

**TIME VARYING COEFFICIENT MODEL FOR
GAP TIMES IN ECOLOGICAL MOMENTARY
ASSESSMENT DATA**

by

Xiaoxue Li

BS, University of Science and Technology of China, China, 2007

Submitted to the Graduate Faculty of
Graduate School of Public Health in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH
GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Xiaoxue Li

It was defended on

October, 17, 2014

and approved by

Stewart J. Anderson, Ph.D., Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Saul Shiffman, Ph.D., Professor, Department of Psychology,
The Dietrich School of Arts and Sciences, University of Pittsburgh

Abdus Wahed, Ph.D., Associate Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Ada Youk, Ph.D., Assistant Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Dissertation Director: Stewart J. Anderson, Ph.D., Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Stewart J. Anderson

**TIME VARYING COEFFICIENT MODEL FOR GAP TIMES IN
ECOLOGICAL MOMENTARY ASSESSMENT DATA**

Li, Xiaoxue, PhD

University of Pittsburgh, 2014

Abstract

Ecological momentary assessment (EMA) studies investigate instantaneous and repeated observations on subjects over time in their everyday life. Such study designs are useful for applications to public health and social sciences because they provide intensive information about the temporal pattern of one's behavior. Throughout this dissertation, we will use an EMA study of intermittent smokers (ITS) to demonstrate our method. In this EMA study, events such as smoking are of primary interest. Here, we focus on a particular temporal pattern when smoking events are clustered in time. The distributions of the time-clusters or smoking "bouts" and covariates that predict such behavior are our interest. Traditional linear mixed effects models are not typically equipped to properly assess this kind of investigation. In this dissertation, we introduce a method of displaying the temporal behavior of subjects via functions of event gap times which allow us to easily identify bouts. We also apply an existing time-varying coefficient model to cumulative log gap times to characterize the time patterns of smoking while concomitantly adjusting for behavioral covariates that may be time varying and related to smoking. The mixed effect model we consider here produces a linear function with coefficients that change over time and hence, can identify meaningful temporal changes both at the subject and population levels. We also apply the inverse probability of weighting methods to weight the observed cases and handle missing data generated by the study design.

Our method has public health significance in that it allows one to identify time patterns (periodic or otherwise) in health event outcomes that can occur multiple times. Hence, one can characterize the time trajectory of these multiply observed events and possibly develop interventions when necessary to alter the time course of such processes.

Keywords: Ecological Momentary Assessment, Intensive Longitudinal Data, Recurrent Events Analysis, Time-varying Coefficient Model, Gap Times.

TABLE OF CONTENTS

PREFACE	ix
1.0 INTRODUCTION	1
1.1 Motivation	1
1.2 Statement of Research Problem	4
2.0 LITERATURE REVIEW	5
2.1 Intensive Longitudinal and Ecological Momentary Assessment (EMA) Studies	5
2.2 Some Special Features of EMA studies	7
2.2.1 Sampling Schemes	7
2.2.2 Situational Variables and Temporal Patterns	8
2.2.3 Missing by Design (MBD) and Analytical Difficulties	9
2.2.4 An Example – EMA Data Structure for Our Data	10
2.3 Traditional Longitudinal Analysis Methods	10
2.3.1 Linear Mixed Effects Models (LMMs)	11
2.3.2 Generalized linear mixed effects models (GLMMs)	12
2.3.3 Generalized Estimating Equation	13
2.3.4 Growth Curve Models	13
2.3.5 Transition Models	14
2.4 Analytical Methods for Intensive Longitudinal Studies	15
2.4.1 Nonparametric splines and polynomial models	16
2.4.2 Time Series Models	17
2.4.3 Analysis of Recurrent Events and Point Process Models	19
2.5 Missing Data	20

3.0 PROCEDURES AND PROPOSED METHODOLOGY	22
3.1 Identification of Bouts	22
3.2 Time Varying Coefficients Cumulative Gap Time Models	23
3.3 Time Varying Coefficients Models with Missing Data	30
4.0 APPLICATION	37
5.0 SIMULATION STUDIES	46
5.1 Picking Up the Functional Forms of the time varying coefficient	46
5.2 Simulation Study with Missing Data	53
6.0 DISCUSSION AND FUTURE WORK	54
APPENDIX: MATLAB CODE FOR MISSING BY DESIGN DATA	57
BIBLIOGRAPHY	67

LIST OF TABLES

2.1	EMA Data Structure	11
5.1	Simulation Results for Time Varying Coefficient Models	49
5.2	Simulation Results for Time Varying Coefficient Models with Missing Data	53

LIST OF FIGURES

2.1	EMA Sampling Scheme	8
2.2	Relationship between Assessed Events, Events and Random Prompts	10
4.1	Event Pattern of One Subject in the Presence of Bouts	41
4.2	Log Gap Times	42
4.3	Cumulative Log Gap Times	42
4.4	Time-varying Coefficients	43
4.5	Time-varying Coefficients	44
4.6	Fitted Curves for Individuals	45
5.1	Simulation Steps Summary	48
5.2	Fitted vs. True Coefficient (N=10) for One Simulated Dataset	50
5.3	Fitted vs. True Coefficient (N=20) for One Simulated Dataset	51
5.4	Fitted vs. True Coefficient (N=50) for One Simulated Dataset	51
5.5	Fitted vs. True Coefficient (N=100) for One Simulated Dataset	52
5.6	Fitted vs. True Coefficient (N=200) for One Simulated Dataset	52

PREFACE

This dissertation is part of my PhD education in the Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh. It is my pleasure to study and doing research here. Finishing this dissertation means a lot to me. It means that I can move to my next step in life. It teaches me to be responsible for my own research. It teaches me to think ahead and broke statistical researches into different steps.

During this process, I read books and papers, I learned a lot outside of the textbooks. At the beginning, I was completely lost when reading papers on different models that are potentially valuable. At the end, I understand the model I am using and the mechanism behind. The process of writing my dissertation changes my way of thinking open-ended problems and the way problems are linked to statistics. It is painful but I believe it is valuable for the rest of my life.

I always feel lucky enough to have worked for four years with Dr. Saul Shiffman and have learned Ecological Momentary Assessment (EMA) studies. It is amazing to see how the EMA data can capture information on peoples everyday real-world behavior. I never doubt that EMA will advance the science and practice of health area.

My dissertation topic is difficult to me. I am trying to explore a special smoking behavior, where consumed cigarettes tended to be clustered in time, among intermittent smokers in the EMA data. In addition, I want to identify covariates that are associated with this behavior. I hypothesize that a weekly pattern exists, i.e. the time-clusters tend to occur on weekends. Thus, we proposed a time varying coefficient model on cumulative gap times to handle this problem. The solution is not limited to smoking data. It can be applied to any recurrent events where events tend to be clustered in time.

I want to thank my adviser, Dr. Stewart Anderson for all of his help and support during my entire PhD education. I want to thank Dr. Saul Shiffman and Dr. Roslyn Stone for their funding support. I want to thank Dr. Wahed for his help in extending the model to handle missing data. I want to thank Dr. Ada York for her help in editing my dissertation document. I want to thank all of my committee members for their patience, encouragement and help during my dissertation work. Finally I would like to thank my family and friends for being helpful and supportive during my time studying in Pittsburgh.

1.0 INTRODUCTION

1.1 MOTIVATION

Many processes of interest involve events occurring over time. These might be occurrences of crimes, heart beats during exercise, certain circadian animal behaviors or the occurrence of storms, which can occur at different times, either regularly or irregularly. Within medical and behavioral science, certain behavioral events are of interest. For example, cigarette smoking occurs in discrete episodes distributed over time. While studies of cigarette smoking have often focused on total consumption, usually expressed as cigarettes per day, less attention has been paid to how those smoking events are distributed over time within a day and across several weeks. For heavy smokers who might consume as many as two cigarettes per hour every day, there is room for only modest variation in how cigarettes are distributed or grouped over the waking hours. Moreover, the dominant theory of smoking behavior posits that smokers strive to minimize variation in nicotine levels, which would call for smoking at regular intervals [3]. At the other extreme, very light smokers, including those who do not smoke every day, do not appear to smoke in order to regulate nicotine levels [29], and have a large range of potential variation in how smoking occasions are distributed over time [28].

Traditionally, tobacco research has been focused on daily smokers (DS). However, in the last twenty years, a substantial increase in the proportion of non-daily, or intermittent smokers (ITS) was found by Centers for Disease Control and Prevention (CDC) and other national surveys. According to CDC records, more than a quarter of US adult smokers are

ITS (CDC, 2008 [6, 7]); in addition, the prevalence increased by 40% from 1996 to 2001 (CDC, 2003 [5]).

ITS and DS populations have different characteristics. ITS tend to be younger, better-educated, with higher income, and are more heterogeneous than DS in their smoking behavior (ref). Some ITS tend to smoke when they are happy, while some others tend to smoke when they are stressed; some tend to smoke at home, while some others tend to smoke at bars; some are more likely to smoke when being alone, while some others are more likely to smoke when socializing. These characteristics of ITS make us think that there are different causalities of smoking between ITS and DS. Shiffman, et al. [28], conducted an *Ecological Momentary Assessment (EMA)* study to assess smoking patterns among ITS and compare them with that of DS. The EMA study, a subclass of the *intensive longitudinal* study, is a longitudinal study where observations are sampled instantaneously and intensively over time for every participant. The measurements can be taken at the time an event occurs in real-world settings. This will lead to more reliable conclusions as compared to traditional studies where participants are asked to recall their past behavior. Typically, the numbers of measurements for each subject can vary from < 10 to $100+$ over the period of observation. Hence, the term “*intensive*” is used to describe the nature of the sampling of the data.

In the Shiffman EMA smoking study we investigated, both DS and ITS individuals were repeatedly measured on their instantaneous location, mood, activities, food and drink over time in real-world settings. They either experienced a “smoking event” or are randomly prompted (RP) at a time when no “event” occurred. Moreover, smoking events were assessed according to pre-calculated probabilities based on the number of smoking events recorded the day before. Some ITS individuals had a special smoking behavior, that is, their reported cigarettes tended to be clustered in time. For this kind of phenomenon, we say that these individuals tended to smoke in time-clusters or “bouts”. Bout smoking behavior is a temporal or micro-pattern of individual’s general smoking behavior. A problem arises naturally as how to characterize this phenomenon. Intuitively, one could simply define bouts as events that

happened close in time. For example, if a bout is defined as a set of events where two or more consecutive events occurred within an hour, then, approximately 28% of the cigarettes, on average, were found to be in bouts periods among ITS and about 60% for DS. The problem with the above definition of a bout is that those defined among DS might be invalid because DS individuals tend to smoke at a consistently high rate and result in consistently short inter-event-times, or “gap time” intervals. Hence, it is of interest to find a more flexible way to identify bouts based on the baseline smoking rates per individual. Adjustment for situational variables, or time-varying covariates should be included because when cigarettes are in bouts, they may be smoked in a similar environment. Conversely, the fact that several events are close in time and may share a similar behavioral environment may lead us to believe that such events occur in one time-cluster. Adjusting for variables that are related to smoking might increase the accuracy of the classification of cigarettes in bouts. For this dissertation, we attempt to identify different patterns for ITS subjects who had bouts. We also examined what factors promoted individuals to smoke in bouts.

We assumed that the distribution of smoking events followed a point Poisson process where the inter-event-time interval was a function of a set of covariates [8]. The model is written as

$$g(T_{ij}) = \mu + \boldsymbol{\beta}'\mathbf{X}_{ij} + U_i + e_{ij}, \quad (1.1)$$

where T_{ij} is the inter-event time and $g(\cdot)$ is a transformation that makes $g(T_{ij})$ normally distributed [25]. The natural logarithmic transformation is frequently used in practice. The term μ is the intercept term, \mathbf{X}_{ij} represents the vector of covariates and $\boldsymbol{\beta}'$ is the coefficient vector associated with covariates, U_i is the frailty term for the random effect, and e_{ij} is the error term. The model 1.1 is referred to as an accelerated gap time model for inter-event time.

1.2 STATEMENT OF RESEARCH PROBLEM

The general research interests to be addressed in this dissertation are:

1. To identify temporal patterns in multiple events data for a cohort of individuals. For our specific data, the particular interest is in identifying bouts or time clusters where smoking intensity is greatly increased.
2. To characterize situational factors that are associated with an event being in a bout period.
3. To handle the cases where observations are missing by a sampling algorithm.

2.0 LITERATURE REVIEW

2.1 INTENSIVE LONGITUDINAL AND ECOLOGICAL MOMENTARY ASSESSMENT (EMA) STUDIES

Longitudinal studies enable investigators to explore processes that can change over time. They are typically studies in which observations are recorded repeatedly over time. Longitudinal studies usually involve two measurement units: one is the subject, and the other is the within subject observation. Longitudinal studies usually have multiple subjects and multiple observations for each subject. The subject unit is also referred to as a “cluster” because observations from the same subject are usually correlated and are more similar to each other as compared to observations measured on other subjects. Hence, the data has a hierarchical structure of two layers: cluster and observation.

With the development of computer techniques, data collection is far easier than before; hence, researchers have the potential to design more complicated studies tailored to their specific research interest. The intensive longitudinal study is one class of longitudinal studies that usually has many more observations per subject and more variation as compared to the usual longitudinal study [31]. Diary data, self-monitoring data, and experience sampling data are types of intensive longitudinal data that have been used historically. Diaries are usually used to capture daily experience on a wide range of behavior; self-monitoring methods are used to target events of interest; experience sampling methods collect data periodically on participants’ behavior.

Ecological Momentary Assessment (EMA) is a recently developed type of longitudinal study that involves large numbers of observations per subject. Recent EMA studies use electronic diaries to intensively measure participants in real time and in a real world setting. Hence, EMA is defined as “the repeated collection of real-time data on participants’ momentary states in a natural environment” [31]. It can be viewed as a sampling collection methodology, which could either be self-reported (i.e. smoking data) or automatically collected by electronic devices.

EMA studies were first introduced by Stone and Shiffman in 1986 [31, 30]. These studies have now become more and more widely used in many disciplines, such as public health, social science, behavior science, and clinical psychology. EMA studies are known to reduce recall bias. Recall bias is a well-known problem for research studies. People may tend to report skewed experience rather than actual experience, or may not be able to accurately remember their past behaviors; how people behave in a research lab might be different from that in the real world [30]. Thus, being able to collect data in real-time and real-world settings, or neutral environments, at the time when events of interest happen is very attractive. EMA studies have such features and they substantially reduce recall bias. In addition, in an EMA study, vast amounts of temporal data are collected, which may give clues on the nature of a behavior. Consequently, EMA has the potential to of being more widely used in the future.

Instantaneous information is also collected in EMA studies and can be used to study micro-processes within a main data stream. With EMA, one is able to assess the “momentary cross sectional associations” rather than the overall cross sectional association where instantaneous associations are smoothed over time [30]. The focus of this dissertation is to model the temporal behavior. Our particular interest is motivated by a real-world problem when analyzing the Shiffman EMA smoking data. The problem is not limited to this specific situation, but rather, could be generalized in broader environment whenever data consistently have “ties” or “peaks”.

2.2 SOME SPECIAL FEATURES OF EMA STUDIES

2.2.1 Sampling Schemes

Despite the attractive features of EMA, the design of EMA assessment may be complicated. EMA studies often involve two components: one is the nature of the data acquisition and the other is the sampling scheme involved. EMA studies often use electronic diary devices to sample the participants and collect data. Two sampling approaches are involved: one is the event-based sampling and the other is a time-based sampling approach [30]. A typical EMA design can involve one or both of the two sampling approaches.

For the event-based design, assessments on participants' instantaneous situation, including questionnaires on location, activities, socializing, mood, food and drink, etc., are taken at the time when events occur. The occurrence of events can be either recorded by self-reported or detected by the devices. Each event has a certain probability of being assessed and sampling probabilities are applied so that participants do not feel overburdened during the study.

On occasions where data are collected through self-report procedures, subjects may be bored and won't even look at the questions when they answer them. Then, to appropriately adjust for compliance, a probability of assessment can be applied so that the participants only answer the questionnaire on a subset of events. Accordingly, researchers can get some assessment of the events without "annoying" participants too much. The probability of assessment is specific for the event-based design.

In a time-based design, electronic diaries prompt randomly at times other than the event times. The devices may prompts over the entire sampling period. When a random prompt occurs, subjects are also given a set of questions on their instantaneous situations. The time-based sampling assessments are referred to as random prompt (RP) assessments, or non-event assessments.

The frequency of the RPs can be determined by the total duration of time available and the expected number of RPs assessments within a single day. The two sampling schemes are depicted in Figure 2.2.1. The horizontal line represents time. The thin blue and thick green verticle lines represent events and random prompts, respectively. The dard red arrow indicates that the events are assessed.

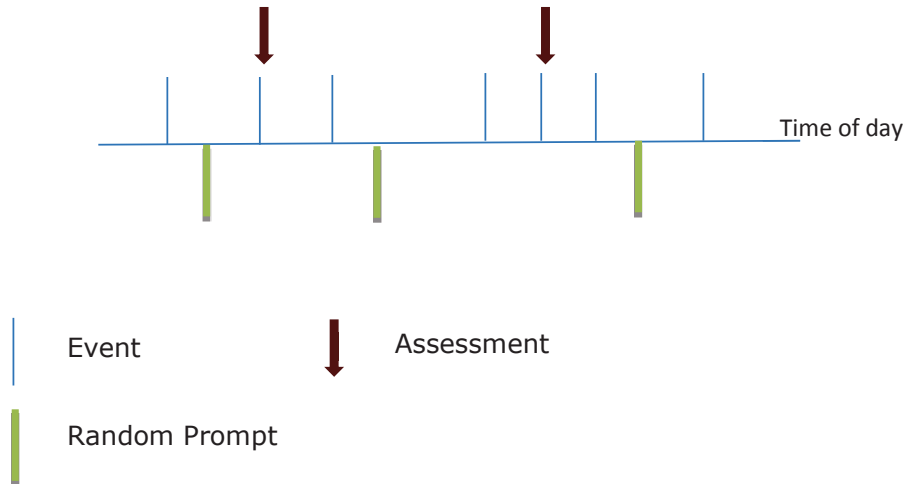


Figure 2.1: EMA Sampling Scheme

2.2.2 Situational Variables and Temporal Patterns

Per the protocol of the Shiffman et al. smoking study, subjects were assessed with a large pool of questions measuring their instantaneous mood, activities, location, socializing and environment. These measurements were referred to as situational variables. A possible assumption is that these situational variables may affect the outcome of interest and thus there might be potential interest to analyze temporal patterns at both the individual and population levels.

