THE UNIVERSITY OF CHICAGO

ECOLOGICAL MOMENTARY ASSESSMENT (EMA) DATA: STATISTICAL METHODS FOR HETEROGENEOUS VARIANCE, MISSING DATA AND LATENT STATE CLASSIFICATION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PUBLIC HEALTH SCIENCES

BY
XIAOLEI LIN

CHICAGO, ILLINOIS
AUGUST 2018

To My Husband, Puxuan

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my mentor, Dr. Donald Hedeker, for his skillful supervision and endless support, introducing me to the academic world of biostatistics, allowing me freedom to explore broad research interests and make my own mistakes, and ultimately leading me to establish a challenging yet bright career path. I also thank Dr. James Dignam, Dr. Robert Gibbons and Dr. Robin Mermelstein for serving on my dissertation committee and providing me with invaluable feedback. The financial and spiritual support from the Department of Public Health Sciences are greatly appreciated.

I would also like to thank all my office mates and fellow students for providing me with great companionship along my PhD journey, particularly Dayana and Eloesa. You have made my PhD journey full of happiness, courage and hope.

A special thanks to my family - my parents, my beloved husband and the newest addition, my son Isaac. This dissertation would not have been possible without their warm love, continued patience, and endless support.

ABSTRACT

Ecological Momentary Assessment (EMA) studies collect self-reported activities, behaviors and emotions intensively throughout the entire study span, and provide valuable information about how subjects' psychological activities evolve over time. While inheriting some similarities from the traditional longitudinal data, the abundance and intensity create unique characteristics for the EMA data and pose new challenges to the existing statistical analysis.

Statistical methodologies investigating the associations between risk factors and mood regulation in EMA studies have not been studied thoroughly, and there is recent evidence that mood variability, together with mood assessment level, are important metrics in understanding subjects tendency to problem behaviors. In this dissertation, a series of three studies was conducted to systematically investigate the effect of certain psychosocial factors on mood regulation in EMA studies, using both metrics of mood variability and mood assessment level. The novel statistical methods can be extended to more general frameworks where a broad spectrum of related statistical and substance problems can be solved.

The methods developed in this dissertation were motivated by an EMA adolescent mood study. First, a three level mixed effect location scale model that includes multiple random subject and wave effects in both the mean and within variance model was developed for hierarchical EMA data where subjects were measured at multiple waves, and at each wave, data were intensively collected over time on each subject. The proposed model allows heterogeneous variance at baseline as well as variance change over time, adjusting for observed covariates. Second, a shared parameter model was framed to address the non-ignorable missing responses in EMA studies, where missing indicator and longitudinal outcomes are jointly modeled by sharing common but unobserved subject level information. The missing indicator was modeled via random intercept logistic regression, and outcome by random location and scale intercept regression, with the three random effects all representing subject

specific traits. The model allows subject's missing propensities to influence his/her mood level and variability. Third, a mixed location scale Hidden Markov model was explored to classify subjects into distinct mood states at each time point, with latent mood states at sequential time points form a Markov Chain. This model allows differential effects on mood regulation for subjects with different mood states.

All models in the above studies were estimated via Bayesian sampling framework by Stan. The model estimation procedures are computational more efficient compared to the maximum likelihood based methods. Extensive simulation studies were conducted to validate the model performance and compare with the existing methods. The proposed models were applied to each of the motivating data set with interpretable results and insightful conclusions. Finally, a discussion of the advantages, limitations as well as future directions were included in the end of each method chapter.

CHAPTER 1

INTRODUCTION

1.1 Overview

Developed by personality/social psychologists, ecological momentary assessment (EMA) sampling methods allow researchers to study people's thoughts, emotions, and behaviors in their natural contexts. Typically they involve self-reports or data collection from individuals over the course of hours, days, and weeks. Such data will permit more sensitive assessments, enable wide-ranging and detailed measures of mood and behavior, and thus provide insights that are not possible using conventional experimental or survey research methods. For patients whose moods fluctuate and exhibit high volatility over time, ordinary data collection methods are not capable of providing an accurate mood estimate over the entire period by only focusing on the recent mood record. Moreover, data might not be collected accurately if family members are not able to report behaviors that he/she has not observed, or interviewers may begin an interview with expectations that preclude the solicitation of relevant information to inform rating judgements. All these limitations motivate the novel EMA data collection methods that relies on intermittent self-reporting data.

Often times, EMA studies are conducted at multiple waves, and within each wave, subjects are measured intensively at multiple occasions. The multi-wave EMA studies create unique hierarchical data structures with special correlation structures at the between subject, within subject between wave and within subject within wave level. These complex correlation structure and abundance of the data pose statistical challenges when analyzing the multi-wave EMA studies. In particular, failure to account for the intermediate wave clustering, ignoring the change rate of the outcome mean/variability over time, or omitting individual differences not explained by the measured covariates would render the models invalid for statistical inference. The major goal of the first study in this disertation is to develop a three level mixed effect location scale model that can overcome these difficulties and provide valid inference for the special data structure often encountered in research.

Since relatively large numbers of measurements per subject are produced during an EMA study, intermittent missingness due to non-responses is likely to occur. Subjects with a substantial proportion of non-responses can be systematically different in terms of the outcomes compared to those without, which can lead to potential issues of informative missing. Methods that only focus on the observed data would lead to invalid statistical inference as the observed sample is a biased representation of the population. The second study of the dissertation focuses on linking the primary outcome which are subject to informative missing with the missing data mechanism by sharing subject specific traits that summarize their own characteristics on the outcome mean and variability, thus correct the bias introduced by informative missingness.

Another major focus of this dissertation is to allow heterogeneous covariate effects on mood regulation for distinct subgroups characterized by subjects' latent mood states. Hidden Markov Models (HMMs) are able to perform subgroup classification based on subjects' latent mood states, therefore allow for model specifications with heterogeneous covariate effects. However, HMMs ignore individual differences on top of the time serial dependence that are both present in the data, and would result in inaccurate classification and thus incorrect effect estimates. Mixed HMMs, which incorporates random effects with HMMs, provide a framework to allow both time serial dependence and individual differences in characterizing heterogeneous covariate effects in mood regulation. The third study in this dissertation extended the mixed HMMs to accout for additional individual differences in within subject variability of the outcome and thus provide a comprehensive strategy for analyzing EMA data when research interests center around the variability of the intensive longitudinal outcomes.

The overall goal of this disertation is to develop comprehensive statistical methodology to better establish the associations between intensively measured risk behaviors and psychological activities, and ultimately shed light on early intervention cues. The three studies and the corresponding statistical methodologies were motivated by three distinct yet related psychological and behavioral research fields with similar data structure but unique modeling challenges. Novel statistical models were developed for each study, which can be generalized to a broad spectrum of research fields. In later sections, I will describe the background, data structure, and existing methods as well as our proposed methods for each research projects.

1.2 Background and Motivation

1.2.1 The Multi-wave Adolescent Smoking Study

Ecological Momentary Assessment (EMA) studies usually produce intensively measured longitudinal data with large numbers of observations per unit, and research interest is often centered around understanding the changes in variation of people's thoughts, emotions and behaviors. Often times, EMA studies are conducted at multiple waves and within each wave, subjects get intensively measured at multiple occasions, creating a three level data structure - occasions nested within waves and waves then nested within subjects, and complex correlation structure - between subjects, within subjects between waves and within subjects within waves correlation. It is important to understand how risk factors can change mood assessments at different levels as well as their trajectories over time.

The adolescent smoking study

The data that motivate the development of the Bayesian three level mixed effect location scale model are from an EMA adolescent smoking study. In the study, 461 adolescents from

9th and 10th grade were recuited. They carried hand held devices for 7 days at each measurement wave, during which they responded to random interviews (approximately 5 times per day) or event recorded any episodes of smoking. At each prompt or smoking episode, participates were asked to answer questions, including location, activities, companionship, mood and other psychological measurements. The study was conducted at 6 waves: baseline, 6 months, 15 months, 2 years, 5 years, and 6 years. Data were collected on age, gender, beep type (smoking event vs random prompt), and positive affect (measure of positive mood). Because of the interest in comparing responses across waves from random prompts vs smoking events, subjects were included if they were measured on at least two waves and had at least two smoking events at each wave, resulting in a sample size of 254 subjects.

Among all the subjects, 51.6% were female, and on average subjects were followed up at three waves with 36 to 51 prompts (including smoking episodes) per wave during the entire study span . A total of 24490 random prompts and 8087 smoking events were obtained, with an approximate average of 96 random prompts and 32 smoking episodes per subject. For the analyses reported, a three level structure of observations (level1) within waves (level 2) within subjects (level 3) was considered.

The outcome is the measure of subjects' positive affect (PA), which consists of the average of several mood items rated from 1 to 10: I felt happy, I felt relaxed, I felt cheerful, I felt confident, and I felt accepted by others. Thus, higher PA levels indicate better mood. The interest is to see whether subjects tend to have higher and more consistent positive affect after smoking compared to random prompts. We are also interested in differentiating the between subject and between subject within wave effect from the within subject within wave effect. That is, the effect of smoking when comparing the same subject at the same wave could be different from that when comparing different subjects, or the same subject across different waves.

Existing methods

Mixed effect regression models (MRM) have been used to analyze longitudinal outcomes. Consider the following MRM for mood assessment y of subject i ($i = 1, 2, ..., N$ subjects) at wave j ($j = 1, 2, ..., J_i$) and occasion k ($k = 1, 2, ..., K_{ij}$) :

$$Y_{ijk} = X_{ijk}^{\top} \ \beta + Z_{ijk}^{\top} \ \gamma_i + V_{ijk}^{\top} \ \nu_{ij} + \epsilon_{ijk}$$

where X_{ijk} is the $p \times 1$ vector of regressors (typically including a column of "1" for the intercept) which contains either subject level, wave level or time level variables, and β is the corresponding regression coefficients. Z_{ijk} is the vector of regressors for random effects γ (usually a subset of X_{ijk}), and γ_i is the subject level random effects, which indicates the influence of individual i on his/her repeated mood assessments. For a random intercept model, Z_{ijk} will be 1, while for a random intercept and slope model, Z_{ijk} will be two dimensional (1, time). Similarly, V_{ijk} is the vector of regressors for random effects ν, and ν_{ij} is the wave level random effects, which indicates the influence of wave on repeated mood assessments of subject i. The population distributions of both γ and ν are assumed to be independent, normal with zero mean and usually constant variance-covariance structure Σ_γ, Σ_ν. The errors ϵ_{ijk} are assumed to be independent and normally distributed with zero mean and variance σ_ϵ^2, and independent of all random effects. Here, Σ_γ represents the between subject variance, Σ_ν represents the between wave variance, and σ_ϵ^2 represents the within variance for all repeated measurements within the same subject at a single wave.

Most of the existing MRM models assume that σ_ϵ is constant for all subjects, i.e., subjects have the same variability for their repeated mood assessments at all waves. This assumption could be easily violated, especially for psychological and social science studies as subjects almost always exhibit strong pattern of heterogeneity at different waves. Previous studies have utilized models that allow the covariates to have an influence on the within variance. For example, several studies included gender and age as covariates in the within variance model when females show relatively unstable mood compared to males, and older people exhibit more variabilitity in their mood assessments.

$$\sigma^2_{\epsilon,ij} = \exp(W_{ijk}^\top \ \alpha)$$

where W_{ijk} is the vector of regressors (usually a subset of X_{ijk}) that has an effect on the within variance and α is the corresponding regression coefficients. Together with the previous model for y, both the mean and within variance can depend on observed covariates such as age, sex as well as prompt type.

However, this model assumes that mood variability for the same subject within a single wave can be fully explained by the set of covariates we observe. In observational studies, this is rarely the case as we almost always have confounding issue due to unmeasured variables.

Proposed methods

In Chapter 2, I will propose a Bayesian three level mixed effect location scale model to better analyze the multi-wave adolsecent smoking study data. The proposed model extends the traditional mixed effect regression model by including multiple random effects at both subject and wave level, and for both outcome mean and within subject within wave variance. The proposed model allows individual differences in the baseline mood as well as in the mood change over time, in terms of both mood level and consistency. Bayesian Markov Chain Monte Carlo (MCMC) sampling method was used in model estimation procedure, which serves as a better approach than likelihood based methods when multiple random effects are present.

Together with a series of simulation studies as well as real data application, I will illustrate the advantages of the proposed model over the existing (reduced) models. Results and conclusions on the effects of smoking at different levels on adolescent mood regulation will also be discussed at the end of Chapter 2.

1.2.2 The Adolescent Mood Study - Missing Data

In EMA studies, subjects' real time behavior and experiences are intensively and repeatedly sampled, resulting in a large number of observations per subject. Intermittent missingness is likely to occur due to non-responses (usually at early morning or late night) and therefore statistical methods need to be developed for various missing data scenarios.

There are several scenarios where data could be missing: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). In MCAR, missingness is independent of both the observed and missing outcome. In MAR, missingness could depend on the observed outcome, but is still independent of the missing outcome, whereas in MNAR, missingness could depend on missing outcomes. A good example of MNAR would be that subjects who had severe drug side effects did not complete the study that aimed to evaluate drug effects, making the statistical inference biased toward fewer adverse events. Most statistical programs deal with missing data by analyzing the observed data and can lead to bias since the available data might not be truely representative of the total population. Commonly used imputation methods including mean value imputation and multiple imputation assume data are missing (completely) at random, and thus not only decrease variability but can introduce bias due to non-ignorable missing mechanism.

Shared parameter models provide one appropriate approach in the case of MNAR by factoring the full data likelihood into two parts: a missing data model and a longitudinal outcome model that are linked by subject specific traits. It is crucial to adapt the shared parameter framework in the context of EMA studies and develop statistical models to recover the true associations between covariates and the outcome of interest.

The adolescent mood study

This research project is motivated by an EMA study investigating the effects of psychosocial factors on mood regulation among adolescents. The entire EMA study was conducted across 6 waves: baseline, 6 months, 15 months, 2 years, 5 years and 6 years. For illustration purposes, we will focus on data from the baseline wave.

At baseline, 461 adolescents (average age 15.6, minimum 14.4, maximum 16.7) from 9th and 10th grade were asked to carry electronic devices and answer questions when randomly prompted during a 7 day study period. Each individual was prompted multiple times within a single day. Questions included location, activities, companionship, mood and other psychological assessments. The primary outcomes of interest are positive affect (PA) as well as negative affect (NA), which consist of the average of several mood items rated from 1 to 10 that measures subject's positive/negative mood. For PA, questions include: I felt happy, I felt relaxed, I felt cheerful, I felt confident, and I felt accepted by others; for NA, questions include: I felt sad, I felt stressed, I felt angry, I felt frustrated, and I felt irritable. Higher PA levels indicates better mood while higher NA indicates worse mood. Each response will be time stamped regardless of missing status. For the analyses presented here, we will consider subject-level covariates of age, gender, smoking status, negative mood regulation, and the occasion-varying indicator of whether they were alone or with others at the time of the prompt.

Intermittent missingness was generated when an individual did not respond to the prompts. Dropout is not as big a concern here as more than 96.5% individuals were still available at the end of the seven day study period. On average, 22.3% of prompts were missing for each individual, with the highest missing proportion per individual being 89.7%. There were a fair number of prompts missing on each study day: 22.6% on day 1, 19.0% on day 2, 20.9% on day 3, 24.6% on day 4, 25.6% on day 5, 25.0% on day 6 and 23.9% on day 7. The proportions of missing prompts were relatively similar during weekdays. In terms of time of day, most missingness occured between 3 am to 9 am (27.2%) and least from 6 pm to 9 pm (20.8%), but the pattern on weekend days was different in that missingness occured mostly between 9 pm to 3 am (52.9% on Saturday and 73.1% on Sunday).

Existing methods

Most software packages throw out the missing observations and use only available data for analysis. However, the available data could be a biased sample of the total population and thus differ systematically in terms of race composition, smoking behavior and mood regulation. In the case of missing completely at random (MCAR) or missing at random (MAR), the available data might be able to recover the valid statistical inference about the population, possibly with less efficiency. On the other hand, if data are missing not at random (MNAR), as is often suspected in a real world situation, the available data analysis used by most statistical packages would introduce bias and yield invalid inference.

Some statistical models assume MCAR, where missingness only depends on observed covariates (e.g., smokers are more likely to be missing compared to non-smokers and smoking status is observed for all subjects at all time points). MRM is more robust to missing data since it assumes MAR, where missingness can depend on observed data (e.g., if a subject has bad mood, he/she will be missing at the next time point). Often times, missing mechanisms are more complicated than MCAR or MAR, since we cannot observed all relevant covariates, or missingness could depend on missing outcomes. For example, a subject's time schedule could influence his/her response rate in the middle of the day, but we have no measurement over his/her schedule, or subjects with bad mood are more likely to miss the response at that particular time point. In all these non-random scenarios, methods described above are no longer valid and more sophisticated statistical models should be developed.

The proposed methods

Motivated by the concept of location scale from the mixed effect location scale model originally proposed by Hedeker et al., I developed a shared parameter model framework designed specifically for EMA studies that are subject to intermittent and informative missingness. In Chapter 3, I will illustrate the detailed model specifications - a random intercept logistic regression model for the binary missing indicator, a mixed location scale model for the intensive longitudinal outcome, and a linear model that links these two models together. The proposed model extends the traditional shared parameter models by incorporating subject effects in the error variance and sharing those with the missing data mechanism, thus provide better and more comprehensive bias correction when data show evidence of individual differences in the outcome variability.

A series of simulation studies as well as real data application will also be conducted and the results are shown in Chapter 3. I will illustrate the advantages of shared parameter models over naive models where only observed data are used for analysis, as well as the advantages of sharing information on both outcome location and scale over the location only case. Results and conclusions on the effects of several covariates on adolescent mood regulation will also be discussed at the end of Chapter 3.

1.2.3 The Adolescent Mood Study - Latent State Classification

EMA studies produce multiple observations over time for each subject, which can also be considered as distinct time series. For sequential data (time series for the same subject), observation at time t is likely to be influenced by its previous value and it also influences the next values. Intuitively, it is reasonable to assume that recent observations are likely to be more informative than more historical observations in predicting future values. That is to say, observations at time t are more accurately predicted by those at time $t-1$ compared to $t-2$ or further away. However, it is also impractical to consider a general dependence of future observations on all previous observations because of the uncontrollable or unlimited complexity. In this sense, Markov models which assume future observations are independent of all but the most recent one would be a good choice.

In Hidden Markov Models (HMMs), one assumes there are sequential latent discrete variables z_1, z_2, ..., z_n, which follow a first order Markov chain, and sequential observed continuous variables x_1, x_2, ..., x_n, which are independent of each other and follow some known distributions conditional on the latent variable z. In the context of the EMA study on adolescent mood study, z_t can be considered as the (latent) mood state (pleasant vs unpleasant), and x_t can be considered as the observed mood assessment at time t. Since z takes on discrete values, subjects can be classified into distinct subgroups based on their observed measurements x. Using HMMs, we can specify various models that allow for heterogeneous covariate effects for different latent subgroups, which is less restrictive and makes more practical sense for research settings.

In psychological and behavioral science, research interests are often in identifying risk factors for despaired mood regulation. Individuals differ considerably in their ability to effectively regulate their mood and can thus be potentially classified into distinct subgroups based on mood regulation ability. It is reasonable to assume different covariate effects on mood regulation for individuals in different subgroups, which creates a perfect setting for HMMs. However, HMMs treat each time series as homogeneous and ignore the individual differences, which is a crucial characteristics of longitudinal data. Incorporating mixed effects in HMMs as a tool for latent state classification and future prediction, especially in the context of intensive longitudinal studies / EMA studies is of great importance and has not been studied thoroughly.

Existing methods

One way to analyze data with the above structure is to treat it as random realizations of one time series and use the ordinary Hidden Markov Model directly. This method ignores the fact that subjects are heterogeneous in many aspects that could influence the outcomes at each time point. Ignoring subject effects would result in biased inference, most noticeably, narrower confidence intervals.

Another common strategy is to utilize a two-stage method: first remove subject effects by a mixed effect model, and then apply ordinary Hidden Markov Model to the residuals which are supposed to be random realizations of time series. This strategy, however, ignores the possibility that the estimation for random effects parameters and HMM parameters could be dependent, i.e., estimating random effects variance depend on HMM parameter estimates and estimating HMM parameters also depends on random effects variance estimates. In addition to the biasness issue, this method also loses efficiency when the sample size is not large enough.

Altman proposed a Mixed Hidden Markov Model by extending the original Hidden Markov Model to longitudinal data setting and jointly estimating the random effect parameters and HMM parameters. However, this model only allows random subject effects in the mean model, ignoring the possibility that subjects could also exhibit heterogeneity in terms of within variance. Ignoring the scale random effects might bias the estimation of variance parameters.

The proposed methods

In the effort of incorporating individual differences into the HMMs, I propose to include bivariate subject random effects (in location and scale) in the conditional model of the observed process. Specifically, the latent states are assumed to form a first order Markov Chain given the random effects, and the observed measurements are independent of each other and follow some known distribution conditional on the random effects and latent states. Differential covariate effects are specified in both the mean model and within subject variance model for different subgroups.

In Chapter 4, I will illustrate the detailed model formulation of the proposed mixed location scale HMM, model estimation procedure, its advantages over simple HMM or mixed HMM and the applicability to real data sets. A discussion of the results and conclusions in applying the model to the adolescent mood study, together will the limitations and future directions will also be included.

1.3 Summary

Motivated by the need for systematic studies and comprehensive statistical methodology that record, analyze, and understand people's instant activities and psychological behaviors over time, I developed this dissertation to better understand several research questions from a statistical perspective. A Bayesian three level mixed effect location scale model was formulated and developed for multi-wave EMA studies that allows for individual differences not explained by covariates at both baseline and change over time, in terms of both outcome mean and variability. A shared parameter model framework was tailored for intensive longitudinal data where intemittent and informative missing often occur due to non-responses, making it possible to do valid statistical inference as well as data imputation. A Hidden Markov Model was explored in combination with bivariate mixed effects to account for both time serial dependence and individual differences, allowing heterogeneous covariates effects and latent state classification on the observed measurements. All three chapters deal with clustered and intensively measured EMA data and share common research interests - identifying risk factors on mood regulation. The three studies were conducted to tackle problems and answer research questions from different statistical perspectives yet related to each other.

CHAPTER 2

A THREE LEVEL BAYESIAN MIXED EFFECTS LOCATION SCALE MODEL WITH AN APPLICATION TO ECOLOGICAL MOMENTARY ASSESSMENT DATA

2.1 Introduction

Modern data collection procedures, such as ecological momentary assessments (EMA), allow researchers to study outcomes with high volatility by repeated sampling of subjects' behaviors and experiences in real time and subjects' natural environments [Shiffman et al., 2008]. Typically these procedures involve self-reported data collection from individuals over the course of hours, days, and weeks, thus yield relatively large numbers of observation per subject [Ebner-Priemer and Trull, 2009]. A particular interest in EMA studies is to identify factors that affect the within subject variance of the intensively measured outcomes, in addition to the overall mean levels [Stone and Shiffman, 1994]. Due to the hierarchical nature of EMA data, random subject effects are usually included in statistical models to account for the correlation among repeated measures for a given subject [Schwartz and Stone, 1998]. Hedeker et al. [Hedeker et al., 2008] developed a mixed effect location scale model that includes an additional random subject effect in the error variance, thus allowing subject variation in terms of both the mean and variance of the intensively measured outcomes. Random effects in both the mean and variance model can be useful in distinguishing the residual variation from unobserved subject-level variables, thus providing more accurate standard errors and valid statistical inference [W.S et al., 2002].

EMA studies are sometimes conducted at multiple measurement waves, resulting in a three level data structure: observations nested within waves, and waves in turn nested within subjects [Piasecki et al., 2014]. For example, a person's mood can be assessed multiple times at each wave and the subject can be followed up at multiple waves. There are three possible sources of mood variation for this type of data: variation between subjects, variation within subject but between waves, and variation within subject within wave. Ignoring any possible sources of variation would lead to misspecification of the correlation structure and invalid statistical inference. Li and Hedeker [Li and Hedeker, 2012] proposed a three level mixed effect location scale model that includes random subject and day intercepts for both the mean and within variance of the outcome. Kapur et al. [Kapur et al., 2015] proposed a similar Bayesian mixed effect location scale model for multivariate outcomes at one EMA wave. However, these models assume that subjects change with a constant rate in terms of both mean and variance. This assumption can be easily violated, especially in psychological and behavioral studies, as subjects almost always exhibit heterogeneous trajectories across time [Dewey et al., 2015]. Using the above mood example, subjects are likely to have different mood variability at baseline, and over time, some can become more consistent while others become more erratic. Rast et al. [Rast et al., 2012], Leckie [Leckie, 2014], and Goldstein et al. [Goldstein et al., 2017] all presented a two level mixed effect location scale model that includes random intercept and slope for both the location and scale model, allowing for heterogeneous trajectories across time. Therefore, a three level model that treats observations within waves within subjects while accounting for subject heterogeneity at baseline and over time for both mean and variance will provide a more comprehensive utilization of the data as well as address more specific questions of interest. However, estimation of such general models with relatively large numbers of random effects using likelihood-based methods can be prohibitive due to computational and numerical complexity [Bates et al., 2015].

In this article, we propose a Bayesian mixed effect location scale model for three level data structures, where observations are nested within waves, and waves further nested within subjects. The proposed model extends the conventional three level mixed effect regression model by including random subject intercept and slope as well as random wave intercept for both the mean and within-subject variance of the outcome. At the mean level, the proposed model allows subjects to have heterogeneity in their baseline responses as well as different growth rates over time. Similarly at the variance level, subjects are allowed to exhibit different variation at baseline and the variation can also change differentially over time. Both the subject and wave level heterogeneity can be explained by observed covariates as well as unobserved variables through specification of random effects. Furthermore, the random location and scale effects are allowed to be correlated. The proposed model is estimated using a Bayesian approach. Specifically, Markov Chain Monte Carlo sampling methods are used to generate samples from the joint posterior distribution, and parameter estimates and credible intervals are obtained by summarizing the corresponding distributions [Bradley and Siddhartha, 1995]. We will demonstrate how Stan (an open source Hamiltonian Monte Carlo sampler) and the Hamiltonian Monte Carlo algorithm can be used to achieve consistent parameter estimates and we provide a detailed syntax example in the Supplemental Materials [Carpenter et al., 2017]. The model is validated via a sequence of simulation studies against several reduced models. Finally the proposed three level model is applied to an EMA adolescent smoking study, where the interest is on identifying risk factors associated with high mood variation as well as exploring the possible mood trajectories.

2.2 Motivating Adolescent Smoking Study Example

The data that motivate the development of the Bayesian three level mixed effect location scale model are from an EMA adolescent smoking study. In the study, 461 adolescents from 9th and 10th grade were recruited. The average age of the participants is 15.6, with the minimun being 14.4 and maximum 16.7. They carried hand held devices for 7 days at each measurement wave, during which they responded to random interviews (approximately 5 times per day) or event recorded any episodes of smoking. At each prompt or smoking episode, participates were asked to answer questions, including location, activities, companionship, mood and other psychological measurements. The study was conducted at

6 waves: baseline, 6 months, 15 months, 2 years, 5 years, and 6 years. Data were collected on age, gender, beep type (smoking event vs random prompt), and positive affect (measure of positive mood). Because of the interest in comparing responses across waves from random prompts vs smoking events, subjects were included if they were measured on at least two waves and had at least two smoking events at each wave, resulting in a sample size of 254 subjects.

