The Pennsylvania State University

The Graduate School

Eberly College of Science

JOINT MODELING OF LONGITUDINAL BINARY AND CONTINUOUS RESPONSES: NEW MODELS, STATISTICAL PROCEDURES AND APPLICATIONS

A Dissertation in

Statistics

by

Esra Kurum

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2012

The dissertation of Esra Kurum was reviewed and approved* by the following:

Runze Li
Professor of Statistics and Professor of Public Health Sciences
Dissertation Advisor and Chair of Committee

Naomi Altman
Professor of Statistics

Stephanie Lanza
Research Associate Professor of Health and Human Development

Zhibiao Zhao
Assistant Professor of Statistics

*Signatures are on file in the Graduate School.

Abstract

Joint Modeling of Longitudinal Binary and Continuous Responses: New Models, Statistical Procedures and Applications

Longitudinal data of mixed (e.g., binary and continuous) type are now very common. In this dissertation, we propose two classes of joint modeling procedures for studying the time-varying association between two intensively measured longitudinal responses: a continuous one and a binary one. Since it is well-established that varying-coefficient models increase the flexibility of parametric models and may reduce the modeling bias, our first procedure is based on time-varying coefficient models. However, it is known that semiparametric models represent an appealing compromise between parametric models and nonparametric models, retaining the explanatory power of parametric models while offering the flexibility of nonparametric models. Hence, in the second part of this dissertation, we propose a semiparametric approach, namely, partially linear models. A major challenge in jointly modeling binary and continuous responses is the lack of a multivariate distribution. We suggest introducing a normal latent variable underlying the binary response and factorizing the model into two components: a marginal model for the continuous response, and a conditional model for the binary response given the continuous response. For both techniques, we develop a two-stage estimation procedure and establish the asymptotic normality of the resulting estimators. We also derive the standard error formulas for

the estimated coefficients. In each part we conduct a Monte Carlo simulation study to assess the finite sample performance of our procedures. The proposed nonparametric method is illustrated by an empirical analysis of smoking cessation data (Shiffman et al., 1996), in which we investigate the association between urge to smoke, the continuous response, and the status of alcohol use, the binary response, and how this association varies over time. We apply our proposed semiparametric methodology to data from the Womens Interagency HIV Study. In this analysis we examine the time-varying association between CD4 cell count, the continuous response, and smoking status, the binary response.

Table of Contents

List of Tables

List of Figures

Acknowledgments

During my study at the Pennsylvania State University, I received support from many people. I would like to thank everyone who encouraged me and helped me.

My foremost thanks go to my advisor Dr. Runze Li. I would like to thank him for his understanding and patience. He has provided me the most generous mentoring in academic and personal life.

I am most indebted to Dr. Damla Şentürk for working with me through my master's thesis. She worked closely with me during my first three years at Penn State University. She provided me the kindest assistance of all sorts and helped me through obstacles in my life.

I would like to express my thanks to Dr. Stephanie Lanza, my committee member from the Methodology Center, for providing me with insightful guidance in analyzing both smoking cessation (Shiffman et al., 1996) and Women's Interagency HIV studies. I wish to express my gratitude to my committee members from the Statistics department, Dr. Naomi Altman and Dr. Zhibiao Zhao, for taking time from their busy schedules to read and react to my dissertation. I would also like to thank all of those at Penn State University who have been supportive as I have been completing this project. I would especially like to mention Ceyda Coruh, Elif Balin, and Hakan Arslan.

I want to express my love and gratitude to my parents, Gulsen and Selahattin Kurum, and my sister, Efsun Kurum, for their unwavering support and encouragement, without which I would truly have been lost. Finally, I want to thank, John Hughes, for his endless supply of support, for which my mere expression of thanks does not suffice.

Esra Kurum

The Pennsylvania State University

August, 2012

Chapter 1

Introduction

Various forms of clustered data, such as repeated measurements in a longitudinal study, occur in many fields such as ecology, medical studies and social sciences. To improve the statistical power and to detect the changes over time in a clinical trial with limited experimental units, the same object is measured at different time points to form a longitudinal data set.

There are two major difficulties in analyzing longitudinal data. One is to incorporate the within-subject correlation structure into the estimation procedure. The other is to handle longitudinal data that are collected at irregular and subject-specific time points. These challenges led statisticians to develop new statistical procedures for longitudinal data.

Parametric modeling is one methodology that has been extended to longitudinal data (Diggle et al., 2002). There are two important assumptions of parametric models. First, both linear regression models and generalized regression models assume a linear predictor. Second, parametric models require every individual to follow the same parametric model. However, these two assumptions are restrictive and may introduce biased estimation if they are violated. Hence, to relax the assumptions of parametric models, nonparametric procedures for longitudinal data have been introduced. Müller (1988) presented a detailed summary of earlier nonparametric procedures developed for longitudinal data.

Another statistical procedure that has been introduced for longitudinal data analysis is time-varying coefficient models. Varying coefficient models increase the flexibility of linear models and reduce the modeling bias. For longitudinal data sets, an obvious option is to allow the regression coefficients to vary over time. Hoover et al. (1998), Wu et al. (1998), Fan and Zhang (2000); Wu and Chiang (2000), Chiang et al. (2001), Huang et al. (2002) are among the authors who proposed statistical methods via time-varying coefficient models for longitudinal data. A detailed review on this approach will be given in Section 2.4.

Early work on modeling longitudinal and clustered data focused on developing methodologies for datasets with a single response. More recent studies have involved multiple responses, often of mixed type, e.g., binary and continuous. The work described in this dissertation was motivated by such a dataset. The data, which were collected intensively during a smoking cessation study (Shiffman et al., 1996), contain multiple responses such as urge to smoke (a continuous response), alcohol use, and presence of other smokers (both binary responses). The latter two responses are of interest because it has been observed that alcohol consumption and the presence of other smokers increase the odds of smoking (Hymowitz et al., 1997; Shiffman and Balabanis, 1995; Shiffman et al., 2002). Our primary interest is to estimate the time-varying association between these responses and urge to smoke so that researchers can understand how the association between these variables changes during the smoking cessation process. To estimate the association between the variables, we need to model the variables jointly.

The major difficulty in modeling binary and continuous responses jointly is the lack of a natural multivariate distribution. To overcome this challenge, Catalano and Ryan (1992) suggested introducing a continuous latent variable underlying the binary response, and assumed that the latent variable and the continuous response follow a joint normal distribution. After introducing the latent variable, Catalano and Ryan (1992) suggested decomposing the joint distribution into two components that can be modeled separately: a marginal distribution for the continuous response, and a conditional distribution for the binary response given the continuous response. The first component is readily obtained, and the second component is obtained using the assumption of joint normality. Cox and Wermuth (1992) compared several joint models based on factorization of the joint model of the discrete and continuous responses.

Unlike the method proposed by Catalano and Ryan (1992), Fitzmaurice and Laird (1995) reversed the order of the conditioning so that the regression parameters for both binary and continuous responses have marginal interpretations. Gueorguieva and Agresti (2001) studied a similar correlated probit model as did Catalano and Ryan (1992). However, Catalano and Ryan (1992) used generalized estimating equations methodology (Liang and Zeger, 1986; Zeger and Liang, 1986), whereas Gueorguieva and Agresti (2001) proposed a Monte Carlo EM algorithm for finding maximum likelihood estimates. The main interest in the papers mentioned above was the effect of a treatment on the multivariate response vector. However, the primary concern in Liu et al. (2009) is to model the association between the two responses. They extend the latent variable approach to longitudinal data assuming that the latent variable and the continuous response follow a multivariate normal distribution. More references on joint modeling will be given in Section 2.6.

