models, respectively; y_1 and y_2 are the observed response matrices at baseline and $t \geq 2$ time points, respectively. Although these two likelihoods seem to be independent, they are connected to each other via the use of the estimates of β^*, $\hat{\beta}^*$, for $t = 2$ due to the marginal constrained equation, see (4.14). We can consider the modeling of $log(\sigma_t)$ for $t = 1, \ldots, T$ instead of directly modeling σ_t or σ_t^2 due to computational aspects such as convergence problems, since taking logarithm of the variance components extends the related parameter space from the interval of $[0, +\infty)$ to the one of $(-\infty, +\infty)$. Turning back to the estimates of σ_t or σ_t^2 is possible by the invariance property of maximum likelihood estimates (MLE) and the use of delta method would help obtaining the related variance estimates.

The likelihood function given in (4.34) needs numerical methods to be maximized while taking integrals over the random effects distribution. The most popular method to numerically solve an integral is the Gauss-Hermite Quadrature (GHQ) method which evaluate the integral as a weighted sum of the integrand function. GHQ requires that the integral has the form of multiplication of a function $f(u)$ and $exp(-u^2)$ and the related approximation has the following form

$$\int_{-\infty}^{+\infty} f(u)exp(-u^2)du \approx \sum_{q=1}^{Q} w_q f(z_q) \tag{4.38}$$

where w_q are the weights and z_q are the quadrature points (also known as zeros) which are tabulated values. These weights and quadrature points can be found in for instance Abramowitz and Stegun (1972) up to $Q = 20$ or can be calculated in standard software by using the formula given in (14.31) on pp. 328 of McCulloch et al. (2008). It is well known that for one-dimensional integrals, i.e. constant random effects over time or independent random effects over time, GHQ is a successful method in terms of approximating an integral. McCulloch et al. (2008, pp. 329) stated that for likelihood functions, 20-point GHQ are enough to achieve an accurate approximation. With normally distributed random effects, the related approximation has the form

$$\int_{-\infty}^{+\infty} f(z)\frac{exp(-z^2/2)}{\sqrt{2\pi}}dz \approx \sum_{q=1}^{Q} w_q \frac{f(\sqrt{2}z_q)}{\sqrt{\pi}} \tag{4.39}$$

which can be obtained by writing z as a function of another random variable v such as $z = \sqrt{2}v$ and some simple calculus. The denominator, $\sqrt{\pi}$, can be neglected in likelihood calculations, since it loses its effect while equating the first partial derivatives of the log-likelihood function to 0. Therefore, neglecting it reduces the extra computational burden.

One restriction on the use of GHQ is that it requires that the integral has a certain form, i.e. random effects follow a normal distribution. Exceptions can be found in Abramowitz and Stegun (1972), Griswold (2005), McCulloch et al. (2008) and Lange (2010). Another restriction is that a linear increase in the dimension of the random effects (correlated random effects) cause exponential increase in the number of required quadrature points. Agresti (2002) stated that approximations with Gauss-Hermite quadratures for integrals with dimension more than 5 are not feasible.

An adaptive GHQ (Liu and Pierce, 1994) method, which includes centering the quadrature points at the mode of the integrated function and scaling them with respect to the estimated curvature at the mode, might increase the efficiency of the approximation and decrease the number of quadrature points, Q (Agresti, 2002, pp. 522). However, their implementation might be very demanding, especially for the complex functions. A comparative illustration of adaptive GHQ method which was motivated by a real life data application can be found in Lesaffre and Spiessens (2001). Crouch and Spiegelman (1990) proposed a method to evaluate the integrals with the form given in (4.38) when the GHQ method fails. The method includes modifying the function, $f(u)exp(-u^2)$ by multiplying a certain one such that $1 - exp[-2\phi i(t - t_0)/h]$ to create a rectangular contour and then integrating of that modified function. However, this method requires more computational time and could be alternative to GHQ for some exceptional cases. Monte Carlo (MC) methods are the usual counterparts of the GHQ to evaluate an integral approximately which relies on drawing random samples from the related density of the integrand function. MC methods are more appropriate than GHQ for approximating high dimensional integrals. Two major disadvantages of MC methods are that they do not guarantee the optimal representation of the true density function due to independence of the psuedo random samples, i.e. sampling error, and cause extra computational burden by the use of large or huge sized random samples due to the law of the large numbers. Quasi Monte Carlo (QMC; Caflish, 1998; González et al., 2006; Pan and Thompson, 2007; Lee et al, 2009) methods alternate the MC methods which guarantee the optimal representation of the psuedo distribution, i.e. reduce the discrepancy between the psuedo and sample distributions, hence also called low-discrepancy numbers. Nonetheless, they still yield computational intensity such as the use of samples with sizes ranging from 10,000 to 100,000 (Pan and Thompson, 2007). For uniform and normal low discrepancy numbers in R, especially Sobol and Halton sequences, see the R package **fOptions** (Wuertz et al., 2011). Some modified MC methods could be found in Chapter 14 of McCulloch et al. (2008) and in Chapter 12 of Agresti (2002). These versions include Monte Carlo Expectation-Maximization Algorithm (MCEM) which treats the random effects as missing data by the fact that both are unobserved and Monte Carlo Newton Raphson Methods (MC-NR) which combines MC and NR methodologies and various alternative methods to handle the integrals such as by using a second-order Taylor Series expansion to solve the integrals (Penalized quasi-Likelihood Estimation; Breslow and Clayton, 1993) and simulating the whole likelihood. Bayesian approaches such as Markov Chain Monte Carlo (MCMC) methods do not require the evaluation of integrals, but they are also computationally intensive. Please note that Ilk and Daniels (2007) applied MCMC methods to MTREM.

Several authors considered Gauss-Hermite Quadratures (GHQ) to evaluate one dimensional integrals having the form given in (4.38). Whereas the related works for generalized linear mixed models can be exemplified by Hedeker and Gibbons (1994), Gibbons and Hedeker (1997), they can be exemplified for multilevel models by Heagerty (1999), Heagerty and Zeger (2000), Wang and Louis (2004), Schildcrout and Heagerty (2007), Ilk and Daniels (2007). Since in PNMTREM we assume that the random effects belonging to different time points follow independent univariate normal distributions, we prefer to use 20-points GHQ to solve the integrals given in (4.34).

First partial derivatives of the log-likelihood functions of PNMTREM do not permit obtaining explicit solutions to the MLE of the parameters. Therefore, optimization techniques are needed. Newton-Raphson Algorithm (N-R) is one of the mostly used optimization methods for finding roots of the likelihood functions and it requires the calculation of first and second partial derivatives of the log-likelihood functions. However, for PNMTREM(1) even the first partial derivatives of the log-likelihoods have very complex forms (will be shown later), hence the use of N-R is not appropriate. Luckily, Fisher-Scoring Algorithm (F-S) does not require the calculation of the second partial derivatives and it solves the log-likelihood functions by using only the first partial derivatives (Kennedy and Gentle, 1980; Demidenko, 2004; Hedeker and Gibbons, 2006). Instead of the Hessian matrix (the matrix of the second partial derivatives; also known as the observed information matrix) in N-R, F-S uses the expected information matrix which only requires the first partial derivatives of the individual contributions to the log-likelihood function. The explicit calculation of the expected information matrix will be introduced later. Another great feature of F-S is that the inverse of the expected information matrix at convergence is a consistent estimator of the large sample variance-covariance matrix of the model parameters. If there is lack of convergence in N-R algorithm especially for functions which have complex forms, reducing the steps by dividing them to the powers of 2, i.e. 2^a where a is a non-negative constant is recommended (Damped Newton-Raphson Algorithm; Conte and de Boor, 1980 and Bose, 2008). Here, $a = 0$ corresponds to no reduce in steps. The detailed functional form of the application of F-S to PNMTREM(1) and damping the related steps will be presented later.

4.3.1 Maximum Likelihood Estimation of the Baseline Parameters (θ_1)

Recall that

$$P^r_{i1j} = P(Y_{i1j}|X_{i1j}, b_{i1}) = \Phi(\Delta^*_{i1j} + \lambda^*_j b_{i1}) \tag{4.40}$$

where $\Delta^*_{i1j} = \sqrt{1 + \lambda^{*2}_j \sigma^2_1}\,(X_{i1j}\beta^*)$ and $b_{i1} \sim N(0, \sigma^2_1)$, $b_{i1} = \sigma_1 z_i$.

If we equate $c_1 = log(\sigma_1)$, then $\sigma_1 = e^{c_1}$ and $\sigma^2_1 = e^{2c_1}$. Re-writing Δ^*_{itj} and c_1 in (4.40) yields

$$P(Y_{i1j}|X_{i1j}, b_{i1}) = \Phi(\sqrt{1 + \lambda^{*2}_j e^{2c_1}}\,(X_{i1j}\beta^*) + \lambda^*_j e^{c_1} z_i) \tag{4.41}$$

For simplicity, throughout let's consider $\sqrt{1 + \lambda^{*2}_j e^{2c_1}}\,(X_{i1j}\beta^*) + \lambda^*_j e^{c_1} z_i = d_{i1j}$. Then, the likelihood function for baseline model (4.36) can be written as

$$L_1(\theta_1|y_1) = \prod_{i=1}^{N} \int \prod_{j=1}^{k} \left(\Phi(d_{i1j})\right)^{y_{i1j}} \left(1 - \Phi(d_{i1j})\right)^{1-y_{i1j}} \phi(z_i) dz_i \tag{4.42}$$

where $\theta_1 = (\beta^*, \lambda^*_j, c_1)$.

Let's denote the individual contribution of subject i to the baseline likelihood, $L_1(\theta_1|y_1)$, by $h(Y_{i1}|\theta_1)$ which is given by

$$h(Y_{i1}|\theta_1) = \int \prod_{j=1}^{k} \left(\Phi(d_{i1j})\right)^{y_{i1j}} \left(1 - \Phi(d_{i1j})\right)^{1-y_{i1j}} \phi(z_i) dz_i \tag{4.43}$$

Moreover, let's denote the joint distribution of the multivariate binary responses at baseline given the random effects by $\ell(Y_{i1}|\theta_1)$ which is given by

$$\ell(Y_{i1}|\theta_1) = \prod_{j=1}^{k} \Big(\Phi(d_{i1j})\Big)^{y_{i1j}} \ \Big(1 - \Phi(d_{i1j})\Big)^{1-y_{i1j}} \tag{4.44}$$

Here, $\ell(Y_{i1}|\theta_1)$ can be re-written as

$$\ell(Y_{i1}|\theta_1) = exp\left[\sum_{j=1}^{k} \Big(Y_{i1j}\log\Big(\Phi(d_{i1jq})\Big) + (1 - Y_{i1j})\log\Big(1 - \Phi(d_{i1jq})\Big)\Big)\right] \tag{4.45}$$

Then, the individual contribution of subject i is equivalent to

$$\begin{aligned} h(Y_{i1}|\theta_1) &= \int \ell(Y_{i1}|\theta_1)\phi(z_i)dz_i \\ &= \int exp\left[\sum_{j=1}^{k} \Big(Y_{i1j}\log\Big(\Phi(d_{i1j})\Big) + (1 - Y_{i1j})\log\Big(1 - \Phi(d_{i1j})\Big)\Big)\right]\phi(z_i)dz_i \end{aligned} \tag{4.46}$$

The above integral is approximated by 20-points Gauss-Hermite quadratures with the quadrature points and weights which are denoted by z_q and w_q, respectively, and $q = 1, \ldots, 20$. These are available in Abramowitz and Stegun (1972). The approximation of the integral yields

$$h(Y_{i1}|\theta_1) \approx \sum_{q=1}^{20} w_q \ exp\left[\sum_{j=1}^{k} \Big(Y_{i1j}\log\Big(\Phi(d_{i1jq})\Big) + (1 - Y_{i1j})\log\Big(1 - \Phi(d_{i1jq})\Big)\Big)\right] \tag{4.47}$$

where d_{i1jq} denotes the expression d_{i1j} in which the z_i's are replaced by the quadrature points, z_q, and the open form of it is given by

$$d_{i1jq} = \sqrt{1 + \lambda_j^{*2} e^{2c_1}} \ (X_{i1j}\beta^*) + \lambda_j^* e^{c_1} \sqrt{2} \ z_q \tag{4.48}$$

The baseline likelihood function , $L_1(\theta_1|y_1)$ given in (4.42) takes the form

$$L_1(\theta_1|y_1) = \prod_{i=1}^{N} h(Y_{i1}|\theta_1) \tag{4.49}$$

Taking natural logarithm of it yields

$$\log\left(L_1(\theta_1|y_1)\right) = \sum_{i=1}^{N} \log h(Y_{i1}|\theta_1) \tag{4.50}$$

The functional form of the first partial derivatives of the log-baseline likelihood with respect to the baseline parameters is given by

$$\frac{\partial \log\left(L_1(\theta_1|y_1)\right)}{\partial\theta_1} = \sum_{i=1}^{N} \frac{1}{h(Y_{i1}|\theta_1)} \frac{\partial h(Y_{i1}|\theta_1)}{\partial\theta_1} \tag{4.51}$$

Here, $\frac{\partial h(Y_{i1}|\theta_1)}{\partial\theta_1}$ is given by

$$\begin{aligned}
\frac{\partial h(Y_{i1}|\theta_1)}{\partial\theta_1} &= \frac{\partial}{\partial\theta_1}\left(\int \underbrace{\prod_{j=1}^{k} \left(\Phi(d_{i1j})\right)^{y_{i1j}} \left(1-\Phi(d_{i1j})\right)^{1-y_{i1j}}}_{\ell(Y_{i1}|\theta_1)} \phi(z_i)dz_i \right) \\
&= \frac{\partial}{\partial\theta_1}\left(\int exp\left[\sum_{j=1}^{k}\left(Y_{i1j}log\left(\Phi(d_{i1j})\right) + (1-Y_{i1j})log\left(1-\Phi(d_{i1j})\right)\right)\right]\phi(z_i)dz_i \right) \\
&= \int \ell(Y_{i1}|\theta_1)\left[\sum_{j=1}^{k}\left(Y_{i1j}\frac{1}{\Phi(d_{i1j})}\phi(d_{i1j})\frac{\partial d_{i1j}}{\partial\theta_1} + (1-Y_{i1j})\frac{1}{1-\Phi(d_{i1j})}(-\phi(d_{i1j}))\frac{\partial d_{i1j}}{\partial\theta_1}\right)\right]\phi(z_i)dz_i \\
&= \int \ell(Y_{i1}|\theta_1)\left\{\sum_{j=1}^{k}\left[\frac{\partial d_{i1j}}{\partial\theta_1}\phi(d_{i1j})\left(\frac{Y_{i1j}}{\Phi(d_{i1j})} - \frac{1-Y_{i1j}}{1-\Phi(d_{i1j})}\right)\right]\right\}\phi(z_i)dz_i \\
&= \int \ell(Y_{i1}|\theta_1)\left\{\sum_{j=1}^{k}\left[\frac{\partial d_{i1j}}{\partial\theta_1}\phi(d_{i1j})\left(\frac{Y_{i1j}-\Phi(d_{i1j})}{\left(\Phi(d_{i1j})\right)\left(1-\Phi(d_{i1j})\right)}\right)\right]\right\}\phi(z_i)dz_i
\end{aligned} \tag{4.52}$$

The above integral is also approximated by 20-points Gauss-Hermite quadrature points and weights which are denoted by z_q and w_q, respectively. The approximation of the integral yields

$$\frac{\partial h(Y_{i1}|\theta_1)}{\partial\theta_1} \approx \sum_{q=1}^{20} w_q \left\{\ell(Y_{i1}|\theta_1)\left\{\sum_{j=1}^{k}\left[\frac{\partial d_{i1jq}}{\partial\theta_1}\phi(d_{i1jq})\left(\frac{Y_{i1j}-\Phi(d_{i1jq})}{\left(\Phi(d_{i1jq})\right)\left(1-\Phi(d_{i1jq})\right)}\right)\right]\right\}\right\} \tag{4.53}$$

where d_{i1jq} denotes the expression d_{i1j} in which the z_i's are replaced by the quadrature points, z_q, and the open form of it is given in (4.48).

As a result, the first partial derivatives of the log-likelihood function for baseline model (4.51) can be approximated by

$$\frac{\partial log\left(L_1(\theta_1|y_1)\right)}{\partial\theta_1} \approx \sum_{i=1}^{N}\frac{1}{h(Y_{i1}|\theta_1)}\sum_{q=1}^{20} w_q \left\{\ell(Y_{i1}|\theta_1)\left\{\sum_{j=1}^{k}\left[\frac{\partial d_{i1jq}}{\partial\theta_1}\phi(d_{i1jq})\left(\frac{Y_{i1j}-\Phi(d_{i1jq})}{\left(\Phi(d_{i1jq})\right)\left(1-\Phi(d_{i1jq})\right)}\right)\right]\right\}\right\} \tag{4.54}$$

The derivatives of d_{i1jq} with respect to $\theta_1 = (\beta^*, \lambda_j^*, c_1)$ are given below

$$\begin{aligned}
\frac{\partial d_{i1jq}}{\partial \beta^*} &= \sqrt{1 + \lambda_j^{*2} e^{2c_1}}(X_{i1j}) \\
\frac{\partial d_{i1jq}}{\partial \lambda_j^*} &= (1 + \lambda_j^{*2} e^{2c_1})^{-1/2} \lambda_j^* e^{2c_1} (X_{i1j}\beta^*) + e^{c_1} \sqrt{2} z_q \\
\frac{\partial d_{i1jq}}{\partial c_1} &= (1 + \lambda_j^{*2} e^{2c_1})^{-1/2} \lambda_j^{*2} e^{2c_1} (X_{i1j}\beta^*) + \lambda_j e^{c_1} \sqrt{2} z_q
\end{aligned} \tag{4.55}$$

As stated earlier, the MLE of baseline parameters are obtained iteratively by Fisher-Scoring Algorithm (F-S) and the related algorithm is given by

$$\theta_1^{(m+1)} = \theta_1^m + I(\theta_1^m)^{-1} \frac{\partial log\left(L_1(\theta_1^m | y_1)\right)}{\partial \theta_1^m} \tag{4.56}$$

where m represents the F-S steps and $I(\theta_1)$ is an empirical and consistent estimator of the baseline information matrix and can be calculated by

$$I(\theta_1) = \sum_{i=1}^{N} h(Y_{i1}|\theta_1)^{-2} \left(\frac{\partial h(Y_{i1}|\theta_1)}{\partial \theta_1}\right)\left(\frac{\partial h(Y_{i1}|\theta_1)}{\partial \theta_1}\right)^T \tag{4.57}$$

The large-sample variance-covariance matrix of the baseline parameters, θ_1, can be obtained as the inverse of $I(\hat{\theta}_1)$ at convergence.

Since λ_j^* are response specific parameters, the calculation of λ_j^* component of both $\frac{\partial log(L_1(\theta_1|y_1))}{\partial \theta_1}$ and $I(\theta_1)$ is different from those for β^* and c_1. All the details are provided in Appendix B.

4.3.2 Maximum Likelihood Estimation of the $t \geq 2$ Parameters (θ_2)

Recall that for $t \geq 2$ model, the last level of PNMTREM(1) is

$$P_{itj}^r = P(Y_{itj} = 1 | X_{itj}, y_{itj}, b_{it}) = \Phi(\Delta_{itj}^* + \lambda_j b_{it}) \tag{4.58}$$

where $\Delta_{itj}^* = \sqrt{1 + \lambda_j^2 \sigma_t^2} \left(A_{itj}(\beta - \beta_0) + B_{itj}(\alpha_{t,1} - \alpha_{t,10}) + Z_{itj,1}\alpha_{t,1} y_{i,t-1,j}\right)$.

Moreover, recall that

$$A_{itj} = -\frac{\left.\frac{\partial F}{\partial \beta}\right|_{(\beta_0, \alpha_{t10}, \Delta_{itj0})}}{\left.\frac{\partial F}{\partial \Delta_{itj}}\right|_{(\beta_0, \alpha_{t,10}, \Delta_{itj0})}}, \quad B_{itj} = -\frac{\left.\frac{\partial F}{\partial \alpha_{t,1}}\right|_{(\beta_0, \alpha_{t10}, \Delta_{itj0})}}{\left.\frac{\partial F}{\partial \Delta_{itj}}\right|_{(\beta_0, \alpha_{t,10}, \Delta_{itj0})}} \tag{4.59}$$

and

$$F(\Delta_{itj}, \beta, \alpha_{t,1}, X_{itj}, X_{it-1j}, Z_{itj}) = \Phi(X_{itj}\beta) - \Phi(\Delta_{itj})(1 - \Phi(X_{i,t-1,j}\beta)) - \Phi(\Delta_{itj} + \alpha_{t,1} Z_{itj,1})\Phi(X_{i,t-1,j}\beta) = 0 \tag{4.60}$$

In (4.58), $b_{it} \sim N(0, \sigma_t^2)$ and $b_{it} = \sigma_t z_i$. If we equate $c_t = log(\sigma_t)$, then $\sigma_t = e^{c_t}$ and $\sigma_t^2 = e^{2c_t}$. Re-writing Δ^*_{itj} and c_t in (4.58) yields

$$P(Y_{itj} = 1|X_{itj}, y_{itj}, b_{it}) = \Phi\Big(\sqrt{1 + \lambda_j^2 e^{2c_t}} \, \Big[A_{itj}(\beta - \beta_0) + B_{itj}(\alpha_{t,1} - \alpha_{t,10}) + Z_{itj,1}\alpha_{t,1}y_{i,t-1,j}\Big] + \lambda_j e^{c_t} z_i\Big) \tag{4.61}$$

which is equivalent to

$$P(Y_{itj} = 1|X_{itj}, y_{itj}, b_{it}) = \Phi(\sqrt{1 + \lambda_j^2 e^{2c_t}} \, \Big[-(A_{itj}\beta_0 + B_{itj}\alpha_{t,10}) + A_{itj}\beta + \alpha_{t,1}(B_{itj} + Z_{itj}y_{it-1j})\Big] + \lambda_j e^{c_t} z_i) \tag{4.62}$$

For simplicity, throughout let's consider $\sqrt{1 + \lambda_j^2 e^{2c_t}} \, \Big[-(A_{itj}\beta_0 + B_{itj}\alpha_{t,10}) + A_{itj}\beta + \alpha_{t,1}(B_{itj} + Z_{itj}y_{it-1j})\Big] + \lambda_j e^{c_t} z_i = d_{itj}$. Then, the likelihood function for $t \geq 2$ model (4.37) can be written as

$$L_2(\theta_2|y_2) = \prod_{i=1}^{N}\prod_{t=2}^{T}\int\prod_{j=1}^{k}\Big(\Phi(d_{itj})\Big)^{y_{itj}}\Big(1 - \Phi(d_{itj})\Big)^{1-y_{itj}}\phi(z_i)dz_i \tag{4.63}$$

where $\theta_2 = (\beta, \alpha_{t,1}, \lambda_j, c_t)$.

Let's denote the individual contribution of subject i at time t to the $t \geq 2$ likelihood, $L_2(\theta_2|y_2)$, by $h(Y_{it}|\theta_2)$ which is given by

$$h(Y_{it}|\theta_2) = \int\prod_{j=1}^{k}\Big(\Phi(d_{itj})\Big)^{y_{itj}}\Big(1 - \Phi(d_{itj})\Big)^{1-y_{itj}}\phi(z_i)dz_i \tag{4.64}$$

Furthermore, let's denote the joint distribution of the multivariate binary responses for $t \geq 2$ given the random effects by $\ell(Y_{it}|\theta_2)$ which is given by

$$\ell(Y_{it}|\theta_2) = \prod_{j=1}^{k}\Big(\Phi(d_{itj})\Big)^{y_{itj}}\Big(1 - \Phi(d_{itj})\Big)^{1-y_{itj}} \tag{4.65}$$

Here, $\ell(Y_{it}|\theta_2)$ can be re-written as

$$\ell(Y_{it}|\theta_2) = exp\left[\sum_{j=1}^{k}\Big(Y_{itj}log\Big(\Phi(d_{itjq})\Big) + (1 - Y_{itj})log\Big(1 - \Phi(d_{itjq})\Big)\Big)\right] \tag{4.66}$$

Then, the individual contribution of subject i is equivalent to

$$\begin{aligned} h(Y_{it}|\theta_2) &= \int \ell(Y_{it}|\theta_2)\phi(z_i)dz_i \\ &= \int exp\left[\sum_{j=1}^{k}\Big(Y_{itj}log\Big(\Phi(d_{itj})\Big) + (1 - Y_{itj})log\Big(1 - \Phi(d_{itj})\Big)\Big)\right]\phi(z_i)dz_i \end{aligned} \tag{4.67}$$

The above integral is approximated by 20-points Gauss-Hermite quadratures with the quadrature points and weights which are denoted by z_q and w_q, respectively, and $q = 1, \ldots, 20$. These are available in Abramowitz and Stegun (1972). The approximation yields

$$h(Y_{it}|\theta_2) \approx \sum_{q=1}^{20} w_q \; exp\left[\sum_{j=1}^{k} \left(Y_{itj} log\left(\Phi(d_{itjq})\right) + (1 - Y_{itj}) log\left(1 - \Phi(d_{itjq})\right)\right)\right] \tag{4.68}$$

where d_{itjq} denotes the expression d_{itj} in which the z_i's are replaced by the quadrature points, z_q, and the open form of it is given by

$$d_{itjq} = \sqrt{1 + \lambda_j^2 e^{2c_t}} [-(A_{itj}\beta_0 + B_{itj}\alpha_{t,10}) + A_{itj}\beta + \alpha_{t,1}(B_{itj} + Z_{itj}y_{it-1j})] + \lambda_j e^{c_t} \sqrt{2} z_q \tag{4.69}$$

The likelihood function for $t \geq 2$, $L_2(\theta_2|y_2)$ (4.35), takes the form

$$L_2(\theta_2|y_2) = \prod_{i=1}^{N} \prod_{t=2}^{T} h(Y_{it}|\theta_2) \tag{4.70}$$

Taking natural logarithm of it yields

$$\log\left(L_2(\theta_2|y_2)\right) = \sum_{i=1}^{N} \sum_{t=2}^{T} \log h(Y_{it}|\theta_2)$$

The functional form of the first partial derivatives of the log-likelihood for $t \geq 2$ with respect to the parameters, θ_2, is given by

$$\frac{\partial \log\left(L_2(\theta_2|y_2)\right)}{\partial \theta_2} = \sum_{i=1}^{N} \sum_{t=2}^{T} \frac{1}{h(Y_{it}|\theta_2)} \frac{\partial h(Y_{it}|\theta_2)}{\partial \theta_2} \tag{4.71}$$

Here, $\frac{\partial h(Y_{it}|\theta_2)}{\partial \theta_2}$ is given by

$$\begin{aligned}
\frac{\partial h(Y_{it}|\theta_2)}{\partial \theta_2} &= \frac{\partial}{\partial \theta_2}\left(\int \underbrace{\prod_{j=1}^{k} \left(\Phi(d_{itj})\right)^{y_{itj}} \left(1 - \Phi(d_{itj})\right)^{1-y_{itj}}}_{\ell(Y_{it}|\theta_2)} \phi(z_i) dz_i\right) \\
&= \frac{\partial}{\partial \theta_2}\left(\int exp\left[\sum_{j=1}^{k} \left(Y_{itj} log\left(\Phi(d_{itj})\right) + (1 - Y_{itj}) log\left(1 - \Phi(d_{itj})\right)\right)\right] \phi(z_i) dz_i\right) \\
&= \int \ell(Y_{it}|\theta_2) \left[\sum_{j=1}^{k} \left(Y_{itj} \frac{1}{\Phi(d_{itj})} \phi(d_{itj}) \frac{\partial d_{itj}}{\partial \theta_2} + (1 - Y_{itj}) \frac{1}{1 - \Phi(d_{itj})} (-\phi(d_{itj})) \frac{\partial d_{itj}}{\partial \theta_2}\right)\right] \phi(z_i) dz_i \\
&= \int \ell(Y_{it}|\theta_2) \left\{\sum_{j=1}^{k} \left[\frac{\partial d_{itj}}{\partial \theta_2} \phi(d_{itj}) \left(\frac{Y_{itj}}{\Phi(d_{itj})} - \frac{1 - Y_{itj}}{1 - \Phi(d_{itj})}\right)\right]\right\} \phi(z_i) dz_i \\
&= \int \ell(Y_{it}|\theta_2) \left\{\sum_{j=1}^{k} \left[\frac{\partial d_{itj}}{\partial \theta_2} \phi(d_{itj}) \left(\frac{Y_{itj} - \Phi(d_{itj})}{\left(\Phi(d_{itj})\right)\left(1 - \Phi(d_{itj})\right)}\right)\right]\right\} \phi(z_i) dz_i
\end{aligned} \tag{4.72}$$

The above integral is also approximated by 20-points Gauss-Hermite quadrature points and weights which are denoted by z_q and w_q, respectively. The approximation of the integral yields

$$\frac{\partial h(Y_{it}|\theta_2)}{\partial \theta_2} \approx \sum_{q=1}^{20} w_q \left\{ \ell(Y_{it}|\theta_2) \left\{ \sum_{j=1}^{k} \left[\frac{\partial d_{itjq}}{\partial \theta_2} \phi(d_{itjq}) \left(\frac{Y_{itj} - \Phi(d_{itjq})}{\left(\Phi(d_{itjq})\right)\left(1 - \Phi(d_{itjq})\right)} \right) \right] \right\} \right\} \tag{4.73}$$

where d_{itjq} denotes the expression d_{itj} in which the z_i's are replaced by the quadrature points, z_q, and the open form of it is given in (4.69).

As a result, the first partial derivatives of the log-likelihood function for $t \geq 2$ model (4.71) can be approximated by

$$\frac{\partial log\,(L_2(\theta_2|y_2))}{\partial \theta_2} \approx \sum_{i=1}^{N} \sum_{t=2}^{T} \frac{1}{h(Y_{it}|\theta_2)} \sum_{q=1}^{20} w_q \left\{ \ell(Y_{it}|\theta_2) \left\{ \sum_{j=1}^{k} \left[\frac{\partial d_{itjq}}{\partial \theta_2} \phi(d_{itjq}) \left(\frac{Y_{itj} - \Phi(d_{itjq})}{\left(\Phi(d_{itjq})\right)\left(1 - \Phi(d_{itjq})\right)} \right) \right] \right\} \right\} \tag{4.74}$$

The derivatives of d_{itjq} with respect to $\theta_2 = (\beta, \alpha_{t1}, \lambda_j, c_t)$ are given below.

$$\begin{aligned}
\frac{\partial d_{itjq}}{\partial \beta} &= \sqrt{1 + \lambda_j{}^2 e^{2c_t}}(A_{itj}) \\
\frac{\partial d_{itjq}}{\partial \alpha_{t1}} &= \sqrt{1 + \lambda_j{}^2 e^{2c_t}}(B_{itj} + Z_{itj} y_{it-1j}) \\
\frac{\partial d_{itjq}}{\partial \lambda_j} &= (1 + \lambda_j{}^2 e^{2c_t})^{-1/2} \lambda_j e^{2c_t} \left(-(A_{itj}\beta_0 + B_{itj}\alpha_{t,10}) + A_{itj}\beta + \alpha_{t,1}(B_{itj} + Z_{itj} y_{it-1j}) \right) + e^{c_t} \sqrt{2} z_q \\
\frac{\partial d_{itjq}}{\partial c_t} &= (1 + \lambda_j{}^2 e^{2c_t})^{-1/2} \lambda_j{}^2 e^{2c_t} \left(-(A_{itj}\beta_0 + B_{itj}\alpha_{t,10}) + A_{itj}\beta + \alpha_{t,1}(B_{itj} + Z_{itj} y_{it-1j}) \right) + \lambda_j e^{c_t} \sqrt{2} z_q
\end{aligned} \tag{4.75}$$

As stated before, the MLEs of the parameters are obtained iteratively by Fisher-Scoring Algorithm (F-S) which is given by

$$\theta_2^{(m+1)} = \theta_2^m + I(\theta_2^m)^{-1} \frac{\partial log\left(L_2(\theta_2^m|y_2)\right)}{\partial \theta_2^m} \tag{4.76}$$

where m represents the F-S steps and $I(\theta_2)$ is an empirical and consistent estimator of the information matrix and can be calculated by

$$I(\theta_2) = \sum_{i=1}^{N} \left(\sum_{t=2}^{T} \frac{1}{h(Y_{it}|\theta_2)} \frac{\partial h(Y_{it}|\theta_2)}{\partial \theta_2} \right) \left(\sum_{t=2}^{T} \frac{1}{h(Y_{it}|\theta_2)} \frac{\partial h(Y_{it}|\theta_2)}{\partial \theta_2} \right)^T \tag{4.77}$$

The large-sample variance-covariance matrix of the parameters at $t \geq 2$, θ_2, can be obtained as the inverse of $I(\hat{\theta}_2)$ at convergence.

Since α_{t1} and c_t are time specific and λ_j^* are response specific parameters, the calculation of $\frac{\partial log(L_2(\theta_2|y_2))}{\partial \theta_2}$ and $I(\theta_2^m)$ for those components are different than the one for β. All the related details are placed in Appendix B.

4.3.3 Parameter Estimation Algorithm of PNMTREM(1)

The algorithm which PNMTREM(1) follows for parameter estimation can be investigated in two parts which are for baseline and for $t \geq 2$ models, respectively.

The Parameter Estimation Algorithm of Baseline Model

The parameter estimation algorithm for the baseline model can be depicted as below.

1. Select the initials of $\theta_1 = (\beta^*, \lambda_j^*, c_1)$. The initials of β^* can be obtained by independent probit regression models or pooled probit regression models (pooled over the multiple responses) for binary response data which are available in standard software. The initials of λ_j^* could be set to 1, since the estimate of λ_j^* is expected to be around 1. The initials of c_1 could be found by trial and error or by educated guesses.
2. Obtain Δ_{i1j}^* by using (4.31).
3. Calculate $\frac{\partial log L_1(\theta_1|y_1)}{\partial \theta_1}$ (4.52) and $I(\theta_1)$ (4.57) with the current values of Δ_{i1j}^*, λ_j^* and c_1 and by approximating the integrals with Gauss-Hermite quadratures.
4. Improve the parameter estimates by a Fisher-Scoring Algorithm step (4.56).
5. Repeat steps 2-4 until convergence is satisfied, i.e. until $\sqrt{(\hat{\theta}_1^{old} - \hat{\theta}_1^{new})^T(\hat{\theta}_1^{old} - \hat{\theta}_1^{new})} \leqslant$ tolerance. Tolerance could be set to 10^{-3}, 10^{-4} or even a smaller value.

The Parameter Estimation Algorithm of $t \geq 2$ Model

The parameter estimation algorithm for the $t \geq 2$ model can be summarized as below.

1. Select the initials of $\theta_2 = (\beta, \alpha_{t,1}, \lambda_j, c_t)$. The initials of β can be obtained by multivariate marginal models (MMM), i.e. the use of `mmm` or `mmm2` functions avaiable in the R package **mmm** via the probit link options. The initials of $\alpha_{t,1}$ could be obtained by probit transition models, such as regressing the responses at time $t-1$ on the responses at time t, by the use of a standard software. The initials of λ_j could be set to 1. The initials of c_t could be found by trial and error or by educated guesses.
2. Set the values of β_0 and $\alpha_{t,10}$ components of the point $P_0 = (\beta_0, \alpha_{t,10}, \Delta_{itj0})$ to 0. This is

around which the implicit function theorem (IFT) searches solutions. Then, Δ_{itj0} for $t \geq 3$ are 0 as well (4.15). Obtain Δ_{i2j0} by Newton-Raphson algorithm for $t = 2$ since $\hat{\beta}^*$ replaces β_0 as the multiplier of lag-1 covariates (4.19).
3. Obtain Δ_{itj} (4.16) and Δ^*_{itj} (4.25).
4. Calculate $\frac{\partial logL_2(\theta_2|y_2)}{\partial\theta_2}$ (4.72) and $I(\theta_2)$ (4.76) with current values of Δ^*_{itj}, λ_j and c_t and by approximating the integrals with Gauss-Hermite quadratures.
5. Improve the parameter estimates by Fisher-Scoring Algorithm.
6. Repeat steps 3-5 until convergence is satisfied, i.e. until $\sqrt{(\hat{\theta_2}^{old} - \hat{\theta_2}^{new})^T(\hat{\theta_2}^{old} - \hat{\theta_2}^{new})} \leqslant$ tolerance. Again, tolerance could be set to 10^{-3}, 10^{-4} or even a smaller value.

Note that tolerances for baseline and $t \geq 2$ models could be set to different values, since the latter is a more complex model compared to the former.

4.3.4 Empirical Bayesian Estimation of Random Effects Coefficients

To calculate the individual probabilities such as $P(Y_{i1j} = 1|X_{i1j}, b_{i1})$ and $P(Y_{itj} = 1|X_{itj}, y_{it-1j}, b_{it})$ for $t \geq 2$, we need the estimates of Δ^*_{i1j}, λ^*_j, b_{i1} for baseline model and Δ^*_{itj}, λ_j, b_{it} for $t \geq 2$ model. Previously, we mentioned about the maximum likelihood estimation (MLE) of those model parameters including σ_1 and σ_t (remember that $b_{i1} = \sigma_1 z_i$ and $b_{it} = \sigma_t z_i$ where $z_i \sim N(0, 1)$). Note that Δ^*_{i1j} and Δ^*_{itj} are deterministic functions of the other model parameters such as $\beta^*, \lambda^*_j, \sigma_1$ and $\beta, \alpha_{t,1}, \lambda_j, \sigma_t$; hence $\hat{\Delta}^*_{i1j}$ and $\hat{\Delta}^*_{itj}$ could be directly obtained by putting the MLEs of them (see 4.31 and 4.25, respectively). The only unestimated parameters are z_i, which denote the characteristics of each subject. These parameters together with σ_1 and σ_t construct the estimates of the subject specific random effects for baseline and $t \geq 2$ models, respectively.

Given the MLE of $\theta_1 = (\beta^*, \lambda^*_j, c_1 = log(\sigma_1))$ and $\theta_2 = (\beta, \alpha_{t,1}, \lambda_j, c_t = log(\sigma_t))$, we can obtain the Empirical Bayes estimators of b_{it}, $\tilde{b}_{it}$ $(t = 1, \ldots, T)$ by solving the posterior score equations of z_i (Heagerty, 1999). Note that for the empirical Bayes estimations, we make no distinction in time such as $t = 1$ vs. $t \geq 2$, since both of the model parameters are already estimated and we are just dealing with the estimation of z_i which does not differ with respect to time but only differ from subject to subject, i.e. they are subject-specific. The posterior distribution of z_i is proportional to the conditional distribution of the observed data, Y_i given z_i, $[Y_i|z_i]$, times the prior distribution of z_i such that

$$
\begin{aligned}
f(z_i|-) &\propto f(Y_i|z_i)f(z_i) \\
&\propto \left\{ \prod_{t=1}^{T} \prod_{j=1}^{k} \left[\Phi(\Delta_{itj}^{*c} + \lambda_j^c \sigma_t^c z_i) \right]^{Y_{itj}^c} \left[1 - \Phi(\Delta_{itj}^{*c} + \lambda_j^c \sigma_t^c z_i) \right]^{1-Y_{itj}^c} \right\} e^{-z_i^2/2}
\end{aligned}
$$

where $\Delta_{itj}^{*c} = (\Delta_{i1j}^{*}, \Delta_{itj}^{*})$, $\lambda_j^c = (\lambda_j^*, \lambda_j)$, $\sigma_t^c = (\sigma_1, \sigma_t)$ and $Y_{itj}^c = (Y_{i1j}, Y_{itj})$ and c corresponds to complete matrices and vectors.