Among all the subjects, 51.6% were female, and on average subjects were followed up at three waves with 36 to 51 prompts (including smoking episodes) per wave during the entire study span. A total of 24490 random prompts and 8087 smoking events were obtained, with an approximate average of 96 random prompts and 32 smoking episodes per subject. For the analyses reported, a three level structure of observations (level1) within waves (level 2) within subjects (level 3) was considered.

The outcome is the measure of subjects' positive affect (PA), which consists of the average of several mood items rated from 1 to 10: I felt happy, I felt relaxed, I felt cheerful, I felt confident, and I felt accepted by others. Thus, higher PA levels indicate better mood. The interest is to see whether subjects tend to have higher and more consistent positive affect after smoking compared to random prompts. We are also interested in differentiating the between subject and within subject between wave effect from the within subject within wave effect. That is, the effect of smoking when comparing different subjects (between subject), the same subject at different waves (within subject between wave), and the same subject at the same wave but different occasions (within subject within wave). Since the three variables contain different information in characterizing subjects' smoking behavior, it is useful to include the decomposed variables in the model and investigate their relative statistical and clinical significance so that further interventions can be done at that level [Piasecki et al., 2014]. Investigation into the PA showed that subjects exhibit different trends across waves in terms

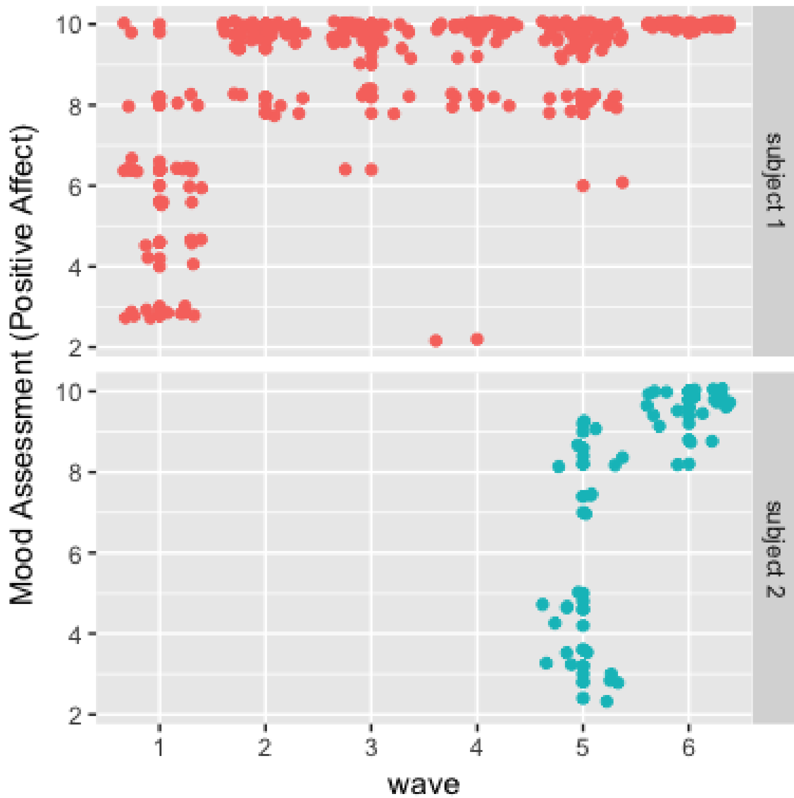

Figure 2.1: Mood assessments: erratic to consistent. (Random subject scale effect estimates are estimated to be $(-0.41, -3.64)$ and $(0.30, -0.25)$)

of both mean and variability, as shown in Figure 2.1 and Figure 2.2.

2.3 Methods

2.3.1 Three Level Mixed Effect Model

Suppose there are $k = 1, ..., n_{ij}$ observations nested within $j = 1, ..., n_i$ waves, and waves are then nested within $i = 1, ..., n$ subjects. Let y_{ijk} denote the outcome for subject i measured at wave j and occasion k. The conventional three level mixed effect model can be expressed as

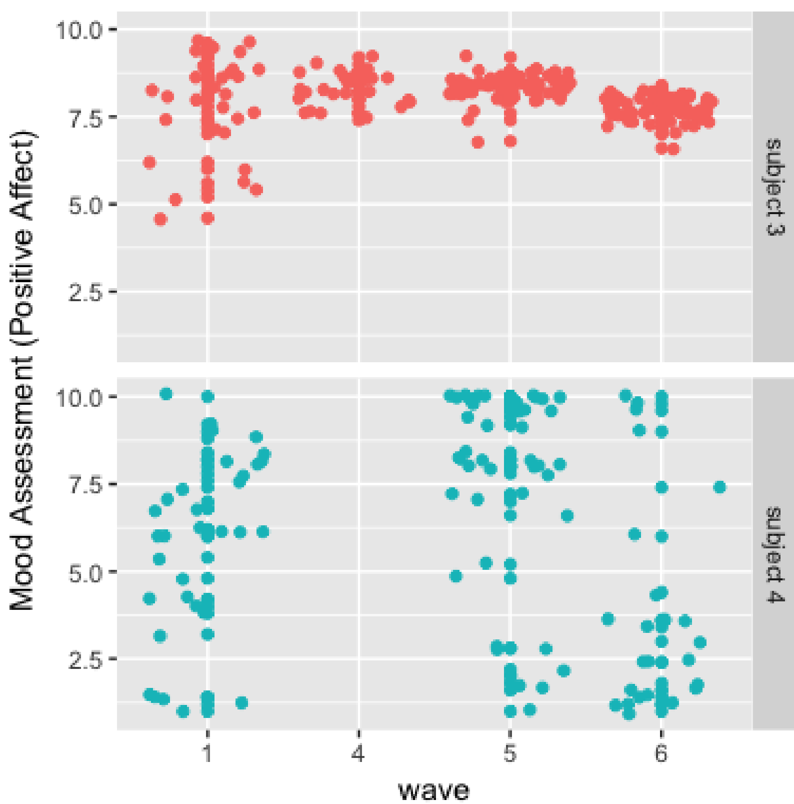

Figure 2.2: Mood assessments: remains consistent or erratic.(Random subject scale effect estimates are estimated to be $(-1.09, -2.52)$ and $(1.36, 1.52)$)

$$Y_{ijk} = X_{ijk}^{\top} \ \beta + Z_{ijk}^{\top} \ \gamma_i + V_{ijk}^{\top} \ \nu_{ij} + \epsilon_{ijk} \tag{2.1}$$

where X_{ijk} is the $p \times 1$ vector of regressors (typically including a column of "1" for the intercept) which can contain either subject, wave or occasion level variables, and β is the corresponding vector of regression coefficients. Z_{ijk} (usually a subset of X_{ijk}) is the vector of regressors for random effect γ_i , and γ_i is the vector of random subject effect, indicating the influence of individual i on his/her repeated mood assessments. Similarly, V_{ijk} (again, usually a subset of X_{ijk}) is the vector of regressors for random effect ν_{ij}, where ν_{ij} represents the vector of random wave effect, indicating the influence of wave j on subject i's repeated mood assessments.

For the EMA adolescent smoking study example, the outcome Y_{ijk} is the positive affect for subject i at wave j and occasion k. Since we are interested in differentiating the within subject within wave effect from the between subject as well as the within subject between wave effects, we will decompose the occasion level variable smk_{ijk} (1 for smoking event and 0 for random prompt) into subject, wave and occasional level variables.

$$\overline{smk}_i = \sum_j \sum_k smk_{ijk} / \sum_j K_{ij}, \ \ \overline{smk}_{ij} = \sum_k smk_{ijk}/K_{ij} - \overline{smk}_i, \ \ \widetilde{smk}_{ijk} = smk_{ijk} - \overline{smk}_{ij} \tag{2.2}$$

Here, K_{ij} is the number of observations for subject i at wave j; $\overline{smk}_i$ is the decomposed subject level variable and represents the average (proportion) of smoking events for subject i; $\overline{smk}_{ij}$ is the decomposed wave level variable and represents the deviation of average smoking events at wave j relative to the subject level average $\overline{smk}_i$; $\widetilde{smk}_{ijk}$, which is computed as the deviation of smoking events at occasion k relative to the subject's wave level average, represents the pure occasion level smoking effect adjusted for his/her subject and wave level average. All three variables will be included in the mean model. For random subject effects, both a random intercept and a random slope over wave will be included since there is interest about subject heterogeneity both at baseline and trajectories over time. So Z_{ijk} will be two dimensional and consists of a column of 1 and wave indicator $wave_{ij}$. Correspondingly, $\gamma_i = \{\gamma_{0,i}, \gamma_{1,i}\}$, with $\gamma_{0,i}$ being the random subject intercept indicating the influence of subject i on his/her baseline mood, and $\gamma_{1,i}$ being random subject slope indicating the influence of subject i on how fast or slow his/her mood changes over time. Since our data have a three level structure with an intermediate wave clustering, an additional random wave effect should be included. For wave, only a random intercept will be considered to indicate the possible influence of wave on subjects' repeated mood assessments: even for the same subject, the mood can be different at different waves and the difference cannot be fully explained by the observed wave level variables. As a result, V_{ijk} will be a column of 1 and ν_{ij} is of dimension 1. Therefore, the mean model for the adolescent mood study example can be expressed explicitly as

$$Y_{ijk} = \beta_0 + \beta_1 \ male_i + \beta_2 \ \overline{smk}_i + \beta_3 \ \overline{smk}_{ij} + \beta_4 \ \widetilde{smk}_{ijk} + \beta_5 \ wave_{ij} + \gamma_{0,i} + \gamma_{1,i} \ wave_{ij} + \nu_{ij} + \epsilon_{ijk} \tag{2.3}$$

The random effects γ and ν are referred to as location random effects since they influence the mean or location of the outcome. Both γ and ν are assumed to be normally distributed with constant variance covariance structure Σ_γ and σ^2_ν, and independent of each other. The size of the diagonal elements in Σ_γ indicates the amount of between subject variability, while size of σ^2_ν indicates the amount of the within subject between wave variability. The random error ϵ_{ijk} are usually assumed to be normally distributed with constant variance σ^2_ϵ.

2.3.2 Three Level Mixed Effect Location Scale Model - An Extension of The Conventional Mixed Effect Model

In the above three level mixed effect model specification, σ_ϵ^2 represents the amount of variability that exists within subjects and within waves. By assuming σ_ϵ^2 constant, we are assuming that the within variance does not vary for different subjects or waves. This assumption can be easily violated in practice, especially for psychological and behavioral studies, where subjects almost always exhibit variation in terms of the consistency in their responses. One approach to relax this assumption is to additionally model σ_ϵ^2 by another mixed effect model through a log-linear representation

$$log(\sigma_{ijk}^2) = \alpha_0 + \alpha_1 \ male_i + \alpha_2 \ \overline{smk}_i + \alpha_3 \ \overline{smk}_{ij} + \alpha_4 \ \widetilde{smk}_{ijk} + \alpha_5 \ wave_{ij} + \lambda_{0,i} + \lambda_{1,i} \ wave_{ij} + \tau_{ij} \tag{2.4}$$

Similar to the mean model 2.2, the within variance model contains both fixed effects α and random effects $\{\lambda, \tau\}$. In addition to the observed variables $\left\{male_i, \overline{smk}_i, \overline{smk}_{ij}, \widetilde{smk}_{ijk} wave_{ij}\right\}$ that can influence the variability of the outcome for certain subject at certain waves, there can also be unmeasured variables contributing to how consistent/erratic the outcome measurements could possibly be. Ignoring the unobserved information would lead to invalid inference about the variance parameters. This motivates the inclusion of random scale effects in the within variance model 2.4. At subject level, $\lambda_{0,i}$ is the random scale intercept and indicates the influence of subject i on his/her mood variability at baseline, and $\lambda_{1,i}$, the random scale slope, indicates the influence of subjects on how the variability changes over time. For example, some subjects may start off with relatively consistent responses (small within variance at baseline), but over time their responses become more and more erratic (positive slope on the within variance over wave), while others may follow some different patterns. This heterogeneity among subjects in terms of the variance trajectories can be captured by the random subject scale intercept λ_0 and slope λ_1. At the wave level, only a random scale

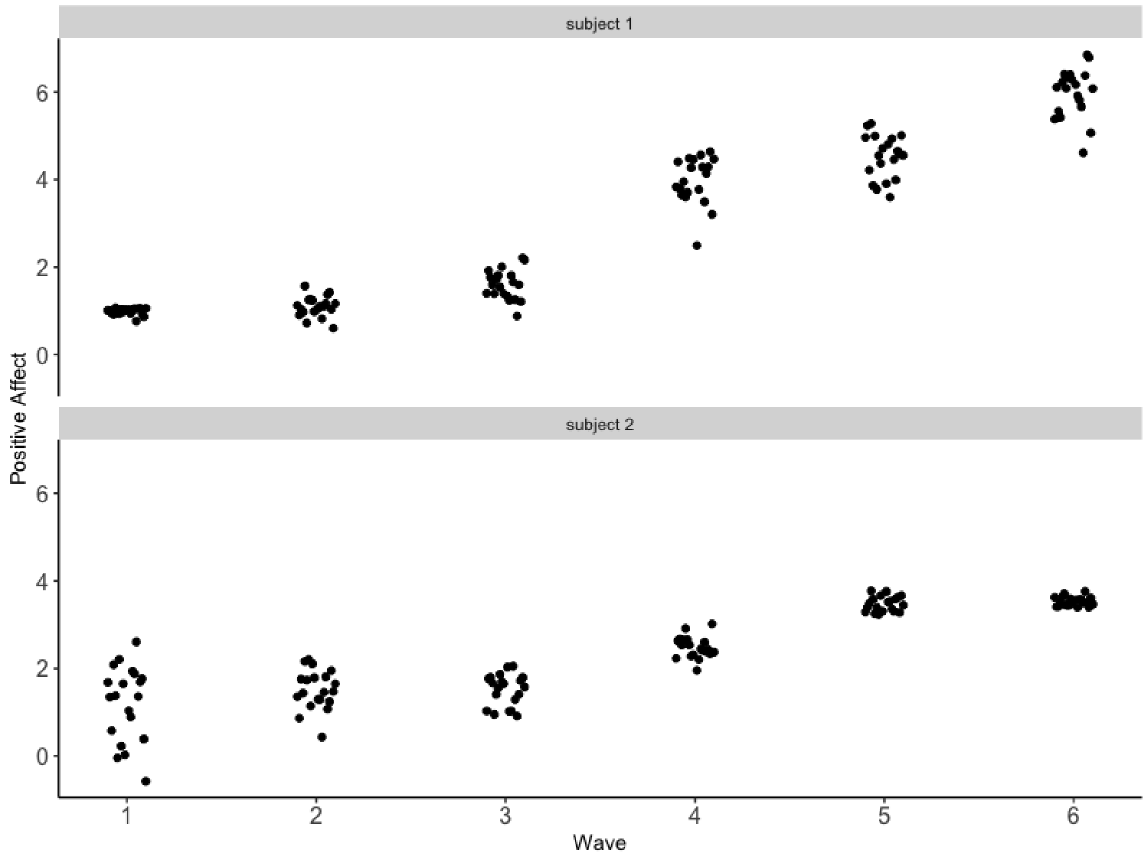

Figure 2.3: Visualization of the model mechanics

intercept τ will be considered to account for the possible effect of wave on the within variance.

An intuituive visualization of the model mechanics is shown in Figure 2.3. There are two hypothetical subjects: subject 1 has both increasing postive affect and mood variation, while subject 2 has increasing positive affect but diminishing mood variaton across wave. We can also visualize the wave effect as different waves exihibit different mean as well as variation. The different patterns suggest different mood trajectories as well as disease progostics from a psychological perspective, which our proposed model is able to captur.

There are six random effects, consisting both subject and wave levels, in terms of both the mean and within variance of the outcome. The population distribution of these random effects is similar to the ordinary mixed effects models in that random subject effects can be possibly correlated but should be independent of the random wave effects. In addition, the random location effects are allowed to be possibly correlated with the random scale effects, as extreme mean values are often accompanied with more consistent variance due to ceiling or floor measurement effects. The distributional assumption for the six random effects can be expressed as

$$\begin{bmatrix} \gamma_{0,i} \\ \gamma_{1,i} \\ \lambda_{0,i} \\ \lambda_{1,i} \end{bmatrix} \sim \mathcal{N}_4 \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\gamma_0} & \text{cov}_{\gamma_0,\gamma_1} & \text{cov}_{\gamma_0,\lambda_0} & \text{cov}_{\gamma_0,\lambda_1} \\ \text{cov}_{\gamma_0,\gamma_1} & \sigma^2_{\gamma_1} & \text{cov}_{\gamma_1,\lambda_0} & \text{cov}_{\gamma_1,\lambda_1} \\ \text{cov}_{\gamma_0,\lambda_0} & \text{cov}_{\gamma_1,\lambda_0} & \sigma^2_{\lambda_0} & \text{cov}_{\lambda_0,\lambda_2} \\ \text{cov}_{\gamma_0,\lambda_1} & \text{cov}_{\gamma_1,\lambda_1} & \text{cov}_{\lambda_0,\lambda_1} & \sigma^2_{\lambda_1} \end{bmatrix} \right) \tag{2.5}$$

$$\begin{bmatrix} \nu_{0,ij} \\ \tau_{0,ij} \end{bmatrix} \sim \mathcal{N}_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\nu_0} & \text{cov}_{\nu_0,\tau_0} \\ \text{cov}_{\nu_0,\tau_0} & \sigma^2_{\tau_0} \end{bmatrix} \right) \tag{2.6}$$

2.3.3 Model Estimation

To estimate the model parameters, Bayesian approaches are favored against maximum likelihood methods which usually involve heavy numerical integration and approximation of the first and second order partial derivatives [McCulloch and Neuhaus, 2005]. For a typical Newton Raphson algorithm to achieve MLEs, one would need to integrate the conditional likelihood over the joint distribution of all random effects in order to compute the marginal likelihood. As a result, the computational load and complexity increases exponentially with the number of random effects, making the estimating procedure infeasible for models with relatively large numbers of random effects [Hedeker and Gibbons, 2006]. Bayesian approaches, on the other hand, perform the estimation by drawing Markov Chain Monte Carlo samples from the joint posterior distribution given the prior that reflects our belief about the parameters before collecting the data [Fong et al., 2010]. Various sampling algorithms can be used, including a mixture of Gibbs sampling, Metropolis-Hastings and Hamiltonian Monte Carlo [Gelman et al., 2013]. Parameter estimates and credible intervals can then be obtained by taking the point estimates and corresponding intervals associated with the posterior, thus avoiding the computational issues associated with numerical integration [Leonard, 1975]. Given flat priors and enough MCMC samples, the Bayesian approach will yield consistent parameter estimates [Hobert and Casella, 1996].

For a full Bayesian approach, parameters (β, α) and random effects $(\gamma_0, \gamma_1, \lambda_0, \lambda_1, \nu_0, \tau_0)$ are both regarded as random quantities while the data Y are regarded as fixed. To simplify the notation, denote $\theta = (\beta, \alpha)$ as the model parameter vector, $(\gamma_i, \lambda_i) = \left(\gamma_{0,i}, \gamma_{1,i}, \lambda_{0,i}, \lambda_{1,i}\right)_{i=1}^{n}$ as the random subject location/scale effects, $\left(\nu_{ij}, \tau_{ij}\right) = \left(\nu_{0,ij}, \tau_{0,ij}\right)_{i,j=1}^{n,n_i}$ as the random wave location/scale effects, and Y_{ij} as the data vector.

Since θ, (γ, λ) and (ν, τ) are all random, they each follow some prior distribution before we get to observe the data Y, which we denote as $\pi(\theta)$, $\pi(\gamma, \lambda)$, and $\pi(\nu, \tau)$ respectively. Since individuals are assumed to be independent and correlations exist due to repeated measurements within the same individual, $\pi(\gamma, \lambda)$ can be written as $\prod_{i=1}^{n} \pi(\gamma_i, \lambda_i)$, the product of the prior for each individual (same applies to $\pi(\nu, \tau)$). A natural choice for $\pi(\gamma_i, \lambda_i)$ and $\pi(\nu_{ij}, \tau_{ij})$ are bivariate standard normals. $\pi(\theta)$ is a bit tricky to choose, but one can specify a separate prior for each of the component in θ provided that a full conditional posterior is obtained for each of them [Casella and George, 1992]. For illustration purposes, we will derive the full conditional posterior for θ, (γ_i, λ_i) and $\left(\nu_{ij}, \tau_{ij}\right)$, instead of every component in each vector.

$$P(\theta \mid \gamma_i, \lambda_i, \nu_{ij}, \tau_{ij}, Y_{ij}) \;\propto\; P(Y_{ij} \mid \theta, \gamma_i, \lambda_i, \nu_{ij}, \tau_{ij})\pi(\theta) \tag{2.7}$$

$$P(\gamma_i, \lambda_i \mid \theta, \nu_{ij}, \tau_{ij}, Y_{ij}) \;\propto\; P(Y_{ij} \mid \theta, \gamma_i, \lambda_i, \nu_{ij}, \tau_{ij})\pi(\gamma_i, \lambda_i) \tag{2.8}$$

$$P(\nu_{ij}, \tau_{ij} \mid \theta, \gamma_i, \lambda_i, Y_{ij}) \;\; \propto \;\; P(Y_{ij} \mid \theta, \gamma_i, \lambda_i, \nu_{ij}, \tau_{ij})\pi(\nu_{ij}, \tau_{ij}) \tag{2.9}$$

$$P(Y_{ij} \mid \theta, \gamma_i, \lambda_i, \nu_{ij}, \tau_{ij}) = \prod_{k=1}^{K_{ij}} \frac{1}{\sqrt{2\pi \exp(X_{ijk}^\top \alpha + Z_{ijk}^\top \lambda_i + \tau_{ij})}} \exp\left(-\frac{\left(y_{ijk} - \left(X_{ijk}^\top \beta + Z_{ijk}^\top \gamma_i + \nu_{ij}\right)\right)^2}{2\exp(X_{ijk}^\top \alpha + Z_{ijk}^\top \lambda_i + \nu_{ij})}\right) \tag{2.10}$$

where $P(Y_{ij} \mid \theta, \gamma_i, \lambda_i, \nu_{ij}, \tau_{ij})$ is the conditional likelihood given in 2.10, and π is the corresponding prior distribution. If the priors are independent for θ, (γ_i, λ_i) and (ν_{ij}, τ_{ij}), we will arrive at the nice posterior distribution in 2.7, 2.8 and 2.9. Once the full conditional posteriors are obtained, we can approximate the joint posterior $P(\theta, \gamma, \lambda, \nu, \tau \mid Y)$ by sampling each variable from its full conditional posterior given in 2.7, 2.8, 2.9 iteratively. A problem that will often arise in Gibbs sampling is that it is difficult to sample from the conditional posteriors if they are not of recognized forms. One solution is to use the Metropolis-Hastings algorithm, which keeps drawing samples from a proposal distribution and decides whether or not to accept the sample as from the conditional posterior with some probability (acceptance ratio) [Chib and Greenberg, 1995]. This is not of concern if we choose the priors as multivariate Gaussian, as normal likelihood will lead to conjugate posterior for multivariate Gaussian priors [Bishop, 2006]. However, in the absence of conjugacy, Metropolis-Hastings should be used to sample from each full conditional.

The detailed Markov Chain Monte Carlo algorithm where (component wise) Metropolis-Hastings is nested in Gibbs sampling is listed below, with q being the corresponding (user defined) proposal distribution. After enough runs, the chains will ultimately converge to the joint posterior and one can summarize the posterior samples to get the parameter estimates as well as credible intervals.

1. Initialize at $(\theta, \gamma, \lambda, \nu, \tau) = \left(\theta^0, \gamma^0, \lambda^0, \nu^0, \tau^0\right)$

2. Sample a single random value iteratively from each full conditional posterior by Metropolis-Hastings algorithm below, for $t = 1, 2, ...$

   (a) Given the current value of $\theta^t = \theta$, generate a proposed new value θ' according to $q_\theta(\theta \to \cdot)$, and accept θ' with probability $A_\theta = min\left(1, \frac{P(\theta'|\gamma^t,\lambda^t,\nu^t,\tau^t,Y)q_\theta(\theta'\to\theta)}{P(\theta|\gamma^t,\lambda^t,\nu^t,\tau^t,Y)q_\theta(\theta\to\theta')}\right)$

   (b) Given the current value of $\left(\gamma^t, \lambda^t\right) = (\gamma, \lambda)$, generate a proposed new pair $\left(\gamma', \lambda'\right)$ according to $q_{\gamma,\lambda}((\gamma, \lambda) \to \cdot)$, and accept $\left(\gamma', \lambda'\right)$ with probability $A_{\gamma,\lambda} = min\left(1, \frac{P((\gamma',\lambda')|\theta^{t+1},\nu^t,\tau^t,Y)q_{\gamma,\lambda}((\gamma',\lambda')\to(\gamma,\lambda))}{P((\gamma,\lambda)|\theta^{t+1},\nu^t,\tau^t,Y)q_{\gamma,\lambda}((\gamma,\lambda)\to(\gamma',\lambda')}\right)$

   (c) Given the current value of $\left(\nu^t, \tau^t\right) = (\nu, \tau)$, generate a proposed new value $\left(\nu', \tau'\right)$ according to $q_{\nu,\tau}((\nu, \tau) \to \cdot)$, and accept $\left(\nu', \tau'\right)$ with probability $A_{\nu,\tau} = min\left(1, \frac{P((\nu',\tau')|\theta^{t+1},\gamma^{t+1},\lambda^{t+1},Y)q_{\nu,\tau}((\nu',\tau')\to(\nu,\tau))}{P((\nu,\tau)|\theta^{t+1},\gamma^{t+1},\lambda^{t+1},Y)q_{\nu,\tau}((\nu,\tau)\to(\nu',\tau'))}\right)$

3. Repeat step a, b c in (2) until convergence.

The trade-off in any MCMC algorithm is the step size versus the acceptance rate. In the above Metropoli-Hastings-Gibbs method, every time we propose a new value from q given the current value of the variable, the closer the newly proposed value moves from the current value, the more likely it will be accepted. However, it will also lead to slow convergence, or, often in practice (within finite time limit), no convergence to the global posterior. Therefore, Hamiltonian Monte Carlo (also referred to as Hybrid Monte Carlo) is often used to balance the trade-off between step size and acceptance rate. It works by reducing the correlation between successive samples by using a Hamiltonian evolution and targeting values with a higher acceptance rate than the observed probability distribution. Thus, we can specify the data generating process for the data Y in Stan, an open source Hamiltonian Monte Carlo sampler, and obtain posterior samples from it [Stan-Development-Team, 2014]. There are other open source samplers that utilize Metropolis-Hastings algorithms [Rast et al., 2012]. Implementation details in Stan and code are provided in 2.4.2.