In a longitudinal setting, regression coefficients may change over time; an ordinary regression model cannot capture this change. Thus, unlike the aforementioned joint modeling techniques, we allow time-varying effects of predictors by introducing time-varying coefficients (Brumback and Rice, 1998; Hoover et al., 1998; Wu et al., 1998) to joint modeling of longitudinal binary and continuous responses. This will allow researchers to investigate how the relationship between the responses and the predictors changes during the smoking cessation process. In addition, varying coefficient models can enhance the flexibility of linear regression models and reduce the modeling bias (Cleveland et al., 1992; Fan and Zhang, 2008; Hastie and Tibshirani, 1993).

In our current work we focus on estimating the association between longitudinal binary and continuous responses measured at the same time point within a subject. This association may change over time, and observing this change would give researchers useful knowledge about the relationship. Hence, our approach allows the association to be time-varying. To further enhance the flexibility of our model, we also assume that the binary and continuous responses for each subject are correlated at all measurement times, and we allow these correlations to be time-varying.

To estimate the time-varying correlation, we employ the above mentioned joint modeling approach. That is, we introduce a continuous latent variable underlying the binary response, and we decompose the joint distribution into two components. This leads to a two-stage estimation procedure. In the first stage we fit the marginal model of the continuous response by using time-varying coefficient models (Brumback and Rice, 1998; Hoover et al., 1998; Wu et al., 1998). In the second stage we use generalized time-varying coefficient models (Cai et al., 2000) to fit the conditional model of the binary response. We establish the asymptotic normality of the proposed estimators. The efficacy of our methodology is demonstrated by a simulation study. We further illustrate the proposed methodology by analyzing the smoking cessation study (Shiffman et al., 1996).

It is known that nonparametric models relax the restrictive assumptions of parametric models, which may reduce the modeling bias but be too flexible to permit concise conclusions. Hence, we further propose a joint modeling framework based on a semiparametric approach, namely, partially linear models. These models are natural extensions of ordinary linear regression models with multiple predictors and nonparametric models with a single covariate, such as time. Thus, they retain the explanatory power of parametric models and the flexibility of nonparametric models. In our approach using the parametric component of the partially linear models, researchers can investigate the relationship between the continuous response and a set of predictors.

Similar to the nonparametric approach, to estimate the time-varying correlation using the semiparametric framework, we introduce a continuous latent response underlying the binary response, and assume that the latent variable and the continuous response follow a joint normal distribution. We factorize the joint distribution of the binary and continuous responses into two components, which results in a two-stage estimation procedure. In the first stage we fit the marginal model for the continuous response using a backfitting algorithm (Zeger and Diggle, 1994). In the second stage we employ generalized time-varying coefficient models (Cai et al., 2000) to fit the conditional model of the binary response. We show the asymptotic normality of the proposed estimators in both stages. The performance of the proposed procedure is examined by using a simulation study. In our modeling scheme, we take into account the fact that the binary and continuous responses for each subject can be correlated at all measurement times, and these correlations can be time-varying. We present results of a simulation study designed to show the finite sample behavior of our estimators. We apply our methodology to a subset from the Women's Interagency HIV Study. This subset contains 205 women recruited in the study between 1994-1995. Our main interest in this data set is to observe the time-varying association between CD4 cell count, which is a continuous response and is used to assess the progress of HIV disease, and smoking status, a binary response.

This dissertation is organized as follows. Chapter 2 provides a detailed review of the statistical concepts that are relevant to the development of our estimation procedure. In Chapter 3 we introduce our nonparametric approach for joint modeling of longitudinal binary and continuous responses. We describe our two-stage estimation procedure by using local linear regression techniques and study asymptotic properties of the resulting estimators. We conduct a simulation study to investigate the finite sample behavior of our estimators and we further illustrate the proposed methodology by analyzing the smoking cessation study (Shiffman et al., 1996). In Chapter 4 we develop our semiparametric approach for joint modeling of longitudinal binary and continuous responses. We propose a two-stage estimation procedure and describe the asymptotic properties of the estimators. We present the results of a simulation study designed to show the finite sample behavior of our estimators and illustrate the proposed methodology by analyzing data from the Women's Interagency HIV Study. In Chapter 5 we give our conclusions and outline some future research topics.

Chapter 2

Literature Review

In this chapter we briefly review statistical concepts that are relevant to the development of our estimation procedures. In Section 2.1 we review generalized linear models. Brief summaries on local polynomial regression and varying coefficient models are given in Sections 2.2 and 2.3, respectively. In Section 2.4 we summarize the time-varying coefficient models for longitudinal data. A short review on partially linear models is provided in Section 2.5. In Section 2.6 we present existing estimation procedures for joint models of binary and continuous responses.

2.1 Generalized Linear Models

Let Y be a response variable and $\mathbf{X}$ be the covariates, a generalized linear model (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989) assumes that the regression function $m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$ satisfies the following relation:

$$\eta(\mathbf{X}) = g\left\{m(\mathbf{X})\right\} = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta},$$

where $g(\cdot)$ is a known link function, which transforms the regression function $m(\mathbf{x})$ into a linear predictor $\eta(\mathbf{X}) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}$. The generalized linear model has three components: the link function $g(\cdot)$, the random component; that is, the distribution $Y|\mathbf{X} = \mathbf{x}$ and the systematic component $\eta(\mathbf{X}) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}$. A fundamental assumption of generalized linear models is that the conditional distribution of Y given $\mathbf{X}$ belongs to the exponential family:

$$f_y(y:\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\},$$

where $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ are specific functions, ϕ is the dispersion parameter and θ is the canonical parameter. The mean function $m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x}) = b'(\theta)$ is a function of θ, hence θ is the parameter of interest and ϕ is referred as the nuisance parameter. Canonical link corresponds to $g(\cdot) = b'^{-1}(\cdot)$ and with this link, we have $\theta = \eta = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}$. For ordinary linear regression models, the link function is $g(t) = t$, which is also the canonical link when the random component is assumed to follow a Gaussian distribution with mean μ and variance σ^2. When the response Y is a binomial response or count response, it is fitted by a binomial distribution or Poisson distribution, respectively. The canonical links for binomial distribution and Poisson distribution are $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$ and $g(\mu) = \log(\mu)$, respectively. McCullagh and Nelder (1989) provides estimation procedures and various types of generalized linear models with examples.

2.2 Local Polynomial Regression

The concept of local polynomial regression has been around for a long time. It was studied by Stone (1977), Cleveland (1979), Fan (1992, 1993) and Fan and Gijbels (1992, 1996). These papers introduced local polynomial fitting and studied its advantages. In this section, we briefly review the estimation procedure and advantages of local polynomial fitting.

Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ is a random sample from a nonparametric regression model,

$$Y = m(X) + \varepsilon,$$

where $E(\varepsilon|X) = 0$, $Var(\varepsilon|X = x) = \sigma^2(x)$, $m(x)$ is assumed to be a smooth function. The aim is to estimate $m(x_0) = E(Y|X = x_0)$ and its derivatives $m'(x_0), m''(x_0), \cdots, m^{(p)}(x_0)$. Assuming that the $(p+1)^{th}$ derivative of $m(x)$ at x_0 exists, we can locally approximate the regression function $m(x)$ in the neighborhood of x_0 via the Taylor expansion,

$$m(x) \approx \sum_{j=0}^{p} \frac{m^{(j)}(x_0)}{j!}(x - x_0)^j \equiv \sum_{j=0}^{p} \beta_j (x - x_0)^j. \tag{2.1}$$

From a statistical point of view, (2.1) suggests modeling $m(x)$ locally. Intuitively, the data points closer to x_0 carries more information about $m(x)$ than the points further away. Hence, the local parameter β_js are estimated by minimizing a locally weighted least squares function,

$$\sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j \right\}^2 K_h(X_i - x_0), \tag{2.2}$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$ with $K(\cdot)$ as the kernel function that assigns weights to each datum point and h is the bandwidth or smoothing parameter that governs the size of the local neighborhood. One of the commonly used kernel functions is the symmetric beta family,

$$K(x) = \frac{1}{Beta(1/2, \gamma + 1)}(1 - x^2)^{\gamma} 1_{\{|x| \le 1\}}, \quad \gamma = 0, 1, 2 \ldots,$$

where $Beta(\cdot, \cdot)$ denotes a beta function. When $\gamma = 0, 1, 2$ and 3, the kernel functions are referred to as the uniform, Epanechnikov, biweight and triweight kernel functions, respectively. Another commonly used kernel function is the Gaussian kernel,

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Denote the solution of (2.2) by $\hat{\beta}_j$. Since, by definition, $m(x) \equiv \beta_0$, then the estimated regression function is $\hat{m}(x) = \hat{\beta}_0$. Furthermore, the estimator for

$m^{(v)}(x)$ is,

$$\hat{m}^{(v)}(x) = v!\hat{\beta}_v.$$

When $p = 0$ the local polynomial regression is referred to as a kernel regression or local constant fitting. This type of fitting does not allow us to estimate the derivatives of the regression function. One of the commonly used kernel estimators is the Nadaraya-Watson estimator (Nadaraya, 1964; Watson, 1964),

$$\hat{m}_h(x) = \frac{\sum_{i=1}^{n} K_h(X_i - x)Y_i}{\sum_{i=1}^{n} K_h(X_i - x)}. \tag{2.3}$$

Another well-known kernel regression estimator is the Gasser-Müller estimator (Gasser and Müller, 1984),

$$\hat{m}_h(x) = \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_h(u - x)du\, Y_i, \tag{2.4}$$

where $s_i = (X_i + X_{i+1})/2$ assuming $X_i < X_{i+1}$, $X_0 = -\infty$ and $X_{n+1} = +\infty$. The denominator in (2.3) is inconvenient when taking the derivatives of the estimator and deriving the asymptotic properties of the estimator. However, the Gasser-Müller estimator does not need a denominator since the sum of the weights in (2.4) equals to one.

When $p = 1$, the local polynomial regression is referred to as local linear regression. Local linear fitting enables us to estimate both the regression function $m(x)$ and its first derivative, $m'(x)$. The local linear estimator is defined as,

$$\hat{m}_0(x) = \frac{\sum_1^n w_i Y_i}{\sum_1^n w_i}, \quad w_i = K_h(X_i - x)\{S_{n,2} - (X_i - x)S_{n,1}\},$$

where $S_{n,j} = \sum_{i=1}^{n}(X_i - x)^j K_h(X_i - x)$. The asymptotic properties of the Nadaraya-Watson estimator, the Gasser-Müller estimator and local linear estimator are demonstrated in Table 2.1, adapted from Fan and Gijbels (1996). In Table 2.1, $b_n = \frac{1}{2}\int_{-\infty}^{+\infty} u^2 K(u)du\, h^2$ and $V_n = \frac{\sigma^2(x)}{f(x)nh}\int_{-\infty}^{+\infty} K^2(u)du$.

Local polynomial fitting has some advantages over the Nadaraya-Watson estimator (2.3) and Gasser-Müller estimator (2.4). According to Table 2.1, for random design models, the Nadaraya-Watson estimator has an undesirable bias form and the Gasser-Müller estimator corrects this bias form; however, it pays a price at the inflation of its variance. Another advantage of local polynomial fitting is that there is no boundary effect: the bias in the interior is of the same order as the one at the boundary, without using specific boundary kernels (Ruppert and Wand, 1994; Fan and Gijbels, 1996). Local polynomial fitting is also appealing, due to its minimax efficiency properties (Fan, 1993), whereas the Nadaraya-Watson estimator has zero minimax efficiency. For further discussion on the advantages of local polynomial regression, see Fan and Gijbels (1996).

Table 2.1: Asymptotic biases and variances for kernel regression estimators

Method	Bias	Variance
Nadaraya-Watson estimator	$\left\{m''(x) + \frac{2m'(x)f'(x)}{f(x)}\right\} b_n$	V_n
Gasser-Müller estimator	$m''(x)b_n$	$1.5V_n$
Local Linear estimator	$m''(x)b_n$	V_n

2.3 Varying Coefficient Models

Let Y be the response variable, and $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$ be the vector of predictors. Then the ordinary regression model can be written as follows:

$$Y = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon, \tag{2.5}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ are the p unknown regression coefficients and $E(\varepsilon|\mathbf{X}) = 0$. Model (2.5) assumes $\boldsymbol{\beta}$ to be constant and this assumption limits the application of linear regression models. For instance, in longitudinal studies and some ecological studies, the regression coefficient may change with some underlying covariates, such as time, temperature or geographical locations and model (2.5) cannot capture this change. To enhance the flexibility of linear regression models and to reduce modeling bias, varying coefficient models allow the regression coefficients to vary over variables of interest, such as time. Hence, these models can be used to explore the dynamic pattern of data. Varying coefficient models are first introduced by Cleveland et al. (1992) and became popular in the statistical literature due to the work by Hastie and Tibshirani (1993). The varying coefficient model is of the following form:

$$Y = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}(U) + \varepsilon, \tag{2.6}$$

where U is a scalar predictor, $\boldsymbol{\beta}(U) = (\beta_1(U), \ldots, \beta_p(U))^{\mathrm{T}}$ consists of the p unknown coefficient functions, and $E(\varepsilon|\mathbf{X}, U) = 0$. There are three main approaches to estimation in varying coefficient models: local polynomial smoothing (Wu et al., 1998; Hoover et al., 1998; Fan and Zhang, 1999; Kauermann and Tutz, 1999), polynomial splines (Huang et al., 2002; Huang and Shen, 2004) and smoothing splines (Hastie and Tibshirani, 1993; Hoover et al., 1998; Chiang et al., 2001). Perhaps, the most natural of the three is kernel smoothing, since varying coefficient models are local linear models (Fan and Zhang, 2008). Hence, in this section we briefly outline the local polynomial smoothing method.

Suppose that $\left\{U_i, \mathbf{X}_i^{\mathrm{T}}, Y_i\right\}$, $i = 1, \ldots, n$, is an independent and identically distributed sample from the varying coefficient model (2.6). To estimate the regression coefficient functions at a fixed point u_0, we locally approximate the functions in a neighborhood of u_0 via the Taylor expansion,

$$\beta_j(u) \approx \beta_j(u_0) + \beta_j'(u_0) \equiv a_j + b_j(u - u_0), \tag{2.7}$$

for $j = 1, \ldots, p$. Let $\mathbf{a} = (a_1, \ldots, a_p)^{\mathrm{T}}$ and $\mathbf{b} = (b_1, \ldots, b_p)^{\mathrm{T}}$. We minimize the following equation:

$$\ell(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{n} \left\{ Y_i - \mathbf{X}_i^{\mathrm{T}} \mathbf{a} - \mathbf{X}_i^{\mathrm{T}} \mathbf{b} (U_i - u_0) \right\}^2 K_h(U_i - u_0),$$

with respect to $(\mathbf{a}, \mathbf{b})$, where $K_h(\cdot) = h^{-1} K(\cdot/h)$ with $K(\cdot)$ as the kernel function and $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$. The solution to the least squares algorithm is,

$$\hat{\boldsymbol{\beta}}(u_0) = (\mathbf{I}_p, \mathbf{0}_p)(\boldsymbol{\Gamma}_u^{\mathrm{T}} \mathbf{W}_u \boldsymbol{\Gamma}_u)^{-1} \boldsymbol{\Gamma}_u^{\mathrm{T}} \mathbf{W}_u Y, \tag{2.8}$$

where $\mathbf{I}_p$ is the identity matrix with size p, $\mathbf{0}_p$ is a size p matrix with each entry equal to zero, $\boldsymbol{\Gamma}_u = (\mathbf{X}, \mathbf{U}_u \mathbf{X})$ with $\mathbf{U}_u = \mathrm{diag}(U_1 - u_0, \ldots, U_n - u_0)$ and $\mathbf{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$; and $\mathbf{W}_u = \mathrm{diag}\big(K_h(U_1 - u_0), \ldots, K_h(U_n - u_0)\big)$ (Fan and Zhang, 2008). The estimator (2.8) is asymptotically normally distributed.