We can obtain the estimates of z_i, $\tilde{z}_i$, based on the MLEs of θ_1 and θ_2 by equating the first partial derivative of the natural logarithm of the posterior distribution of z_i with respect to z_i to 0 and then by solving the score equations for z_i. The related operations are given below. Note that for the sake of simplicity, we refer $\Delta_{itj}^{*c} + \lambda_j^c \sigma_t^c z_i$ as d_{itj}^c.

$$
\begin{aligned}
log(f(z_i|-)) &= \left\{ \sum_{t=1}^{T} \sum_{j=1}^{k} \left[Y_{itj}^c log\left(\Phi(d_{itj}^c)\right) + (1 - Y_{itj}^c) log\left(1 - \Phi(d_{itj}^c)\right) \right] \right\} - \frac{z_i^2}{2} \\
\frac{\partial log(f(z_i|\_))}{\partial z_i} &= \left\{ \sum_{t=1}^{T} \sum_{j=1}^{k} \left[Y_{itj}^c \frac{1}{\Phi(d_{itj}^c)} \Phi'(d_{itj}^c) + (1 - Y_{itj}^c) \frac{1}{1 - \Phi(d_{itj}^c)} (-\Phi'(d_{itj}^c)) \right] \right\} - z_i
\end{aligned}
$$

Note that,

$$
\begin{aligned}
&\Phi(d_{itj}^c) = \Phi(\Delta_{itj}^{*c} + \lambda_j^c \sigma_t^c z_i) \quad \text{and} \\
&\Phi'(d_{itj}^c) = \Phi'(\Delta_{itj}^{*c} + \lambda_j^c \sigma_t^c z_i) = \phi(\Delta_{itj}^{*c} + \lambda_j^c \sigma_t^c z_i) \lambda_j^c \sigma_t^c = \lambda_j^c \sigma_t^c \phi(d_{itj}^c)
\end{aligned}
$$

$$
\frac{\partial log(f(z_i|\_))}{\partial z_i} = \left\{ \sum_{t=1}^{T} \sum_{j=1}^{k} \left[\frac{Y_{itj}^c \lambda_j^c \sigma_t^c \phi(d_{itj}^c)}{\Phi(d_{itj}^c)} - \frac{(1 - Y_{itj}^c) \lambda_j^c \sigma_t^c \phi(d_{itj}^c)}{1 - \Phi(d_{itj}^c)} \right] \right\} - z_i
$$

Finally, we can obtain $\tilde{z}_i$ by solving

$$
0 = \left\{ \sum_{t=1}^{T} \sum_{j=1}^{k} \frac{\hat{\lambda}_j^c \hat{\sigma}_t^c \phi(\hat{d}_{itj}^c) \left(Y_{itj}^c - \Phi(\hat{d}_{itj}^c)\right)}{\Phi(\hat{d}_{itj}^c)\left(1 - \Phi(\hat{d}_{itj}^c)\right)} \right\} - z_i \tag{4.78}
$$

for z_i where $\hat{d}_{itj}^c = \hat{\Delta}_{itj}^{*c} + \hat{\lambda}_j^c \hat{\sigma}_t^c z_i$ and $\hat{\Delta}_{itj}^{*c}$ are obtained by using the MLEs θ_1 and θ_2. To solve (4.78) with respect to z_i, we need to apply some optimization methods such as Newton-Raphson (N-R) Algorithm, since the equation does not provide close solutions for z_i. To apply N-R algorithm, we need the second derivative of the log-posterior distribution of z_i. The related calculations are placed in Appendix B.

4.4 Simulation Study

We conducted a Monte Carlo simulation study to confirm the mathematical derivations and the written R code for the PNMTREM(1). This simulation study allows us to examine the bias and variance of the model parameters as well. Mean, bias and Mean Squared Error (MSE) are reported for these purposes.

4.4.1 Data Generation

In each replications of the simulation study, we simulated datasets under PNMTREM(1) which include bivariate binary responses and two associated covariates for 500 subjects with 4 follow-ups. We considered different set of covariates for baseline and $t \geq 2$ time points. Moreover, we considered varying effects of the common covariates for these time points, i.e. $\beta^* \neq \beta$.

For $t = 1$, we considered a model with the framework given by

$$P(Y_{i1j} = 1|X_{i1j}) = \Phi(\beta_0^* + \beta_1^* X_{i1j1}) \tag{4.79}$$

$$P(Y_{i1j} = 1|X_{i1j}, b_{i1}) = \Phi(\Delta_{i1j}^* + \lambda_j^* b_{i1}) \tag{4.80}$$

where $\beta^* = (\beta_0^*, \beta_1^*) = (-1, 1.9)$, $\lambda_j^* = (\lambda_1^*, \lambda_2^*) = (1, 1.07)$ and $b_{i1} \sim N(0, \sigma_1^2)$, $\sigma_1 = 0.7$. X_1 is generated from $Uniform(0, 1)$. We assumed that the intercept and the slope are common for the first and the second responses. Moreover, we assumed that the conditional probability of success ($P(Y_{i1j} = 1|X_{i1j}, b_{i1})$) is greater for the second response compared to the first one, i.e. $\lambda_2^* > \lambda_1^*$. Δ_{i1j}^* were obtained depending on the other model parameters such that

$$\Delta_{i1j}^* = \sqrt{1 + \lambda_j^{*2} \sigma_1^2}\, (\beta_0^* + \beta_1^* X_{i1j1} + \beta_2^* X_{i1j2})$$

The binary responses were obtained by the random effects probabilities specified by the last level of the model (4.80).

For $t \geq 2$, we consider a model with the framework given by

$$P(Y_{itj} = 1|X_{itj}) = \Phi(\beta_0 + \beta_1 X_{itj1} + \beta_2 X_{itj2}) \tag{4.81}$$

$$P(Y_{itj} = 1|y_{i,t-1,j}, X_{itj}) = \Phi(\Delta_{itj} + \alpha_{t,1} Z_{itj} y_{i,t-1,j}) \tag{4.82}$$

$$P(Y_{itj} = 1|y_{i,t-1,j}, X_{itj}, b_{it}) = \Phi(\Delta^*_{itj} + \lambda_j b_{it}) \tag{4.83}$$

where $\beta = (\beta_0, \beta_1, \beta_2) = (-1, 2, 0.2)$, $\alpha_{t,1} = (\alpha_{21,1}, \alpha_{31,1}, \alpha_{41,1}) = (0.5, 0.7, 0.9)$, $\lambda_j = (\lambda_1, \lambda_2) = (1, 1.05)$ and $b_{it} \sim N(0, \sigma_t^2)$, $\sigma_t = (\sigma_2, \sigma_3, \sigma_4) = (0.66, 0.63, 0.60)$. X_1 is assumed to be a time independent variable, i.e. $X_{itj1} = X_{i1j1}$. X_2 is a response type indicator variable for which while the first response takes 1, the second one takes 0. By the inclusion of response type as a covariate, we allow bivariate responses having different intercepts, i.e. while for the first response the intercept is $\beta_0 + \beta_2 = -1 + 0.2 = -0.8$, for the second response the intercept is $\beta_0 = -1$. Additionally, the effect of X_1 was assumed to be shared across the responses. Moreover, we assumed that the effects of the past responses on the current ones are not interacted by any covariates, i.e. $Z_{itj} = 1$. This also indicates that the transition parameters are shared across responses, since Z_{itj} do not include response type indicator variables. We also assumed that the serial dependence is slightly increasing by time, i.e. $\alpha_{4,1} > \alpha_{3,1} > \alpha_{2,1}$. As for $t = 1$, we assumed that the conditional probability of success is greater for the second response compared to the first one, i.e. $\lambda_2 > \lambda_1$. To reflect the common expectancy that more variability is expected at baseline and the variability decreases by time, we assumed that $\sigma_1 > \sigma_2 > \sigma_3 > \sigma_4$.

Δ_{itj} were obtained by a function of β and $\alpha_{t,1}$ such that

$$\Delta_{itj} = A_{itj}(\beta - \beta_0) + B_{itj}(\alpha_{t,1} - \alpha_{t,10}) \tag{4.84}$$

where β_0 and $\alpha_{t,10}$ are the first two components of implicit function theorem (IFT) point, $P_0 = (\beta_0, \alpha_{t,10}, \Delta_{itj0})$. β_0 and $\alpha_{t,10}$ are set to 0, i.e. $\beta_0 = (\beta_{00}, \beta_{10}, \beta_{20}) = (0, 0, 0)$ and $\alpha_{t,10} = (\alpha_{2,10}, \alpha_{3,10}, \alpha_{4,10}) = (0, 0, 0)$. Δ_{itj0} are equal to 0 for $t \geq 3$. However, Δ_{i2j0} were obtained by Newton-Raphson Algorithm (N-R) by solving (4.15) given β_0 and $\alpha_{t,10}$, since for $t = 2$, β^* replaces β as the multiplier of the lag-1 covariates. The full details of obtaining (4.84) and the related arguments could be found in Section 4.2.2.1.

Δ^*_{itj} were obtained as a function of Δ_{itj}, $\alpha_{t,1}$, λ_j and σ_t such that

$$\Delta^*_{itj} = \sqrt{1 + \lambda_j^2 \sigma_t^2}\, (\Delta_{itj} + \alpha_{t,1} Z_{itj,1} y_{i,t-1,j}) \tag{4.85}$$

Then, the binary responses for $t \geq 2$ were obtained by the random effects probabilities calculated by the third level of the model (4.83).

We replicate the simulation study 100 times. In each of these replications, the initial values of the parameters to start the Fisher-Scoring Algorithm were obtained by various methods to mimic the real life. The initials of β^* were obtained by probit binary regression models which were pooled over the bivariate responses. While the initial of λ_2^* was set to 1, the initial of $\log(\sigma_1)$ was set to $\log(0.5)$. The initials of β were obtained by multivariate marginal models with response specific parameters by using the `mmm` function (**mmm** package; Asar and Ilk, 2012). The initials of $\alpha_{t,1}$ were obtained by time specific transition models with probit link which were pooled over the bivariate responses. To exemplify, transition models include regressing the responses at time $t - 1$ on the ones at time t. As for the baseline time point, while the initial of λ_2 was set to 1, the initials of $\log(\sigma_t)$ were set to $\log(0.5)$ for $t \geq 2$. Setting the initials of λ_j^* and λ_j for $j \geq 2$ to 1 is reasonable, since we assume λ_1^* and λ_1 are equal to 1. Additionally, the related hypothesis testing places equality of λ_j^* and λ_j for $j \geq 2$ to 1 in the null hypothesis. Equating the initials of σ_t to 0.5 was one of the possible choices and we selected that value to accelerate the convergence. Some smaller and larger values were tried and no convergence problem was observed. The β_0 and $\alpha_{t,10}$ components of the IFT point P_0 were set to 0 and Δ_{i2j0} were obtained by Newton-Raphson methods. The tolerances were selected as 0.0001 for both baseline and $t \geq 2$ models. Analysis of one simulated data (the last one) by PNMTREM(1) took 13.8 minutes on a PC with 2.00 GB RAM and 2.79 GHz processor.

While constructing the simulation study, we tried several combinations regarding the components of the IFT point, P_0, for both data generation and model building processes. We brought up that setting the β_0 and $\alpha_{t,10}$ components of P_0 to 0 is more reasonable compared to other combinations. Then, Δ_{itj} for $t \geq 3$ is equal to 0 as well and for $t = 2$, we need to obtain Δ_{i2j0} by N-R method.

4.4.2 Simulation Results

The simulation results confirmed the correctness of the written codes and the derivations done for parameter estimation of PNMTREM(1) (Table 4.1). The marginal mean parameters were estimated very well for the baseline model. They reflect ignorable bias and small MSE values. Whereas these quantities were -0.004 and 0.007 for the intercept, they were 0.005 and 0.022 for X_1 which was a covariate coming from the standard uniform distribution. Similarly, the response specific association parameter λ_2^* were estimated with low bias and low MSE, the corresponding values are 0.038 and 0.004, respectively. The related estimates indicated the truth about the relative success probabilities of the bivariate responses which was the second response had greater success probabilities compared to the first one, i.e. $\hat{\lambda}_2^* > \hat{\lambda}_1^*$. On the other hand, some negative and sizable bias was observed for the random effects variance, σ_1: the mean over the 100 replications were approximately 0.62 (= $exp(-0.484)$). Nonetheless, it was estimated with a small variability (MSE=0.026).

As for the baseline model, marginal mean parameters for the $t \geq 2$ model were estimated very well. Again, the maximum MSE were observed for X_1 coming from the standard uniform distribution as 0.009. While the intercept and response type indicator variable were estimated as almost unbiased, they were estimated with lower variability compared to X_1; the corresponding MSE values were 0.004 and 0.002, respectively. The transition parameters, $\alpha_{t,1}$ were estimated as unbiased and with MSEs close to 0 as well. Whereas the maximum bias was seen for the transition parameter at time 3 as -0.013, the maximum MSE was seen for the transition parameter at times 3 and 4 as 0.009. Actually, the bias and MSE values for different transition parameters were very close to each other. However, the bias for $\alpha_{4,1}$ was -0.001 which is very close to zero. Sizable bias were seen for the parameters of the third level of the model. The response specific parameter for the second response, λ_2 reflected overestimation with a bias of 0.097. Nonetheless, it had a small MSE as 0.019 which is comparable with the ones for the mean and transition parameters. On the other hand, the estimates of λ_2 indicated higher success probability for the second response which is actually the ground truth in our simulation study. As for the baseline model, the random effects variance parameters for the $t \geq 2$ model reflected underestimation. While the true parameter values were $(\sigma_2, \sigma_3, \sigma_4) = (0.66, 0.63, 0.60)$ the corresponding mean values of 100 estimates were $(0.601, 0.548, 0.508)$. Besides, those estimates had small MSE values with a maximum MSE for $\log(\sigma_3)$ as 0.063. Nevertheless, the estimates of random effects variances reflected the truth, i.e. $\hat{\sigma}_2 > \hat{\sigma}_3 > \hat{\sigma}_4$.

Table 4.1: Simulation results for baseline and $t \geq 2$ models over 100 replications.

Baseline				
Parameter	True	Mean	Bias	MSE
β_0^*	-1.000	-1.004	-0.004	0.007
β_1^*	1.900	1.905	0.005	0.022
λ_2^*	1.070	1.108	0.038	0.004
$\log(\sigma_1)$	$\log(0.7) = -0.357$	-0.484	-0.128	0.026
t ≥ 2				
Parameter	True	Mean	Bias	MSE
β_0	-1.000	-1.009	-0.009	0.004
β_1	2.000	2.016	0.016	0.009
β_2	0.200	0.202	-0.002	0.002
$\alpha_{2,1}$	0.500	0.498	0.010	0.007
$\alpha_{3,1}$	0.700	0.687	-0.013	0.009
$\alpha_{4,1}$	0.900	0.899	-0.001	0.009
λ_2	1.050	1.147	0.097	0.019
$\log(\sigma_2)$	$\log(0.66) = -0.416$	-0.509	-0.094	0.027
$\log(\sigma_3)$	$\log(0.63) = -0.462$	-0.602	-0.140	0.053
$\log(\sigma_4)$	$\log(0.60) = -0.511$	-0.678	-0.168	0.063

4.5 Real Life Data Application

4.5.1 Data

The dataset used to illustrate the validity of PNMTREM(1) in real life came from the Iowa Youth and Families Project (IFYP) which was discussed in Section 1.3.2 in details. The main aim of IYFP was to investigate the long term effects of the farm crisis (began in 1980s in America) on the well being of family members living in rural parts of the country. 451 families from rural parts of Iowa were included in this study. The focus was on 7th graders with two biological parents and a sib within 4 years of age. The project was conducted between the period of 1989 and 1999 including an 11-year follow-up. Whereas it was conducted yearly until 1992, it was continued at 1994, 1995, 1997 and 1999. Therefore, here we considered only the first 4-year follow-up (a fully constrained portion) of the dataset. At each follow-up, information of both young people and their families were collected (Table 1.4). The variables regarding young people were gender and the negative life experiences of them such as having a close friend who moved away. On the other hand, the variables regarding the household information included negative economic events experiences such as replacing the current job for a worse one and economical cutback experiences such as moving to a cheaper and worse place. The main aim of collecting household level information together with individual level information is that economic hardship might influence young people indirectly such as harsh parenting. The distress levels of young people were measured by a symptom checklist which included several criterion such as shakiness, nervousness and feeling low in energy and etc. These distress variables were gathered under three main variables which were anxiety, hostility and depression and they were later dichotomized regarding whether any symptom was present (Ilk and Daniels, 2007). Since some of the variables (NEE and cut) include more than two levels and we included response type indicator variables and follow-up times as additional covariates, we reconstructed the dataset which was presented in Section 1.3.2 and the details about the response variables and the covariates regarding the new dataset were presented in Table 4.2.

Table 4.2: Variable list of IYFP used in PNMTREM(1).

Variable	Explanation
anxiety	whether the young person had symptoms: 0=absence, 1=presence
hostility	whether the young person had symptoms: 0=absence, 1=presence
depression	whether the young person had symptoms: 0=absence, 1=presence
gender	gender of the young person: 0=male, 1=female
NLE1	first indicator variable for negative life event experiences of young people: 1=some, 0=none or lots of
NLE2	second indicator variable for negative life event experiences of young people: 1=lots of, 0=none or some
NEE	whether the household had any negative economical event: 0=no, 1=yes
cut1	first indicator variable for financial cutback experiences of the household: 1=between 1 and 5, 0= none or more than 5
cut2	second indicator variable for financial cutback experiences of the household: 1=more than 5, 0= none or between 1 and 5
rtype1	first response type indicator variable: 1=hostility, 0=anxiety or depression
rtype2	second response type indicator variable: 1=depression, 0=hostility or anxiety
time1	first indicator variable for follow-up time: 1=1991, 0=1990 or 1992
time2	second indicator variable for follow-up time: 1=1992, 0=1990 or 1991

While there were no missing observations at baseline (1989), there were varying amounts of missing observations for both the responses and the covariates at later time points. These missing data include both intermittent (non-monotone) and drop-out (monotone) missingness. The responses have missing percentages of 6%, 9.8% and 10.6% for the years of 1990, 1991 and 1992, respectively. Among the covariates NLE, NEE and cut are time-varying covariates and all of them had missing percentages of 6%, 9.8% for both 1990 and 1991, respectively. Additionally, the related missing percentages were 10.6%, 10.4% and 10.4% at 1992, respectively. Lorenz et al. (1997, cited in Ilk and Daniels, 2007) noted that there were no differences between the observed responses or measured covariates of the people who completed the study and the ones who dropped-out. Moreover, Ilk and Daniels (2007) reported that missing data occurred mostly due to relocation of the respondents (due to a job change or leaving home for higher education) based on their discussions with the investigators of IFYP. Both of these information about the nature of the missing data arose in IYFP study indicate that the reason why people dropped-out might most probably be due to some reasons other than the missing observations themselves. Therefore, the related missing mechanism can be assumed to be at random (MAR). Since our aim in this study is not on data analysis under missing data, we imputed these missing observations before our analyses. We utilized random imputation by mimicking the observed proportions of the available data regarding the period of 1989 to 1999. Note that only the time dependent covariates and responses have missing observations and all of them are categorical. To illustrate, while imputing the missing data in anxiety, we generated random numbers from Bernoulli distribution with success probability of 0.67 starting with a random seed number, since the observed proportion of the presence in anxiety between 1989 and 1999 is approximately 0.67. Moreover, we imputed NLE by generating random numbers from its categorical (pseudo) distribution based on the empirical proportions of the data by using the Accept-Reject method. The related proportions were 0.09, 0.58 and 0.32 for categories 0, 1 and 2, respectively. Discrete uniform distribution with boundaries 0 and 2 were used in this approach. The imputed data is available upon request from the authors.

PNMTREM(1) enables us to answer several questions which would be asked for both the comparison of the sub-groups of young people and/or their families and for some specific young persons. Moreover, it permits making different statistical inferences for $t = 1989$ and $t \geq 1990$ periods. For instance, we can compare the distress levels of males and females by the first levels of both baseline and $t \geq 2$ models. The inclusion of the interaction between gender and response type indicator variables in the design matrices permits response specific comparison of the gender, i.e. comparison of anxiety, hostility and depression levels of males and females separately. We can measure the effect of the past year's distress status on the current ones by the second level of $t \geq 2$ model. The inclusion of the interaction between the lag-1 responses and the response type indicator variables allows us to make response specific inferences about the transition probabilities. For instance, we can measure the effect of the anxiety status of young people at 1990 on the ones at 1991. Furthermore, we can draw subject-specific inferences for 1989 by the second level of the baseline model and for $t \geq 1990$ by the third level of $t \geq 2$ model. For instance, we can calculate the probability of being depressed for subject 223 at year 1990. Note that this probability is subject, time and response specific.

4.5.2 Results

In this section, we presented the PNMTREM(1) modeling results on the IYFP dataset. Specifically, we built two different models. While the marginal regression parameters of these models are same, they differ in terms of separating the effects of the distress status histories on the current distress status for multiple responses. Put another way, second model (Model 2 in Table 4.5) includes response type indicator variables in the design matrix Z_{itj}, i.e. $Z_{itj} = [1 \quad rtype1 \quad rtype2]$. On the other hand, the first model (Model 1 in Table 4.5) includes only ones in the design matrix Z_{itj}, i.e. $Z_{itj} = [1]$. Hence, Model 1 assumed that the effect of the past year's distress status on the current ones are shared across anxiety, hostility and depression. Since baseline models are same for Model 1 and Model 2, we presented only one baseline modeling result in Table 4.4. In Table 4.4, pooled probit GLM results (pooled over the multiple responses) are presented as well. Note that probit GLM results ignores the multivariate response dependency. Indeed, these are the estimates which we used to start the Fisher-Scoring (F-S) algorithm for the baseline model. The results of the $t \geq 2$ models are presented in Table 4.5. Additionally, in Table 4.5 we include the results of multivariate marginal models with shared regression parameters (MMM2) as well. Note that MMM2 accommodates both the serial and multivariate response dependencies and the related estimates were used to start the F-S algorithm for the $t \geq 2$ model. Since Model 1 and Model 2 are nested models, we can compare them by likelihood ratio test (LRT). Note that both Model 1 and Model 2 comprise of baseline and $t \geq 2$ models. Therefore, the corresponding maximized log-likelihoods are the summation of the ones for baseline and $t \geq 2$ models: -1535.15 $(= -207.89 - 1327.26)$ and -1539.46 $(= -207.89 - 1331.57)$ for Models 1 and 2, respectively. The LRT statistics for the comparison of these models is 8.62 $(= -2*(-1539.46-(-1535.15)))$ with p-value of 0.20 which indicates that there is not enough evidence to conclude that Model 2 better explains the data compared to Model 1 with 95% confidence level, $(\chi^2_{6,0.95} = 12.59)$.

In the light of this result, throughout we only consider Model 1 while making interpretations about the parameter estimates.

Since all the covariates included Model 1 and Model 2 are binary variables (Table 4.2), we had some difficulties during the convergence of the related model fitting algorithms. We overcame this issue by applying variable base standardization of all the covariates for the periods of $t = 1989$ and $t \geq 1990$ separately. Note that the interactions were initially calculated by using the binary values, then they are standardized as well. The values which the binary variables took after the standardization are presented in Table 4.3.

Table 4.3: The standardized values of the independent variables.

Variable	Binary Code	Standardized Values	
		t = 1989	$t \geq 1990$
gender	0 (1)	-1.0473 (0.9541)	-1.0476 (0.9544)
NLE1	0 (1)	-0.9669 (1.0334)	-1.2175 (0.8212)
NLE2	0 (1)	-0.9799 (1.0198)	-0.7271 (1.3750)
NEE	0 (1)	-1.3036 (0.7666)	-1.1526 (0.8674)
cut1	0 (1)	-0.7775 (1.2852)	-0.8149 (1.2269)
cut2	0 (1)	-0.9044 (1.1049)	-0.8326 (1.2008)
rtype1	0 (1)	-0.7069 (1.4137)	-0.7070 (1.4140)
rtype2	0 (1)	-0.7069 (1.4137)	-0.7070 (1.4140)
time1	0 (1)		-0.7070 (1.4140)
time2	0 (1)		-0.7070 (1.4140)
gender*rtype1	0 (1)	-0.4595 (2.1748)	-0.4596 (2.1753)
gender*rtype2	0 (1)	-0.4595 (2.1748)	-0.4596 (2.1753)
rtype1*time1	0 (1)		-0.3535 (2.8281)
rtype1*time2	0 (1)		-0.3535 (2.8281)
rtype2*time1	0 (1)		-0.3535 (2.8281)
rtype2*time2	0 (1)		-0.3535 (2.8281)

We checked the existence of multicollinearity problem by the Variance Inflation Factor (VIF) regarding pooled logistic regression analyses via the `vif` function available under the R package **HH** (Heiberger, 2009). Results showed that there were no such a problem, i.e. none of the VIF values were greater than 10 for the data at 1989 and none of the VIF values were greater than 5 for the data at the period of 1990-1992.

As is stated in Section 4.2, PNMTREM accommodates only exogenous time dependent covariates. Ilk and Daniels (2007) checked this assumption for IYFP by regressing the time dependent covariates, which are negative life and economic events and financial cutbacks, on the history of these variables, the history of responses and adjusting for the time independent covariate (gender) and time. They reported that the time varying covariates were not predicted by the previous responses, i.e. all the confidence intervals regarding the odds-ratios covered 1.

Table 4.4: PNMTREM(1) and probit GLM results on IYFP data for $t = 1989$. $H_0 : \lambda_j = 1$ for $j = 2, 3$; others are tested for 0.

	PNMTREM(1)				GLM			
Parameter	Est.	SE	Z	P	Est.	SE	Z	P
β_0^*	1.22	0.05	23.36	0.00	1.16	0.05	25.10	0.00
β_{gender}^*	0.04	0.08	0.53	0.60	0.02	0.07	0.29	0.77
β_{NLE1}^*	0.20	0.13	1.58	0.12	0.19	0.12	1.64	0.10
β_{NLE2}^*	0.41	0.13	3.23	0.00	0.39	0.12	3.24	0.00
β_{NEE}^*	0.03	0.05	0.73	0.47	0.03	0.05	0.67	0.50
β_{cut1}^*	0.07	0.07	1.15	0.25	0.07	0.06	1.11	0.27
β_{cut2}^*	-0.01	0.07	-0.08	0.94	0.01	0.06	0.12	0.91
β_{rtype1}^*	0.04	0.07	0.53	0.60	0.03	0.07	0.49	0.63
β_{rtype2}^*	0.35	0.09	3.87	0.00	0.33	0.08	3.91	0.00
$\beta_{gender*rtype1}^*$	-0.06	0.08	-0.76	0.45	-0.06	0.08	-0.72	0.47
$\beta_{gender*rtype2}^*$	-0.13	0.10	-1.30	0.19	-0.13	0.09	-1.47	0.14
$\lambda_{hostility}^*$	1.11	0.74	0.15	0.88				
$\lambda_{depression}^*$	1.04	0.65	0.06	0.95				
$\log(\sigma_1)$	-0.33	0.38						
Max. loglik.	-207.89				-511.98			

Baseline results (Table 4.4) indicate that only the intercept, one of the negative life events indicator variables (NLE2) and one of the response type indicator variables (rtype2) were significant at 95% confidence level in 1989. We can interpret our *probit* marginal mean parameters as in the case of *logit* estimates in terms of the odds-ratios by using the JKB constant (Johnson et al. 1995, pp. 113-163 cited in Griswold, 2005 pp. 85-96). Related approach offers an approximate and deterministic relationship between the *probit* and *logit* estimates, i.e. $\beta_{logit} = c * \beta_{probit}$ where $c = (15/16)(\pi/\sqrt{3}) = 1.700437$. Young people who experienced lots of negative life events were while approximately 4.03 ($= exp(1.700437 * ((0.20 * -0.9669 + 0.41 * 1.0198) - (0.20 * -0.9669 + 0.41 * -0.9799)))$) times more likely to be distressed compared to the ones who did not have such an experience, these young people were 2.04 ($= exp(1.700437 * ((0.20 * -0.9669 + 0.41 * 1.0198) - (0.20 * 1.0334 + 0.41 * -0.9799)))$) times more likely to be distressed compared to the ones who had some such an experience. Moreover, teenagers were more likely to be depressed compared to them being anxious or hostile; the related odds-ratios of being depressed with respect to anxiety and hostility were 3.53 ($= exp(1.700437 * ((0.04 * -0.7069 + 0.35 * 1.4137) - (0.04 * -0.7069 + 0.35 * -0.7069)))$) and 3.06 ($= exp(1.700437 * ((0.04 * -0.7069 + 0.35 * 1.4137) - (0.04 * 1.4137 + 0.35 * -0.7069)))$), respectively. There are not enough evidence to say that the pairwise correlations between anxiety, hostility and depression are significantly different; corresponding p-values of $\lambda^*_{hostility}$ and $\lambda^*_{depression}$ are 0.88 and 0.95. For the structure of the pairwise correlations see the formulation given in (4.4). The standard deviation of the random effects distribution at 1989 was estimated as 0.72 (= $exp(-0.33)$). Although PNMTREM(1) and probit GLM results seem to be in agreement, the serious difference between the maximized log-likelihoods of these models (related values were -207.89 and -511.98 for the former and latter, respectively) indicates that fitting a marginalized random effects model (baseline model of PNMTREM(1)) better explains the data at 1989. Note that usual likelihood ratio test (LRT) does not hold here, since the related model comparison includes the equality of the variance component to its lower bound, i.e. $\sigma_1 = 0$, in the null hypothesis (Molenberghs and Verbeke, 2007).

For the later time points (1990 − 1992), intercept, gender, both of the negative life events indicator variables (NLE1, NLE2), one of the response type indicator variables (rtype1), one of the time indicator variables (time2) and some of the interactions were significant at 95% confidence level (Table 4.5). Young people who were females and had negative life and economic experiences seem to be more likely to be distressed. Note that gender was found to be insignificant at 1989 and this result was supported by Ge et al. (2001, cited in Ilk, 2008) and Ilk (2008). Females were (approximately) 3.05 ($= exp(1.700437*((0.13*0.9544+0.15*2.1753+0.02*(-0.4596))-(0.13*(-1.0476)+0.15*(-0.4596)+0.02*(-0.4596)))))$ times more likely to report feeling hostile compared to males. Young people who experienced lots of negative life events had high probability of being distressed. For instance, the ones who experienced lots of negative life events are while approximately 3.14 ($= exp(1.700437*((0.13*-1.2175+0.32*1.3750)-(0.13*-1.2175+0.32*-0.7271))))$ times more likely to be distressed compared to the ones who did not have such an experience and they are 2 ($= exp(1.700437*((0.13*-1.2175+0.32*1.3750)-(0.13*0.8212+0.32*-0.7271))))$ times more likely to be distressed compared to the ones who had some such an experience. Note that these approximate odds ratios were higher at baseline. Nonetheless, there is not enough evidence to conclude that negative economic event experience (NEE) and financial cutbacks (cut1 and cut2) had significant effects on distress status of young people with 95% confidence. The distress probabilities of young people seem to be decreasing with time ($\hat{\beta}_{time1} = -0.06$, $\hat{\beta}_{time2} = -0.14$). At 1992, young people were more likely to be hostile compared to the previous years ($\hat{\beta}_{rtype1*time2} = 0.13$).

Table 4.5: PNMTREM(1) and probit MMM2 results on IYFP data for $t \geq 1990$. $H_0 : \lambda_j = 1$ for $j = 2, 3$; others are tested for 0.

	PNMTREM(1)								MMM2		
	Model 1				Model 2						
Parameter	Est.	SE	Z	P	Est.	SE	Z	P	Est.	SE	Z
β_0	0.78	0.03	30.01	0.00	0.78	0.03	26.72	0.00	0.71	0.05	15.21
β_{gender}	0.13	0.04	2.91	0.00	0.13	0.04	3.03	0.00	0.14	0.05	2.96
β_{NLE1}	0.13	0.04	3.13	0.00	0.13	0.04	2.14	0.00	0.14	0.06	2.39
β_{NLE2}	0.32	0.04	7.38	0.00	0.32	0.04	7.40	0.00	0.31	0.06	5.07
β_{NEE}	0.04	0.02	1.56	0.12	0.04	0.02	1.59	0.11	0.03	0.03	1.10
β_{cut1}	0.05	0.03	1.56	0.12	0.05	0.03	1.55	0.12	0.04	0.04	1.11
β_{cut2}	0.03	0.03	0.97	0.33	0.03	0.03	0.92	0.36	0.03	0.04	0.79
β_{rtype1}	0.31	0.06	5.21	0.00	0.32	0.06	5.19	0.00	0.40	0.05	8.05
β_{rtype2}	0.04	0.07	0.55	0.58	0.04	0.07	0.63	0.53	-0.002	0.05	-0.03
β_{time1}	-0.06	0.05	-1.30	0.19	-0.06	0.05	-1.20	0.23	-0.06	0.04	-1.51
β_{time2}	-0.14	0.06	-2.43	0.02	-0.13	0.06	-2.18	0.03	-0.14	0.04	-3.63
$\beta_{gender*rtype1}$	0.15	0.04	3.83	0.00	0.15	0.04	4.01	0.00	0.13	0.05	2.60
$\beta_{gender*rtype2}$	0.02	0.06	0.35	0.73	0.02	0.06	0.38	0.71	-0.03	0.05	-0.59
$\beta_{rtype1*time1}$	0.03	0.05	0.64	0.52	0.02	0.05	0.31	0.75	0.03	0.05	0.61
$\beta_{rtype1*time2}$	0.13	0.05	2.75	0.01	0.12	0.05	2.22	0.03	0.12	0.05	2.35
$\beta_{rtype2*time1}$	0.02	0.05	0.43	0.67	0.02	0.05	0.37	0.71	0.02	0.05	0.49
$\beta_{rtype2*time2}$	0.05	0.06	0.97	0.33	0.06	0.07	0.90	0.37	0.04	0.05	0.89
$\alpha_{21,1}$	0.78	0.11	7.49	0.00	0.71	0.12	5.84	0.00			
$\alpha_{22,1}$					0.19	0.11	1.68	0.01			
$\alpha_{23,1}$					-0.003	0.15	-0.02	0.99			
$\alpha_{31,1}$	0.98	0.11	10.57	0.00	0.94	0.10	9.38	0.00			
$\alpha_{32,1}$					0.17	0.10	1.77	0.08			
$\alpha_{33,1}$					0.02	0.13	0.14	0.89			
$\alpha_{41,1}$	0.79	0.09	8.91	0.00	0.62	0.10	6.15	0.00			
$\alpha_{42,1}$					0.09	0.10	0.93	0.35			
$\alpha_{43,1}$					-0.07	0.13	-0.54	0.59			
$\lambda_{hostility}$	1.30	0.48	0.63	0.53	1.29	0.48	0.60	0.55			
$\lambda_{depression}$	1.26	0.42	0.62	0.54	1.30	0.43	0.69	0.49			
$\log(\sigma_2)$	-0.67	0.25			-0.69	0.25					
$\log(\sigma_3)$	-0.70	0.24			-0.69	0.24					
$\log(\sigma_4)$	-0.68	0.24			-0.65	0.24					
Max. loglik	-1327.26				-1331.57						

Note: Standard error and Z estimates of MMM2 are the robust ones. MMM2 was fitted under unstructured working correlation matrix.

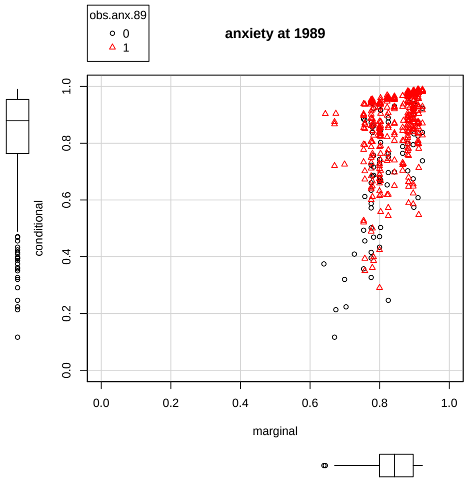
The transition parameter estimates indicated that young people who were distressed at year $t-1$ were more likely to be distressed at year t compared to the ones who were not distressed at year $t-1$. Although Model 2 resulted that young people who had hostility at 1989 were more likely to have hostility at 1990 compared to his/her anxiety and depression statuses, LRT indicated that there is not enough evidence to fit Model 2 instead of Model 1 to better explain the IFYP data. As for the baseline model, there are not enough evidence to say that the pairwise correlations between anxiety, hostility and depression are significantly different; corresponding p-values were 0.53 and 0.54 for hostility and depression, respectively. The estimates of the standard deviations of the random effects distributions were found to be 0.51

(= $exp(-0.67)$), 0.50 (= $exp(-0.70)$) and 0.51 (= $exp(-0.68)$), respectively at 1990, 1991 and 1992 which indicate that the individual variations are very close to each other at these years. Note that the estimate of σ_1 ($\hat{\sigma}_1$=0.72) is relatively larger than these estimates which indicates more variability among the young people at 1989 and this variability decreased after the baseline.

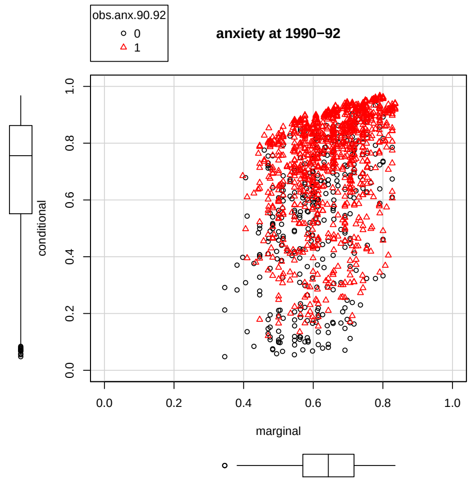
PNMTREM(1) and MMM2 marginal regression parameter estimates seem to be in agreement and they are even same. Two exceptions were observed: 1) for one of the response type indicator variables (rtype2); the related estimates were 0.04 and -0.002 for PNMTREM(1) and MMM2, respectively, 2) for interaction by gender and one of the response type indicator variables (gender*rtype2); the related estimates were 0.02 and -0.03 for PNMTREM(1) and MMM2, respectively. Nevertheless, both of the modeling results for these estimates were statistically insignificant. Actually, similar results between these models are expected since MMM2 takes into account both serial, multivariate and cross-temporal response dependencies. Nonetheless, MMM2 does not give reliable idea about the dependence structures and does not permit subject-specific interpretation. Moreover, MMM2 only used the data belonging to the period $t \geq 1990$, since it does not permit specifying different covariate sets for baseline and later time points. On the other hand, PNMTREM(1) took into account the data at 1989 via the connection of baseline and $t \geq 2$ models with the marginal constraint equation, see (4.10, 4.19). Our PNMTREM(1) results for the IYFP datasets mostly coincide with the ones reported by Ilk (2008). Note that while Ilk (2008) used data augmentation (Tanner and Wong, 1987 cited in Ilk, 2008) to impute the missing data, we used random imputation method. We observed that Model 1 and Model 2 produced very similar, mostly same marginal regression parameter estimates and the related Z statistics and p-values were close to each other. This is natural due to the fact that marginalized models are insensitive to the misspecification of the dependence structures (Heagety and Kurland, 2001). Moreover, Heagerty (2002b) proved that the parameters of the first and second levels of marginalized transition models (MTM) are orthogonal and note that the first and second levels of PNMTREM are MTM indeed.