2.4 Simulation

2.4.1 Results and Conclusions

To validate the proposed model as well as the estimation procedure, a simulation study was conducted. A series of 100 data sets, each with 10000 observations (100 subjects, each subject measured at 10 waves and 10 occasions within each wave) were generated under the three level location scale model with three covariates ($X_i \sim \mathcal{N}(0,1)$, $X_{ij} \sim \mathcal{N}(0,1)$, $X_{ijk} \sim \mathcal{N}(0,1)$). The true parameter values for the six random effects variances/covariances are $COV\left(\begin{bmatrix}\gamma_0\\ \gamma_1\\ \lambda_0\\ \lambda_1\end{bmatrix}\right) = \begin{bmatrix} 1.0 & -0.1 & 0 & 0\\ -0.1 & 0.25 & 0 & 0\\ 0 & 0 & 0.25 & -0.025\\ 0 & 0 & -0.025 & 0.065\end{bmatrix}$, and $COV\left(\begin{bmatrix}\nu_0\\ \tau_0\end{bmatrix}\right) = \begin{bmatrix}1.0 & 0.25\\ 0.25 & 1.0\end{bmatrix}$.
For each generated data set, a series of four candidate models were considered: a two level (subject and occasion level) mixed effect regression model with heterogeneous variance (MRM HV), a two level mixed effect location scale model (MLS), a three level (subject, wave and occasion level) mixed effect regression model with heterogeneous model, and the proposed three level mixed effect location scale model. The first three models are considered to be reduced models relative to the last one since they ignore either the clustering due to the intermediate wave or unobserved variables in the variance.

Two level MRM HV:

$$Y_{ik} = X_{ik}^{\top}\ \beta + \gamma_{0,i} + \gamma_{1,i}\ wave_{ik} + \epsilon_{ik} \tag{2.11}$$

$$\sigma^2_{\epsilon,ik} = \exp(W_{ik}^{\top}\ \alpha) \tag{2.12}$$

Two level MLS:

$$Y_{ik} = X_{ik}^{\top} \ \beta + \gamma_{0,i} + \gamma_{1,i} \ wave_{ik} + \epsilon_{ik} \tag{2.13}$$

$$\sigma^2_{\epsilon,ik} = \exp(W_{ik}^{\top} \ \alpha + \lambda_{0,i} + \lambda_{1,i} \ wave_{ik}) \tag{2.14}$$

Three level MRM HV:

$$Y_{ijk} = X_{ijk}^{\top} \ \beta + \gamma_{0,i} + \gamma_{1,i} \ wave_{ij} + \nu_{0,ij} + \epsilon_{ijk} \tag{2.15}$$

$$\sigma^2_{\epsilon,ijk} = \exp(W_{ijk}^{\top} \ \alpha) \tag{2.16}$$

Three level MLS:

$$Y_{ijk} = X_{ijk}^{\top} \ \beta + \gamma_{0,i} + \gamma_{1,i} \ wave_{ij} + \nu_{0,ij} + \epsilon_{ijk} \tag{2.17}$$

$$\sigma^2_{\epsilon,ijk} = \exp(W_{ijk}^{\top} \ \alpha + \lambda_{0,i} + \lambda_{1,i} \ wave_{ij} + \tau_{0,ij}) \tag{2.18}$$

The four candidate models were compared in terms of both mean and variance parameter estimates as well as credible intervals. Bias, average 95% credible interval width (AIW) and average coverage rate (COV) out of 100 data sets were obtained to evaluate model performance. The results are represented in Table 2.1 and Table 2.2.

Table 2.1: Results from 100 simulations under the three level mixed effects location scale model : Mean Model Parameters

Model	$\beta^{intercept} = 1$			$\beta^{subj} = 1$			$\beta^{wave} = 1$			$\beta^{obs} = 1$		
	Bias	AIW	COV	Bias	AIW	COV	Bias	AIW	COV	Bias	AIW	COV
2L MRM	-0.0274	0.3973	46%	-0.0005	0.3953	91%	-0.0081	0.2075	17%	0.0003	0.0613	96%
2L MLS	-0.0292	0.4050	52%	0.0009	0.4000	94%	-0.0075	0.2103	19%	0.0011	0.0577	96%
3L MRM	-0.0288	1.6993	98%	-0.0018	0.3976	95%	-0.0006	1.8039	97%	-0.0019	0.0510	98%
3L MLS	-0.0257	1.6873	99%	-0.0014	0.3989	93%	-0.0016	1.8150	95%	-0.0010	0.0404	94%

AIW: average 95% credible interval; COV: 95% coverage rate out of 100 simulations.

Table 2.2: Results from 100 simulations under the three level mixed effects location scale model : Variance Model Parameters

Model	$\alpha^{intercept} = 0.3$			$\alpha^{subj} = 0.2$			$\alpha^{wave} = 0.1$		
	Bias	AIW	COV	Bias	AIW	COV	Bias	AIW	COV
Two Level MRM HV	0.6109	0.0608	1	-0.0609	0.0567	19	-0.0096	0.0712	16
Two Level MLS	0.5350	0.1499	5	-0.0730	0.1439	49	0.0071	0.1106	17
Three Level MRM HV	0.2410	0.0605	5	-0.0004	0.0570	47	-0.0402	0.0672	12
Three Level MLS	-0.0038	0.8596	97	0.0040	0.2081	95	-0.0155	0.9114	99

AIW: average 95% credible interval; COV: 95% coverage rate out of 100 simulations.

In Table 2.1, $\beta^{intercept}$, β^{subj}, β^{wave}, β^{obs} are the mean model regression coefficients for the intercept, subject, wave and occasion level covariates, respectively. The four models all did relatively well in estimating β as can be seen from the small bias. But they do perform different in terms of estimating the uncertainties associated with the coefficients: the three level models (three level MRM HV and three level MLS) produced wider and more correct intervals (1.6873/1.6993 and 1.8105/1.8039) for the intercept and wave covariate compared to the two level models (0.4050/0.3973 and 0.2103/0.2075). This is due to the fact that neither two level MRM HV nor two level MLS account for the possible unobserved variables at baseline or the intermediate wave level by including random intercept or random wave effect(s), which in turn over states the certainty around the point estimates. Although the point estimates in all four models show small bias, only the three level models yield credible intervals closer to the correct 95% level.

In Table 2.2, $\alpha^{intercept}$, α^{subj} and α^{wave} are the corresponding regression coefficients associated with the intercept, subject and wave level covariates in the log-linear representation of the error variance model. All three reduced models have $\alpha^{intercept}$ estimates biased upwards with narrower credible intervals and insufficient coverage. One explanation is that, when one omits the wave level covariates or random scale effects (or both) in the log linear error variance model, all the variations unexplained by the existing covariates has to be absorbed by $\alpha^{intercept}$, which makes $\alpha^{intercept}$ biased towards the population averaged effect rather than subject specific effects. Leckie [Leckie, 2014] had similar findings regarding the variance model intercept in a two level random intercept location scale model. In terms of α^{subj} and α^{wave}, since neither MRMs included random scale effects, they produced narrower and incorrect credible intervals. Also, the two level MLS under covers α^{wave} due to the fact that it failed to include a random scale effect at the wave level.

In summary, none of the reduced models are comparable to the three level mixed effects location scale model in terms of unbiasedness and correct coverage. If one were to analyze a three level structure data sets where both location and scale random effects are present, using the reduced models would yield invalid statistical inference and arrive at possibly false positive results.

2.4.2 Software Implementation

Implementation of Data Generating Process in Stan:

To facilitate the computation, it is convenient to represent the random effects in standardized form as multivariate standard normals. To standardize the random wave effects $\nu_{0,ij}$ and $\tau_{0,ij}$, we need to find $\mathbf{C}$, a lower triangular matrix, and $\delta_{1,ij}$ and $\delta_{2,ij}$, independent bivariate standard normals, such that

$$\begin{bmatrix} \nu_{0,ij} \\ \tau_{0,ij} \end{bmatrix} = \mathbf{C} \begin{bmatrix} \delta_{1,ij} \\ \delta_{2,ij} \end{bmatrix} = \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} \delta_{1,ij} \\ \delta_{2,ij} \end{bmatrix} \tag{2.19}$$

It can easily be shown that $\mathbf{C}$ is the Cholesky decomposition of the covariance matrix Σ_{ν_0,τ_0}.

$$\mathrm{COV}\begin{pmatrix}\nu_{0,ij}\\ \tau_{0,ij}\end{pmatrix} = \mathrm{COV}\left(\mathbf{C}\begin{bmatrix}\delta_{1,ij}\\ \delta_{2,ij}\end{bmatrix}\right) = \mathbf{C}\mathrm{COV}\begin{pmatrix}\delta_{1,ij}\\ \delta_{2,ij}\end{pmatrix}\mathbf{C}^{\top} = \mathbf{C}\mathbf{C}^{\top} \tag{2.20}$$

$$\implies \Sigma_{\nu_0,\tau_0} = \mathbf{C}\mathbf{C}^{\top} \tag{2.21}$$

Similarly, we can standardize the random subject effects $\gamma_{0,i}$, $\gamma_{1,i}$, $\lambda_{0,i}$, $\lambda_{1,i}$

$$\begin{bmatrix}\gamma_{0,i}\\ \gamma_{1,i}\\ \lambda_{0,i}\\ \lambda_{1,i}\end{bmatrix} = \mathbf{D}\begin{bmatrix}\theta_{1,i}\\ \theta_{2,i}\\ \theta_{3,i}\\ \theta_{4,i}\end{bmatrix} = \begin{bmatrix}d_{11} & 0 & 0 & 0\\ d_{21} & d_{22} & 0 & 0\\ d_{31} & d_{32} & d_{33} & 0\\ d_{41} & d_{42} & d_{43} & d_{44}\end{bmatrix}\begin{bmatrix}\theta_{1,i}\\ \theta_{2,i}\\ \theta_{3,i}\\ \theta_{4,i}\end{bmatrix} \tag{2.22}$$

where $\begin{pmatrix}\theta_{1,i}\\ \theta_{2,i}\\ \theta_{3,i}\\ \theta_{4,i}\end{pmatrix} \sim \mathcal{N}_4\left(\mathbf{0},\ \mathbb{1}_4\right)$, and $\mathbf{D}\mathbf{D}^{\top} = \Sigma_{\gamma_0,\gamma_1,\lambda_0,\lambda_1}$

Once the random effects are standardized, we can re-express our models in terms of the standard normals θ and δ.

$$\begin{aligned}Y_{ijk} =& \beta_0 + \beta_1 \times s\bar{m}k_i + \beta_2 \times s\bar{m}k_{ij} + \beta_3 \times s\tilde{m}k_{ijk} + \beta_4 \times wave_{ij}\\ & d_{11} \times \theta_{1,i} + (d_{21} \times \theta_{1,i} + d_{22} \times \theta_{2,i}) \times wave_{ij} +\\ & c_{11} \times \delta_{1,ij} +\\ & \epsilon_{ijk}\end{aligned}$$

$$
\begin{aligned}
log(\sigma^2_{\epsilon,ijk}) =& \alpha_0 + \alpha_1 \times s\bar{m}k_i + \alpha_2 \times s\bar{m}k_{ij} + \alpha_3 \times sm\tilde{k}_{ijk} + \alpha_4 \times wave_{ij} \\
& (d_{31} \times \theta_{1,i} + d_{32} \times \theta_{2,i} + d_{33} \times \theta_{3,i}) + \\
& (d_{41} \times \theta_{1,i} + d_{42} \times \theta_{2,i} + d_{43} \times \theta_{3,i} + d_{44} \times \theta_{4,i}) \times wave_{ij} + \\
& (c_{21} \times \delta_{1,ij} + c_{22} \times \delta_{2,ij})
\end{aligned}
$$

R code to simulate three level data and run the proposed model:

Below is sample code in R to simulate a three level data structure to be used below since the adolescent smoking data cannot be shared publicly at this time.

```
DataGen <- function(nsubj, nwave, nobs,
beta0, beta1, beta2, beta3,
alpha0,alpha1, alpha2,
sd.subj_loc, sd.subj_scale,
sd.wave_loc, sd.wave_scale,
cor.subj_loc = matrix(c(1,-0.2,-0.2,1), nrow = 2, byrow = TRUE),
cor.subj_scale = matrix(c(1,-0.2,-0.2,1), nrow = 2, byrow = TRUE)){

## Generate subject wave and time indicators
N <- nsubj * nwave * nobs
subject <- gl(n = nsubj, k = nwave * nobs)
wave <- rep(gl(n = nwave,k = nobs), nsubj)
obs <- rep(seq_len(nobs), nsubj * nwave)

## Generate subject, wave and obs level covariates (observed).
x.subj <- rep(rbinom(nsubj, 1, 0.5), each = nwave * nobs)
x.wave <- wave
```

```
x.obs <- rbinom(nobs, 1, 0.5)


## Generate subject, wave level random effects
chol.subj_loc <- t(chol(cor.subj_loc))
sigma.sd.subj_loc <- diag(sd.subj_loc) %*% chol.subj_loc
subj_loc <- mvrnorm(nsubj, mu = c(0, 0),
Sigma = sigma.sd.subj_loc %*% t(sigma.sd.subj_loc))
chol.subj_scale <- t(chol(cor.subj_scale))
sigma.sd.subj_scale <- diag(sd.subj_scale) %*% chol.subj_scale
subj_scale <- mvrnorm(nsubj, mu = c(0, 0),
Sigma = sigma.sd.subj_scale %*% t(sigma.sd.subj_scale))
subj1.loc <- rep(subj_loc[, 1], each = nwave * nobs)
subj2.loc <- rep(subj_loc[, 2], each = nwave * nobs)
subj1.scale <- rep(subj_scale[, 1], each = nwave * nobs)
subj2.scale <- rep(subj_scale[, 2], each = nwave * nobs)
wave_loc <- rnorm(nwave, mean = 0, sd = sd.wave_loc)
wave_scale <- rnorm(nwave, mean = 0, sd = sd.wave_scale)
wave.loc <- rep(rep(wave_loc, each = nobs), nsubj)
wave.scale <- rep(rep(wave_scale, each = nobs), nsubj)


## Generate the mean
y.mean <- beta0 + beta1 * x1 + beta2 * x2 + beta3 * x3 +
subj1.loc + subj2.loc * x2 + wave.loc
## Generate the sd
y.sd <- sqrt(exp(alpha0 + alpha1 * x1 + alpha2 * x2 +
subj1.scale + subj2.scale * x2 + wave.scale))
## Generate response
```

```
y <- rnorm(N, mean = y.mean, sd = y.sd)


## Return a data frame
df.LSME <- data.frame(subject = subject, wave = wave, time = time,
gender = x.subj, wave = x.wave, smk = x.obs, y = y)
return(df.LSME)
}
```

Below is the code in R to prepare data to be used in Stan.

```
# Simulate data by calling DataGen function
df.LSME <- DataGen(nsubj = 100, nwave = 10, ntime = 10,
beta0 = 1, beta1 = 1, beta2 = 1, beta3 = 1,
alpha0 = 0.3,alpha1 = 0.2, alpha2 = 0.1,
sd.subj_loc = c(1, 0.5),
sd.subj_scale = c(0.5, 0.25),
sd.wave_loc = 1,
sd.wave_scale = 0.5,
cor.subj_loc = matrix(c(1,   -0.2,
-0.2,  1),
nrow = 2,
byrow = TRUE),
cor.subj_scale = matrix(c(1,   -0.2,
-0.2,  1),
nrow = 2,
byrow = TRUE))
# Prepare data for stan
nsubj <- 100
```

```
nwave <- 10
N <- nrow(df.LSME)
ni <- 10
subject <- df.LSME$subject
wave <- df.LSME$wave
# Decompose x3 into subject, wave and observation level
x3 <- df.LSME$x3
x3.i <- as.vector(tapply(df.LSME$x3, df.LSME$subject, mean))
x3.subj <- rep(x3.i, each = 100)
x3.ij <- t(tapply(df.LSME$x3, list(df.LSME$subject, df.LSME$wave), mean))
x3.wave <- rep(x3.ij, each = 10) - x3.subj
x3.obs <- x3 - x3.wave
# Other variables
x1 <- df.LSME$x1
x2 <- df.LSME$x2
wave <- df.LSME$wave
X_mean <- unname(model.matrix(~ 1 + x1 + x2 + x3.subj + x3.wave + x3.obs))
X_var <- unname(model.matrix(~ 1 + x1 + x2 + x3.subj + x3.wave))
Z_subj <- unname(model.matrix(~ 1 + x2))
y <- df.LSME$y
# Incorporate into a list
StanDat <- list (
N = N,
nsubj = nsubj,
nwave = nwave,
subject = subject,
wave = wave,
```

```
X_mean = X_mean,
X_var = X_var,
Z_subj = Z_subj,
y = y)
# Get initial values for stan
lme_fit <- lmer(y ~ x1 + x2 + x3.subj + x3.wave + x3.obs +
(1 + x2 | subject) +
(1 | subject:wave))
sum <- summary(lme_fit)

MLS_init <- function(){
list(
beta = sum$coefficients[, "Estimate"],
alpha = c(log(sum$sigma), 0, 0, 0, 0),
C_subject_loc = diag(2),
C_subject_scale = diag(2),
sigma_subj_loc = sqrt(diag(VarCorr(lme_fit)$subject)),
sigma_subj_scale = c(0.1, 0.1),
sigma_wave_loc = sqrt(diag(VarCorr(lme_fit)$"subject:wave")),
sigma_wave_scale = 0.1,
z_subj_loc = array(rnorm(2*nsubj), dim = c(nsubj, 2)),
z_subj_scale = array(rnorm(2*nsubj), dim = c(nsubj, 2)),
z_wave_loc = array(rnorm(nwave*nsubj), dim = c(nsubj, nwave)),
z_wave_scale = array(rnorm(nwave*nsubj), dim = c(nsubj, nwave))
)
}
```

Below is sample code in Stan to estimate the model presented in this manuscript. Save this stan file in the same working directory as LSME_model.stan.

```
data {
int<lower=1> N;                             //number of data points
int<lower=1> nsubj;                         //number of subjects
int<lower=1> nwave;                         //number of waves
int<lower=1, upper=nsubj> subject[N];       //indicator for subjects
int<lower=1, upper=nwave> wave[N];          //indicator for waves
row_vector[6] X_mean[N];                    //design matrix for fixed effect
row_vector[5] X_var[N];                     //design matrix for fixed effect
row_vector[2] Z_subj[N];                    //design matrix for subject level RE
real y[N];                                  //outcome
}

parameters {
vector[6] beta;                           //fixed effect for mean
vector[5] alpha;                          //fixed effect for variance
cholesky_factor_corr[2] C_subj_loc;       //cholesky components for RE
cholesky_factor_corr[2] C_subj_scale;     //cholesky components for RE
vector<lower=0>[2] sigma_subj_loc;        //se for RE at subject level
vector<lower=0>[2] sigma_subj_scale;      //se for RE at subject level
real<lower=0> sigma_wave_loc;             //se for RE at wave level
real<lower=0> sigma_wave_scale;           //se for RE at wave level
vector[2] z_subj_loc[nsubj];              //RE at subject level
vector[2] z_subj_scale[nsubj];            //RE at subject level
real z_wave_loc[nwave];                   //RE at wave level
real z_wave_scale[nwave];                 //RE at wave level
```

```
}

transformed parameters {
matrix[2,2] Sigma_subj_loc;
matrix[2,2] Sigma_subj_scale;
vector[2] var_subj_loc;
vector[2] var_subj_scale;
real var_wave_loc;
real var_wave_scale;
vector[2] nu_subj_loc[nsubj];
vector[2] nu_subj_scale[nsubj];
real nu_wave_loc[nwave];
real nu_wave_scale[nwave];
{
Sigma_subj_loc = diag_pre_multiply(sigma_subj_loc, C_subj_loc);
Sigma_subj_scale = diag_pre_multiply(sigma_subj_scale, C_subj_scale);
var_subj_loc = sigma_subj_loc .* sigma_subj_loc;
var_subj_scale = sigma_subj_scale .* sigma_subj_scale;
var_wave_loc = sigma_wave_loc * sigma_wave_loc;
var_wave_scale = sigma_wave_scale * sigma_wave_scale;

for (j in 1 : nsubj) {
nu_subj_loc[j] = Sigma_subj_loc * z_subj_loc[j];
nu_subj_scale[j] = Sigma_subj_scale * z_subj_scale[j];
}
for (k in 1 : nwave) {
nu_wave_loc[k] = sigma_wave_loc * z_wave_loc[k];
```

```
nu_wave_scale[k] = sigma_wave_scale * z_wave_scale[k];
}
}
}

model {
//priors
C_subj_loc ~ lkj_corr_cholesky(1);
C_subj_scale ~ lkj_corr_cholesky(1);
//REs
for (j in 1 : nsubj) {
z_subj_loc[j] ~ normal(0, 1);
z_subj_scale[j] ~ normal(0, 1);
}
for (k in 1 : nwave) {
z_wave_loc[k] ~ normal(0, 1);
z_wave_scale[k] ~ normal(0, 1);
}
//likelihood
for (i in 1 : N) {
y[i] ~ normal(X_mean[i] * beta +
Z_subj[i] * nu_subj_loc[subject[i]] +
nu_wave_loc[wave[i]],
sqrt(exp(X_var[i] * alpha +
Z_subj[i] * nu_subj_scale[subject[i]] +
nu_wave_scale[wave[i]])));
```

```
}
}
```

Below is sample code in R to run the stan model and extract the MCMC samples.

```
# Run the stan model
MLS.fit <- stan(data = StanDat,
file = "LSME_model.stan",
init = MLS_init,
iter = 1000,
warmup = 500,
chains = 4,
cores = 10)
# Examine traceplot
traceplot(MLS.fit, pars = "beta")
stan_dens(MLS.fit, pars = "beta", inc_warmup =  TRUE)
# Extract MCMC samples, get point estimate and CI
MCMC.beta <- extract(MLS.fit, pars = "beta")$beta
beta.est <- colMeans(MCMC.beta)
beta.sd <- apply(MCMC.beta, 2, sd)
beta.lower <- apply(MCMC.beta, 2, quantile, probs = 0.025)
beta.upper <- apply(MCMC.beta, 2, quantile, probs = 0.975)
```

2.5 Application to Adolescent Smoking Study

The proposed Bayesian three level mixed effect location scale model was applied to the EMA adolescent smoking study introduced in section 2. For comparison purposes, results from a three level mixed effect regression model, as well as a three level mixed effect regression model with heterogeneous variance were also listed. The focus was on identifying risk factors associated with lowered and unstable mood assessments, with special interest in separating the within subject within wave effect from the between subject and within subject between wave effects of smoking events vs random prompts. The outcome is positive affect, which is a measure of a subjects' positive mood as described in the motivating example section. The occasion level covariate smk (1 if the response is from a smoking event, or 0 if from a random prompt) was decomposed into subject, wave and occasion level variables as described earlier since we are interested in identifying the most significant level of smoking effect. Wave is a continuous variable with values from 0 (baseline) to 6 (6 years after baseline); to facilitate computation, we made one unit equal to five calendar years so that it takes values from 0 to 1.2. In the supplemental materials, we have included R code to simulate a similar three level data set as well as run the Stan program from R.

Results are summarized in Table 2.3, for both mean and variance models. Since parameters were estimated using a Bayesian approach, Hamiltonian Monte Carlo samples were obtained from the posterior distributions for all parameters. The point estimates were obtained as the mean of the posterior distribution for regression coefficients β and α (since their posterior distributions are approximately symmetric), and as the mode of the posterior for random effect variances σ^2 (since their posterior distributions are skewed and mode would be most similar to the MLEs if one were to do likelihood estimation methods). The 95% credible intervals were bounded by 2.5% and 97.5% quantiles of the posterior for all parameters. The first two columns list the parameter estimates and corresponding credible intervals of the three level mixed effect regression model (MRM) which assumes homogeneous error variance and includes random subject location intercept and slope as well as random wave location intercept; the 3rd and 4th columns list results of the three level MRM which has the same random effects specification, but allows the error variance to depend on observed covariates; the final two columns list results of the proposed three level mixed effect location scale model (MLS), which, in addition to the random location effects, also includes the random scale effects and further allows the error variance to depend on both observed and unobserved covariates. The top two panels list regression coefficients for the mean β and within error variance α, with α on the natural log scale; the third panel lists the variances and covariances of the random effects, both for location and scale; and the bottom lists the model selection criteria, the expected log pointwise predictive density, or *elpd*, for all three models. $elpd_{LOO}$ is a measure of how well the model fits the data controlling for the model complexity, and is often used for Bayesian model comparison [Vehtari et al., 2017]. According to Vehtari et al. [Vehtari et al., 2017], $elpd_{LOO}$ is prefered over deviance information criterion (DIC) since it evaluates the likelihood over the entire posterior distribution, works for singular models and is invariant to parametrization. Higher $elpd_{LOO}$ indicates better model fit adjusting for the model complexity.

Table 2.3: Application of the Bayesian Three Level Mixed Effects Location Scale Model on Adolescent Smoking Study

Parameters	MRM		MRM $with$ HV		MLS	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
β^{intcpt}	6.6678	(6.3530, 7.0017)	6.6533	(6.3174, 6.9810)	6.6739	(6.3507, 6.9960)
β^{Male}	-0.0409	(-0.3158, 0.2401)	-0.0415	(-0.3353, 0.2464)	-0.0455	(-0.3166, 0.2121)
$\beta^{smk^{subj}}$	0.5538	(-0.5657, 1.6776)	0.5886	(-0.5824, 1.7126)	0.7113	(-0.4083, 1.8347)
$\beta^{smk^{wave}}$	0.0327	(-0.6442, 0.7056)	0.0278	(-0.6352, 0.6695)	0.0385	(-0.6297, 0.7041)
$\beta^{smk^{obs}}$	0.2310	(0.1927, 0.2685)	0.2260	(0.1896, 0.2615)	0.0965	(0.0689, 0.1238)
β^{wave}	0.4498	(0.2594, 0.6428)	0.4465	(0.2594, 0.6445)	0.4463	(0.2523, 0.6411)
α^{intcpt}	0.6850	(0.6700, 0.6997)	1.0930	(1.0513, 1.1355)	0.8728	(0.6868, 1.0700)
α^{Male}	—	—	-0.1693	(-0.2007, -0.1375)	-0.1379	(-0.3180, 0.0270)
$\alpha^{smk^{subj}}$	—	—	-0.7555	(-0.8816, -0.6342)	-0.5698	(-1.2555, 0.1118)
$\alpha^{smk^{wave}}$	—	—	-0.3535	(-0.5465, -0.1635)	-0.0578	(-0.5989, 0.4788)
$\alpha^{smk^{obs}}$	—	—	-0.0666	(-0.1043, -0.0284)	-0.0600	(-0.1030, -0.0192)
α^{wave}	—	—	-0.2060	(-0.2380, -0.1734)	-0.3211	(-0.4645, -0.1800)
$loc : \sigma^2_{subj\ \ int}$	0.9808	(0.7293, 1.2807)	0.9689	(0.7087, 1.2978)	1.0020	(0.7583, 1.3068)
$loc : \sigma^2_{subj\ \ slope}$	0.6393	(0.3389, 1.0178)	0.6449	(0.3106, 1.1147)	0.6902	(0.3711, 1.0580)
$loc : \sigma_{int\ \ slope}$	-0.1205	(-0.3740, 0.0930)	-0.1171	(-0.4175, 0.1108)	-0.1434	(-0.4051, 0.0698)
$location : \sigma^2_{wave}$	0.4080	(0.3597, 0.5058)	0.4582	(0.3563, 0.4919)	0.3794	(0.3201, 0.4505)
$scale : \sigma^2_{subj\ \ int}$	—	—	—	—	0.2755	(0.1785, 0.3935)
$scale : \sigma^2_{subj\ \ slope}$	—	—	—	—	0.4298	(0.2518. 0.6570)
$scale : \sigma_{int\ \ slope}$	—	—	—	—	-0.0953	(-0.2263, 0.0135)
$scale : \sigma^2_{wave}$	—	—	—	—	0.2729	(0.2218, 0.3138)
$elpd_{LOO}$	-57755.1 (182.8)		-57532.3 (184.1)		-53721.5 (204.7)	

From Table 2.3, all random effect variances in the three level MRM, three level MRM with heterogeneous variance, and three level MLS are estimated to be greater than 0. But since the variance parameters are bounded, a preferred way to judge the significance would be to compare the $elpd_{LOO}$ of the current model to those without corresponding random effects. The model selection criteria $elpd_{LOO}$ strongly favors the three level MLS relative to either the three level MRM or three level MRM with heterogeneous variance. This provides clear evidence that the homogeneous error variance assumption is violated, and observed information is insufficient to explain the amount of variation either at the subject level or at the wave level. Subjects do exhibit heterogeneity in terms of both mood and mood variation, and the heterogeneity in mood variation can be explained by some unmeasurable variables that are absorbed into random subject and wave effects. Specifically, subjects mood variation differs significantly at baseline and changes with different rates over time. The negative covariance between the scale intercept and slope indicates that subjects with more erratic mood at baseline exhibit greater mood stabilization across time, though this is not quite statistically significant as the credible interval includes zero.