Theorem 2.1. *Under the conditions given in Zhang and Lee (2000),*

$$\mathrm{cov}^{-1/2}(\hat{\boldsymbol{\beta}}(u_0)) \left[\hat{\boldsymbol{\beta}}(u_0) - \boldsymbol{\beta}(u_0) - bias(\hat{\boldsymbol{\beta}}(u_0)) \right] \xrightarrow{D} \mathcal{N}(0, I_p),$$

where
$bias(\hat{\boldsymbol{\beta}}(u_0)) = \frac{1}{2} \mu_2 \, \boldsymbol{\beta}^{(2)}(u_0) \, h^2$, $\mathrm{cov}(\hat{\boldsymbol{\beta}}(u_0)) = \left\{ n \, h \, c(u_0) E(\mathbf{X}\mathbf{X}^T | U = u) \right\}^{-1} \nu_0 \, \sigma^2$, $\mu_i = \int u^i K(u) du$, $v_i = \int u^i K^2(u) du$ *and* $c(u)$ *is the marginal density of* U.

Zhang and Lee (2000) also derived the estimator of the conditional bias and covariance, which are important in hypothesis testing. Let $\mathcal{D} = \big(U_1, \mathbf{X}_1^{\mathrm{T}}, \ldots, U_n, \mathbf{X}_n^{\mathrm{T}}\big)$, using Taylor expansion we have the following:

$$E(\hat{\boldsymbol{\beta}}(u_0) | \mathcal{D}) - \boldsymbol{\beta}(u_0) \approx (\mathbf{I}_p, \mathbf{0}_p)(\boldsymbol{\Gamma}_u^{\mathrm{T}} \mathbf{W}_u \boldsymbol{\Gamma}_u)^{-1} \boldsymbol{\Gamma}_u^{\mathrm{T}} \mathbf{W}_u \psi,$$

where the i^{th} element of ψ is $2^{-1} X_i^{\mathrm{T}} \left\{ \boldsymbol{\beta}^{(2)}(u_0)(U_i - u_0)^2 + 3^{-1} \boldsymbol{\beta}^{(3)}(u_0)(U_i - u_0)^3 \right\}$. This leads the following estimators of the conditional bias and covariance:

$$\begin{aligned} \widehat{\mathrm{bias}}(\hat{\boldsymbol{\beta}}(u_0) | \mathcal{D}) &= (\mathbf{I}_p, \mathbf{0}_p)(\boldsymbol{\Gamma}_u^{\mathrm{T}} \mathbf{W}_u \boldsymbol{\Gamma}_u)^{-1} \boldsymbol{\Gamma}_u^{\mathrm{T}} \mathbf{W}_u \hat{\psi}, \\ \widehat{\mathrm{cov}}(\hat{\boldsymbol{\beta}}(u_0) | \mathcal{D}) &\approx (\mathbf{I}_p, \mathbf{0}_p)(\boldsymbol{\Gamma}_u^{\mathrm{T}} \mathbf{W}_u \boldsymbol{\Gamma}_u)^{-1} \left(\boldsymbol{\Gamma}_u^{\mathrm{T}} \mathbf{W}_u^2 \boldsymbol{\Gamma}_u \right) (\boldsymbol{\Gamma}_u^{\mathrm{T}} \mathbf{W}_u \boldsymbol{\Gamma}_u)^{-1} (\mathbf{I}_p, \mathbf{0}_p)^{\mathrm{T}} \hat{\sigma}^2(u_0), \end{aligned}$$

where $\hat{\psi}$ is ψ with $\boldsymbol{\beta}^{(j)}(u_0)$ replaced by $\hat{\boldsymbol{\beta}}^{(j)}(u_0)$ and $j = 2, 3$. Further information on conditional bias and variance can be found in Carroll et al. (1998) and Fan and Zhang (1999).

Generalized varying coefficient models, introduced by Cai et al. (2000), are extensions of the model (2.6),

$$\eta(U, \mathbf{X}) = g\{m(U, \mathbf{X})\} = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}(U), \tag{2.9}$$

where $g(\cdot)$ is a link function and $m(U, \mathbf{X})$ is the mean regression function of the response variable Y given the covariates $U = u$ and $\mathbf{X} = \mathbf{x}$. The local maximum likelihood method is employed to estimate the coefficient functions, $\boldsymbol{\beta}(U)$, since it has many nice properties as mentioned in Section 2.2. Suppose that $\boldsymbol{\beta}(U)$ has a continuous second derivative, then we locally approximate $\boldsymbol{\beta}(U)$ for u in a neighborhood of u_0 by the Taylor expansion as in (2.7). The local maximum likelihood estimator $\left(\hat{\mathbf{a}}^{\mathrm{T}}, \hat{\mathbf{b}}^{\mathrm{T}}\right)$ of $\left(\mathbf{a}^{\mathrm{T}}, \mathbf{b}^{\mathrm{T}}\right)$ is the maximizer of the local log-likelihood function

$$\ell_n(\mathbf{a}, \mathbf{b}) = \frac{1}{n}\sum_{i=1}^{n} \ell\left(g^{-1}\left[\sum_{j=1}^{p}\{a_j + b_j(U_i - u_0)\}X_{ij}\right], Y_i\right) K_h(U_i - u_0). \tag{2.10}$$

For varying coefficient models, computing the local maximum likelihood estimate (MLE) can be computationally intensive, since the local maximum likelihood (2.10) needs to be maximized for hundreds of distinct values of u_0 and each maximization requires an iterative algorithm. Hence, Cai et al. (2000) proposed a one-step Newton-Raphson estimator.

Denote the gradient and Hessian matrix of the local log-likelihood $\ell_n(\boldsymbol{\beta})$ by $\ell_n'(\boldsymbol{\beta})$ and $\ell_n''(\boldsymbol{\beta})$, respectively. The one-step Newton-Raphson algorithm results in the following updated estimator:

$$\hat{\boldsymbol{\beta}}_{OS} = \hat{\boldsymbol{\beta}}_0 - \{\ell_n''(\boldsymbol{\beta}_0)\}^{-1}\ell_n'(\boldsymbol{\beta}_0),$$

where $\hat{\boldsymbol{\beta}}_0 = \hat{\boldsymbol{\beta}}_0(u_0)$ is the initial estimator.

Cai et al. (2000) proved that one-step local maximum likelihood estimate, $\hat{\boldsymbol{\beta}}_{OS}$, have the same asymptotic distribution as the fully iterative local maximum likelihood estimate, $\hat{\boldsymbol{\beta}}_{MLE}$, as long as the initial estimator $\hat{\boldsymbol{\beta}}_0$ satisfies,

$$H(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) = \mathrm{O}_p \left\{ h^2 + (nh)^{-1/2} \right\},$$

where $\mathbf{H} = \mathrm{diag}\,(1, h) \bigotimes \mathbf{I}_p$ with $\bigotimes$ denoting the Kronecker product.