2.2.3 Missing by Design (MBD) and Analytical Difficulties

In the Shiffman smoking study, both events-based and time-based sampling schemes were applied. For the events-based design, not all events can be assessed. For reasons involving compliance considerations (participants may not adhere to the protocol if they are asked to answer too many questions), only part of the events are assessed. For DS, the expected number of assessments was 4.5 per day and the probabilities of a cigarette being assessed were calculated as a function of the number of cigarettes consumed the prior day. For ITS, the probabilities of assessment are 1 for the first observation, 0.5 for the second to fourth observations, and 0.25 for the rest.

The time-based sampling was conditional on the event-based scheme where the electronic devices occasionally beep smokers and asked questions on situational variables while they were not smoking ([28]). Here, the electronic diary would not prompt within 15 minutes of a smoking episode. The relationship between Assessed events (Assessed-E), Events and RP can be summarized in the Figure 2.2.3. Both RP and the probabilities of assessments are conditional on events. When observations are not assessed, we refer the observations as missing “by design” (MBD). MBD may complicate the statistical analysis. Other types of missingness may also be present.

In addition to the missing data issue, the complexity of the sampling schemes introduces potential analytical difficulties and hence, classic mixed effect methods, such as generalized estimating equation (GEE) techniques and random effect models need to be used with great caution. One reason for this is that waiting time of a random prompt may need to be incorporated in the analyses. When comparing the outcome and covariates from events-based and randomly collected sampling schemes, the absolute value of coefficients may be subject to the number of random prompt assessments. In addition, because the events-based and RP samples are inherently different, the above-mentioned models may not adequately account for within subject correlation and also may not adequately characterize covariate and event time temporal patterns.

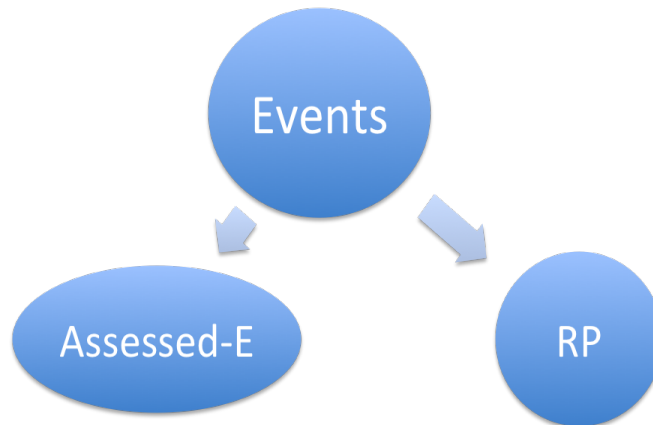


Figure 2.2: Relationship between Assessed Events, Events and Random Prompts

2.2.4 An Example – EMA Data Structure for Our Data

The data structure from our EMA study is briefly summarized in Table 2.1. The values are arbitrary. EMA data can be indexed by subject and time. In the data set, each row represents one observation, which is associated with either an event record type or a random prompt record type. An event could be either assessed or non-assessed. If the event is non-assessed, then corresponding time-varying covariates are missing by design (MBD). All of the random prompts are assessed.

2.3 TRADITIONAL LONGITUDINAL ANALYSIS METHODS

Traditional analytical tools originally designed for longitudinal studies with small numbers of observations per subject are often used for analyzing EMA data [30, 28]. In this section, we

Table 2.1: EMA Data Structure

Subj	Time	Record	P(Ass)	Assess	Covariate
1	0.2	E	1	1	xxx
1	0.4	E	0.5	0	MBD
1	0.7	RP	1	1	xxx
1	0.8	E	0.5	1	MAR
1	0.9	E	0.25	0	MBD

review these classic analytical tools and in the next section, we will introduce some modern methodologies that are modified for EMA data.

2.3.1 Linear Mixed Effects Models (LMMs)

Probably, the most widely used modern technique for analyzing continuous longitudinal data involves linear mixed effects models. In linear mixed models, the covariates could be either time-varying and time-invariant. The parameters are efficiently estimated via likelihood-based methods. Restricted maximum likelihood (REML) is frequently used for parameter estimates and making inferences because the bias of parameter estimates with small sample size on maximum likelihood-based approaches can be reduced by REML. The linear mixed effects models can handle unbalanced data with both time-varying a time-invariant covariates. The structure of covariance matrix is pre-specified. A linear mixed effects model is written as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \quad (2.1)$$

where \mathbf{Y}_i is the continuous outcome vector for subject i , \mathbf{X}_i contains the fixed covariates for the i th subject, \mathbf{Z}_i contains the random covariates, $\boldsymbol{\beta}$ is the “fixed” (population) parameter,

\mathbf{b}_i is the random trajectory component for individual i , $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{G})$ and $\boldsymbol{\epsilon}_i$ is the error term, $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \mathbf{R})$. Because of the linearity of the fixed effects, only polynomial models can be considered for summarizing population effects. This is a limitation especially when individuals exhibit periodic behavior patterns.

2.3.2 Generalized linear mixed effects models (GLMMs)

Generalized linear effect models, as its name, are a generalization of linear mixed effects models. These models release the assumption that the outcome is continuous but still allow one to assume random effects. Generalized linear mixed effects model [GLMMs] [24] can be written as

$$f(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \quad (2.2)$$

where $f(\cdot)$ is a link function. Examples of link functions are *logit* function for binary data and *log* function for count data. Other parameters are the same as described in the linear mixed effects models.

In generalized linear mixed effects models, the mean response of the transformed outcomes is conditional on both the observed predictors and the random effects. The random effects are assumed to be independent of covariates. A beta-binomial distribution model is an example of GLMM. This model can be conceptualized as a two-stage model. In the first stage, the binary responses are conditionally independent and identically distributed. In the second stage, these binary probabilities are independently distributed as a beta distribution. This model can account for over-dispersion of binomial variance and can account for the clustered feature in longitudinal data.

GLMMs have some limitations. For example, they assume the same marginal distribution at each measurement occasion. However, this may not always be the case because in a longitudinal study, the mean response may change over time. linear mixed effects model can be thought as a special case of generalized models where the identity link is applied.

2.3.3 Generalized Estimating Equation

To handle the situation where each subject has multiple observations, both LMMs and GLMMs assume subjects are randomly chosen from a pre-specified distribution. The *generalized estimating equation* (GEE) method, proposed by Liang and Zeger in 1986, is an alternative approach where GEE uses a correlation matrix to represent the within subject observations ([22]). GEE methodology uses a quasi-likelihood approach, which incorporates additional nuisance parameters for the covariance matrix of the response vector. GEE can handle unbalanced data and mixture of discrete and continuous predictors. GEE requires missing completely at random. In addition, parameter estimates in GEE are robust to misspecification of within subject correlation. But note that although the estimators are consistent under different within subject correlation, the standard errors are not robust to misspecification of within subject correlations. The empirical estimates of standard errors are asymptotically robust, i.e. robust for large sample sizes.

A key difference between GEE and GLMMs is that the mean response and within-subject association are modeled separately in GLMMs with the random covariance matrix and the within subject residual matrix, where in GEE, the cluster effect is adjusted for using a correlation matrix. The separation of the modeling of the mean response and the association among responses has important implications for interpretation of the parameter estimation. For GEE, the interpretation of the parameter estimates is population averaged, whereas in LMMs and GLMMs, each subject can have his/her own coefficients.

2.3.4 Growth Curve Models

A third approach to analyzing longitudinal data is growth curve models. This approach is also known as trajectory analysis. These types of models can assess the pattern of the outcome of interest over time for a single population, or classify subjects into different subgroups. Growth curve models are empirical models, which can be linear or nonlinear. In

longitudinal settings, individuals may have different trajectories, which can be clustered into different classes. The clustering procedure allows parameters to vary for different classes and suggests that the trajectories classes are latent ([1]). Each class has a group-level overall growth curve and this model is then called growth mixture models. Growth curve models are a generalization of mixed effects models, while growth mixture models are an extension of growth curve models.

2.3.5 Transition Models

Transition models are yet another approach to analyzing longitudinal categorical outcomes. Unlike GEE, in the transition models, the mean responses of the outcome depend on both the predictors and the history of each subject. This is adopted with the sequential time nature of longitudinal data. The conditional distribution is modeled on both the predictors and the previous responses. The dependencies of repeated measures can be taken into account by conditioning on the previous responses [23]. One difficulty that transition model encounter is that the initial values of states should be incorporated in the conditional probability; however, in most cases, the distribution of the initial responses cannot be estimated/obtained.

There are several limitations of the transition model. One disadvantage is that the models are appropriate when time is equally spaced with no missing values. Moreover, the interpretation will differ if conditioning on the previous one response or previous two responses. If a covariate is a significant predictor to mean responses at all of the time points, then conditioning the current response on previous responses may result in biased parameter estimates.

2.4 ANALYTICAL METHODS FOR INTENSIVE LONGITUDINAL STUDIES

Traditional methodologies focus on the global pattern of effects and may wash out potential micro patterns captured in EMA studies. These methods may oversimplify the structure of EMA data because they simply assume the observations are repeated measurements and ignore how different sources of repeated measurements are obtained (event-based or random-prompt). They also ignore the fact that each subject has more observations and the data may contain useful temporal behavior information. Consequently, such multilevel analysis methods do not have the ability to detect temporal patterns among subjects and hence, special analytical tools are needed.

Two important features need to be considered simultaneously while analyzing EMA data. The first is the comparison between individual subjects and the entire population (subject dimension); the second is the comparison between the local and the global time frame (time dimension). A successful analytical methodology for EMA data should be a balance between the two dimensions. In our example (described later), we found that for some subjects, behaviors were sometimes clustered in time. For these temporal behavior, or micro-processes, we define them as time-clusters. Now we have two types of clusters here that we want to model them together. Further details about methods that can capture both of the above features are discussed in later subsections and in Chapter 3.

There are several ways of looking at EMA data. One way is to look at the temporal patterns using traditional methodology combined with non-parametric splines [36]. A second way is to consider each participant as a time series [27]. A third way is to view the data as stochastic processes [26]. Each of these three perspectives can accommodate different types of complexity in temporal data and will be discussed in the subsections 2.5.1–2.5.3. The specific EMA analytical methods being discussed fall into the following three categories:

- (1) Nonparametric splines and polynomial models;

- (2) Time Series Analysis; and
- (3) Recurrent Event Analysis .

2.4.1 Nonparametric splines and polynomial models

Non-parametric techniques have been used to study patterns among data. When the raw data is heterogeneous, such techniques may result in complicated patterns that are difficult to interpret. Mixed effects models and non-parametric techniques can be combined to partially address this problem.

One example is a time-varying coefficient model proposed by Wu and Zhang in 2005 [36]:

$$y_i(t) = c_i(t) + \mathbf{X}_i(t)' \mathbf{B}(t) + e_i(t), i = 1, 2, 3, \dots, n \text{ subjects,}$$

where $y_i(t)$ is the continuous outcome for i^{th} subject at time t , $c_i(t)$ is a time-varying intercept, $\mathbf{X}_i(t)$ is a vector of covariates and $\mathbf{B}(t)$ is a time-varying coefficient vector associated with the corresponding covariates in $\mathbf{X}_i(t)$. The error term, $\mathbf{e}_i(t)$, is normally distributed as $N(\mathbf{0}, \Gamma_{\mathbf{e}})$. Wu and Zhang assume that the time-varying slopes are linear combinations of time-bases: $\mathbf{B}(t) = (b_1(t), b_2(t), \dots, b_p(t))' = (\Psi_1 \boldsymbol{\beta}_1, \Psi_2 \boldsymbol{\beta}_2, \dots, \Psi_p \boldsymbol{\beta}_p)'$, where p is the number of total covariates. $\Psi_{p_i} = (\psi_{i1}, \psi_{i2}, \dots, \psi_{i r_{p_i}})'$ is the time basis for the p_i th covariate, and r_{p_i} is the number of elements in this basis. Then let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_p)'$. Due to the martingale properties of Gaussian process, the above model could be transformed into the following formulae and analyzed using mixed models techniques:

$$y_{ij} = \mathbf{z}'_{ij} \boldsymbol{\alpha}_i + \mathbf{x}'_{ij} \boldsymbol{\beta} + e_{ij}, \tag{2.3}$$

where $\boldsymbol{\alpha}_i \sim N(\mathbf{0}, \mathbf{D})$, $e_{ij} \sim N(0, R_i)$, where z_{ij} represent the random effects covariates z_{ij} and \mathbf{x}_{ij} represents fixed covariates. z_{ij} and \mathbf{x}_{ij} are function of time bases and covariates. Theoretically, the model allows each covariate to have its own time basis, but the computation would be difficult. Consequently, in practice, a shared time basis is usually assumed.

In Wu and Zhang’s model, continuous time is partitioned into several pieces, forming a time basis. The estimated coefficients are time-varying, forming a linear combination of the time basis. One could also assume a random intercept and/or slope for each subject, and then individual time-varying processes are obtained. All subjects share the common time basis but have different coefficients of the time-basis. The subject-random effect and time-random effects are independent. The random covariance matrix, G , for the random effect, is the outer product of the two types of random effects. The model is a series of splines models when conditional on the subject random effects. The estimation method for this model is through maximum likelihood. Because this is the key model used in this dissertation, further analytical detail will be provided later.

This type of model combines a penalized splines component with the mixed effect component. Moreover, all of the methodologies discussed in the previous section have the potential to be generalized to dynamic versions when considering the time-varying relationship between the outcome and covariates.

We will consider the dynamic models that are not computationally difficult but still allow the coefficients to vary over time. Dynamic generalized linear models were introduced by West et al. [34] and Hastie and Tibshirani [17]. The models were first applied to longitudinal data by Brumback and Rice, Hoover et al. [4, 14]. Early versions of these models ignored the within subject correlation. Liang and Wu extended the models to incorporate the within subject correlation for general penalized splines [21] and a smoothing spline model [35]. Recently, the penalized version has been applied to EMA data by Tan et al. via a linear time varying mixed effects model [32]. For our application, we apply the penalized spline models to gap time intervals observed between smoking event times.

2.4.2 Time Series Models

Time series approaches are widely used in economics to study long temporal patterns and can detect small changes in time, i.e., the flow of stocks. Accordingly, it seems reasonable

to borrow ideas from time series analytical methods and apply them to EMA data. Rovine et al. [27] used the p^{th} order auto-regressive $AR(p)$ series to model the between subject differences on a daily-alcohol-usage data set. In their analysis, they assumed that the outcome of interest (daily alcohol consumption), y_{ti} , followed an $AR(p)$ process, $p = 1, 2$, and also assumed the coefficients in the auto-regressive series were functions of the covariates of interest. While modeling the coefficients, they proposed linear mixed effects models. Their model is written as a two level model summarized below.

Level 1: Assume the mean response y_{ti} is an $AR(1)$ process

$$y_{ti} = \alpha_{0i} + \alpha_{1i}y_{t-1,i} + r_{ti} \quad (2.4)$$

where r_{ti} is a white noise process that follows three conditions:

1. $E(\epsilon_{i(t)}) = 0$
2. $E(\epsilon_{i(t)}^2) = \sigma_i^2$
3. $E(\epsilon_{i(t)}\epsilon_{i(s)}) = 0 \forall t \neq s$.

Level 2: Model the coefficients in the $AR(1)$ process

$$\alpha_{0i} = \boldsymbol{\beta}'\mathbf{x}_i + \omega_{0i} \quad (2.5)$$

$$\alpha_{1i} = \boldsymbol{\beta}'\mathbf{x}_i + \omega_{1i} \quad (2.6)$$

where ω_{0i} and ω_{1i} jointly followed a bi-variate normal distribution with mean $\mathbf{0}$, and variance-covariance matrix

This method basically adds a time-series component to a mixed effects model. It allows each individual have its own AR(1) process, and this variation depends on the own individuals' own characteristics.

2.4.3 Analysis of Recurrent Events and Point Process Models

Because events occur over time for each subject, smoking behavior can also be viewed as a stochastic point process. There are three main ways of looking at point process data. The first is to model the instantaneous intensity function, which is defined as the limit of the number of events occurring within a short period, Δt , around a given time t . This instantaneous intensity function is conditional on past events. Rathbun and Shiffman first introduced a point process model to EMA data [26].

Events are assumed to follow a point process:

$$Prob\left\{N[t_1, t_2] = n\right\} = \frac{e^{-\int_{t_1}^{t_2} \lambda(t) dt}}{n!} \left\{\int_{t_1}^{t_2} \lambda(t) dt\right\}^n \quad (2.7)$$

The intensity function is the key to this approach. It is defined as

$$\lambda(t; H_t) = \lim_{\delta \rightarrow 0} \frac{E\{N[t, t + \delta] | H_t\}}{\delta}, H_t = \{t_i : t_i < t\}. \quad (2.8)$$

The intensities can be viewed as the average number of the events at each time unit. When the time unit approaches 0, the average count approaches the instantaneous intensity.

Rathbun, et al., first applied point process models in social sciences and modeled smoking intensities as a function of mood. Rathbun used a modulated Poisson process model and assumed the intensity function was a linear function of time-varying covariates (Rathbun et al., 2006, Cox, 1972). The parametric form of the intensity function is:

$$\lambda_i(t; \boldsymbol{\beta}) = \exp\{\boldsymbol{\beta}' \mathbf{x}_i(t)\} \quad (2.9)$$

The second method is using a rate function, which, in our case, can be defined as the integral of the instantaneous intensity function, from the beginning of the study until a given time t . The differences between the second method and the previous one is that the rate is an integral function in second approach and may be more difficult to estimate as compared to the instantaneous intensity approach.

A third approach is to use the time between events, commonly referred to as “gap times”. This is the approach that we use in this work. This approach attempts to characterize the time patterns of smoking for subjects as well as the population(s) of interest, which is a feature that the other two approaches do not have. A gap time model can be written as [10]

$$g(T_{ij}) = \mu_0 + \boldsymbol{\beta}^T \mathbf{X}_{ij} + U_i + e_{ij} , \quad (2.10)$$

where T_{ij} is the inter-event time and $g(\cdot)$ is a transformation that makes $g(T_{ij})$ normally distributed. The natural logarithmic transformation is frequently used in practice. The term μ_0 is the intercept term, \mathbf{X}_{ij} represents the vector of covariates, $\boldsymbol{\beta}^T$ is a coefficient vector associated with the covariates, U_i is the individual intercept for random effects e_{ij} is the error term. Model (2.10) is also called an accelerated failure time model for inter-event times.

2.5 MISSING DATA

Large amounts of missing information may be present in an EMA data set. This missingness in general can be characterized into three categories: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). MCAR assumes that the missingness is not related to any covariates, MAR assumes that the missingness is random conditional on a particular covariate. MNAR refers to situations where the missingness is conditional on both covariates and outcomes. For example, people may be lost to follow up because they encounter poor outcomes. In addition, missingness can be introduced by the missing by design (MBD) sampling scheme described previously.