When comparing the mean effects β among the models, all three models give similar results except for smk^{obs}, where the two MRMs yield a larger marginal effect compared to MLS. For all three models, smk^{obs} and $wave$ are seen to be statistically significant. For smk^{obs}, the point estimate is positive with 95% credible interval not including 0. This suggests that if we compare the same subject at the same wave, the subject tends to have better mood after a smoking event compared to after a random prompt. For $wave$, the point estimate and credible interval are both positive, indicating that across waves, subjects' mood tends to improve. Although the 95% credible interval contains 0, the results for smk^{subj} and smk^{wave} suggest that, for different subjects, heavier smokers tend to have higher mean mood; for the same subject across different waves, his mood tend to be better after a smoking event compared to a random prompt. Similar results among the three models suggests that, if the main interest is in the mean effects or changes in the mean, the ordinary MRM, MRM with heterogeneous error variance, and MLS all provide valid results.

When comparing the variance effects α among the models, the three level MRM assumes homogeneous error variance and thus only provides an intercept estimate; the three level MRM with heterogeneous variance, on the other hand, has a log linear representation of the error variance, and thus provides a point estimate and corresponding credible interval for each observed covariate. Additionally, the three level MLS further permits unobserved variables to affect the error variance by including the random scale intercept and slope, thus the three level MLS provides α for covariates as well as variances of the random scale effects. Results and conclusions from the latter two models differ, as can be expected based on the simulation study. Since the three level MRM with heterogeneous variance does not include the random scale effects, the point estimates for α might be reasonable, but the credible intervals will likely be too narrow. As can be seen from Table 2.3, the effects of *male*, smk^{subj} and smk^{wave} all tend to be significant by the three level MRM with heterogeneous variance, but not the three level MLS, due to the narrow credible intervals of the former. These positive effects are likely to be false positives since the three level MRM with heterogeneous variance tends to underestimate the uncertainty associated with α. Based on the three level MLS, smk^{obs} and *wave* have negative effects on the variance and are seen to be statistically significant. For smk^{obs}, if we compare the same subject at the same wave, the subject tends to have more consistent mood after a smoking event compared to a random prompt. For *wave*, subjects' mood tends to become more consistent across time, as can be depicted in Figure 2.1. Although the 95% credible interval contains 0, the results for smk^{subj} and smk^{wave} suggest that, for different subjects, heavier smokers tend to have more stable mood; for the same subject across different waves, his mood tends to be more stable after a smoking event compared to a random prompt. The dramatic differences in the results of α among the three models indicate that, neither MRMs provides adequate information or valid statistical inference if the main interest is centered around the variance effects or change of variation, in which case one should consider the proposed location scale model.

Data from four representative subjects were plotted to illustrate the subject and wave heterogeneity. In Figure 2.1, subject 1 was measured at all six waves while subject 2 was measured only at the last two waves. Subject 1 entered the study with relatively bad and unstable mood, but over time, his/her mood became better and more consistent. At the last wave, he/she provided very consistent high positive affect responses. This is consistent with the random subject scale effects estimates $(-0.41, -3.64)$, where the slope effect is estimated to be far below the population average. Subject 2, with the random scale slope estimated to be -0.25, showed a somewhat similar pattern: he/she entered the study at wave 5 with unstable mood assessments, but became more consistent at wave 6. In Figure 2.2, subject 3 was measured at baseline and the last three waves, while subject 4 was measured at baseline and the last two waves. Subject 3 entered the study with relatively stable mood and then remained to be consistent throughout the study. Alternatively, subject 4, who entered the study with erratic mood assessments, then remained erratic until the end, which can also be depicted from his/her random scale intercept estimate $(1.36, 1.52)$, which is above the population average. These four subjects showed distinct patterns in terms of baseline mood variation as well as mood variation trajectories over time.

2.6 Discussion

In this chapter we have extended the existing two level mixed effects location scale model proposed by Hedeker et al. [Hedeker et al., 2008] to a three level structure and additionally allowed for multiple random location and scale effects. The three level mixed effect location scale model allows covariates to influence both the mean and within variance of the outcome, and thus relaxes the homogeneous error variance assumption. This model also includes random effects in both the mean and variance model, allowing variation in the outcome that cannot be fully explained by the covariates. The multiple random effects at the subject and wave levels allows variation in outcome trajectories among subjects and across waves, and provides more realistic assumptions as opposed to simpler random effect models. The magnitude of the random effect variance can help to reveal the degree to which heterogeneity is due to subjects and/or waves. Markov Chain Monte Carlo sampling methods were used to estimate the model parameters and to avoid numerical computation problems caused by the large number of random effects. Our example using the adolescent smoking data showed that subjects experience systematic mood variation at baseline as well as change over time.

The proposed model can be generalized to various research settings where the interest is in both the mean and variation of the outcome and where multiple levels of data clustering are present, such as smoking cessation [Gwaltney et al., 2008] or substance addiction [Lukasiewicz et al., 2007] studies. Since EMA studies produce relatively large number of observations per subject, the location scale model with both mean and variance modeling not only relaxes the constant error variance assumption but also permits more valuable information in terms of the outcome and subject (wave) heterogeneity. Furthermore, the proposed method can also be modified and used in a non-EMA setting where data are collected from a series of hierarchical units instantly. For example, in clinical settings, glucose levels are often measured multiple times per day for type II diabetes patients at possibly multiple waves [Duncan, 2012]. The research interest often involves comparing the possible trajectories as glucose levels evolve with or without insulin pumps and thus infer the effectiveness of insulin pump therapies.

In this chapter, we only considered the possible effects of covariates on the within variance. However, one can also expand our model to additionally allow covariates to influence the between subject, as well as the within subject between wave variance [Hedeker et al., 2012]. To do this, we need to include another set of between variance models. Specifically, let γ_i

denote the random subject location effects, and Σ_{γ_i} be the variance covariance matrix of γ_i. Then the model for the diagonal elements in Σ_γ can be expressed as $\exp(X_i \times \eta)$, where X_i is the set of subject level covariates that have an effect on the between subject variance, and η is the corresponding regression coefficient. Similarly, we can include $\exp(X_{ij} \times \rho)$ to model the diagonal elements in the within subject between wave covariance matrix. Leckie [Leckie, 2014] discussed the option for modeling the two by two between subject covariance matrix by specifying a log link function for the variances and inverse tanh link for the correlation. But it is trickier to extend well to higher order covariance matrices.

Our current work focuses on continuous outcomes only. Future work could therefore extend the current model and estimation framework to ordinal outcomes as well as count outcomes, by including a scale model representation for the overdispersion [Hedeker et al., 2009]. Since ordinal and/or count outcomes generally provide less information compared to continuous outcomes, one might need to collect more data points in order to achieve relatively equal statistical power.

CHAPTER 3

A SHARED PARAMETER LOCATION SCALE MIXED EFFECT MODEL FOR EMA DATA SUBJECT TO INFORMATIVE MISSING

3.1 Introduction

Many scientific investigations generate longitudinal data with missing values, either by intermittent missing or in the form of drop out [Laird, 1988]. In either case, if subjects with missing values behave different in terms of the primary outcome compared to those without, the conventional statistical methods that assumes MCAR (missing completely at random) would yield invalid inference. Furthermore, if missing observations have different distributions conditional on the observed data, models assuming MAR (missing at random) would also not be valid. Modern data collection procedures, such as ecological momentary assessments (EMA), allow researchers to study psychological and behavioral outcomes by repeated sampling in real time fashion [Shiffman et al., 2008]. Typically these procedures involve instant short surveys from individuals over the course of hours, days, and weeks, where relatively large numbers of measurements per subject are produced and intermittent missingness due to non-responses can be an issue [Sokolovsky et al., 2014]. Subjects with a substantial proportion of non-responses could be systematically different in terms of the outcomes compared to those without. An intuitive suspicion would be that, subjects with worse behavioral outcomes or at the occasions when they are experiencing higher levels of stress might respond less often. In psychological and behavioral sciences, within subject variation is another critical metric in characterizing the primary mental outcomes [Martin and Hofer, 2004]. Therefore, within subject variation of the primary outcome can further diverge depending on the missingness. For instance, subjects with unstable outcomes could respond less often compared to those with relatively consistent outcomes. However, research about the informative and intermittent missingness with respect to both the mean and within subject variation of the primary outcomes is rather limited.

In general, three modeling frameworks can be used under the scenario of informative missing. Selection models, first originated in econometrics by Heckman [Heckman, 1979], and later formulated in a longitudinal setting by Diggle and Kenward [Diggle and Kenward, 1994], assume that the observed outcomes are subject to selection bias and a drop out model is introduced to correct for this bias. Pattern mixture models proposed by Little [Little, 1993], partition the joint distribution of the primary outcome and missing process into distinct missing patterns, and compute the joint likelihood conditional on each pattern. Shared parameter models assume that there is a set of latent variables U shared between the primary outcome and missing process, which are conditionally independent given U [Vonesh et al., 2006]. Unlike time to event studies where there are relatively small numbers of missing patterns, in EMA studies intermittent missingness is common and the number of missing patterns could be intrinsically large, making it hard to employ pattern mixture models. We thus resort to shared parameter models, which turn out to make more intuitive sense. Consider the scenario where the primary outcome of interest is mood assessments and data can be intermittently missing due to non-responses. The shared parameter model would assume that some common but unobserved information is shared between subject's mood and the probability of missing, and conditional on the latent information, mood can be modeled independently of the missing process. Recent work has shown that sharing subject's specific location traits between the primary mood outcome and the missing process can significantly improve the model fit [Cursio et al.,]. Our model extends this work by additionally allowing covariates to influence the within subject variance, by including random subject scale (variance) effects, by allowing the random location and scale effects to influence missingness, and by adopting a Bayesian model estimation framework instead of maximum likelihood methods. In particular, no work has been done to explore the possible effect of a subject's scale on the missing process. Furthermore, we investigate the use of the model to impute the missing responses.

Shared parameter models can be difficult to implement due to marginalization of the random effects. Wu and Carroll utilized a maximum likelihood estimation method by numerical integration [Wu and Carroll, 1988]. Follmann and Wu approximated the generalized linear model by conditioning on the data that describes missingness [Follmann and Wu, 1995]. Pulkstenis et al. derived a closed form expression of the marginal likelihood by specifying conjugate random effects for both the outcome and missing process [Pulkstenis et al., 1998]. Unlike these methods that are either sensitive to starting values or restrictive in terms of the random effects distributional assumptions, we propose to estimate the model by a full Bayesian approach without undue restrictions on the distribution of the primary outcome, missing process, or random effects.

In this paper, we develop a comprehensive Bayesian approach to shared parameter models, where the primary outcome and missing process can be related in terms of both the subject's location and scale random effects. The expansion of sharing to include additional scale information makes practical sense in the context of EMA and psychological studies [Nesselroade, 2004]. The primary outcomes are allowed to follow a variety of distributions with heterogeneous error variance, where the heterogeneity is characterized by a location and scale random effect, respectively [Hedeker et al., 2008]. The intermittent missing process is modelled by a logistic regression model including a random subject effect. The missingness random effect is then linked with the outcome location and scale random effects to allow for informative missingness. With the full Bayesian approach, posterior distributions of the model parameters as well as random effects are obtained by MCMC [Bradley and Siddhartha, 1995].

3.2 Motivating Adolescent Mood Study Example

This research is motivated by an EMA study investigating the effects of psychosocial factors on mood regulation among adolescents. The entire EMA study was conducted across 6 waves: baseline, 6 months, 15 months, 2 years, 5 years and 6 years. For illustration purposes, we will focus on data from the baseline wave.

At baseline, 461 adolescents (average age 15.6, minimum 14.4, maximum 16.7) from 9th and 10th grade were asked to carry electronic devices and answer questions when randomly prompted during a 7 day study period. Each individual was prompted multiple times within a single day. Questions included location, activities, companionship, mood and other psychological assessments. The primary outcomes of interest are positive affect (PA) as well as negative affect (NA), which consist of the average of several mood items rated from 1 to 10 that measures subject's positive/negative mood. For PA, questions include: I felt happy, I felt relaxed, I felt cheerful, I felt confident, and I felt accepted by others; for NA, questions include: I felt sad, I felt stressed, I felt angry, I felt frustrated, and I felt irritable. Higher PA levels indicates better mood while higher NA indicates worse mood. Each response will be time stamped regardless of missing status. For the analyses presented here, we will consider subject-level covariates of age, gender, smoking status, negative mood regulation, and the occasion-varying indicator of whether they were alone or with others at the time of the prompt.

Intermittent missingness was generated when an individual did not respond to the prompts. Dropout is not as big a concern here as more than 96.5% individuals were still available at the end of the seven day study period. On average, 22.3% prompts were missing for each individual, with the highest missing proportion per individual being 89.7%. There were a fair number of prompts missing on each study day: 22.6% on day 1, 19.0% on day 2, 20.9% on day 3, 24.6% on day 4, 25.6% on day 5, 25.0% on day 6 and 23.9% on day 7. The proportions of missing prompts were relatively similar during weekdays. In terms of time of day, most missingness occured between 3 am to 9 am (27.2%) and least from 6 pm to 9 pm (20.8%), but the pattern on weekend days was different in that missingness occured mostly between 9 pm to 3 am (52.9% on Saturday and 73.1% on Sunday).

In what follows, we propose a shared parameter model that links the primary outcome with the missing process through a set of random subject effects. For simplicity, we will illustrate the model framework using a normally distributed outcome and binary missing indicator in the context of the example EMA study. However, it can also be extended to other outcome types (e.g., binary or Poisson). We then illustrate a full Bayesian approach for model estimation using MCMC. A series of simulation studies are presented to validate the model estimation procedure, and to examine the use of the model for imputation of missing observations. Finally, the proposed model is applied to the adolescent mood EMA study and results are compared to naive analyses where only the observed data are used.

3.3 Methods

We present the methodology along the lines of Follmann and Wu [Follmann and Wu, 1995], but for a normally distributed and intensively measured longitudinal outcome and binary intermittent missing indicator, which often arise in the context of EMA studies. The approach is to specify the outcome and missing models that share the same set of random effects for each individual.

3.3.1 Model for Intensively Measured Longitudinal Outcomes

Let Y_{ij} be the outcome for individual i at occasion j, where $i = 1, ..., n$, and $j = 1, ..., n_i$ (we allow different individuals to have different number of measurements by subscript n with i). Examples might include mood assessments (PA/NA), craving for food, depression scores, or other psychological measurements. Since Y is measured intensively over time, most of the n_i would be large (usually 20 to 40) compared to traditional longitudinal studies. We specify a mixed effect location scale model for Y_{ij} as described in Hedeker et al. [Hedeker et al., 2008]:

$$Y_{ij} \mid \{\nu_{1,i}, \nu_{2,i}\} \sim \mathcal{N}\left(X_{ij}^\top \beta + P_{ij}^\top \nu_{1,i} \ ; \ \exp(Z_{ij}^\top \alpha + Q_{ij}^\top \nu_{2,i})\right) \tag{3.1}$$

where X_{ij} and Z_{ij} are the fixed effect covariate vectors in the mean and within subject variance model. Both can include subject and occasion level covariates, and usually Z_{ij} contain a subset of variables in X_{ij}; β and α are the corresponding fixed effect coefficient vectors (α in log scale), which indicate the population average effect of the covariates on the mean and (log of) the within subject variability of the outcome. Similarly, P_{ij} and Q_{ij} are the random effect covariate vectors in the mean and variance model, with $\nu_{1,i}$ and $\nu_{2,i}$ being the corresponding random subject location and scale effects, indicating the effect of subject i on his/her mean and within subject variability of the repeated measurements. Usually P and Q are subsets of X and Z. In the case of a random intercept location scale model, P and Q both consist of a column of 1's. The reason to include both location and scale random effects is to allow for subject heterogeneity in both the mean and within subject variability of the outcome that cannot be fully explained by covariates. This relaxes the homogeneous error variance assumption adopted by most statistical methods.

In the context of the EMA adolescent study example, we will model the mood outcomes (PA and NA) as:

$$Y_{ij} = \beta_0 + \beta_1 smoke_i + \beta_2 gender_i + \beta_3 NMR_i + \beta_4 GPA_i + \beta_5 AloneWS_{ij} + \beta_6 AloneBS_i + \nu_{1,i} + \epsilon_{ij} \tag{3.2}$$

where smoke (0 = non-smoker, 1 = smoker), gender (0 = Female , 1 = Male) , NMR (negative mood regulation) and GPA are all subject level covariates. The occasion level covariate Alone is further decomposed into AloneBS and AloneWS, which are the between-subject and within-subject component of Alone, respectively. The reason for the decomposition is to investigate how the effect of being alone on mood differs when comparing the same subject at different occasions (AloneWS) to different subjects averaged over all occasions (AloneBS). This might help to indicate the appropriate level for clinical interventions to be performed at Piasecki et al. [Piasecki et al., 2014]. $\nu_{1,i}$ is the random subject location intercept and reflects the influence of subject i on his/her mood assessments. ϵ_{ij} is the random error and reflects the uncertainty in measuring subject i's mood at occasion j relative to the subject average. The variance of ϵ_{ij} reflects the mood consistency measured for subject i, thus the smaller the variance is, the more stable subject i behaves in terms of his/her mood. To account for the fact that individuals usually exhibit distinct patterns for mood consistency, we additionally model the error variance by

$$\log(\sigma^2_{\epsilon_{ij}}) = \alpha_0 + \alpha_1 smoke_i + \alpha_2 gender_i + \alpha_3 NMR_i + \alpha_4 GPA_i + \alpha_5 AloneWS_{ij} + \alpha_6 AloneBS_i + \nu_{2,i} \tag{3.3}$$

where $\nu_{2,i}$ is the random subject scale intercept and reflects the influence of subject i on the variability of his/her repeated mood assessments. The log function ensures that the error variance is strictly positive. Here we have the same set of covariates in the mean and variance model since the interest is to understand how these covariates affect mood levels as well as the within subject mood variability. The random effects $\{\nu_{1,i}, \nu_{2,i}\}$ are assumed to follow a bivariate normal distribution with mean 0 and some covariance structure. Conditional on $\{\nu_{1,i}, \nu_{2,i}\}$, the mood measurements y_{ij} are *i.i.d.*

3.3.2 Model for the Missing Process

We propose a random intercept logistic regression model for the binary missing prompt indicators. Let M_{ij} be the missing indicator for subject i at occasion j, where M_{ij} is 0 if the subject responds to the prompt and 1 if missing. Since all responses are time stamped, we can investigate whether time of day influences these prompt indicators. Here we assume a typical day starts at 3 am in the morning and divide each day into five time bins: 3 am to 9 am, 9 am to 3 pm, 3 pm to 6 pm, 6 pm to 9 pm and 9 pm to 3 am, and use these time bins as covariates in our modeling of the missing process. Empirical analyses indicate that students tend to behave similar during the weekdays, but quite differently on Saturday and Sunday. Thus, to simplify and facilitate model estimation, we combine each of the five time bins from Monday to Friday, resulting in a total of 15 bins: 5 during weekdays, 5 on Saturday and another 5 on Sunday. The proposed random intercept logistic regression model is given by

$$\log\left(\frac{Pr(M_{ij}=1)}{1-Pr(M_{ij}=1)}\right) = \tau_0 + \sum_{k=2}^{k=15} \tau_k \cdot T_{ij}^k + \lambda_i \tag{3.4}$$

where $k = 2, ..., 15$ is the time bin index and T_{ij}^k is the indicator of the k_{th} time bin for the prompt individual i received at occasion j. For the purpose of model identifiability, the first bin T_{ij}^1 is treated as the reference time bin; τ_0 is the fixed intercept, indicating the log odds of missing a response for an individual with $\lambda_i = 0$ during 3 am to 9 am on a weekday when he/she received a prompt. λ_i is subject i's random intercept, indicating the influence of subject i on his/her log odds of missing prompts. Similar to the model in section 3.3.1, conditional on λ_i, the missing indicators M_{ij} are assumed to be *i.i.d* following a Bernoulli distribution with missing probability $p_{ij} = \frac{\exp(\tau_0+\sum_{k=2}^{k=15}\tau_k\cdot T_{ij}^k+\lambda_i)}{1+\exp(\tau_0+\sum_{k=2}^{k=15}\tau_k\cdot T_{ij}^k+\lambda_i)}$. Here p_{ij} is modelled by both observed and latent information, with time bins being explicitly measured and the random effect λ_i accounting for all unobserved information at the subject level.

3.3.3 Parameter Sharing and Joint Model

Up to this point, the outcome and missing process are still separate. It is possible that there exists some common but unobserved information that contributes to both the outcome and missing process. A legitimate example in the above adolescent mood study would be that, an individual's work schedule cannot be measured but is related to both the mood assessments and missing propensity. For example, individuals with tight schedules might have worse and unstable mood, and they might also be less likely to respond to the prompts. In this case, the unmeasurable work schedule might be represented using the random subject effect λ_i extracted from the missing model. This leads to the parameter sharing below:

$$\nu_{1,i} = \gamma \cdot \lambda_i + \eta_{1,i} \tag{3.5}$$

$$\nu_{2,i} = \delta \cdot \lambda_i + \eta_{2,i} \tag{3.6}$$

where $\{\nu_{1,i}, \nu_{2,i}\}$ and λ_i are random subject effects in the outcome and missing model, and can both be regarded as traits specific to individual i. A set of linear models are used to link $\{\nu_{1,i}, \nu_{2,i}\}$ with λ_i. In equation 3.5, individuals i's location random intercept $\nu_{1,i}$ can be represented by his/her personal missing trait λ_i and an error term $\eta_{1,i}$ that absorbs the residual variation orthogonal to λ_i. The shared parameter coefficient γ indicates the effect of missingness on the subject's mean outcome. Similarly in equation 3.6, δ represents the effect of missingness on the within subject variability of the outcome. By specifying equation 3.5 and 3.6, the primary longitudinal outcome is linked with the missing process through the random subject effects and informative missing can be taken into account. Thus, we call λ the shared random subject effect between the outcome and missingness, and η_1 and η_2 as the residual random subject location and scale effects.

By substituting equation 3.5 and 3.6 into equation 3.2 and 3.3, we can re-express the shared parameter outcome model for this example below as:

$$Y_{ij} = \beta_0 + \sum_{k=1}^{6} \beta_k \cdot x_{ij}^k + \gamma \cdot \lambda_i + \eta_{1,i} + \epsilon_{ij} \tag{3.7}$$

$$\log(\sigma^2_{\epsilon_{ij}}) = \alpha_0 + \sum_{k=1}^{6} \alpha_k \cdot x_{ij}^k + \delta \cdot \lambda_i + \eta_{2,i} \tag{3.8}$$

3.3.4 Model Estimation

Maximum Likelihood Estimation Method

The shared parameter location scale model assumes that conditional on the set of random subject effects $\{\lambda_i, \eta_{1,i}, \eta_{2,i}\}$, the primary outcome Y_{ij} and missing indicator M_{ij} are independent. Thus, we can explicitly write out the conditional joint likelihood of $\mathcal{L}(Y, M \mid \lambda, \eta_1, \eta_2)$ as

$$\mathcal{L}(Y, M \mid \lambda, \eta_1, \eta_2) = \prod_{i=1}^{n} \prod_{j=1}^{n_i} f_N(y_{ij} \mid x_{ij}, \lambda_i, \eta_{1,i}, \eta_{2,i}) \cdot f_B(m_{ij} \mid t_{ij}, \lambda_i) \tag{3.9}$$

where f_N and f_B denote the probability density (mass) function of a Normal and Bernoulli random variable. Specifically, $y_{ij} \mid \lambda_i, \eta_{1,i}, \eta_{2,i} \sim \mathcal{N}\left(x_{ij}^\top \beta + \gamma \cdot \lambda_i + \eta_{1,i}, \exp(x_{ij}^\top \alpha + \delta \cdot \lambda_i + \eta_{2,i})\right)$, and $m_{ij} \mid \lambda_i \sim \mathcal{B}\left(p = \frac{\exp(t_{ij}^\top \tau + \lambda_i)}{1+\exp(t_{ij}^\top \tau + \lambda_i)}\right)$. The marginal joint likelihood $\mathcal{L}(Y, M)$ is then obtained by integrating the conditional joint likelihood $\mathcal{L}(Y, M \mid \lambda, \eta_1, \eta_2)$ over the joint distribution of the random effect vector $\{\lambda, \eta_1, \eta_2\}$.

$$\mathcal{L}(Y, M) = \int \prod_{i=1}^{n} \left\{ \prod_{j=1}^{n_i} f_N(y_{ij} \mid x_{ij}, \lambda_i, \eta_{1,i}, \eta_{2,i}) \cdot f_B(m_{ij} \mid t_{ij}, \lambda_i) \right\} d\mathcal{F}\left(\lambda_i, \eta_{1,i}, \eta_{2,i}\right) \tag{3.10}$$

where equation 3.10 can be further simplified by model assumptions. As mentioned in section 3.3.3, λ_i is independent of $\{\eta_{1,i}, \eta_{2,i}\}$, while $\eta_{1,i}$ and $\eta_{2,i}$ are allowed to be correlated.

Therefore, $d\mathcal{F}\left(\lambda_i, \eta_{1,i}, \eta_{2,i}\right)$ can be factored into $d\mathcal{F}_N(\lambda_i)d\mathcal{F}_N(\eta_{1,i}, \eta_{2,i})$. Once the marginal joint likelihood is computed, optimization can proceed by obtaining the first and second partial derivatives of the (log) marginal likelihood with respect to all model parameters. However, this procedure involves multi-dimensional numerical integration over the random effect distribution and can be computationally challenging. Also, computing the inverse of the Hessian Matrix can be difficult due to the large number of parameters in the joint model (35 in the above adolescent mood study example).

Full Bayesian Estimation Approach

Due to the difficulties in evaluating the marginal joint likelihood as described in MLE based methods, we switch to a full Bayesian estimation approach, where parameters and random effects are regarded as random quantities while data are regarded as fixed. To simplify notation, denote $\theta = (\beta, \alpha, \tau, \gamma, \delta)$ as the model parameter vector, $\lambda = \{\lambda_i\}_{i=1}^n$ as the random subject effects for the missing process, $\eta = \left\{\eta_{1,i}, \eta_{2,i}\right\}_{i=1}^n$ as the random subject effect vector in the outcome model and $D = \{Y_i, M_i\}_{i=1}^n$ as the data.