The asymptotic properties of the local maximum likelihood estimator is provided in the following theorem.

Theorem 2.2. *Under the conditions given in Cai et al. (2000),*

$$\sqrt{nh} \left\{ \hat{\boldsymbol{\beta}}_{MLE}(u_0) - \boldsymbol{\beta}(u_0) - \frac{1}{2}\mu_2 \boldsymbol{\beta}''(u_0)h^2 \right\} \xrightarrow{D} \mathcal{N} \left\{ \mathbf{0}_{p\times 1}, \boldsymbol{\Sigma}\, \nu_0 / c(u_0) \right\},$$

where

$$\boldsymbol{\Sigma} = \left[E \left\{ E \left(\frac{\partial^2 \ell \left[g^{-1} \left\{ \mathbf{X}^T \boldsymbol{\beta}(u_0) \right\}, y \right]}{\partial \left\{ \mathbf{X}^T \boldsymbol{\beta}(u_0) \right\}^2} \middle| \mathbf{X}, U \right) \mathbf{X}\mathbf{X}^T \middle| U = u \right\} \right]^{-1}.$$

Cai et al. (2000) provided an estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}_{MLE}$ using conventional techniques,

$$\widehat{\mathrm{cov}}\big(\hat{\boldsymbol{\beta}}(u_0)\big) = (\mathbf{I}_p, \mathbf{0}_p) \hat{\boldsymbol{\Gamma}}(u_0)^{-1} \hat{\boldsymbol{\Lambda}}(u_0) \hat{\boldsymbol{\Gamma}}(u_0)^{-1} (\mathbf{I}_p, \mathbf{0}_p)^{\mathrm{T}},$$

where $\mathbf{I}_p$ is the identity matrix with size p, and $\mathbf{0}_p$ is a size p matrix with each entry equal to zero,

$$\hat{\boldsymbol{\Gamma}}(u_0) = -\frac{1}{n} \sum_{i=1}^{n} q_2 \left[\sum_{j=1}^{p} \left\{ \hat{a}_j X_{ij} + \hat{b}_j (U_i - u_0) \right\}, Y_i \right] K_h(U_i - u_0) \begin{pmatrix} \mathbf{X}_i \\ \mathbf{X}_i (U_i - u_0)/h \end{pmatrix}^{\bigotimes 2},$$

$$\hat{\boldsymbol{\Lambda}}(u_0) = \frac{h}{n} \sum_{i=1}^{n} q_1^2 \left[\sum_{j=1}^{p} \left\{ \hat{a}_j X_{ij} + \hat{b}_j (U_i - u_0) \right\}, Y_i \right] K_h^2(U_i - u_0) \begin{pmatrix} \mathbf{X}_i \\ \mathbf{X}_i (U_i - u_0)/h \end{pmatrix}^{\bigotimes 2}$$

with $q_j(s,y) = (\partial^j/\partial s^j)\ell\{g^{-1}(s),y\}$ and $A^{\bigotimes 2}$ denotes AA^T for a matrix or vector A.

The extension of varying coefficient models to a nonlinear setting leads to nonlinear varying coefficient models (Kurum et al., 2011),

$$Y = f\left\{\mathbf{X}, \boldsymbol{\beta}(U)\right\} + \varepsilon,$$

where $f(\cdot,\cdot)$ is a pre-specified function, $\boldsymbol{\beta}(U) = (\beta_1(U),\cdots,\beta_p(U))^{\mathrm{T}}$ are the p unknown coefficient functions and $E(\varepsilon|\mathbf{X},U) = 0$. Similar to the estimation procedure in varying coefficient models, we locally approximate the regression coefficient functions in a neighborhood of u_0 via the Taylor expansion as in (2.7). We minimize the following equation with respect to local parameters $(\mathbf{a},\mathbf{b})$:

$$\ell(\mathbf{a},\mathbf{b}) = \sum_{i=1}^{n}\left[Y_i - f\left\{X_i, \mathbf{a} + \mathbf{b}(U_i - u_0)\right\}\right]^2 K_h(U_i - u_0). \tag{2.11}$$

The estimators of the regression coefficient functions are $\hat{\boldsymbol{\beta}}(u_0) = \hat{\mathbf{a}}$. Since there is not a closed form for $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$, an iterative linear regression algorithm is proposed to obtain the values that solves the local least squares function (2.11). For details of this algorithm, see Kurum et al. (2011). It was established that the local estimators are asymptotically normally distributed (Kurum et al., 2011). Define $\boldsymbol{\theta}(u_0) = (a_1,\ldots,a_p,b_1,\ldots,b_p)^{\mathrm{T}}$ and $\hat{\boldsymbol{\theta}}(u_0) = (\hat{\mathbf{a}}^{\mathrm{T}},\hat{\mathbf{b}}^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{H} = \mathrm{diag}\,(1,h)\bigotimes\mathbf{I}_p$ with $\bigotimes$ denoting the Kronecker product and $\mathbf{I}_p$ as the $p\times p$ identity matrix. Let

$$\begin{aligned}
\Gamma_1(u_0) &= E\{f'\{\mathbf{X};\boldsymbol{\beta}(u_0)\}[f'\{\mathbf{X};\boldsymbol{\beta}(u_0)\}]^{\mathrm{T}}|U=u_0\}_{p\times p},\\
\Gamma_2(u_0) &= E\{\sigma^2(u_0,\mathbf{X})f'\{\mathbf{X};\boldsymbol{\beta}(u_0)\}[f'\{\mathbf{X};\boldsymbol{\beta}(u_0)\}]^{\mathrm{T}}|U=u_0\}_{p\times p}.
\end{aligned}$$

Theorem 2.3. *Under the regularity conditions given in Kurum et al. (2011), we have the following result for* $\hat{\theta}(u_0)$*:*

$$\sqrt{nh} \quad \left[\mathbf{H}\{\hat{\boldsymbol{\theta}}(u_0) - \boldsymbol{\theta}(u_0)\} - \frac{h^2}{2\left(\mu_2 - \mu_1^2\right)} \begin{pmatrix} \left(\mu_2^2 - \mu_1\mu_3\right) \boldsymbol{\beta}''(u_0) \\ \left(\mu_3 - \mu_1\mu_2\right) \boldsymbol{\beta}''(u_0) \end{pmatrix} + o_p\left(h^2\right) \right]$$
$$\xrightarrow{\mathcal{D}} \quad N\left(0, \Delta^{-1}\Lambda\Delta^{-1}\right) \quad \text{as } n \to \infty,$$

where

$$\Delta = c(u_0) \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \Gamma_1(u_0) \ \text{ and } \ \Lambda = c(u_0) \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix} \otimes \Gamma_2(u_0).$$

Further details on nonlinear varying coefficient models can be found in Kurum et al. (2011).