Up to here, we have drawn population-averaged statistical inferences regarding PNMTREM(1)-Model 1. Besides, we can draw individual-level inferences by using PNMTREM(1), i.e. via the second level of the baseline model and the third level of the $t \geq 2$ model. To illustrate, we calculated the success probabilities regarding anxiety, hostility and depression of each person at each year by using the last levels of these models. In addition to these random effects probabilities, we calculated marginal probabilities by using the first levels of these models. The related probabilities are summarized in Figure 4.2. Note that the related graphics were obtained by using the `scatterplot` function under the R package **car** (Fox and Weisberg, 2011). In these graphics, while the observed values equal to 0 were represented by circles, the ones equal to 1 were represented by triangulars. Whereas the conditional probabilities ranged almost between the lower and upper probability bounds, the marginal counterparts ranged in a narrower interval. For instance, while the marginal probabilities of having hostility at the period of 1990-1992 took only the values in the interval of (0.70,0.98), the conditional probabilities ranged between 0.13 and 1 (Figure 4.2d). This means that even the young people who had actually no hostility for that period were assigned more probability of being hostile by the marginal models which would yield wrong decisions. On the other hand, we observed that the conditional probabilities spread widely and yielded higher rates of correct decisions. For instance, in Figure 4.2d, the circular points (observing no hostility for a person) were associated with lower conditional probabilities and this is what we expected. The associated box-plots reflected the location and scale information of these marginal and conditional probabilities as well. For instance, whereas the box-plot of the conditional probabilities reflected a spread distribution and lots of outlying probabilities, the marginal counterparts reflected a stacked and narrow distribution. Marginal models only rely on how well the covariates explain the variation of the responses and ignores the individual characteristics. Put another way, two young people with same covariates but different unobserved features would have the same probability of being anxious based on the results of marginal models. However, in random effects models these individual features are accounted by the random effects parameters in addition to the covariate effects. The reason that marginal probabilities stacked in a narrower interval and tend to assign high probabilities to the cases in which distress variables were absent is most probably due to these facts.

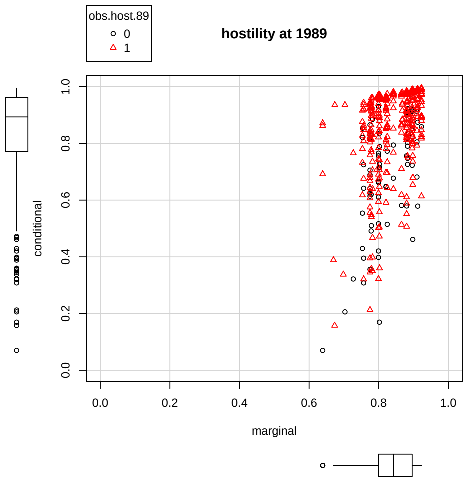
We built simple linear regression models considering the probit of the conditional probabilities, $\Phi^{-1}(P^r(Y_{i1j}))$ and $\Phi^{-1}(P^r(Y_{itj}))$, as dependent variables and the probit of the marginal probabilities, $\Phi^{-1}(P^m(Y_{i1j}))$ and $\Phi^{-1}(P^m(Y_{itj}))$ as independent ones to measure how much the variation in the responses were explained by the covariates. R-squares of these models were presented in Table 4.6. We observed that covariates in the IYFP data does not explain the individual characteristics well, since only 30% of the individual variations were explained by the



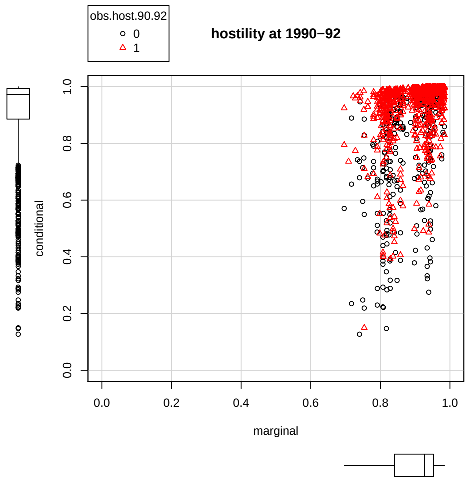
(a) Anxiety at 1989



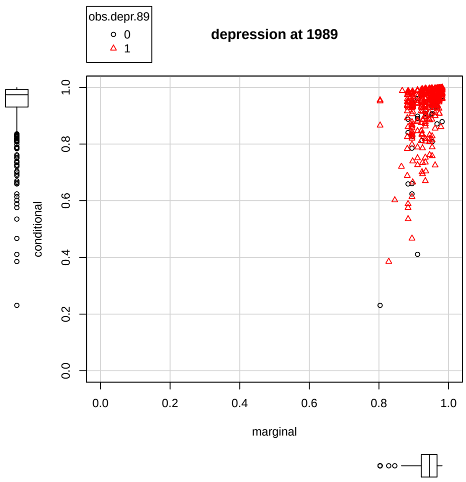
(b) Anxiety at 1990-92



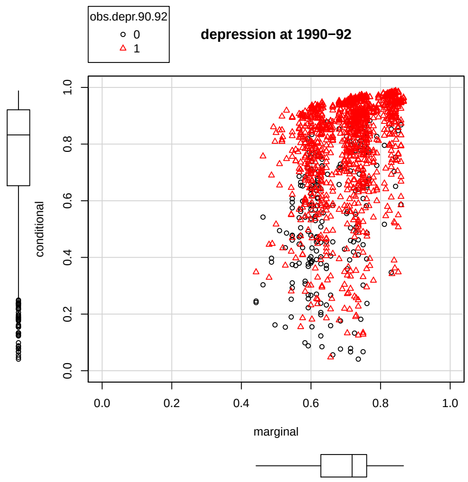
(c) Hostility at 1989



(d) Hostility at 1990-92



(e) Depression at 1989



(f) Depression at 1990-92

Figure 4.2: Scatter and box plots of marginal vs. conditional probabilities at 1989 and 1990-1992.

covariates at 1989 for response=depression (R-square=0.30) and only 31% of the individual variations were explained by the covariates at 1990-1992 response=hostility (R-square=0.31).

Table 4.6: R-squares of the simple linear models which were constructed considering $\hat{\Delta}^*_{i1j} + \hat{\lambda}_j^{\,*}\hat{b}_{i1j}$ and $\hat{\Delta}^*_{itj} + \hat{\lambda}_j\hat{b}_{itj}$ as dependent variables and $X\hat{\beta}^*$ and $X\hat{\beta}$ as independent variables.

Response	1989	1990-1992
Anxiety	0.26	0.20
Hostility	0.25	0.31
Depression	0.30	0.21

Interactive graphics, specifically GGobi (Cook and Swayne, 2007), might help to identify interesting people. In Figure 4.3, we illustrated this identification with an example. For instance, the young person with ID=223 actually reported no hostility at 1990. For this person, while the marginal model estimated a high probability of having hostility (0.94), the conditional counterpart estimated a low probability (0.28). Additionally, we observed that points near that identified point in Figure 4.3 include other observations of the same person. We concluded that for this person while conditional model tend to yield correct inference, the marginal model tend to yield wrong inference. Similarly, we identified some other young people for whom marginal and conditional models yielded different inferences again by using GGobi such as the one with ID=359.

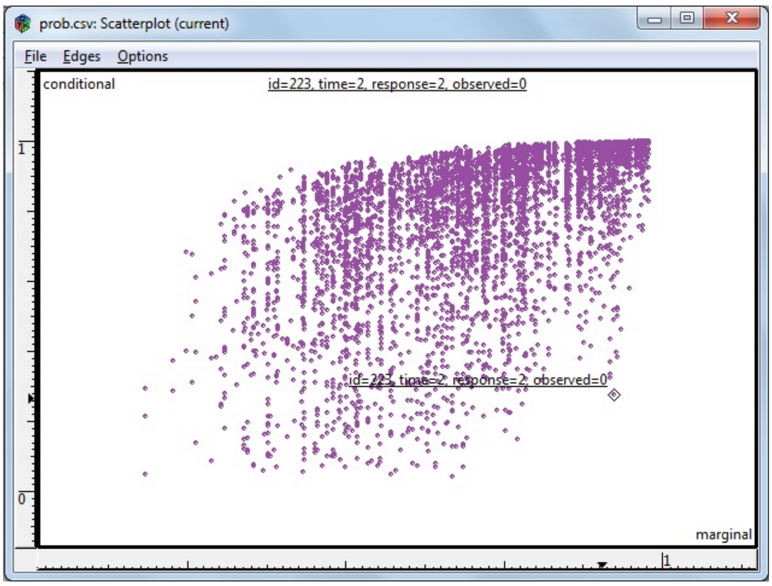


Figure 4.3: Interactive scatter plot of marginal versus conditional probabilities regarding the period of 1989-1992. An illustration of identifying interesting people.

Conditional probabilities can also be calculated by assuming that the person is an average person, i.e. $b_{it} = 0$. In this case, these probabilities could be calculated by $P^r(Y_{itj} = 1|X_{itj}, y_{it-1j}, b_{it} = 0) = \Phi(\Delta^*_{itj})$. They are still subject/time/response specific probabilities, since Δ^*_{itj} holds subject/time/response specific information. Here, we refer this type of calculation of the conditional probabilities as Conditional* and the usual calculation of the conditional probabilities as Conditional. The marginal, and these two types of conditional probabilities for young people with ID's 223 and 359 alongside with the associated covariates are displayed in Table 4.7. Here, the person with ID=223 is a female with some negative life events, no economic events and some cutbacks (except at 1992) who actually never had any distress at the period of 1989-1992. On the other hand, the other person (ID=359) is a male with some negative life events, negative economic event only at 1989 and some cutbacks (except at 1992) who had all the three types of distress at every year. The empirical Bayesian estimates of the individual characteristics for these people were $\hat{z}_{223} = -1.69$ and $\hat{z}_{359} = 1.01$, respectively. While $\hat{z}_{223}$ indicates that the person tends to be less likely to report distress, $\hat{z}_{359}$ indicates that the person tends to be more likely to report distress. Actually, we can order the subjects by these empirical Bayesian estimates and detect the subjects who are most likely and less likely to report distress. For the person with ID=223 the marginal model estimated high probability of distress. However, the conditional model assigned low probability of having distress for the same person (Conditional). For instance, whereas at 1990 the probability of having depression for that person was estimated as 0.74 by the marginal model, this probability is calculated as 0.04 by the conditional model. Moreover, the conditional probability assuming that the person is an average person (Conditional*) was estimated as 0.46. Here, while the marginal model indicates a wrong decision, the conditional probabilities even the restricted one (Conditional*) indicate correct decision. For the person with ID=359, the marginal model indicate moderate probability of having distress, the conditional model indicate strong distress probabilities. For instance, at 1991 while the probability of depression was estimated as 0.60 by the former, the same probability was estimated as 0.94 by the conditional model. Additionally, the restricted conditional probability (Conditional*) was found to be 0.81.

Finally, we considered two different accuracy measures to compare the results of the first and last levels of baseline and $t \geq 2$ models of Model 1. Specifically, we considered Expected Proportion of Correct Prediction (ePCP; Herron, 1999) and Area Under the Receiver Operating

Table 4.7: Illustration of marginal and conditional probabilities for some specific people, ID=223 and 359.

ID	Time	Response	Gender	NLE	NEE	cut	Observed	Marginal	Conditional	Conditional*
223	1989	Anxiety	1	1	0	1	0	0.83	0.25	0.88
223	1989	Hostility	1	1	0	1	0	0.80	0.17	0.86
223	1989	Depression	1	1	0	1	0	0.91	0.41	0.93
223	1990	Anxiety	1	1	0	1	0	0.69	0.07	0.43
223	1990	Hostility	1	1	0	1	0	0.94	0.28	0.87
223	1990	Depression	1	1	0	1	0	0.74	0.04	0.46
223	1991	Anxiety	1	1	0	1	0	0.64	0.08	0.45
223	1991	Hostility	1	1	0	1	0	0.94	0.33	0.89
223	1991	Depression	1	1	0	1	0	0.71	0.07	0.54
223	1992	Anxiety	1	1	0	0	0	0.55	0.06	0.38
223	1992	Hostility	1	1	0	0	0	0.94	0.40	0.92
223	1992	Depression	1	1	0	0	0	0.66	0.06	0.51
359	1989	Anxiety	0	1	1	1	1	0.82	0.97	0.87
359	1989	Hostility	0	1	1	1	1	0.84	0.98	0.90
359	1989	Depression	0	1	1	1	1	0.95	1.00	0.98
359	1990	Anxiety	0	1	0	1	1	0.60	0.83	0.67
359	1990	Hostility	0	1	0	1	1	0.82	0.97	0.89
359	1990	Depression	0	1	0	1	1	0.63	0.86	0.67
359	1991	Anxiety	0	1	0	1	1	0.55	0.88	0.75
359	1991	Hostility	0	1	0	1	1	0.81	0.99	0.95
359	1991	Depression	0	1	0	1	1	0.60	0.94	0.81
359	1992	Anxiety	0	1	0	0	1	0.45	0.79	0.61
359	1992	Hostility	0	1	0	0	1	0.83	0.99	0.94
359	1992	Depression	0	1	0	0	1	0.55	0.89	0.73

Note: While Conditional corresponds to the random effects probabilities calculated by $\Phi(\hat{\Delta}^*_{itj} + \hat{\lambda}_j \hat{b}_{itj})$, Conditional* corresponds to random effects probabilities calculated by $\Phi(\hat{\Delta}^*_{itj})$. The empirical Bayesian estimates of individual characteristics are: $\hat{z}_{223} = -1.69$ and $\hat{z}_{359} = 1.01$, respectively.

Characteristics (AUROC) curve. ePCP considers the calculation of the average probability of estimating the actual observations, i.e. it considers $(1 - \hat{p}_i)$ for 0's and $\hat{p}_i$ for 1's where $\hat{p}_i$ are the estimated success probabilities. On the other hand, AUROC considers all the possible cut-off values (between 0 and 1) and dichotomizes the estimated success probabilities as 0 or 1 with respect to these possible cut-off values. Then, the area under the curve which is drawn by placing false positive rate (FPR) on the x-axis and true positive rate (TPR) on the y-axis is considered as the corresponding accuracy measure. Here, while TPR is calculated by the ratio of the number of cases which are assigned as positive (1, here) and was actually observed to be 1 to the total number of actual positives, FPR is calculated by the ratio of the number of cases which are assigned as positive (1, here) and was actually observed to be 0 to the total number of actual negatives. Note that these accuracy measures will be illustrated in details in Section 5.3 and AUROC measures were calculated by the `somers2` function under the R package **Hmisc** (Harrell et al., 2010). In the light of these measures, we concluded that the inference drawn from conditional models out performed the one drawn from the marginal models. This difference is apparent especially in terms of AUROC. For instance, while the value of the AUROC value for response=depression at 1990-1992 was found to be 0.684 for marginal models, this value was found to be 0.844 for the conditional models.

Table 4.8: EPCP and AUROC results for baseline and $t \geq 1990$ periods under Model 1.

1989				
	EPCP		AUROC	
Response	Marginal	Conditional	Marginal	Conditional
Anxiety	0.736	0.770	0.642	0.774
Hostility	0.732	0.769	0.612	0.800
Depression	0.877	0.886	0.618	0.802
1990-1992				
	EPCP		AUROC	
Response	Marginal	Conditional	Marginal	Conditional
Anxiety	0.589	0.679	0.638	0.787
Hostility	0.712	0.761	0.602	0.813
Depression	0.660	0.750	0.684	0.844

4.6 pnmtrem: An R Package for Probit Normal Marginalized Transition Random Effects Models

We proposed an R function to fit PNMTREM(1) called `pnmtrem1` which is placed under an R package **pnmtrem**. pnmtrem1 depends on the R package **MASS** and loads it automatically. Note that **MASS** is installed within R by default. The default function of `pnmtrem1` has the following form

```
R> pnmtrem1(covmat1,covmat2,respmat1,respmat2,z,nsubj,nresp,param01,param02,beta0,
+  alpha0,tol1=0.0001,tol2=0.0001,maxiter1=50,maxiter2=50,tun1=1,tun2=1,x01=0,
+  eps1=10^-10,x02=0,eps2=10^-10,silent=TRUE,delta.print=FALSE,deltastar.print=FALSE)
```

Before beginning to illustrate the details of the arguments of the `pnmtrem1` function, we shall note that the arguments which include the observed data, i.e. covariates or responses, will be illustrated under the assumption that there are N subjects in the longitudinal study for whom k different multiple binary responses are collected over T time points. Also, we shall note that the covariate sets included in the baseline and $t \geq 2$ models might be different, hence we denote the number of covariates included in these models with different notations, such as p_1

and p_2 for the baseline and $t \geq 2$ models, respectively. Moreover, the number of covariates included in the second level of the $t \geq 2$ model, i.e. in Z_{itj}, is denoted by p_3.

The argument `covmat1` is a $(p_1 + 1) \times N \times k$ matrix or data frame, which has the design matrix form, for the baseline time point ($t = 1$). The general form of `covmat1` is illustrated in Table 4.9. Here, X_{i1jl_1} corresponds to the l_1^{th} ($l_1 = 1, \ldots, p_1$) covariate for response j ($j = 1, \ldots, k$) of the subject i ($i = 1, \ldots, N$) at the first time point ($t = 1$).

Table 4.9: An illustration of covmat1 argument of pnmtrem1.

Ones	Cov_1	Cov_2	$\cdots$	Cov_{p_1}
1	X_{1111}	X_{1112}	$\cdots$	X_{111p_1}
1	X_{2111}	X_{2112}	$\cdots$	X_{211p_1}
.	.	.	.	.
1	X_{N111}	X_{N112}	$\cdots$	X_{N11p_1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	X_{11k1}	X_{11k2}	$\cdots$	X_{11kp_1}
1	X_{21k1}	X_{21k2}	$\cdots$	X_{11kp_1}
.	.	.	.	.
1	X_{N1k1}	X_{N1k2}	$\cdots$	X_{N1kp_1}

The argument `covmat2` is a $(p_2+1)\times N\times k\times(T-1)$ matrix or data frame, which has the design matrix form, for $t \geq 2$. The general form of `covmat2` is illustrated in Table 4.10. Here, X_{i1jl_2} corresponds to the l_2^{th} ($l_2 = 1, \ldots, p_1$) covariate for response j ($j = 1, \ldots, k$) of the subject i ($i = 1, \ldots, N$) at time t ($t = 2, \ldots, T$).

The argument `respmat1` is an $(N * k) \times 1$ matrix or data frame for the multiple responses at baseline. The general form of it can be depicted as $\text{Respmat}_1 = (Y_{.11}, \ldots, Y_{.1k})^T$ where $Y_{.1j} = (Y_{11j}, \ldots, Y_{N1j})$.

The argument `respmat2` is an $(N*k*T)\times 1$ matrix or data frame for the multiple responses for $t \geq 1$. The general form of it can be illustrated as $\text{Respmat}_2 = (Y_{.11}, \ldots, Y_{.1k}, \ldots, Y_{.T1}, \ldots, Y_{.Tk})^T$ where $Y_{.tj} = (Y_{1tj}, \ldots, Y_{Ntj})$.

The argument `z` is a $(p_3+1)\times N\times k\times(T-1)$ matrix or data frame to be included in the second level of the $t \geq 2$ model. `z` typically includes a subset of covariates and the general form of it could be found in Table 4.10. Note that while illustrating the `z` argument, p_3 replaces p_2 in Table 4.10.

The `nsubj` argument is an integer which defines the number of subjects in the study.

Table 4.10: An illustration of covmat2 argument of pnmtrem1.

Ones	Cov_1	Cov_2	$\cdots$	Cov_{p_2}
1	X_{1211}	X_{1212}	$\cdots$	X_{121p_2}
1	X_{2211}	X_{2212}	$\cdots$	X_{221p_2}
.	.	.	.	.
1	X_{N211}	X_{N212}	$\cdots$	X_{N21p_2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	X_{12k1}	X_{12k2}	$\cdots$	X_{12kp_2}
1	X_{22k1}	X_{22k2}	$\cdots$	X_{22kp_2}
.	.	.	.	.
1	X_{N2k1}	X_{N2k2}	$\cdots$	X_{N2kp_2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	X_{1T11}	X_{1T12}	$\cdots$	X_{1T1p_2}
1	X_{2T11}	X_{2T12}	$\cdots$	X_{2T1p_2}
.	.	.	.	.
1	X_{NT11}	X_{NT12}	$\cdots$	X_{NT1p_2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	X_{1Tk1}	X_{1Tk2}	$\cdots$	X_{1Tkp_2}
1	X_{2Tk1}	X_{2Tk2}	$\cdots$	X_{2Tkp_2}
.	.	.	.	.
1	X_{NTk1}	X_{NTk2}	$\cdots$	X_{NTkp_2}

The `nresp` is an integer which defines the number of multiple binary responses.

`param01` argument is a length of $[(p_1 + 1) + (k - 1) + 1]$ vector where p_1 is the number of covariates included in the baseline model and k is number of multiple responses. `param01` is used to start the Fisher-Scoring (FS) algorithm for the baseline model. The general form of it can be given as param01 = $(\beta^*, \lambda_j^*, c_1)$, where $j = 2, \ldots, k$.

`param02` argument is a length of $[(p_2+1)+(p_3+1)*(T-1)+(k-1)+(T-1)]$ vector where p_2 is the number of covariates included in the first level of $t \geq 2$ model, p_3 is the number of covariates included in the second level of it, k is the number of multiple responses and T is the number of repeated measurements per subject. `param02` is used to start the FS algorithm for the $t \geq 2$ model. The general form of it can be given as param02 = $(\beta, \alpha_{t,1}, \lambda_j, c_t)$, where $\alpha_{t,1} = (\alpha_{21,1}, \ldots, \alpha_{2p_3,1}, \ldots, \alpha_{T1,1}, \ldots, \alpha_{Tp_3,1})$ and $j = 2, \ldots, k$ and $t = 2, \ldots, T$.

`beta0` argument is a $(p_2 + 1) \times 1$ matrix for which all the elements are set to 0. It corresponds to the β_0 component of the implicit function theorem (IFT) point, P_0.

`alpha0` argument is a $(p_3 + 1) \times (T - 1)$ matrix for which all the elements are set to 0. It corresponds to the $\alpha_{t,10}$ component of the P_0.

`tol1` is the amount of tolerance for the convergence of the FS algorithm for baseline model. The default is set to 0.0001.

`tol2` is the amount of tolerance for the convergence of the FS algorithm for $t \geq 2$ model. The default is set to 0.0001.

`maxiter1` is the maximum number of iterations expected to be consumed by the FS algorithm for baseline model. The default is set to 50.

`maxiter2` is the maximum number of iterations expected to be consumed by the FS algorithm for $t \geq 2$ model. The default is set to 50.

`tun1` is the tuning parameter for baseline model need to be chosen preferably as integer to decrease the FS steps in each iteration in cases where the algorithm might miss the convergence of the parameters. The default is set to 1.

`tun2` is the tuning parameter for $t \geq 2$ model to decrease the FS steps in each iteration as in the case of `tun1`. The default is set to 1.

`x01` is an integer defined for the initial values of the Newton-Raphson (N-R) algorithm to obtain Δ_{i2j0}. The default is set to 0.

`eps1` is the amount of tolerance for the convergence of N-R algorithm to obtain Δ_{i2j0}. The default is set to 10^{-10}.

`x02` is an integer defined for the initial values of the Newton-Raphson (N-R) algorithm to obtain the empirical Bayesian estimates of the individual characteristics, $\hat{z}_i$. The default is set to 0.

`eps2` is the tolerance defined for the convergence of N-R algorithm to obtain $\hat{z}_i$. The default is set to 10^{-10}.

`silent` is a logical statement to decide whether the details of the FS algorithm details for both the baseline and $t \geq 2$ models to be printed. The default is set to TRUE which means not printing these details.

`delta.print` is a logical statement to decide the print of the estimates of Δ_{itj} where $t = 2,\ldots,T$ together with the modeling outputs. The default is set to FALSE which means not printing these estimates.

`deltastar.print` is a logical statement to decide the print of the estimates of Δ^*_{itj} where $t = 1,\ldots,T$ together with the modeling outputs. The default is set to FALSE which means not printing these estimates.

`pnmtrem1` prints the modeling output of baseline and $t \geq 2$ models and the associated maximized log-likelihood values. Additionally, it automatically prints the empirical Bayesian estimates of the individual characteristics, $\hat{z}_i$.

To illustrate the usage of `pnmtrem1`, let's consider Model 1 (Table 4.5). We achieved the related model fitting by using the following R script

```
R> z<-matrix(rep(1,451*3*3),ncol=1)
R> param01<-c(bsinit,1,1,log(0.5))
R> param02<-c(binit,alpinit,1,1,log(0.5),log(0.5),log(0.5))
R> beta0<-matrix(rep(0,length(binit)),ncol=1)
R> alpha0<-matrix(rep(0,length(alpinit)),ncol=1,byrow=T)
R> library("pnmtrem")
R> fit<-pnmtrem1(covmat1=covmat1.std,covmat2=covmat2.std,respmat1=resp1,
+  respmat2=respmat2,z=z,nsubj=451,nresp=3,param01=param01,param02=param02,
+  beta0=beta0,alpha0=alpha0,tol1=0.001,tol2=0.001,maxiter1=50,maxiter2=50,
+  tun1=20,tun2=30,x01=0,eps1=10^-10,x02=0,eps2=10^-10,silent=FALSE,
+  delta.print=TRUE,deltastar.print=TRUE)
```

Here, `bsinit` includes the estimates of the probit GLM presented in Table 4.4 and `binit` includes the estimates of the probit MMM2 presented in Table 4.5. `alpinit` are the estimates obtained by probit transition models which were achieved by the following R script

```
R> glm2<-glm(resp2[1:1353,]~resp1,family=binomial(link=probit))
R> glm3<-glm(resp2[1354:2706,]~resp2[1:1353,],family=binomial(link=probit))
R> glm4<-glm(resp2[2707:4059,]~resp2[1354:2706,],family=binomial(link=probit))
R> alpinit<-c(glm2$coef[2],glm3$coef[2],glm4$coef[2])
```

Note that the data arguments such as `covmat1.std`, `covmat2.std`, `resp1` and `resp2` are available from the authors. The related output is displayed by

```
R> fit
$title
[1] "FIRST ORDER PROBIT NORMAL MARGINALIZED TRANSITION RANDOM EFFECTS MODELS"

$version
[1] "Package version 1.0"

$date
[1] "Fri Mar 23 13:59:46 2012"

$title1
[1] "Baseline Model"

$output1
                Estimate Std. Error           Z      P value
betastar0    1.222094334 0.05232616 23.35532089 0.0000000000
betastar1    0.041411058 0.07784052  0.53199874 0.5947268688
betastar2    0.199715656 0.12670730  1.57619697 0.1149804159
betastar3    0.410140667 0.12716729  3.22520579 0.0012588215
betastar4    0.034147967 0.04687341  0.72851460 0.4662986343
betastar5    0.074133400 0.06473948  1.14510344 0.2521663089
betastar6   -0.005494449 0.06704500 -0.08195165 0.9346851610
betastar7    0.038616401 0.07341884  0.52597403 0.5989062531
betastar8    0.350529663 0.09050577  3.87300910 0.0001074998
betastar9   -0.063324386 0.08334349 -0.75980002 0.4473741289
betastar10  -0.125523953 0.09659743 -1.29945432 0.1937880587
lambdastar2  1.112870515 0.74048204  0.15242843 0.8788490345
lambdastar3  1.037759540 0.64761989  0.05830510 0.9535056089
logsigma1   -0.329508372 0.37584298          NA           NA

$maxloglik1
         [,1]
[1,] -207.888

$title2
[1] "t>=2 Model"

$output2
             Estimate Std. Error          Z      P value
beta0      0.77976966 0.02598416 30.0094273 0.000000e+00
```

```
beta1      0.12527429 0.04304348  2.9104128 3.609517e-03
beta2      0.12819470 0.04101522  3.1255398 1.774791e-03
beta3      0.31480731 0.04266281  7.3789634 1.594280e-13
beta4      0.03736849 0.02389118  1.5641120 1.177913e-01
beta5      0.04608468 0.02946247  1.5641828 1.177746e-01
beta6      0.03014254 0.03095078  0.9738863 3.301130e-01
beta7      0.30966349 0.05942483  5.2110117 1.878136e-07
beta8      0.03727770 0.06761410  0.5513303 5.814073e-01
beta9     -0.06207370 0.04760834 -1.3038409 1.922878e-01
beta10    -0.13656233 0.05615176 -2.4320223 1.501478e-02
beta11     0.14844838 0.03871817  3.8340754 1.260375e-04
beta12     0.02008421 0.05780979  0.3474189 7.282766e-01
beta13     0.02994182 0.04681878  0.6395258 5.224809e-01
beta14     0.12968889 0.04715932  2.7500161 5.959235e-03
beta15     0.02038692 0.04701754  0.4336023 6.645773e-01
beta16     0.05423579 0.05601189  0.9682907 3.328992e-01
alpha21    0.78409001 0.10465004  7.4924957 6.750156e-14
alpha31    0.97614706 0.09235593 10.5694030 0.000000e+00
alpha41    0.78531625 0.08810364  8.9135504 0.000000e+00
lambda2    1.30322677 0.48289520  0.6279350 5.300466e-01
lambda3    1.25686692 0.41519106  0.6186716 5.361327e-01
logsigma2 -0.67105829 0.24471726         NA           NA
logsigma3 -0.70174522 0.23622749         NA           NA
logsigma4 -0.68261675 0.23985242         NA           NA

$maxloglik2
          [,1]
[1,] -1327.256

$delta
[1] -0.0525453453  0.2392343987  0.2392343987 -0.3466904902  0.0122832756 . . .
.
.
.
[4057] -0.1398495843 -0.1398495843 -0.2153332720

$deltastar
[1]  1.035073373  1.553218608  1.553218608  1.035073373  1.364046485 . . .
.
.
```

```
.
[5410]  0.764635388  0.764635388  0.675215590

$empbayes
[1]  0.5070126096 -0.4727353922 -0.6191321520  0.2094997341 -0.6124262135 . . .
.
.
.
[449] -0.7372977293  0.1448241539 -0.3144903984
```

The desired components could be extracted by the $ symbol (as common in R software). For instance, the outputs for the baseline and $t \geq 2$ models could be obtained by the following script

```
R> fit$output1
R> fit$output2
```

4.7 Discussion and Conclusion

Multivariate longitudinal data introduce two main association structures: 1) within-subject association, 2) multivariate response association at a given time point. In this chapter, we proposed a marginalized multivariate model for multivariate longitudinal binary data, namely Probit Normal Marginalized Transition Random Effects Models (PNMTREM) which accommodates these two types of associations while modeling the covariate effects. Our model is an extension of MTREM proposed by Ilk and Daniels (2007) in terms of link functions and parameter estimation methodology. Since PNMTREM includes a transition structure and for $t = 1$ no history data is available, we constructed two different marginalized multivariate models for $t = 1$ and $t \geq 2$ periods and called these models as baseline and $t \geq 2$ models, respectively. Whereas baseline model is a two-level probit regressions model, $t \geq 2$ model is a three-level probit regressions model. These models are connected via the marginalization constraint introduced in the $t \geq 2$ model. Therefore, the features of data at $t = 1$ are reflected in the modeling of the data at $t \geq 2$ period. In the first levels of baseline and $t \geq 2$ models, we can build models with shared effects of the covariates on multiple responses as well as different effects of them on different responses. Similarly, in the second level of $t \geq 2$ model, we can build models with shared transition parameters for different multiple responses as well as we can allow these transition parameters separated for these responses. Moreover, in this level we could interact the effect of the past responses on the current ones by some certain covariates. We proposed the use of implicit function theorem to explicitly link the first and second level parameters of $t \geq 2$ model of PNMTREM. It is the first proposal in marginalized structured models to explicitly link the related marginal and transition probability levels. Specifically, we considered first order PNMTREM, PNMTREM(1) and conducted a simulation study to confirm the mathematical derivations and the written code and to examine the bias and variance of model parameters as well. The results confirmed these derivations and the related code and relieved us in terms of model parameters. Nonetheless, we observed some amount of over and underestimation for the parameters in the third level of $t \geq 2$ model. This is the case which is also valid for the original proposal, MTREM. Yalcinoz (2008) reported the same behavior for these parameters in a simulation study. We proposed an R function `pnmtrem1` for the PNMTREM(1) which is placed under an R package **pnmtrem**. Moreover, we provided the related details of the function in details and illustrated model output and the related manipulations. **pnmtrem** will be submitted to CRAN. We illustrated the validity of PNMTREM(1) in real life via an application to the IYFP dataset and discussed related parameter meaning and interpretations. Moreover, we illustrated the population-averaged and subject-specific inferences which could be drawn by PNMTREM at the same time.

A natural extension of our work here would be higher order PNMTREM, PNMTEM(p). Following Ilk and Daniels(2007) and Ilk (2008), the within-subject association structures of IYFP data could be better captured by a PNMTREM(2). Another possible extension of our work would be modeling missing data mechanisms as well. Specifically, missing not at random data (MNAR) models via sensitivity analyses could be studied. We only considered likelihood ratio test (LRT) for model selection purposes among the nested models. Since the random effects are parameters themselves, random effects models do not permit direct use of conventional Akaike Information Criterion (AIC). Instead of AIC, conditional Akaike Information Criterion (cAIC; Vaida and Blanchard, 2005) could be developed for random effects models with binary response data. Hypothesis tests regarding the random effects variance components could be developed, since the ones for variance components place a constraint on null hypothesis which states the equality of the component to its lower bound ($\sigma = 0$) and the test statistic follows some mixtures of the chi-square distributions with some certain weights (Molenberghs and Verbeke, 2007). The random effects variances in the second level of baseline model and in the third level of $t \geq 2$ model could be permitted to be modified by a subset of covariates, i.e. $\log(\sigma_t) = M_{itj}\,\omega_t$ where M_{itj} is a possible subset of covariates and ω_t are the related parameters. This might handle the over and underestimation issues observed in PNMTREM(1). Alternatively, to overcome this issue, the random effects might be assumed to have a multivariate normal distribution, i.e. $b_{it} \sim N(0, D)$ where D is a $T \times T$ matrix. In PNMTREM, we considered probit links for all the levels. However, following Diggle et al. (2002) and Griswold (2005), we can specify logit link for the first level of $t \geq 2$ model while keeping the others probit in terms of easing the interpretations of the population-averaged parameters. These are on-going works.

CHAPTER 5

FORECASTING

This chapter discusses the possible answers to the question: "Why one forecast?", especially with longitudinal data and the related methodologies to achieve this goal. It is divided into three sections. In Section 5.1, we motivate our study on forecasting longitudinal data together with the literature about time series and longitudinal data forecasting studies. In Section 5.2, we provide the forecasting methodologies which are specific to each of the models we mentioned in early chapters of this thesis, together with forecasting independent variables. Section 5.3 provides the accuracy measures which are necessary to confirm the validness of our forecast both for dependent (binary, here) and independent (continuous, here) variables.

5.1 Motivation

Forecasting might be regarded as the prediction of future events and it is one of the perspectives of statistical data analyses regarding time-dimensioned data such as time series and longitudinal data. It mainly relies on the quality of the available data at hand and the model built for that data which is thought to be the best one. For instance, if the data were collected from the randomized subjects and include no measurement and coding errors and no missing observation, the quality of forecast would be possibly high. Similarly, if a good model such as a simple one which fits well the data could be built and selected among the candidate models, then again obtaining high-quality forecast would be more likely. Furthermore, the main assumption of forecasting is that the system which evidenced by the available data continue for the forecasting period (Hyndman, 2010). Here, the system can be clarified as the serial dependence and variance structures of the responses, the relationships between the dependent and independent variables, from the longitudinal data perspective.

In forecasting, one of the key cautions might be constructing simple models in terms of the uncertainty of the model, for the forecasting period. For instance, although many covariates can be included in the model constructed for the available data, due to the fact that the effects of some of them might change in future and inclusion of these covariates' information might be unnecessary for the forecasting period, this might cause inefficient forecasting. The other key suggestion might be forecasting near future's data due to the fact that if the forecast time point diverge too much from the available data, the validity of the built model for that time point diverge from actual. Note that the data between that forecast time point and observed time points are forecast indeed. Diggle (1990, pp. 189) showed that the variance of the forecast is increasing as the forecast time point diverge from the last point of the observed data for simple time series linear regression model only including time as a covariate which might be generalized to forecasting with more complex model but the fact of increasing the variance of the forecast would still hold.

Forecasting is common in time series literature but it is rare in longitudinal studies (Baltagi, 2008). This might be due to the lengths of the histories of these two research areas, i.e. it is well-known that the former has a longer history and facing more modeling difficulties in the latter such as missing data. Moreover, in some time series studies, the main aim of data collection and model building is making forecast such as in applied macroeconomics and financial econometrics (Harris and Sollis, 2003, pp. 10). On the other hand, in longitudinal studies the key questions are about the effects of independent variables on the dependent ones for which the parameter estimation is achieved by taking into account the association structures of the repeated measurements. Nonetheless, forecasting with longitudinal data might be more informative and rich compared to time series forecasting since the former consists of both cross-sectional (includes more than one subjects/units) and time series data structures (the subjects/units are followed through time). Put another way, forecasting with longitudinal studies permits both prediction of future events and explaining these forecasts due to other factors such as independent variables and/or random effects. Here, it is beneficial to emphasize some of the main differences between time series data and longitudinal data; for the details see Section 1.1 of this thesis and the related references cited therein. Zeger et al. (2006) pointed out that while time series data is a realization of a single stochastic process which is defined as an infinite sequence of random variables, longitudinal data includes multiple and short series. They further stated that while the key sample size for the former is the length of the series, it is the number of subjects for the latter. Moreover, while in time series data models, the related parameter estimation is evidenced over one unit such as a country or a firm, in longitudinal data models, parameter estimation is evidenced over multiple subjects. Additionally, random effects parametrization in longitudinal models helps capturing subject-specific unobserved characteristics.