Since θ, λ and η are all random, they each follow some prior distribution before we get to observe the data D, and we denote the priors as $\pi(\theta)$, $\pi(\lambda)$, and $\pi(\eta)$ respectively. Since individuals are assumed to be independent, $\pi(\lambda)$ can be written as $\prod_{i=1}^n \pi(\lambda_i)$, and similarly for $\pi(\eta)$. Natural choices for $\pi(\lambda_i)$ and $\pi(\eta_i)$ are a univariate standard normal and bivariate standard normal, respectively. For $\pi(\theta)$, one can specify a separate prior for each component in θ provided that a full conditional posterior is obtained for each of them. Given independent priors, one can derive the conditional posterior as

$$P(\theta \mid \lambda_i, \eta_i, D_i) \propto P(D_i \mid \theta, \lambda_i, \eta_i)\pi(\theta) \tag{3.11}$$

$$P(\lambda_i \mid \theta, \eta_i, D_i) \propto P(D_i \mid \theta, \lambda_i, \eta_i)\pi(\lambda_i) \tag{3.12}$$

$$P(\eta_i \mid \theta, \lambda_i, D_i) \propto P(D_i \mid \theta, \lambda_i, \eta_i)\pi(\eta_i) \tag{3.13}$$

$P(D_i \mid \theta, \lambda_i, \eta_i)$ is the conditional joint likelihood given in equation 3.9, and π is the corresponding prior. Once the full conditional posteriors are obtained for θ, λ and η, we can approximate their joint posterior by sampling each variable from its full conditional posterior iteratively using Gibbs sampling [Casella and George, 1992]. In the case where the conditional posterior is not of a recognized form, one can use the Metropolis-Hastings algorithm, which keeps drawing samples from a proposal distribution and decides whether or not to accept the sample as from the conditional posterior with some acceptance rate [Chib and Greenberg, 1995]. We devised a MCMC sampling algorithm where component-wise Metropolis-Hastings algorithms are nested within Gibbs sampling. The detailed Markov Chain Monte Carlo algorithm where (component wise) Metropolis-Hastings is nested within Gibbs sampling is listed below, with q being the corresponding (user defined) proposal distribution. After enough runs, the chains will ultimately converge to the joint posterior and one can summarize the posterior samples to get the parameter estimates as well as credible intervals.

1. Initialize at $(\theta, \lambda, \eta) = \left(\theta^0, \lambda^0, \eta^0\right)$

2. Sample a single random value iteratively from each full conditional posterior by Metropolis-Hastings algorithm below, for $t = 1, 2, ...$

   (a) Given the current value of $\theta^t = \theta$, generate a proposed new value θ' according to $q_\theta(\theta \to \cdot)$, and accept θ' with probability $A_\theta = min\left(1, \frac{P(\theta'|\lambda^t,\eta^t,D)q_\theta(\theta'\to\theta)}{P(\theta|\lambda^t,\eta^t,D)q_\theta(\theta\to\theta')}\right)$

   (b) Given the current value of $\lambda^t = \lambda$, generate a proposed new value λ' according to $q_\lambda(\lambda \to \cdot)$, and accept λ' with probability $A_\lambda = min\left(1, \frac{P(\lambda'|\theta^{t+1},\eta^t,D)q_\lambda(\lambda'\to\lambda)}{P(\lambda|\theta^{t+1},\eta^t,D)q_\lambda(\lambda\to\lambda')}\right)$

   (c) Given the current value of $\eta^t = \eta$, generate a proposed new value η' according to $q_\eta(\eta \to \cdot)$, and accept η' with probability $A_\eta = min\left(1, \frac{P(\eta'|\theta^{t+1},\lambda^{t+1},D)q_\eta(\eta'\to\eta)}{P(\eta|\theta^{t+1},\lambda^{t+1},D)q_\eta(\eta\to\eta')}\right)$

3. Repeat step a, b c in (2) until convergence.

However, a better approach can be taken using Stan, an open source Hamiltonian Monte Carlo sampler, since it can better deal with the trade off between step size and acceptance rate by reducing the correlation between successive samples using a Hamiltonian evolution and target values with a higher acceptance rate than the observed probability distribution [Stan-Development-Team, 2014]. Both the MCMC derivation and Stan implementation details are provided in the Supplemental Materials. The Hamiltonian Monte Carlo sampling uses improper uniform priors (uniform on $(-\infty, +\infty)$) for regression coefficients, improper bounded uniform priors (uniform on $(0, +\infty)$) for random effect variances, and an LKJ prior for the random effect correlation matrix.

3.4 Simulation

To validate the proposed model and estimation procedure, we conducted a series of simulation studies and present the results here. Because of the heavy computation load, we limited the number of simulations to 100 under each scenario. Generally, results became quite consistent after the first 25 simulations and remained consistent till the end.

For each simulation, an intensively measured longitudinal outcome Y was generated via a location scale process for 100 individuals at a total of 30 occasions. Covariates included gender (subject level) and time stamps (occasional level). Once the complete data were generated, observations were set to intermittent missing via a Bernoulli process, where the missing probability was simulated under two scenarios: 1) missing does not depend on potential outcomes (MCAR or MAR), and 2) missing depends on potential outcomes (MNAR). For each scenario, analyses were conducted using two candidate methods: a) a naive model which assumes MAR and utilizes only the observed outcome, and b) the proposed model that shares random subject location and scale effects between the outcome and missing process.

The detailed model specifications are shown in equation 3.14 and 3.15 for the naive model and equation 3.16 - 3.18 for the proposed model, where Y_{ij} denotes the outcome and M_{ij} denotes the missing indicator.

Naive Model:

$$Y_{ij} = \beta^{int} + \beta^{gender} \cdot gender_i + \beta^{time} \cdot time_{ij} + \nu_{1,i} + \epsilon_{ij} \tag{3.14}$$

$$\log(\sigma^2_{\epsilon_{ij}}) = \alpha^{int} + \alpha^{gender} \cdot gender_i \tag{3.15}$$

Proposed Model:

$$\log\left(\frac{Pr(M_{ij} = 1)}{1 - Pr(M_{ij} = 1)}\right) = \tau^{int} + \tau^{gender} \cdot gender_i + \tau^{time} \cdot time_{ij} + \lambda_i \tag{3.16}$$

$$Y_{ij} = \beta^{int} + \beta^{gender} \cdot gender_i + \beta^{time} \cdot time_{ij} + \gamma \cdot \lambda_i + \eta_{1,i} + \epsilon_{ij} \tag{3.17}$$

$$\log(\sigma^2_{\epsilon_{ij}}) = \alpha^{int} + \alpha^{gender} \cdot gender_i + \delta \cdot \lambda_i + \eta_{2,i} \tag{3.18}$$

Results are summarized in Table 3.1. β are the regression coefficients for the outcome mean model, and α are for the within subject variance model (on log scale). $\sigma^2_{\nu_1}$ and $\sigma^2_{\nu_2}$ are the variances for the random subject location and scale effects. γ and δ are the coefficients for the shared parameters, indicating the influence of missingness on the mean and within subject variance, as described in section 3.3.3. The point estimates are obtained as the posterior mean for regression coefficients β, α, γ and δ since their posteriors are approximately symmetric, and as mode for random effect variances $\sigma^2_{\nu_1}$ and $\sigma^2_{\nu_2}$ since their posteriors are relatively skewed. Bias is computed for each parameter as the average point deviation from the true value: $Bias = \sum_{k=1}^{100}(\hat{\theta}_k - \theta)/100$, where $\hat{\theta}_k$ denotes the posterior mean for $(\beta, \alpha, \gamma, \delta)$ and mode for $\left(\sigma^2_{\nu_1}, \sigma^2_{\nu_2}\right)$ from the k_{th} simulation. AIW (average interval width) is computed as the average range between the 97.5% and 2.5% quantile of the posterior: $AIW = \sum_{k=1}^{100}(\theta^U_k - \theta^L_k)/100$, where θ^U_k and θ^L_k are the 97.5% and 2.5% quantile of the posterior distribution from the k_{th} simulation. For each parameter, we also calculate the number of times out of 100 that the 95% credible interval contains its true value, thus providing the coverage rate as $COV = \sum_{k=1}^{100} \mathbb{1}\left\{\theta_k^L \leqslant \theta \leqslant \theta_k^U\right\}/100$.

Under MAR, the naive model and proposed model are expected to perform similar as the missing process is independent of the potential outcome. This is also confirmed in Table 3.1 from the small bias, reasonable AIW and correct COV for both models. Under MNAR, however, the two models diverge in terms of the inference for α. The shared parameter model has both smaller bias and much better coverage rate for α^{intcp} and α^{gender} compared to the naive model. Furthermore, the small bias and correct coverage of $\sigma^2_{\nu_2}$, γ and δ also provide evidence of the validity of the proposed model. This clearly shows that the naive analyses, which ignore the association between the primary outcome and the missing process, can lead to invalid inference, particularly for the variance model parameters. The mean model parameters and random location effect variance seem less likely to be affected as all bias is absorbed into the error variance components. Overall, the proposed shared parameter model achieves good estimation precision, correct interval length and asymptotic coverage rate, yet provides insightful information about the missing mechanisms.

In addition, we imputed the missing values under the two scenarios by both candidate models and compared the imputated values with the true (missing) outcomes. Under MAR, both the naive and shared parameter models achieve small imputation bias (0.0005 vs 0.0007) and correct coverage rate (94.8% vs 94.9%). Under MNAR, however, the naive model produces greater imputation bias (0.162 vs 0.005) and an inadequate coverage rate (86.5% vs 94.9%) as compared to the shared parameter model.

Table 3.1: Simulation Results under two Scenarios: MCAR/MAR and MNAR

Parameter	True Value	Naive Model			Shared Parameter Model		
		Bias	AIW	COV	Bias	AIW	COV
Scenario MCAR/MAR:							
β^{int}	0.30	$-2.2 \cdot 10^{-2}$	0.836	98%	$-2.3 \cdot 10^{-2}$	0.842	95%
β^{gender}	0.20	$8.5 \cdot 10^{-4}$	1.162	97%	$1.6 \cdot 10^{-3}$	1.171	96%
β^{time}	0.01	$-7.1 \cdot 10^{-5}$	0.012	92%	$-9.2 \cdot 10^{-5}$	0.012	92%
α^{int}	0.30	$-4.7 \cdot 10^{-3}$	0.181	94%	$-7.2 \cdot 10^{-3}$	0.192	97%
α^{gender}	0.10	$1.5 \cdot 10^{-3}$	0.253	96%	$2.2 \cdot 10^{-3}$	0.265	96%
$\sigma^2_{\nu_1}$	2.00	$1.26 \cdot 10^{-1}$	1.266	94%	$1.46 \cdot 10^{-1}$	1.276	93%
$\sigma^2_{\nu_2}$	0	—	—	—	$7.7 \cdot 10^{-3}$	0.032	—
γ	0	—	—	—	$8.9 \cdot 10^{-4}$	0.468	97%
δ	0	—	—	—	$1.34 \cdot 10^{-2}$	1.886	96%
Scenario MNAR:							
β^{int}	0.30	$4.6 \cdot 10^{-2}$	0.863	96%	$-8.1 \cdot 10^{-3}$	0.858	97%
β^{gender}	0.20	$3.0 \cdot 10^{-2}$	1.185	95%	$2.9 \cdot 10^{-2}$	1.195	93%
β^{time}	0.01	$5.0 \cdot 10^{-4}$	0.014	96%	$3.1 \cdot 10^{-4}$	0.009	95%
α^{int}	0.30	$1.76 \cdot 10^{-1}$	0.185	29%	$1.2 \cdot 10^{-3}$	0.622	95%
α^{gender}	0.10	$2.9 \cdot 10^{-2}$	0.260	43%	$8.3 \cdot 10^{-3}$	0.880	95%
$\sigma^2_{\nu_1}$	2.00	$1.34 \cdot 10^{-1}$	1.326	93%	$1.27 \cdot 10^{-1}$	1.378	99%
$\sigma^2_{\nu_2}$	1.00	—	—	—	$8.1 \cdot 10^{-2}$	0.739	95%
γ	-1.00	—	—	—	$-5.8 \cdot 10^{-2}$	0.656	97%
δ	1.00	—	—	—	$5.6 \cdot 10^{-2}$	0.671	93%

3.5 Application to Adolescent Mood Study Example

3.5.1 Results and Conclusions

In this section, we revisit the example introduced in section 3.2. An important aim of the study is to identify factors that can potentially influence the mean and within subject variance of individuals' positive/negative mood. Informative missing is likely to occur since mood can only be assessed if individuals respond to the prompt, and whether or not they decide to respond may be affected by their mood at the time of the prompt. For every prompt individual i received, we record the missing indicator vector M_i (1 if missing, 0 if respond), mood assessment vector Y_i (PA or NA) if not missing and the time window T_i ($T^k_{ij} = 1$ if the prompt occurred in the k_{th} window for $k = 1, ..., 15$).

Three candidate models are applied to the example EMA data: 1) a random intercept model with heterogeneous variance (HV) that provides covariance adjustment for the correlation among repeated measurements and allows covariates to affect the within subject variance, 2) the shared location model that not only provides inference for the outcome mean and within subject variance, but also allows missingness and the outcome mean to be correlated through shared random subject effect, and 3) the proposed shared location scale model that associates each individual's missingness with both the mean and within subject variance of the primary outcome. The detailed model specifications are shown below, where $X_{ij} = (1, smk_i, gender_i, NMR_i, GPA_i, AloneBS_i, AloneWS_{ij})$ is the covariate vector for the mean and within subject variance model, $\beta = (\beta^{int}, \beta^{smk}, \beta^{gender}, \beta^{NMR}, \beta^{GPA}, \beta^{AloneBS}, \beta^{AloneWS})$ is the regression coefficient vector for the mean model, and $\alpha = (\alpha^{int}, \alpha^{smk}, \alpha^{gender}, \alpha^{NMR}, \alpha^{GPA}, \alpha^{AloneBS}, \alpha^{AloneWS})$ is the regression coefficient vector for the within subject variance model.

Random Intercept HV Model:

$$Y_{ij} = X_{ij}^{\top}\beta + \nu_{1,i} + \epsilon_{ij} \tag{3.19}$$

$$\log(\sigma^2_{\epsilon_{ij}}) = X_{ij}^{\top}\alpha + \nu_{2,i} \tag{3.20}$$

Shared Location Model:

$$\log\left(\frac{Pr(M_{ij}=1)}{1-Pr(M_{ij}=1)}\right) = \tau_0 + \sum_{k=2}^{k=15} \tau_k \cdot T_{ij}^k + \lambda_i \tag{3.21}$$

$$Y_{ij} = X_{ij}^{\top}\beta + \gamma \cdot \lambda_i + \eta_{1,i} + \epsilon_{ij} \tag{3.22}$$

$$\log(\sigma^2_{\epsilon_{ij}}) = X_{ij}^{\top}\alpha + \nu_{2,i} \tag{3.23}$$

Shared Location Scale Model:

$$\log\left(\frac{Pr(M_{ij}=1)}{1-Pr(M_{ij}=1)}\right) = \tau_0 + \sum_{k=2}^{k=15} \tau_k \cdot T_{ij}^k + \lambda_i \tag{3.24}$$

$$Y_{ij} = X_{ij}^\top \beta + \gamma \cdot \lambda_i + \eta_{1,i} + \epsilon_{ij} \tag{3.25}$$

$$\log(\sigma^2_{\epsilon_{ij}}) = X_{ij}^\top \alpha + \delta \cdot \lambda_i + \eta_{2,i} \tag{3.26}$$

To better compare the parameter estimates and credible intervals, we performed the Bayesian approach in section 3.3.4 for all three candidate models. Results are summarized in Table 3.2 for Positive Affect and Table 3.3 for Negative Affect. Again, the point estimates are obtained as the posterior mean for regression coefficients β, α, γ and δ, and as mode for random effect variances $\sigma^2_{\nu_1}$ and $\sigma^2_{\nu_2}$. The 95% credible intervals (CI) are obtained as the 2.5% and 97.5% posterior quantiles for all parameters. The model selection criteria $elpd_{LOO}$, proposed by Vehtari et al. [Vehtari et al., 2017], estimates the pointwise leave one out (LOO) prediction accuracy from a fitted Bayesian model by evaluating the log likelihood over the posterior samples. It is prefered over the deviance information criterion (DIC) since it accounts for the entire posterior distribution, works for singular models and is invariant to parametrization. Higher $elpd_{LOO}$ indicates better model fit adjusting for the model complexity.

For covariate effects on the mean of the mood outcome, the shared location model and shared location scale model give relatively similar estimates that are different from the random intercept HV model, except for $AloneWS$. Specifically for Positive Affect, higher negative mood regulation is significantly associated with higher PA, while higher GPA and being alone (both at subject and occasion levels) are associated with lower PA. Although not statistically significant, the trend indicates that smokers tend to have lower PA compared to non-smokers and males tend to have better PA than females, adjusting for all other covariates. For all three candidate models, the magnitude of the effect of $AloneBS$ on PA is estimated to be three times as big as $AloneWS$, suggesting that the between- and within-subject effects are not equal, though of the same sign. Subjects who are alone more often report lower average

PA (between-subject effect), and when subjects are alone they also report lower PA (within-subject effect). For Negative Affect, in Table 3.3, the mean effect estimates are of opposite sign and lead to similar conclusions, as higher NA indicates worse mood.

For covariate effects on the within subject variance, PA and NA models provide relatively similar coefficient estimates since both the within subject variance of PA and NA reflect individuals' mood consistency/inconsistency. The random intercept HV model and shared location model give similar effect estimates as well as narrower credible intervals that are different from the proposed shared location scale model. This is as expected since neither of the two former models include scale random effects for the within subject variance and there is no parameter sharing between the outcome variation and the missing process. Therefore we will refer to the proposed shared location scale model for coefficient interpretation. Specifically for Positive Affect, higher negative mood regulation is significantly associated with more stable mood, which is in agreement with the theory that higher negative mood regulation indicates better mood control. Males tend to have more stable PA compared to females, and subjects with lower GPA tend to have more stable PA. Although not statistically significant, the trend indicates that smokers tend to have more erratic PA compared to non-smokers adjusting for all other covariates. The three candidate models disagree on the effect of *Alone*, where the random intercept HV model and shared location model indicate that the between subject component $AloneBS$ contributes significantly, while the shared location scale model indicates increased variation for the within subject component $AloneWS$. Thus, comparing responses from the same subject at different occasions, the subject's PA is more variable when he/she is alone compared to when he/she is with others.

As an extra benefit from the shared parameter model, there seems to be a negative association between the missingness and PA mean (positive for NA), as indicated by the estimate of γ. This is in agreement with our hypothesis that a lower response rate is related to worse mood. Although not significant at the 5% level, the positive estimate of δ on PA and NA indicates that a lower response rate is also associated with unstable mood. For PA, there is a negative association between the random location and scale random effect possibly due to a ceiling effect (i.e., subjects with high PA means tend to have lower scale due to the ceiling of measurement). This association is positive for NA indicating that subjects with lower means have lower variability, possibly due to a floor effect of measurement. The shared location scale model achieves great improvement in terms of the model fit over the other two models adjusting for complexity, as is shown by $elpd_{LOO}$, and is thus preferred.

Table 3.4 summarizes the estimated effects in the missingness model for all 15 time windows (5 time bins for weekdays as well as for Saturday and Sunday) for both PA and NA. Generally, one would expect to obtain very similar estimates for both PA and NA since, when prompted, students are most likely to answer them both or neither. During weekdays, the most prompts were answered late in the day (from 6 pm to 9 pm) and students become least attentive early in the day (from 3 am to 9 am). For the weekend days, students answer most prompts from 9 am to 3 pm on Saturday and from 6 pm to 9 pm on Sunday. Students generally behave less responsive on Saturday, and a bit more responsive on Sunday (except from 3 am to 9 am) during the same time frame as compared with weekdays.

In addition to the above analyses, we also performed cross validation using the adolescent mood data. Specifically, observations were set to missing according to various missing scenarios, then missing observations were imputed by candidate models trained from the available data (as well as missing patterns), and finally imputed values were compared with the true values based on posterior prediction accuracy as measured by the model likelihood ($elpd_{CV}$). As before, candidate models include 1) a random intercept HV model that assumes MAR, 2) a shared parameter location model that only links the outcome mean with the missing process, and 3) a shared parameter location scale model that links both PA mean and variance with the missing process. In Table 3.5, γ and δ are the coefficients in the parameter sharing model and denote the effect of missingness on PA mean and variance, respectively. In the first scenario, where both γ and δ were set to 0 and missingness does not depend on potential outcomes (MAR or MCAR), the first two candidate models should achieve similar posterior prediction accuracy (since sharing location or scale does not contribute to model fit). The shared location scale model achieves slightly higher prediction accuracy due to the inclusion of random scale effects. In the second scenario, where δ is set to 0 and missingness only depends on the outcome mean, the shared location model should perform better than the random intercept HV model since it corrects for MNAR by linking information between the PA mean with the missingness. Again, the shared location scale model performs slightly better than the shared location model due to the inclusion of random scale effects. In the third scenatio, where δ is allowed to vary and missingness depends on both the PA mean and variance, only the shared location scale model is able to capture the correct missing pattern. The further away δ moves from 0, the greater benefit the shared location scale model achieves over the random intercept HV and shared location models.

3.5.2 Software Implementation

The details of implementing the three candidate models in Stan and R (using the package "RStan") are provided below.

Candidate model 1: Random Intercept HV model

```
data {
int<lower=1> N_outcome;
int<lower=1> nsubj;
int<lower=1, upper=nsubj> subject_outcome[N_outcome];
row_vector[7] X_outcome[N_outcome];
```

Table 3.2: Comparison of parameter estimates and credible intervals between the random intercept model with heterogeneous variance, and shared parameter location model, and shared parameter location scale model: Positive Affect.

Parameter	Random Intercept HV Model		Shared Location Model		Shared LS Model	
	Estimate	CI	Estimate	CI	Estimate	CI
β^{int}	6.084	(5.463, 6.742)	6.152	(5.532, 6.826)	6.115	(5.422, 6.789)
β^{smk}	-0.135	(-0.330, 0.071)	-0.143	(-0.358, 0.055)	-0.146	(-0.363, 0.076)
β^{gender}	0.204	(-0.011, 0.440)	0.210	(-0.153, 0.412)	0.212	(-0.019, 0.428)
β^{NMR}	0.640	(0.491, 0.803)	0.647	(0.466, 0.811)	0.658	(0.500, 0.821)
β^{GPA}	-0.148	(-0.302, -0.011)	-0.170	(-0.315, -0.022)	-0.172	(-0.319, -0.020)
$\beta^{AloneBS}$	-1.070	(-1.597, -0.536)	-1.059	(-1.64, -0.450)	-0.999	(-1.529, -0.450)
$\beta^{AloneWS}$	-0.328	(-0.387, -0.270)	-0.329	(-0.391, -0.268)	-0.256	(-0.307, -0.205)
α^{int}	1.518	(1.367, 1.670)	1.512	(1.371, 1.661)	1.286	(0.868, 1.713)
α^{smk}	0.006	(-0.043, 0.054)	0.006	(-0.043, 0.059)	0.033	(-0.109, 0.170)
α^{gender}	-0.220	(-0.274, -0.166)	-0.220	(-0.271, -0.167)	-0.270	(-0.421, -0.114)
α^{NMR}	-0.176	(-0.214, -0.136)	-0.175	(-0.211, -0.140)	-0.160	(-0.263, -0.050)
α^{GPA}	-0.082	(-0.114, -0.049)	-0.081	(-0.114, -0.048)	-0.080	(-0.179, 0.027)
$\alpha^{AloneBS}$	0.321	(0.189, 0.451)	0.320	(0.191, 0.447)	0.330	(-0.021, 0.676)
$\alpha^{AloneWS}$	0.049	(-0.010, 0.105)	0.051	(-0.007, 0.109)	0.107	(0.047, 0.168)
$\sigma^2_{\nu_1}$	1.025	(0.998, 1.333)	1.093	(1.013, 1.317)	1.215	(1.013, 1.326)
$\sigma^2_{\nu_2}$	—	—	—	—	0.530	(0.415, 0.572)
ρ_{ν_1,ν_2}	—	—	—	—	-0.294	(-0.358, -0.275)
γ	—	—	-0.109	(-0.242, 0.006)	-0.134	(-0.254, -0.006)
δ	—	—	—	—	0.061	(-0.0030, 0.150)
$elpd_{LOO}$	-23866	—	-23863	—	-22815	—

```
real outcome[N_outcome];

}

parameters {

vector[7] beta;

vector[7] alpha;

real<lower=0> sigma_subj_loc;

real z_subj_loc[nsubj];

}

transformed parameters {

real vars_subj_loc;

real nu_subj_loc[nsubj];

{

vars_subj_loc = sigma_subj_loc * sigma_subj_loc;

for (j in 1 : nsubj) {
```

Table 3.3: Comparison of parameter estimates and credible intervals between the random intercept model with heterogeneous variance, and shared parameter location model, and shared parameter location scale model: Negative Affect.

Parameter	Random Intercept HV Model		Shared Location Model		Shared LS Model	
	Estimate	CI	Estimate	CI	Estimate	CI
β^{int}	4.550	(3.756, 5.293)	4.591	(3.818, 5.349)	4.516	(3.773, 5.307)
β^{smk}	0.374	(0.086, 0.639)	0.366	(0.082, 0.626)	0.359	(0.100, 0.616)
β^{gender}	-0.354	(-0.612, -0.092)	-0.377	(-0.652, -0.105)	-0.384	(-0.640, -0.123)
β^{NMR}	-0.825	(-1.020, -0.628)	-0.863	(-1.036, -0.685)	-0.853	(-1.052, -0.663)
β^{GPA}	0.241	(0.071, 0.415)	0.262	(0.087, 0.424)	0.275	(0.101, 0.440)
$\beta^{AloneBS}$	0.172	(-0.516, 0.901)	0.175	(-0.509, 0.850)	0.106	(-0.569, 0.770)
$\beta^{AloneWS}$	0.199	(0.132, 0.268)	0.201	(0.136, 0.263)	0.090	(0.045, 0.135)
α^{int}	1.803	(1.654, 1.945)	1.803	(1.663, 1.955)	1.689	(1.105, 2.215)
α^{smk}	0.132	(0.082, 0.184)	0.132	(0.078, 0.185)	0.204	(0.020, 0.374)
α^{gender}	-0.236	(-0.286, -0.185)	-0.236	(-0.289, -0.185)	-0.370	(-0.570, -0.169)
α^{NMR}	-0.183	(-0.221, -0.146)	-0.182	(-0.221, -0.144)	-0.299	(-0.436, -0.150)
α^{GPA}	-0.078	(-0.111, -0.045)	-0.078	(-0.111, -0.045)	-0.047	(-0.160, 0.069)
$\alpha^{AloneBS}$	-0.030	(-0.164, 0.103)	-0.030	(-0.170, 0.113)	0.062	(-0.446, 0.546)
$\alpha^{AloneWS}$	-0.001	(-0.061, 0.057)	-0.001	(-0.060, 0.056)	0.023	(-0.041, 0.088)
$\sigma^2_{\nu_1}$	1.625	(1.544, 2.069)	1.643	(1.551, 2.061)	1.714	(1.527, 2.033)
$\sigma^2_{\nu_2}$	—	—	—	—	0.758	(0.745, 1.015)
ρ_{ν_1,ν_2}	—	—	—	—	0.413	(0.375, 0.443)
γ	—	—	0.176	(0.017, 0.343)	0.175	(0.017, 0.338)
δ	—	—	—	—	0.086	(-0.031, 0.207)
$elpd_{LOO}$	-25339	—	-25339	—	-23840	—

```
nu_subj_loc[j] = z_subj_loc[j] * sigma_subj_loc;

}

}

}

model {

for (j in 1 : nsubj) {

z_subj_loc[j] ~ normal(0, 1);

}

for (i in 1 : N_outcome) {

outcome[i] ~ normal(X_outcome[i] * beta + nu_subj_loc[subject_outcome[i]],

sqrt(exp(X_outcome[i] * alpha)));

}

}

generated quantities{
```

Table 3.4: Parameter estimates and credible intervals for the missing process model: Positive Affect and Negative Affect.