2.4 Time-Varying Coefficient Models

Varying coefficient models have been popular in longitudinal data analysis. Of interest is to study the relationship between the predictors and the response variable. Hence, the following linear model is generally used:

$$Y(t) = \mathbf{X}^{\mathrm{T}}(t)\boldsymbol{\beta} + \varepsilon(t), \tag{2.12}$$

for predictors and response variable collected at time t. Although, the predictors and the response variable vary over time, model (2.12) does not allow the relationship between them to vary over time. Thus, Brumback and Rice (1998) and Hoover et al. (1998) proposed the following time-varying coefficient model:

$$Y(t) = \mathbf{X}^{\mathrm{T}}(t)\boldsymbol{\beta}(t) + \varepsilon(t), \tag{2.13}$$

where $\boldsymbol{\beta}(t) = \left(\beta_0(t), \ldots, \beta_p(t)\right)^{\mathrm{T}}$ are arbitrary smooth functions of t, and $\mathbf{X}(t)$ and $\varepsilon(t)$ are assumed to be independent. If the data set comes from an independent cross-sectional study, model (2.13) reduces to a varying coefficient model

(Hastie and Tibshirani, 1993). Hoover et al. (1998) introduced two nonparametric estimation procedures for model (2.13), namely smoothing splines and locally weighted polynomials. First, we summarize the spline method. Assuming the functions $\beta_0(t), \ldots, \beta_p(t)$ are twice continuously differentiable and their second derivatives are bounded and square integrable, a smoothing spline estimator of $\boldsymbol{\beta}(t)$ (2.13) minimizes,

$$J(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[Y_{ij} - \left\{ \sum_{l=0}^{p} X_{ijl} \beta_l(t_{ij}) \right\} \right]^2 + \sum_{l=0}^{p} \lambda_l \int \{\beta_l''(t)\}^2 \, dt, \qquad (2.14)$$

where $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_p)^{\mathrm{T}}$ are the positive-valued smoothing parameters to penalize the roughness of regression coefficient functions. The minimizers of (2.14) are natural cubic splines with knots located at the distinct values of t_{ij}. An adequate smoothing parameter vector λ plays a very important role in the practical implementation of the smoothing splines. If λ_l is too large, it gives an excessive penalty for the roughness of β_l and thus an oversmoothed estimator, whereas if λ_l is too small, it results in an undersmoothed estimator.

In addition to proposing the smoothing spline estimation procedure, Hoover et al. (1998) also extended the existing local polynomial estimation techniques to longitudinal data. Let $W_{ij}(t)$ be the weight functions of t_{ij} and t for $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. The weight function is generally selected as a kernel function $K_h(t - t_{ij}) = h^{-1} K(t - t_{ij}/h)$ with h as the bandwidth. Let $\mathcal{W} = \mathrm{diag}\big(W_{i1}(t), \ldots, W_{in_i}(t)\big)$ be the diagonal weight matrix. The local polynomial fit minimizes the following equation:

$$L(t) = \sum_{i=1}^{n} \left[Y_i - \sum_{l=0}^{p} \{X_{il} \mathcal{B} b_l(t)\} \right]^{\mathrm{T}} \mathcal{W} \left[Y_i - \sum_{l=0}^{p} \{X_{il} \mathcal{B} b_l(t)\} \right], \qquad (2.15)$$

where $\mathcal{B}_i$ is the $n_i \times d$ basis matrix with the entry $(t_{iq} - 1)^{r-1}$ as the $(q, r)^{th}$ element, $\hat{b}(t) = \big(\hat{b}_0(t), \ldots, \hat{b}_p(t)\big)$ with $\hat{b}_l(t) = \big(\hat{b}_{1l}(t), \ldots, \hat{b}_{dl}(t)\big)^{\mathrm{T}}$ is the local polynomial fit and d denotes the degree of the polynomial. The normal equations corresponding to the locally weighted sum of squares (2.15) is

$$\mathcal{N}(t)\hat{b}(t) = \mathcal{M} \circ Y, \tag{2.16}$$

where $\mathcal{M}_{il}(t) = (X_{il}\mathcal{B}_i)^{\mathrm{T}} \mathcal{W}_i(t)$ and $\mathcal{N}(t)$ is a matrix with the entry $\sum_{i=1}^{n} (X_{il}\mathcal{B}_i)^{\mathrm{T}} \mathcal{W}_i(t) (X_{ir}\mathcal{B}_i)$ as the $(r, l)^{th}$ block and r, $l = 0, \ldots, p$.

In addition to the local polynomial and smoothing spline methods proposed for the estimation in time-varying coefficient models, Fan and Zhang (2000) proposed a two-step estimation procedure, which is computationally faster and more effective in performance than the smoothing splines method. The two-step estimation procedure can be applied to unbalanced longitudinal data sets. Denote the distinct time points, where the data were collected, by t_j with $j = 1, \ldots, T$. In the first step a standard linear model is fitted to obtain the estimates $\mathbf{b}(t_j) = (b_1(t_j), \ldots, b_p(t_j))^T$ of the regression coefficient functions $\boldsymbol{\beta}(t_j) = (\beta_1(t_j), \ldots, \beta_p(t_j))^T$. These estimates are usually not smooth. Hence, in the second step a smoothing technique is applied to $[\{t_j, \mathbf{b}(t_j)\}, j = 1, \ldots, T]$ to obtain smooth estimates. In the second step any smoothing technique can be used. Another benefit of this estimation procedure is that different degrees of smoothness can be applied to different regression coefficient functions. Asymptotic properties of the local polynomial estimators (2.16) are derived by Wu et al. (1998). Methods on constructing confidence regions based on kernel methods are studied by Wu et al. (1998) and Wu and Chiang (2000). Huang et al. (2002) derived the confidence bands using the polynomial spline methods.

We can extend model (2.13) and define the generalized time-varying coefficient models,

$$\eta(t, \mathbf{X}) = g\{m(t, \mathbf{X})\} = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}(t),$$

where t denotes time as a covariate. This model is a version of the generalized varying coefficient model (2.9) with time as the covariate. The estimation techniques defined in Cai et al. (2000) can easily be used to estimate the regression coefficient functions in this model. For further details on statistical procedures with varying coefficient models, refer to Fan and Zhang (2008).

2.5 Partially Linear Models

Partially linear models are natural extensions of ordinary linear regression models with multiple predictors and nonparametric models with a single covariate. A partially linear model is defined as

$$Y = \alpha(U) + \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon, \tag{2.17}$$

where Y is the response variable, U and $\mathbf{X}$ are the covariates, $\alpha(\cdot)$ is a nonparametric smooth baseline function, $\boldsymbol{\beta}$ is an unknown regression coefficient vector, and ε is a random error with $E(\varepsilon|U,\mathbf{X}) = 0$.

In literature this model has been well studied for independent data (Engle et al., 1986; Härdle et al., 2000; Heckman, 1986; Speckman, 1988). Heckman (1986) used splines to estimate $\alpha(\cdot)$ and $\boldsymbol{\beta}$. Moreover, Heckman (1986) established that under mild conditions on the covariates and the error term, $\hat{\boldsymbol{\beta}}$ is asymptotically normally distributed in balanced cases of analysis of covariance and showed that its bias is asymptotically negligible. Rice (1986) studied asymptotic properties of $\hat{\boldsymbol{\beta}}$. Rice proved that in unbalanced cases, unless the nonparametric component is undersmoothed, the bias of $\hat{\boldsymbol{\beta}}$ can asymptotically dominate the variance when covariates U and $\mathbf{X}$ are correlated. This undesirable result motivated Speckman (1988) to introduce two estimation procedures for (2.17). The first one is related to smoothing splines and the second one is referred as a partial residual approach or more generally a profile least squares method. Let us briefly describe this approach. Note that

$$E(Y|U) = \alpha(U) + E(\mathbf{X}|U)^{\mathrm{T}}\boldsymbol{\beta},$$

and then using (2.17), we obtain

$$Y - E(Y|U) = \{\mathbf{X} - E(\mathbf{X}|U)\}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon. \tag{2.18}$$

Speckman (1988) proposed smoothing the response Y and the covariate $\mathbf{X}$ over the covariate U to obtain estimates of $E(Y|U)$ and $E(\mathbf{X}|U)$. After plugging these estimates into (2.18), we can estimate $\boldsymbol{\beta}$. We substitute $\hat{\boldsymbol{\beta}}$ into (2.17) and then use a nonparametric estimation technique to estimate the nonparametric function $\alpha(\cdot)$. Speckman (1988) also observed the negative result pointed out by Rice (1986) about the estimators obtained by partial smoothing spline approach. However, the bias of the profile least squares estimator is of lower order and asymptotically negligible with no need of undersmoothing the nonparametric component (Speckman, 1988).