To deal with missing data practically, one should first determine the potential sources of missingness. Data can be missing due to various reasons; for example, subjects may randomly skip a couple of questions, or “skip-patterns” may exist in the questionnaires.

To analyze EMA data with missingness, one can

- impute missing data (i.e. multiple imputation); or
- use Expectation-Maximization (EM) algorithm; or
- use the inverse probability of weighting (IPW) method.

For our data, both MAR and MBD are assumed to be present. With MBD cases, each missing observation is associated with a pre-calculated probability of missingness. Thus, we think the IPW method is a good choice when MBD is present.

3.0 PROCEDURES AND PROPOSED METHODOLOGY

In this chapter, a graphical way of identifying and displaying bouts is first proposed. In the second section, we introduce an existing linear mixed effects time varying coefficient model and explain how it is applied to our EMA data. We model the cumulative gap time outcomes as a function of situational covariates. In the third section, we extend this model to handle the cases where missingness occurs due to the study design.

3.1 IDENTIFICATION OF BOUTS

Before qualitatively assessing what factors stimulate smoking bouts, we first need to identify such behavior mathematically. Bouts behavior is not commonly studied and not much information can be found in the literature. Here, we propose a graphical method to identify bouts. Intuitively, a potential time-cluster, or a potential bout, is a group of smoking events that occur within a very short time period. Thus, the bout can be determined when the time period is specified.

We assume that a bout is a period where two or more consecutive events occur within a pre-specified time interval, d . In our application, we applied a log transformation to the gap-times as $\log\left(\frac{T_i(t_{ij})}{d}\right)$, where $T_i(t_{ij})$ is the j^{th} gap time for subject i . There are two reasons for doing this: (1) the log transformed times are less skewed; and (2) to construct

a useful graphical presentation. Hence, we plot $\log\left(\frac{T_i(t_{ij})}{d}\right)$ versus time, t , so that we can easily identify bouts whenever $\log\left(\frac{T_i(t_{ij})}{d}\right) < 0$.

To model the effect of covariates on gap times when bouts occur, we sum over the log transformed gap time and obtain $\sum_{t_{ij} \leq t} \log\left(\frac{T_i(t_{ij})}{d}\right)$. Hence, bouts can be identified whenever the resulting slope is negative. We require that the pre-defined gap-time designation for the identification of bouts, d , to be the same all individuals. d can be considered as the “scale” of bouts, i.e. how intensively the consumed cigarettes are considered as being in bouts. This is related to the overall distribution of gap time intervals. Moreover, when d is specified, the transformation is equivalent to subtracting a constant from $\log(T_i(t))$. We assume that d is a pre-specified period of bout time that is clinically meaningful. Here, to explain our method, we set $d = 1$ hour. Thus, the constant we use for our application is $\log(d) = \log(1) = 0$. We use these definitions to help visualize smoking patterns over time. In the next section, we will illustrate how to model the cumulative gap-times to explore the relationship between covariates and gap times when bouts occurred.

3.2 TIME VARYING COEFFICIENTS CUMULATIVE GAP TIME MODELS

In the recurrent-events survival analysis framework, a model for cumulative inter-event time, or gap time can be written as:

$$y_i(t_{ij}) = \sum_{j^*=1}^j \log\left(\frac{T_i(t_{ij^*})}{d}\right) = \alpha_{0i}(t_{ij}) + \mathbf{x}_i(t_{ij})^T \boldsymbol{\alpha}_1(t_{ij}) + \epsilon_i(t_{ij}) \quad (3.1)$$

$$\alpha_{0i}(t) \sim N(\mu(t), \Gamma) \text{ at time } t, \quad \epsilon_i(t) \sim N(0, \sigma_R^2)$$

$$j = 1, 2, \dots, m_i, \quad i = 1, 2, \dots, n;$$

where i refers to subjects, t refers to the time, t_{ij} refers to the time when the j^{th} event of i^{th} subject occurs, $T_i(t_{ij})$ is the gap time for subject i at time t_{ij} , $\mathbf{x}_i(t)$ is a column vector of

covariates, and $\boldsymbol{\alpha}_1(t)$ is the column vector of corresponding coefficients. Note that T_{ij} , where j represents the j^{th} event of i^{th} subject, is equivalent to $T_i(t_{ij})$, as was used earlier. When only one covariate is present, $\mathbf{x}_i(t_{ij})^T$ and $\boldsymbol{\alpha}_1(t)$ become scalar quantities $x_i(t_{ij})$ and $\alpha_1(t)$. For simplicity, we will model only one covariate.

Model (3.1) can be viewed as a dynamic version of a random intercept model for gap times. It is different from a typical accelerated gap time model in the way the outcome is defined, that is, we use the cumulative log transformed gap time instead of the instantaneous gap time. We do this because the cumulative version is more stable to model. However, mathematically, the estimation procedure remains the same. Another reason for using the cumulative log gap times is that this function displays a sawtooth wave pattern during times when bouts exist, whereas by using instantaneous log gap times, distinct patterns at bout times cannot be as easily visualized.

Following Zhang [36], the continuous time t is divided into several partitions, referred to as a time basis [36] and the time-varying coefficients are obtained by a linear combination of a spline basis, $\boldsymbol{\Phi} = (\phi_1, \dots, \phi_l)^T$. That is,

$$\mu(t) = \sum_k \mu_{0k} \phi_k = \boldsymbol{\Phi}^T \boldsymbol{\mu}_0, \quad (3.2)$$

$$\alpha_{0i}(t) = \sum_k a_{0ik} \phi_k = \boldsymbol{\Phi}^T \mathbf{A}_{0i}, \text{ and} \quad (3.3)$$

$$\alpha_1(t) = \sum_k a_{1k} \phi_k = \boldsymbol{\Phi}^T \mathbf{A}_1. \quad (3.4)$$

Then, the discrete version of equation (3.1) is given by

$$\mathbf{Y} = \sum_{k \leq j} \log \left(\frac{T_i(t_{ik})}{d} \right) = \mathbf{Z} \mathbf{A}_0 + \mathbf{X} \mathbf{A}_1 + \boldsymbol{\epsilon} \quad (3.5)$$

where

$$\mathbf{Y} = (y_{11}, \dots, y_{nm_n})^T, \quad \mathbf{A}_1 = (\alpha_1 1, \dots, \alpha_1 l), \quad (3.6)$$

$$\mathbf{A}_0 = (\mathbf{A}_{01}^T, \dots, \mathbf{A}_{0n}^T)^T, \quad \mathbf{A}_{0i} = (\alpha_{0i} 1, \dots, \alpha_{0i} l), \quad (3.7)$$

$$\mathbf{A}_{0i} \sim N(\mathbf{M}, \mathbf{G}), \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R}), \quad (3.8)$$

$$j = 1, 2, \dots, m_i, \quad i = 1, 2, \dots, n. \quad (3.9)$$

Because the parameter function, $\alpha_{0i}(t)$, is normal at any arbitrary time t , it is natural to assume that at each component of the time basis, $\mathbf{A}_{0i} \sim N(\mathbf{M}, \mathbf{G})$. Then $\alpha_{0i}(t)$ is merely a linear combination of normal distributions so that the relationship between the random effect covariance matrices \mathbf{G} and $\mathbf{\Gamma}$ is $\mathbf{G} = \mathbf{\Phi}^T \mathbf{\Gamma} \mathbf{\Phi}$. In our specific example, only the intercept effect varies by subject; thus, \mathbf{G} is a scalar and is denoted as G henceforth.

The model assumes that the time-varying coefficients and errors are independent. All subjects share a common time basis for both fixed and random effects. It is also possible to assume different time bases for fixed and random effects. However, such an assumption would greatly increase the number of parameters to be estimated. Hence, to simplify technical difficulties and potential nonidentifiability, and to focus on the dynamic model itself, we assume that only one basis is present. For each component of the basis, individuals may have their own associated coefficients. To derive the design matrix for the dynamic population effect, we note that

$$\begin{aligned} \mathbf{x}(t)\alpha_1(t) &= \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1m_1} \\ \vdots \\ x_{n1} \\ x_{n2} \\ \vdots \\ x_{nm_1} \end{bmatrix}_{N \times 1} \left(\begin{bmatrix} \phi_1, \phi_2, \dots, \phi_l \end{bmatrix}_{1 \times l} \begin{bmatrix} \alpha_{11} \\ \alpha_{12} \\ \vdots \\ \alpha_{1l} \end{bmatrix}_{l \times 1} \right) \\ &= \begin{bmatrix} x_{111}, x_{112}, \dots, x_{11l} \\ \vdots \\ x_{1m_11}, x_{1m_12}, \dots, x_{1m_1l} \\ \vdots \\ x_{n11}, x_{n12}, \dots, x_{n1l} \\ \vdots \\ x_{nm_11}, x_{nm_12}, \dots, x_{nm_1l} \end{bmatrix}_{N \times l} \begin{bmatrix} \alpha_{11} \\ \alpha_{12} \\ \vdots \\ \alpha_{1l} \end{bmatrix}_{l \times 1} \end{aligned}$$

$$= \begin{pmatrix} \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1m_1} \\ \vdots \\ x_{n1} \\ x_{n2} \\ \vdots \\ x_{nm_n} \end{bmatrix}_{N \times 1} \otimes \begin{bmatrix} \phi_1, \phi_2, \dots, \phi_l \end{bmatrix}_{1 \times l} \end{pmatrix} \begin{bmatrix} \alpha_{11} \\ \alpha_{12} \\ \vdots \\ \alpha_{1l} \end{bmatrix}_{l \times 1}, \quad (3.10)$$

where $\mathbf{x} = (x_{11}, x_{12}, \dots, x_{1m_1}, \dots, x_{n1}, x_{n2}, \dots, x_{nm_n})$ is a covariate vector, $N = \sum_{i=1}^n m_i$, $x_{ijk} = x_{ikj_k} = x_{ij}(t_{ij} - t_k) +$ when j_k is the j^{th} measurement in k^{th} element in the time basis, and $x_{ijk} = 0$ when t_{ij} is not in the k^{th} partition of the time basis.

The design matrix for the dynamic individual effect in model (3.5) is derived in a similar way:

$$\mathbf{z} \boldsymbol{\alpha}_0(t) = \begin{bmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \dots & & & \\ 0 & \dots & 0 & J_n \end{bmatrix} \begin{pmatrix} \boldsymbol{\Phi}^T \mathbf{A}_{01} \\ \boldsymbol{\Phi}^T \mathbf{A}_{02} \\ \vdots \\ \boldsymbol{\Phi}^T \mathbf{A}_{0n} \end{pmatrix} = \begin{pmatrix} J_1 \otimes \boldsymbol{\Phi}^T & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & J_2 \otimes \boldsymbol{\Phi}^T & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & J_n \otimes \boldsymbol{\Phi}^T \end{pmatrix} \begin{bmatrix} \mathbf{A}_{01} \\ \mathbf{A}_{02} \\ \vdots \\ \mathbf{A}_{0n} \end{bmatrix} \quad (3.11)$$

where $\mathbf{z} = \begin{bmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \dots & & & \\ 0 & \dots & 0 & J_n \end{bmatrix}$ is the design matrix for the random intercept in model (3.1), $J_i = (1, 1, \dots, 1)^T$ is a $n_i \times 1$ vector of ones, $\mathbf{A}_{0i} = (\alpha_{0i1}, \alpha_{0i2}, \dots, \alpha_{0il})$, $i = 1, 2, \dots, n$.

The dynamic model described here is equivalent to estimating $(l - 1)$ more parameters in a linear mixed effects model where l is the number of components in the time basis. When the number of dynamic effects is large, computation can be a problem. Thus, how to choose an appropriate basis is very important. This question involves two considerations:

one is to choose a moderate number of knots and the other is to choose a degree of polynomial (linear, quadratic, cubic etc.) for components of the spline basis. One should note that the number of components increases as the degree increases. In our example, a penalized linear spline basis was applied and the basis is given by:

$$\Phi = (\phi_1, \dots, \phi_l)^T = (1, t, (t - t_1)_+, (t - t_2)_+, \dots, (t - t_{l-2})_+)^T \quad (3.12)$$

where

$$(t - t_k)_+ = \begin{cases} t - t_k & \text{if } t > t_k \\ 0 & \text{otherwise} \end{cases}, \quad k = 1, 2, \dots, l.$$

Consequently, in our linear spline model, a basis with l components and $l - 2$ knots need to be selected.

The coefficients of the terms $(t - t_1)_+$, $(t - t_2)_+$, \dots , $(t - t_l)_+$ introduce extra terms for the time varying intercepts and time varying coefficients in the likelihood function. Then, for mixed effects models, following the penalized generalized log-likelihood (PGLL) criteria [19], we need to minimize:

$$\begin{aligned} Q = & (\mathbf{Y} - \mathbf{Z}\mathbf{A}_0 - \mathbf{X}\mathbf{A}_1)^T \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{Z}\mathbf{A}_0 - \mathbf{X}\mathbf{A}_1) \\ & + \mathbf{A}_0^T \mathbf{G}_*^{-1} \mathbf{A}_0 + \lambda_1 \mathbf{A}_1^T \mathbf{E} \mathbf{A}_1 + \lambda_0 \mathbf{A}_0^T \mathbf{E} \mathbf{A}_0 \end{aligned} \quad (3.13)$$

where $\mathbf{R}^{-1} = \text{diag}(\mathbf{R}_1^{-1}, \dots, \mathbf{R}_n^{-1})$ is a block diagonal matrix with n blocks, $\mathbf{R}_i = \sigma_R^2 \otimes \mathbf{I}_{n_i}$, where \mathbf{I}_{n_i} is a $n_i \times n_i$ identity matrix, n_i is the observation number for subject i , $\mathbf{G}_*^{-1} = \text{diag}(G^{-1}, \dots, G^{-1})$ is a $n \times n$ matrix, $\mathbf{E} = \text{diag}(0, \dots, 0, \dots, 1, \dots, 1)$ is a $l \times l$ matrix with last $(l - 2)$ components being 1, n is the number of total subjects. To minimize (3.13), \mathbf{A}_0 and \mathbf{A}_1 can be estimated by

$$\hat{\mathbf{A}}_1 = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} + \lambda_1 \mathbf{E})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}, \quad \text{and} \quad (3.14)$$

$$\hat{\mathbf{A}}_0 = \mathbf{D} \mathbf{Z}^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}^T \hat{\mathbf{A}}_1) \quad (3.15)$$

where $\mathbf{D} = (\mathbf{G}_*^{-1} + \lambda_0 \mathbf{E})^{-1}$ acts as the role of random effect covariance matrix when the penalty due to the linear penalized spline time basis is present, $\mathbf{V} = \mathbf{Z} \mathbf{D} \mathbf{Z}^T + \mathbf{R}$ acts as the

overall sum of square errors to the model. When no fixed dynamic coefficient is present, the formula for \mathbf{A}_1 reduces to an estimation for fixed covariates in a linear random effects model.

Then, the coefficient of the k^{th} element in the time basis is obtained by

$$\begin{aligned}\hat{\alpha}_{1k} &= \mathbf{H}_k(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} + \lambda_1 \mathbf{E})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}, \\ \hat{\alpha}_{0ik} &= (\mathbf{H}_i \otimes \mathbf{H}_k) \mathbf{DZ}^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}^T \hat{\mathbf{A}}_0), \text{ for subject } i,\end{aligned}\tag{3.16}$$

where $\mathbf{H}_k = \text{diag}(0, \dots, 0, 1, 0, \dots, 0)$, is a $l \times l$ diagonal matrix where the k^{th} diagonal element equals to 1 and otherwise 0, $\mathbf{H}_i \otimes \mathbf{H}_k$ is a $nl \times nl$ diagonal matrix where the ik^{th} diagonal element equals to 1 and otherwise 0, $ik = i \times k$, l is the total number of elements in the time basis and n is the total number of subjects.

Finally, the time varying coefficients are given as

$$\hat{\alpha}_1(t) = \Phi^T \hat{\mathbf{A}}_1 = \sum_{k=1}^l \phi_k \hat{\alpha}_{1k} = \sum_{k=1}^l \phi_k \mathbf{H}_k (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} + \mathbf{E})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}, \text{ and}\tag{3.17}$$

$$\begin{aligned}\hat{\alpha}_{0i}(t) &= \Phi^T \hat{\mathbf{A}}_{0i} = \sum_{k=1}^l \phi_k \hat{\alpha}_{0ik} = \sum_{k=1}^l \phi_k (\mathbf{H}_i \otimes \mathbf{H}_k) \mathbf{DZ}^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}^T \hat{\mathbf{A}}_0), \\ &\text{for subject } i,\end{aligned}\tag{3.18}$$

where $\Phi = (\phi_1, \dots, \phi_l)^T = (1, t, (t - t_1)_+, (t - t_2)_+, \dots, (t - t_{l-2})_+)^T$.

To estimate the standard errors, Zhang and Wu use a Bayesian approach as follows:

$$\mathbf{Y} | \mathbf{A}_0, \mathbf{A}_1 \sim N(\mathbf{X}\mathbf{A}_1 + \mathbf{Z}\mathbf{A}_0, \mathbf{R}),\tag{3.19}$$

where \mathbf{A}_0 and \mathbf{A}_1 are independent, with priors:

$$\mathbf{A}_1 \sim N(\mathbf{0}, (\lambda_1 \mathbf{E})^{-1}), \mathbf{A}_0 \sim N(\mathbf{0}, \mathbf{D}).\tag{3.20}$$

Then the following theorems hold [37]:

Theorem 3.2.1. *Under the conditions of equations 3.19 and 3.20, the following properties hold:*

$$\mathbf{A}_1|\mathbf{Y} \sim N\left(\hat{\mathbf{A}}_1, (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} + \lambda_1\mathbf{E})^{-1}\right) \quad (3.21)$$

$$\mathbf{A}_0|\mathbf{Y} \sim N\left(\hat{\mathbf{A}}_0, \mathbf{D} - \mathbf{DZ}^T\mathbf{V}^{-1}\mathbf{ZD}\right) \quad (3.22)$$

$$\boldsymbol{\epsilon}|\mathbf{Y} \sim N\left(\hat{\boldsymbol{\epsilon}}, \mathbf{R} - \mathbf{RW}_R\mathbf{R}\right) \quad (3.23)$$

where $\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{ZA}_0 - \mathbf{XA}_1$, $\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \mathbf{Z}\hat{\mathbf{A}}_0 - \mathbf{X}\hat{\mathbf{A}}_1$, and $\mathbf{W}_R = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} + \lambda_1\mathbf{E})^{-1}\mathbf{X}^T\mathbf{V}^{-1}$.