Actually, forecasting with time series data and related methods are beyond the scope of this thesis. However, we provided some information about these topics in comparison with the longitudinal data and methods to best illustrate our aims and the necessity of our studies. There is a wealth of literature about time series models and forecasting with time series data. Interested reader may refer to the following books: Diggle (1990), Wei (1990), Box et al. (1994), Chatfield (1996), Harris and Sollis (2003). Among these Diggle (1990) mainly concentrated on time series methods with a biostatical perspective. Another reference which mainly focuses on biostatical time series data is Zeger et al. (2006). In addition to these references, a short review about forecasting with biostatistical time series could be found in Chatfield (2005). On the other hand, the other references concentrate on more general topics. These are the references we benefited during our forecasting studies.

Forecasting might be life-saver and/or increase the quality of life, such as in medical and social studies. In Mother's Stress and Children's Morbidity study (MSCM; for details see Section 1.3.1), we are able to forecast the stress status of a mother (absence or presence) and the illness status of her child (absence or presence), since these variables are measured across time; daily in this study. Moreover, the effect of employment status of mother (unemployed or employed) could be incorporated while achieving forecasting of these variables by the help of longitudinal models. Note that for simplicity, we only considered mother's employment status among all independent variables. By the help of forecasting, necessary precautions might be taken. For instance, in the MSCM study by the forecasting of stress and morbidity statuses, the mothers and children who are more likely to be stressed or ill could be identified and their independent variables which have significant effect on stress/illness statuses could be controlled to decrease the risk of being stressed/ill. Therefore, while the infants would be precluded to have illness, mothers' quality of life would be increased. Similarly, forecasting the distress statuses of teenagers in Iowa Youth and Families Project (IYFP; for details see Section 1.3.2) via longitudinal models might be life-saver. It is well-known that high-level and constant distress might cause suicide. Understanding the factors that might cause distress together with predicting the future distress statuses of each young people would help saving their life. Note that due to the fact that IYFP dataset is a constrained longitudinal data, i.e. with irregular and common time points, we do not consider forecasting with this data in this study. However, in this thesis, we consider forecasting with fully constrained longitudinal data, i.e. with regular and common time points, such as MSCM dataset.

As emphasized earlier, forecasting is not common in longitudinal data literature. Most of the available studies are in the concept of econometric longitudinal data. A literature survey of such studies up to 2008 could be found in Baltagi (2008). Moreover, most of the available studies are for continuous response and all of them are for univariate response.

Baadsgaard et al. (2004) considered forecasting health statuses of 15 Danish pig herds for the period of January, 2001 - December, 2001 with a 12-month follow-up. They mainly considered three different veterinary clinical health outcomes, which are systematic signs of unhealthy conditions such as poor body condition or dullness, respiratory tract and diarrhoea. The scientific interest was on forecasting prevalence proportions of these health indicators for the aforementioned 15 different herds. Related methods included two main approaches: 1) a simple moving average method which assigns the average of past two observations as the current one, 2) a Bayesian state space model which captures the serial dependence structure via an AR-1 correlation assumption. This research concluded that these two approaches performed similarly in terms of forecasting the aforementioned clinical signs. It might be a preliminary and a motivational study to us in terms of being a medical one, since the aforementioned models are more similar to the ones for time series data, even the considered Bayesian state space model. Moreover, the data was not divided into model building and forecasting time periods. Instead, the researchers preferred using the observed data at time $t-1$ and t to forecast the one at time $t+1$ by their naive forecasting method, even if these former data points were actually forecasts. They further preferred building the state space model using all the available data. The forecasting via the state space model was achieved by a two-stage simulation algorithm in a Bayesian way. Additionally, the effects of independent variables on the aforementioned health variables and the association structure between these health outcomes were not taken into account.

The work of Frees and Miller (2004) concentrated on forecasting Wisconsin lottery sales. This study used a continuous longitudinal response in the econometric part of longitudinal data. Related data comprised lottery sales in 40 weeks (April, 1998 - January, 1999) from 50 different postal codes, located in Wisconsin. The models used to forecast the data included mainly the Laird-Ware random effects models (Laird and Ware, 1982) and their subject and time-wise modifications. They followed two main steps in this study: 1) model selection, 2) forecasting. Their main methodology to select the model which best explained the data included separating the 40-week data into model building (35 weeks) and validation (5 weeks) parts. Then, they fitted and validated different models via different accuracy measures and selected the best model. Results indicated that most complex models did not outperformed the less complex ones in terms of validation even though they outperformed the others in terms of model building. The actual forecasting procedure is somehow problematic such that they built their best model by utilizing the 40-week data and forecast the lottery sales of the period week 36-40 based on the parameter estimates of this model. Moreover, they did not consider forecasting the independent variables and the forecasting of time dependent parameters were not explained clearly. This procedure does not reflect the reality of forecasting.

Aslan (2010) considered a simulation study on forecasting univariate longitudinal binary data. Specifically, their simulation study was on comparison of the forecasting performances of 21 different methods which could be used to forecast such data. These methods include very simple ones such as moving averages, medians and modes and complex models such as marginal, transition, random effects and marginalized transition models (MTM; Heagerty, 2002b). They considered a model independent data simulation which permits the models to fairly compete. They further considered forecasting the covariates as well to better mimick the real life. Results indicated that random intercept models and transition models could be used to forecast univariate longitudinal binary data. On the other hand, the most complex model, MTM gave the worst results among all the models.

In this study, we consider forecasting *multivariate* longitudinal binary data in biostatistical perspective and we concentrate on forecasting performances of different models that might be used to forecast such data. Specifically, these models include univariate marginal models (UMM), multivariate marginal models with response specific regression parameters (MMM1), multivariate marginal models with shared regression parameters (MMM2), marginalized multivariate random effects models (MMREM; Lee et al., 2009) and probit normal marginalized transition random effects models (PNMTREM). Following Aslan (2010), we consider model independent data simulation and forecast the covariates. The forecasting performances of our models are illustrated via a real life dataset: Mother's Stress and Children's Morbidity (MSCM) data.

5.2 Forecasting Methodology

We consider forecasting multivariate longitudinal binary data with five different models, specifically Univariate Marginal Models (UMM), Multivariate Marginal Models with Response Specific and Shared Regression Parameters (MMM1 and MMM2, respectively), Marginalized Multivariate Random Effects Models (MMREM) and Probit-Normal Marginalized Transition Random Effects Models (PNMTREM). Each of these models have different characteristics, hence the related forecasting methodologies are different. For instance, while UMM, MMM1 and MMM2 do not include any time-varying parameters, MMREM and PNMTREM include time-varying parameters and require forecasting of these parameters in addition to forecasting of independent and dependent variables. Moreover, whereas UMM permits modeling the effects of different covariate sets on different responses, MMM2 and PNMTREM consider same set of covariates for multiple responses which might include some covariates taking values which are specific to each multiple responses (by response type indicator variables and related interactions by other independent variables) and the other models (MMM1 and MMREM) consider same set of covariates (which are identical for multiple responses) having different effects on different responses. In this section, we present the forecasting methodologies which are specific to each models. While illustrating these methodologies, we assume that there are N subjects for whom bivariate longitudinal binary data are collected over T time points. For the sake of simplicity, we consider bivariate data since in both of our simulation study and real life data application, there are bivariate longitudinal binary data. Nonetheless, extension of these methodologies to more than two responses is straightforward. We further assumed that we require forecasting these bivariate longitudinal data for the future m time points such as for the period of $((T+1), \ldots, (T+m))$. Following the general notation setup of this thesis, throughout this section i denotes the subject index where $i = 1, \ldots, N$; t denotes the time index where $t = 1, \ldots, T$ for the available data and $t = (T+1), \ldots, (T+m)$ for forecasting period; j is the response index where $j = 1, 2$ due to bivariate responses. Two new notations are r_1 and r_2, where while r_1 corresponds to the number of independent variables for which the effects are modeled on the first response, r_2 corresponds to the number of independent variables for which the effects are modeled on the second response. For some cases, these numbers of independent variables are same for multiple responses, hence $r_1 = r_2 = r$. Here, we shall note that the accuracy measures for checking the forecasts of dependent (binary) and independent (continuous) variables will be presented in the next section, Section 5.3.

Before beginning to illustrate the aforementioned forecasting methodologies, it is beneficial to depict forecasting longitudinal independent variables due to the fact that while illustrating these methodologies, we talk about complete design matrices for the forecasting period.

Forecasting Independent Variables

One of the major goals of our forecasting study on multivariate longitudinal binary response data is to forecast the binary response variables. In literature, the common methodology followed in forecasting studies is forecasting the future response variables by using the true values of the independent variables at the forecasting periods (for e.g. Baadsgaard et al., 2004 and Frees and Miller, 2004). However, in real life, the values of the independent variables are unknown as well as the dependent variables. Aslan (2010) considered forecasting the independent variables together with the dependent ones. In this study, to best mimic the real life, we consider that the independent variables are unknown in the forecasting period, $(T + 1), \ldots, (T + m)$, as well, following Aslan (2010).

As it was mentioned in Section 1.1, longitudinal independent variables might be changing by time (time varying) and/or might not be changing with time (time invariant). We do not need to forecast the future values of the latter type, since the values are constant across time. Nonetheless, the future values for the first type of independent variables are random variables indeed and need to be forecasted. However, forecasting is not needed for some time varying independent variables which are deterministic functions of time, such as age. The real life dataset considered in this thesis in the scope of forecasting, namely Mother's Stress and Child's Morbidity (MSCM, for details see Section 1.3.1) data do not include any time-varying independent variables. On the other hand, the dataset considered in the simulation study on forecasting (for details, see Section 7.1) include two time-varying and two time-invariant independent variables. Therefore, we only consider forecasting longitudinal independent variables in our simulation studies.

Methods relying only on the history of independent variables might be the best methods to forecast the longitudinal independent variables. Alternatively, methods which accommodate the effects of other independent variables might be used while forecasting an independent variable. Nonetheless, if these independent variables include any time-dependent variable, we need to have the future values of this variable. On the other hand, the effects of time-independent covariates (for which the future values are known, in advance) might be accommodated while forecasting an independent variable. However, this method might not be appropriate due to the fact that low correlation between independent variables is expected. Nevertheless, in real life, this option should be investigated carefully, since real life often surprises against such expectations. One of the common features of longitudinal data is that it include short series compared to time series data. For instance, in our simulation study, there are eight time points for which first four time points are for model building and last four are separated for forecast validation. Therefore, the stationarity and seasonality assumptions of times series data do not need to be considered in typical longitudinal data (including short series). Since in our simulated datasets, we assumed low correlation between independent variables and the dataset include four time points for model building period, we considered methods which only rely on the history of independent variables to be forecasted, such as first and second order transition models (TM(1) and TM(2), respectively) to forecast our independent variables. Aslan (2010) considered a similar longitudinal binary data simulation process for *univariate* longitudinal binary data, and considered TM(1), TM(2), and simple moving average which relies on assigning the averages of the observations at time t, $(t-1)$, $(t-2)$ and $(t-3)$ (since the related model building process include 4 time points) to forecast the value of $(t+1)$, and random intercept model which includes time, as a covariate. In that study, it was reported that TM(1) and TM(2) are superior in terms of forecasting independent variables, compared to simple moving average and random intercept methods. Following her work, we will also consider TM(1) and TM(2) in this thesis. Here, we shall note that she called these models as AR(1) and AR(2) in her thesis, but here we prefer to call them as TM(1) and TM(2), since we think that the latter names suit better to the related methods. The mathematical representations of TM(1) and TM(2) methods are given below.

First Order Transition Models

The modeling formulation of first order transition models, TM(1), could be illustrated by

$$X_{it} = \beta_0 + \beta_1 X_{it-1} + \epsilon_{it} \tag{5.1}$$

where $\epsilon_{it} \sim N(0, \sigma^2)$; X_{it} denotes an independent variable for subject i where $i = 1, \ldots, N$, at time t where $t = 2, \ldots, T$; X_{it-1} denotes the (same) independent variable for that subject at time $(t-1)$ and β_0 and β_1 are the related intercept and slope parameters, respectively.

Second Order Transition Models

The modeling formulation of second order transition models (TM(2)) could be illustrated by

$$X_{it} = \beta_0 + \beta_1 X_{it-1} + \beta_2 X_{it-2} + \epsilon_{it} \tag{5.2}$$

where $\epsilon_{it} \sim N(0, \sigma^2)$; X_{it} denotes an independent variable for subject i where $i = 1, \ldots, N$, at time t where $t = 3, \ldots, T$; X_{it-1} denotes the (same) independent variable for that subject at time $(t-1)$; X_{it-2} denotes the independent variable for the same subject at time $(t-2)$ and β_0, β_1 and β_2 are the related intercept and slope parameters, respectively.

The forecasting methodology for independent variables using TM(1) and TM(2), could be illustrated as below.

1. Estimate β based on the data which are collected in the period of $(1, \ldots, T)$, where $\beta = (\beta_0, \beta_1)$ for TM(1) and $\beta = (\beta_0, \beta_1, \beta_2)$ for TM(2).
2. Set $t = T + 1$, and put the estimates of β, $\hat{\beta}$, in (5.1, 5.2) for TM(1) and TM(2), respectively, and obtain the forecast values of X_{it+1}, $\hat{X}_{it+1}$. Here, $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ and $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ for TM(1) and TM(2), respectively.
3. Repeat step 2 by setting $t = t + 1$, until $t = T + m$.

Regardless we use TM(1) or TM(2) to forecast the independent variables, we would obtain a complete design matrix for the forecasting time period, $(T + 1), \ldots, (T + m)$. Note that the best method to forecast the independent longitudinal variables, among TM(1) and TM(2) will be decided by simulation studies and based on some accuracy measures. Results of these simulation studies will be presented in Section 7.2 and the accuracy measures will be presented in the next section, Section 5.3.

After completing the design matrix for the forecasting time period, we could step to the forecasting methodologies followed for multivariate longitudinal binary responses. These methodologies are presented below.

Univariate Marginal Models

Univariate marginal models (UMM) permit considering different independent variable sets on bivariate responses, i.e. X_{it1} and X_{it2} might be non-nested variable sets with differing dimensions, r_1 and r_2, where X_{it1} and X_{it2} are the independent variable sets for the first and second responses. For the details of UMM, interested reader may refer to Section 2.3 of this thesis and the references therein. The forecasting methodology with UMM could be illustrated in the following steps.

1. Obtain the estimates of β_1 and β_2, $(\hat{\beta}_1, \hat{\beta}_2)$, where β_1 and β_2 are $r_1 \times 1$ and $r_2 \times 1$ matrices belonging to first and the second responses, respectively.
2. Obtain the forecasts of the success probabilities, $\hat{p}_{it1} = \frac{exp(\hat{X}_{it1}\hat{\beta}_1)}{1+exp(\hat{X}_{it1}\hat{\beta}_1)}$ and $\hat{p}_{it2} = \frac{exp(\hat{X}_{it2}\hat{\beta}_2)}{1+exp(\hat{X}_{it2}\hat{\beta}_2)}$ where $\hat{X}_{it1}$ and $\hat{X}_{it2}$ are the forecasted independent variable sets for the first and second responses, respectively, and $t = (T + 1), \ldots, (T + m)$.

Multivariate Marginal Models with Response Specific Regression Parameters

Multivariate marginal models with response specific regression parameters (MMM1) assume that the sets of independent variables for different responses are identical, i.e. $X_{it1} \equiv X_{it2}$ and $r_1 = r_2 = r$, but their effects on these responses are different ($\beta_1 \neq \beta_2$). Note that MMM1 was illustrated in Chapter 3, in details. The related forecasting methodologies could be summarized in the following steps.

1. Obtain the estimates of β_1 and β_2, $(\hat{\beta}_1, \hat{\beta}_2)$, where β_1 and β_2 are both $r \times 1$ matrices belonging to first and the second responses, respectively.
2. Obtain the forecasts of the success probabilities, $\hat{p}_{it1} = \frac{exp(\hat{X}_{it}\hat{\beta}_1)}{1+exp(\hat{X}_{it}\hat{\beta}_1)}$ and $\hat{p}_{it2} = \frac{exp(\hat{X}_{it}\hat{\beta}_2)}{1+exp(\hat{X}_{it}\hat{\beta}_2)}$ where $\hat{X}_{it}$ are the forecasted independent variable sets for the first and second responses and $t = (T + 1), \ldots, (T + m)$.

Multivariate Marginal Models with Shared Regression Parameters

Multivariate marginal models with shared regression parameters (MMM2) assume a single set of independent variables for multiple responses where some of the covariates might take different values with respect to different responses such as response type indicator variable and related interactions, hence $X_{it1} \equiv X_{it2}$ or $X_{it1} \not\equiv X_{it2}$. On the other hand, MMM2 assumes that the effects of the independent variables are shared across multiple responses, i.e. $\beta_1 = \beta_2 = \beta$. Note that MMM2 was introduced in Chapter 3. The methodology followed while forecasting with MMM2 could be depicted by the following steps.

1. Obtain the estimates of β ($\hat{\beta}$), where β is an $r \times 1$ matrix.
2. Obtain the forecasts of the success probabilities, $\hat{p}_{it1} = \frac{exp(\hat{X}_{it1}\hat{\beta})}{1+exp(\hat{X}_{it1}\hat{\beta})}$ and $\hat{p}_{it2} = \frac{exp(\hat{X}_{it2}\hat{\beta})}{1+exp(\hat{X}_{it2}\hat{\beta})}$ where $\hat{X}_{it1}$ and $\hat{X}_{it2}$ are the forecasted independent variable sets for the first and second responses, respectively, and $t = (T+1), \ldots, (T+m)$

Marginalized Multivariate Random Effects Models

Likewise MMM1, marginalized multivariate random effects models (MMREM) assume that the sets of independent variables for different responses are identical, i.e. $X_{it1} \equiv X_{it2}$ and $r_1 = r_2 = r$, but their effects on these responses are different ($\beta_1 \neq \beta_2$). Details of MMREM could be found in Section 2.4. The related forecasting methodology could be presented below.

1. Obtain the estimates of β_1, β_2, α, s_{11}, s_{12}, s_{22} and z_i ($\hat{\beta}_1$,$\hat{\beta}_2$, $\hat{\alpha}$, $\hat{s}_{11}$, $\hat{s}_{12}$, $\hat{s}_{22}$, $\hat{z}_i$), where β_1 and β_2 are both $r \times 1$ matrices, z_i is a length of N vector and α, s_{11}, s_{12} and s_{22} are all single values.
2. Extend $\Sigma_1^{1/2}$ from a $T \times T$ matrix to a $(T+m) \times (T+m)$ matrix, based on the estimate of $\hat{\alpha}$.
3. Obtain $\Sigma^{1/2} = \Sigma_1^{1/2} \otimes \Sigma_2^{1/2}$ where $\otimes$ denotes Kronecker product.
(For the details of the symbols and operations presented in steps 1-3, see pages 37 and 38 of this thesis or the references cited therein)
4. Obtain the estimates of Δ_{itj}, $\hat{\Delta}_{itj}$, by solving the non-linear equation given in (2.25) via 40-points Gauss-Hermite Quadratures and Newton-Raphson algorithm in terms of Δ_{itj} based on $\hat{X}_{it}$, $\hat{\beta}_j$, $\hat{\alpha}$, $\hat{s}_{11}$, $\hat{s}_{12}$ and $\hat{s}_{22}$ where $\hat{X}_{it}$ are forecasted design matrix and $t = (T+1), \ldots, (T+m)$.
5. Obtain the success probabilities by using the re-parametrized second level of MMREM given in (2.26), i.e. $\hat{p}_{it1} = \frac{exp(\hat{\Delta}_{it1}+c^{2(t-1)+1}\hat{z}_i)}{1+exp(\hat{\Delta}_{it1}+c^{2(t-1)+1}\hat{z}_i)}$ and $\hat{p}_{it2} = \frac{exp(\hat{\Delta}_{it2}+c^{2(t-1)+2}\hat{z}_i)}{1+exp(\hat{\Delta}_{it2}+c^{2(t-1)+2}\hat{z}_i)}$ where $c^{k(t-1)+j}$ is the $(k(t-1)+j)^{th}$ row vector of $\Sigma^{1/2}$ for $j = 1, 2$.

An alternative method of forecasting with MMREM would be using only the columns $(T+1)$ to $(T+m)$ instead of the whole elements of the rows defined by $c^{2(t-1)+1}$, in step 6 of the above algorithm. The estimation of individual characteristics, z_i were not derived in the original article of MMREM (Lee et al., 2009) and were not readily available in the proposed R code by the authors. Therefore, we derived the Empirical Bayesian estimator of z_i and prepared the related R code. In addition to the forecasting method relies on estimation of z_i's illustrated above, we considered two alternative ways of forecasting with MMREM without estimating these parameters. First of these method could be illustrated below. Note that we continue from step 5 of the above algorithm.

6. Generate $K(K = 30, 50, 80, 100, 120, 150, 180, 200, 250, 300, 400, 500, 750, 1000)$ independent standard normal random variables, $\hat{z}_i$ for each subject and obtain K different success probability estimates which are calculated by $\hat{p}_{itj} = \frac{exp(\hat{\Delta}_{itj}+c^{k(t-1)+j}\hat{z}_i)}{1+exp(\hat{\Delta}_{itj}+c^{k(t-1)+j}\hat{z}_i)}$ where $T = 1, \ldots, (T + m)$.
7. Take mean or median of these K $\hat{p}_{itj}$. Exploratory analysis of randomly sampled subjects (say size of 10) would help on deciding whether the mean or median of these K probability estimates is to be taken.
8. Calculate the accuracy measures for each value of K.
9. Study the successive percentage changes of these replications, e.g. $K = 30$ vs. $K = 50$, ... $K = 750$ vs. $K = 1000$.
10. Select the amount of replication for which the successive percentage change with the next amount of replication is reasonably small, as the optimal amount of replication.

The second alternative forecasting method with MMREM would be using only the subject/time/response specific intercepts, Δ_{itj} while calculating the probability estimates such as $\hat{p}_{itj} = \frac{exp(\hat{\Delta}_{itj})}{1+exp(\hat{\Delta}_{it1})}$. This approach considers all the people are average ones such that all of them have $z_i = 0$. The related results would differ from the results to be obtained from marginal models, since Δ_{itj} contains subject/time/response information.

Related forecasting results will also be presented in Section 7.2.2 with application to MSCM data.

Probit Normal Marginalized Transition Random Effects Models

Likewise MMM2, probit normal marginalized transition random effects models (PNMTREM) assume a single set of independent variables for multiple responses where some of the covariates might take changing values with respect to different responses such as response type indicator variable and related interactions, i.e. $X_{it1} \equiv X_{it2}$ or $X_{it1} \not\equiv X_{it2}$, but their effects on these responses are different ($\beta_1 \neq \beta_2$). On the other hand, PNMTREM assumes that the effects of the independent variables are shared across multiple responses, i.e. $\beta_1 = \beta_2 = \beta$. Note that PNMTREM was introduced in Chapter 4. Here, we consider first order probit normal marginalized transition random effects models, PNMTREM(1). The related forecasting methodology could be illustrated in the following steps.

1. Obtain the estimates of β, $\alpha_{t,1}$, c_t, z_i ($\hat{\beta}$, $\hat{\alpha}_{t,1}$, $\hat{c}_t$, $\hat{z}_i$), where β is an $r \times 1$ matrix, $\alpha_{t,1}$ is an $(T-1) \times s$ matrix where s is the number of independent variables to interact with response history, c_t is a vector of length T and z_i is a vector of length N and $t = 2, \ldots, T$ for $\alpha_{t,1}$ and $t = 1, \ldots, T$ for c_t.
2. Obtain forecasts of $\alpha_{t,1}$ and c_t for $t = (T+1), \ldots, (T+m)$ by exponential smoothing methods for MSCM data (8 time points for model building) and simple moving averages method for the simulation study (4 time points for model building).
3. Obtain forecasts of A_{itj} and B_{itj} as given in (4.21), by taking the components of implicit function theorem (IFT) point, $P_0 = (\beta_0, \alpha_{t,0}, \Delta_{itj0})$, all equal to 0, based on $\hat{X}_{it-1j}$ and $\hat{X}_{itj}$ for $t = (T+1), \ldots, (T+m)$. Note that $\hat{X}_{it-1j}$ reduces to X_{iTj} at $t = T+1$ for $t = (T+1), \ldots, (T+m)$.
4. Obtain the estimates of Δ^*_{itj}, $\hat{\Delta}^*_{itj}$, as given in (4.27).
5. Set $t = T+1$ and obtain the forecasts of the success probabilities, $\hat{p}_{it1} = \Phi(\hat{\Delta}^*_{it1} + \lambda_1 \hat{\sigma}_t \hat{z}_i)$ and $\hat{p}_{it2} = \Phi(\hat{\Delta}^*_{it2} + \hat{\lambda}_2 \hat{\sigma}_t \hat{z}_i)$. Here, note that $\lambda_1 = 1$ by the model formulation of PNMTREM.
6. Dichotomize $\hat{p}_{it1}$ and $\hat{p}_{it2}$ by considering a classification rule as: classify $\hat{p}_{itj}$ as $\hat{y}_{itj} = 1$ if $\hat{p}_{itj} \geq cutoff_j$ and 0 otherwise for $j = 1, 2$, since the calculation of Δ^*_{itj} requires the binary past response (4.27).
7. Repeat steps 5 and 6 by setting $t = t+1$ until $t = T+m$.

For the discussion and details of exponential smoothing methods, interested reader may refer to Hyndman and Khandar (2008). This paper also illustrates the implementation of these methods and forecasting with them in R via using the R package **forecast** (Hyndman et al., 2012).

Similar to MMREM, we considered an alternative forecasting method for PNMTREM, in addition to the one illustrated in above steps. This alternative method assumes that all the people are average people, i.e. $z_i = 0$, and replaces the step 5 (of the above algorithm) as given by

5. Set $t = T+1$ and obtain the forecasts of the success probabilities, $\hat{p}_{it1} = \Phi(\hat{\Delta}^*_{it1})$ and $\hat{p}_{it2} = \Phi(\hat{\Delta}^*_{it2})$.

Our findings on MMREM showed that replication and assuming $z_i = 0$ give similar results. Therefore, we did not consider the replication method which has been illustrated for MMREM in terms of getting rid of computational burden.

Several *cutoff* values could be studied to see the effect of dichotomizing $\hat{p}_{itj}$ on forecasting performance of PNMTREM. We considered selecting the *cutoff* as the empirical proportions of 1's for different responses, obtained from model building portion of the dataset, selecting the *cutoff*=0.5 and simulating the responses from Bernoulli distribution with the predicted success probabilities, $\hat{p}_{itj}$. Additionally, we consider using the observed responses at $(T + 1), \ldots, (T + m)$ instead of dichotomizing $\hat{p}_{itj}$ as an alternative forecasting method with PNMTREM. Regarding all of these methodologies, we considered forecasting with the estimates of z_i, $\hat{z}_i$, and assuming $\hat{z}_i = 0$.

5.3 Accuracy Measures

One of the important aspects of forecasting is the validation of the forecast accuracy via several measures/indexes over multiple candidate forecast methods and/or models. This requires researchers to know the true forecast values. In this study, we are able to check the accuracy of our forecast, since we keep last four time points of our datasets for forecasting check purposes, in both simulation study and real life data application. Note that in this study, we both forecast the multiple binary responses and the independent variables (continuous, here) to best mimic the real life. Although there is a wealth of literature for accuracy measures for continuous data, there are relatively less such measures for binary data. We mainly calculate these accuracy measures to compare the forecasting performances of our models.

In conventional statistical modeling setup, the scientific interest is on accommodating the effects of some factors such as independent variables and/or random effects on the mean response, i.e. modeling the mean response conditional on some other factors. For a binary response (a variable only takes 0=failure or 1=success), this approach takes the form of modeling the success probability, p (the probability of being 1) which is the mean of a Bernoulli distribution, since a binary data follows such a distribution. Therefore, by a binary data model we can obtain the estimate of probability of being 1, $\hat{p}$ for a subject/unit and hence the probability of being 0, $1 - \hat{p}$ for the same subject (by the fact that these probabilities add up to 1). While the high success probabilities indicate higher chance of occurring towards 1, such low probabilities indicate high chance of occurring towards 0. For instance, let's consider stress status of a mother in the Mother's Stress and Children's Morbidity (MSCM, details could be found in Section 1.3.1) study, where this variable only takes 0=absence of stress and 1=presence of stress. In an imaginary modeling, a mother with $\hat{p} = 0.99$ is indicated that she would be stressed 99 out of 100 times and a mother with $\hat{p} = 0.01$ is indicated that she would be stressed only 1 out of 100 times. Additionally, a mother with $\hat{p} = 0.5$ is indicated that she would be stressed 50 out of 100 times, i.e. absence and presence of stress for that mother would occur equally likely.

One of the commonly considered accuracy measures in literature for binary data is the *proportion of correct prediction* (PCP). PCP mainly relies on the classification of predicted success probabilities, $\hat{p}$ such as classifying $\hat{p}$ to either $\hat{y} = 0$ or $\hat{y} = 1$ depending on a selected cut-off value, c. Here, $\hat{y}$ denotes the estimate of a binary response, y. This classification procedure can be illustrated as classifying $\hat{p}$ as $\hat{y} = 1$ if $\hat{p} \geq c$, classifying $\hat{p}$ as $\hat{y} = 0$ otherwise. When the classified and the observed responses are cross-tabulated, then there arise four possible combinations (Table 5.1). The related representation is also known as the *confusion matrix* (Fawcett, 2006). Let's denote y or $\hat{y} = 0$ as negative (N), and y or $\hat{y} = 1$ as positive (P), by following the literature. Here, whereas true negative (TN) and true positive (TP) correspond to correct classifications, false negative (FN) and false positive (FP) correspond to wrong classifications.

Table 5.1: An illustration of possible cases when the observed and the classified values are compared, cross-tabulated. Assuming 0: Negative (N) and 1: Positive (P); TN: True Negative, FP: False Positive, FN: False Negative, TP: True Positive.

		Classified	
		0	1
Observed	0	TN	FP
	1	FN	TP

PCP is then defined as the ratio of the correct decisions to the total decisions M, which is calculated by

$$PCP = \frac{TN + TP}{TN + FN + FP + TP} \tag{5.3}$$

Related $(1 - \alpha) \times 100\%$ (approximate) confidence interval (CI) could be calculated by

$$\begin{aligned} LB &= PCP - Z_{\frac{\alpha}{2}} \sqrt{\frac{PCP * (1 - PCP)}{M}} \\ UB &= PCP + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{PCP * (1 - PCP)}{M}} \end{aligned} \tag{5.4}$$

where LB, UB are the lower and upper bounds of the $(1 - \alpha) \times 100\%$ CI, respectively; Z is the quantile of the cumulative distribution function of standard normal distribution and M is the total number of observations to be forecasted. The maximum value which PCP can take is 1 and larger values of it indicate better performance.

Here we shall note that, M denotes the total number of observations and can be equal to the number of subjects in the study, N, if forecasting is considered for a single time point; and can be equal to $\sum_{i=1}^{N} m_i$, where m_i is the total number of repeated measures to be forecasted for subject i, if forecasting is considered for multiple time points. Note that for fixed number of maximum repeat measures for each subjects in the study, i.e. when $m_i = m$ for all i, the total number of observations to be forecasted is equal to $M = N * m$, for multiple forecast time points. These generalizations of M are valid for all the accuracy measures considered in this study.

There are two major drawbacks of PCP: 1) it largely depends on the choice of the cut-off value, c, 2) related calculation ignores the separate contributions of wrong decisions, FN and FP. Kutner et al. (2005, Section 14.10) discussed three possible ways to choose c. The first way is choosing c as 0.5 by which one assumes that failures (0's) and successes (1's) are equally likely to occur and the related incorrect classification costs are approximately same. The second way is investigating different c values and selecting the one which yields the lowest incorrect classification rate, which is calculated by $(1 - PCP)$, as the optimal one. This way relies on the facts that data is a random sample from the relevant population, i.e. reflects the population proportions of 0's and 1's, and the incorrect classification costs for 0's and 1's are approximately same. The third way is using prior information about the proportions and incorrect classification costs to choose the optimal c which might be impossible or incorrect in most of the real life problems. The authors reported that this option can be used for the cases in which data is not a random sample from the relevant population.

Regardless the method which is used to decide the optimal cut-off value, c, the calculation of PCP depends on the chosen c through classification of the predicted success probabilities, $\hat{p}$. For instance, suppose $c = 0.5$ and two mothers with success probabilities, $\hat{p}_1 = 0.49$ and $\hat{p}_2 = 0.51$ will be classified as not stressed ($\hat{y}_1 = 0$) and stressed ($\hat{y}_1 = 1$), although their predicted stress probabilities are very similar to each other. Put another way, the first mother is indicated that she would be stressed 49 out of 100 times, the other is indicated that she would be stressed 51 out of 100 times. Another typical illustration would be two imaginary mothers with success probabilities, $\hat{p}_1 = 0.51$ and $\hat{p}_2 = 0.99$ who will be classified as stressed ($\hat{y}_1 = \hat{y}_2 = 1$), although their predicted success probabilities are very different. In the first example, we would treat very similar probabilities as they indicate different response categories, and we would treat very different probabilities as they indicate same response categories. These two typical cases are valid for any value of c.

Another problematic case with PCP would be that if all the $\hat{p}$'s are classified either into $\hat{y}_1 = 0$ or $\hat{y}_1 = 1$, yet we might obtain very high PCP values. For instance, let's again consider the MSCM study in which the observed proportions of $y = 1$'s for the mother's stress variable is very low, approximately 0.13. If all the $\hat{p}$'s are classified into $\hat{y} = 0$ such as under $c = 0.5$, we would obtain a PCP of 0.87 even though all the $y = 1$'s are classified as wrongly as $\hat{y} = 0$. In this case, the true positive rate ($TPR = TP/(FN + TP)$), also known as *sensitivity*, of our classification would be 0, although the true negative rate ($TNR = TN/(TN + FP)$), also known as *specificity*, of this classification would be 1. This means all of the stressed mothers ($y = 1$) are classified into the category of not stressed ($\hat{y} = 0$), and all of the not stressed mothers ($y = 0$) are classified into the category of not stressed ($\hat{y} = 0$). In short, the high PCP value, 0.87 would mislead the researcher. Alternatively, instead of reporting the PCP alone, the *sensitivity* and *specificity* of the results based on the related classification should be investigated.

In literature, there are some other versions of PCP such as *proportional reduction in error* (PRE; Hagle and Glenn, 1990 and Herron, 1999) which can be calculated by

$$PRE = \frac{PCP - PMC}{1 - PMC} \tag{5.5}$$

where PMC is the proportion of observations in the modal category, i.e. proportion of 1's in the available data. Nonetheless, PRE depends on the choice of cut-off value, c through the calculation of PCP. The maximum value which PRE can take is 1 and larger values of it indicate better performance.

Another accuracy measure to check the performance of a model is the calculation of the area under the receiver operator characteristics ($AUROC$) curve. This measure also depends on the classification of the predicted success probabilities, $\hat{p}$ to either $\hat{y} = 0$ or $\hat{y} = 1$ by the aforementioned classification rule of classifying $\hat{p}$ as $\hat{y} = 1$ if $\hat{p} \geq c$, classifying $\hat{p}$ as $\hat{y} = 0$ otherwise. However, it relies on the choices of all the possible c values between 0 and 1. The procedure to calculate $AUROC$ could be illustrated in the following steps:

1. Select the cut-off value, c between 0 and 1
2. Classify $\hat{p}$ as $\hat{y} = 1$ if $\hat{p} \geq c$, classify $\hat{p}$ as $\hat{y} = 0$ otherwise
3. Calculate the *sensitivity* and *specificity* values
4. Repeat steps 1-3 for all the possible c values and save the related *sensitivity* and *specificity* values
5. Draw a line graph by putting 1-*specificity* on the x-axis and *sensitivity* on the y-axis which yields the receiver operator characteristics (*ROC*) curve
6. Calculate the area under the related curve

The maximum value which $AUROC$ can take is 1 and larger values of it indicate better performance. $AUROC$ is a better accuracy measure than both PCP and PRE, since instead of a single cut-off value, c it uses all the possible such values. Therefore, it represents the model performances in a more fair sense. Interested reader may refer to Section 6.2.6 of Agresti (2002), Section 14.10 of Kutner et al. (2005) and Fawcett (2006) for the details of AUROC.

An alternative and free of the cut-off value choice accuracy measure is the expected proportion of correct prediction ($ePCP$) proposed by Herron (1999). In its original proposition it was named as expected percentage of correctly predictions, but we re-named it here, since its proposal was based on the calculation of proportions rather than percentages. Besides, we consider the same measure with Herron (1999), here. Instead of classifying the predicted success probabilities, it uses the raw probabilities themselves. It mainly relies on the fact that a good model is expected to yield a high success probability, $\hat{p}$ for the observed response $y = 1$ and a low success probability, $\hat{p}$ for the observed response $y = 0$, and the related calculation considers the average of the predicted success probabilities, $\hat{p}$ for $y = 1$ and predicted failure probabilities $1 - \hat{p}$ for $y = 0$. Related calculation can be illustrated by

$$ePCP = \frac{1}{M}\left(\sum_{y_i=1}\hat{p}_i + \sum_{y_i=0}(1 - \hat{p}_i)\right) \tag{5.6}$$

where M is the total number of observations. Related $(1-\alpha)\times 100\%$ (approximate) confidence interval (CI) could be calculated by

$$\begin{aligned} LB &= ePCP - Z_{\frac{\alpha}{2}}\sqrt{\frac{ePCP * (1 - ePCP)}{M}} \\ UB &= ePCP + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{ePCP * (1 - ePCP)}{M}} \end{aligned} \tag{5.7}$$

where LB, UB are the lower and upper bounds of the $(1 - \alpha) \times 100\%$ CI, respectively; Z is the quantile of the cumulative distribution function of standard normal distribution. Note that Herron (1999, pp. 92) proposed an alternative way to calculate the CI's for $ePCP$ which includes simulating all the parameters in a model, from a normal distribution with mean $\hat{\theta}$ and variance-covariance matrix of $\hat{\theta}$ which is estimated from the available data, as S times, calculating S different $ePCP$ values and calculating the empirical quantiles of them. Here, θ is a vector containing all the model parameters and $\hat{\theta}$ is the estimate of θ. We replaced this method by the approximate confidence interval calculation, since it is computationally cumbersome, such as Herron (1999) considered $S = 5000$.

Since $ePCP$ do not use the classification of $\hat{p}$, it precludes classifying very close probabilities into different response categories such as $\hat{p} = 0.49$ and $\hat{p} = 0.51$ into $\hat{y} = 0$ and $\hat{y} = 1$, respectively; and classifying very different probabilities into same response category such as $\hat{p} = 0.51$ and $\hat{p} = 0.99$ into $\hat{y} = 1$, respectively (assuming the cut-off value, $c = 0.5$). The maximum value which $ePCP$ can take is 1 and larger values of it indicate better performance.