Zeger and Diggle (1994) and Moyeed and Diggle (1994) extended the partially linear models to longitudinal data settings. For the i^{th} subject, $i = 1, \ldots, n$, denote the response measured at time point t_{ij} by $Y_i(t_{ij})$ and the vector of predictors by $\mathbf{X}_i(t_{ij}) = (X_{i1}(t_{ij}), \ldots, X_{ip}(t_{ij}))^{\mathrm{T}}$, where $j = 1, \ldots, n_i$. A partially linear model for longitudinal data is defined as follows:

$$Y_i(t_{ij}) = \alpha(t_{ij}) + \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\beta} + \varepsilon_i(t_{ij}). \tag{2.19}$$

Zeger and Diggle (1994) proposed a backfitting algorithm to estimate $\alpha(\cdot)$ and $\boldsymbol{\beta}$. Each iteration of the algorithm has the following steps:

1. Given the generalized least squares estimate $\hat{\boldsymbol{\beta}}^{(k)}$ at the k^{th} iteration, we fit a nonparametric model to the residuals $Y_i(t_{ij}) - \mathbf{X}_i^{\mathrm{T}}(t_{ij})\hat{\boldsymbol{\beta}}^{(k)}$, and obtain $\hat{\alpha}^{(k)}(\cdot)$.

2. Given $\hat{\alpha}^{(k)}(\cdot)$, we calculate $\widetilde{Y}_i(t_{ij}) = Y_i(t_{ij}) - \hat{\alpha}^{(k)}(t_{ij})$. To obtain the generalized least squares estimate $\hat{\beta}^{(k+1)}$, we fit the model $\widetilde{Y}_i(t_{ij}) = \mathbf{X}_i^{\mathrm{T}}(t_{ij})\hat{\boldsymbol{\beta}} + \varepsilon_i(t_{ij})$.

Iterate this procedure until it converges. Moyeed and Diggle (1994) suggested an estimation procedure that improves the backfitting algorithm, which extends the partial residual approach (Speckman, 1988). By undersmoothing the estimate of $\boldsymbol{\alpha}(t)$ and by adjusting the covariate $\mathbf{X}$ for the dependence on covariate t, their method achieves a lower order bias compared to the backfitting estimator.

Hu et al. (2004) presented theoretical comparisons of the two estimators: backfitting estimator (Zeger and Diggle, 1994) and the profile least squares estimator (Lin and Carroll, 2001). They assumed a working independence correlation matrix for the nonparametric estimation. Hu et al. (2004) showed that for the independent data, as Opsomer and Ruppert (1999) proved, the two estimators have the same asymptotic variance. However, for the clustered data, the asymptotic variances are equal if and only if the smoother matrix has a specific structure, for instance the nonparametric smoother proposed by Wang (2003).

Martinussen and Scheike (1999) and Lin and Ying (2001) proposed the counting process technique to estimate $\alpha(\cdot)$ and $\boldsymbol{\beta}$ in (2.19). The number of observations taken on the i^{th} subject by time t is characterized by the counting process: $N_i(t) = \sum_{j=1}^{n_i} I(t_{ij} \leq t)$, where $I(\cdot)$ is the indicator function. These observation times are considered as realizations from an arbitrary counting process that is censored at the end of follow-up. Specifically, $N_i(t) = N_i^*(t \wedge C_i)$, where $N_i^*(t)$ is a counting process in discrete or continuous time, C_i is the follow-up or censoring time, and $a \wedge b = min(a, b)$. Both the process $Y_i(t)$ and time-varying covariates $\mathbf{X}_i(t)$ were observed only at the jump points of $N_i(t)$. The censoring time C_i is allowed to depend on the covariates $\mathbf{X}_i(\cdot)$ in an arbitrary manner.

Lin and Ying (2001) required the censoring to be noninformative in the sense that $E\{Y_i(t)|\mathbf{X}_i(t), C_i \geq t\} = E\{Y_i(t)|\mathbf{X}_i(t)\}$. Lin and Ying (2001) proposed minimizing the weighted least squares function

$$\sum_{i=1}^{n} \int_0^{\infty} w(t)\{Y_i(t) - \alpha(t) - \mathbf{X}_i^{\mathrm{T}}(t)\boldsymbol{\beta}\}^2 dN_i(t), \tag{2.20}$$

where $w(\cdot)$ is possibly a data-dependent weight function. Lin and Ying (2001) developed estimation procedures for two situations depending on whether the observation times depend on covariates or not. When the observation times depend on the covariates, it is assumed that

$$E\{dN_i^*(t)|\mathbf{X}_i(t), Y_i(t), C_i \geq t\} = \exp\{\gamma \mathbf{X}_i(t)\} d\Lambda(t), \tag{2.21}$$

where γ is a vector of unknown parameters, $\Lambda(\cdot)$ is an arbitrary nondecreasing function, and $i = 1, \ldots, n$. The observation times are independent of covariate, when $\gamma = 0$. Let

$$\begin{aligned} \bar{\mathbf{x}}(t,\gamma) &= \frac{\sum_{i=1}^{n} \xi_i(t) \exp\{\gamma^{\mathrm{T}} \mathbf{X}_i(t)\} \mathbf{X}_i(t)}{\sum_{i=1}^{n} \xi_i(t) \exp\{\gamma^{\mathrm{T}} \mathbf{X}_i(t)\}} \quad \text{and} \\ \bar{y}(t,\gamma) &= \frac{\sum_{i=1}^{n} \xi_i(t) \exp\{\gamma^{\mathrm{T}} \mathbf{X}_i(t)\} \mathbf{Y}_i(t)}{\sum_{i=1}^{n} \xi_i(t) \exp\{\gamma^{\mathrm{T}} \mathbf{X}_i(t)\}}, \end{aligned}$$

where $\xi_i(t) = I(C_i \geq t)$. By replacing $dN_i(t)$ in (2.20) with its expectation (2.21), (2.20) becomes

$$\sum_{i=1}^{n} \int_0^{\infty} w(t)\{Y_i(t) - \alpha(t) - \mathbf{X}_i^{\mathrm{T}}(t)\boldsymbol{\beta}\}^2 \xi_i(t) \exp\{\gamma^{\mathrm{T}} \mathbf{X}_i(t)\} d\Lambda(t).$$

We minimize the equation above for each given $\boldsymbol{\beta}$ and γ, which is equivalent to minimizing it at each given time t. Thus, we estimate the baseline function as follows:

$$\hat{\alpha}(t; \boldsymbol{\beta}, \gamma) = \bar{y}(t,\gamma) - \bar{\mathbf{x}}(t,\gamma)^{\mathrm{T}} \boldsymbol{\beta}.$$

Replacing $\alpha(t)$ in (2.20) with $\hat{\alpha}(t;\boldsymbol{\beta},\gamma)$ results in

$$\ell(\boldsymbol{\beta},\gamma)=\sum_{i=1}^{n}\int_0^{\infty} w(t)\left[\{Y_i(t)-\bar{y}(t,\gamma)\}-\{\mathbf{X}_i(t)-\bar{\mathbf{x}}(t,\gamma)\}^{\mathrm{T}}\boldsymbol{\beta}\right]^2 dN_i(t), \tag{2.22}$$

where γ is unknown, but can be consistently estimated by solving

$$\sum_{i=1}^{n}\int_0^{\infty} w(t)\{\mathbf{X}_i(t)-\bar{\mathbf{x}}(t,\gamma)\}dN_i(t)=0.$$

Given $\hat{\gamma}$, $\hat{\boldsymbol{\beta}}$ can be obtained. Note that (2.22) requires the response $Y_i(t)$ and the covariate $\mathbf{X}_i(t)$ to be fully observable; however, in reality this assumption would not hold. Hence, Lin and Ying (2001) modified their method by replacing $Y_i(t)$ and $\mathbf{X}_i(t)$ with their corresponding values at the nearest time where their values are observed. This adjustment helps in practice, but it introduces nonnegligible bias due to the nearest neighborhood approximations.