Parameter	Positive Affect		Negative Affect	
	Estimate	CI	Estimate	CI
τ_0	-1.281	(-1.453, -1.103)	-1.285	(-1.477, -1.085)
τ : Weekday 9 am - 3 pm	-0.245	(-0.428, -0.064)	-0.240	(-0.424, -0.053)
τ : Weekday 3 pm - 6 pm	-0.112	(-0.310, 0.078)	-0.104	(-0.300, 0.089)
τ : Weekday 6 pm - 9 pm	-0.270	(-0.465, -0.078)	-0.264	(-0.460, -0.059)
τ : Weekday 9 pm - 3 am	-0.144	(-0.360, 0.062)	-0.138	(-0.354, 0.080)
τ : Saturday 3 am - 9 am	0.847	(0.290, 1.412)	0.845	(0.277, 1.421)
τ : Saturday 9 am - 3 pm	-0.100	(-0.361, 0.159)	-0.096	(-0.359, 0.157)
τ : Saturday 3 pm - 6 pm	0.117	(-0.171, 0.404)	0.126	(-0.151, 0.397)
τ : Saturday 6 pm - 9 pm	0.032	(-0.279, 0.326)	0.041	(-0.240, 0.324)
τ : Saturday 9 pm - 3 am	0.251	(-0.057, 0.543)	0.256	(-0.043, 0.538)
τ : Sunday 3 am - 9 am	1.725	(1.043, 2.463)	1.739	(1.005, 2.521)
τ : Sunday 9 am - 3 pm	-0.234	(-0.495, 0.032)	-0.229	(-0.481, 0.042)
τ : Sunday 3 pm - 6 pm	-0.341	(-0.673, -0.033)	-0.333	(-0.667, -0.025)
τ : Sunday 6 pm - 9 pm	-0.465	(-0.789, -0.158)	-0.459	(-0.777, -0.157)
τ : Sunday 9 pm - 3 am	-0.313	(-0.647, 0.013)	-0.303	(-0.640, 0.032)
σ^2_λ	0.767	(0.666, 0.959)	0.717	(0.636, 0.945)

```
vector[N_outcome] log_lik;

for(i in 1:N_outcome){

log_lik[i] = normal_lpdf(outcome[i] |

X_outcome[i] * beta + nu_subj_loc[subject_outcome[i]],

sqrt(exp(X_outcome[i] * alpha)));

}

}
```

Candidate model 2: Shared location model

```
data {

int<lower=1> N_miss;

int<lower=1> N_outcome;

int<lower=1> nsubj;
```

Table 3.5: Comparison of Cross Validated Prediction Accuracy (measured by $elpd_{CV}$) under Various Missing Scenarios.

Scenarios γ	δ	Random Intercept HV Model	Shared Location Model	Shared LS Model
0	0	-6183.44	-6183.82	-5953.89
-1	0	-7597.93	-7507.31	-7316.31
-2	0	-9697.15	-8879.17	-8606.67
-4	0	-12398.75	-10617.76	-10468.86
-1	1	-9133.91	-8950.74	-8500.87
-1	2	-10396.42	–10165.195	-9595.23
-1	4	-12899.99	-12299.93	-11621.36

```
int<lower=1, upper=nsubj> subject_miss[N_miss];

int<lower=1, upper=nsubj> subject_outcome[N_outcome];

row_vector[15] X_miss[N_miss];

row_vector[7] X_outcome[N_outcome];

int miss[N_miss];

real outcome[N_outcome];

}

parameters {

vector[7] beta;

vector[7] alpha;

vector[15] tau;

real gamma;

real<lower=0> sigma_subj_loc;

real<lower=0> sigma_miss;

real z_subj_loc[nsubj];

real z_miss[nsubj];

}

transformed parameters {
```

```
real vars_subj_loc;
real vars_miss;
real nu_subj_loc[nsubj];
real nu_miss[nsubj];
{
vars_subj_loc = sigma_subj_loc * sigma_subj_loc;
vars_miss = sigma_miss * sigma_miss;
for (j in 1 : nsubj) {
nu_subj_loc[j] = z_subj_loc[j] * sigma_subj_loc + gamma * z_miss[j];
nu_miss[j] = z_miss[j] * sigma_miss;
}
}
}
model {
for (j in 1 : nsubj) {
z_subj_loc[j] ~ normal(0, 1);
z_miss[j] ~ normal(0, 1);
}
for (i in 1 : N_miss){
miss[i] ~ bernoulli_logit(X_miss[i] * tau + nu_miss[subject_miss[i]]);
}
for (i in 1 : N_outcome) {
outcome[i] ~ normal(X_outcome[i] * beta + nu_subj_loc[subject_outcome[i]],
sqrt(exp(X_outcome[i] * alpha)));
}
}
generated quantities{
```

```
vector[N_outcome] log_lik;
for(i in 1:N_outcome){
log_lik[i] = normal_lpdf(outcome[i] |
X_outcome[i] * beta + nu_subj_loc[subject_outcome[i]],
sqrt(exp(X_outcome[i] * alpha)));
}
}
```

Candidate model 3: Shared location Scale model

```
data {
int<lower=1> N_miss;
int<lower=1> N_outcome;
int<lower=1> nsubj;
int<lower=1, upper=nsubj> subject_miss[N_miss];
int<lower=1, upper=nsubj> subject_outcome[N_outcome];
row_vector[15] X_miss[N_miss];
row_vector[7] X_outcome[N_outcome];
int miss[N_miss];
real outcome[N_outcome];
}
parameters {
vector[7] beta;
vector[7] alpha;
vector[15] tau;
real gamma;
real delta;
real<lower=0> sigma_subj_loc;
real<lower=0> sigma_subj_scale;
```

```
real<lower=0> sigma_miss;
real z_subj_loc[nsubj];
real z_subj_scale[nsubj];
real z_miss[nsubj];
}
transformed parameters {
real vars_subj_loc;
real vars_subj_scale;
real vars_miss;
real nu_subj_loc[nsubj];
real nu_subj_scale[nsubj];
real nu_miss[nsubj];
{
vars_subj_loc = sigma_subj_loc * sigma_subj_loc;
vars_subj_scale = sigma_subj_scale * sigma_subj_scale;
vars_miss = sigma_miss * sigma_miss;
for (j in 1 : nsubj) {
nu_subj_loc[j] = z_subj_loc[j] * sigma_subj_loc + gamma * z_miss[j];
nu_subj_scale[j] = z_subj_scale[j] * sigma_subj_scale + delta * z_miss[j];
nu_miss[j] = z_miss[j] * sigma_miss;
}
}
}
model {
for (j in 1 : nsubj) {
z_subj_loc[j] ~ normal(0, 1);
z_subj_scale[j] ~ normal(0, 1);
```

```
z_miss[j] ~ normal(0, 1);
}
for (i in 1 : N_miss){
miss[i] ~ bernoulli_logit(X_miss[i] * tau + nu_miss[subject_miss[i]]);
}
for (i in 1 : N_outcome) {
outcome[i] ~ normal(X_outcome[i] * beta + nu_subj_loc[subject_outcome[i]],
sqrt(exp(X_outcome[i] * alpha + nu_subj_scale[subject_outcome[i]])));
}
}
generated quantities{
vector[N_outcome] log_lik;
for(i in 1:N_outcome){
log_lik[i] = normal_lpdf(outcome[i] |
X_outcome[i] * beta + nu_subj_loc[subject_outcome[i]],
sqrt(exp(X_outcome[i] * alpha + nu_subj_scale[subject_outcome[i]])));
}
}
```

3.6 Discussion

In this paper, we have developed a general shared parameter framework for normally distributed and intensively measured longitudinal outcomes with informative missingness. We exploit a mixed effect location scale model proposed by Hedeker et al. [Hedeker et al., 2008] for the primary outcome, and a random intercept logistic model for the intermittent missing process. The two models are connected by the random subject location and scale effects, which are assumed to capture the common but unmeasurable information at the subject level that contributes to both the primary outcome and missing process. Due to the computational burden in multi-dimensional numerical integration, we propose a Bayesian MCMC estimation approach where the joint posterior distribution can be approximated by samples drawn from a conventional Metropolis-Hasting-Gibbs algorithm. Various improvements such as Hamiltonian Monte Carlo can be used to balance the trade off between step size and acceptance rate, and ultimately make the chains converge faster.

For simplicity, we have assumed a linear relationship between the outcome location/scale random effects and the missingness random effect with some measurement error. However, this relationship could be made more general. Since the scale random effects are on the log metric, it could be the case that they are related with the missingness random intercept on the original scale, i.e, $\exp(\nu_{2,i}) = \delta \cdot \lambda_i + \eta_{2,i}$. However, this might introduce additional difficulties in model estimation. Also, we have assumed a simple random intercept logistic model for the missing process, which might not be a good fit in some situations. As an extension, one might try a random slope model for the missingness where each individual has a bivariate random effect vector $\{intercept_i, slope_i\} = \left\{\lambda_{1,i}, \lambda_{2,i}\right\}$, representing the influence of individual i on his/her baseline missing propensity as well as the rate of change. In this case, a more sophisticated sharing mechanism would be required to connect $\left\{\nu_{1,i}, \nu_{2,i}\right\}$ with $\left\{\lambda_{1,i}, \lambda_{2,i}\right\}$, which might not be trivial.

The simulation results under the MAR assumption show almost no difference between the naive and proposed model. One might prefer the naive method given it is more parsimonious and easier to estimate. However, one cannot know whether data are missing at random or not in practice when the true underlying mechanism is unknown. As is shown in the adolescent mood study example, the naive model underperforms in terms of the out of sample prediction accuracy relative to the proposed approach. Also, the estimated γ and δ coefficients both suggest evidence towards non-random missingness. Therefore, the shared parameter model would seem to be a safer choice and might be preferred in practice.

In psychological and behavioral sciences, research interests are usually focused on both the actual magnitude of the outcome as well as the within subject variability [Steven et al., 2014]. For example, clinician or patient rated average levels of depression, as well as their variability, are both critical aspects in characterizing depressed and bipolar patients. Identifying factors that can potentially influence the mean and within subject variation of the psychological outcomes can jointly provide deep insights for clinical intervention. Further, as both the outcome mean and variability can be correlated with individuals' propensity of responding, it is essential to share both the location and scale random effects between the outcome and missingness models. For binary and Poisson outcomes, where the variability is a 1-1 function of the mean, one can replace the bivariate $\left\{\nu_{1,i}, \nu_{2,i}\right\}$ vector in the sharing model with the univariate $\nu_{1,i}$ scaler. An alternative model specification when there is evidence for over-dispersion is to model the over-dispersion parameter with a set of scale random effects, and adopt a similar sharing mechanism as in the proposed model [Hedeker et al., 2009].

CHAPTER 4

MIXED LOCATION SCALE HIDDEN MARKOV MODEL WITH AN APPLICATION TO ECOLOGICAL MOMENTARY ASSESSMENT DATA

4.1 Introduction

Hidden Markov Models (HMMs) are often used to describe the relationship between two stochastic processes: an underlying hidden (latent) process that is assumed to follow a Markov chain, and an observed process that are modeled as some independent distributions conditional on the hidden process. Individual differences are to be expected in many psychological applications, and thus HMMs with random effects in the observed process, also referred to as mixed HMMs, offer a flexible alternatives in analyzing longitudinal data when a person alternates between discrete states in a wide range of situations [Altman, 2007]. Using mixed HMMs, researchers are able to identify discrete affective states (e.g., pleasant-unpleasant mood, calm-tense mood), cognitions (e.g., appraisals, self-esteem) and behaviors (e.g., treatment compliance, drinking baheviors) as well as how these latent states fluctuate over time, in the presents of subject heterogeneity [Crayen et al., 2012]. Few HMMs have been considered in the context of EMA studies where intensively measured longitudinal outcomes provide opportunities to extract state-switch information on both the outcome mean and within subject variability [Shiffman et al., 2008].

To account for individual differences, both observed covariates and subject specific random effects are incorporated in the conditional distributions for the observed process. The inclusion of subject specific random effects are crucial in latent state classification as it reduces the bias arising from unobserved subject level variables. The addition of random effects in the error variance is a natural extension of the mixed HMMs and provides an effective way to valid statistical inference and latent state classification with regard to within subject variability, which is another aspect of importance for psychological and behaviral research [Hedeker et al., 2008].

In this article, we focus on extending the mixed HMMs to include additional random effects in error variance of the conditional distribution, which applies in the context of intensive longitudinal data and provides a general framework for working in EMA context. The augmented model allows subject heterogeneity in both the mean and within subject variability of the outcomes in the observed process. This article is organized as follows. In section 2, an EMA adolescent mood study example is introduced and described to motivate the proposed model. In section 3, both mixed HMMs and our proposed model are explained in the context of the motivating example, and Bayesian estimation methods are briefly mentioned. In section 4, we will compare the performance of HMM, mixed HMM and mixed location scale HMM using a series of simulation studies. Finally, the article concludes with the results of applying the models to the motivating example and a discussion.

4.2 Motivating Example

This research is motivated by an EMA study investigating the effects of psychosocial factors on mood regulation among adolescents. The entire EMA study was conducted across 6 waves: baseline, 6 months, 15 months, 2 years, 5 years and 6 years. For illustration purposes, we will focus on data from the baseline wave. The detailed data description can be found in a previous study [Lin et al., 2018].

At baseline, 461 adolescents (average age 15.6, minimum 14.4, maximum 16.7) from 9th and 10th grade were asked to carry electronic devices and answer questions when randomly prompted during a 7 day study period. Each individual was prompted multiple times within a single day. Questions included location, activities, companionship, mood and other psychological assessments. The primary outcomes of interest are positive affect (PA) as well as negative affect (NA), which consist of the average of several mood items rated from 1 to 10 that measures subject's positive/negative mood. For PA, questions include: I felt happy, I felt relaxed, I felt cheerful, I felt confident, and I felt accepted by others; for NA, questions include: I felt sad, I felt stressed, I felt angry, I felt frustrated, and I felt irritable. Higher PA levels indicates better mood while higher NA indicates worse mood. For the analyses presented here, we are interested in the effects of subject level variable smoking status (indicator of whether a person smokes during the study period) and occasion level variable alone (indicator of whether a person is alone or with others at the time of the prompt) on PA. Since mood assessments are taken at multiple random time intervals within each day during a seven day study period and traditional HMMs require the time points to be equally spaced, we aggregate the observations within each day and calculate the average PA as well as alone status on each day and use those in the following model formulation.

Mood states (e.g., pleasant - unpleasant) fluctuates and the pattern of fluctuation is partly due to mood regulation behavior. Research demonstrated that adolescents differ considerably in their ability to effectively regulate their mood, and even for the same individual, mood regulation can change quite a lot over time [Rocke et al., 2009]. For instance, at a certain time point, individuals can be classified into distinct subgroups based on their mood regulation ability; the same individual can be classified into very different subgroups at different time points. Mixed Hidden Markov Models are well suited to investigate both the interindividual and intraindividual differences since these models take into account the time serial dependence of mood states for the same individual while adjusting for the individual differences at each time point [MacDonald and Zucchini, 1997]. For example, an adolescent is more likely to be in a pleasant mood state if he or she has a pleasant mood state at the previous time point; some adolecsents have the tendency to have better mood compared to others though they might have the same measured characteristics.

In what follows, we propose a more comprehensive mixed HMM that takes into account of individual differences in both the mean and within subject variability of the outcomes. For simplicity, we will illustrate the model framework using a normally distributed outcome, one subject level covariate (smoking status) and one occasion level covariate (being alone or not) in the context of the example EMA study.

4.3 Methods

In this section, we propose a mixed location scale Hidden Markov Model on top of the existing mixed HMMs. We start by introducing the mixed HMMs developed for longitudinal data settings. We then extend the mixed HMMs in the context where the longitudinal outcomes are intensively measured for each subject.

4.3.1 Mixed Hidden Markov Models

Let Y_{it} denote the observed outcome, and Z_{it} denote the hidden state associated with subject i, $i = 1, ..., N$, at time t, $t = 1, ..., T_i$. Y_i denotes the T_i dimensional vector of observed time series for subject i, and Y the vector of observed time series for all subjects. The vector of hidden states, Z_i, denotes the latent group membership for subject i from time 1 all the way to time T_i.

We assume that, given the random effects, $\{Z_{it}\}_{t=1}^{T_i}$ are first order Markov Chains, such that

$$z_{t+1} \perp\!\!\!\perp z_{t-1} \mid z_t \tag{4.1}$$

Thus, the conditional distribution of the hidden state Z_t can be written explicitly in the form

$$P(z_t|z_{t-1}, A) = \prod_{k=1}^{K}\prod_{j=1}^{K} A_{jk}^{z_{t-1,j}, z_{t,k}} \tag{4.2}$$

where z_t is the latent state at time t and z_{t-1} is the latent state at time $t-1$. A is the transition probability matrix and is assumed to be the same across all subjects at any successive time points. A_{jk} is the probability that subjects transit from state j at time $t-1$ to state k at time t. The initial latent state z_1 is a special case and its marginal distribution can be represented by

$$P(z_1|\pi) = \prod_{k=1}^{K} \pi_k^{z_{1,k}} \tag{4.3}$$

We also assume that, given the random effects, the i_{th} process, $\{Y_{it}\}_{t=1}^{T_i}$ is a HMM, and observed time series on different processes are independent. Conditional on the random effects ν_i and hidden states z_i, Y_{it} are independently and normally distributed,

$$Y_{it}|z_{it} = k, X_{it}, \nu_i, \beta \;\; \sim \;\; \mathcal{N}(\mu_{itk}, \sigma_{ik}^2) \tag{4.4}$$

and

$$\mu_{itk} = X_{it}\beta_k + \nu_i, \;\; \sigma_{ik}^2 = \exp(W_{it}\alpha_k) \tag{4.5}$$

Here X_{it} is the covariate vector for the outcome mean, and includes intercept, smoking status (subject level variable) and alone (occasion level variable) in the above adolescent mood study example; W_{it} is the covariate vector for the within subject variance and is usually a subset of X_{it}. β_k is the fixed effect regression coefficient vector in the outcome mean for latent state k, and α_k is the fixed effect regression coefficient vector in the outcome within subject variance for latent state k. ν_i is the subject specific random effect, representing the influence of subject i on his or her mood assessment. It is assumed to be the same across all time points and normally distributed with mean 0 and constant variance σ_ν^2.

The conditional distribution of the observed outcome given latent state and random effects are called emission probability. The above emission model indicates that, at time t, the observed outcomes depend on both subject effects ν_i, and the latent state k they are in. For example, as in the case of pleasant - unpleasant mood category, the observed mood assessments in the pleasant mood category will be higher than those in the unpleasant category. In addition to the group effect, there could be significant subject effects such that some subjects tend to have better and consistent mood compared to others.

4.3.2 Mixed Location Scale Hidden Markov Model - An Extension of Mixed Hidden Markov Models

Here we only focus on the addition of subject random effects in the conditional model for the observed process and assume that subject specific differences do not appear in the model for the hidden process. In particular, the conditional model include both random subject effects in the outcome mean (location) and within subject variance (scale) and thus inherit the name mixed location scale HMMs. Also, we continue to assume that the hidden process is homogeneous for all subjects with common initial probabilities π_k and transition probability $A_{j,k}$. In the above adolescent mood study example, the proposed model would allow both the mean mood assessment and mood variability to vary among subjects.

Using the same notation as in 4.3.1, let $(\nu_{1,i}, \nu_{2,i})$ denote the random location and scale effect vector for subject i. The proposed mixed location scale HMMs can be expressed by two process: the hidden process has the exact same model specifications as in 4.3.1, while the observed process can be expressed as below

$$Y_{it}|z_{it} = k, X_{it}, \nu_{1,i}, \nu_{2,i}, \beta \;\sim\; \mathcal{N}(\mu_{itk}, \sigma^2_{ik}) \tag{4.6}$$

and

$$\mu_{itk} = X_{it}\beta_k + \nu_{1,i}, \;\; \sigma^2_{ik} = \exp(W_{it}\alpha_k + \nu_{2,i}) \tag{4.7}$$

X_{it} and W_{it} are the covariate vector for the mean and within subject variance model, as in 4.3.1 and include an intercept, subject level variable variable smoking status and occasion level variable alone. β_k and α_k are the regression coefficient vectors for the above two sets of models. In addition to the random subject location effect $\nu_{1,i}$ which is also included in the conditional model in 4.3.1, the proposed model also has another set of random effects $\nu_{2,i}$, which represent the specific influence of subject i on his or her within subject variance of the outcome, and are thus called random subject scale effect. The bivariate random subject location and scale effects (ν_1, ν_2) are assumed to be time invariant and normally distributed with mean vector 0 and covariance matrix Σ_ν.

Similar to mixed HMMs, the conditional model in mixed location scale HMMs indicates that, at time t, the observed outcomes depend on both subject location and scale effects $\left(\nu_{1,i}, \nu_{2,i}\right)$ and the latent state k they are in. In the adolescant mood study example, if subjects are to be classofied into two distinct subgroups based on their mood assessments - pleasant versus unpleasant mood category, individuals in group one will have higher mood mean and lower mood variability compared to those in group two. In addition to the group effect, there could be significant subject effects such that some subjects tend to have better and more consistent mood, despite the fact that they might have the same measured covariates. The addition of random subject scale effects in the novelty of the proposed model and is of great importance. Omitting the random scale effects would lead to incorrect subgroup classification, especially when the interest is in the within subject variance, as the true group membership will be confounded by the individual differences introduced by unmeasured subject level variables.

4.3.3 Model Estimation for Mixed Location Scale Hidden Markov Model

Parameters in HMMs are often estimated by the expectation - maximumization (EM) algorithm [Cappe, 2012]. Since random effects are introduced for mixed HMMs, a revised EM algorithm can be used for model estimation where the conditional likelihood in the expectation step is marginalized over the distribution of random effects. In the case of bivariate random effects in mixed location scale HMM, let θ denote the 2-dimensional random subject location and scale effect vector, where $\theta = \{\nu_1, \nu_2\}$, and ϕ denote the model parameters, where $\phi = \{\beta, \alpha, \Sigma_\nu\}$. The margimal likelihood for this model can be written as

$$\begin{aligned}
&\mathcal{L}(\phi; y) \\
&= \int_\theta \sum_z f(y|z, \theta, \phi) f(z; \phi) f(\theta; \phi) d\theta \\
&= \int_\theta \sum_z \left\{ \prod_{i=1}^{N} \prod_{t=1}^{T_i} f(y_{it}|z_{i,t}, \theta, \phi) \right\} \times \left\{ \prod_{i=1}^{N} \pi_{z_{i1}} \prod_{t=2}^{t_i} P_{z_{i,t-1}, z_{i,t}} \right\} f(\theta; \phi) d\theta \\
&= \int_\theta \prod_{i=1}^{N} \left\{ \sum_{z_i} \pi_{z_{i1}} f(y_{i1}|z_{i1,\theta,\phi}) \times \prod_{t=2}^{T_i} P_{z_{i,t-1}, z_{i,t}} f(y_{it}|z_{it}, \theta, \phi) \right\} f(\theta; \phi) d\theta
\end{aligned}$$

For a given value of θ, denote C_i^1 as the vector with K elements, where $c_{i,k}^1 = \pi_k f(y_{i1}|z_{i1} = k, \theta)$. Similarly, denote C_i^t as the matrix in K by K dimension, with the k, l element $c_{i,kl}^t = A_{k,l} f(y_{it}|z_{it} = l, \theta)$ for $t >= 2$. Then the likelihood can be re-written as

$$\mathcal{L}(\phi; y) = \int_\theta \prod_{i=1}^{N} \left\{ \left(C_i^1\right)^\top \left(\prod_{t=2}^{T_i} C_i^t \right) \mathbb{1}_K \right\} f(\theta; \phi) d\theta \tag{4.8}$$

The likelihood associated with each i is simply the conditional likelihood of a single HMM given the bivariate random effect. And for a single subject, C_i^1 is the joint likelihood of being in the K latent state at time 1, while C_i^t is the joint likelihood of transitioning to the K latent state at time t from the K latent state at time $t-1$, conditional on the random effects. For the E-M algorithm estimation, let $l = log(\mathcal{L})$ denote the log likelihood, ϕ^{old} denote the parameter estimates from previous iteration. First we need to derive the joint posterior distribution of $\{z, \theta\}$ given the observed data and parameter estimates from previous iteration.

$$
\begin{aligned}
&p(z, \theta \mid y, \phi_{old}) \\
&= \frac{f(y \mid z, \theta, \phi^{old}) \ f(z \mid \theta, \phi^{old}) \ f(\theta; \phi^{old})}{f(y; \phi^{old})} \\
&= \frac{f(y \mid z, \theta, \phi^{old}) \ f(z \mid \theta, \phi^{old}) \ f(\theta; \phi^{old})}{\int_\theta \sum_z f(y \mid z, \theta, \phi^{old}) \ f(z \mid \theta, \phi^{old}) \ f(\theta; \phi^{old}) d\theta}
\end{aligned}
$$

The integral can be computed using Monte Carlo Gibbs sampling [McCulloch, 1997]. Specifically, we generate B samples from $f(\theta; \phi^{old})$, and compute the posterior of z given each generated θ and observed data y,

$$
f(z \mid \theta^b, \phi^{old}) = \frac{f(y \mid z, \theta^b, \phi^{old}) \ f(z \mid \theta^b, \phi^{old})}{\sum_{l=1}^B \sum_z f(y \mid z, \theta^l, \phi^{old}) \ f(z \mid \theta^l, \phi^{old})} \tag{4.9}
$$

Then we can approximate the expectation of log likelihood conditional on $\{z, \theta\}$ by

$$
Q(\phi, \phi^{old}) = \mathbb{E}(\phi; y, z, \theta) \mid y, \phi^{old} \approx \frac{1}{B} \sum_{b=1}^B \sum_z l(\phi^{old}; y, z, \theta^b) \ f(z \mid \theta^b, \phi^{old}) \tag{4.10}
$$

Once we have the expectation of the log likelihood, we can maximize it with respect to the model parameters and update ϕ^{old}. Iterate the Expectation and Maximization steps until the model parameters converge by a pre-specified threshold. An open source Hamiltonian Monte Carlo sampler Stan can be used to estimate HMMs.

4.4 Simulation Study

To validate the proposed mixed location scale HMM and estimation procedure, we conducted a series of simulation studies under two scenarios and present the results here. Because of the heavy computation load, we limited the number of simulations to 100 under each scenario.