Theorem 3.2.2. *Under the conditions of equations 3.19 and 3.20, the following properties hold:*

$$\mathbf{A}_0|\mathbf{Y}, \mathbf{A}_1 \sim N\left(\mathbf{DZ}^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{XA}_1), \mathbf{D} - \mathbf{DZ}^T\mathbf{V}^{-1}\mathbf{ZD}\right) \quad (3.24)$$

$$\boldsymbol{\epsilon}|\mathbf{Y}, \mathbf{A}_1 \sim N\left(\mathbf{RV}^{-1}(\mathbf{Y} - \mathbf{XA}_1), \mathbf{R} - \mathbf{RV}^{-1}\mathbf{R}\right) \quad (3.25)$$

where $\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{ZA}_0 - \mathbf{XA}_1$.

Theorem 3.2.3. *Under the conditions of equations 3.19 and 3.20, the following properties hold:*

$$\mathbb{E}(\hat{\sigma}_R^2|\mathbf{Y}) = \frac{1}{n} \sum_{i=1}^n \left(\hat{\boldsymbol{\epsilon}}_i \hat{\boldsymbol{\epsilon}}_i^T + \sigma_R^2 (\mathbf{I}_{n_i} - \sigma_R^2 \text{tr}(\mathbf{W}_i^{-1})) \right), \text{ and} \quad (3.26)$$

$$\mathbb{E}(\hat{\mathbf{D}}|\mathbf{Y}) = \frac{1}{n} \sum_{i=1}^n \left(\hat{\mathbf{A}}_{0i} \hat{\mathbf{A}}_{0i}^T + \mathbf{D}_i - \mathbf{D}_i \mathbf{Z}_i^T \mathbf{W}_i^{-1} \mathbf{Z}_i \mathbf{D}_i \right), \quad (3.27)$$

where $\hat{\boldsymbol{\epsilon}}_i = \mathbf{R}_i \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i^T \hat{\mathbf{A}}_{1i})$, $\mathbf{W}_{Ri} = \mathbf{V}_i^{-1} - \mathbf{V}_i^{-1} \mathbf{X}_i (\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i + \lambda_1 \mathbf{E}_i)^{-1} \mathbf{X}_i^T \mathbf{V}_i^{-1}$ for subject i , $i = 1, 2, \dots, n$.

Theorems 3.2.1, 3.2.2 and 3.2.3, as well as the proofs are given in [37]. The standard errors are then estimated by the Expectation-Maximization (EM) algorithm using equations 3.26 and 3.27 [37].

3.3 TIME VARYING COEFFICIENTS MODELS WITH MISSING DATA

Zhang's original time varying coefficient model did not accommodate missing data. However, in our study, as introduced in previous sections, observations may be missing due to two reasons: missing by design (MBD) and missing at random (MAR). To extend Zhang's model, we use the inverse probability of weighting (IPW) method to handle missing data. For MBD cases, each non-missing observation is associated with a pre-defined probability of assessment, $q_{ij}^{(MBD)}$. For MAR cases, a probability of non-missingness, $\hat{q}^{(MAR)}$, can be estimated through the following logistic regression:

$$\text{logit}(q^{(MAR)}) = \beta_0 + X_1^* \beta_1 + \dots + X_r^* \beta_r, \quad (3.28)$$

where r is the number of predictors that are associated with the probability of observing the data, or non-missingness.

The logistic regressions are conducted only when the observations are assessed. Hence, when both MBD and MAR are present, the probability of non-missingness is $q_{ij}^{(MBD)} \times \hat{q}^{(MAR)}$. This assumes that the MAR and MBD are independent. The missing mechanism can be quite complicated. In this section, we will only discuss the MBD cases.

Rewriting the probability of assessment as q_{ij} allows each observation to be associated with only one value of q_{ij} . To estimate the model, we weight the observation with the inverse of q_{ij} , and minimize the following function:

$$\begin{aligned} Q &= (\mathbf{Y} - \mathbf{Z}\mathbf{A}_0 - \mathbf{X}\mathbf{A}_1)^T \mathbf{W}\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{Z}\mathbf{A}_0 - \mathbf{X}\mathbf{A}_1) \\ &\quad + \mathbf{A}_0^T \mathbf{G}_*^{-1} \mathbf{A}_0 + \lambda_1 \mathbf{A}_1^T \mathbf{E}\mathbf{A}_1 + \lambda_0 \mathbf{A}_0^T \mathbf{E}\mathbf{A}_0 \\ &= (\mathbf{Y} - \mathbf{Z}\mathbf{A}_0 - \mathbf{X}\mathbf{A}_1)^T \mathbf{W}^{1/2} \mathbf{R}^{-1} \mathbf{W}^{1/2} (\mathbf{Y} - \mathbf{Z}\mathbf{A}_0 - \mathbf{X}\mathbf{A}_1) \\ &\quad + \mathbf{A}_0^T \mathbf{G}_*^{-1} \mathbf{A}_0 + \lambda_1 \mathbf{A}_1^T \mathbf{E}\mathbf{A}_1 + \lambda_0 \mathbf{A}_0^T \mathbf{E}\mathbf{A}_0, \end{aligned} \quad (3.29)$$

where $\mathbf{W} = \text{diag}(\mathbf{W}_1, \dots, \mathbf{W}_n)$ is a block diagonal matrix with n blocks of weights, $\mathbf{W}_i = \text{diag}(1/q_{i1}, 1/q_{i2}, \dots, 1/q_{im_i})$ is a $m_i \times m_i$ diagonal matrix, n is the number of total

subjects. Thus, the weight matrix \mathbf{W} is a full rank diagonal matrix and all of its diagonal terms are greater than 0. Then, $\mathbf{W}^{1/2} = \text{diag}(\mathbf{W}_1^{1/2}, \dots, \mathbf{W}_n^{1/2})$ is also a diagonal matrix, where $\mathbf{W}_i^{1/2} = \text{diag}(1/q_{i1}^{1/2}, 1/q_{i2}^{1/2}, \dots, 1/q_{im_i}^{1/2})$.

We let,

$$\mathbf{R}_{new_i} = \mathbf{W}_i^{-1/2} \mathbf{R}_i \mathbf{W}_i^{-1/2}. \quad (3.30)$$

Then,

$$\mathbf{R}_{new}^{-1} = (\mathbf{W}_i^{-1/2} \mathbf{R} \mathbf{W}_i^{-1/2})^{-1} = \mathbf{W}_i^{1/2} \mathbf{R}^{-1} \mathbf{W}_i^{1/2}, \text{ for subject } i, \quad (3.31)$$

Then stacking all subjects, we obtain

$$\mathbf{R}_{new} = \mathbf{W}^{-1/2} \mathbf{R} \mathbf{W}^{-1/2}, \text{ and} \quad (3.32)$$

$$\mathbf{R}_{new}^{-1} = \mathbf{W}^{1/2} \mathbf{R}^{-1} \mathbf{W}^{1/2}. \quad (3.33)$$

Here, accordingly to the PGLL criteria, we need to minimize the function

$$\begin{aligned} Q_{new} &= (\mathbf{Y} - \mathbf{Z}\mathbf{A}_0 - \mathbf{X}\mathbf{A}_1)^T \mathbf{R}_{new}^{-1} (\mathbf{Y} - \mathbf{Z}\mathbf{A}_0 - \mathbf{X}\mathbf{A}_1) \\ &\quad + \mathbf{A}_0^T \mathbf{G}_*^{-1} \mathbf{A}_0 + \lambda_1 \mathbf{A}_1^T \mathbf{E} \mathbf{A}_1 + \lambda_0 \mathbf{A}_0^T \mathbf{E} \mathbf{A}_0 \\ &= \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{Z}_i \mathbf{A}_{0i} - \mathbf{X}_i \mathbf{A}_1)^T \mathbf{R}_{new_i}^{-1} (\mathbf{Y}_i - \mathbf{Z}_i \mathbf{A}_{0i} - \mathbf{X}_i \mathbf{A}_1) \\ &\quad + \sum_{i=1}^n \mathbf{A}_{0i}^T \mathbf{G}_*^{-1} \mathbf{A}_{0i} + \lambda_1 \mathbf{A}_1^T \mathbf{E} \mathbf{A}_1 + \sum_{i=1}^n \lambda_0 \mathbf{A}_{0i}^T \mathbf{E} \mathbf{A}_{0i} \end{aligned} \quad (3.34)$$

where $\mathbf{R}_{new}^{-1} = \text{diag}(\mathbf{R}_{new_1}^{-1}, \dots, \mathbf{R}_{new_n}^{-1}) = \text{diag}(\mathbf{W}_1^{1/2} \mathbf{R}_1^{-1} \mathbf{W}_1^{1/2}, \dots, \mathbf{W}_n^{1/2} \mathbf{R}_n^{-1} \mathbf{W}_n^{1/2}) = \mathbf{W}^{1/2} \mathbf{R}^{-1} \mathbf{W}^{1/2}$ is a block diagonal matrix with n blocks, $\mathbf{R}_i = \sigma_R^2 \otimes \mathbf{I}_{n_i}$, where \mathbf{I}_{n_i} is a $n_i \times n_i$ identity matrix, n_i is the observation number for subject i , $\mathbf{W} = \text{diag}(\mathbf{W}_1, \dots, \mathbf{W}_n)$ is a block diagonal matrix with n blocks of weights, $\mathbf{G}_*^{-1} = \text{diag}(G^{-1}, \dots, G^{-1})$ is a $n \times n$ matrix, $\mathbf{E} = \text{diag}(0, \dots, 0, \dots, 1, \dots, 1)$ is a $l \times l$ matrix with last $(l-2)$ components being 1, n is the number of total subjects.

Correspondingly, to minimize (3.29), we take the derivatives of Q_{new} with respect to \mathbf{A}_0 and \mathbf{A}_1 and make them equal to 0's, we get:

$$-\sum_{i=1}^n \mathbf{X}_i^T \mathbf{R}_{new_i}^{-1} (\mathbf{Y}_i - \mathbf{Z}_i \mathbf{A}_{0i} - \mathbf{X}_i \mathbf{A}_1) + \lambda_1 \mathbf{E} \mathbf{A}_1 = 0 \quad (3.35)$$

$$-\mathbf{Z}_i^T \mathbf{R}_{new_i}^{-1} (\mathbf{Y}_i - \mathbf{Z}_i \mathbf{A}_{0i} - \mathbf{X}_i \mathbf{A}_1) + \mathbf{G}_*^{-1} \mathbf{A}_{i0} + \lambda_0 \mathbf{E} \mathbf{A}_{0i} = 0 \quad (3.36)$$

for $i = 1, 2, \dots, n$.

From 3.35, we get

$$\left(\sum_{i=1}^n \mathbf{X}_i^T \mathbf{R}_{new_i}^{-1} \mathbf{X}_i + \lambda_1 \mathbf{E} \right) \mathbf{A}_1 = \sum_{i=1}^n \mathbf{X}_i^T \mathbf{R}_{new_i}^{-1} (\mathbf{Y}_i - \mathbf{Z}_i \mathbf{A}_{0i}). \quad (3.37)$$

From 3.36, we get

$$(\mathbf{Z}_i^T \mathbf{R}_{new_i}^{-1} \mathbf{Z}_i + \mathbf{G}_*^{-1} + \lambda_0 \mathbf{E}) \mathbf{A}_{0i} = \mathbf{Z}_i^T \mathbf{R}_{new_i}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \mathbf{A}_1) \quad (3.38)$$

and

$$\mathbf{A}_{0i} = (\mathbf{Z}_i^T \mathbf{R}_{new_i}^{-1} \mathbf{Z}_i + \mathbf{D}^{-1})^{-1} \mathbf{Z}_i^T \mathbf{R}_{new_i}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \mathbf{A}_1) \quad (3.39)$$

where $\mathbf{D} = (\mathbf{G}_*^{-1} + \lambda_0 \mathbf{E})^{-1}$. Note that

$$\begin{aligned} (\mathbf{Z}_i^T \mathbf{R}_{new_i}^{-1} \mathbf{Z}_i + \mathbf{D}^{-1})^{-1} &= \mathbf{D} - \mathbf{D} \mathbf{Z}_i^T \mathbf{R}_{new_i}^{-1} (\mathbf{R}_{new_i}^{-1} + \mathbf{R}_{new_i}^{-1} \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T \mathbf{R}_{new_i}^{-1})^{-1} \mathbf{R}_{new_i}^{-1} \mathbf{Z}_i \mathbf{D} \\ &= \mathbf{D} - \mathbf{D} \mathbf{Z}_i^T \left(\mathbf{R}_{new_i} (\mathbf{R}_{new_i}^{-1} + \mathbf{R}_{new_i}^{-1} \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T \mathbf{R}_{new_i}^{-1}) \mathbf{R}_{new_i} \right)^{-1} \mathbf{Z}_i \mathbf{D} \\ &= \mathbf{D} - \mathbf{D} \mathbf{Z}_i^T (\mathbf{R}_{new_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T)^{-1} \mathbf{Z}_i \mathbf{D} \\ &= \mathbf{D} - \mathbf{D} \mathbf{Z}_i^T \mathbf{V}_{new}^{-1} \mathbf{Z}_i \mathbf{D} \end{aligned} \quad (3.40)$$

where $\mathbf{V}_{new} = \mathbf{Z} \mathbf{D} \mathbf{Z}^T + \mathbf{R}_{new}^{-1}$.

Then, equation 3.39 becomes

$$\begin{aligned}
\mathbf{A}_{0i} &= (\mathbf{D} - \mathbf{DZ}_i^T \mathbf{V}_{new}^{-1} \mathbf{Z}_i \mathbf{D}) \mathbf{Z}_i^T \mathbf{R}_{new_i}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \mathbf{A}_1) \\
&= \mathbf{DZ}_i^T \mathbf{R}_{new_i}^{-1} - \mathbf{DZ}_i^T \mathbf{V}_{new}^{-1} \mathbf{Z}_i \mathbf{DZ}_i^T \mathbf{R}_{new_i}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \mathbf{A}_1) \\
&= \mathbf{DZ}_i^T (\mathbf{I}_{n_i} - \mathbf{V}_{new}^{-1} \mathbf{Z}_i \mathbf{DZ}_i^T) \mathbf{R}_{new_i}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \mathbf{A}_1) \\
&= \mathbf{DZ}_i^T \mathbf{V}_{new}^{-1} \mathbf{R}_{new_i} \mathbf{R}_{new_i}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \mathbf{A}_1) \\
&= \mathbf{DZ}_i^T \mathbf{V}_{new}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \mathbf{A}_1)
\end{aligned}$$

Likewise, equation 3.35 becomes

$$-\sum_{i=1}^n \mathbf{X}_i^T \mathbf{R}_{new_i}^{-1} \left(\mathbf{Y}_i - \mathbf{Z}_i \mathbf{DZ}_i^T \mathbf{V}_{new}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \mathbf{A}_1) - \mathbf{X}_i \mathbf{A}_1 \right) + \lambda_1 \mathbf{E} \mathbf{A}_1 = 0, \quad (3.41)$$

that is,

$$\begin{aligned}
&-\sum_{i=1}^n \mathbf{X}_i^T \mathbf{R}_{new_i}^{-1} \left((\mathbf{I}_{n_i} - \mathbf{Z}_i \mathbf{DZ}_i^T \mathbf{V}_{new}^{-1}) \mathbf{Y}_i - (\mathbf{I}_{n_i} - \mathbf{Z}_i \mathbf{DZ}_i^T \mathbf{V}_{new}^{-1}) \mathbf{X}_i \mathbf{A}_1 \right) + \lambda_1 \mathbf{E} \mathbf{A}_1 = 0 \\
&\iff -\sum_{i=1}^n \mathbf{X}_i^T \mathbf{R}_{new_i}^{-1} \left((\mathbf{R}_{new_i} \mathbf{V}_{new}^{-1}) \mathbf{Y}_i - (\mathbf{R}_{new_i} \mathbf{V}_{new}^{-1}) \mathbf{X}_i \mathbf{A}_1 \right) + \lambda_1 \mathbf{E} \mathbf{A}_1 = 0 \\
&\iff -\sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_{new_i}^{-1} \left(\mathbf{Y}_i - \mathbf{X}_i \mathbf{A}_1 \right) + \lambda_1 \mathbf{E} \mathbf{A}_1 = 0 \\
&\iff \left(\sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_{new_i}^{-1} \mathbf{X}_i + \lambda_1 \mathbf{E} \right) \mathbf{A}_1 = \sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_{new_i}^{-1} \mathbf{Y}_i
\end{aligned} \quad (3.42)$$

Thus,

$$\hat{\mathbf{A}}_{1w} = \left(\sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_{new_i}^{-1} \mathbf{X}_i + \lambda_1 \mathbf{E} \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_{new_i}^{-1} \mathbf{Y}_i \right), \quad (3.43)$$

which is equivalent to

$$\hat{\mathbf{A}}_1 = (\mathbf{X}^T \mathbf{V}_{new}^{-1} \mathbf{X} + \lambda_1 \mathbf{E})^{-1} \mathbf{X}^T \mathbf{V}_{new}^{-1} \mathbf{Y}. \quad (3.44)$$

Then, from equation 3.41,

$$\hat{\mathbf{A}}_0 = \mathbf{DZ}^T \mathbf{V}_{new}^{-1} (\mathbf{Y} - \mathbf{X}^T \hat{\mathbf{A}}_1) \quad (3.45)$$

where $\mathbf{D} = (\mathbf{G}_*^{-1} + \lambda_0 \mathbf{E})^{-1}$ is the same as unweighted version, and $\mathbf{V}_{new} = \mathbf{ZDZ}^T + \mathbf{W}^{-1/2} \mathbf{R} \mathbf{W}^{-1/2} = \mathbf{ZDZ}^T + \mathbf{R}_{new}^{-1}$ is modified from the unweighted version.

To estimate the standard errors, we first note that

$$\mathbf{Y}_w = \mathbf{W}^{1/2} \mathbf{Y} | \mathbf{A}_0, \mathbf{A}_1 \sim N(\mathbf{W}^{1/2} \mathbf{X} \mathbf{A}_1 + \mathbf{W}^{1/2} \mathbf{Z} \mathbf{A}_0, \mathbf{R}) \quad (3.46)$$

where \mathbf{A}_0 and \mathbf{A}_1 are independent, with priors:

$$\mathbf{A}_1 \sim N(\mathbf{0}, (\lambda_1 \mathbf{E})^{-1}), \text{ and } \mathbf{A}_0 \sim N(\mathbf{0}, \mathbf{D}).$$

For $\boldsymbol{\epsilon} = \mathbf{Y}_w - \mathbf{W}^{1/2} \mathbf{X} \mathbf{A}_1 - \mathbf{W}^{1/2} \mathbf{Z} \mathbf{A}_0$, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R})$.