Another cut-off free accuracy measure is the likelihood function of Bernoulli distribution which utilizes the raw predicted probabilities, $\hat{p}$. Since the likelihood function is in the multiplication form and $\hat{p}$ range between 0 and 1, it possibly takes values which are close to 0. Usual approach is taking the natural logarithm of the likelihood function which makes it taking values between 0 and $-\infty$. Since all the aforementioned accuracy measures take positive values, we take minus sign of it and make it ranging between 0 and ∞. The calculation of negative log-likelihood (NLL) is given by

$$NLL = -\sum_{i=1}^{M} (y_i log(\hat{p}_i) + (1 - y_i) log(1 - \hat{p}_i)) \tag{5.8}$$

where M is the total number of observations. Lower values of NLL indicates better model performance. Likewise $ePCP$, NLL works only with the $\hat{p}$, hence it precludes classification of close $\hat{p}$ values to different response categories and different $\hat{p}$ values to same response category.

The information criterion, such as Akaike Information Criterion (AIC; Akaike, 1973; deLeeuw, 1992), Bayesian Information Criterion (BIC; Schwarz, 1978) and Information Complexity (ICOMP; Bozdogan, 1998) are all likelihood based. Though, they all calculate lack of fit as $-2 * loglikelihood$, they differ in terms of the penalization terms (For details, see the references). On the other hand, Deviance Information Criterion (DIC; Spiegelhalter, 2002) takes the number of effective parameters into account in a Bayesian fashion. Quasi Information Criterion (QIC; Pan, 2002) is proposed for marginal models with GEE as an alternative of AIC. In addition, the common property of those information criterion is that they are built to compare the same model with different features such as the case of omitting covariates from the same model.

In addition to the previous accuracy measures, these information criterion might be used to compare the forecasting performances of the models. Although the models considered in this study are applied on the same data sets, they have different parameter estimating procedures such as generalized estimating equations (GEE) for UMM, MMM1 and MMM2 and maximum likelihood estimation (MLE) for MMREM and PNMTREM. In addition, these models differ in terms of taking into account the association between the multivariate responses at a given time point, i.e. they have univariate (e.g. for UMM) and multivariate modeling structures (e.g. for MMM1, MMM2, MMREM, PNMTREM) and some models estimate common parameters for multivariate responses. For instance, MMREM estimates the correlation parameter, α, as common for the multivariate responses. On the other hand, some models have time specific parameters, e.g. Δ_{itj} and b_{itj} for MMREM, and Δ_{itj}, Δ^*_{itj} and b_{it} for PNMTREM; hence, one should forecast these parameters as well. The information criterion mentioned in the previous paragraph do not meet these extraordinary situations. Moreover, MMREM and

PNMTREM are indeed random effects structured models and the direct use of the usual information criterion (AIC, BIC, ICOMP and QIC), i.e. by counting the variance parameters of the random effects distribution as additional free parameters, for these models is not correct (Vaida and Blanchard, 2005).

As is mentioned in the beginning of this section, the literature on the accuracy measures for continuous data is richer compared to the one on such measures for binary data. Again, as is mentioned in the beginning of this section, in the conventional statistical modeling, the interest is on modeling the mean response conditional on some other factors such as independent variables and/or random effects. Unlike the binary data models in which the success probability is modeled and the fitted values are probabilities indeed, in the continuous data models the fitted values, $\hat{x}$, are quantities which are in the same scale with the observed responses, x. Note that since we consider continuous data modeling and forecasting only for our covariates in this study, we denote related variables by x instead of y. Therefore, there is no classification problem, hence no need to spend effort to select a best cut-off value, when one predicts a continuous response. The related accuracy measures mainly uses the difference of the observed and predicted values which is also known as prediction/forecast error such as $e = x - \hat{x}$ and they differ in terms of the used function of e to construct their own measures.

As is mentioned in the above paragraph, we are interested in forecasting continuous data only for our covariates. Since our real life data, MSCM data, do not include time-varying covariates, we only consider forecasting of covariates in our simulation study, which will be introduced in Chapter 7. In that simulation study, there are four covariates among which two of them are time-varying. These time-varying covariates are assumed to have normal distributions with means 0 for both of them and variances 2.5 and 25. We are interested in both the comparison of models in terms of forecasting a single covariate and the comparison of models in terms of multiple covariates such as model performances to forecast variables with low and large variances, 2.5 and 25, i.e. forecast accuracy between data. We believe that these little and early information about the covariates to be forecasted is mandatory in terms of better understanding the selection of the proper accuracy measures.

Hyndman and Koehler (2006) and Hyndman (2006) described the accuracy measures for continuous data under four different categories: 1) scale-dependent measures, 2) percentage-error measures, 3) relative-error measures and 4) scale-free error measures. Here, we mainly illustrate the ones which we used in our studies and give the related formulas, since they are too many to mention all here and are readily available in these references and the ones cited therein.

Scale-dependent measures heavily depend on data scale, hence they can be used only for method/model comparison for single data. The most common scale-dependent accuracy measures are: 1) mean squared error (MSE), 2) root mean squared error ($RMSE$), 3) mean absolute error (MAE) and 4) median absolute error ($MdAE$). Among these measures, only MSE is not in the same scale of data, but in the squared of it. On the other hand, MSE and $RMSE$ are more sensitive to outliers compared to the latter two measures, since they work with the squared forecast errors, e^2. We do not prefer MSE and $RMSE$ here, since they are harder to understand compared to the latter two measures and more sensitive to outliers. We prefer the use of MAE rather than $MdAE$, since while the former yields an averaged result, the latter picks only one forecast error value, the $50th$ forecast error quantile, which might mask the truth especially when the accuracy measure is calculated for the forecast of multiple time points. MAE can be calculated by

$$MAE = \frac{1}{M}\sum_{i=1}^{M}|e_i| \tag{5.9}$$

where $e_i = x_i - \hat{x}_i$ and M is the total number of observations to be forecasted.

Percentage-error measures are independent of data scale and can be used either for comparison of multiple methods on single data or forecast accuracy between data. These measures mainly use the percentage error which is defined by $e * 100/x$. There are two main disadvantages of these measures: 1) when the observed data, x takes 0, these measures take infinity, and they take very large values when x takes values close to 0, and 2) when both error and x take 0, these measures are undefined due to $0/0$. Relative-error measures are useful for comparison of naive and relatively complex methods. They mainly rely on the use of relative errors defined by $r = e/e^*$ where e^* is the forecast error of a naive method. The aforementioned two disadvantages of percentage-error measures are also valid for the relative-error measures, since 1) e^* might be equal to 0 or close to 0, and 2) both e and e^* might be 0. Since our covariates have normal distributions with means 0, these disadvantages are most likely to occur. Therefore, use of percentage-error and relative-error methods are not preferred in this study.

Scale-free error measures are also independent of data scale and can be used for comparison of multiple methods on single data and forecast accuracy between data. Hyndman and Koehler (2006) proposed mean absolute scaled error ($MASE$) as a scale-free error measure and showed that it is superior compared to the other scale-free methods. Original proposal of it can be found in Hyndman and Koehler (2006). Here, we mention about the modified form of it for longitudinal data forecasting, which can be calculated by

$$MASE = \frac{1}{M} \sum_{t=1}^{m_i} \sum_{i=1}^{N} \frac{|e_{it}|}{\frac{1}{n_i - 1} \sum_{h=2}^{n_i} |x_{ih} - x_{ih-1}|} \tag{5.10}$$

where m_i is the maximum number of forecast time points for subject i; N is the sample size; n_i is the maximum number of repeated measures for subject i at the model building period; e_{it} is the forecast error for subject i at the forecast time t and M is the total number of observations to be forecasted. Note that for equal number of repeated measurements for each subjects at the model building period, n_i is replaced by n, and for equal number of repeated measurements for each subjects at the forecasting period, m_i is replaced by m. For a single forecast time point, $MASE$ can be calculated by

$$MASE = \frac{1}{N} \sum_{i=1}^{N} \frac{|e_i|}{\frac{1}{n_i - 1} \sum_{h=2}^{n_i} |x_{ih} - x_{ih-1}|} \tag{5.11}$$

In (5.10, 5.11), the denominator expression, $\sum_{h=2}^{n_i} |x_{ih} - x_{ih-1}|$ could be read in two different fashions: 1) scaling term based on a simple method, last observation carried forward in longitudinal data context (Hyndman and Koehler (2006) defined it as one-step naive forecast in time series context), and 2) averaged difference to measure the change in successive measurements in the correlated (longitudinal) data context. When $MASE$ is less than 1, it gives better results compared to the naive method, last observation carried forward. Similarly, $MASE$ values grater than 1 indicate worse forecast compared to that naive method.

CHAPTER 6

FORECASTING APPLICATIONS ON MOTHER'S STRESS AND CHILDREN'S MORBIDITY STUDY

This chapter provides the results of modeling and forecasting for the models considered in this study, specifically Univariate Marginal Models (UMM), Multivariate Marginal Models with Response Specific and Shared Regression Parameters (MMM1 and MMM2, respectively), Marginalized Multivariate Random Effects Models (MMREM) and Probit Normal Marginalized Transition Random Effects Models (PNMTREM) on Mother's Stress and Children's Morbidity (MSCM) data. A brief overview of MSCM dataset is provided in Section 6.1. While the modeling and forecasting results are provided in Section 6.2, the discussion and conclusion regarding these results are provided in Section 6.3.

6.1 Brief Overview of Mother's Stress and Children's Morbidity Study

In this section, we provide a brief data description of Mother's Stress and Children's Morbidity (MSCM) study, since the study were presented in Section 1.3.1 in details. Also, an overview of MSCM data could be found in 3.4.1 and the related variable list and explanations could be found in Table 3.5.

In MSCM study, 167 mothers and their pre-school children (aged between 18 months-5 years) were enrolled with the main aim of investigating the relationship between mother's employment status and pediatric care usage. In a baseline interview, some demographic and family information were collected through the following variables: the marriage status (marriage), education level (education), employment status (employed) of mothers, children's race (race) and gender (gender), the health status of both mothers (mhlth) and children (chlth) at baseline and the size of the household (housize). Here, the words in parentheses are the abbreviations of the related variables which were used in Sections 1.3.1 and 3.4.1 and will be used throughout this chapter, as well. After the baseline interview, mothers were asked to keep the records of their stress status and their children's illness status in a 28-day health diary. These variables were dichotomized later as stress status of mothers (stress: 0=absence, 1=presence) and illness status of children (illness: 0=absence, 1=presence) at day t. Whereas the demographic and family variables did not include any missing observation, the health diary variables included very low percentages of such observations; 0.97% and 1.42% for stress and illness, respectively. These missing data were imputed by naive methods such as mode imputation due to two main reasons: 1) the related percentages are very low, and 2) data analysis under missing data is beyond the scope of this study.

Empirical investigation of the within-subject association structures of both mother's stress and child's illness variables suggested very weak serial dependencies in the period of day 1 to 16. Therefore, we only considered the period of day 17 to 28, for our forecasting studies. Nonetheless, to capture the individual characteristics of mothers and their children, we calculated the averages of mother's stress and child's illness statuses, for each mother and child for the period of day 1 to 16, and considered these as two new independent variables. These variables were abbreviated as bstress and billness, respectively. Standardized study time (week), which is calculated by (day-22)/7, were also included as a new covariate.

Since we added the averages of mother's stress and child's illness statuses at the period day 17 to 24 while we had mothers' health and children's illness statutes at baseline (gathered in a preliminary interview), we suspected about possible multicollinearity problems. We checked the existence of such problem by Variance Inflation Factor (VIF) via pooled logistic regression models by using the `vif` function under the R package **car** (Fox and Weisberg, 2011). Results showed that none of the VIF values are grater than 1.394 which indicate inclusion of these additional independent variables would not create multicollinearity problem.

As is mentioned in Section 5.3, forecasting accuracies should be checked by (possibly) several accuracy measures and this requires the researcher to know the true forecast values. Therefore, we divided the aforementioned portion of MSCM data (days 17 to 28) into two main parts: 1) model building (days 17 to 24), and 2) forecast validation (days 25 to 28) parts. Once the appropriate models were built for the former time period, forecasting were done based on the systems (effects of the independent variables), indicated by these models.

6.2 Results

6.2.1 Modeling

Univariate Marginal Models

The framework of univariate marginal models (UMM, for details see Section 2.3) considers separate modeling of the effects of independent variables on multiple responses, hence they ignore the multivariate response dependence at a given time point. Nonetheless, they take the within-subject association into account. On the other hand, they permit assuming different working correlation structures and considering different set of independent variables for different longitudinal responses.

During the model building process with UMM, we assumed different working correlation matrices such as *exchangeable*, *AR-1* and *unstructured* for both mother's stress and child's illness variables. Furthermore, different independent variable sets were considered for both of these responses. We used all of these model combinations (working correlation matrix and independent variable set combinations) during our forecasting studies. Since very similar forecasting results were obtained, here we only report the results of the models which give the best results in terms of model building and forecasting (Table 6.1). Here, we shall note that the UMM's were built by utilizing the `gee` function under the R package **gee** (Carey, 2011a).

Our UMM results for MSCM data are mostly in agreement with the results of Zeger and Liang (1986, Table 2, pp. 127) and Diggle et al. (2002, Table 12.2 in pp. 251). But note that whereas the former considered the period of days 1 to 9 and only the child's illness as a response, the latter considered the period of days 1 to 28 and modeled mother's stress and child's illness separately under independence working correlation assumption.

Multivariate Marginal Models with Response Specific Regression Parameters

Multivariate marginal models with response specific regression parameters (MMM1, for details see Chapter 3) consider joint modeling of multiple responses, i.e. they both allow capturing within-subject and multivariate response associations. The related parameter estimation

Table 6.1: Modeling results of UMM on MSCM data.

	Response=Stress			Response=Illness		
Parameter	Est.	SE	Z	Est.	SE	Z
Intercept	-1.86	0.49	-3.82	-1.20	0.50	-2.41
married				0.68	0.29	2.37
employed	-0.50	0.25	-1.99			
chlth	-0.37	0.15	-2.40	-0.42	0.18	-2.25
mhlth				-0.14	0.17	-0.80
housize	-0.10	0.30	-0.33	-0.81	0.31	-2.63
bstress	4.49	0.71	6.34			
billness				1.55	1.02	1.52
week	0.35	0.43	0.81	3.38	1.08	3.13
mhlth*week				-0.97	0.38	-2.55
housize*week	-0.98	0.57	-1.70	-1.31	0.76	-1.72
billness*week				-5.41	2.01	-2.68

Note: The results were obtained under exchangeable and AR-1 working correlation assumptions for response=stress and response=illness, respectively. Only robust standard error (SE) and Z (Z) estimates were reported.

relies on the use of a single working correlation matrix to accommodate the aforementioned associations. MMM1 framework further considers same set of independent variables but allows for different effects of them on multiple longitudinal responses.

Based on the modeling results of UMM, we built different MMM1's with different working correlation matrices such as *exchangeable* and *AR-1*. However, we could not achieve model fitting via *unstructured* working correlation matrix choice, since related model requires the estimation of 120 (=$\binom{8*2}{2}$, where 8 is the number of time points, and 2 is the number of multiple responses) different association parameters. We considered forecasting with both MMM1's under exchangeable and AR-1 working correlation matrices. Likewise UMM, similar forecasting results were obtained and here we preferred to report the modeling results of the MMM1 which gives best forecast results (Table 6.2). The models were fitted by using the `mmm` function under the R package **mmm** (Asar and Ilk, 2012).

Multivariate Marginal Models with Shared Regression Parameters

Multivariate marginal models with shared regression parameters (MMM2, details could be found in Chapter 3) also allow joint modeling of multiple longitudinal responses, i.e. accommodates both within-subject and multivariate response dependencies. The related modeling framework assumes that each covariate effects are same on different multiple responses.

Table 6.2: Modeling results of MMM1 on MSCM data.

	Response=Stress			Response=Illness		
Parameter	Est.	SE	Z	Est.	SE	Z
Intercept	-2.06	0.51	-4.05	-1.22	0.54	-2.27
married	-0.08	0.26	-0.31	0.71	0.30	2.36
employed	-0.44	0.25	-1.76	-0.26	0.31	-0.86
chlth	-0.32	0.16	-1.98	-0.35	0.18	-1.93
mhtlh	0.04	0.16	0.22	-0.16	0.18	-0.89
housize	-0.08	0.31	-0.27	-0.82	0.31	-2.68
bstress	4.39	0.79	5.54	0.36	0.91	0.40
billness	0.49	0.89	0.55	1.24	0.92	1.34
week	-0.15	0.78	-0.20	3.16	1.07	2.96
mhlth*week	0.27	0.34	0.78	-0.89	0.39	-2.30
housize*week	-0.92	0.58	-1.61	-1.25	0.77	-1.62
billness*week	0.58	1.71	0.34	-5.60	2.01	-2.78

Note: The results were obtained under exchangeable working correlation assumption. Only robust standard error (SE) and Z (Z) estimates were reported.

Nonetheless, the inclusion of response type indicator variable as a new covariate permits the multiple responses to have their own intercepts. Furthermore, the inclusions of interactions of covariates by response type indicator variable allow multiple responses to have their own slopes.

Likewise MMM1, MMM2 permits studying with several working correlation matrices. However, here we built the MMM2 based on the results of the MMM1 given in Table 6.2, i.e. *exchangeable* working correlation matrix. The results of MMM2 are presented in Table 6.3 where Model 1 is an equivalent model with the MMM1 for which the results are presented in Table 6.2. After fitting this equivalent model, a new MMM2 (Model 2 in Table 6.3) were built by omitting the insignificant interactions by response type indicator variables. Forecasting with MMM2 was achieved by using Model 2. Here, insignificant interactions by response type indicator variables indicate that the effects of the related covariates (such as employed, chlth, billness and housize*week) on stress and illness variables are not significantly different. The MMM2's were fitted by using the `mmm2` function under the R package **mmm2** (Asar and Ilk, 2012b) and response type indicator variable (rtype) takes 0 for response=stress, and takes 1 for response=illness by the default systematic of `mmm2` on creating response type indicator variables (see Chapter 3). Whereas UMM and MMM1 took less than a second to fit the models for MSCM data, MMM2 took a few seconds.

Marginalized Multivariate Random Effects Models

Table 6.3: Modeling results of MMM2 on MSCM data.

	Model 1			Model 2		
Parameter	Est.	SE	Z	Est.	SE	Z
Intercept	-2.06	0.51	-4.05	-1.93	0.46	-4.19
married	-0.08	0.26	-0.31	-0.06	0.26	-0.24
employed	-0.44	0.25	-1.76	-0.37	0.22	-1.65
chlth	-0.32	0.16	-1.98	-0.35	0.15	-2.44
mhlth	0.04	0.16	0.22	-0.06	0.13	-0.42
housize	-0.08	0.31	-0.27	-0.10	0.31	-0.31
bstress	4.39	0.79	5.54	4.20	0.78	5.38
billness	0.49	0.89	0.55	0.86	0.69	1.25
week	-0.15	0.78	-0.20	0.08	0.68	0.11
mhlth*week	0.27	0.34	0.78	0.12	0.29	0.44
housize*week	-0.92	0.58	-1.61	-1.07	0.50	-2.17
billness*week	0.58	1.71	0.34	0.91	1.44	0.63
rtype	0.84	0.57	1.46	0.57	0.34	1.69
married*rtype	0.79	0.35	2.24	0.82	0.34	2.38
employed*rtype	0.18	0.33	0.54			
chlth*rtype	-0.04	0.17	-0.20			
mhlth*rtype	-0.20	0.22	-0.89			
housize*rtype	-0.73	0.38	-1.95	-0.70	0.35	-2.01
bstress*rtype	-4.03	1.09	-3.68	-3.56	1.10	-3.23
billness*rtype	0.74	1.16	0.64			
week*rtype	3.32	1.30	2.56	2.87	0.91	3.16
mhlth*week*rtype	-1.16	0.52	-2.24	-0.88	0.39	-2.23
housize*week*rtype	-0.33	0.94	-0.35			
billness*week*rtype	-6.18	2.62	-2.36	-6.86	1.91	-3.60

Note: The results were obtained under exchangeable working correlation assumption. Only robust standard error (SE) and Z (Z) estimates were reported. Response type indicator variable (rtype) takes 0 for response=stress, and takes 1 for response=illness.

Marginalized multivariate random effects models (MMREM; Lee et al., 2009) permit joint modeling of multiple longitudinal binary responses, hence accommodate both within-subject and multivariate response dependencies. The framework includes two different logistic regression models for marginal effects of the independent variables on mean response and relies on different covariate effects on different responses in the first level. On the other hand, the second level includes a correlation and a variance-covariance matrices for serial dependence and multivariate response dependence, respectively as formers of the variance-covariance matrix of the random effects distribution. While the former matrix have an AR-1 structure, i.e. includes only one unknown correlation parameter, the latter includes $(k \times (k+1)/2)$ covariance parameters, where k is the number of multiple responses. Interested reader may refer to either Section 2.4 of this thesis or Lee et al. (2009) for the details.

Based on the results of UMM and MMM's (MMM1 & MMM2), we considered a set of independent variables and built an MMREM for the MSCM data, Model 1 in Table 6.4. An alternative MMREM, which excludes the main effect of mother's health at baseline (mhlth) and its interaction with time (mhlth*week) from Model 1, were built with the aim of fitting four less parameters. The results of this new MMREM could be found under Model 2, in Table 6.4. Since Model 1 and Model 2 are nested models and the parameter estimation of MMREM is achieved via maximum likelihood estimation (MLE), likelihood ratio test (LRT) could be used to compare these models. The LRT statistics was found to be $9.41 = (-2 * (-807.1623 - (-802.456)))$ which is to be compared with the theoretical distribution quantile, ($\chi^2_{4,0.95} = 9.49$) and the related $p - value$ was found to be 0.052. LRT indicated that there is almost enough evidence to conclude that Model 1 explains the MSCM data better than Model 2. Therefore, we considered Model 1, while forecasting MSCM data with MMREM. We calculated the computational times consumed by these models: while Model 1 took 7.1 minutes for model fitting, Model 2 took 8.6 minutes for model fitting, on a PC with 4.00 GB RAM and 2.53 GHz processor.

Since almost all of the covariates were binary variables, we had some difficulties in terms of convergence of the model fitting algorithm (Fisher-Scoring Algorithm). To get rid of this issue, we applied variable based standardization to all of the independent variables. Therefore, the MMREM modeling results on MSCM data might not be comparable with the ones of UMM and MMM's, since the latter ones were built on binary independent variables. However, the incomparability of parameter estimates is not a problem, since our goal is comparing models in terms of their forecasting abilities. MMREM was fitted by using the R function `findmle` together with the related FORTRAN .dll files which are available from the authors of the original article (Lee et al., 2009).

Probit Normal Marginalized Transition Random Effects Models

Probit normal marginalized transition random effects models (PNMTREM) assume two different models for baseline time ($t = 1$) and later time points ($t \geq 2$), i.e. baseline and $t \geq 2$ models. This is due to the fact that at $t = 1$, we do not have history of data. Each of these models have marginalized modeling structures and while the former includes two probit regression models, the latter includes three probit regression models. They both accommodate the marginal covariate effects on mean response in the first levels and similar to MMM2, the

Table 6.4: Modeling results of MMREM on MSCM data.

MODEL 1	Response=Stress			Response=Illness			Association Parameters			
Parameter	Est.	SE	Z	Est.	SE	Z	Parameter	Est.	SE	Z
Intercept	-2.19	0.16	-13.42	-2.17	0.25	-8.85	s_{11}	2.07	0.40	
married	-0.07	0.15	-0.44	0.33	0.16	2.00	s_{12}	1.66	0.63	
employed	-0.25	0.16	-1.55	-0.11	0.15	-0.71	s_{22}	4.56	1.51	
chlth	-0.29	0.18	-1.66	-0.40	0.17	-2.31	α	0.28	0.05	5.46
mhtlh	0.06	0.19	0.30	0.16	0.17	0.94				
housize	0.15	0.15	1.01	-0.32	0.16	-1.97				
bstress	0.69	0.15	4.54	0.13	0.18	0.73				
billness	0.04	0.15	0.24	0.28	0.18	1.52				
week	-0.09	0.11	-0.75	0.06	0.11	0.57				
mhlth*week	0.07	0.11	0.63	-0.15	0.09	-1.63				
housize*week	-0.16	0.12	-1.35	-0.23	0.09	-2.72				
billness*week	0.08	0.13	0.61	-0.23	0.07	-3.16				
Max.loglik.	-802.46									
MODEL 2	Response=Stress			Response=Illness			Association Parameters			
Parameter	Est.	SE	Z	Est.	SE	Z	Parameter	Est.	SE	Z
Intercept	-2.08	0.14	-14.42	-2.07	0.23	-8.96	s_{11}	1.94	0.40	
married	-0.03	0.13	-0.24	0.34	0.16	2.20	s_{12}	1.06	0.52	
employed	-0.20	0.15	-1.40	-0.03	0.14	-0.20	s_{22}	4.84	1.50	
chlth	-0.26	0.13	-2.07	-0.25	0.15	-1.68	α	0.32	0.06	5.45
housize	0.12	0.13	0.87	-0.30	0.15	-2.03				
bstress	0.60	0.12	5.21	0.12	0.16	0.75				
billness	0.02	0.13	0.14	0.26	0.15	1.69				
week	-0.06	0.11	-0.57	0.07	0.11	0.59				
housize*week	-0.13	0.12	-1.15	-0.25	0.08	-3.01				
billness*week	0.06	0.12	0.49	-0.26	0.07	-3.44				
Max.loglik.	-807.16									

covariate effects are initially assumed to be shared across multiple responses, and including response type indicator variables and related interactions by covariates permits building models with response specific intercepts and slopes, respectively. Within and between response associations are captured in other levels of these models. For instance, multivariate response dependence at $t = 1$ is captured in the second level of baseline model. Moreover, whereas the serial dependence of the responses is captured in the second level of $t \geq 2$ model, the multivariate response dependence for $t \geq 2$ is captured in the third level of this model. These models are indeed connected to each other via a marginal constraint equation (4.10). For the details of PNMTREM, see Chapter 4 of this thesis.

We specifically consider first order probit normal marginalized transition random effects models (PNMTREM(1)). To build PNMTREM(1)'s with appropriate set of independent variables, we investigated the MSCM data at the period of days 17 to 24 by dividing into two parts: 1) day 17, and 2) days 18 to 24. Several MMM2's were built for these time periods (results are not shown here) and the appropriate set of independent variables were decided. After deciding on the independent variable sets for the baseline and $t \geq 2$ models, we build two different PNMTREM(1). While the first one (Model 1 in Table 6.6) assumes that the effects of lag-1 responses are shared across response=stress and response=illness, i.e. $Z = [\ 1\ ]$, the second one (Model 2 in Table 6.6) permits separating these effects with respect to different responses, i.e. $Z = [\ 1\ rtype\ ]$, where rtype corresponds to response type indicator variable. Since these models are nested ones, we can compare them via Likelihood Ratio Test (LRT). This comparison yields the LRT statistic to be 30.18 ($= -2 * ((-10.49 - 58.01) - (-10.49 - 42.92))$) which is to be compared with $\chi^2_{7,0.95} = 14.07$. The LRT for comparison of Model 1 vs. Model 2 indicates that Model 2 explains the MSCM data better at 95% confidence level. Therefore, we consider Model 2 to forecast MSCM data. The were fitted by the `pnmtrem1` function under the R package **pnmtrem**. We calculated computational times consumed by these models: while Model 1 took 27.4 minutes to converge, Model 2 took 43.3 minutes. Note that the models were fitted on a PC with 4.00 GB RAM and 2.53 GHz processor.

Here, we shall note that similar to MMREM, we faced convergence problems due to binary covariates. To get rid of this issue, variable based standardization was applied. Due to standardization of the covariates, assuming two different models for $t = 1$ and $t \geq 2$ and using probit link, the modeling results of PNMTREM(1) might not be directly comparable with the previous models, UMM, MMM1, MMM2 and MMREM.

Table 6.5: Modeling results of PNMTREM(1) on MSCM data at day=17.

Marginal					Association				
Parameter	Est.	SE	Z	P	Parameter	Est.	SE	Z	P
Intercept	-1.37	0.12	-11.18	0.00	$\lambda^*_{illness}$	-1.48	0.45	-5.55	0.00
employed	-0.07	0.11	-0.63	0.53	$\log(\sigma_{17})$	-1.23	1.39		
chlth	-0.29	0.12	-2.38	0.02					
mhtlh	0.22	0.15	1.49	0.14					
housize	0.11	0.11	1.00	0.32					
bstress	0.06	0.13	0.48	0.63					
billness	0.24	0.11	2.20	0.03					
rtype	-0.13	0.12	-1.09	0.28					
employed*rtype	0.21	0.11	1.94	0.05					
mhlth*rtype	0.25	0.13	1.91	0.06					
bstress*rtype	-0.20	0.15	-1.37	0.17					
billness*rtype	0.32	0.12	2.65	0.01					
Max.loglik.	-10.49								

Note: Since baseline models of Model 1 and Model 2 are identical, we report the baseline results only once. $\lambda^*_{illness}$ is tested against 1, the others are tested against 0. rtype takes -0.9999 for response=stress, and 0.9999 for response=illness after standardization.

The baseline results are provided in in Table 6.5. Since the baseline models of Models 1 & 2 (Table 6.6) are identical, we only reported these baseline results once. In Table 6.5, $\hat{\lambda}^*_{illness} = -1.48$ means that there is a negative correlation between mother's stress and children's illness at day 17. For an approximate functional form of this correlation, see (4.4). Empirical correlation analysis, of these responses at day 17 supports the results, for e.g. the estimate of the Spearman Rank Correlation was found to be -0.03. The estimate of the random effects variance at day 17, $\hat{\sigma}_{17}$ was found to be 0.29 (= $exp(-1.23)$).

The results for $t \geq 2$ modeling results are provided in Table 6.6. The estimates of the variances of the random effects distributions were found to be: $(\hat{\sigma}_{18}, \hat{\sigma}_{19}, \hat{\sigma}_{20}, \hat{\sigma}_{21}, \hat{\sigma}_{22}, \hat{\sigma}_{23}, \hat{\sigma}_{24}) = (0.47, 0.45, 0.53, 0.50, 0.53)$. The random effects variances seem to be higher compared to the one in day 17. The correlation between mother's stress and children's illness seem to be positive in the period of day 18 to day 24, i.e. $\hat{\lambda}_{illness} = 0.92$. Note that this correlation was found to be negative at day 17.

6.2.2 Forecasting

The forecasting results of the MSCM data based on the models mentioned in Section 6.2.1 are provided in Tables 6.9 & 6.10. Before beginning to interpret the forecasting results pre-

Table 6.6: Modeling results of PNMTREM(1) on MSCM data at day 18 to 24.

	Model 1				Model 2			
Parameter	Est.	SE	Z	P	Est.	SE	Z	P
Intercept	-1.25	0.06	-22.20	0.00	-1.24	0.06	-20.92	0.00
married	0.06	0.04	1.62	0.11	0.06	0.04	1.45	0.15
employed	-0.09	0.05	-1.75	0.08	-0.09	0.05	-1.74	0.08
chlth	-0.12	0.05	-2.69	0.01	-0.12	0.05	-2.65	0.01
mhlth	-0.04	0.06	-0.69	0.49	-0.04	0.06	-0.61	0.54
housize	-0.10	0.06	-1.85	0.07	-0.11	0.06	-1.83	0.07
bstress	0.21	0.05	4.14	0.00	0.20	0.05	3.84	0.00
billness	0.06	0.06	1.00	0.32	0.06	0.06	0.90	0.37
week	0.003	0.08	0.04	0.97	0.005	0.09	0.06	0.95
mhlth*week	-0.01	0.08	-0.16	0.87	-0.01	0.09	-0.11	0.92
housize*week	-0.10	0.08	-1.28	0.20	-0.11	0.09	-1.22	0.22
billness*week	-0.12	0.09	-1.35	0.18	-0.11	0.09	-1.18	0.24
rtype	-0.01	0.05	-0.10	0.92	0.003	0.05	0.06	0.95
married*rtype	0.10	0.04	2.22	0.03	0.09	0.05	1.98	0.05
housize*rtype	-0.08	0.04	-1.86	0.06	-0.08	0.05	-1.79	0.07
bstress*rtype	-0.14	0.04	-3.11	0.00	-0.14	0.04	-3.07	0.00
mhlth*week*rtype	-0.04	0.08	-0.48	0.63	-0.03	0.08	-0.35	0.73
billness*week*rtype	-0.09	0.07	-1.20	0.23	-0.09	0.08	-1.06	0.29
$\alpha_{18,1;1}$	1.25	0.27	4.69	0.00	1.30	0.31	4.26	0.00
$\alpha_{18,2;1}$					0.62	0.30	2.04	0.04
$\alpha_{19,1;1}$	1.33	0.28	4.76	0.00	1.36	0.30	4.57	0.00
$\alpha_{19,2;1}$					0.32	0.24	1.35	0.18
$\alpha_{20,1;1}$	1.45	0.21	6.96	0.00	1.57	0.21	7.48	0.00
$\alpha_{20,2;1}$					0.70	0.19	3.69	0.00
$\alpha_{21,1;1}$	1.44	0.26	5.62	0.00	1.54	0.27	5.72	0.00
$\alpha_{21,2;1}$					0.46	0.25	1.83	0.07
$\alpha_{22,1;1}$	0.77	0.23	3.39	0.00	0.74	0.23	3.16	0.00
$\alpha_{22,2;1}$					-0.21	0.27	-0.80	0.43
$\alpha_{23,1;1}$	1.42	0.23	6.26	0.00	1.42	0.24	5.96	0.00
$\alpha_{23,2;1}$					0.35	0.23	1.55	0.12
$\alpha_{24,1;1}$	1.30	0.25	5.15	0.00	1.27	0.26	4.98	0.00
$\alpha_{24,2;1}$					-0.10	0.20	-0.47	0.64
$\lambda_{illness}$	0.92	0.51	-0.15	0.88	0.91	0.59	-0.15	0.88
$\log(\sigma_{18})$	-0.76	1.20			-0.80	1.43		
$\log(\sigma_{19})$	-0.79	1.22			-0.91	1.59		
$\log(\sigma_{20})$	-0.64	0.53			-0.61	0.55		
$\log(\sigma_{21})$	-0.85	1.30			-0.95	1.64		
$\log(\sigma_{22})$	-0.64	0.71			-0.60	0.70		
$\log(\sigma_{23})$	-0.70	0.82			-0.61	0.85		
$\log(\sigma_{24})$	-0.63	0.67			-0.55	0.71		
Max.loglik.	-58.01				-42.92			

$\lambda_{illness}$ is tested against 1, the others are tested against 0. rtype takes -0.9999 for response=stress, and 0.9999 for response=illness, after standardization.

sented in these tables, we need to mention a few things about forecasting with MMREM and PNMTREM to clarify the methods and the contents of these tables.

As is mentioned in Section 5.2, we considered an alternative methodology for forecasting with MMREM. This methodology includes generating the individual characteristics, z_i, from independent standard normal distributions K ($K = 30, 50, 80, 100, 120, 150, 180, 200, 250, 300, 400, 500, 750, 1000$) times and calculating K different individual success probabilities by using the re-parametrized conditional level of MMREM given in (2.26). Exploratory analysis such as individual histograms (not shown here) of a subset of subjects for these amount of replications showed that these individual specific multiple estimates have a right-skewed distribution, hence we decided taking the median of K success probabilities as individual forecasted success probability. Percentage changes (Table 6.7) in the accuracy measures directed us selecting $K = 150$ for response=stress and $K = 250$ for response=illness. We believe that finding K for response=illness relatively larger than the one of response=stress is most probably related to the different individual variations with respect to these responses. For instance, the modeling results of MMREM (Table 6.4) showed that children reflect more variation in terms of their illness statuses compared to mothers in terms of their stress statuses; related variance parameter estimates were found to be $s_{11} = 2.07$ and $s_{22} = 4.56$, based on Model 1 (Table 6.4).

The percentage changes, presented in Table 6.7, for ePCP and the related 95% confidence bounds were seem to be very low for both response=stress and response=illness, for all the time points. This is true for smaller values of K (not shown here) as well. While the highest changes for AUROC and NLL of response=stress were seen at day=28 as 3.18% , -5.23% , respectively; the highest changes for these measures of response=illness were seen at day=28 and day=26 as 4.41% and -13.34% , respectively. Nonetheless, these changes are reasonably small. For instance, -13.34% change in NLL corresponds to a decrease of amount -10.72 ($= 80.36 - 91.08$) where 80.36 is the value of NLL at day=26 for $K = 250$ and 91.08 is the value of NLL at this time point for $K = 300$. Similarly, 4.41% change in AUROC corresponds to an increase of amount 0.03 ($= 0.68 - 0.65$) where 0.68 is the value of AUROC for $K = 250$ and 0.65 is the value of AUROC for $K = 300$. The forecasting results of $K = 150$ for response=stress and $K = 250$ for response=illness are denoted by MMREM3 in Tables 6.9 & 6.10, to distinguish them from the other forecasting methods with MMREM. Note that whereas the forecasting method which relies on using the estimates of the individual

Table 6.7: Percentage changes for the alternative forecasting method with MMREM.

response=stress					
Day	ePCP	ePCP.l	ePCP.u	AUROC	NLL
17 to 24	0.42	0.45	0.39	-0.68	1.07
25	0.23	0.28	0.18	1.05	0.89
26	-0.17	-0.21	-0.13	0.23	3.91
27	0.70	0.89	0.53	-0.44	-2.73
28	0.88	1.10	0.68	3.18	-5.23
25 to 28	0.42	0.47	0.37	0.99	-0.47
response=illness					
Day	ePCP	ePCP.l	ePCP.u	AUROC	NLL
17 to 24	0.21	0.22	0.19	3.67	-5.68
25	0.49	0.61	0.39	2.59	-6.47
26	0.67	0.84	0.51	3.34	-13.34
27	0.56	0.71	0.42	0.34	-10.52
28	0.04	0.05	0.03	4.41	-10.17
25 to 28	0.44	0.49	0.39	3.04	-9.58

Note: The changes between $K = 150$ and $K = 180$ and the ones between $K = 250$ and $K = 300$ are reported for response=stress and response=illness, respectively. The headers, ePCP.l and ePCP.u corresponds to lower and upper 95% confidence bounds for ePCP. Percentage changes were calculated by $((AM_{K1} - AM_{K2})/AM_{K1}) * 100$ where AM denotes “Accuracy Measures”, and $K1 = (150, 250)$ and $K2 = (180, 300)$.

characteristics, z_i, $\hat{z}_i$, and columns 1 to $(T + m)$ of the Cholesky factor, $\hat{\Sigma}^{1/2}$, for forecasting time period is denoted by MMREM1, the one which also relies on using $\hat{z}_i$ but uses columns $(T + 1)$ to $(T + m)$ of $\hat{\Sigma}^{1/2}$ is denoted by MMREM2. Also, as is mentioned in Section 5.2, we considered forecasting with only $\hat{\Delta}_{itj}$'s which corresponds to assuming all the subjects are average subjects, i.e. $z_i = 0$. Note that the results of this method is denoted by MMREM4.