For each simulation, an intensively measured longitudinal outcome Y was generated according to 1) a mixed HMM model and 2) a mixed location scale HMMs for 100 individuals at a total of 10 equally spaced time points. For both scenarios, each subject has equal chance to be placed into two distinct subgroups at t_0 (initial probability $p_1 = p_2 = 0.5$), where those in group one have relatively lower mean and higher within subject variability compared to group two ($\mu_1 = 1, \sigma_{\epsilon,1} = 2, \mu_2 = 2, \sigma_{\epsilon,2} = 1$. At later time points $t > 1$, the probability of staying in group one given already in group one at $t-1$ is set to be 0.9, and 0.8 for group two (transition probability matrix $A = \begin{bmatrix} tp_{11} & tp_{12} \\ tp_{21} & tp_{22} \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$. For scenario 1, subject random location effect variance $\sigma_{\nu,loc} = 1$; for scenario 2, subject random location scale random effects covariance matrix $\Sigma_\nu = \begin{bmatrix} 1 & -0.3\sqrt(3) \\ -0.3\sqrt(3) & 3 \end{bmatrix}$. Once the data are generated, they are analyzed by three candidate models: 1) HMM with no random effects, 2) HMM with random subject location effects in the conditional model (mixed HMM), 3) HMM with both random subject location and scale effects in the conditional model (mixed location scale HMM). For each candidate model, point estimate as well as 2.5% and 97.5% quantiles are recorded for p_1, tp_{11}, tp_{22}, μ_1, μ_2, $\sigma_{\epsilon,1}$, $\sigma_{\epsilon,2}$, $\sigma_{\nu,loc}$ and Σ_ν. Finally, bias, average interval width and coverage rate are calculated for the above parameters using results from the 100 simulations.

Results are summarized in Table 4.1. p_1 is the initial probability of being in group one. tp_{11} and tp_{22} are the probabilities of being in group one or group two given subjects already in that same group at the previous time point. μ_1 and $\sigma_{\epsilon,1}$ are the mean and within subject variance for subjects in group one, and the same thing applies to μ_2 and $\sigma_{\epsilon,2}$. $\sigma_{\nu,loc}$ and $\sigma_{\nu,scale}$ are the variances for random subject location and scale effects. Bias is computed for each parameter as the average point deviation from the true value: $Bias = \sum_{k=1}^{100}(\hat{\theta}_k - \theta)/100$, where $\hat{\theta}_k$ denotes the posterior mean for $(p_1, tp_{11}, tp_{22}, \mu_1, \mu_2)$ and mode for $\left(\sigma_{\epsilon,1}, \sigma_{\epsilon,2}, \sigma_{\nu,loc}, \sigma_{\nu,scale}\right)$ from the k_{th} simulation. AIW (average interval width) is computed as the average range between the 97.5% and 2.5% quantile of the posterior: $AIW = \sum_{k=1}^{100}(\theta_k^U - \theta_k^L)/100$, where θ_k^U and θ_k^L are the 97.5% and 2.5% quantile of the posterior distribution from the k_{th} simulation. For each parameter, we also calculate the number of times out of 100 that the 95% credible interval contains its true value, thus providing the coverage rate as $COV = \sum_{k=1}^{100} \mathbb{1}\left\{\theta_k^L \leqslant \theta \leqslant \theta_k^U\right\}/100$.

If the data are generated under the first scenarion where subjects only exhibit heterogeneity in terms of location, the last two candidate models - mixed HMM and mixed location scale HMM behave similar interms of the bias and coverage rate while HMM produces large bias and insufficient coverage rate. This can be shown by Figure 4.1, where true parameter values and bias from all three candidate models are plotted against each other. Although mixed location scale HMM produces slightly larger bias compared to mixed HMM, the estimates are still accurate enough and the coverage rates show well-bahaved consistency. Under the second scenario where subjects are allowed to exhibit heterogeneity in both location and scale, only the proposed mixed location scale HMM provide valid parameter estimates and correct coverage rate. Neither HMM nor mixed HMM perform well as can be shown by Figure 4.2. Overall, the proposed mixed location scale HMM achieves good estimation accuracy,

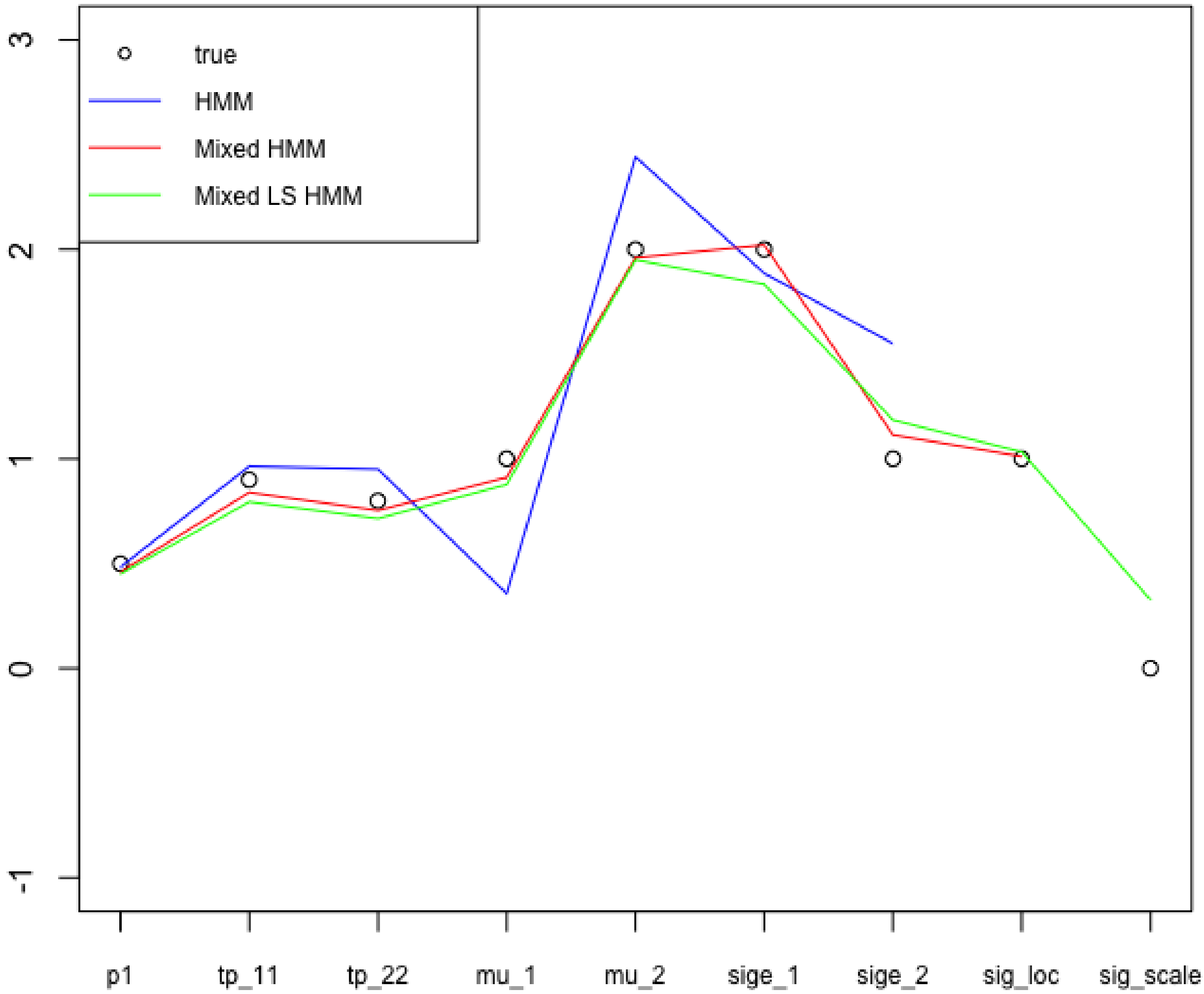

Figure 4.1: Scenario 1: Comparison of Parameter Estimates from HMM, Mixed HMM and Mixed Location Scale HMM

reasonable interval width and asymptotic coverage rate under both scenarios and provides insightful information about individual differences in various perspectives.

4.5 Application to Adolescent Mood Study Example

In this section, we revisit the example introduced in section 4.2. One question of interest is to investigate the effect of smoking and being alone in regulating adolescents mood activity, including both their mean mood level and mood consistency over time. Subjects may fall

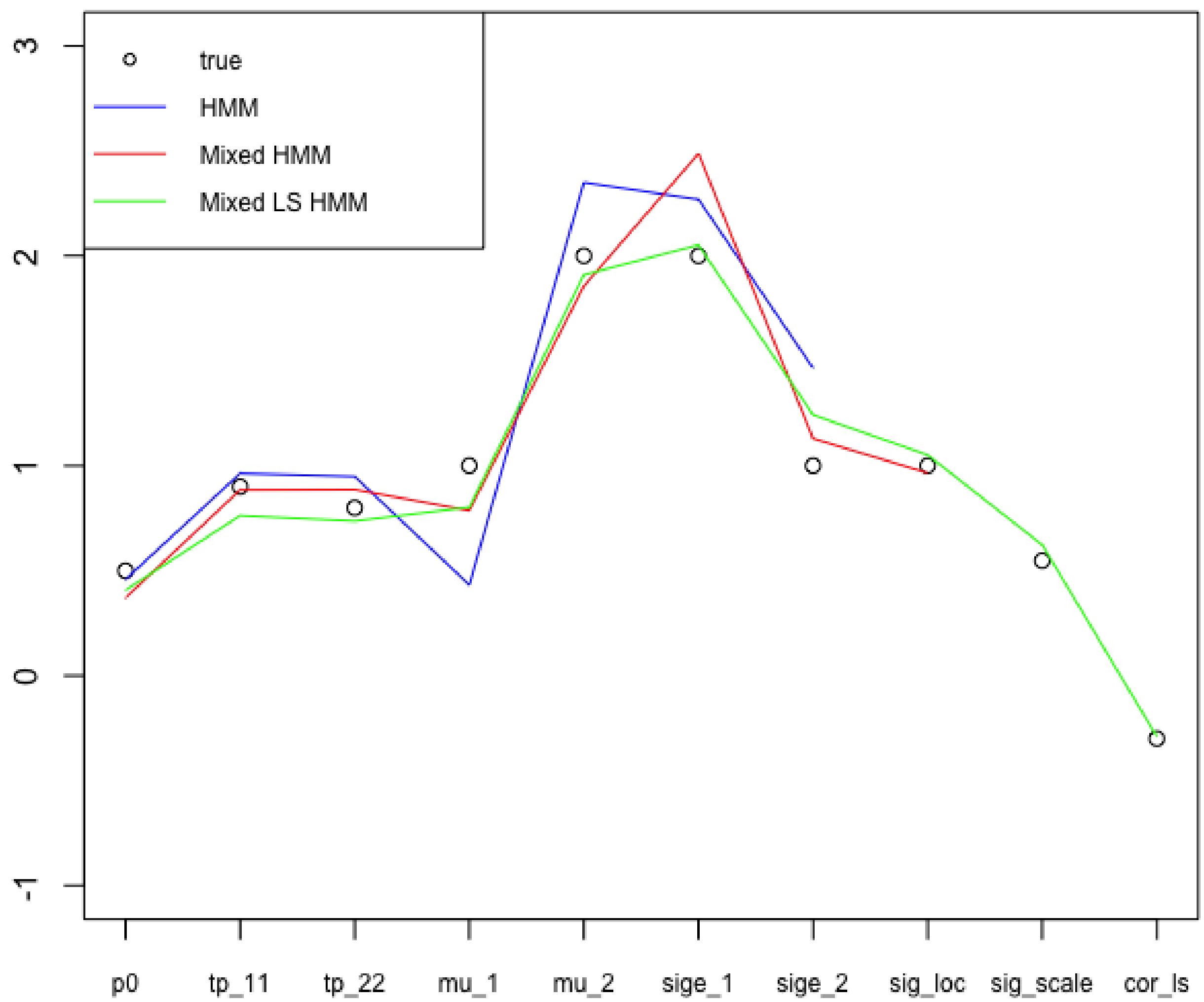

Figure 4.2: Scenario 2: Comparison of Parameter Estimates from HMM, Mixed HMM and Mixed Location Scale HMM

Table 4.1: Simulation Results under two Scenarios: location process / location & scale process

Parameter	True Value	HMM			mixed HMM			mixed LS HMM		
		Bias	AIW	COV	Bias	AIW	COV	Bias	AIW	COV
Scenario 1:										
p_1	0.50	-0.017	0.322	83%	-0.039	0.478	94%	-0.049	0.565	98%
tp_{11}	0.90	0.064	0.076	25%	-0.060	0.319	97%	-0.106	0.468	96%
tp_{22}	0.80	0.151	0.087	0%	-0.046	0.336	93%	-0.085	0.450	96%
μ_1	1.0	-0.643	0.623	4%	-0.088	0.874	96%	-0.123	1.169	95%
μ_2	2.0	0.442	0.559	23%	-0.040	0.715	93%	-0.050	0.798	94%
$\sigma_{\epsilon,1}$	2.0	-0.114	0.279	69%	0.021	0.442	89%	-0.167	0.562	96%
$\sigma_{\epsilon,2}$	1.0	0.550	0.302	0%	0.114	0.489	91%	0.185	0.606	92%
$\sigma_{\nu,loc}$	1.0	—	—	—	0.011	0.386	97%	0.033	0.403	97%
$\sigma_{\nu,scale}$	0	—	—	—	—	—	—	0.327	0.393	—
Scenario 2:										
p_1	0.50	-0.041	0.337	84%	-0.129	0.365	67%	-0.092	0.657	97%
tp_{11}	0.90	0.063	0.083	33%	-0.015	0.193	96%	-0.138	0.554	91%
tp_{22}	0.80	0.149	0.084	1%	0.086	0.151	44%	-0.064	0.513	98%
μ_1	1.0	-0.568	0.720	13%	-0.213	0.882	84%	-0.198	1.440	98%
μ_2	2.0	0.349	0.515	38%	-0.144	0.594	78%	-0.090	0.871	89%
$\sigma_{\epsilon,1}$	2.0	0.269	0.422	37%	0.487	0.585	14%	0.051	0.443	94%
$\sigma_{\epsilon,2}$	1.0	0.465	0.332	3%	0.129	0.347	62%	0.075	0.480	96%
$\sigma_{\nu,loc}$	1.0	—	—	—	-0.035	0.380	97%	0.033	0.403	97%
$\sigma_{\nu,scale}$	0.548	—	—	—	—	—	—	0.009	0.536	94%

into distinct subgroups that differ significantly in terms of their mood regulating abilities, and assuming a homogeneous effect thoughout the entire population would lead to undefinite inference and incomplete conclusions. In addition, individual heterogeneity is likely to exist since not all information related to mood regulation can be measured and some adolescents appear to be consistently happier than others despite the fact that they carry the same measurable characteristics. Failing to account for these individual heterogeneity would lead to incorrect subgroup classification.

Let Y_{it} denote the intensively measured longitudinal outcome - positive effect (PA), smk_i (1 for smoker and 0 for non-smoker) and $alone_{it}$ (1 if alone and 0 if with others) the subject and ocassion level covariates, and Z_{it} the latent mood state associated with subject i, $i = 1, ..., N$, at time t, $t = 1, ..., T_i$. Here we are interested in the pleasant - unpleasant (for the mean) and consistent - erratic (for the within subject variance) mood state and thus Z_{it} has two categories ($K = 2$). π_k denotes the initial probability for subject to be in mood state k with $\sum_{k=1}^{K} \pi_k = 1$. A denotes the transition probability matrix for the hidden process, and the element A_{jk} denotes the probability of transition from mood state j at time point $t-1$ to mood state k at time t.

Three candidate models are applied to the example EMA data: 1) HMM (without random subject effect in the conditional model), 2) mixed HMM (with subject location random effect in the conditional model), and 3) mixed location scale HMM (with subject random location and scale effect in the conditional model). The three candidate models have the same model specifications for the hidden process, which can be expressed as

$$z_{t+1} \perp\!\!\!\perp z_{t-1} \mid z_t \tag{4.11}$$

$$P(z_t|z_{t-1}, A) = \prod_{k=1}^{K}\prod_{j=1}^{K} A_{jk}^{z_{t-1,j}, z_{t,k}} \tag{4.12}$$

and

$$P(z_1|\pi) = \prod_{k=1}^{K} \pi_k^{z_{1,k}} \tag{4.13}$$

but different observed process (conditional model). Specifically, for the proposed mixed locations scale HMM,

$$\begin{aligned} Y_{it}|z_{it} = k, smk_i, alone_{it}, \beta_0, \beta_{1,k}, \beta_{2,k}, \alpha_0, \alpha_0 1, k, \alpha_0 2, k \\ \sim \mathcal{N}(\beta_0 + smk_i\beta_{1,k} + alone_{it}\beta_{2,k} + \nu_{1,i}, \exp(alpha_0 + smk_i\alpha_{1,k} + alone_{it}\alpha_{2,k} + \nu_{2,i})) \end{aligned} \tag{4.14}$$

$$Y_{it}|z_{it} = k, X_{it}, \nu_{1,i}, \nu_{2,i}, \beta \sim \mathcal{N}(\mu_{itk}, \sigma_{ik}^2) \tag{4.15}$$

and

$$\mu_{itk} = X_{it}\beta_k + \nu_{1,i}, \;\; \sigma^2_{ik} = \exp(W_{it}\alpha_k + \nu_{2,i}) \tag{4.16}$$

Here β and α are the regression coefficients in the mean and within subject variance model. The two mood categories have the same β_0 and α_0, but different β_1, α_1, β_2 and α_2, indicating that individuals in different subgroups are allowed to have the same intercept but different smoking and alone effects on mood regulation. $\nu_{1,i}$ and $\nu_{2,i}$ are the subject random location and scale effects, indicating the influence of subject i on his or her mood mean and consistency.

For the reduced models, mixed HMM only has random subject lcation effect $\nu_{1,i}$ in the conditional model, while HMM has neither location nor scale random effects. For comparison purposes, results are summarized in Table 4.2. The point estimates are obtained as the posterior mean for regression coefficients (β and α) and probabilities (π_1, tp_{11} and tp_{22}) since their posteriors are approximately symmetric, and as the mode for random effect variances $\sigma^2_{\nu_1}$ or / and $\sigma^2_{\nu_2}$ since their posteriors are relatively skewed. The 95% credible intervals (CI) are obtained as the 2.5% and 97.5% posterior quantiles for all parameters. The model selection criteria $elpd_{LOO}$, proposed by Vehtari et al. [Vehtari et al., 2017], estimates the pointwise leave one out (LOO) prediction accuracy from a fitted Bayesian model by evaluating the log likelihood over the posterior samples. It is prefered over the deviance information criterion (DIC) since it accounts for the entire posterior distribution, works for singular models and is invariant to parametrization. Higher $elpd_{LOO}$ indicates better model fit adjusting for the model complexity.

The three models provide different estimates and credible intervals for the probability parameters (π_1, tp_{11} and tp_{22}), indicating that they have identified subgroups with different individuals and these individuals have distinct trends evolving over time. The proposed mixed location scale HMM has the lowest model selection criteria $elpd_{LOO}$, indicating that altough it has the largest number of parameters in the model, it still fits the data significantly better compared to the other models. Therefore, we will focus on the results and conclusions from the mixed location scale HMM in the example EMA study context.

The proposed model has identified two subgroups with different smoking and alone effects on mood regulation. Specifically 1) for group one, being a smoker (versus non-smoker) and being alone (vesus with others) both make a subject more likely to have unpleasant and more erratic mood as shown by the negative estimates (and credible intervals) of β_1^1 and β_2^1 and positive estimates of α_1^1 and α_2^1, and 2) for group two, being a smoker and being alone makes a subject more likely to have consistent mood but does not change their mood mean significantly since neither the credible intervals of β_1^2 nor β_2^2 exclude 0, while α_1^2 and α_2^2 are both estimated to be negative and their credible intervals exclude 0. The model also estimates significant subject random effects in location and scale, which are negatively correlated as is often the case due to floor or ceiling effects.

Both mixed HMM and mixed locations scale HMM can successfully identify the two distinct subgroups with β^1 and α^1 separated well from β^2 and α^2. However, mixed HMM generally have more separated estimates compared to mixed locations scale HMM, especially in terms of the variance coefficients α, leading to over classification and possibly false positive findings. When subject heterogeneity does exist in the within subject variability, models that ignore it when performing the latent state classification would attribute more of the total differences to subgroup differences while it can be explained partly by the individual differences represented by random subject scale effects. This is also true if we compare mixed HMM and HMM in terms of the separation between β_1 and β_2. Therefore, including both location and scale random effects in the conditional model increases the latent state classification efficiency and accuracy in the adolescent mood study example.

Table 4.2: Comparison of parameter estimates and credible intervals between HMM, mixed HMM, and mixed location scale HMM.

Parameter	HMM		mixed HMM		mixed LS HMM	
	Estimate	CI	Estimate	CI	Estimate	CI
π_1	0.465	(0.383, 0.549)	0.472	(0.281, 0.644)	0.688	(0.474, 0.878)
tp_{11}	0.974	(0.948, 0.994)	0.823	(0.718, 0.897)	0.837	(0.740, 0.905)
tp_{22}	0.986	(0.969, 0.998)	0.908	(0.862, 0.947)	0.858	(0.762, 0.930)
β_0	7.167	(7.070, 7.264)	7.143	(6.968, 7.308)	7.135	(6.946, 7.313)
β_1^1	-0.980	(-1.177, -0.787)	-1.028	(-1.653, -0.591)	-0.619	(-1.002, -0.279)
β_1^2	0.246	(0.102, 0.394)	0.142	(-0.132, 0.417)	0.102	(-0.240, 0.416)
β_2^1	-1.659	(-1.936, -1.391)	-0.900	(-1.362, -0.485)	-0.940	(-1.227, -0.669)
β_2^2	0.416	(0.204, 0.642)	-0.129	(-0.384, 0.121)	0.217	(-0.055, 0.477)
α_0	0.661	(0.557, 0.763)	-0.124	(-0.238, -0.008)	-0.390	(-0.569, -0.211)
α_1^1	0.075	(-0.106, 0.255)	0.585	(0.270, 0.847)	0.467	(0.157, 0.799)
α_1^2	-0.354	(-0.538, -0.180)	-0.683	(-0.987, -0.367)	-0.468	(-0.924, -0.026)
α_2^1	-0.291	(-0.571, -0.002)	0.268	(-0.133, 0.752)	-0.109	(-0.520, 0.296)
α_2^2	-0.521	(-0.787, -0.257)	-0.821	(-1.303, -0.356)	-1.025	(-1.990, -0.238)
$\sigma^2_{\nu_1}$	—	—	1.095	(1.000, 1.190)	1.159	(1.066, 1.292)
$\sigma^2_{\nu_2}$	—	—	—	—	0.766	(0.594, 0.824)
ρ_{ν_1,ν_2}	—	—	—	—	-0.350	(-0.474, -0.300)
$elpd_{LOO}$	-4095	—	-3471	—	-3359	—

Finally, we took a closer look at the two subgroups of observations identified by the proposed mixed effects location scale HMM. Subgroup one, for whom smoking and being alone both have detrimental effects on positive mood regulation, has lower positive affect, higher negative affect (not in the model) and lower negative mood regulation (not in the model) compared to subgroup two, on all 7 study days, as shown in 4.3, 4.4 and 4.5. Further investigations are being done to explore the possible associations between subgroup affiliation and future risk behaviors.

4.6 Discussion

In this paper, we have developed a mixed location scale hidden markov model and shown that the addition of both subject random effects in the mean and within subject variance of the intensively measured longitudinal outcome to the conditional model for the observed process yields a model that can be well interpreted and estimated, yet provides a better fit for the data in the context of EMA studies. The application of this model to the EMA adolescent mood study example demonstrated the advantages of such an approach over mixed HMM and

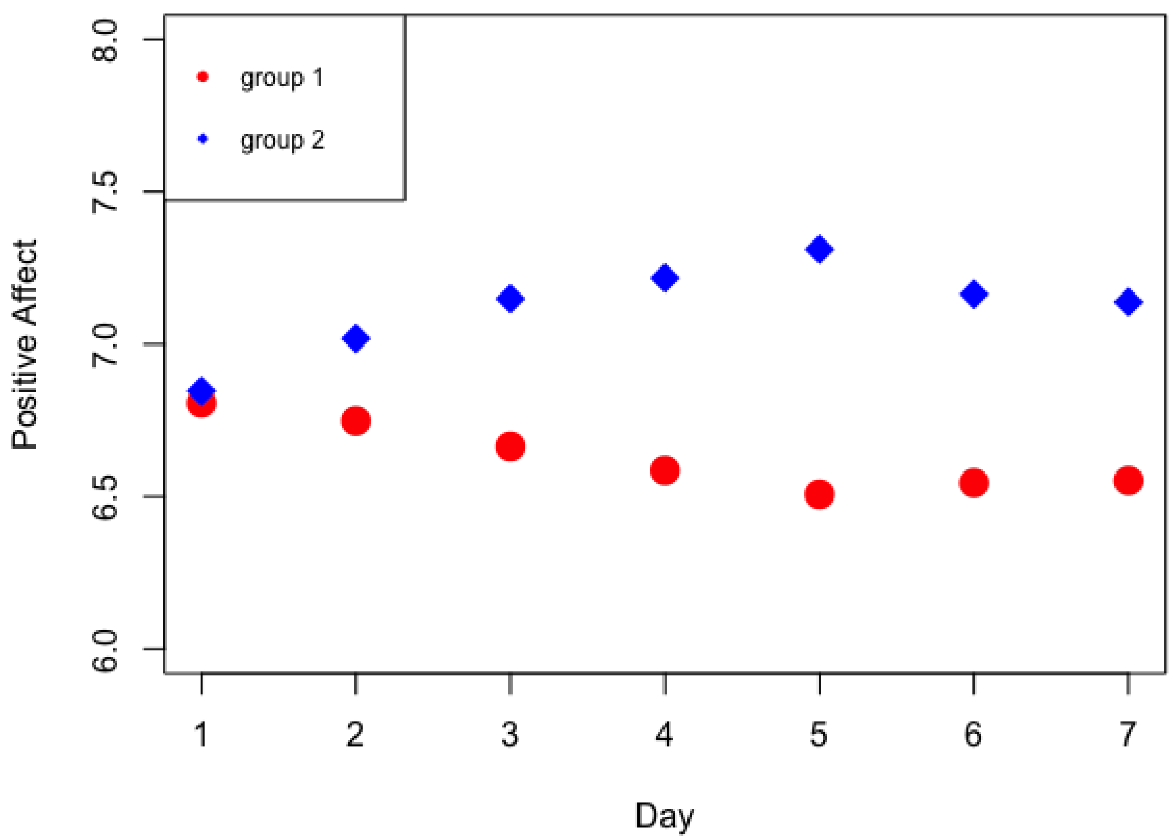

Figure 4.3: Separation of Positive Affect by two subgroups on all 7 study days.