Let $\boldsymbol{\epsilon}_* = \mathbf{W}^{-1/2} \boldsymbol{\epsilon}$, then $\boldsymbol{\epsilon}_* \sim N(\mathbf{0}, \mathbf{R}_{new})$, where $\mathbf{R}_{new} = \mathbf{W}^{-1/2} \mathbf{R} \mathbf{W}^{-1/2}$, and \mathbf{W} is the diagonal weight matrix. Then equation 3.46 is equivalent to

$$\mathbf{Y} | \mathbf{A}_0, \mathbf{A}_1 \sim N(\mathbf{X} \mathbf{A}_1 + \mathbf{Z} \mathbf{A}_0, \mathbf{R}_{new}) \quad (3.47)$$

with the same prior distributions.

Turning to $\mathbf{Y}_* = \mathbf{Y} - \mathbf{X} \mathbf{A}_1 = \mathbf{Z} \mathbf{A}_0 + \boldsymbol{\epsilon}_*$, we can conclude that \mathbf{Y}_* follows a normal distribution with mean $\mathbf{0}$, and

$$\text{Cov}(\mathbf{Y}_*, \mathbf{A}_0) = \text{Cov}(\mathbf{Z} \mathbf{A}_0 + \boldsymbol{\epsilon}_*, \mathbf{A}_0) = \mathbf{W}^{1/2} \mathbf{Z} \text{Var}(\mathbf{A}_0),$$

$$\text{Cov}(\mathbf{Y}_*, \boldsymbol{\epsilon}_*) = \text{Cov}(\mathbf{Z} \mathbf{A}_0 + \boldsymbol{\epsilon}_*, \boldsymbol{\epsilon}_*) = \mathbf{R}_{new},$$

$$\text{Cov}(\mathbf{Y}_*, \mathbf{Y}_*) = \mathbf{Z} \text{Var}(\mathbf{A}_0) \mathbf{Z}^T + \text{Var}(\boldsymbol{\epsilon}_*) = \mathbf{V}_{new},$$

where $\mathbf{D} = (\mathbf{G}_*^{-1} + \lambda_0 \mathbf{E})^{-1}$, $\mathbf{V}_{new} = \mathbf{ZDZ}^T + \mathbf{W}^{1/2} \mathbf{R}^{-1} \mathbf{W}^{1/2}$.

Thus,

$$\begin{bmatrix} \mathbf{A}_0 \\ \boldsymbol{\epsilon}_* \\ \mathbf{Y}_* \end{bmatrix} \sim N \left[\mathbf{0}, \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{DZ}^T \\ \mathbf{0} & \mathbf{R}_{new} & \mathbf{R}_{new} \\ \mathbf{ZD} & \mathbf{R}_{new} & \mathbf{V}_{new} \end{pmatrix} \right]. \quad (3.48)$$

For the conditional distribution $\mathbf{A}_0|\mathbf{Y}_*$,

$$\mathbb{E}(\mathbf{A}_0|\mathbf{Y}_*) = 0 + \mathbf{D}\mathbf{Z}^T\mathbf{V}_{new}^{-1}\mathbf{Y}_* = \mathbf{D}\mathbf{Z}^T\mathbf{V}_{new}^{-1}(\mathbf{Y} - \mathbf{X}^T\hat{\mathbf{A}}_1) = \hat{\mathbf{A}}_0,$$

and the variance

$$\text{Var}(\mathbf{A}_0|\mathbf{Y}_*) = \mathbf{D} - \mathbf{D}\mathbf{Z}^T\mathbf{V}_{new}^{-1}\mathbf{Z}\mathbf{D}.$$

Similarly, for the conditional distribution $\boldsymbol{\epsilon}_*|\mathbf{Y}_*$,

$$\mathbb{E}(\boldsymbol{\epsilon}_*|\mathbf{Y}_*) = 0 + \mathbf{R}_{new}\mathbf{V}_{new}^{-1}\mathbf{Y}_* = \mathbf{R}_{new}\mathbf{V}_{new}^{-1}(\mathbf{Y} - \mathbf{X}^T\hat{\mathbf{A}}_1),$$

and

$$\text{Var}(\boldsymbol{\epsilon}_*|\mathbf{Y}_*) = \mathbf{R}_{new} - \mathbf{R}_{new}\mathbf{V}_{new}^{-1}\mathbf{R}_{new}.$$

We rewrite the conditional distributions $\mathbf{A}_0|\mathbf{Y}_*$ and $\boldsymbol{\epsilon}_*|\mathbf{Y}_*$ for each individual and obtain

$$\begin{aligned} \mathbf{A}_{0i}|\mathbf{Y}, \mathbf{A}_1 = \mathbf{A}_{0i}|\mathbf{Y}_* &\sim N\left(\mathbf{D}_i\mathbf{Z}_i^T\mathbf{V}_{new_i}^{-1}(\mathbf{Y}_i - \mathbf{X}_i^T\hat{\mathbf{A}}_{1i}), \mathbf{D}_i - \mathbf{D}_i\mathbf{Z}_i^T\mathbf{V}_{new_i}^{-1}\mathbf{Z}_i\mathbf{D}_i\right) \\ \boldsymbol{\epsilon}_{*i}|\mathbf{Y}, \mathbf{A}_1 = \boldsymbol{\epsilon}_{*i}|\mathbf{Y}_* &\sim N\left(\mathbf{R}_{new_i}\mathbf{V}_{new_i}^{-1}(\mathbf{Y}_i - \mathbf{X}_i^T\hat{\mathbf{A}}_{1i}), \mathbf{R}_{new_i} - \mathbf{R}_{new_i}\mathbf{V}_{new_i}^{-1}\mathbf{R}_{new_i}\right) \end{aligned} \quad (3.49)$$

where $\mathbf{D} = (\mathbf{G}_*^{-1} + \lambda_0\mathbf{E})^{-1}$, $\mathbf{V}_{new} = \mathbf{Z}\mathbf{D}\mathbf{Z}^T + \mathbf{W}^{1/2}\mathbf{R}^{-1}\mathbf{W}^{1/2}$.

Thus,

$$\text{Var}(\boldsymbol{\epsilon}_{*i}|\mathbf{Y}, \mathbf{A}_1) = \mathbf{R}_{new_i} - \mathbf{R}_{new_i}\mathbf{V}_{new_i}^{-1}\mathbf{R}_{new_i} = \sigma_R^2\mathbf{W}_i(\mathbf{I}_{n_i} - \sigma_R^2\mathbf{W}_i\mathbf{V}_i^{-1}),$$

and

$$\begin{aligned} \mathbb{E}\left(\boldsymbol{\epsilon}_{*i}^T\boldsymbol{\epsilon}_{*i}|\mathbf{Y}, \mathbf{A}_1 = \hat{\mathbf{A}}_1\right) &= \text{tr}\left[\mathbb{E}(\boldsymbol{\epsilon}_{*i}^T\boldsymbol{\epsilon}_{*i}|\mathbf{Y}, \mathbf{A}_1 = \hat{\mathbf{A}}_1)\right] \\ &= \text{tr}\left[\mathbb{E}(\boldsymbol{\epsilon}_{*i}|\mathbf{Y}, \hat{\mathbf{A}}_1)\mathbb{E}(\boldsymbol{\epsilon}_{*i}|\mathbf{Y}, \hat{\mathbf{A}}_1)^T + \text{Var}(\boldsymbol{\epsilon}_{*i}|\mathbf{Y}, \hat{\mathbf{A}}_1)\right] \\ &= \text{tr}\left[\hat{\boldsymbol{\epsilon}}_{*i}\hat{\boldsymbol{\epsilon}}_{*i}^T + \sigma_R^2\mathbf{W}_i(\mathbf{I}_{n_i} - \sigma_R^2\mathbf{W}_i\mathbf{V}_i^{-1})\right] \\ &= \hat{\boldsymbol{\epsilon}}_{*i}\hat{\boldsymbol{\epsilon}}_{*i}^T + \sigma_R^2\left(\sum_{j=1}^{m_i} w_{ij} - \sigma_R^2\text{tr}(\mathbf{W}_i\mathbf{V}_i^{-1}\mathbf{W}_i)\right). \end{aligned} \quad (3.50)$$

Note that equation 3.50 holds because $\boldsymbol{\epsilon}_{*i}^T \boldsymbol{\epsilon}_{*i} | \mathbf{Y}, \mathbf{A}_1 = \hat{\mathbf{A}}_1$ is a scalar.

Similarly,

$$\text{Var}(\mathbf{A}_{0i} | \mathbf{Y}, \mathbf{A}_1) = \mathbf{D}_i - \mathbf{D}_i \mathbf{Z}_i^T \mathbf{V}_{new_i}^{-1} \mathbf{Z}_i \mathbf{D}_i,$$

and

$$\begin{aligned} \mathbb{E} \left(\mathbf{A}_{0i}^T \mathbf{A}_{0i} | \mathbf{Y}, \mathbf{A}_1 = \hat{\mathbf{A}}_1 \right) &= \text{tr} \left[\mathbb{E}(\mathbf{A}_{0i}^T \mathbf{A}_{0i} | \mathbf{Y}, \mathbf{A}_1 = \hat{\mathbf{A}}_1) \right] \\ &= \text{tr} \left[E(\mathbf{A}_{0i} | \mathbf{Y}, \hat{\mathbf{A}}_1) E(\mathbf{A}_{0i} | \mathbf{Y}, \hat{\mathbf{A}}_1)^T + \text{Var}(\mathbf{A}_{0i} | \mathbf{Y}, \hat{\mathbf{A}}_1) \right] \\ &= \hat{\mathbf{A}}_{0i} \hat{\mathbf{A}}_{0i}^T + \mathbf{D}_i - \mathbf{D}_i \mathbf{Z}_i^T \mathbf{V}_{new_i}^{-1} \mathbf{Z}_i \mathbf{D}_i \end{aligned} \quad (3.51)$$

Thus, the maximum likelihood estimators (MLE) of the variance and covariance matrices can be obtained by

$$\begin{aligned} \mathbb{E}(\hat{\sigma}_R^2 | \mathbf{Y}) &= \frac{1}{N} \sum_{i=1}^n \left(\hat{\boldsymbol{\epsilon}}_{*i} \hat{\boldsymbol{\epsilon}}_{*i}^T + \sigma_R^2 \left(\sum_{j=1}^{m_i} w_{ij} - \sigma_R^2 \text{tr}(\mathbf{W}_i \mathbf{V}_i^{-1} \mathbf{W}_i) \right) \right), \text{ and} \\ \mathbb{E}(\hat{\mathbf{D}} | \mathbf{Y}) &= \frac{1}{n} \sum_{i=1}^n \left(\hat{\mathbf{A}}_{0i} \hat{\mathbf{A}}_{0i}^T + \mathbf{D}_i - \mathbf{D}_i \mathbf{Z}_i^T \mathbf{V}_{new_i}^{-1} \mathbf{Z}_i \mathbf{D}_i \right), \end{aligned} \quad (3.52)$$

where $N = \sum_{i=1}^n m_i$ is the total number of observations, n is the total number of subjects, $\hat{\boldsymbol{\epsilon}}_{*i} = \mathbf{R}_{new_i} \mathbf{V}_{new_i}^{-1} (\mathbf{Y}_i - \mathbf{X}_i^T \hat{\mathbf{A}}_{1i})$, \mathbf{A}_0 , \mathbf{A}_1 are given by equations 3.45 and 3.44, respectively. Then we can follow the EM-algorithm outlined by Zhang and Wu, using equation 3.52 for calculation.

Note that for MAR cases, the probability of non-missingness is random, thus, to estimate the standard errors, the above method is no longer appropriate. To estimate the standard errors for MAR cases is not the focus here but can be explored in the future.

4.0 APPLICATION

We applied our gap time model to data obtained from the Shiffman smoking study [28]. This EMA study contained a total of 212 ITS subjects. Because the occurrence of bouts is the focus of our analysis, we limited our subject cohort to those who had such behavior. As outlined in the procedure and method sections, gap times were considered as the outcomes of interest. We “started” everyone on Monday. Furthermore, we assumed no censoring, which meant days in the monitoring period were consecutive and time was continuous. No partial days were included unless at the start or end of the study. After data elimination, the final data set contained 102 ITS subjects. We used MATLAB code from Wu and Zhang [36], and applied the code to our data while ignore missing data.

For the purpose of illustration, we only included one covariate in the model, noted as a variable called “Positive Affect (PA) Score”. The PA scores were one of the scales that summarized sixteen mood adjectives (able to focus, active, angry/frustrated, bored, calm/relaxed, difficulty concentrating, enthusiastic, happy, irritable, miserable, nervous/tense, quiet/sleepy, restless, sad, overall mood and overall arousal level). These continuous variables are summarized into four factors and the PA subscale is one of the four. The range was between 0 and 100. Subjects had positive mood when the values of the corresponding variables were high. Clinically, it was surmised that this variable was associated with individuals’ desire to smoke. This because it was thought likely that ITS would smoke intensively in positive social situations such as parties. The observations were collected via assessed smoking

events only. Clinically, we hypothesized that the effect of the variable PA to the cumulative log gap time might be different for weekdays and weekends.

The mean and median of the gap times across the 102 subjects were 7.28 (SD=14.68) and 2.07 hours, respectively. The distribution of the gap times was highly skewed. We illustrated our graphical methods in Figures 4 - 4. Figure 4 shows the graph of the smoking behavior for one subject. In all three panels, the x-axes represent the days in the study (starting from Monday). The first panel displays the stochastic process of the occurrence of smoking episodes. Here, each vertical line represents a cigarette. Panels (b) and (c) give other ways of displaying the event process by interpolating consecutive observations for an individual. In both (b) and (c), the gap time was defined as the time to the next smoking event in hours. We plotted $\log\left(\frac{T_i(t_{ij})}{D}\right)$ and $\sum_{t_{ij} \leq t} \log\left(\frac{T_i(t_{ij})}{D}\right)$ versus the time index t , respectively. The bout periods are denoted with “×” symbols. In (b), bouts can be easily identified whenever $\log\left(\frac{T_i(t_{ij})}{D}\right) < 0$. In (c), events in bouts have a negative slope in the interpolated plots. One advantage of using the cumulative log transformed gap time is that it can qualitatively and quantitatively characterize bouts without loss of the global perspective. Thus, it is more robust and stable for modeling, and less sensitive to the situational variables, as compared to $\log\left(\frac{T_i(t_{ij})}{D}\right)$.

Figures 4 and 4 show graphs of log gap times, and cumulative log gap times, plotted over time for five subjects. Comparing the two plots, Figure 4 expresses the differences between the five subjects more clearly. As stated previously, the cumulative log gap times follow a sawtooth wave. Moreover, different individuals are distinct with each other when looking at the distribution of cumulative log gap times over time (Figure 4), whereas in a plot of instantaneous gap time versus time, the patterns are not obvious (Figure 4).

Figure 4 shows the results of parameter estimates of the time-varying coefficient model described in equation (3.1), assuming that no missingness is present. The first panel shows the time-varying intercept and the second plot shows the effect of the covariate, that is, the PA variable. As outlined in equation (3.1), the intercept here is a random effect

with mean $\mu(t)$; thus, panel (a) is a plot of $\mu(t)$. The slope of $\mu(t)$ is positive over time because the log gap time is cumulative. In both panels, the dashed lines represent the 95% confidence intervals. In the second panel, the blue vertical lines highlight the periods between the beginning of Thursday and the end of Saturday which we refer to as “weekends”. The population model does not seem to support our hypothesis of having a consistent weekly relationship between positive mood and smoking over the entire cohort. Furthermore, the PA coefficient does not seem to vary significantly over time as indicated by the lack of the point-wise 95% confidence intervals including 0 through out nearly the entire time period. The lack of a weekend pattern in the cohort could be due to wide across subject variability. Had such a pattern of the PA effect existed, our method could detect it without making parametric assumptions.

When missing by design was present, Figure 4 shows the results of parameter estimates of the corresponding time-varying coefficient model. Because the weights in the smoking data were not large, the results do not differ much when applying weights or not. Figure 4 gives the fitted patterns of six individuals.

In general, it is difficult to interpret a time varying coefficient model. It is important to phase one’s hypothesis in terms of why s/he wants to fit such a model. In our particular case, we hypothesized that a weekly pattern was present as bouts were more frequent on weekends. When the slope is positive or near 0, the model implies that the smoking pattern is not in a bout at that time. On the other hand, when the slope is negative, the model implies that cigarettes seem to be consumed in a bout. In addition, the absolute value of the slopes reflect how intensive (when the slope is negative) or how widespread (when the slope is positive) the cigarettes are consumed over an appropriate interval of time. If our underlying assumption of the weekly pattern is met, we might observe that the slopes become negative around weekends and become to positive or near 0 during weekdays.

Due to the EMA study design, one may have the opportunity to assess the temporal patterns at both population and individual levels. In this dissertation, the graphical

display can give an empirical way of looking at individuals' overall pattern together with the time-clusters. The time varying coefficient model also gives an opportunity to evaluate the relationship between situational variables and the cumulative gap times, when the data are complete or missing by design. Our method is not limited to Shiffman smoking data, but can be applied to any situations when trying to identify patterns in health event outcomes.