While forecasting with PNMTREM, following Hyndman and Khandakar (2008) we considered several exponential smoothing (ES) methods to forecast $\alpha_{t1;1}$, $\alpha_{t,2;1}$ and σ_t for $t = 25, 26, 27, 28$. For each of the ES methods, several accuracy measures such as root means squared error, mean absolute (percentage) error and mean absolute scaled error, and some model selection criterion such as Akaike Information Criterion, its small sample correction and Bayesian Information Criterion were calculated through the R package **forecast** (Hyndman et al., 2012). These measures and criterion values seem to be very close to each other for $\alpha_{t1,1}$, $\alpha_{t,2;1}$ and σ_t indeed (results not shown here). We forecasted these parameters via all the ES methods and most reasonable ones are selected as the forecast values of these parameters. For instance, for $\alpha_{t,2;1}$, we could build two different ES models: 1) ES with additive error,

Table 6.8: Forecast values of $\alpha_{t,1;1}$, $\alpha_{t,2;1}$ and σ_t.

Day (t)	$\alpha_{t,1;1}$	$\alpha_{t,2;1}$	σ_t
25	1.31	0.31	0.61
26	1.31	0.31	0.64
27	1.31	0.31	0.68
28	1.31	0.31	0.71

Note: The forecast methods are ES with additive error, no trend, no seasonality for $\alpha_{t,1;1}$ and $\alpha_{t,2;1}$, additive error, additive trend, no seasonality for σ_t.

no trend, no seasonality, and 2) ES with additive error, additive trend, no seasonality. While the first ES resulted positive forecast values for $\alpha_{t,2;1}$, the second one resulted all negative forecast values. We checked the empirical Spearman Rank Correlation values for the period of days 25 to 28 one-by-one for response=stress and response=illness, and observed that all of these correlation values are positive indeed. Therefore, the first method, ES with additive error, no trend, no seasonality were used to forecast $\alpha_{t,2;1}$. Moreover, while ES with additive error, no trend, no seasonality was used to forecast $\alpha_{t,1;1}$, ES with additive error, additive trend, no seasonality was used to forecast σ_t. The forecast results of these parameters are given in Table 6.8. Note that we did not consider ES models with seasonality (either additive or multiplicative), since we do not expect our data reflect such a behavior, and actually there is not enough time points to expect this behavior.

Several PNMTREM(1)'s were considered to forecast MSCM data mainly based on the choices of the *cutoff* value to dichotomize the $\hat{p}_{itj}$'s. For each of these choices, two additional forecasting methodology combinations were considered; forecasting with considering $\hat{z}_i$ as estimated from the model and with $\hat{z}_i = 0$ (all the subjects are average persons). Among these methods, whereas we call the method with *cutoff*$= 0.5$ and forecasting with $\hat{z}_i$ as PNMTREM1, we call the one with *cutoff*=0.5 and forecasting with $\hat{z}_i = 0$ as PNMTREM2 (Tables 6.9 & 6.10). The calling systematic of the other methods are as follows: PNMTREM3: using actual values of y_{itj} (not dichotomizing $\hat{p}_{itj}$) at $day = 25, \ldots, 28$ with $\hat{z}_i$, PNMTREM4: using actual values of y_{itj} at $day = 25, \ldots, 28$ with $\hat{z}_i = 0$; PNMTREM5: obtaining *cutoff* as empirical proportions of 1's by using the model building portion of data for response=stress (*cutoff*=0.125) and response=illness (*cutoff*=0.113) separately and using $\hat{z}_i$, PNMTREM6: *cutoff* as empirical proportions of 1's with $\hat{z}_i = 0$; PNMTREM7: simulating the responses from Bernoulli distributions with predicted success probabilities, $\hat{p}_{itj}$ and using $\hat{z}_i$, PNMTREM8: simulating the responses based on $\hat{p}_{itj}$ via using $\hat{z}_i = 0$. The results of PNMTREM3-PNTREM8 could be found in Tables 6.11 & 6.12.

Among these methods for PNMTREM, PNMTREM3 and PNMTREM4 do not reflect the real life cases since we do not know the actual values of the responses to be forecasted. Nonetheless, we considered these cases to see the effect of dichotomizing the $\hat{p}_{itj}$'s. Since PNMTREM1 and PNMTREM2 gave the best results for forecasting with PNMTREM, we included them in Tables 6.9 & 6.10 together with other models for comparison purposes. The other PNMTREM (PNMTREM3-PNTREM8) methods are presented in Tables 6.11 & 6.12 for illustration purposes.

In Tables 6.9, 6.10, 6.11 & 6.12, we preferred reporting the expected proportion of correct prediction (ePCP; Herron, 1999) and related 95% confidence intervals, area under the receiver operating characteristics (AUROC) curve and negative log-likelihood (NLL) to summarize the forecasting results. Although, we calculated proportion of correct prediction (PCP), sensitivity and specificity values, we did not prefer to report these measures here. As we mentioned in Section 5.3, we observed that these measures yielded wrong decisions such as very high PCP with sensitivity=0 or low values for both of them due to choices of cut-off values for datasets with unexpected behaviors such as very low percentages of 1's. For instance, while the observed proportion of 1's is 0.125 for response=stress, it is 0.113 for response=illness at the period of day 17 to 24, these values are 0.088 and 0.096 at the period of days 25 to 28 for response=stress and response=illness, respectively. Last but not least, since all the independent variables in the MSCM study were collected in a preliminary interview, all of them are time-invariant. Therefore, we do not need to forecast them.

For model building period, days 17 to 24, the marginalized models, MMREM and PMTREM(1) seem to performed better compared to UMM, MMM1 and MMM2. For instance, for response=stress (Table 6.9) MMREM3 and MMREM4 outperformed these marginal models; the corresponding ePCP values are 0.844 and 0.842, respectively and related 95% CI's do not intersect the ones for these marginal models. Moreover, the AUROC values of MMREM1 and MMREM2 were found to be 0.821 which indicate these models over-performed compared to the others. Note that MMREM1 and MMREM2 are identical models at days 17 to 24 indeed. PNMTREM1 follows these models with an AUROC value of 0.804 (response=stress). A similar behavior was observed in terms of NLL, corresponding values are 406.3, 406.3 and 422.1 for MMREM1, MMREM2 and PNMTREM1, respectively. However, worst NLL were seen for MMREM3, MMREM4 and PNMTREM2; corresponding values are 522.5, 513.5 and 511.2, respectively. Similar model ranking was observed for the period of days 17 to 24, for response=illness (Table 6.10). For instance, in terms of ePCP all the MMREM's and PNMTREM1 seem to outperform the others, corresponding values are 0.868, 0.868, 0.881, 0.882 and 0.855, respectively. In terms of AUROC and NLL, MMREM1, MMREM2 and PNMTREM1 seem to be the best models. Whereas the corresponding AUROC values are 0.878, 0.878 and 0.814; the NLL values are 331.6, 331.6 and 394.7, respectively. Again, MMREM3, MMREM4 and PNMTREM2 seem to be the worst performing models in terms of NLL; corresponding values are 762.8, 778.3 and 521.2, respectively.

For forecasting time period, days 25 to 28, PNMTREM1 and PNMTREM2 performed worst in terms of ePCP and NLL. Nevertheless, in terms of AUROC, these models especially PNMTREM2 outperformed the other models. For instance, while the corresponding ePCP values for response=stress (Table 6.9) in the period of days 25 to 28 (overall forecasting period) are 0.665 and 0.730, the NLL values are 491.2 and 293.9, respectively. Nonetheless, their AUROC values are found to be 0.736 and 0.759, respectively. Similarly, for response=illness (Table 6.10), while the ePCP values of PNMTREM1 and PNMTREM2 are found to be 0.556 and 0.621, their NLL values are found to be 921.4 and 531.3, respectively in the period of days 25 to 28. Again, similar to response=stress, their AUROC values are the highest ones, corresponding values are 0.701 and 0.765 for PNMTREM1 and PNMTREM2, respectively. In terms of ePCP, MMREM's, especially, MMREM3 and MMMREM4 seem to be best performed models. Nonetheless, although this is always true compared to PNMTREM's such that related 95% CI's do not intersect, this better performance seem not to be statistically significant compared to UMM and MMM's (MMM1 & MMM2) in most of the cases. For instance, for response=stress at day 26, the 95% CI of ePCP is found to be $(0.614, 0.755)$ for PNMTREM1, the 95% CI for MMREM3 is found to be $(0.820, 0.922)$. On the other hand, for response=illness at day 25, the 95% CI's for ePCP are found to be $(0.720, 0.845)$ and $(0.808, 0.913)$ for MMM1 and MMREM4, respectively. Nonetheless, MMREM3 and MMREM4 seem to significantly better performed compared to UMM and MMM's for response=illness at days 27 and 28. UMM and MMM's seem usually having the lowest NLL statistics. For instance, for response=illness the NLL values of UMM, MMM1 and MMM2 are found to be 227.3, 224.7 and 222.4, respectively for overall forecasting period. MMREM1 and MMREM2 performed worst for response=stress at day 28 in terms of AUROC, corre-

Table 6.9: Forecast results for mothers's stress.

Model	Day	ePCP (95% CI)	AUROC	NLL	Day	ePCP (95% CI)	AUROC	NLL
UMM (Exch)	17 to 24	0.799 (0.778,0.821)	0.726	456.0	27	0.858 (0.805,0.911)	0.790	36.7
MMM1 (Exch)		0.799 (0.778,0.821)	0.726	455.2		0.855 (0.802,0.909)	0.747	37.8
MMM2 (Exch)		0.800 (0.779,0.821)	0.722	455.8		0.861 (0.809,0.914)	0.780	35.8
MMREM1		0.828 (0.808,0.849)	**0.821**	**406.3**		0.851 (0.797,0.905)	0.757	43.7
MMREM2		0.828 (0.808,0.849)	**0.821**	406.3		0.884 (0.835,0.933)	0.754	37.7
MMREM3		**0.844** (0.824,0.863)	0.712	**522.5**		0.903 (0.856,0.948)	0.735	37.3
MMREM4		**0.842** (0.823,0.862)	0.721	**513.3**		0.900 (0.854,0.945)	0.719	37.8
PNMTREM1		0.824 (0.803,0.844)	**0.804**	**422.1**		**0.662** (0.590,0.734)	0.764	**130.4**
PNMTREM2		0.829 (0.809,0.850)	0.712	**511.2**		**0.754** (0.689,0.819)	**0.843**	65.3
UMM (Exch)	25	0.823 (0.765,0.881)	0.678	52.4	28	0.846 (0.791,0.901)	0.709	44.0
MMM1 (Exch)		0.821 (0.763,0.880)	0.641	53.3		0.842 (0.787,0.898)	0.670	45.3
MMM2 (Exch)		0.823 (0.765,0.881)	0.657	53.1		0.845 (0.790,0.900)	0.690	44.9
MMREM1		0.831 (0.774,0.888)	**0.775**	47.4		0.830 (0.773,0.887)	**0.585**	63.9
MMREM2		0.866 (0.814,0.917)	**0.759**	51.3		0.859 (0.806,0.912)	**0.595**	56.0
MMREM3		0.868 (0.817,0.919)	0.675	57.8		0.899 (0.841,0.937)	0.661	48.1
MMREM4		0.867 (0.815,0.918)	0.695	56.1		0.886 (0.838,0.934)	0.651	48.7
PNMTREM1		**0.688** (0.618,0.758)	**0.761**	**92.6**		**0.625** (0.552,0.699)	0.608	**153.8**
PNMTREM2		**0.723** (0.655,0.791)	**0.743**	69.7		**0.732** (0.665,0.799)	**0.735**	75.8
UMM (Exch)	26	0.829 (0.772,0.886)	0.704	49.6	25 to 28	0.839 (0.811,0.867)	0.718	182.7
MMM1 (Exch)		0.828 (0.770,0.885)	0.704	49.6		0.841 (0.813,0.869)	0.687	186.0
MMM2 (Exch)		0.829 (0.772,0.886)	0.715	49.7		0.840 (0.812,0.868)	0.708	183.5
MMREM1		0.836 (0.780,0.892)	**0.811**	44.1		0.837 (0.809,0.865)	0.729	199.1
MMREM2		0.865 (0.813,0.916)	**0.803**	45.7		0.868 (0.843,0.894)	0.711	190.6
MMREM3		0.871 **(0.820,0.922)**	0.728	53.5		0.883 (0.858,0.907)	0.699	196.6
MMREM4		0.869 (0.818,0.920)	0.725	53.2		0.880 (0.856,0.905)	0.698	195.8
PNMTREM1		**0.685 (0.614,0.755)**	**0.819**	**114.4**		**0.665** (0.629,0.701)	**0.736**	**491.2**
PNMTREM2		**0.711** (0.642,0.780)	**0.731**	**83.2**		**0.730** (0.696,0.764)	**0.759**	**293.9**

Note: MMREM1 and MMREM2 are identical for model building periods.

sponding values are 0.585 and 0.595, respectively. Additionally, UMM, MMM1 and MMM2 performed worst for response=illness at day 27 again in terms of AUROC, corresponding values are 0.560, 0.562 and 0.575, respectively. MMREM3 and MMREM4 performed worst for response=illness at day 25 in terms of AUROC; corresponding values are 0.570 and 0.582, respectively.

As expected, the results of PNMTREM3 and PNMTREM4 (Tables 6.11 & 6.12) seem to be better than the other methods regarding PNMTREM, since these methods do not include dichotomizing the predicted success probabilities, $\hat{p}_{itj}$. The results of PNMTREM1 and PNMTREM2 which uses $cutoff = 0.5$ follow these methods. PNMTREM7 and PNMTEM8 (simulating the responses via $\hat{p}_{itj}$) seem to follow PNMTREM1 and PNMTREM2. Interestingly, PNMTREM5 and PNMTREM6 (empirical cutoff) seem to be the worst methods which yielded ePCP of 0.198 and 0.192 at day 26 for response=illness (Table 6.12). They are even worst at day 28 for response=illness, corresponding values are 0.175 and 0.172, respectively.

Table 6.10: Forecast results for children's illness.

Model	Day	ePCP (95% CI)	AUROC	NLL	Day	ePCP (95% CI)	AUROC	NLL
UMM (AR-1)	17 to 24	0.815 (0.794,0.836)	0.719	428.2	27	0.793 **(0.732,0.855)**	**0.560**	56.6
MMM1 (Exch)		0.815 (0.794,0.836)	0.720	426.7		0.800 **(0.739,0.861)**	**0.562**	55.5
MMM2 (Exch)		0.815 (0.794,0.835)	0.719	427.4		0.800 **(0.739,0.861)**	**0.575**	53.9
MMREM1		**0.868** (0.850,0.886)	**0.878**	**331.6**		0.759 (0.694,0.824)	0.656	107.2
MMREM2		**0.868** (0.850,0.886)	**0.878**	**331.6**		0.873 (0.822,0.923)	0.677	65.0
MMREM3		**0.881** (0.864,0.899)	0.697	**762.8**		**0.907 (0.863,0.951)**	0.662	68.0
MMREM4		**0.882** (0.865,0.900)	0.705	**778.3**		**0.908 (0.864,0.952)**	0.665	72.5
PNMTREM1		**0.855** (0.836,0.874)	**0.814**	**394.7**		**0.526** (0.450,0.602)	0.686	**265.7**
PNMTREM2		0.851 (0.831,0.870)	0.691	**521.2**		**0.605** (0.531,0.679)	**0.741**	**146.4**
UMM (AR-1)	25	0.779 (0.716,0.842)	0.640	**65.4**	28	0.799 **(0.738,0.860)**	0.687	50.4
MMM1 (Exch)		0.782 **(0.720,0.845)**	0.632	**65.7**		0.807 **(0.747,0.866)**	0.685	49.2
MMM2 (Exch)		0.781 (0.718,0.843)	0.608	**65.9**		0.805 **(0.745,0.865)**	0.674	48.6
MMREM1		0.785 (0.722,0.847)	**0.736**	103.8		0.761 (0.697,0.826)	0.622	**120.6**
MMREM2		0.856 (0.803,0.910)	0.674	110.2		0.853 (0.799,0.907)	0.647	73.7
MMREM3		0.857 (0.804,0.910)	**0.570**	128.6		**0.911 (0.868,0.954)**	0.683	58.4
MMREM4		0.860 **(0.808,0.913)**	**0.582**	130.4		**0.917 (0.876,0.959)**	0.686	59.2
PNMTREM1		**0.629** (0.556,0.702)	**0.798**	125.8		**0.514** (0.439,0.590)	0.638	**287.7**
PNMTREM2		**0.652** (0.580,0.724)	**0.783**	98.3		**0.605** (0.531,0.679)	**0.772**	**157.4**
UMM (AR-1)	26	0.797 (0.736,0.858)	0.623	**54.9**	25 to 28	0.792 (0.761,0.823)	0.617	**227.3**
MMM1 (Exch)		0.802 (0.742,0.863)	0.627	**54.3**		0.798 (0.767,0.828)	0.616	**224.7**
MMM2 (Exch)		0.801 (0.741,0.862)	0.614	**54.0**		0.800 (0.766,0.827)	0.608	**222.4**
MMREM1		0.764 (0.699,0.828)	0.677	104.3		0.767 (0.735,0.799)	0.675	435.9
MMREM2		0.875 (0.825,0.926)	**0.725**	69.6		0.864 (0.838,0.890)	0.657	318.6
MMREM3		0.893 (0.846,0.940)	0.670	80.4		**0.892** (0.869,0.916)	0.636	335.4
MMREM4		0.895 (0.849,0.942)	0.688	83.1		**0.895** (0.872,0.918)	0.645	345.1
PNMTREM1		**0.555** (0.479,0.630)	**0.713**	**242.1**		**0.556** (0.518,0.594)	**0.701**	**921.4**
PNMTREM2		**0.624** (0.550,0.697)	**0.791**	129.3		**0.621** (0.585,0.658)	**0.765**	**531.3**

Note: MMREM1 and MMREM2 are identical for model building periods.

Table 6.11: Alternative trials for PNMTREM-Forecast results for mothers's stress.

Model	Day	ePCP (95% CI)	AUROC	NLL	Day	ePCP (95% CI)	AUROC	NLL
PNMTREM3	17 to 24	0.824 (0.803,0.844)	0.804	422.1	27	0.694 (0.624,0.764)	0.772	90.3
PNMTREM4		0.829 (0.809,0.850)	0.712	511.2		0.750 (0.684,0.816)	0.831	58.9
PNMTREM5		0.824 (0.803,0.844)	0.804	422.1		0.486 (0.411,0.562)	0.751	196.4
PNMTREM6		0.829 (0.809,0.850)	0.712	511.2		0.546 (0.471,0.622)	0.792	136.5
PNMTREM7		0.824 (0.803,0.844)	0.804	422.1		0.606 (0.532,0.680)	0.760	149.6
PNMTREM8		0.829 (0.809,0.850)	0.712	511.2		0.662 (0.591,0.734)	0.790	97.8
PNMTREM3	25	0.688 (0.618,0.758)	0.761	92.6	28	0.680 (0.609,0.750)	0.676	90.8
PNMTREM4		0.723 (0.655,0.791)	0.743	69.7		0.756 (0.691,0.821)	0.771	58.1
PNMTREM5		0.688 (0.618,0.758)	0.761	92.6		0.464 (0.389,0.540)	0.584	214.2
PNMTREM6		0.723 (0.655,0.791)	0.743	69.7		0.533 (0.458,0.609)	0.715	143.1
PNMTREM7		0.688 (0.618,0.758)	0.761	92.6		0.550 (0.474,0.625)	0.547	180.8
PNMTREM8		0.723 (0.655,0.791)	0.743	69.7		0.632 (0.559,0.706)	0.675	114.5
PNMTREM3	26	0.700 (0.631,0.770)	0.826	81.2	25 to 28	0.690 (0.655,0.726)	0.762	355.0
PNMTREM4		0.731 (0.664,0.798)	0.752	66.5		0.740 (0.707,0.773)	0.772	253.9
PNMTREM5		0.521 (0.446,0.597)	0.827	176.7		0.540 (0.502,0.578)	0.703	679.9
PNMTREM6		0.554 (0.478,0.629)	0.734	135.1		0.589 (0.552,0.626)	0.720	484.3
PNMTREM7		0.663 (0.591,0.734)	0.821	114.6		0.627 (0.590,0.663)	0.713	537.6
PNMTREM8		0.677 (0.606,0.748)	0.725	89.9		0.674 (0.638,0.709)	0.717	371.8

Note: PNMTREM3, PNMTREM5 and PNMTREM7 are identical at days 17-24 and 25. Similarly, PNMTREM4, PNMTREM6 and PNMTREM8 are identical at these days.

Table 6.12: Alternative trials for PNMTREM-Forecast results for children's illness.

Model	Day	ePCP (95% CI)	AUROC	NLL	Day	ePCP (95% CI)	AUROC	NLL
PNMTREM3	17 to 24	0.855 (0.836,0.874)	0.814	394.7	27	0.608 (0.534,0.682)	0.804	114.4
PNMTREM4		0.851 (0.831,0.870)	0.691	521.2		0.655 (0.583,0.727)	0.815	88.4
PNMTREM5		0.855 (0.836,0.874)	0.814	394.7		**0.187** (0.128,0.247)	0.657	452.2
PNMTREM6		0.851 (0.831,0.870)	0.691	521.2		**0.180** (0.122,0.239)	0.634	391.8
PNMTREM7		0.855 (0.836,0.874)	0.814	394.7		0.395 (0.321,0.469)	0.679	326.9
PNMTREM8		0.851 (0.831,0.870)	0.691	521.2		0.424 (0.349,0.499)	0.678	258.4
PNMTREM3	25	0.629 (0.556,0.702)	0.798	125.8	28	0.595 (0.520,0.669)	0.706	123.1
PNMTREM4		0.652 (0.580,0.724)	0.783	98.3		0.650 (0.578,0.723)	0.776	90.4
PNMTREM5		0.629 (0.556,0.702)	0.798	125.8		**0.175** (0.117,0.232)	0.594	475.1
PNMTREM6		0.652 (0.580,0.724)	0.783	98.3		**0.172** (0.115,0.229)	0.655	405.4
PNMTREM7		0.629 (0.556,0.702)	0.798	125.8		0.326 (0.255,0.397)	0.608	389.9
PNMTREM8		0.652 (0.580,0.724)	0.783	98.3		0.378 (0.304,0.452)	0.675	297.4
PNMTREM3	26	0.597 (0.523,0.672)	0.786	129.1	25 to 28	0.607 (0.570,0.644)	0.777	492.4
PNMTREM4		0.640 (0.567,0.713)	0.814	97.7		0.649 (0.613,0.686)	0.798	374.7
PNMTREM5		**0.198** (0.138,0.259)	0.651	437.5		**0.297** (0.263,0.332)	0.637	1490.5
PNMTREM6		**0.192** (0.132,0.251)	0.639	378.9		**0.299** (0.264,0.334)	0.616	1274.4
PNMTREM7		0.468 (0.392,0.544)	0.720	274.5		0.455 (0.417,0.492)	0.677	1117.1
PNMTREM8		0.505 (0.430,0.581)	0.641	194.3		0.490 (0.452,0.528)	0.669	848.4

Note: PNMTREM3, PNMTREM5 and PNMTREM7 are identical at days 17-24 and 25. Similarly, PNMTREM4, PNMTREM6 and PNMTREM8 are identical at these days.

6.3 Discussion and Conclusion

In this chapter, we considered forecasting MSCM data, specifcally mother's stress and children's illness statuses, via several models. These models include univariate and multivariate marginal models (UMM, MMM1 and MMM2), and two different marginalized models (MMREM and PNMTREM, specifically PNMTREM(1)). The latter models are very complex models compared to the former ones. For instance, PNMTREM includes time-varying association parameters (α_t, σ_t), hence one needs to forecast these parameters as well. Similarly, these models took more time compared to the former ones for parameter estimation. We built several models regarding these 5 different models and considered several versions of them.

Different accuracy measures yielded different ranking of the models. Nevertheless, in the model building period, days 17 to 24, MMREM and PNMTREM seem to be superior compared to univariate and multivariate marginal models in terms of all the measures. However, for the forecasting time period the ranking could be changing. In terms of ePCP, MMREM's, especially MMREM3 and MMREM4 which rely on forecasting with only $\hat{\Delta}_{itj}$ and simulating z_i from independent standard normal distribution via some amount of replication, seem to be the best ones. Actually, these methods performed very similarly which is expected beforehand due the law of large numbers such that large amount of standard normal random variables tend to have mean 0 which is not necessarily different than considering only $\hat{\Delta}_{itj}$ for forecasting, i.e. $\hat{z}_i = 0$. Again in terms of ePCP, the performance of MMREM3 and MMREM4 seem not be significantly better than UMM, MMM1, MMM2, MMREM1 and MMREM2, i.e. for most cases the 95% CI's are overlapping. Nonetheless, both PNMTREM1 and PNMTREM2 seem to be coming in the bottom of the model list, in terms of ePCP. In terms of AUROC, PNMTREM's seem to be the best models to forecast the MSCM data and MMREM's, especially MMREM1 and MMREM2 follow them. UMM, MMM1 and MMM2 seem to be worst among others in terms of AUROC. In terms of NLL, UMM, MMM1 and MMM2 seem to be better than the others and PNMTREM's seem to be the worst ones.

The marginal models, UMM, MMM1 and MMM2, seem to perform similarly for the MSCM data. Actually, one might expect that MMM1 and/or MMM2 should outperform the UMM due to the fact that the former ones accommodate both within and between response dependencies. However, the between response associations of mother's stress and children's illness are weak indeed, such that the Spearman rank correlations range between -0.03 and 0.29 in the period of days 17 to 28. Note that these correlation have an average of 0.13 at this time period. For datasets with larger such correlations MMM1 and/or MMM2 might yield better forecasts. The MMM2 considered for MSCM data includes fitting 5 less parameters compared to its parent MMM1. Although this resulted in efficiency gain for parameter estimates, only little improvement is observed for forecasting. For datasets with lots of multiple responses and lots of independent variables which have similar effects on these responses, one might expect MMM2 could yield better forecasts, since MMM2 would fit many less parameters.

MMREM with forecasting methodology of using the columns $(T+1), \ldots, (T+m)$ of $\hat{\Sigma}^{1/2}$ together with using $\hat{z}_i$ (i.e. MMREM2) and MMREM with forecasting only with $\hat{\Delta}_{itj}$ (i.e. $\hat{z}_i = 0$; MMREM4) could be studied together. Similarly, PNMTREM with $\hat{z}_i$ (i.e. PNMTREM1) and its corresponding methodology that only used $\hat{\Delta}^*_{itj}$ (i.e. PNMTREM2) could be studied together. Actually, PNMTREM with only $\hat{\Delta}^*_{itj}$ seem to be better in terms of forecasting compared to the use of $\hat{z}_i$ for most of the cases in MSCM data. This might be most probably due to the reasons such that the σ_t's might not be well forecasted, and the estimates of $\hat{z}_i$ might not hold for forecasting time period. Moreover, the estimates of λ_j, $\hat{\lambda}_j$, which is obtained from the model building portion of data, might not well represent the multiple response correlation for forecasting time periods. We observed that choice of cutoff to dichotomize the predicted probabilities of PNMTREM heavily effects forecasting results. Several cutoff value choices could be studied together by considering a portion of the available data as a validation part through time.

Unlike the common expectancy that the accuracy measures decrease as the lag between the last time point of model building data (T) and the forecast time point increase ($T + m$), i.e. as m increases, accuracy measures increased for MSCM data at some time points as the aforementioned time lag increased. We checked the effects of the independent variables on the responses via time-by-time logistic regression models for days 25 to 28 (at the forecasting time period) by calculating the accuracy measures (considered for the evaluation of forecasts), and observed similar trend with the forecast accuracy measures found for MSCM data. This directed us to conclude this unexpected behavior might be due to changing effects of independent variables with respect to time. Moreover, none of the models except PNMTREM rely on dichotomizing the forecasted success probabilities for later time point forecasts. Actually, PNMTREM reflected decreasing accuracy measures as the aforementioned time lag increased for almost all of the forecast time points.

Marginal models, both UMM, MMM1 and MMM2, resulted in low (i.e. better) NLL values compared to the other methods. This is due to the fact that these models do not include random effects structures, i.e. do not take into account individual differences. Therefore, they yielded similar amounts of success probabilities and low NLL. This observation is in agreement with the one we experienced during the PNMTREM analysis of Iowa Youth and Families project data; for instance see Section 4.5.2 and Figure 4.2.

In this study, we consdiered first-order PNMTREM, PNMTREM(1) to forecast the MSCM data. Higher-order PNMTREM's could result better forecasts depending on the association structure of the datasets such as datasets with strong lag-2 dependencies. Empirical correlation structure analysis of mother's stress and children's illness indicate an unstructured type correlation matrix for response=stress and an AR-1 type correlation structure for response=illness. Nonetheless, stressing the choice of dependence structure is a disadvantage of the marginalized models considered in this study, i.e. PNMTREM(1) only takes the lag-1 response into account and MMREM assumes an AR-1 correlation structure which is common for multiple responses. However, UMM and MMM's permit working with different correlation structures. Moreover, PNMTREM and MMREM could result better forecast results for datasets with strong between-response dependencies compared to UMM. We forecasted the

Δ_{itj} of MMREM via 40-points Gauss-Hermite Quadratures and Newton-Raphson root finding method. This can be compared to smoothing methods.

Forecasting studies with real life datasets with more than 2 responses and moderate and/or strong between response dependencies would be a possible future work regarding our forecasting study with MSCM dataset.

CHAPTER 7

SIMULATION STUDY ON FORECASTING

In this chapter, we present a simulation study on forecasting multivariate longitudinal binary data. While in Section 7.1 we provide the details of the data generation process, in Section 7.2 we provide the forecasting results of our models. Discussions and conclusions are presented in Section 7.3.

7.1 Data Generation

It is well-known that real life datasets are full of surprises. This can be exemplified from a longitudinal data perspective that real life datasets might reflect very strange association structures. For instance, in MSCM data while the correlation between mother's stress and children's illness is negative at day 17, this correlation is positive at later time points. These diverse behaviors inevitably have an important role in terms of conclusions. For instance, forecasting results of MSCM data suggest that marginalized models might yield improved forecasts. But can we trust this conclusion hundred percent? For instance, we stated in Chapters 3 & 6 that the between response dependence of MSCM data is weak indeed and this might make the multivariate models unnecessary, since they include estimation of extra between-response association parameter(s) which would yield inefficiency with no doubt. Actually, the univariate and multivariate marginal modeling results seem to be similar in terms of forecasting MSCM data. This creates the following questions. What if a dataset with stronger multivariate response associations is available? How the forecasting results would be affected? Similarly, the strength of individual variations would be effective regarding the forecasting conclusions. Put another way, the strength of individual variations would effect both the modeling and forecasting results such as lower such variations would tend to support the marginal models. Moreover, the differing modeling frameworks would differ in terms of modeling and forecasting performances. For instance, while PNMTREM considers changing individual variations over time but shared for multiple responses, MMREM considers a constant individual variation over time, but separate for multiple responses. To illustrate this example, modeling results of PNMTREM on MSCM data showed that the individual variations over time seems not to be very different, which might yield inefficiency of the related parameter estimates and would affect the related forecasting results.

The well-known cure to these questions is conducting simulation studies, in which we know the truth and might set up several scenarios. Therefore, the behavior of competitor models under different conditions might be drawn. The typical simulation study considered in literature includes generating the data from a specified model via known parameters. However, these datasets would include the association structures imposed by the data generation model and most probably tend to make the data generating model superior compared to the others. Therefore, we considered a simulation study which relies on model-independent data generation. We mainly built on the methodology of Aslan (2010) who considered model-independent data generation for *univariate* longitudinal binary data with *four time-dependent covariates*. Specifically, we considered model-independent data generation for *bivariate* longitudinal binary data with *two time-dependent and two time-independent covariates*. Although there are several choices of scenarios to consider for our forecasting study, here we considered only one scenario which reflect the common expectancies from a multivariate longitudinal binary data, based on our experiences with such data. While designing the data generation scenario, we tried to consider a more general dataset which would not create biases towards any of the models. Differing model-independent data generation scenarios might be the topic of future studies.

We mainly assume that there are 500 subjects ($i = 1, \ldots, 500$), who are followed repeatedly over 8 time points ($t = 1, \ldots, 8$). At each time points, six different variables are assumed to be measured for these subjects. Among these variables, two of them are considered to be the bivariate binary responses, i.e. $k = 2$ and four of them are to be the independent variables, $p = 4$. Among the independent variables, whereas two of them are assumed to be time-independent, other two are assumed to be time-dependent. At first all the variables, including the dependent ones, are assumed to be continuous and they are generated from a psuedo multivariate normal distribution via a specific mean, μ, and variance-covariance, Σ, combination.

Specifically, all the variables are assumed to have mean 0, i.e. $\mu = (0, \ldots, 0)_{1\times 6}$, which is constant over time. To specify the desired correlation structures of our bivariate longitudinal binary data, such as the ones between longitudinal responses, between dependent and independent variables, within dependent and independent variables, a well-known decomposition of variance-covariance matrix is considered. This decomposition can be illustrated by

$$\Sigma = V^{1/2}\,\rho\,V^{1/2} \tag{7.1}$$

where $V^{1/2}$ is a 34×34 diagonal matrix including the square root of the variance parameters in its diagonals, and ρ is a 34×34 symmetric matrix which includes the correlation parameters. This decomposition can be found in standard multivariate statistical references, such as Johnson and Wichern (1998, cited in Aslan, 2010). Here, we shall note that both V and ρ have dimensions of 34×34 rather than 48×48. This is due to the fact that we generate the time independent variables only once at the first time point. Specifically, the first 6 rows and columns corresponds to the variables at first time point with the first two are the dependent variables, and next four are the independent ones. Also, the ones between rows and columns 7 and 10 corresponds to the variables at second time point with while the $7th$ and $8th$ are the dependent variables, the $9th$ and $10th$ are the time-varying independent variables. The order of the rest of the matrices follows these examples. The general form of them could be depicted as below.

$$V^{1/2} = \begin{bmatrix} \sqrt{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{34}^2} \end{bmatrix} \quad , \quad \rho = \begin{bmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,34} \\ \rho_{2,1} & 1 & \cdots & \rho_{2,34} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{34,1} & \rho_{34,2} & \cdots & 1 \end{bmatrix}$$

Here, σ_i^2 denotes the variance of the i^{th} variable, and $\rho_{i,j}$ is the correlation between the variables i and j, where $\rho_{i,j} = \rho_{j,i}$ for $i, j = 1, \ldots, 34$. Specifically, the continuous version of the dependent variables, Y_1^* and Y_2^* are assumed to have variances of 1.5 and 2.5. Moreover, the independent variables, X_1, X_2, X_3 and X_4 are assumed to have variances of 8, 2.5, 15, 25, respectively. Among them while X_1 and X_3 are time-independent, X_2 and X_4 are time dependent. These variances are arbitrary indeed, and selected via several trial and errors. Nonetheless, we wanted to consider the bivariate responses having different variances, to see

Table 7.1: A summary of the assumed correlation structure.

Time	$w - Y_j^*$	$w - X_l$	Y_j^*, X_l	Y_1^*, Y_2^*	$X_l, X_{l'}$
t, t	1.00	1.00	0.80	0.60	0.20
$t, t-1$	0.90	0.88	0.70	0.55	0.18
$t, t-2$	0.80	0.76	0.60	0.45	0.16
$t, t-3$	0.70	0.64	0.50	0.40	0.14
$t, t-4$	0.60	0.52	0.40	0.35	0.12
$t, t-5$	0.50	0.40	0.30	0.30	0.10
$t, t-6$	0.40	0.28	0.20	0.25	0.08
$t, t-7$	0.30	0.16	0.10	0.20	0.06

Note: $w - Y^*$ & $w - X$ correspond to within response and within covariate correlations, respectively. Values are correlations for $t = 1, \ldots, 8$. In this table while $j = 1, 2$, $l = 1, 2, 3, 4$.

the effect of response variance on forecasting results. Moreover, we wanted to assign low and large variances to time-dependent variables to see the effect of variance in terms of independent variable forecasting. Time-independent variables seem to be assigned moderate to large variances which were selected after the trial and errors to satisfy the desired correlations between and within the variables. Since presenting all the elements of ρ here requires long space and is redundant, we preferred reporting only a summary of it which reports the necessary correlations of bivariate longitudinal data (Table 7.1). We specifically assumed that ρ has an AR-1 type of correlation structure which is common in longitudinal studies (Aslan, 2010). Since we pre-specify a complex correlation form, the variance-covariance matrix may not be a positive-definite one. To prevent this problem, we consider singular value decomposition of the variance-covariance matrix.

Our trials with different correlation and variance choices showed that it is only possible to attain smaller correlations after data generations compared to assumed ones. Therefore, we assumed some of the correlations such as $cor(Y_1, Y_2)$ and $cor(X, Y)$ a bit high. We checked the resulted correlation structure of data by generating the data 10,000 times and calculating the average correlations. These results showed that the datasets reflect the correlation structures which we try to create. The independent variables are assumed to be exogenous such that the within-correlation of the covariates are strong, hence we do not need to accommodate the relationship between the covariate history and the current responses. To prevent multicollinearity problem, low correlations are assumed between the independent variables, such as $cor(X_l, X_{l'}) \leq 0.20$. Additionally, some initial analyses were conducted on a few datasets to control the data and whether the independent variables have significant relationships with the responses. Results showed that the data passed all the tests.

After generating the datasets, the dependent variables are dichotomized by the following rule: classify Y_{itj} as 0 if $p_{itj} < 0.5$ and 1 otherwise, where p_{itj}'s are calculated by

$$p_{itj} = \frac{exp(Y^*_{itj})}{1 + exp(Y^*_{itj})} \quad (7.2)$$

The period of times 1 to 4 are considered for model building and last four time points were considered for forecasting validation. The idea of dichotomizing continuous variables are in accordance with the idea of Hedeker and Gibbons (1997). They assumed that continuous values underline the binary variables. Last but not least, we shall note that the data are simulated via the `rmvnorm` function under the R package **mvtnorm** (Genz et al., 2011). Singular value decomposition method is achieved via setting the `method` option of `rmvnorm` to `"svd"`.

7.2 Results

In this section, we present the forecasting results of independent and dependent variables, respectively.

7.2.1 Independent Variables

We first consider forecasting the independent variables of the generated datasets. Remember that we have two different time-varying independent variables, X_2 and X_4, with variances of 2.5 and 25, respectively. First and second-order transition models (TM(1) and TM(2), respectively) are considered to forecast these variables. Mean absolute error (MAE) and mean absolute scaled error (MASE) are considered to evaluate the forecasting performances of these models. Note that small MAE and MASE values indicate better performance. Also note that the details of TM(1) and TM(2) and MAE and MASE could be found in Sections 5.2 and 5.3 of this thesis, respectively.