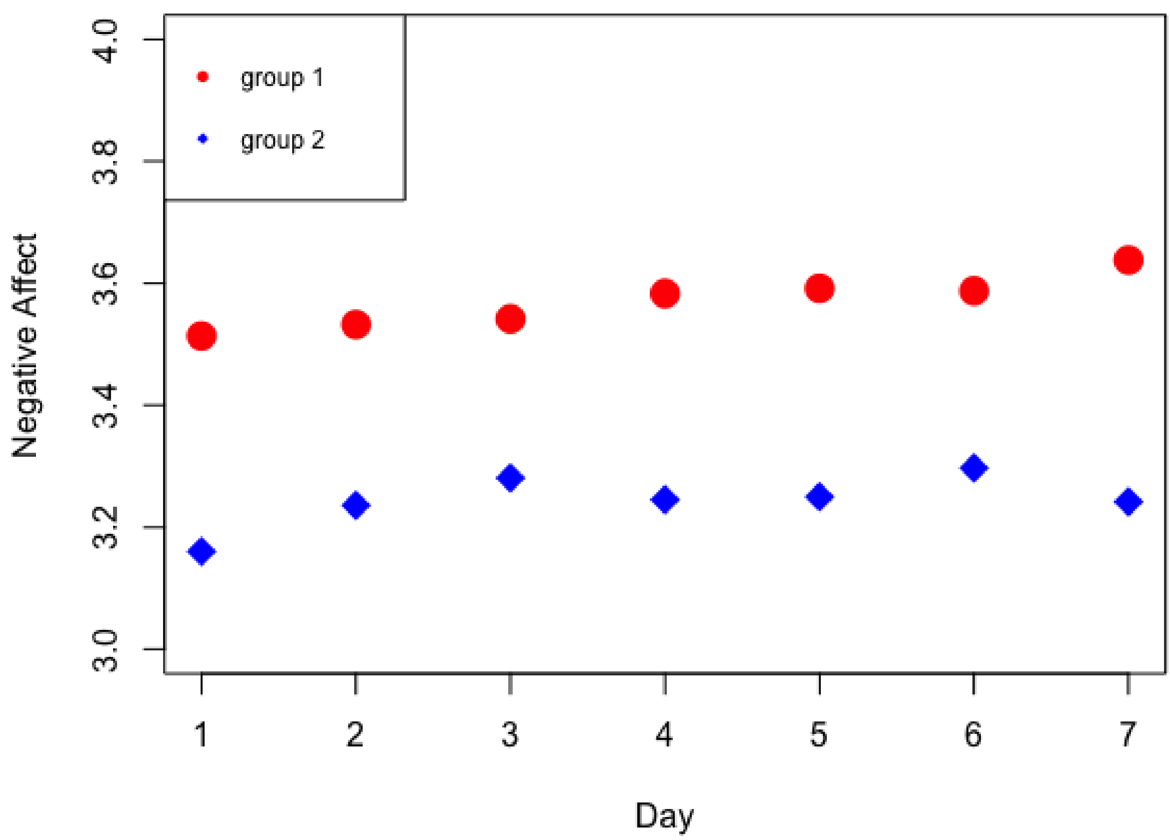

Figure 4.4: Separation of Negative Affect by two subgroups on all 7 study days.

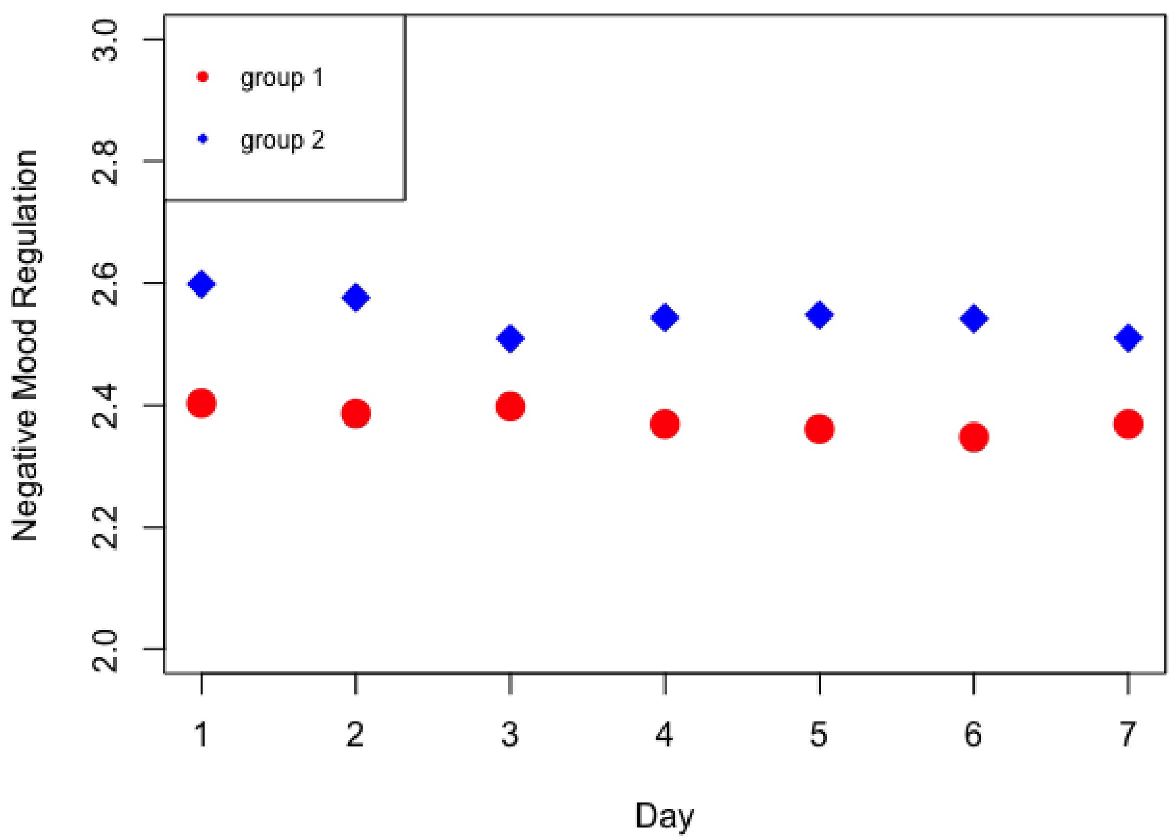

Figure 4.5: Separation of Negative Mood Regulation by two subgroups on all 7 study days.

HMM in extracting subject specific information as well as providing accurate and efficient subgroup identification regarding the effects of smoking and being alone on mood regulation.

In EMA studies where outcomes get intensively measured over a relatively short period of time, research interests are often surrounded around both the effects of certain baheviors or interventions on subjects mood level and how consistent / erratic their mood changes over time [Stone and Shiffman, 1994]. Therefore, the mixed location scale approach provides a rigorous way to address these questions by introducing subject random location and scale effects which represent subjects' specific influence on their own mean and within subject variability, up or below the population average. What's more, the distinct subgroups identified by the proposed model and the transition probability over time estimated by the markov chain enables one to investigate factors that play a role in separating mood regulation patterns. Combining mixed location scale effects with hidden markov model presents a novel yet efficient approach for latent state classification for EMA data.

For simplicity, we have assumed a constant transition probability matrix for all subjects. However, it is also possible to include random effects in the model for the hidden process if subjects show unexplained heterogeneity in terms of their transitions [Jackson et al., 2015]. But such models would be more difficult to estimate as one generally has less information about the hidden process compared to the observed process. Therefore, whether random effects need to be included in the hidden process depend on the research interests and abundance of the data. In order to carry out the proposed mixed location scale HMM (or even the mixed HMM), we have to assume the time points are equally spaced. However, this might not be true in practice as some EMA studies have measurements that are randomly spaced. In this case, a better alternative is to switch to a continuous-time formulation of the Markov model where the temporal structure of the state-switching process is captured in a transition intensity matrix instead of a probability matrix [Liu et al., 2015]. The original data from the example EMA study are hierarchical - measurements were taken at random time intervals during a day and days are then nested within a week - thus future work could focus on establishing separate transition matrices within a day as well as within the week, and link these two transition matrices together to better describe the latent process [Shirley et al., 2010].

CHAPTER 5

CONCLUSIONS AND FUTURE PLANS

5.1 Summary and Conclusions

In this dissertation, I focused on developing comprehensive statistical methods for ecological momentary assessment data where subjects are intensively measured over a relatively short period of time. First, a three level Bayesian mixed effect location scale model was developed for a multi-wave EMA study that allows for both intermediate wave clustering and individual heterogeneity in the mean and variability of the outcome. Second, a shared parameter model framework was designed for informative missingness introduced by non-responses that links the primary outcome with the missing data mechanism by subject specific traits. Finally, a mixed effect location scale hidden Markov model was explored for latent state classification and heterogeneous covariate effect that incorporates both time serial dependence and individual differences in the mean and variance model.

The methods developed in these dissertation projects were motivated by the EMA study that aims to investigate the association between psychosocial factors on mood regulation among adolescents. Compared to traditional longitudinal data, EMA studies produce intensively measured longitudinal outcomes with unique characteristics and pose special challenges in terms of statistical analysis. First, EMA studies are often conducted at multiple waves and within each wave, subjects are intensively measured at multiple occasions, resulting in multiple layers of correlation structure. Besides, individual differences are likely to be present both at baseline and change over time, in terms of both the mean and within subject within wave variability. Second, non-responses usually occur intermittently due to intensive sampling and non-responses. Subjects with worse/erratic mood are less likely to respond compared to those with better/stable mood, leading to informative missing in the observed sample. Third, research demonstrated that adolescents differ considerably in their ability to effectively regulate their mood, and even for the same individual, mood regulation can change quite a lot over time. Therefore, heterogeneous covariate effects for distinct mood subgroups together with heterogeneous subject specific effects need to be accounted for when analyzing EMA data. The methods developed in this dissertation are well suited for the above challenges and can also be generalized to other areas of research with similar research interests and data structures.

One difficulty in implementing the above methods is the model estimation. Due to the relative large numbers of random effects (and random effects in the variance model), maximum likelihood based methods that involve numerical integrals over the distribution of random effects would be computationally infixible. Alternatively, I derived a Bayesian model estimation framework and relied on Markov Chain Monte Carlo sampling methods for parameter estimation. A better approach with Stan using Hamiltonian Monte Carlo was also illustrated. Model comparisons between the proposed model and existing (reduced) models were conducted based on the model selection criteria $elpd_{LOO}$ developed by Vehtari et al. [Vehtari et al., 2017] that evaluates the leave-one-out likelihood evaluated over the entire posterior samples. In all three projects, the proposed models were shown to fit the data better than existing ones adjusting for the model comlexity.

From psychological and behavioral science perspectives, the initial aims of the projects were to better understand the associations between psychosocial factors - smoking, being alone, gender, negative mood regulation ability (NMR), etc - and mood regulation among adolescents. Application of the proposed methods to the EMA adolescent mood study in the first two dissertation projects showed consistent conclusions regarding the effects of the smoking, being alone, gender and NMR on mood regulation. In particular, smoking, being male, higher NMR score are associated with better and more stable mood, while being alone, being female and lower NMR score are associated with worse and more erratic mood, adjusting for the possible informative missing. Results from the third dissertation projects showed evidence of heterogeneous smoking and alone effects on mood regulation - smoking and being alone play dramatically different roles for subjects in distinct subgroups characterized by mood regulation ability.

5.2 Future Plans

An interesting and powerful use of the framework underlying my methodological work in this dissertation is the inclusion of the within-subject variance for performing statistical analysis to better utilize the abundance of EMA data. In the future, I am interested in extending the current mixed effect location scale models to multivariate outcomes as well as extending the framework together with Bayesian MCMC techniques to develop more comprehensive methodologies for clustering/classification at different hierarchies of EMA outcomes. Here I discuss several possible research projects or directions.

5.2.1 Bivariate Shared Parameter Location Scale Mixed Effect Model

In Chapter 3, I developed a shared parameter model framework that incorporates a mixed effect logistic regression model for the binary missing process, a mixed effect location scale model for the intensively measured univariate longitudinal outcome and a parameter sharing model. When the primary outcome consists of multivariate measurements, such as positive affects (PA) and negative affects (NA), developing a multivariate version of the share parameter framework by incorporating correlations between multivariate outcomes will provide more accurate and efficient statistical inference.

In the context of the EMA adolescent mood study example, let $Y_{ij} = \big(PA_{ij}, NA_{ij}\big)$ be the bivariate outcome vector for individual i at occasion j, where $i = 1, ..., n$, and $j = 1, ..., n_i$.

We specify a bivariate mixed effect location scale model for Y_{ij} on top of the univariate model described in Chapter 3,

$$Y_{ij} = X_{ij}^{\top}\beta + \nu_{loc,i} + \epsilon_{ij} \tag{5.1}$$

$$\log(\sigma^2_{\epsilon_{ij}}) = X_{ij}^{\top}\alpha + \nu_{scale,i} + \nu_{scale,i} \tag{5.2}$$

where $X_{ij} = \big(smoke_i, gender_i, NMR_i, GPA_i, aloneWS_{ij}, aloneBS_i\big)$ is the fixed effect covariate vectors in the mean and within subject variance model. β and α are the corresponding fixed effect coefficient matrices (α in log scale), which indicate the population average effect of the covariates on the mean and (log of) the within subject variability of the outcome, for PA and NA respectively. $\nu_{loc,i}$ is the bivariate random subject location effect vector, indicating the effect of subject i on his/her mean of the PA/NA measurements; $\nu_{scale,i}$ can be either bivariate vector or scaler, depending on model fit. For example, if data suggest PA and NA have similar individual differences in within subject variability, then we can include $\nu_{scale,i}$ as a scaler whereas assume $\nu_{scale,i}$ is bivariate vector if individual heterogeneity show different patterns for PA and NA. Conditional on $\nu_{loc,i}$ and $\nu_{scale,i}$, the bivariate outcome PA_{ij} and NA_{ij} are independent.

The bivariate version of shared parameter model has the exact same model specification for the missing process. However, in the parameter sharing section, we need to establish new link between $\nu_{loc,i}$, which is now a vector, and λ_i, the random effect from the missing model.

$$\nu_{loc,i} = \gamma \cdot \lambda_i + \eta_{loc,i} \tag{5.3}$$

$$\nu_{scale,i} = \delta \cdot \lambda_i + \eta_{scale,i} \tag{5.4}$$

where γ and $\eta_{loc,i}$ are now a bivariate vector, indicating the corresponding component for PA and NA. The new version is designed specifically for multivariate EMA outcomes subject to informative missing and can achieve more estimation as well as interpretation efficiency, such as tighter credible intervals and higher power.

5.2.2 A Shared Parameter Location-Scale Item Response Theory (IRT) Model for Repeated Ordinal Questionnaire Data

In Chapter 3, the shared parameter model was developed for a normally distributed and univeriate primary outcome. But often in practice, researchers need to deal with repeated questionaire data that have a natural ordering associated with them. For example, interviewees were asked about 10 items from the Nicotine Dependence Syndrome Scale (NDSS) on 4-point scale: 1=not at all true, 2=not very true, 3=fairly true, 4=very true. In this case, ordinal location scale models is a better modeling approach compared to those that treat the outcomes as nominal variables. One possible future direction is to extend the shared parameter framework for this special data type by including a location scale item response theory model for the primary outcome and establish a link between the outcome and missing outcome. In particular, let Y_{ij} denote the ordinal response of subject i on item j; $P_{ijc} = Pr(y_{ij} \leqslant c)$ the cumulative probabilities for C categories of Y; X_j the vector of item indicators; β the vector of item difficulty parameters; σ the vector of item discrimination parameters; θ_i is the standardized location random effects, $\theta_i \sim \mathcal{N}(0,1)$; ω_i is the scale random effects, $\omega_i \sim \mathcal{N}(0, \sigma_\omega^2)$; $\alpha_1 < \alpha_2 ... \alpha_{C-1}$ the strictly increasing thresholds or intercepts (for identification $\alpha_1 = 0$). Then the location scale item response theory model can be expressed as

$$z_{ijc} = \log(\frac{p_{ijc}}{1-p_{ijc}}) = \frac{\alpha_c - (x_j \top \beta + x_j \top \delta \theta_i)}{\sigma_{\epsilon_j}} \tag{5.5}$$

$$\sigma_{\epsilon_j}^2 = \exp(x_j \top \gamma + \omega_i) \tag{5.6}$$

5.2.3 Extension of The Mixed Location Scale Hidden Markov Model

In Chapter 4, I explored a mixed location scale Hidden Markov Model that incorporates random location and scale effects into the conditional model of the Hidden Markov Model. However, this model has its own limitation since it requires the time points to be equally spaced, which might not be suited for a lot of EMA studies. Therefore, a continuous-time formulation of the Markov model where the temporal structure of the state-switching process is captured in a transition intensity matrix instead of a probability matrix provides to be a better and more general approach.

Another limitation of the mixed location scale Hidden Markov Model is that measurements within a single day need to be aggregated before applying the model, which lead to information loss and inefficient estimation. Since mood assessments were taken at random time points within a day, and at seven consecutive days in a week, a better approach can be taken by specifying separate transition matrices within a day as well as within the week, and link these two transition matrices together to better describe the latent process.

5.2.4 A Latent State Classification Algorithm with Subject Specific Mixture Proportion

This is a special case of the finite Gaussian mixture model with subject specific mixture probabilities. Let Y_{ij} denote the repeated measurement for subject i at time j, X_{ij} denote the covariate vector. Define Z_{ij} as the latent group membership, where $z_{ij} = 1$ if $\left(x_{ij}, y_{ij}\right)$ belongs to component 1 and $z_{ij} = 0$ if belong to component 2. Assume both components are normally distributed with distinct mean and variance, then the likelihood can be written as

$$y_{ij} = p_i \mathcal{N}(x_{ij} \top \beta_1, \sigma_1^2) + (1 - p_i) \mathcal{N}(x_{ij} \top \beta_2, \sigma_2^2) \tag{5.7}$$

Using the latent variable z_{ij}, the likelihood can be re-written as

$$y_{ij} = \mathcal{N}(x_{ij}\top\beta_1, \sigma_1^2)^z_{ij} + \mathcal{N}(x_{ij}\top\beta_2, \sigma_2^2)^{1-z_{ij}} \tag{5.8}$$

Here, p_i is the mixture proportion of component 1 for subject i and is the novelty of the model compared to the traditional finite Gaussian mixture model. There can be a fixed effect or a random effect version of the new model. In particular, for a fixed effect version, p_i is a fixed parameter and is estimated explicitly for each individual; for a random effect version, $p_i = p + \theta_i$ where p is the population average mixture probability and is a parameter to be estimated, θ_i is the random subject effect and indicates the effect of subject i on his/her mixture probability. A (modified) E-M algorithm can be used to estimates these two models.

REFERENCES

[Altman, 2007] Altman, R. (2007). Mixed hidden markov models: An extension of the hidden markov model to the longitudinal data setting. *Journal of the American Statistical Association*, 102(477):201–210.

[Bates et al., 2015] Bates, D., Machler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

[Bishop, 2006] Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag, New York.

[Bradley and Siddhartha, 1995] Bradley, P. C. and Siddhartha, C. (1995). Bayesian model choice via markov chain monte carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(3):473–484.

[Cappe, 2012] Cappe, O. (2012). Online em algorithm for hidden markov models. *Journal of Computational and Graphical Statistics*, 20(3):728–749.

[Carpenter et al., 2017] Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32.

[Casella and George, 1992] Casella, G. and George, E. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174.

[Chib and Greenberg, 1995] Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hasting algorithm. *American Statistical Association*, 49(4):327–335.

[Crayen et al., 2012] Crayen, C., EID, M., Lischetzke, T., Courvoisier, D., and Vermunt, J. (2012). Exploring dynamics in mood rgulation - mixture latent markov modeling of ambulatory assessment data. *Psychosomatic Medicine*, 74:366–376.

[Cursio et al.,] Cursio, J., Mermelstein, R., and Hedeker, D. Latent trait shared-parameter mixed-models for missing ecological momentary assessment data. Manuscript submitted for publication.

[Dewey et al., 2015] Dewey, D., McDonald, M., Brown, W., Boyd, S., Bunnell, B., and Schuldberg, D. (2015). The impact of ecological momentary assessment on posttraumatic stress symptom trajectory. *Psychiatry Research*, 230(2):300–303.

[Diggle and Kenward, 1994] Diggle, P. and Kenward, M. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society. Series C*, 43:49–93.

[Duncan, 2012] Duncan, A. (2012). Hyperglycemia and perioperative glucose management. *Current Pharmaceutical Design*, 18(38):6195–6203.

[Ebner-Priemer and Trull, 2009] Ebner-Priemer, U. and Trull, T. (2009). Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological Assessment*, 21(4):463–475.

[Follmann and Wu, 1995] Follmann, D. and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics*, 51:151–168.

[Fong et al., 2010] Fong, Y., Rue, H., and Wakefield, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics*, 11(3):397–412.

[Gelman et al., 2013] Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis*. CRC Press, Florida.

[Goldstein et al., 2017] Goldstein, H., Leckie, G., Charlton, C., Tilling, K., and Browne, W. (2017). Multilevel growth curve models that incorporate a random coefficient model for the level 1 variance function. *Statistical Methods in Medical Research*, Epub:962280217706728.

[Gwaltney et al., 2008] Gwaltney, C., Bartolomei, R., Colby, S., and Kahler, C. (2008). Ecological momentary assessment of adolescent smoking cessation. *Nicotine & Tobacco Research*, 10(7):1185–1190.

[Heckman, 1979] Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47:153–161.

[Hedeker et al., 2009] Hedeker, D., Demirtas, H., and Mermelstein, R. (2009). A mixed ordinal location scale model for analysis of ecological momentary assessment (ema) data. *Statistics and Its Interface*, 2:391–402.

[Hedeker and Gibbons, 2006] Hedeker, D. and Gibbons, R. (2006). *Longitudinal Data Analysis*. Wiley, New Jersey.

[Hedeker et al., 2008] Hedeker, D., Mermelstein, R., and Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (ema) data. *Biometrics*, 64:627–634.

[Hedeker et al., 2012] Hedeker, D., Mermelstein, R., and Demirtas, H. (2012). Modeling between- and within-subject variance in ecological momentary assessment (ema) data using mixed-effects location scale models. *Statistics in Medicine*, 31:3328–3336.

[Hobert and Casella, 1996] Hobert, J. and Casella, G. (1996). The effect of improper priors on gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91(436):1461–1473.

[Jackson et al., 2015] Jackson, J., Albert, P., and Zhang, Z. (2015). A two-state mixed hidden markov model for risky teenage driving behavior. *The Annals of Applied Statistics*, 9(2):849–865.

[Kapur et al., 2015] Kapur, K., Li, X., Blood, E., and Hedeker, D. (2015). Bayesian mixed-effects location and scale models for multivariate longitudinal outcomes: an application to ecological momentary assessment data. *Statistics in Medicine*, 34(4):630–651.

[Laird, 1988] Laird, N. (1988). Missing data in longitudinal studies. *Statistics in Medicine*, 7:305–315.

[Leckie, 2014] Leckie, G. (2014). Modeling heterogeneous variance-covariance components in two-level models. *Journal of Educational and Behavioral Statistics*, 39(5):307–332.

[Leonard, 1975] Leonard, T. (1975). A bayesian approach to the linear model with unequal variances. *Technometrics*, 17:95–102.

[Li and Hedeker, 2012] Li, X. and Hedeker, D. (2012). A three-level mixed-effects location scale model with an application to ecological momentary assessment (ema) data. *Statistics in Medicine*, 31:3192–3210.

[Lin et al., 2018] Lin, X., Mermelstein, R., and Hedeker, D. (2018). A shared parameter location scale mixed effect model for ema data subject to informative missing. *Health Services and Outcomes Research Methodology*, pages 1–17.

[Little, 1993] Little, J. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88:125–134.

[Liu et al., 2015] Liu, Y., Li, S., Li, F., Song, L., and Rehg, J. (2015). Efficient learning of continuous-time hidden markov models for disease progression. *Advances in Neural Information Processing Systems*, 28:3599–3607.

[Lukasiewicz et al., 2007] Lukasiewicz, M., Fareng, M., Benyamina, A., Blecha, L., Reynaud, M., and Falissard, B. (2007). Ecological momentary assessment in addiction. *Expert Review of Neurotherapeutics*, 7(8):939–950.

[MacDonald and Zucchini, 1997] MacDonald, I. and Zucchini, W. (1997). *Hidden Markov Models and Other Models for Discrete-Value Time Series*. Chapman & Hall, London.

[Martin and Hofer, 2004] Martin, M. and Hofer, S. (2004). Intraindividual variability, change, and aging: Conceptual and analytical issues. *Gerontology*, 50:7–11.

[McCulloch, 1997] McCulloch, C. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92:162–170.

[McCulloch and Neuhaus, 2005] McCulloch, C. and Neuhaus, J. (2005). *Generalized Linear Mixed Models*. Wiley, New Jersey.

[Nesselroade, 2004] Nesselroade, J. (2004). Intraindividual variability and short-term change. *Gerontology*, 50:44–47.

[Piasecki et al., 2014] Piasecki, T., Trela, C., Hedeker, D., and Mermelstein, R. (2014). Smoking antecedents: separating between- and within-person effects of tobacco dependence in a multiwave ecological momentary assessment investigation of adolescent smoking. *Nicotine & Tobacco Research*, S:119–126.

[Pulkstenis et al., 1998] Pulkstenis, E., TenHave, T., and Richard, J. (1998). Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *Journal of the American Statistical Association*, 93:438–450.

[Rast et al., 2012] Rast, P., Hofer, S., and Sparks, C. (2012). Modeling individual differences in within-person variation of negative and positive affect in a mixed effects location scale model using bugs/jags. *Multivariate Behavioral Research*, 47:177–200.

[Rocke et al., 2009] Rocke, C., Li, S., and Smith, J. (2009). Intraindividual variability in positive and negative affect over 45 days: do older adults fluctuate less than young adults? *Psychology and Aging*, 24:863–878.

[Schwartz and Stone, 1998] Schwartz, J. and Stone, A. (1998). Strategies for analyzing ecological momentary assessment data. *Health Psychology*, 17(1):6–16.

[Shiffman et al., 2008] Shiffman, S., Stone, A., and Hufford, M. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4:1–32.

[Shirley et al., 2010] Shirley, K., Small, D., Lynch, K., Maisto, S., and Oslin, D. (2010). Hidden markov models for alcoholism treatment trial data. *The Annals of Applied Statistics*, 4(1):366–395.

[Sokolovsky et al., 2014] Sokolovsky, A., Mermelstein, R., and Hedeker, D. (2014). Factors predicting compliance to ecological momentary assessment among adolescent smokers. *Nicotine & Tobacco Research*, 16:351–358.

[Stan-Development-Team, 2014] Stan-Development-Team (2014). Stan modeling language: User's guide and reference manual. Technical report, Bell Labs.

[Steven et al., 2014] Steven, M., Matthew, R., Paul, E., Elizabeth, K., and Daniel, F. (2014). Mood instability and psychosis: Analyses of british national survey data. *Schizophrenia Bulletin*, 40(2):269–277.

[Stone and Shiffman, 1994] Stone, A. and Shiffman, S. (1994). Ecological momentary assessment (ema) in behavioral medicine. *Annals of Behavioral Medicine*, 16(3):199–202.

[Vehtari et al., 2017] Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27(5):1413–1432.

[Vonesh et al., 2006] Vonesh, E., Greene, T., and Schluchter, M. (2006). Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine*, 15:143–163.

[W.S et al., 2002] W.S, C., L, D., and C, L. (2002). Random location and scale effects: model building methods for a general class of models. *Computing Science and Statistics*, 32:3–10.

[Wu and Carroll, 1988] Wu, M. and Carroll, R. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44:175–188.