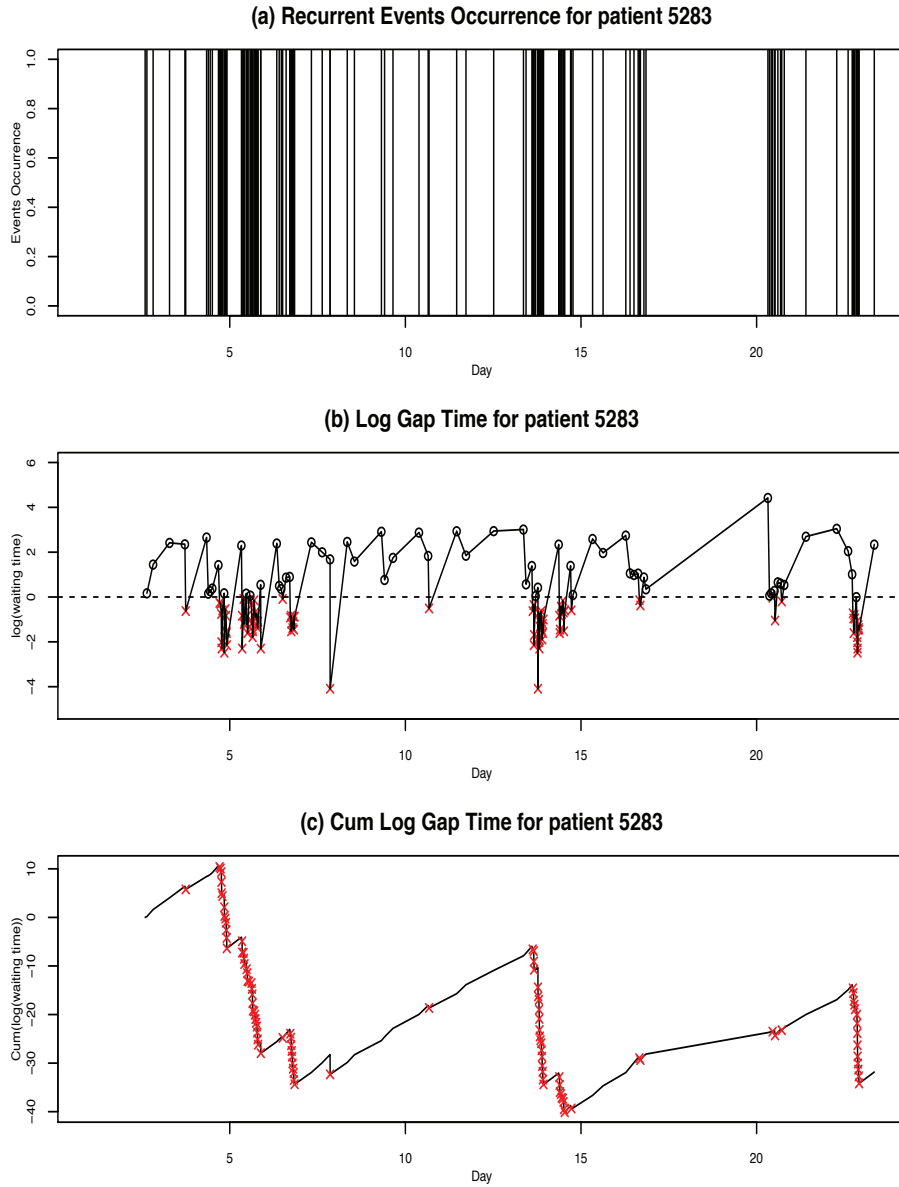


Figure 4.1: Event Pattern of One Subject in the Presence of Bouts

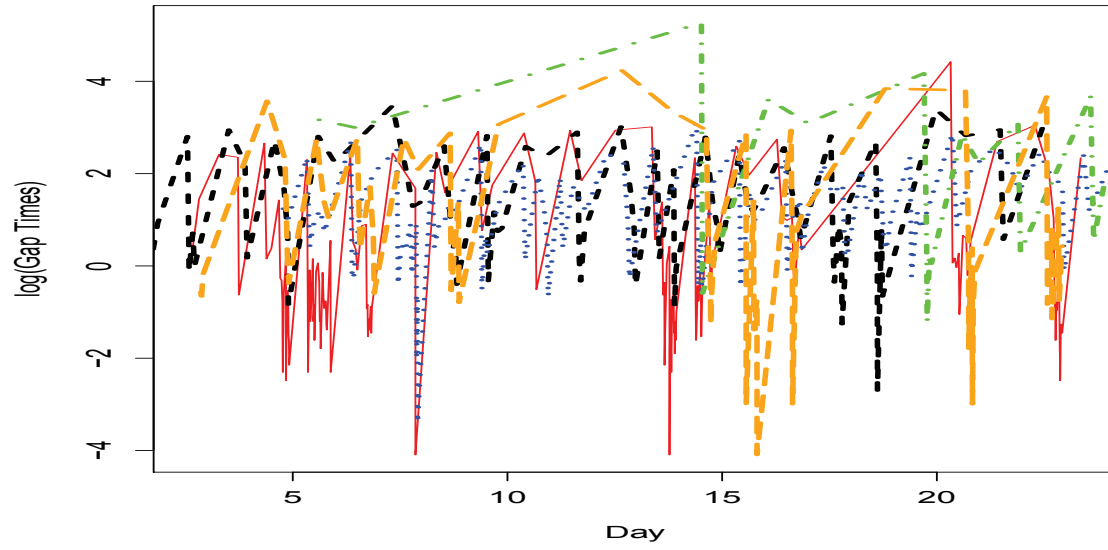


Figure 4.2: Log Gap Times

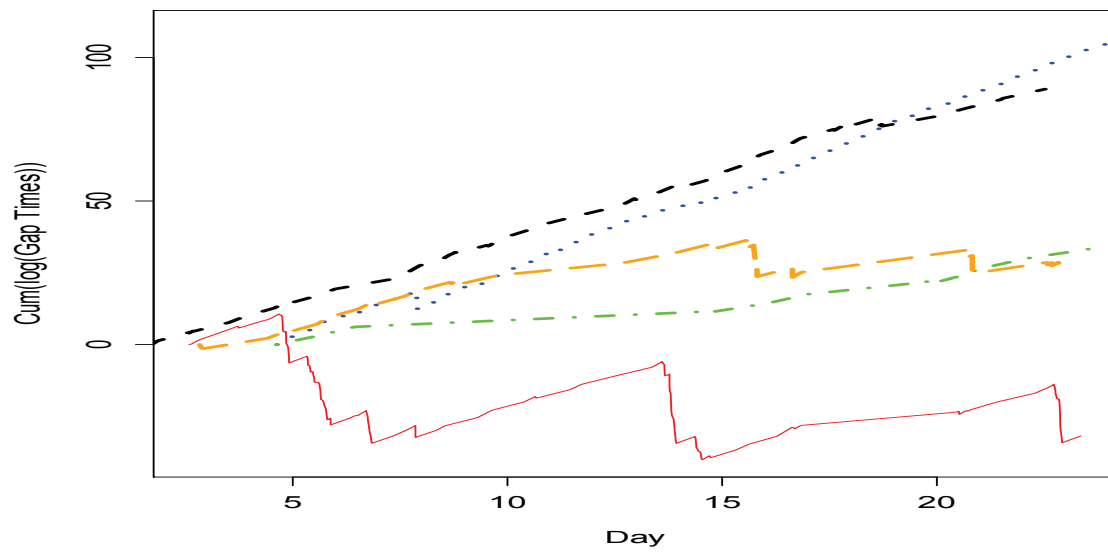


Figure 4.3: Cumulative Log Gap Times

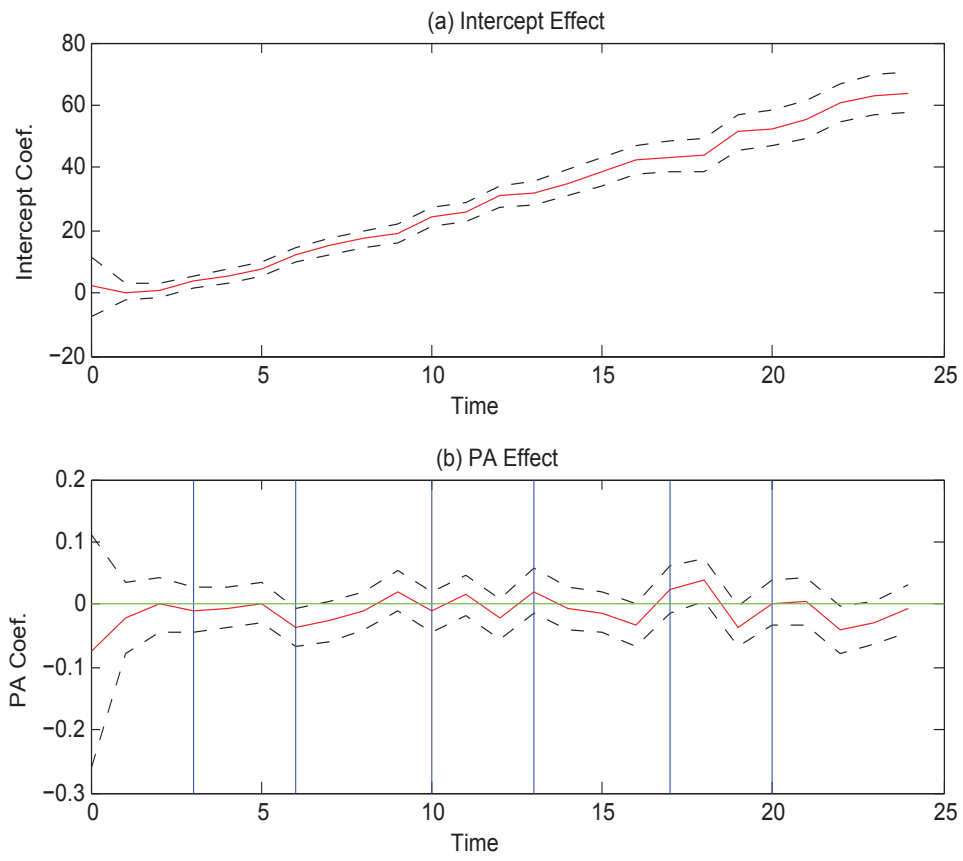


Figure 4.4: Time-varying Coefficients

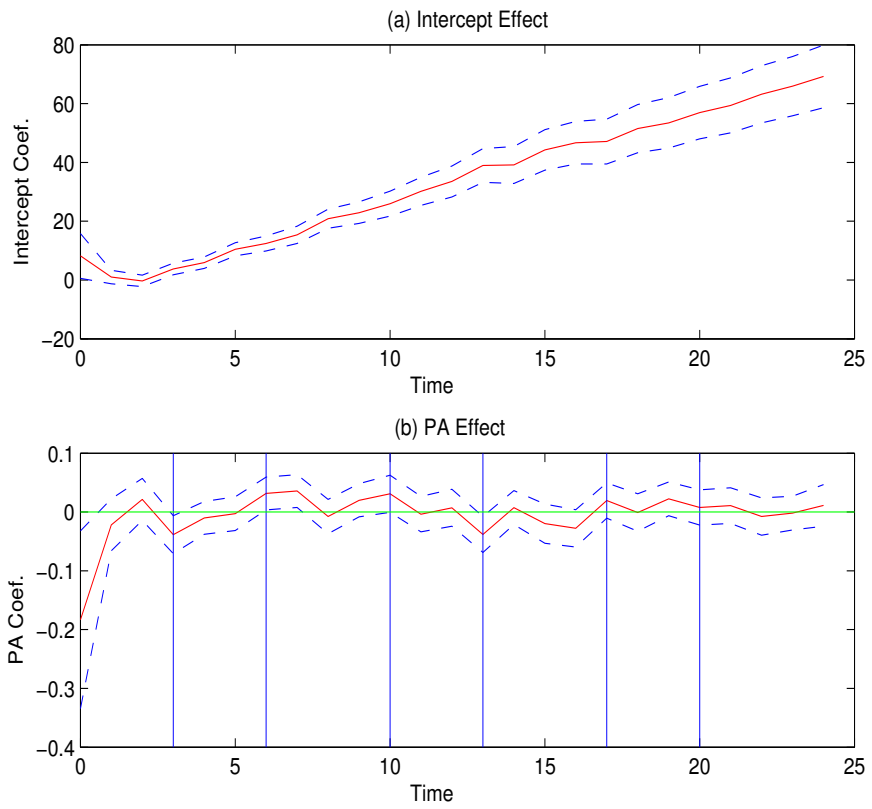


Figure 4.5: Time-varying Coefficients

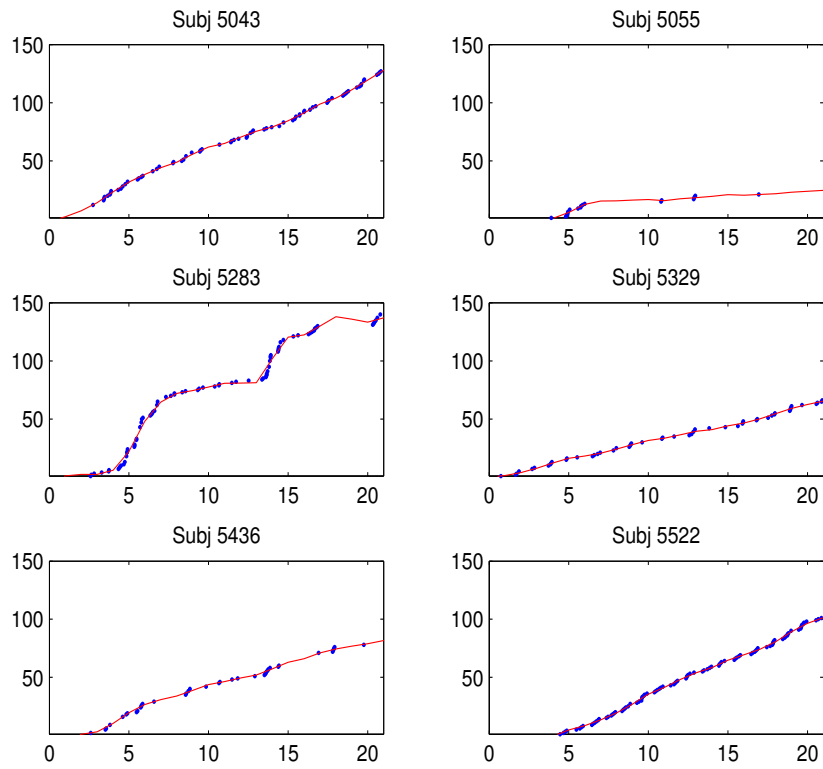


Figure 4.6: Fitted Curves for Individuals

5.0 SIMULATION STUDIES

The purpose of our simulation study was to evaluate how the cumulative log gap time model performed. This evaluation sought to address two questions:

1. Can the model pick up the functional form of a complex time varying coefficient?
2. How does the model perform with missing data?

5.1 PICKING UP THE FUNCTIONAL FORMS OF THE TIME VARYING COEFFICIENT

Here, we specified a function for the coefficient of interest, simulate the data, fit the model, and compare the fitted coefficient with the true coefficient. In this simulation, we used five scenarios with different conditions. We used the simulation studies to explore how the models performed under different situations. We only simulated one covariate. The five scenarios are:

- Scenario 1: the slope of the covariate follows a sawtooth wave of time with no subject variation; the time varying intercept is linear function of time with small subject variation;
- Scenario 2: the time varying slope follows the same sawtooth wave of time, with small subject variation on the vertical scale ($\text{var}=0.1$); the intercept remains the same criteria;

- Scenario 3: the time varying slope follows the same sawtooth wave of time, with small subject variation on where periods start (var=0.1), same period length; the intercept remains the same;
- Scenario 4: the time varying slope is a sawtooth wave, with small subject variation on the period length (7+- var = 1 days); the intercept remains the same;
- Scenario 5: the slope of the covariate follows a sawtooth wave of time with no subject variation; the time varying intercept is linear function of time with small subject variation; the within subject error is t distribution($0.2 \times t(\text{df}=10)$);

Simulations were repeated for different numbers of subjects with variation on different parameters of the time varying coefficients. Robustness was also assessed by applying different within subject error distributions (t-distribution).

Assumptions included:

- The unit of time was an hour
- The total numbers of subjects were denoted by n , $n = 10, 20, 50, 100, 200$.
- The one covariate of interest was distributed as $X \sim N(50, 10)$
- The within subject error variance was small $\epsilon_{ij} \sim N(0, 0.1)$.
- The random intercepts were linear functions over time, with the mean function as $a_{00} + a_{01} \times t = 0.1 \times t$, where a_{00} and a_{01} may vary by subjects $a_{00} \sim N(0, 0.1)$ and $a_{01} \sim N(0.1, 0.01)$.
- The time varying coefficient of X was: $\alpha_1(t) = a_{10} + a_{11} * (t/\text{period} - \text{floor}(t/\text{period}) - 0.5)$, where $a_{10} = 0$, $a_{11} = 0.25$, period = 24×7 hours.
- The baseline gap time was distributed as $T \sim \exp(6)$.
- We setted $d = 4.8$ hours.

A single simulation is represented in figure 5.1 and summarized into the following steps:

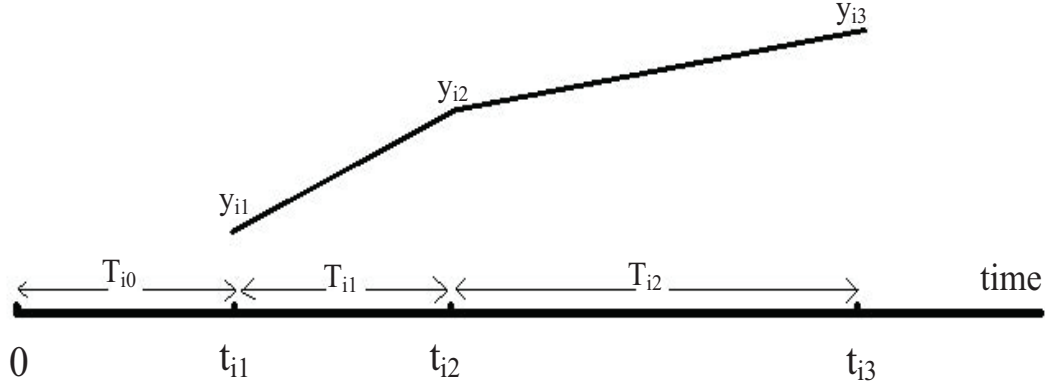


Figure 5.1: Simulation Steps Summary

Step 1. For a single subject i , simulate the parameters of the dynamic random intercept $a_{00i} \sim N(0,0.1)$ and $a_{01i} \sim N(0,0.01)$.

Step 2. Simulate the initial time to next event variable $T_{i0} \sim \exp(6)$, the first value of the outcome is then $y_{i0} = \log(T_{i0})$

Step 3. At time $t_{i1} = T_{i0}$, calculate the subject's time varying intercept as $\alpha_{0i}(t_{i1}) = a_{00i} + a_{01i} \times t_{i1}$.

Step 4. Also at the same time t_{i1} , calculate the time varying coefficient from the prespecified function and obtain $\alpha_1(t_{i1})$, simulate the corresponding covariate $x_{i1} \sim N(50, 10)$, and simulate the error term $\epsilon_{i1} \sim N(0,0.1)$

Step 5. Calculate $y_{i1} = \alpha_{0i}(t_{i1}) + \alpha_1(t_{i1})x_{i1} + \epsilon_{i1}$ be the second outcome

Step 6. Calculate the corresponding second gap time $T_{i1} = \exp(y_{i1} - y_{i0})$ and time when the second event occurs $t_{i2} = t_{i1} + T_{i1}$.

Step 7. Repeat Steps 3 – 6, while replacing the subscripts i_1 with i_{j-1} and i_2 with i_j for $j = 3, 4, \dots, m_i$, where m_i is the number of total observations for subject i .

Step 7. Repeat Steps 1 – 6 for subject $i = 1, 2, \dots, n$, where n is the total number of subjects

To compare between the fitted and the true time varying coefficient, we calculated the coefficient values at each knot, calculated the squared difference between the coefficient values, took the square root of the mean of the squared differences over the number of knots, and took the means across simulation iterations. We refer this comparison statistics as the rooted mean squared differences (RMS). The RMS values were used here to indicate how fitted values are different from the true values. Thus, small RMS values may imply that the model is able to pick up the pre-specified functional form. However, it is still necessary to compare the fitted and true curves on a plot (Figure 5.1). For each scenario, we simulated the data 100 times. The reason for such a small number of replications was that each scenario took many hours to compute.