The forecasting results of the independent variables are presented in Tables 7.2 & 7.3 for X_2 and X_4, respectively. Results are calculated over 10,000 replications and obtaining all of the results presented in Tables 7.2 & 7.3 took 26.3 minutes on a PC with 4.00 GB RAM and

Table 7.2: Forecasting results of X_2, variance of 2.5, over 10,000 replications.

TM(1)					TM(2)				
	MAE		MASE			MAE		MASE	
Time	Mean	SE	Mean	SE	Time	Mean	SE	Mean	SE
2 to 4	0.671	0.013	0.993	0.012	3-4	0.668	0.016	1.046	0.057
5	0.671	0.023	1.242	0.068	5	0.670	0.023	1.531	0.175
6	**0.917**	**0.031**	1.692	0.092	6	**0.914**	**0.031**	2.086	**0.326**
7	1.084	0.037	1.996	0.111	7	1.079	0.037	2.457	**0.317**
8	1.208	0.041	**2.223**	**0.123**	8	1.201	0.041	**2.728**	**0.335**
5 to 8	0.970	0.025	**1.788**	0.085	5-8	0.966	0.025	2.200	**0.262**

Table 7.3: Forecasting results of X_4, variance of 25, over 10,000 replications.

TM(1)					TM(2)				
	MAE		MASE			MAE		MASE	
Time	Mean	SE	Mean	SE	Time	Mean	SE	Mean	SE
2 to 4	1.900	0.037	0.995	0.011	3-4	1.894	0.045	1.042	0.036
5	1.902	0.063	1.248	0.069	5	1.899	0.064	1.539	0.187
6	2.605	0.088	**1.705**	**0.096**	6	2.598	0.088	**2.101**	**0.258**
7	3.085	0.103	2.017	0.112	7	3.074	0.103	2.481	0.286
8	3.443	0.116	2.248	0.125	8	3.425	0.116	2.760	0.326
5 to 8	**2.759**	**0.071**	**1.804**	0.087	5-8	**2.749**	**0.071**	2.220	**0.241**

2.53 GHz processor. Mean and standard error (SE) of 10,000 MAE and MASE values are calculated and reported in these tables.

MAE indicate that TM(1) and TM(2) performed very similar in terms of forecasting X_2 (Table 7.2). While the mean values are very close to each other, standard errors seem to be same. For instance, whereas for $t = 6$ the mean MAE was found to be 0.917 for TM(1), it is found to be 0.914 for TM(2). Moreover, their standard errors were found to be 0.031. Note that though the standard errors of MAE of TM(1) and TM(2) for model building periods seem to be slightly different, they are not directly comparable, since while the former considers 2^{nd}, 3^{rd} and 4^{th} time points in the model building period, i.e. 1,500 observations, the latter considers only 3^{rd} and 4^{th} time points in this period, i.e. 1,000 observations. MASE indicate that TM(1) performed better in terms of forecasting X_2 compared to TM(2) (Table 7.2). For instance, for $t = 8$, the mean of MASE was found to be 2.223 for TM(1), it was found to be 2.728 for TM(2). Moreover, related standard errors were found to be 0.123 and 0.335, respectively. An unexpected decrease was observed for the standard error of MASE of TM(2) at $t = 7$ (Table 7.2). Whereas the standard error of MASE for TM(2) at $t = 6$ was found to be 0.326, it decreased to 0.317 at $t = 7$. Nonetheless, the mean MASE for TM(2) at $t = 7$ seem to be larger than the one at $t = 6$, corresponding values are 2.457 and 2.086, respectively. We controlled this issue by controlling our codes, but no error was found in our codes.

Similar to the MAE results of X_2 (Table 7.2), the MAE results of X_4 (Table 7.3) indicate that TM(1) and TM(2) performed very similar in terms of forecasting X_4. For instance, for the overall forecasting time period, $t = 5$ *to* 8, whereas the mean MAE values of TM(1) and TM(2) were found to be 2.759 and 2.749, respectively, and related standard errors were found to be 0.071. Also similar to X_2, the MASE results of X_4 indicate that TM(1) performed better in terms of forecasting X_4 compared to TM(2). For instance, at $t = 6$ while the mean and standard error of the MASE values were found to be 1.705 and 0.096 for TM(1), respectively, these values were found to be 2.101 and 0.258 for TM(2).

Increase in the variance of the independent variables do not seem to effect the performance of TM(1) and TM(2). Although the mean MAE and MASE values of X_4 seem to be larger compared to the ones of X_2 for both of these models, whereas MAE results indicate that TM(1) and TM(2) perform very similar, MASE results indicate that TM(1) performed better compared to TM(2). An interesting observation is that the standard error of the MASE values for X_2 resulted by TM(2) seem to be larger than the ones of X_4 in all time points except $t = 5$. For instance, while the standard error of the MASE values for X_2 yielded by TM(2) was found to be 0.262, this value was found to be 0.241 for X_4. Since MASE is a scale-independent measure, it can be used to compare the performance of same model on different variables. The mean of the MASE values for X_4 seem to be larger than the ones for X_2. This is an expected behavior indeed, since the variance of X_4 is larger than the variance of X_2. For instance, the MASE of TM(1) for X_2 was found to be 1.788 for $t = 5$ *to* 8, it was found to be 1.804 for X_4.

To sum up, although MAE indicated similar performances for TM(1) and TM(2), this observation might be regarded preference towards TM(1) to forecast the independent variables, since TM(2) estimates one more parameter compared to TM(1) and estimation of this extra parameter seem to be redundant since it does not contribute to the forecasting results. Moreover, MASE indicated that TM(1) performed better compared to TM(2). In the light of these facts, TM(1) might be the preferred model in terms of forecasting the independent variables during our simulation studies. These results are in agreement with the ones reported in Aslan (2010) in which the results were based only on mean squared error (MSE). Here we shall note that the aforementioned unexpected and interesting observations in MASE results, specifically decreases in the standard errors of MASE for TM(2), would not effect our conclusion about preferring TM(1) to forecast the independent variables. However, these might be investigated in a future study in details.

7.2.2 Dependent Variables

After forecasting the independent variables, we considered forecasting the bivariate longitudinal binary responses, Y_1 and Y_2. Based on the forecasting performances of the models on MSCM data, we selected univariate marginal models (UMM), multivariate marginal models with response specific parameters (MMM1), marginalized multivariate random effects models with forecasting methodology of considering the columns of $(T+1)$ to $(T+m)$ of $\hat{\Sigma}^{1/2}$ via using the $\hat{z}_i$'s (MMREM2), marginalized multivariate random effects models with forecasting methodology of considering all the subjects are average ones, i.e. $\hat{z}_i = 0$ (MMREM4), first order probit normal marginalized transition random effects models with selecting $cutoff = 0.5$ to dichotomize the $\hat{p}_{itj}$'s and using the $\hat{z}_i$'s (PNMTREM1), and its reduced form of assuming $\hat{z}_i = 0$ (PNMTREM2). We did not consider multivariate marginal models with shared regression parameters (MMM2) for the simulation study on forecasting, since the results of MMM1 and MMM2 seemed to be very similar for the MSCM data. Moreover, model fitting of MMM2 requires more computational time compared to MMM1 (for comparison of computational times of these models, see Chapter 3 of this thesis). On the other hand, the **mmm** package (Asar and Ilk, 2012a) already fits UMM's for each of the multiple responses, i.e. both MMM1 and UMM's are fitted at the same time. MMM1 and UMM's were fitted under unstructured working correlation matrix assumption.

Simulation results on forecasting bivariate longitudinal binary response data are presented in Tables 7.4 and 7.5 for Y_1 and Y_2, respectively. The simulation study were replicated as 100 times and one replication (the last one) took 11.7 minutes on a PC with 4.00 GB RAM and 2.53 GHz processor. Mean and standard error (SE) of the expected proportion of correct prediction (ePCP), area under the receiver operating characteristic (AUROC) curve and negative log-likelihood (NLL) values of these 100 simulation replications are displayed in these tables.

In the model building period, times 1 to 4, MMREM2 and PNMTREM1 seem to be the best methods in terms of all accuracy measures. For instance, for Y_1 while the ePCP, AUROC and NLL values of the former model were found to be 0.811, 0.924 and 814.133, these values for the latter model were found to be 0.766, 0.905 and 784.949 (Table 7.4). These results seem to be even better for Y_2 especially for MMREM2, the corresponding ePCP, AUROC and NLL values for MMREM2 were found to be 872, 966 and 505.515 (Table 7.5). For forecasting time periods, different accuracy measures indicate different rankings in terms of model performances. Specifically, ePCP indicates that while MMREM2 & MMREM4 are the best performed methods, PNMTREM1 and PNTMREM2 are the worst ones. For instance, for Y_2 at time 7, while the ePCP of MMREM2 was found to be 0.745, it was found to be 0.608 for UMM & MMM1, and 0.553 and 0.543 for PNMTREM1 & PNMTREM2, respectively (Table 7.5). Nonetheless, for one-step forecasts regarding the model building period, i.e. time=5, PNMTREM's seem to perform similar to UMM and MMM1. For instance, for Y_1, while the ePCP's of UMM & MMM1 were found to be 0.625 and 0.626, respectively, the ePCP's of PNMTREM1 & PNMTREM2 were found to be 0.639 and 0.632, respectively. In terms of AUROC, MMREM2 and PNMTREM1 seem to be the best models. For instance, for Y_2 at time=8, the AUROC values of these models were found to be 0.745 and 0.776, respectively (Table 7.5). In one-step forecasts regarding the model building period, i.e. time=5, PNMTREM1 and PNMTREM2 seem to be better than MMREM2. For instance, for Y_1, the AUROC values were 0.891 and 0.884 for PNMTREM1 & PNMTREM2, this value were found to be 0.812 for MMREM2 (Table 7.4). In general, the other models seem to be performing similar. For instance, for Y_1 the AUROC values of UMM, MMM1, MMREM4 and PNMTREM2 were found to be 0.713, 0.713, 0.712 and 0.714, respectively (Table 7.4). In terms of NLL, while UMM and MMM1 seem to be the best models, PNMTREM's seem to be the worst ones. For instance, the NLL values and their standard errors were found to be 1178.890 and 33.718 and 1179.806 and 34.119 for UMM and MMM1 for Y_2 at the overall forecasting period, i.e. times 5 to 8 (Table 7.5). On the other hand, related values of PNMTREM1 and PNMTREM2 were found to be 5275.101 and 382.435 and 5585.716 and 389.440, respectively. The standard errors of ePCP and AUROC values for all of the models seem to be similar. Forecasting results of Y_2 (Table 7.5) seem to better than the ones for Y_1 (Table 7.4).

Table 7.4: Forecasting results of Y_1 over 100 replications.

		ePCP		AUROC		NLL	
Time	Model	Mean	SE	Mean	SE	Mean	SE
1 to 4	UMM (Uns)	0.648	0.011	0.810	0.013	1056.247	29.948
	MMM1 (Uns)	0.649	0.011	0.810	0.014	1056.657	30.073
	MMREM2	**0.811**	0.011	**0.924**	0.008	**814.133**	64.581
	MMREM4	0.721	0.013	0.810	0.014	2042.314	322.925
	PNMTREM1	**0.766**	0.009	**0.905**	0.008	**784.949**	31.626
	PNMTREM2	0.740	0.010	0.870	0.009	948.175	36.846
5	UMM (Uns)	**0.625**	0.012	0.778	0.019	**282.462**	**9.865**
	MMM1 (Uns)	**0.626**	0.012	0.778	0.019	**282.655**	**9.991**
	MMREM2	**0.720**	0.016	**0.812**	0.017	450.774	74.342
	MMREM4	**0.694**	0.017	0.778	0.019	554.678	94.879
	PNMTREM1	**0.639**	0.016	**0.891**	0.015	644.336	60.145
	PNMTREM2	**0.622**	0.016	**0.884**	0.016	676.016	59.049
6	UMM (Uns)	0.605	0.011	0.747	0.020	**297.783**	**10.007**
	MMM1 (Uns)	0.605	0.011	0.747	0.020	**298.047**	**10.119**
	MMREM2	**0.708**	0.016	**0.800**	0.018	441.371	68.156
	MMREM4	**0.670**	0.016	0.747	0.020	584.209	99.749
	PNMTREM1	0.566	0.020	**0.802**	0.019	**1465.514**	**121.154**
	PNMTREM2	0.555	0.020	0.748	0.021	**1573.465**	**122.980**
7	UMM (Uns)	0.585	0.012	**0.713**	0.025	**313.033**	**10.855**
	MMM1 (Uns)	0.585	0.012	**0.713**	0.025	**313.406**	**10.995**
	MMREM2	**0.683**	0.018	**0.768**	0.023	484.794	73.437
	MMREM4	**0.644**	0.019	**0.712**	0.024	616.905	109.443
	PNMTREM1	0.552	0.021	**0.764**	0.023	**1548.175**	**125.661**
	PNMTREM2	0.544	0.020	**0.714**	0.025	**1638.187**	**127.634**
8	UMM (Uns)	0.569	0.012	0.683	0.027	**324.399**	**10.963**
	MMM1 (Uns)	0.569	0.012	0.683	0.027	**324.880**	**11.101**
	MMREM2	**0.658**	0.019	**0.734**	0.026	553.458	88.941
	MMREM4	**0.622**	0.020	0.683	0.027	635.427	115.898
	PNMTREM1	0.547	0.021	**0.755**	0.029	**1576.690**	**136.890**
	PNMTREM2	0.539	0.020	0.706	0.029	**1663.865**	**134.937**
5 to 8	UMM (Uns)	0.596	0.011	0.732	0.019	**1217.676**	**32.759**
	MMM1 (Uns)	0.596	0.010	0.732	0.019	**1218.988**	**33.210**
	MMREM2	**0.692**	0.013	**0.779**	0.016	1930.396	280.700
	MMREM4	**0.658**	0.015	0.732	0.018	2391.219	404.287
	PNMTREM1	0.573	0.016	**0.801**	0.016	**5304.666**	**398.245**
	PNMTREM2	0.563	0.016	0.761	0.017	**5613.059**	**400.303**

Note: Uns denotes unstructured working correlation matrix assumption.

Table 7.5: Forecasting results of Y_2 over 100 replications.

Time	Model	ePCP		AUROC		NLL	
		Mean	SE	Mean	SE	Mean	SE
1 to 4	UMM (Uns)	0.681	0.013	0.842	0.013	976.192	36.695
	MMM1 (Uns)	0.681	0.013	0.842	0.014	976.539	36.795
	MMREM2	**0.872**	0.016	**0.966**	0.009	**505.515**	72.736
	MMREM4	0.752	0.013	0.842	0.013	2312.382	403.725
	PNMTREM1	**0.776**	0.011	**0.917**	0.009	**734.509**	38.892
	PNMTREM2	0.751	0.012	0.883	0.011	890.456	44.784
5	UMM (Uns)	0.655	0.013	0.809	0.018	**266.210**	**10.586**
	MMM1 (Uns)	0.656	0.013	0.809	0.018	**266.338**	**10.619**
	MMREM2	**0.758**	0.016	**0.851**	0.016	479.819	91.508
	MMREM4	**0.722**	0.017	0.809	0.018	651.489	125.151
	PNMTREM1	0.647	0.017	**0.900**	0.013	624.577	59.599
	PNMTREM2	0.630	0.017	**0.892**	0.014	654.605	59.293
6	UMM (Uns)	0.631	0.012	0.774	0.019	**286.926**	**11.362**
	MMM1 (Uns)	0.631	0.012	0.774	0.019	**287.129**	**11.425**
	MMREM2	**0.745**	0.018	**0.833**	0.018	496.702	90.394
	MMREM4	**0.695**	0.018	0.773	0.019	712.735	136.030
	PNMTREM1	0.572	0.022	**0.824**	0.018	**1443.499**	**123.252**
	PNMTREM2	0.560	0.022	0.774	0.019	**1551.711**	**127.501**
7	UMM (Uns)	0.608	0.012	0.738	0.021	**305.214**	**10.955**
	MMM1 (Uns)	0.608	0.012	0.738	0.021	**305.481**	**11.051**
	MMREM2	**0.714**	0.020	**0.792**	0.022	600.063	99.422
	MMREM4	**0.668**	0.018	0.738	0.021	762.594	132.258
	PNMTREM1	**0.553**	0.022	**0.784**	0.020	**1549.836**	**123.364**
	PNMTREM2	**0.544**	0.022	0.738	0.021	**1640.916**	**124.542**
8	UMM (Uns)	0.588	0.013	0.705	0.024	**320.541**	**12.026**
	MMM1 (Uns)	0.588	0.013	0.705	0.025	**320.859**	**12.149**
	MMREM2	**0.676**	0.023	**0.745**	0.028	763.340	136.593
	MMREM4	**0.641**	0.020	0.705	0.025	801.863	141.667
	PNMTREM1	0.549	0.022	**0.776**	0.028	**1577.819**	**137.404**
	PNMTREM2	0.540	0.022	0.729	0.028	**1667.327**	**136.041**
5 to 8	UMM (Uns)	0.621	0.011	0.758	0.016	**1178.890**	**33.718**
	MMM1 (Uns)	0.621	0.011	0.758	0.016	**1179.806**	**34.119**
	MMREM2	**0.723**	0.015	**0.805**	0.016	2339.924	380.648
	MMREM4	**0.682**	0.015	0.758	0.016	2928.681	513.199
	PNMTREM1	0.577	0.017	**0.817**	0.013	**5275.101**	**382.435**
	PNMTREM2	0.566	0.017	0.780	0.014	**5585.716**	**389.440**

Note: Uns denotes unstructured working correlation matrix assumption.

7.3 Discussion and Conclusion

In this chapter, we presented a simulation study on forecasting bivariate longitudinal binary data. We specifically considered a model independent data generation scenario from multivariate normal distribution. We further considered forecasting the independent variables in addition to the bivariate responses. Results for independent variables showed that first-order transition models (TM(1)) might be preferred to the second order transition models (TM(2)) to forecast the independent variables of the datasets with within-independent variable correlation structures like the one we considered while generating our data. Forecasting results regarding the responses mostly coincide with the ones obtained while forecasting the MSCM data. Different accuracy measures indicate different model rankings. In general, marginalized models, especially MMREM with using the estimates of individual characteristics, $\hat{z}_i$, (i.e. MMREM2) seem to yield better forecasting results. For one-step forecasting, i.e. at $(T + 1)$, PNMTREM might yield better forecasting results compared to MMREM. Decreasing performance of PNMTREM at later time points is due to dichotomizing the forecasted success probabilities and treating them as the observed ones while forecasting the responses of the next time point. Although for MSCM data, for some cases the marginalized models with assuming $\hat{z}_i = 0$ might yield better performance compared to their complete model specification versions, i.e. use of $\hat{z}_i$, simulation results showed that the latter always performs at least as good as the former. Univariate and multivariate marginal models (UMM and MMM1, respectively) seem to be performing very similar in terms of forecasting as in the case of forecasting with MSCM data. Although the forecasting results seem to be better for Y_2, i.e. higher ePCP and AUROC values and lower NLL values compared to Y_1, this does not change the model rankings. This is most probably due to the fact that Y_2 has a larger variance compared to Y_1, since the only difference between these variables is in terms of their variances. Specifically, having a larger variance might yield larger correlations between the independent variables. Actually such patterns were observed during our trial and errors for setting the correlation and variance structures of the variables. In contrast to the forecasting results of MSCM data, in the simulation study all the accuracy measures are getting worse as the time lag increases.

Forecasting the independent variables affects the forecasting results of the dependent variables with no doubt. Aslan (2010) reported that this yields only decreases in the accuracy measures but do not yield changes in the model rankings. Therefore, we did not consider the use of observed independent variables for the forecasting time period. We only considered one model generation scenario. Further simulation studies regarding different model generation scenarios in terms of different and diverse correlation structures and datasets with more than two longitudinal responses could be studied.

CHAPTER 8

CONCLUSION AND DISCUSSION

This chapter can be regarded as an overview of the all thesis, since all the chapters have their own conclusion and discussion parts and we prefer to avoid unnecessary repetitions. Here, we briefly review what this thesis contributed mainly to the longitudinal data literature, and what can be done in future.

In Chapter 1, we provided an extended overview of longitudinal data regarding several aspects, such as main differences of this type of data from cross-sectional and time series data, dependent and independent variable types of it, importance of related missing data, and the related models - marginal, transition, random effects and marginalized multilevel models. An extended and up to date literature were provided in Chapter 1 as well. Two different real life datasets, Mother's Stress and Children's Morbidity (MSCM) and Iowa Youth and Families Project (IYFP) datasets, which were considered throughout the thesis were discussed in Chapter 1 for motivational purposes.

In Chapter 2, we began by introducing univariate and multivariate longitudinal data types to motivate our reader, since longitudinal data analysis is somehow new for many researchers of statistics and related fields. In Chapter 2, we provided an extended literature and discussion for marginal models together with quasi-likelihood inference (Wedderburn, 1974) and generalized estimating equations (GEE; Liang and Zeger, 1986). A marginalized multilevel model, namely marginalized multivariate random effects models (MMREM, Lee et al., 2009) were presented in Chapter 2 as well. Our discussion about MMREM was indeed a detailed and through discussion which made some of the aspects of the model easier to understand. We contributed this model by deriving its empirical Bayes estimator of individual characteristics, z_i and providing related R code. We can surely say that our reader most probably do not need to look at another reference while reading our longitudinal data structure, marginal model, quasi-likelihood inference, GEE and MMREM discussions. It would be beneficial providing some examples about data applications of marginal models with GEE in Chapter 2, but we did not prefer this due to the fact that almost all the references include such examples and we directed our reader to these references. Nonetheless, some examples of these models on MSCM data could be found in Chapters 3 & 6 of this thesis.

Marginal models are very popular among the longitudinal data analysts, especially while working with univariate longitudinal discrete data. In the present day, datasets are becoming very intensive and include lots of dependent variables. Therefore, multivariate models are becoming the essential tools for data analysis. In the light of these facts, multivariate marginal models for multivariate longitudinal data are an open area to be worked on. In Chapter 3, we considered two different multivariate marginal models, multivariate marginal models with response specific and shared regression parameters, MMM1 and MMM2, respectively. These models have broad perspectives such as they are not restricted to any response family, but can handle several of them such as Binomial, Poisson, Gaussian and Gamma families. Two R packages, **mmm** and **mmm2**, were proposed to fit these models, among which the latter is available from CRAN and the former will be publicly available from CRAN after the related paper is published. Data analysis under missing data is one of the issues of real life datasets, especially real life longitudinal datasets. Unfortunately, our multivariate marginal models are only robust under missing completely at random data. A natural study could be on extension of these models to missing at random and missing not at random data, in terms of improving our works on these models.

Marginalized multilevel models are becoming very popular to analyze longitudinal data due to their outstanding advantages in terms of including several traditional longitudinal data models, such as marginal, transition and/or random effects models, at the same time in different levels and being more robust to dependence structure misspecification, compared to their single level counterparts such as random effects models. Yet, many perspectives of these models are waiting to be studied on. In Chapter 4, we mainly built on the marginalized transition random effects models (MTREM) of Ilk and Daniels (2007) in terms of link functions and parameter estimation procedure and developed probit normal marginalized transition random effects models (PNMTREM). Our model accommodate both within and between response dependencies and permit population averaged and subject specific inference at the same time.

The model features were illustrated with an application to Iowa Youth and Families Project (IYFP) data. We proposed the use of implicit function theorem to explicitly link levels of marginalized models with transition structures for the first time. Furthermore, we eased the computational burden of MTREM caused due to logit link function and Bayesian Methods such as Markov Chain Monte Carlo, via use of probit link and maximum likelihood estimation. Last but not least, an R package **pnmtrem** is proposed to fit the model and the package will be available from CRAN after the related paper is published. To the best of our knowledge, our R packages are the first ones that are ever proposed for multivariate longitudinal data.

Forecasting with longitudinal data is a rarely studied topic, although it is essentially needed. As is mentioned in Chapter 5, it might increase the quality of life and even save life. We mainly considered forecasting multivariate longitudinal binary data via five different models including univariate and multivariate and marginalized models. To the best of our knowledge, our forecasting study is the first one which considers multivariate longitudinal binary data. In Chapter 5, we mainly discuss the need of forecasting studies with a longitudinal data perspective, and review the previous works related to longitudinal data forecasting. In Chapter 5, we further give the details of the forecasting methodologies which we followed during our forecasting studies. Accuracy measures are the essential tools of evaluating forecasting performances of different methods. In Chapter 5, we provided the details of the accuracy measures which we consider in our forecasting studies as well. Again to the best of our knowledge, previous forecasting literature except Aslan (2010) assumed that the independent variables in the forecasting period are known, which is not reasonable for real life in which similar to the dependent variables the independent variables are unknown as well. In this study, we considered forecasting the independent variables as well.

MSCM data were used to compare the forecasting abilities of our models in real life and the related modeling and forecasting results were provided in Chapter 6. A simulation study was conducted to assess the comparison of forecasting ability of the aforementioned five models and the results were provided in Chapter 7. Since the independent variables in the MSCM data were all time-independent, we did not consider forecasting them. On the other hand, we considered forecasting these variables in our simulation study on forecasting. In this simulation study, following Aslan (2010) we considered the generation of datasets which are not based on a model. We believe that model independent data simulation would permit fairer competition conditions. Results showed that more complex models might provide improved forecasts.

To sum up, an MS thesis with a broad perspective were prepared and illustrated here. Nonetheless, there are still lots of work to do. For instance, forecasting can be considered as a missing data imputation methodology. Automatic longitudinal data forecasting tools such as R packages for this purpose could be developed, such as Hyndman and Khandakar (2008) did. Actually, we prepared the R codes for forecasting with longitudinal data regarding the models mentioned in this study and they will be proposed in an R package in due course, since related process needs some time and effort. This would be very helpful especially for applied researchers. Forecasting of independent variables in longitudinal studies could be studied in more details, especially in real life datasets. In this thesis, we could not be able to consider forecasting such variables in real life datasets, since MSCM data did not include any time-dependent covariate. Similarly, simulation studies on forecasting of independent variables with several scenarios regarding the correlation structures of them would be very useful. Marginalized models for multivariate mixed responses could be developed such as a mix of binary and continuous longitudinal responses, since multiple responses do not need to belong necessarily to the same response family. However, this might most probably be a general research topic, rather than specific to marginalized models. Our research on accuracy measures showed that there is still need for such measures for binary data. Alternative measures could be developed for the evaluation of binary data models. Our research during this thesis taught us that model selection is an important aspect of statistical modeling which is somehow neglected, for instance for marginal models with GEE and random effects models. Further work on model selection is an on-going work.

REFERENCES

Abramowitz, K. M., Stegun, I. A. (1972). *Handbook of Mathematical Functions*. New York: Dover Publications.

Adams, R. A., Essex, C. (2010). *Calculus: A Complete Course, seventh edition*. Toronto: Pearson Publications.

Agresti, A. (2002). *Categorical Data Analysis, second edition*. New Jersey: John Wiley & Sons.

Akaike, H. (1973). Information Theory and An Extension of The Maximum Likelihood Principle. *Proceedings of the Second International Symposium on Information Theory, B. N. Petrov and F. Csaki (eds), 267–281*. Budapest: Akademiai Kiado.

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

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

Asar, O., Ilk, O. (2012a). mmm: An R package for Multivariate Marginal Models. Submitted.

Asar, O., Ilk, O. (2012b). mmm2: Multivariate Marginal Models with Shared Regression Parameters. R package version 1.0. URL http://CRAN.R-project.org/package=mmm2.

Asar, O., Ilk, O. (2012c). OOmisc: Ozgur-Ozlem Miscellaneous. R package version 1.0. URL http://CRAN.R-project.org/package=OOmisc.

Aslan, Y. (2010). *Which Method Gives the Best Forecast for Longitudinal Binary Response Data?: A Simulation Study*. MS Thesis, Middle East Technical University.

Baadsgaard, N. P., Højsgaard, S., Gröhn, Y. T. and Schukken, Y. H. (2004). Forecasting Clinical Disease in Pigs: Comparing a Naive and a Bayesian Approach. *Preventive Veterinary Medicine, 64, 85–100*.

Ballinger, G. A. (2004). Using Generalized Estimating Equations for Longitudinal Data Analysis. *Organizational Research Methods, 7, 127–150*.

Baltagi, B. H. (2008). Forecasting with Panel Data. *Journal of Forecasting, 27, 153–173*.

Bartolucci, F., Farcomeni, A. (2009). A Multivariate Extension of the Dynamic Logit Model for Longitudinal Data Based on a Latent Markov Heterogeneity Structure. *Journal of American Statistical Association, 104, 816–831*.

Bose, K. S. (2008). *Numeric Computing in Fortran*. Alpha Science.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control, third edition*. New Jersey: Prentice Hall.

Bozdogan, H., Haughton, D. M. A. (1998). Information Complexity Criteria for Regression Models. *Computational Statistics & Data Analysis, 28, 51–76*.

Breslow, N. E., Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of American Statistical Association, 88, 125–134*.

By, K., Qaqish, B. F., Preisser, J. S. (2011). orth: Multivariate Logistic Regression Using Orthogonalized Residuals. R package version 1.5.1. URL http://CRAN.R-project.org/package=orth.

Caflish, R. E. (1998). Monte Carlo and quasi-Monte Carlo Methods. *Acta Numerica, 7, 1–49*.

Carey, V. J., Zeger, S. L., Diggle, P. (1993). Modeling Multivariate Binary Data with Alternating Logistic Regression. *Biometrika, 80, 517–526*.

Carey, V. J. (2011a). gee: Generalized Estimating Equation Solver. R package version 4.13-16; Ported to R by Thomas Lumley (versions 3.13, 4.4) and Brian Ripley (version 4.13). URL http://CRAN.R-project.org/package=gee.

Carey, V. J. (2011b). yags: Yet Another GEE Solver. R package version 6.1-10/r42. URL http://R-Forge.R-project.org/projects/yags/.

Carey, V. J., Wang, Y-G. (2011). Working Covariance Model Selection for Generalized Estimating Equations. *Statistics in Medicine, 30, 3117–3124*.

Chatfield, C. (1996). *The Analysis of Time Series: An Introduction*. London: Chapman & Hall/CRC.

Chatfield, C. (2005). Forecasting. In *Encyclopedia of Biostatistics, second edition*, Eds. Armitage, P., Colton, T. John Wiley & Sons.

Cook, D., Swayne, D. F. (2007). *Interactive and Dynamic Graphics for Data Analysis with R and GGobi*. New York: Springer - Verlag.

Conte, S. D., de Boor, C. (1980). *Elementary Numerical Analysis: An Algorithmic Approach, third edition*. New York: McGraw-Hill Publications.

Crouch, E. A. C., Spiegelman, D. (1990). The Evaluation of Integrals of the Form $\int_{-\infty}^{+\infty} f(t)exp(-t^2)$: Application to Logistic-Normal Models. *Journal of American Statistical Association, 85, 464–469*.

deLeeuw, J (1992). Introduction to Akaike (1973) ‘Information Theory and an Extension of the Maximum Likelihood Principle’. In *Breakthroughs in Statistics, Volume I*, Ed. Kotz, S., Johnson, N. L., pp. 599–609. New York: Springer - Verlag.

Demidenko, E. (2004). *Mixed Models: Theory and Applications*. New Jersey: John Wiley & Sons.

Diggle, P. J. (1990). *Time Series: A Biostatistical Introduction*. New York: Oxford University Press.

Diggle, P. J. (1997). Introduction to Liang and Zeger (1986) 'Longitudinal Data Analysis Using Generalized Linear Models'. In *Breakthroughs in Statistics, Volume III*, Ed. Kotz, S., Johnson, N. L., pp. 470–382. New York: Springer - Verlag.

Diggle, P. J., Heagerty, P., Liang, K-Y., Zeger, S. L. (2002). *Analysis of Longitudinal Data, second edition*. New York: Oxford University Press.

Diggle, P. J., Farewell, D., Henderson, R. (2007). Analysis of Longitudinal Data With Drop-Out: Objectives, Assumptions and a Proposal (with comments and a rejoinder by the authors). *Journal of the Royal Statistical Society, Series C (Applied Statistics), 56, 499-550*.

Erhardt, V. (2009). corcounts: Generate Correlated Count Random Variable. R package version 1.4. URL http://CRAN.R-project.org/package=corcounts.

Fawcett, T. (2006). An Introduction to ROC Analysis. *Pattern Recognition Letters, 27, 861–874*.

Fitzmaurice, G. M., Laird, N. M. (1993). A Likelihood-Based Method for Analyzing Longitudinal Binary Data. *Biometrika, 80(1), 141–151*.

Fitzmaurice, G. M., Laird, N. M., Ware, J. H. (2004). *Applied Longitudinal Analysis*. New York: John Wiley & Sons.

Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G. (2009). *Longitudinal Data Analysis*. New York: Chapman & Hall/CRC.

Fox, J., Weisberg, S. (2011). *An R Companion to Applied Regression, second edition*. Thousand Oaks, California: Sage Publications. URL http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Frees, E. W., Miller, T. W. (2004). Sales Forecasting Using Longitudinal Data Models. *International Journal of Forecasting, 20, 99–114*.

Gardiner, J. C., Luo, Z., Roman, L. A. (2009). Fixed Effects, Random Effects and GEE: What Are the Differences?. *Statistics in Medicine, 28, 221–239*.

Gentle, J. E., Härdle, W., Mori, Y. (2004). *Handbook of Computational Statistics*. Heidelberg: Springer - Verlag.

Gentle, J. E. (2009). *Computational Statistics*. New York: Springer - Verlag.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Hothorn, T. (2011). mvtnorm: Multivariate Normal and t Distributions. R package version 0.9-96. URL http://CRAN.R-project.org/package=mvtnorm.

Ghisletta, P., Spini, D. (2004). An Introduction to Generalized Estimating Equations and an Application to Assess Selectivity Effects in a Longitudinal Study on Very Old Individuals. *American Journal of Political Sciences, 45(2), 470-490*.

Gibbons, R. D., Hedeker, D. (1994). Application of Random-Effects Probit Regression Models. *Journal of Consulting and Clinical Psychology, 62, 285–296*.

Gibbons, R. D., Hedeker, D., Charles, S., Frisch P. (1994). A Random-Effects Probit Model for Predicting Medical Malpractice Claims. *Journal of the American Statistical Association, 89, 760–767*.

Gibbons, R. D., Hedeker, D. (1997). Random Effects Probit and Logistic Regression Models for Three-Level Data. *Biometrics, 53, 1527–1537*.

González, J., Tuerlinckx, F., De Boeck, P., Cools, R. (2006). Numerical Integration in Logistic-Normal Models. *Computational Statistics & Data Analysis, 51, 1535–1548*.

Griswold, M. (2005). *Complex Distributions, Hmmmm... Hierarchical Mixtures of Marginalized Multilevel Models*. PhD Thesis, Johns Hopkins University.

Gueorguieva, R. V., Agresti, A. (2001). A Correlated Probit Model for Joint Modeling of Clustered Binary and Continuous Responses. *Journal of American Statistical Association, 96(455), 1102–1112*.

Halekoh, U., Højsgaard, S., Yan, J. (2006). The R package geepack for Generalized Estimating Equations. *Journal of Statistical Software, 15(2), 1–11*.

Hagle, M. T., Glenn, E. M. II. (1992). Goodness-of-Fit Measures for Probit and Logit. *American Journal of Political Science, 36, 762–784*.

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

Harrell, F. E. Jr. and with contributions from many other users. (2010). Hmisc: Harrell Miscellaneous. R package version 3.8-3. URL http://CRAN.R-project.org/package=Hmisc.

Harris, R., Sollis, R. (2003). *Applied Time Series Modelling and Forecasting*. Chichester: John Wiley & Sons.

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. New York: Springer - Verlag.

Heagerty, J. P. (1999). Marginally Specified Logistic-Normal Models for Longitudinal Binary Data. *Biometrics, 45(3), 688–698*.

Heagerty, J. P., Zeger, S. L. (2000). Marginalized Multilevel Models and Likelihood Inference (with comments and a rejoinder by the authors). *Statistical Science, 15, 1–26*.

Heagerty, J. P., Kurland, B. F. (2001). Misspecified Maximum Likelihood Estimates and Generalised Linear Mixed Models. *Biometrika, 88(4), 973–985*.

Heagerty, J. P. (2002a). *Analysis of Longitudinal Data, second edition*. URL http://faculty.washington.edu/heagerty/Books/AnalysisLongitudinal/. Last access: 29/12/2011.

Heagerty, J. P. (2002b). Marginalized Transition Models and Likelihood Inference for Longitudinal Categorical Data. *Biometrics, 58, 342–351*.

Hedeker, D., Gibbons, R. D. (1994). A Random-Effects Ordinal Regression Model for Multilevel Analysis. *Biometrics, 50(4), 933–944*.

Hedeker, D., Gibbons, R. D. (1997). Random Effects Probit and Logistic Regression Models for Three-Level Data. Biometrics, 53, 1527–1537.

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

Heiberger, R. M. (2009). HH: Statistical Analysis and Data Display: Heiberger and Holland. R package version 2.1-32. URL http://CRAN.R-project.org/package=HH.

Herron, M. (1999). Postestimation Uncertainty in Limited Dependent Variable Models. *Political Analysis, 8, 83–98*.

Hin, L. Y., Wang, Y. G. (2009). Working Correlation Structure Identification in Generalized Estimating Equations. *Statistics in Medicine, 28, 642–658*.

Horrocks, J., Van Den Heuvel, M. J. (2009). Prediction of Pregnancy: A Joint Model for Longitudinal and Binary Data. *Bayesian Analysis, 4(3), 523–538*.

Hosmer, D. W., Hosmer, T., le Cessie, S., Lemeshow, S. (1997). A Comparison of Goodness-of-fit Tests for The Logistic Regression Model. *Statistics in Medicine, 16, 965–980*.

Hyndman, R. J. (2006). Another Look at Forecast-Accuracy Metrics for Intermittent Demand. *Foresight: The International Journal of Applied Forecasting, 4, 43–46*.

Hyndman, R. J., Koehler, A. B. (2006). Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting, 22, 679–688*.

Hyndman, R. J., Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software, 27, 1–22*.

Hyndman, R. J. (2010). Forecasting Overview. In *International Encyclopedia of Statistical Science*, Ed. Lovric, M. Springer - Verlag.

Hyndman, R. J., with contributions from Razbash, S. and Schmidt, D. (2012). forecast: Forecasting Functions for Time Series and Linear Models. R package version 3.21. URL http://CRAN.R-project.org/package=forecast.

Ibrahim, J. G., Molenberghs, G. (2009). Missing Data in Longitudinal Studies: A Review (with comments and a rejoinder by the authors). *Test, 18, 1–43*.

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

Ilk, O. (2008). *Multivariate Longitudinal Data Analysis: Models for Binary Response and Exploratory Tools for Binary and Continuous Response*. Saarbrücken: Verlag Dr. Müller (VDM).