Table 5.1: Simulation Results for Time Varying Coefficient Models

	RMS				
N subjects	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
10	0.1292	0.1317	0.1779	0.1534	0.1292
20	0.1230	0.1240	0.1274	0.1477	0.1258
50	0.1232	0.1225	0.1260	0.1448	0.1229
100	0.1227	0.1218	0.1237	0.1444	0.1227
200	0.1226	0.1221	0.1214	0.1422	0.1229

The model was able to pick the functional form when both the subject level and within subject variations were not large. We also simulated the data with larger between subject variances. We applied this increase to Scenario 2 with larger variation among subjects on the vertical scale (var=1) of the sawtooth slope. The fitted slope was flat even when the number of subject is 200. This might explain why no pattern was observed in Shiffman’s

smoking data in the application section. Sample size and power problems may also be worth further investigation.

Figure 5.1 - Figure 5.1 show examples on how sample size can affect the fit of the model. When the sample size was greater than 50, the model captured the discontinuity of the true coefficient function. This is a nice feature because a spline model usually does not have such property. A disadvantage of the model was found when we compared the five figures. That is, the model tended to “shift” the peak forward. This tendency did not seem to go away as the number of subject increased. Increasing the number of observations per subject may reduce such tendency.

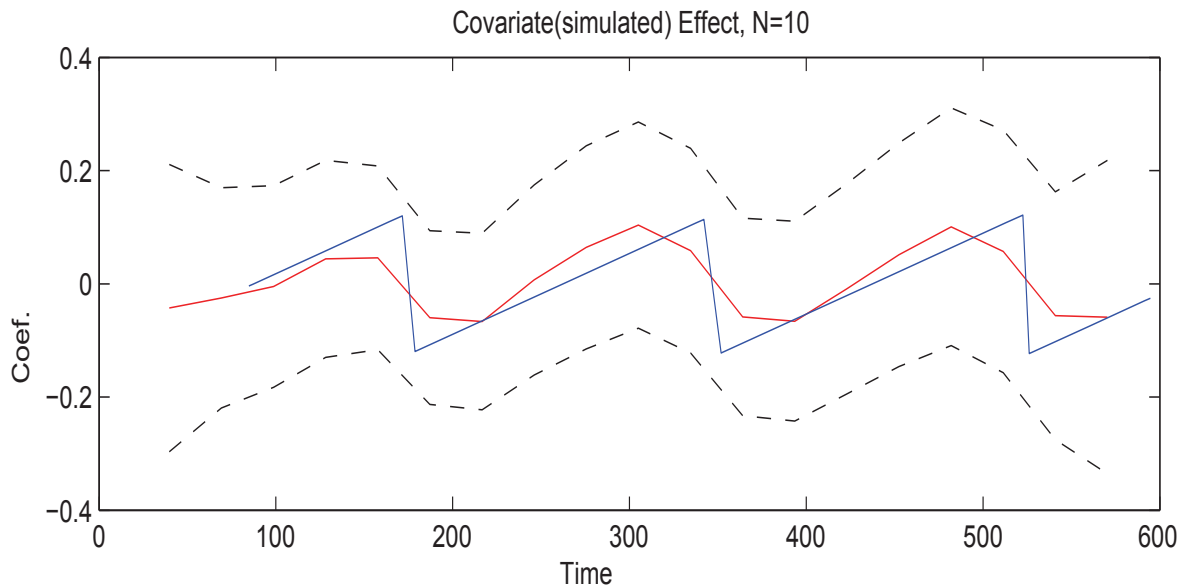


Figure 5.2: Fitted vs. True Coefficient (N=10) for One Simulated Dataset

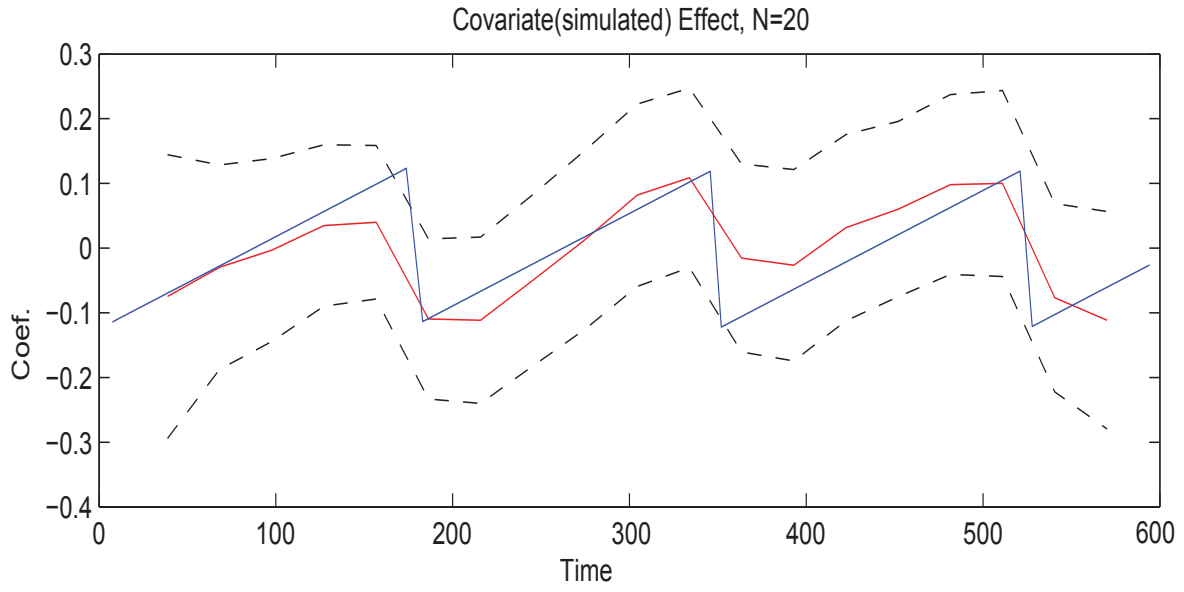


Figure 5.3: Fitted vs. True Coefficient (N=20) for One Simulated Dataset

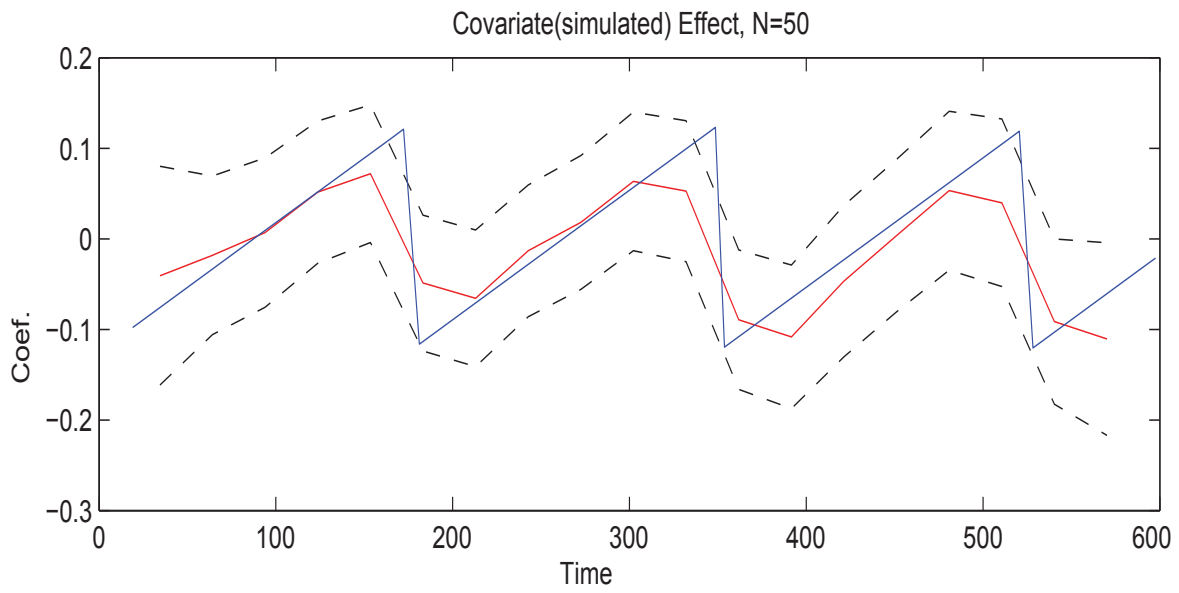


Figure 5.4: Fitted vs. True Coefficient (N=50) for One Simulated Dataset

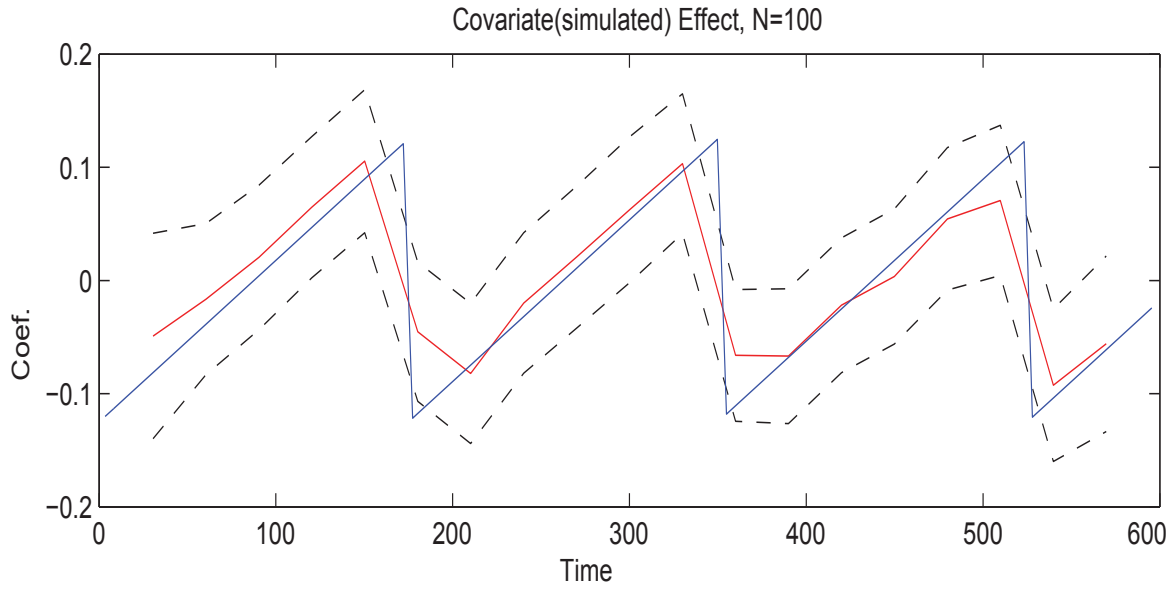


Figure 5.5: Fitted vs. True Coefficient (N=100) for One Simulated Dataset

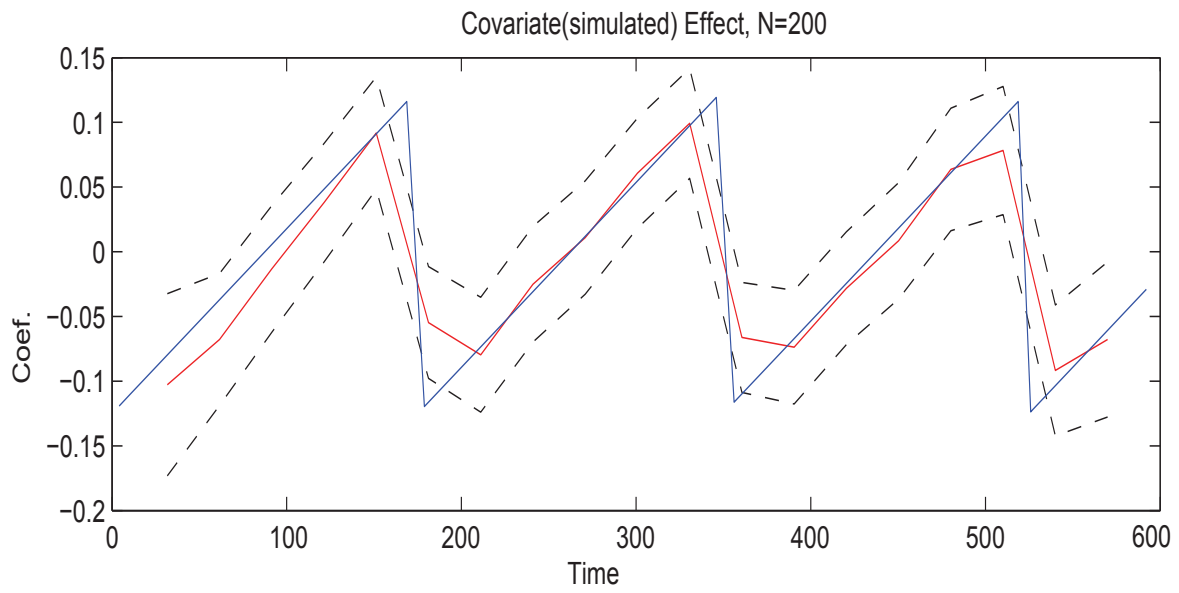


Figure 5.6: Fitted vs. True Coefficient (N=200) for One Simulated Dataset

5.2 SIMULATION STUDY WITH MISSING DATA

The complete data sets were simulated in the first scenario of previous simulation procedure.

We assumed that the probabilities of assessment followed two systematic strategies:

- The first observation of each day was always observed, the probability of assessment reduced to 0.5 for the 2nd - 4th observation, and 0.25 for the rest.
- Probability of assessment was equal to a known pre-specified random number between $[0, 1]$.

Table 5.2: Simulation Results for Time Varying Coefficient Models with Missing Data

N subjects	RMS		
	CompleteData	Missing 1	Missing 2
10	0.1292	0.1416	0.1308
20	0.1230	0.1348	0.1262
50	0.1232	0.1266	0.1244
100	0.1227	0.1298	0.1219
200	0.1226	0.1241	0.1215

We can see that when the missing pattern followed a systematic distribution, the RMS values between the fitted value and the true value were larger as compared to the full data. The RMS values reduced as the sample size increased.

6.0 DISCUSSION AND FUTURE WORK

The contributions of this dissertation are threefold. First, we provide a method for easily visualizing time clusters when examining recurrent event data. Second, we apply time varying coefficient models in a novel setting, that is, gap times associated with recurrent event data. Third, we extend the time varying coefficients model to handle missing by design (MBD) situation.

Our model allowed us to investigate a special phenomenon for our particular application to ITS individuals. In the Chapter 4, Figures 4-4 illustrated a graphical technique that allowed us to identify events that are in smoking “bouts” without losing the ability to characterize the overall pattern of all of the smoking events for a cohort of individuals. We then applied a time varying coefficients model as formulated by Wu [36] to the cumulative log transformed gap time data. Tan, et al. [32], first introduced these models as a way of characterizing continuous EMA data. Here, we focused on gap times and were particularly interested in the phenomenon where events were clustered in time. Because of the EMA design, information on temporal patterns could be explored. Accordingly, time varying models applied to this information could be used to identify temporal effects such as periodicity of covariates of interest.

In our model as applied to gap time data, there were several potential limitations worth further investigation. One limitation was that the results were dependent on the parameter, d . Thus, practically, it would be useful to perform with sensitivity analyses to

see how results change with different values of d . In addition, d can vary by subject, denoted as d_i . Clinically, one may choose d_i as the mean gap time of a given subject.

We can see that the use of d and d_i raise two questions. When we use d , i.e. the same value for all subjects, we explore how intensively subjects smoke based on a pre-specified value using experts' opinion of what the value of d should be. When covariates are included, an underlying assumption is that the effects of covariates, on the pre-specified magnitude of intensively smoking, are the same for all individuals. On the other hand, when d_i is used, we are trying to explore that how subjects intensively smoke as compared to their individual general smoking pace. When covariates are included, we assume that the covariates affect how intensively individuals smoke as compared to their own average smoking pattern.

As stated in the application section, we were interested in the weekly pattern of bouts. In this dissertation, we fixed the number of knots in the basis used in our model. The number is nineteen. However, this value can be changed according to a particular research interest. It is also of clinical interest to explore within day bouts. In the latter case, a large number of knots are needed.

We investigated the situations where missingness occurred due to the study design (missing by design: MBD). Further exploration can be done to adjust for missing at random (MAR). As shown in the formulas in Chapter 3, we can obtain the parameter estimators for MAR cases the same ways as for MBD cases, but cannot for the standard errors. Thus, further work is needed.

Another issue is that occasionally, the exact event times may not be known in practice. For instance, individuals may consume cigarettes overnight or in a meeting but they suspend the diary and do not report them in the next morning or after the meeting. Our model and analysis cannot handle these cases and they were excluded in the analysis. A third possible limitation is the subject random effects and within subject errors are assumed to be normally distributed. We made this assumption based on the presumption that the transformation of the gap times allowed us to approximate normality. The robustness of these

models to miss-specified error distributions could be further explored or other distributions more commonly used in gap time settings (e.g., gamma or Weibull distributions) may be considered.

Other topics of future interest may include:

1. To Explore how well this model performs with different numbers of knots;
2. To use two or more continuous covariates;
3. To use discrete or binary time-varying covariates;
4. To estimate the standard errors for missing at random;
5. To explore power and sample size for this model; and
6. To compare our model with existing stochastic models for characterizing patterns in gap times and associated covariates.