Ilk, O. (2011a). *Multivariate Longitudinal Data Analysis: Models for Binary Response and Exploratory Tools for Binary and Continuous Response*. URL http://www.metu.edu.tr/˜oilk/mlda.html. Last access: 29/12/2011.

Ilk, O. (2011b). *R Yazılımına Giriş* [Introduction to R Language]. ODTÜ Yayıncılık [METU Press].

Kennedy, W. J., Gentle, J. E. (1980). *Statistical Computing*. New York: Marcel Dekker.

Kutner, M. H., Nachtsheim C. J., Neter J., Li, W. (2005). *Applied Linear Statistical Models, fifth edition*. New York: McGraw-Hill Irwin Publication.

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

Laird, N. M., Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics, 38, 963–974*.

Lange, K. (2010). *Numerical Analysis for Statisticans*. New York: Springer - Verlag.

Lee, K., Daniels M. J. (2008). Marginalized Models for Longitudinal Ordinal Data with Application to Quality of Life Studies. *Statistics in Medicine, 27(21), 4359–4380*.

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

Leisch, F. (2008). Creating R Packages: A Tutorial. *Technical Report 36, University of Munich - Department of Statistics*.

Lesaffre, E., Spiessens, B. (2001). On the Effect of the Number of Quadrature Points in a Logistic Random-Effects Model: An Example. *Journal of the Royal Statistical Society, Series C (Applied Statistics), 50(3), 325-335*.

Liang, K.-Y., Zeger, S. L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika, 73(1), 13–22*.

Little, R. J. A., Rubin, D. B. (2002). *Statistical Analysis with Missing Data, second edition*. New York: John Wiley & Sons.

Liu, Q., Pierce, D. A. (1994). A Note on Gauss-Hermite Quadrature. *Biometrika, 81, 624–629*.

Martus, P., Stroux, A., Jünemann, A. M., Korth, M., Jonas, J. B., Horn, F. K., Ziegler, A. (2004). GEE Approaches to Marginal Regression Models for Medical Diagnostic Tests. *Statistics in Medicine, 23, 1377–1398*.

McCullagh, P., Nelder, J. A. (1989). *Generalized Linear Models, 2nd edition*. London: Chapman & Hall/CRC.

McCulloch, C. E., Searle, S. .R, Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models, 2nd edition*. New Jersey: John Wiley & Sons.

Mills, J. E., Field, C. A., Dupuis, D. J. (2002). Marginally Specified Generalized Linear Mixed Models: A Robust Approach. *Biometrics, 58, 727–734*.

Molenberghs, G. (2010). Generalized Estimating Equations: Notes on the Choice of Working Correlation Structure. *Methods of Information in Medicine, 49, 419–420*.

Molenberghs, G., Kenward, M. G. (2007). *Missing Data in Clinical Studies*. Chichester: John Wiley & Sons.

Molenberghs, G., Kenward, M. G. (2010). Semi-Parametric Marginal Models for Hierarchical Data and Their Corresponding Full Models. *Computational Statistics & Data Analysis, 54, 585–597*.

Molenberghs, G., Verbeke, G. (2004). Meaningful Statistical Model Formulations for Repeated Measures. *Statistica Sinica, 14, 989–1020*.

Molenberghs, G., Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer - Verlag.

Molenberghs, G., Verbeke, G. (2007). Likelihood Ratio, Score, and Wald Tests In a Constrained Parameter Space. *The American Statistician, 61, 22–27*.

Neal, R. M. (1996). *Bayesian Learning from Neural Networks, Lecture Notes in Statistics 118*. New York: Springer - Verlag.

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

Pan, W. (2001). Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics, 57(1), 120–125*.

Pan, W. (2002). Goodness-of-fit Tests for GEE with Correlated Binary Data. *Scandinavian Journal of Statistics, 29, 101–110*.

R Core Development Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. URL http://www.R-project.org/.

Ratcliffe, S. J., Shults, J. (2008). GEEQBOX: A MATLAB Toolbox for Generalized Estimating Equations and Quasi-Least Squares. *Journal of Statistical Software, 25(14), 1–14*.

Robins, J. M., Rotnitzky, A., Zhao, L. P. (1995). Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. *Journal of American Statistical Association, 89, 846–866*.

Rotnitzky, A., Robins, J. M., Scharfstein O. (1998). Semiparametric Regression for Repeated Outcomes with Non-ignorable Non-response. *Journal of American Statistical Association, 93, 1321–1339*.

Sabo, R. T., Chaganty, N. R. (2010). What Can Go Wrong When Ignoring Correlation Bounds in the Use of Generalized Estimating Equations. *Statistics in Medicine, 29, 2501–2507*.

SAS Institute Inc (2009). *SAS/STAT User's Guide, Version 9.2, second edition*. Cary: SAS Institute Inc. URL http://www.sas.com/.

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

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics, 6(2), 461–464*.

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

Shults, J., Sun, W., Tu, X., Amsterdam, J., Hilbe, J. M., Ten-Have, T. (2009). A Comparison of Several Approaches for Choosing Between Working Correlation Structures in Generalized Estimating Equation Analysis of Longitudinal Binary Data. *Statistics in Medicine, 28, 2338–2355*.

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

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology), 64(4), 583–649*.

StataCorp LP (2011). *Stata 12 Longitudinal/Panel Data Reference Manual*. Texas: Stata Press.

Stiratelli, R., Laird, N., Ware, J. H. (1984). Random Effects Models for Serial Observations with Binary Responses. *Biometrics, 40, 961–971*.

The MathWorks Inc (2010). *MATLAB - The Language of Technical Computing, Version 1.10.0 (R2010a)*. Massachusetts: The MathWorks Inc. URL http://www.mathworks.com/products/matlab/.

Vaida, F., Blanchard, S. (2005). Conditional Akaike Information For Mixed-Effects Models. *Biometrika, 92, 351–370*.

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

Wedderburn, R. W. M. (1974). Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika, 61, 439–447*.

Wei, W. W. S. (1990). *Time Series Analysis: Univariate and Multivariate Methods*. Redwood City: Addison - Wesley Publishing.

Weiss, R. E. (2005). *Modeling Longitudinal Data*. New York: Springer - Verlag.

Wuertz, D., many others and see the SOURCE file (2011). fOptions: Basics of Option Valuation. R package version 2140.79. URL http://CRAN.R-project.org/package=fOptions.

Yalcinoz, Z. (2008). *A Simulation Study On Marginalized Transition Random Effects Models for Multivariate Longitudinal Binary Data*. MS Thesis, Middle East Technical University.

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

Zeger, S. L., Irizarry, R., Peng, R. D. (2006). On Time Series Analysis of Public Health and Biomedical Data. *Annual Review of Public Health, 27, 57–79*.

Ziegler A. (2011). *Generalized Estimating Equations, Lecture Notes in Statistics 204*. New York: Springer - Verlag.

Zorn, C. J. W. (2001). Generalized Estimating Equation Models for Correlated Data: A Review with Applications. *Journal of Educational and Behavioral Statistics, 29, 421–437*.

APPENDIX A

ABBREVIATION LIST

Table A.1: Abbreviation List

AIC	Akaike Information Criterion
AUROC	Area Under the Receiver Operating Characteristics Curve
BIC	Bayesian Information Criterion
CDF	Cumulative Distribution Function
CRAN	Comprehensive R Archive Network
CSD	Cross-Sectional Data
DIC	Deviance Information Criterion
ePCP	Expected proportion of Correct Prediction
ES	Exponential Smoothing
FN	False Negative
FP	False Positive
F-S	Fisher-Scoring Algorithm
GEE	Generalized Estimating Equations
GEE1	First-Order Generalized Estimating Equations
GEE2	Second-Order Generalized Estimating Equations
GHQ	Gauss-Hermite Quadratures
GLM	Generalized Linear Models
GLMM	Generalized Linear Mixed Models
ICOMP	Information Complexity
IFT	Implicit Function Theorem
IYFP	Iowa Youth and Families Project
LB	Lower Bound
LD	Longitudinal Data
MAR	Missing at Random
MCAR	Missing Completely at Random
MC	Monte Carlo
MCEM	Monte Carlo Expectation Maximization
MCMC	Markov Chain Monte Carlo
MC-NR	Monte Carlo Newton Raphson
MAE	Mean Absolute Error
MASE	Mean Absolute Scaled Error
MD	Missing Data
MdAE	Median Absolute Error

Table A.2: Abbreviation List Continued

MLCD	Multivariate Longitudinal Count Data
MLGD	Multivariate Longitudinal Continuous (Gaussian) Data
MLE	Maximum Likelihood Estimation
MMM	Multivariate Marginal Models
MMM1	Multivariate Marginal Models with Response Specific Regression Parameters
MMM2	Multivariate Marginal Models with Shared Regression Parameters
MMREM	Marginalized Multivariate Random Effects Models
MNAR	Missing not at Random
MSCM	Mother's Stress and Children's Morbidity Study
MSE	Mean Squared Error
MSLNM	Marginally Specified Logistic Normal Models
MTREM	Marginalized Transition Random Effects Models
N	Negative
NLL	Negative Log-Likelihood
N-R	Newton Raphson Root Finding Algorithm
P	Positive
PCP	Proportion of Correct Prediction
PMC	Proportion of Observations in the Model Category
PNMTREM	Probit-Normal Marginalized Transition Random Effects Models
PNMTREM(1)	First-Order Probit-Normal Marginalized Transition Random Effects Models
PRE	Proportional Reduction in Error
QIC	Quasi Information Criterion
RMSE	Root Mean Squared Error
ROC	Receiver Operating Characteristics
TM(1)	First Order Transition Model
TM (2)	Second Order Transition Model
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
TSD	Time Series Data
UB	Upper Bound
UMM	Univariate Marginal Models
VIF	Variance Inflation Factor

APPENDIX B

DERIVATIONS REGARDING PNMTREM(1)

B.1 Linking Level 2 and Level 3 of $t \geq 2$ Model

In Section 4.2, we claim the following

$$\int \Phi(\Delta^*_{itj} + \lambda_j b_{it}) f(b_{it}) db_{it} = \Phi\left(\frac{\Delta^*_{itj}}{\sqrt{1+\lambda_j^2\sigma_t^2}}\right)$$

where $b_{it} \sim N(0, \sigma_t^2)$ and $b_{it} = z_i\sigma_t$, $z_i \sim N(0, 1)$ The related proof which is modified from Griswold (2005), is given below. Let $W_i \perp z_i$, where $W_i \sim N(0, 1)$,

Then,

$$\begin{aligned}
W_i/(\lambda_j\sigma_t) &\sim N(0, (\lambda_j\sigma_t)^{-2}) \\
W_i/(\lambda_j\sigma_t) - z_i &\sim N(0, 1 + (\lambda_j\sigma_t)^{-2}) \\
\tfrac{W_i/(\lambda_j\sigma_t) - z_i}{\sqrt{1+(\lambda_j\sigma_t)^{-2}}} &\sim N(0, 1)
\end{aligned}$$

and

$$\begin{aligned}
\int \Phi(\Delta^*_{itj} + \lambda_j b_{it}) f(b_{it}) db_{it} &= \int_{-\infty}^{+\infty} \Phi(\Delta^*_{itj} + \lambda_j z_i \sigma_1)\phi(z_i) dz_i \\
&= \int_{-\infty}^{+\infty} P(W_i \leq \Delta^*_{itj} + \lambda_j z_i \sigma_t)\phi(z_i) dz_i \\
&= \int_{-\infty}^{+\infty} P\left(\frac{W_i/(\lambda_j\sigma_t) - z_i}{\sqrt{1 + (\lambda_j\sigma_t)^{-2}}} \leq \frac{\Delta^*_{itj}/(\lambda_j\sigma_t)}{\sqrt{1 + (\lambda_j\sigma_t)^{-2}}}\right)\phi(z_i) dz_i \\
&= P\left(\frac{W_i/(\lambda_j\sigma_t) - z_i}{\sqrt{1 + (\lambda_j\sigma_t)^{-2}}} \leq \frac{\Delta^*_{itj}/(\lambda_j\sigma_t)}{\sqrt{1 + (\lambda_j\sigma_t)^{-2}}}\right) = \Phi\left(\frac{\Delta^*_{itj}}{\sqrt{1 + (\lambda_j\sigma_t)^2}}\right)
\end{aligned}$$

B.2 Linking Levels of Baseline Model

In Section 4.2, we also claim the following

$$\int \Phi(\Delta^*_{i1j} + \lambda^*_j b_{i1}) f(b_{i1}) db_{i1} = \Phi\left(\frac{\Delta^*_{i1j}}{\sqrt{1+\lambda^{*2}_j \sigma^2_1}}\right)$$

where $b_{i1} \sim N(0, \sigma^2_1)$ and $b_{i1} = z_i \sigma_1$, $z_i \sim N(0, 1)$

The related proof is given below.

Let $W_i \perp z_i$, where $W_i \sim N(0, 1)$,

Then,

$$\begin{gathered}
W_i/(\lambda^*_j \sigma_1) \sim N(0, (\lambda^*_j \sigma_1)^{-2}) \\
W_i/(\lambda^*_j \sigma_1) - z_i \sim N(0, 1 + (\lambda^*_j \sigma_1)^{-2}) \\
\frac{W_i/(\lambda^*_j \sigma_1) - z_i}{\sqrt{1+(\lambda^*_j \sigma_1)^{-2}}} \sim N(0, 1)
\end{gathered}$$

and

$$\begin{aligned}
\int \Phi(\Delta^*_{i1j} + \lambda^*_j b_{i1}) f(b_{i1}) db_{i1} &= \int_{-\infty}^{+\infty} \Phi(\Delta^*_{i1j} + \lambda^*_j z_i \sigma_1) \phi(z_i) dz_i \\
&= \int_{-\infty}^{+\infty} P(W_i \leq \Delta^*_{i1j} + \lambda^*_j z_i \sigma_1) \phi(z_i) dz_i \\
&= \int_{-\infty}^{+\infty} P\left(\frac{W_i/(\lambda^*_j \sigma_1) - z_i}{\sqrt{1 + (\lambda^*_j \sigma_1)^{-2}}} \leq \frac{\Delta^*_{i1j}/(\lambda^*_j \sigma_1)}{\sqrt{1 + (\lambda^*_j \sigma_1)^{-2}}}\right) \phi(z_i) dz_i \\
&= P\left(\frac{W_i/(\lambda^*_j \sigma_1) - z_i}{\sqrt{1 + (\lambda^*_j \sigma_1)^{-2}}} \leq \frac{\Delta^*_{i1j}/(\lambda^*_j \sigma_1)}{\sqrt{1 + (\lambda^*_j \sigma_1)^{-2}}}\right) = \Phi\left(\frac{\Delta^*_{i1j}}{\sqrt{1 + (\lambda^*_j \sigma_1)^2}}\right)
\end{aligned}$$

B.3 Calculations of The First Partial Derivatives of The Log-Likelihood Function and Information Matrix for Baseline Model

Recall that $L_1(\theta_1|y_1)$ is the likelihood function for the baseline model. The related first partial derivatives of this log-likelihood function with respect to baseline parameters, θ_1, where $\theta_1 = (\beta^*, \lambda^*_j, c_1)$ and $c_1 = log(\sigma_1)$ and y_1 is the observed responses at baseline, are approximated by 20-point Gauss-Hermite quadratures. The related functional forms are given by

$$\frac{\partial \log\left(L_1(\theta_1|y_1)\right)}{\partial \theta_1} \approx \sum_{i=1}^{N} \frac{1}{h(Y_{i1}|\theta_1)} \frac{\partial h(Y_{i1}|\theta_1)}{\partial \theta_1} \tag{B.1}$$

where

$$h(Y_{i1}|\theta_1) \approx \sum_{q=1}^{20} w_q \; exp\left[\sum_{j=1}^{k}\left(Y_{i1j}\log\left(\Phi(d_{i1jq})\right) + (1 - Y_{i1j})\log\left(1 - \Phi(d_{i1jq})\right)\right)\right] \tag{B.2}$$

$$\frac{\partial h(Y_{i1}|\theta_1)}{\partial \theta_1} \approx \sum_{q=1}^{20} w_q \; \left\{\ell(Y_{i1}|\theta_1)\left\{\sum_{j=1}^{k}\left[\frac{\partial d_{i1jq}}{\partial \theta_1}\phi(d_{i1jq})\left(\frac{Y_{i1j} - \Phi(d_{i1jq})}{\left(\Phi(d_{i1jq})\right)\left(1 - \Phi(d_{i1jq})\right)}\right)\right]\right\}\right\} \tag{B.3}$$

$$d_{i1jq} = \sqrt{1 + \lambda_j^{*2} e^{2c_1}} \; (X_{i1j}\beta^*) + \lambda_j^* e^{c_1} \sqrt{2} \; z_q \tag{B.4}$$

In (B.3), $\ell(Y_{i1}|\theta_1)$ is given by

$$\ell(Y_{i1}|\theta_1) = exp\left[\sum_{j=1}^{k}\left(Y_{i1j}\log\left(\Phi(d_{i1jq})\right) + (1 - Y_{i1j})\log\left(1 - \Phi(d_{i1jq})\right)\right)\right] \tag{B.5}$$

Note that (z_q, w_q) for $q = 1, \ldots, 20$ are Gauss-Hermite quadrature points and weights, respectively which are available in Abramowitz and Stegun (1972). Also, note that $\frac{\partial d_{i1jq}}{\partial \theta_1}$ are illustrated below for necessary θ_1 components.

Moreover, recall that the expected baseline information matrix, $I(\theta_1)$, is calculated by

$$I(\theta_1) = \sum_{i=1}^{N} h(Y_{i1}|\theta_1)^{-2} \left(\frac{\partial h(Y_{i1}|\theta_1)}{\partial \theta_1}\right)\left(\frac{\partial h(Y_{i1}|\theta_1)}{\partial \theta_1}\right)^T \tag{B.6}$$

Here, $h(Y_{i1}|\theta_1)$ and $\frac{\partial h(Y_{i1}|\theta_1)}{\partial \theta_1}$ are given in (B.2, B.3), respectively.

Note that since λ_j^* are response specific parameters, the calculations of the first partial derivatives of the log-likelihood function and the information matrix for these parameters are different than the ones for β^* and c_1. The related details are provided below.

The first partial derivatives of the baseline log-likelihood function with respect to β^*, $\frac{\partial log(L_1(\beta^*|y_1))}{\partial \beta^*}$, are calculated by replacing θ_1 with β^* in (B.1). On the other hand, $\frac{\partial d_{i1jq}}{\partial \beta^*}$ (after the related replacement of θ_1 and β^*) is given by

$$\frac{\partial d_{i1jq}}{\partial \beta^*} = \sqrt{1 + \lambda_j^{*2} e^{2c_1}}(X_{i1j}) \tag{B.7}$$

The first partial derivatives of the baseline log-likelihood function with respect to λ_j^*, $\frac{\partial log\left(L_1(\lambda_j^*|y_1)\right)}{\partial \lambda_j^*}$, are calculated by replacing θ_1 with λ_j^* in (B.1). However, the form of $\frac{\partial h(Y_{i1}|\theta_1)}{\partial \lambda_j^*}$ is given by

$$\frac{\partial h(Y_{i1}|\lambda_j^*)}{\partial \lambda_j^*} = \sum_{q=1}^{20} w_q \left\{ \ell(Y_{i1}|\lambda_j^*) \left[\frac{\partial d_{i1jq}}{\partial \lambda_j^*} \phi(d_{i1jq}) \left(\frac{Y_{i1j} - \Phi(d_{i1jq})}{\left(\Phi(d_{i1jq})\right)\left(1 - \Phi(d_{i1jq})\right)} \right) \right] \right\} \tag{B.8}$$

where $\ell(Y_{i1}|\lambda_j^*)$ is calculated by replacing θ_1 with λ_j^* in (B.5) and d_{i1jq} is same with the expression given in (B.4). Additionally, $\frac{\partial d_{i1jq}}{\partial \lambda_j^*}$ is given by

$$\frac{\partial d_{i1jq}}{\partial \lambda_j^*} = (1 + \lambda_j^{*2} e^{2c_1})^{-1/2} \lambda_j^* e^{2c_1} (X_{i1j}\beta^*) + e^{c_1} \sqrt{2} z_q \tag{B.9}$$

The first partial derivative of the baseline log-likelihood function with respect to c_1, $\frac{\partial log(L_1(c_1|y_1))}{\partial c_1}$, is calculated by replacing θ_1 with c_1 in (B.1). Furthermore, $\frac{\partial d_{i1jq}}{\partial c_1}$ is given by

$$\frac{\partial d_{i1jq}}{\partial c_1} = (1 + \lambda_j^{*2} e^{2c_1})^{-1/2} \lambda_j^{*2} e^{2c_1} (X_{i1j}\beta^*) + \lambda_j e^{c_1} \sqrt{2} z_q \tag{B.10}$$

B.4 Calculations of The First Partial Derivatives of The Log-Likelihood Function and Information Matrix for $t \geq 2$ Model

Recall that $L_2(\theta_2|y_2)$ is the likelihood function for the $t \geq 2$ model. The related first partial derivatives of the log-likelihood function with respect to $t \geq 2$ parameters, θ_2, where $\theta_2 = (\beta, \alpha_{t,1}, \lambda_j, c_t)$ and $c_t = log(\sigma_t)$ and y_2 is the observed responses for $t \geq 2$, are approximated by 20-point Gauss-Hermite quadratures. The related functional forms are given by

$$\frac{\partial log\,(L_2(\theta_2|y_2))}{\partial\theta_2} \approx \sum_{i=1}^{N}\sum_{t=2}^{T}\frac{1}{h(Y_{it}|\theta_2)}\frac{\partial h(Y_{it}|\theta_2)}{\partial\theta_2} \tag{B.11}$$

where

$$h(Y_{it}|\theta_2) \approx \sum_{q=1}^{20} w_q\; exp\left[\sum_{j=1}^{k}\left(Y_{itj}log\left(\Phi(d_{itjq})\right) + (1-Y_{itj})log\left(1-\Phi(d_{itjq})\right)\right)\right] \tag{B.12}$$

$$\frac{\partial h(Y_{it}|\theta_2)}{\partial\theta_2} \approx \sum_{q=1}^{20} w_q\; \left\{\ell(Y_{it}|\theta_2)\left\{\sum_{j=1}^{k}\left[\frac{\partial d_{itjq}}{\partial\theta_2}\phi(d_{itjq})\left(\frac{Y_{itj}-\Phi(d_{itjq})}{\left(\Phi(d_{itjq})\right)\left(1-\Phi(d_{itjq})\right)}\right)\right]\right\}\right\} \tag{B.13}$$

$$d_{itjq} = \sqrt{1+\lambda_j^2 e^{2c_t}}[-(A_{itj}\beta_0 + B_{itj}\alpha_{t,10}) + A_{itj}\beta + \alpha_{t,1}(B_{itj} + Z_{itj}y_{it-1j})] + \lambda_j e^{c_t}\sqrt{2}z_q \tag{B.14}$$

In (B.13), $\ell(Y_{it}|\theta_2)$ is given by

$$\ell(Y_{it}|\theta_2) = exp\left[\sum_{j=1}^{k}\left(Y_{itj}log\left(\Phi(d_{itjq})\right) + (1-Y_{i1j})log\left(1-\Phi(d_{itjq})\right)\right)\right] \tag{B.15}$$

A_{itj} and B_{itj} in (B.14) are illustrated in (4.21). Note that (z_q, w_q) for $q = 1, \ldots, 20$ are Gauss-Hermite quadrature points and weights, respectively. Also, note that $\frac{\partial d_{i1tjq}}{\partial\theta_2}$ are illustrated below for $\theta_2 = (\beta, \alpha_{t,1}, \lambda_j, c_t)$.

Moreover, recall that the expected information matrix for $t \geq 2$ is calculated by

$$I(\theta_2) = \sum_{i=1}^{N}\left(\sum_{t=2}^{T}\frac{1}{h(Y_{it}|\theta_2)}\frac{\partial h(Y_{it}|\theta_2)}{\partial\theta_2}\right)\left(\sum_{t=2}^{T}\frac{1}{h(Y_{it}|\theta_2)}\frac{\partial h(Y_{it}|\theta_2)}{\partial\theta_2}\right)^T \tag{B.16}$$

Here, $h(Y_{it}|\theta_2)$ and $\frac{\partial h(Y_{it}|\theta_2)}{\partial\theta_2}$ are given in (B.12, B.13), respectively.

Note that since $\alpha_{t,1}$ and c_t are time specific and λ_j are response specific parameters, the calculation of the first partial derivatives of the log-likelihood function and the information matrix for these parameters are different than the ones for β. Below, we illustrate the related details.

The first partial derivatives of the log-likelihood function for β, $\frac{\partial log(L_2(\beta|y_2))}{\partial\beta}$, is obtained by replacing θ_2 with β in (B.11). On the other hand, $\frac{\partial d_{itjq}}{\partial\beta}$ is given by

$$\frac{\partial d_{itjq}}{\partial \beta} = \sqrt{1+\lambda_j{}^2 e^{2c_t}}(A_{itj}) \tag{B.17}$$

Similarly, the β components of the expected baseline information matrix could be obtained by replacing θ_2 with β in (B.16).

The first partial derivatives of the log-likelihood function with respect to $\alpha_{t,1}$, $\frac{\partial log\left(L_2(\alpha_{t,1}|y_2)\right)}{\partial \alpha_{t,1}}$, are given below.

$$\frac{\partial log\left(L_2(\alpha_{t,1}|y_2)\right)}{\partial \alpha_{t,1}} \approx \sum_{i=1}^{N} \frac{1}{h(Y_{it}|\alpha_{t,1})} \frac{\partial h(Y_{it}|\alpha_{t,1})}{\partial \alpha_{t,1}} \tag{B.18}$$

where $h(Y_{it}|\alpha_{t,1})$ and $\frac{\partial h(Y_{it}|\alpha_{t,1})}{\partial \alpha_{t,1}}$ are obtained by replacing θ_2 with $\alpha_{t,1}$ in (B.12, B.13), respectively. Additionally, $\frac{\partial d_{itjq}}{\partial \alpha_{t,1}}$ is given by

$$\frac{\partial d_{itjq}}{\partial \alpha_{t,1}} = \sqrt{1+\lambda_j{}^2 e^{2c_t}}(B_{itj} + Z_{itj} y_{it-1j}) \tag{B.19}$$

On the other hand, the $\alpha_{t,1}$ components of the expected information matrix have the following forms

$$I(\alpha_{t,1}) = \sum_{i=1}^{N} \left(\frac{1}{h(Y_{it}|\alpha_{t,1})} \frac{\partial h(Y_{it}|\alpha_{t,1})}{\partial \alpha_{t,1}} \right) \left(\frac{1}{h(Y_{it}|\alpha_{t,1})} \frac{\partial h(Y_{it}|\alpha_{t,1})}{\partial \alpha_{t,1}} \right)^T \tag{B.20}$$

where $h(Y_{it}|\alpha_{t,1})$ and $\frac{\partial h(Y_{it}|\alpha_{t,1})}{\partial \alpha_{t,1}}$ are obtained by replacing θ_2 with $\alpha_{t,1}$ in (B.12, B.13), respectively. Note that, $\frac{\partial d_{itjq}}{\partial \alpha_{t,1}}$ in $\frac{\partial h(Y_{it}|\alpha_{t,1})}{\partial \alpha_{t,1}}$ is given in (B.19).

The first partial derivatives of the log-likelihood function with respect to λ_j could be obtained by replacing θ_2 with λ_j in (B.11). Here, $\frac{\partial h(Y_{it}|\lambda_j)}{\partial \lambda_j}$ is obtained by

$$\frac{\partial h(Y_{it}|\alpha_{t,1})}{\partial \lambda_j} \approx \sum_{q=1}^{20} w_q \left\{ \ell(Y_{it}|\lambda_j) \left\{ \left[\frac{\partial d_{itjq}}{\partial \lambda_j} \phi(d_{itjq}) \left(\frac{Y_{itj} - \Phi(d_{itjq})}{\left(\Phi(d_{itjq})\right)\left(1-\Phi(d_{itjq})\right)} \right) \right] \right\} \right\} \tag{B.21}$$

Additionally, $\frac{\partial d_{itjq}}{\partial \lambda_j}$ is given by

$$\frac{\partial d_{itjq}}{\partial \lambda_j} = (1 + \lambda_j{}^2 e^{2c_t})^{-1/2} \lambda_j e^{2c_t} (-(A_{itj}\beta_0 + B_{itj}\alpha_{t,10}) + A_{itj}\beta + \alpha_{t,1}(B_{itj} + Z_{itj}y_{it-1j}) + e^{c_t}\sqrt{2}z_q \tag{B.22}$$

The λ_j components of the expected baseline information matrix can be calculated by replacing θ_2 with λ_j in (B.16). Note that the calculation of $\frac{\partial h(Y_{it}|\lambda_j)}{\partial \lambda_j}$ illustrated in (B.21) is also valid for the information matrix calculations of λ_j.

The functional forms of the first partial derivatives of the log-likelihood function with respect to c_t are same with the ones for $\alpha_{t,1}$ given in (B.18), but note that in these functions $\alpha_{t,1}$ are replaced with c_t. The main difference would be in the calculation of $\frac{\partial d_{itjq}}{\partial c_t}$ which is given by

$$\frac{\partial d_{itjq}}{\partial c_t} = (1 + \lambda_j{}^2 e^{2c_t})^{-1/2} \lambda_j{}^2 e^{2c_t} (-(A_{itj}\beta_0 + B_{itj}\alpha_{t,10}) + A_{itj}\beta + \alpha_{t,1}(B_{itj} + Z_{itj}y_{it-1j}) + \lambda_j e^{c_t}\sqrt{2}z_q \tag{B.23}$$

Similarly, the c_t components of the expected baseline information matrix have the same forms with the ones for $\alpha_{t,1}$ given in (B.20). In these functions, $\alpha_{t,1}$ are replaced with c_t as well.

B.5 Empirical Bayes Estimation of the Random Effects

To estimate the z_i, we need to solve (4.78). Since it does not provide close solutions to z_i, we need to apply Newton-Raphson Algorithm; hence second derivative of the log-posterior distribution of z_i, $f(z_i|\_)$ is needed.

Remember that

$$log(f(z_i|\_)) = \left\{ \sum_{t=1}^{T} \sum_{j=1}^{k} \left[Y_{itj}^c log\left(\Phi(d_{itj}^c)\right) + (1 - Y_{itj}^c) log\left(1 - \Phi(d_{itj}^c)\right) \right] \right\} - \frac{z_i^2}{2}$$

$$\frac{\partial log(f(z_i|\_))}{\partial z_i} = \left\{ \sum_{t=1}^{T} \sum_{j=1}^{k} \frac{\lambda_j^c \sigma_t^c \phi(d_{itj}^c)\left(Y_{itj}^c - \Phi(d_{itj}^c)\right)}{\Phi(d_{itj}^c)\left(1 - \Phi(d_{itj}^c)\right)} \right\} - z_i$$

where $d_{itj}^c = \Delta_{itj}^{*c} + \lambda_j^c \sigma_t^c z_i$ and $\Delta_{itj}^{*c} = (\Delta_{i1j}^*, \Delta_{itj}^*)$, $\lambda_j^c = (\lambda_j^*, \lambda_j)$, $\sigma_t^c = (\sigma_1, \sigma_t)$ and $Y_{itj}^c =$

(Y_{i1j}, Y_{itj}) and c corresponds to complete matrices and vectors. Then the second partial derivative of $log(f(z_i|\_))$ is as follows

$$\frac{\partial^2 log(f(z_i|\_))}{\partial z_i \partial z_i} = \sum_{t=1}^{T}\sum_{j=1}^{k} \frac{\left\{\lambda_j^c\sigma_t^c\phi'(d_{itj}^c)Y_{itj}^c - \left[\lambda_j^c\sigma_t^c\phi'(d_{itj}^c)\Phi(d_{itj}^c) + \lambda_j^c\sigma_t^c\phi(d_{itj}^c)\Phi'(d_{itj}^c)\right]\right\}\left\{\Phi(d_{itj}^c)\left(1-\Phi(d_{itj}^c)\right)\right\} - \left\{\lambda_j^c\sigma_t^c\phi(d_{itj}^c)Y_{itj}^c - \lambda_j^c\sigma_t^c\phi(d_{itj}^c)Y_{itj}^c\Phi(d_{itj}^c)\right\}\left\{\Phi'(d_{itj}^c)\left(1-\Phi(d_{itj}^c)\right) + \Phi(d_{itj}^c)\left(-\Phi'(d_{itj}^c)\right)\right\}}{\left(\Phi(d_{itj}^c)\left(1-\Phi(d_{itj}^c)\right)\right)^2} - 1$$

$$= \sum_{t=1}^{T}\sum_{j=1}^{k} \frac{\left\{\lambda_j^c\sigma_t^c\phi'(d_{itj}^c)Y_{itj}^c - \lambda_j^c\sigma_t^c\phi'(d_{itj}^c)\Phi(d_{itj}^c) - \lambda_j^c\sigma_t^c\phi(d_{itj}^c)\Phi'(d_{itj}^c)\right\}\left\{\Phi(d_{itj}^c)\left(1-\Phi(d_{itj}^c)\right)\right\} - \left\{\lambda_j^c\sigma_t^c\phi(d_{itj}^c)\left(Y_{itj}^c - \Phi(d_{itj}^c)\right)\right\}\left\{\Phi'(d_{itj}^c) - \Phi(d_{itj}^c)\Phi'(d_{itj}^c) - \Phi(d_{itj}^c)\Phi'(d_{itj}^c)\right\}}{\left(\Phi(d_{itj}^c)\left(1-\Phi(d_{itj}^c)\right)\right)^2} - 1$$

$$= \sum_{t=1}^{T}\sum_{j=1}^{k} \frac{\lambda_j^c\sigma_t^c\left[\left\{\phi'(d_{itj}^c)Y_{itj}^c - \phi'(d_{itj}^c)\Phi(d_{itj}^c) - \phi(d_{itj}^c)\Phi'(d_{itj}^c)\right\}\left\{\Phi(d_{itj}^c)\left(1-\Phi(d_{itj}^c)\right)\right\} - \left\{\phi(d_{itj}^c)\left(Y_{itj}^c - \Phi(d_{itj}^c)\right)\right\}\left\{\Phi'(d_{itj}^c)\left(1-2\Phi(d_{itj}^c)\right)\right\}\right]}{\left(\Phi(d_{itj}^c)\left(1-\Phi(d_{itj}^c)\right)\right)^2} - 1$$

$$= \sum_{t=1}^{T}\sum_{j=1}^{k} \frac{\lambda_j^c\sigma_t^c\left[\left\{\phi'(d_{itj}^c)\left(Y_{itj}^c - \Phi(d_{itj}^c)\right) - \phi(d_{itj}^c)\Phi'(d_{itj}^c)\right\}\left\{\Phi(d_{itj}^c)\left(1-\Phi(d_{itj}^c)\right)\right\} - \left\{\phi(d_{itj}^c)\left(Y_{itj}^c - \Phi(d_{itj}^c)\right)\right\}\left\{\Phi'(d_{itj}^c)\left(1-2\Phi(d_{itj}^c)\right)\right\}\right]}{\left(\Phi(d_{itj}^c)\left(1-\Phi(d_{itj}^c)\right)\right)^2} - 1$$

- $\phi(d_{itj}^c) = \phi(\Delta_{itj}^{*c} + \lambda_j^c\sigma_t^c z_i) = \frac{1}{\sqrt{2\pi}}exp\left(-\frac{(\Delta_{itj}^{*c} + \lambda_j^c\sigma_t^c z_i)^2}{2}\right)$
- $\phi'(d_{itj}^c) = \underbrace{\frac{1}{\sqrt{2\pi}}exp\left(-\frac{(\Delta_{itj}^{*c} + \lambda_j^c\sigma_t^c z_i)^2}{2}\right)}_{\phi(\Delta_{itj}^{*c} + \lambda_j^c\sigma_t^c z_i)} [-\underbrace{\left(\Delta_{itj}^{*c} + \lambda_j^c\sigma_t^c z_i\right)}_{d_{itj}^c}]\lambda_j^c\sigma_t^c = \lambda_j^c\sigma_t^c(-d_{itj}^c)\phi(d_{itj}^c)$
- $\Phi(d_{itj}^c) = \Phi(\Delta_{itj}^{*c} + \lambda_j^c\sigma_t^c z_i)$
- $\Phi'(d_{itj}^c) = \Phi'(\Delta_{itj}^{*c} + \lambda_j^c\sigma_t^c z_i) = \phi(\Delta_{itj}^{*c} + \lambda_j^c\sigma_t^c z_i)\lambda_j^c\sigma_t^c = \lambda_j^c\sigma_t^c\phi(d_{itj}^c)$

$$= \sum_{t=1}^{T} \sum_{j=1}^{k} \frac{\lambda_j^c \sigma_t^c \left[\left\{ -\lambda_j^c \sigma_t^c \phi(d_{itj}^c) d_{itj}^c \left(Y_{itj}^c - \Phi(d_{itj}^c) \right) \right\} \left\{ \Phi(d_{itj}^c) \left(1 - \Phi(d_{itj}^c) \right) \right\} - \left\{ \phi(d_{itj}^c) \left(Y_{itj}^c - \Phi(d_{itj}^c) \right) \right\} \left\{ \lambda_j^c \sigma_t^c \phi(d_{itj}^c) \left(1 - 2\Phi(d_{itj}^c) \right) \right\} \right]}{\left(\Phi(d_{itj}^c) \left(1 - \Phi(d_{itj}^c) \right) \right)^2} - 1$$

$$= \sum_{t=1}^{T} \sum_{j=1}^{k} \frac{\lambda_j^c \sigma_t^c \left[\left\{ -\lambda_j^c \sigma_t^c \phi(d_{itj}^c) \left(d_{itj}^c \left(Y_{itj}^c - \Phi(d_{itj}^c) \right) + \phi(d_{itj}^c) \right) \right\} \left\{ \Phi(d_{itj}^c) \left(1 - \Phi(d_{itj}^c) \right) \right\} - \lambda_j^c \sigma_t^c \phi(d_{itj}^c) \left\{ Y_{itj}^c - \Phi(d_{itj}^c) \right\} \left\{ 1 - 2\Phi(d_{itj}^c) \right\} \right]}{\left(\Phi(d_{itj}^c) \left(1 - \Phi(d_{itj}^c) \right) \right)^2} - 1$$

$$= \sum_{t=1}^{T} \sum_{j=1}^{k} \frac{-(\lambda_j^c \sigma_t^c)^2 \phi(d_{itj}^c) \left[\left\{ d_{itj}^c \left(Y_{itj}^c - \Phi(d_{itj}^c) \right) + \phi(d_{itj}^c) \right\} \left\{ \Phi(d_{itj}^c) \left(1 - \Phi(d_{itj}^c) \right) \right\} + \left\{ \phi(d_{itj}^c) \left(Y_{itj}^c - \Phi(d_{itj}^c) \right) \right\} \left\{ 1 - 2\Phi(d_{itj}^c) \right\} \right]}{\left(\Phi(d_{itj}^c) \left(1 - \Phi(d_{itj}^c) \right) \right)^2} - 1$$