APPENDIX

MATLAB CODE FOR MISSING BY DESIGN DATA

To run the model, use the code:

```
[mbeta,mbsig]=vcpsmemiss(data2,spar,tfit,param,wght2);

%% First Function

function [mbeta,mbsig,vfit,hcrit,vhcrit,param]
=vcpsmemiss(data,spar,xfit,param,wght)
% add missing weight
% Varying Coefficient ME Models Using P-Splines
%  $y_i(t)=x_i(t)'*beta(t)+v_i(t)+e_i(t)$ ,  $i=1,2,\dots,n$ 
%
% Originally Designed by Jin-Ting Zhang, Stat & Applied Prob.,
% NUS, Singapore, 2005
% and Hulin, Wu, University of Rochester,
%
% Xiaoxue Li added a weight to handle missing by deisng (MBD)
% problem, 2014
%

[n,m]=size(data);
if m==3,
    X=ones(n,1);p=1;
else
    X=data(:,4:m);p=m-3;
end
subj=data(:,1); usubj=unique(subj);nsubj=length(usubj);

x=data(:,2);[ux,flag1,flag2]=unique(x); nux=length(ux);
y=data(:,3);
```

```

%Deal with weights, if no weights are assigned
if length(wght)==0, wght=eye(n); end

if nargin<4|length(param)==0,
    K=-1; dpoly=2; nkflag=1; Gflag=0; bflag=3;
elseif length(param)==1,
    K=param(1); dpoly=2; nkflag=1; Gflag=0; bflag=3;
elseif length(param)==2,
    K=param(1); dpoly=param(2); nkflag=1; Gflag=0; bflag=3;
elseif length(param)==3,
    K=param(1); dpoly=param(2); nkflag=param(3);
    Gflag=0; bflag=3;
elseif length(param)==4,
    K=param(1); dpoly=param(2); nkflag=param(3);
    Gflag=param(4); bflag=3;
elseif length(param)==5,
    K=param(1); dpoly=param(2); nkflag=param(3);
    Gflag=param(4); bflag=param(5);
end

if nargin<3|length(xfit)==0, xfit=ux; end

if K==-1,
    K=ksel(nux, dpoly);
    %K = 10;
end

if nargin<2|length(spar)<=p,
% find the range of spar
    Z0=ones(n,1);
    for ii=1:dpoly,
        Z0=[Z0,x.^ii];
    end
    if K==0,
        Z=Z0;
    else
        if nkflag==0,
            xmin=min(ux); xmax=max(ux);
            knotsq=xmin+(xmax-xmin)*[1:K]/(K+1);
        elseif nkflag==1,
            knotsq=prctile(x,[1:K]/(K+1)*100); % quantiles as knots
            knotsq=unique(knotsq);
            K=length(knotsq);
        end
    end
end

```

```

        knotsq=knotsq(:)'; % so that knots is a row vector
    end

    Z1=x*ones(1,K)-ones(n,1)*knotsq; %Knots matrix: n x nknot
    Z1=(Z1.*(Z1>0)).^dpoly;
    Z=[Z0,Z1];
end
q=size(Z,2);
G=diag([zeros(1,dpoly+1),ones(1,K)]);

% form a big design matrix
if p==1,
    P=(X*ones(1,q)).*Z;GG=G;
else
    P=(X(:,1)*ones(1,q)).*Z;GG=G;
    for r=2:p,
        P=[P,(X(:,r)*ones(1,q)).*Z];
        GG=matdiag(GG,G);
    end %P: Nx(pq) matrix
end
[tmpdim1,tmpdim2]=size(Z);

%nspar=tmpdim2+1;
nspar=20;%number of knots;
[U,D]=eig(P'*P);
d=sort(abs(diag(D)));
A=U*diag(d.^(-1/2))*U';
A=A*GG*A';
[U,D]=eig(A);
d=sort(diag(D));

hmax=(K/.01-1)/d(p*(dpoly+1)+1);%% original dd(3)
hmin=max([1/n,(K/ (.8*q)-1)/d(p*q)]);%% original dd(p)

hmax
hmin

[hmin,hmax]
vspar0=logspace(log10(hmin),log10(hmax),nspar)';

vspar0

if length(spar)==p,
    vspar=[ones(nspar,1)*spar,vspar0];
else
    vspar=vspar0*ones(1,p+1);
end

```

```

% Selection of Smoothing parameter
if length(spar)<=p,
%p = number of covariates, including intercept
  vhcrit=[];
  for i=1:nspar,
    spar0=vspar(i,:);
    params=struct('spar',spar0,...
                  'K',K,...
                  'dpoly',dpoly,...
                  'nkflag',nkflag,...
                  'xfit',xfit);
    [mbeta,mbsig,vfit,hcrit]=vcpsme0miss(data,params,wght);
    disp(hcrit((p+1):(p+4)))
    vhcrit=[vhcrit;hcrit];
  end

  [gcv,fflag]=min(vhcrit(:,bflag+p+1));
  spar=vspar(fflag,:);
  cflag=1;
end

else
  cflag=0;
end

param=[K,dpoly,nkflag,Gflag,bflag];

%Refitting
params=struct('spar',spar,...
              'K',K,...
              'dpoly',dpoly,...
              'nkflag',nkflag,...
              'xfit',xfit);
[mbeta,mbsig,vfit,hcrit]=vcpsme0miss(data,params,wght);

if cflag==0,vhcrit=hcrit; end

%% Second Function

function [mbeta,mbsig,vfit,crit,D,sigma2]
=vcpsme0miss(data,paramstruct,wght)
% Varying Coefficient Mixed-Zffects Models Using P-Splines
%  $y_i(t)=x_i(t)'*beta(t)+v_i(t)+e_i(t)$ ,  $i=1,2,\dots,n$ 
% beta(t) varying coefficient curve
% vi(t) the i-th subject effect

```

```

%          vi ~N(0,D)
%          ei ~N(0,sigma2 Ini)
%
% Originally Designed by Jin-Ting Zhang, Stat & Applied Prob.,
%                   NUS, Singapore, 2005
%                   and Hulin, Wu, University of Rochester,
%
% Xiaoxue Li added a weight to handle missing by deisng (MBD)
%   problem, 2014
%

[n,m]=size(data);
subj=data(:,1); usubj=unique(subj);nsubj=length(usubj);
x=data(:,2);y=data(:,3);
z=data(:,m); %random effect

[ux,flag1,flag2]=unique(x); nux=length(ux);
if m==3,
    X=ones(n,1);p=1; % no covariates;
else

    X= data(:,4:m);p=m-3;
end

%Deal with weights,
if length(wght)==0, wght=eye(n); end

% Now update parameters as specified,
% by parameter structure (if it is used)
%

if nargin > 1 ;    % then paramstruct is an argument

    if isfield(paramstruct,'spar') ;
        spar = getfield(paramstruct,'spar') ;
    end ;

    if isfield(paramstruct,'dpoly') ;
        dpoly= getfield(paramstruct,'dpoly') ;
    else
        dpoly=2;
    end ;

    if isfield(paramstruct,'K') ;
        K = getfield(paramstruct,'K') ;
    else

```

```

    K=-1;
end ;

if K==-1,
    K=kssel(nux,dpoly);
end

if isfield(paramstruct,'nkflag');
    nkflag=getfield(paramstruct,'nkflag');
else
    nkflag=1;
end

if isfield(paramstruct,'Z') ;
    Z = getfield(paramstruct,'Z') ;
    q=size(Z,2);
else

    Z0=ones(n,1);
    for ii=1:dpoly,
        Z0=[Z0,x.^ii]; %%Cubic polynomials
    end
    if K==0,
        Z=Z0;
    else
        if nkflag==0,
            xmin=min(ux);xmax=max(ux);
            knotsq=xmin+(xmax-xmin)*[1:K]/(K+1);
        elseif nkflag==1,
            knotsq=prctile(x,[1:K]/(K+1)*100); %% quantiles as knots
            knotsq=unique(knotsq);
            K=length(knotsq);
            knotsq=knotsq(:)'; %% so that knots is a row vector
        end

        Z1=x*ones(1,K)-ones(n,1)*knotsq; %%Knots matrix: n x nknot
        Z1=(Z1.*(Z1>0)).^dpoly;
        Z=[Z0,Z1];
    end
    q=size(Z,2);%%
end ;

if isfield(paramstruct,'xfit') ;
    xfit = getfield(paramstruct,'xfit') ;
    xflag=1;
else
    xfit=ux;

```



```

    xflag=0; % Only intercept is included;
end ;

if isfield(paramstruct,'G') ;
    G = getfield(paramstruct,'G') ;
else
    G=diag([zeros(1,dpoly+1),ones(1,K)]);
end;

    if isfield(paramstruct,'D') ;
        D = getfield(paramstruct,'D') ;
    else
        D=eye(q);
    end ;

if isfield(paramstruct,'sigma2') ;
    sigma2 = getfield(paramstruct,'sigma2') ;
else
    sigma2 = 1;
end ;

if isfield(paramstruct,'Niter') ;
    Niter = getfield(paramstruct,'Niter') ;
else
    Niter = 10;
end ;

if isfield(paramstruct,'Diter') ;
    Diter = getfield(paramstruct,'Diter') ;
else
    Diter = .01; % original 0.001;
end ;

if isfield(paramstruct,'indfig') ;
    indfig = getfield(paramstruct,'indfig') ;
else
    indfig=0;
end ;

end ; % input parameters

%%%%%%%%%%%%%%

% form a big design matrix

```

```

if p==1,
    P=(X*ones(1,q)).*Z;GG=spar(1)*G;
else
    P=(X(:,1)*ones(1,q)).*Z;
    GG=spar(1)*G;
    for r=2:p,
        P=[P,(X(:,r)*ones(1,q)).*Z];
        GG=matdiag(GG,spar(r)*G);
    end %%P: Nx(pq) matrix
end

kflag=1;dflag=1;
while (kflag<Niter)&(dflag>Diter),

    %disp(['Iteration ', num2str(kflag)])
    A=0;B=0;
    % Estimate beta
    DD=pinv(pinv(D)+spar(p+1)*G);
    for i=1:nsubj,
        flagi=(subj==usubj(i));
        ni=sum(flagi);
        yi=y(flagi);

        Zi=Z(flagi,:);
        Pi=P(flagi,:);
        Wi=wght(flagi,flagi);

        Vi=Zi*DD*Zi'+sigma2*Wi;%*eye(ni);
        temp=pinv(Vi);

        A=A+Pi'*temp*Pi;B=B+Pi'*temp*yi;
    end
    beta=pinv(A+GG)*B;

    % Estimate b, variance componets sigma2 and D using EM algorithm
    % v:nsubjxq

    SSE1=0;SSE2=0;SSE3=0;SSE4=0;
    for i=1:nsubj,
        flagi=(subj==usubj(i));
        ni=sum(flagi);
        Zi=Z(flagi,:);
        Pi=P(flagi,:);
        yi=y(flagi);
        Wi=wght(flagi,flagi);

        Vi=Zi*DD*Zi'+sigma2*Wi;%eye(ni);

```

```

temp=pinv(Vi);
temp0=Pi*beta;

b(:,i)=DD*Zi'*temp*(yi-temp0);
ei=yi-temp0-Zi*b(:,i);

SSE1=SSE1+ei'*ei;
SSE2=SSE2+b(:,i)*b(:,i)';
SSE3=SSE3+(trace(Wi)-trace(Wi*temp*Wi))*sigma2);
SSE4=SSE4+(DD-DD*Zi'*temp*Zi*DD);
end

%Update D and sigma2 using EM algorithm
D1=(SSE2+SSE4)/nsubj;
sigma2=(SSE1+SSE3*sigma2)/n;

dflag=max(max(abs(D1-D)));
D=D1; kflag=kflag+1;
end % end for EM updating

%Compute the AIC, BIC
%Computing the Loglik
A=0;B=[];
% Estimate beta
DD=pinv(pinv(D)+spar(p+1)*G);
for i=1:nsubj,
    flagi=(subj==usubj(i));
    ni=sum(flagi);
    Zi=Z(flagi,:);
    Pi=P(flagi,:);
    Wi=wgth(flagi,flagi);

    Vi=Zi*DD*Zi'+sigma2*Wi;%eye(ni);
    temp=pinv(Vi);
    A=A+Pi'*temp*Pi;B=[B,Pi'*temp];
end
AA=pinv(A+GG);BB=AA*B;
beta=BB*y;
%betaSig2=AA;
betaSig2=AA*A*AA;
df=trace(P*BB);

Loglik=0;A=0;B=0;SSE=0;
for i=1:nsubj,
    flagi=(subj==usubj(i));
    ni=sum(flagi);
    Zi=Z(flagi,:);

```

```

    Pi=P(flagi ,:);
    yi=y(flagi);
    Wi=wght(flagi ,flagi);

    Vi=Zi*DD*Zi'+sigma2*Wi;%eye(ni);
    temp=pinv(Vi);
    temp0=Pi*beta;
    b(:,i)=DD*Zi'*temp*(yi-temp0);

    A=A+Zi'*temp*Zi;
    B=B+Pi'*temp*(Zi*DD*Zi')*temp*Pi;
    SSE=SSE+sum((yi-temp0-Zi*b(:,i)).^2);
    Loglik=Loglik-.5*(yi-temp0)'*temp*(yi-temp0)
        +.5*log(det(temp));
end

dfv=trace(DD*A)-trace(AA*B);
vdf=[df,dfv];
gcv=SSE/(1-sum(vdf)/n)^2/n;
crit=-2*Loglik+[2,log(nsubj)]*sum(vdf);
crit=[spar,gcv,crit,Loglik,vdf];

%% Outputs
Zmat=Z(flag1 ,:);
for r=1:p,
    fflag=[(1+(r-1)*q):(r*q)];
    mbeta(:,r)=Zmat*beta(fflag);
    mbsig(:,r)=sqrt(diag(Zmat*betaSig2(fflag,fflag)*Zmat'));
end
vfit=Zmat*b;% nux x n matrix

if xflag==0,
    mbeta=[ux,mbeta];
    mbsig=[ux,mbsig];
else
    for r=1:p,%p is number of parameters
        mbeta0(:,r)=spline(ux,mbeta(:,r),xfit);
        mbsig0(:,r)=spline(ux,mbsig(:,r),xfit);
    end
    mbeta=[xfit,mbeta0];
    mbsig=[xfit,mbsig0];
    for i=1:nsubj,
        vfit0(:,i)=spline(ux,vfit(:,i),xfit);
    end
    vfit=vfit0;
end
end

```

BIBLIOGRAPHY

- [1] Asparouhov T. and Muthén B. (2008). Multilevel mixture models. *Advances in Latent Variable Mixture Models*. Charlotte, NC: Information Age Publishing.
- [2] Bahadur R., (1961). Multivariate probit analysis. *Biometrics*, 26:535–546.
- [3] Benowitz N. (2010). Nicotine addiction. *New England Journal of Medicine*, 362:2295–2203.
- [4] Brumback B., Rice A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of American Statistical Association* 93: 961–994.
- [5] Centers for Disease Control and Prevention (CDC, 2003). Prevalence of current cigarette smoking among adults and changes in prevalence of current and some day smoking-United States, 1996-2001. *Centers for Disease Control and Prevention, Morbidity and Mortality weekly report*, 52:303–307.
- [6] Centers for Disease Control and Prevention (CDC, 2008a). *Behavioral Risk Factor Surveillance System Survey Data*, Atlanta, GA: U.S. Department of Health and Human Services.
- [7] Centers for Disease Control and Prevention (CDC, 2008b). Cigarette smoking among adults-United States, 2007. *Morbidity and Mortality Weekly Report*, 57:1221–1226.
- [8] Cook R.; Lawless J. (2007). *The Statistical Analysis of Recurrent Events*, Springer, 128–133.
- [9] Cox D. (1972). Stochastic Point Processes. *The statistical analysis of dependencies in point processes*, New York: Wiley., 55–66.
- [10] Duchateau L., Janssen P. (2008). The Frailty Model. *Springer* 30–31.
- [11] Fitzmaurice G., Marie D., Verbeke G. and Molenberghs G. (2008). *Longitudinal Data Analysis*, CRC press, 1–22, 31–146.
- [12] Grizzle J., Starmer C. and Koch G. (1969). Analysis of categorical data by linear models. *Biometrics*, 15: 489–504.

- [13] Gumbel E., (1961). Bivariate logistic distribution. *Journal of the American Statistical Association*, 56: 335–349.
- [14] Hoover D., Wu C. and Yang L. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 85, 809–822.
- [15] Harville D. (1976). Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects. *Annals of Statistics*, 4(2): 384–395.
- [16] Harville D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358): 320–338.
- [17] Hastie T. and Tibshirani R. (1993). Varying-coefficients models. *Journal of Royal Statistical Society, Series B* 55, 757–796.
- [18] Hedeker D., Mermelstein R., Berbaum M. and Campbell R. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) Data. *Biometrics*, 64:627–634.
- [19] Ke C. and Wang Y. (2001). Semiparametric nonlinear mixed-effects models and their applications. *Journal of American Statistical Association* 86:1042–1050.
- [20] Kim H. and Golub G. and Park H. (2005). Missing values estimation for DNA microarray gene expression data: local least squares imputation. *Bioinformatics*, 21(2):187–198.
- [21] Liang H., Wu H., and Carroll R. (2003). The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient semiparametric models with measurement error. *Biostatistics* 4, 297–312.
- [22] Liang K. and Zeger S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73: 13–22.
- [23] Muenz L. and Rubinstein L. (1985). Markov Models for Covariate Dependence of Binary Sequences. *Biometrics*, 41: 91–101.
- [24] Nelder J. and Wedderburn R., (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A* 135, 370–384.
- [25] Rathbun S., Shiffman S. and Gwaltney C. (2006). Point process models for event history data. *Models for intensive longitudinal data*, Oxford: Oxford university press, 224–228; 247–250.
- [26] Rathbun S., Shiffman S. and Gwaltney C. (2007). Modeling the effects of partially observed covariates on Poisson process intensity. *Biometrika*, 94(1):627–634.

- [27] Rovine M. and Walls T. (2006). Multilevel autoregressive modeling of interindividual differences. *Models for intensive longitudinal data*, Oxford: Oxford university press, 136–137.
- [28] Shiffman S., Dunbar M., Li X., School S., Tindle H., Anderson S. and Ferguson S. (2014). Smoking Patterns and Stimulus Control in Intermittent and Daily Smokers. *PLOS ONE*, 9(3).
- [29] Shiffman S., Tindle H., Li X., Scholl S., Dunbar M. and Mitchell-Miland C. (2012). Characteristics and Smoking Patterns of Intermittent Smokers. *Experimental and Clinical Psychopharmacology*, 20(4): 264–277.
- [30] Shiffman S., Stone A. and Hufford M. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, 94(1): 627–634.
- [31] Stone A., Shiffman S., Atienza A. and Nebeling L. (2007). *The Science of Real-time Data Capture*, Oxford: Oxford university press, 1–5, 76–110.
- [32] Tan X., Li R., Li Y., Dierker L. (2012). A time-varying effect model for intensive longitudinal data. *Psychol Methods* 17(1):61–77.
- [33] Walls T. and Hoppner B. and Goodwin M. (2007). Statistical Issues in Intensive Longitudinal Data Analysis. *Statistical Issues in Intensive Longitudinal Data Analysis*, Oxford: Oxford university press, 348–356.
- [34] West M., Harrison P. and Migon H. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of American Statistical Association* 80:73–83.
- [35] Wu H. and Liang H. (2004). Backfitting random varying-coefficient models with time-dependent smoothing covariates. *Scandinavian of Statistics* 31, 3–19.
- [36] Wu H. and Zhang J. (2006). Nonparametric Regression Methods for Longitudinal Data Analysis. *Wiley-Interscience*: 300–302.
- [37] Wu H. and Zhang J. (2006). Nonparametric Regression Methods for Longitudinal Data Analysis. *Wiley-Interscience*: 214–215.