Fan and Li (2004) proposed two estimation procedures, namely, difference-based method and profile least squares method, to improve the efficiency and eliminate nonnegligible bias of Lin and Ying's method. Here, we adopt Fan and Li's notation. Dropping the subscript j, the observed data can be expressed as $\{t_i,\mathbf{X}_i^{\mathrm{T}},\mathbf{Y}_i\}$ with $i=1,\ldots,n^*$ and $n^*=\sum_{i=1}^{n} n_i$. These data are ordered according to time t_{ij}. We can rewrite the partially linear model with this new notation as follows:

$$Y_i=\alpha(t_i)+\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}+\varepsilon_i, \tag{2.23}$$

where $E(\varepsilon_i|\mathbf{X}_i)=0$, and the covariates $\mathbf{X}_i$ and t_i are assumed to be independent. Let us briefly present the difference-based method. Observe that

$$Y_{i+1}-Y_i=\alpha(t_{i+1})-\alpha(t_i)+(\mathbf{X}_{i+1}-\mathbf{X}_i)^{\mathrm{T}}\boldsymbol{\beta}+e_i, \tag{2.24}$$

where $e_i=\varepsilon_{i+1}-\varepsilon_i$ and $i=1,\ldots,n^*-1$. In (2.24), the term $\{\alpha(t_{i+1})-\alpha(t_i)\}$ is negligible, since under some mild conditions, the spacing between $\{(t_{i+1})-(t_i)\}$

is $O(1/n)$. The estimate of $\boldsymbol{\beta}$ can be obtained by fitting the following linear model:

$$Y_{i+1} - Y_i = \alpha_0 + \alpha_1(t_{i+1} - t_i) + (\mathbf{X}_{i+1} - \mathbf{X}_i)^{\mathrm{T}}\boldsymbol{\beta} + e_i.$$

According to the simulation results presented in Fan and Li (2004), this method performs better than the technique proposed in Lin and Ying (2001). Next, we briefly describe the profile least squares method.

For a given $\boldsymbol{\beta}$, let $Y^*(t) \equiv Y(t) - \mathbf{X}(t)^{\mathrm{T}}\boldsymbol{\beta}$. Then rewrite the partially linear model (2.19) as

$$Y^*(t) = \alpha(t) + \varepsilon(t),$$

which is a nonparametric regression problem. Fan and Li (2004) suggested using a local linear regression technique (Fan and Gijbels, 1996) to obtain $\hat{\alpha}(\cdot)$. Explicit expressions for $\hat{\alpha}(.)$ and $\hat{\boldsymbol{\beta}}$ are given in matrix form. The model in (2.23) can be written in matrix form as follows:

$$\mathbf{Y} = \boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.25}$$

where $\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \ldots, \mathbf{Y}_n^{\mathrm{T}})$ with $\mathbf{Y}_i = (Y_i(t_{i1}), \ldots, Y_i(t_{in_i}))^{\mathrm{T}}$, $\mathbf{X} = (\mathbf{X}_1^{\mathrm{T}}, \ldots, \mathbf{X}_n^{\mathrm{T}})^{\mathrm{T}}$ with $\mathbf{X}_i = (\mathbf{X}_i(t_{i1}), \ldots, \mathbf{X}_i(t_{in_i}))^{\mathrm{T}}$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^{\mathrm{T}}, \ldots, \boldsymbol{\alpha}_n^{\mathrm{T}})^{\mathrm{T}}$ with $\boldsymbol{\alpha}_i = (\alpha(t_{i1}), \ldots, \alpha(t_{in_i}))^{\mathrm{T}}$, and $\boldsymbol{\varepsilon}$ is the vector of error terms. Since the local linear fit is linear in $y_i^*(t_{ij})$ (Fan and Gijbels, 1996), the estimate of $\alpha(t)$ is linear in $\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$. Thus, the estimate for $\boldsymbol{\alpha}$ is $\hat{\boldsymbol{\alpha}} = \mathbf{S}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$, where $\mathbf{S}$ is called a smoothing matrix of the local linear smoother. $\mathbf{S}$ depends on the observation times and the bandwidth h. Let $\mathbf{I}$ be the identity matrix of order n^*, substitute $\hat{\boldsymbol{\alpha}}$ into (2.25),

$$(\mathbf{I} - \mathbf{S})\mathbf{Y} = (\mathbf{I} - \mathbf{S})\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{2.26}$$

By applying the weighted least squares to model (2.26), the estimator of $\boldsymbol{\beta}$ is obtained as

$$\hat{\boldsymbol{\beta}} = \{\mathbf{X}^{\mathrm{T}}(\mathbf{I} - \mathbf{S})^{\mathrm{T}}\mathbf{W}(\mathbf{I} - \mathbf{S})\mathbf{X}\}^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{I} - \mathbf{S})^{\mathrm{T}}\mathbf{W}(\mathbf{I} - \mathbf{S})\mathbf{Y},$$

where $\mathbf{W}$ is the weight matrix, called a working covariance matrix. $\mathbf{W}$ is a diagonal matrix, when working independence is assumed. The estimate for covariance matrix $\hat{\boldsymbol{\beta}}$ is derived as

$$\text{cov}\{\hat{\boldsymbol{\beta}}|t_{ij}, \mathbf{X}_i(t_{ij})\} = \mathbf{D}^{-1}\mathbf{V}\mathbf{D}^{-1},$$

where $\mathbf{D} = \mathbf{X}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})^{\mathrm{T}}\mathbf{W}(\mathbf{I}-\mathbf{S})\mathbf{X}$ and $\mathbf{V} = \text{cov}\{\mathbf{X}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})^{\mathrm{T}}\mathbf{W}\boldsymbol{\varepsilon}\}$. $\mathbf{V}$ can be estimated by

$$\hat{\mathbf{V}} = \mathbf{X}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})^{\mathrm{T}}\mathbf{W}\mathbf{C}\mathbf{W}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\mathbf{X},$$

where $\mathbf{C} = \text{diag}(\hat{\boldsymbol{\varepsilon}}_1\hat{\boldsymbol{\varepsilon}}_1^{\mathrm{T}}, \ldots, \hat{\boldsymbol{\varepsilon}}_n\hat{\boldsymbol{\varepsilon}}_n^{\mathrm{T}})$ with $\hat{\boldsymbol{\varepsilon}}_i$ as the residual vector for the i^{th} subject. Fan and Li (2004) showed the asymptotic normality of the profile least squares estimator.

2.6 Joint Models for Binary and Continuous Responses

A number of models have been developed for the joint distribution of continuous and binary response variables. For joint modeling, a commonly used technique is to introduce a partly observed random variable following a bivariate normal distribution. One of the components of this random variable defines the continuous response and the second, the latent component defines the binary response through a probit transformation. Catalano and Ryan (1992), Cox and Wermuth (1992), Fitzmaurice and Laird (1995), Sammel et al. (1997), Regan and Catalano (1999), Dunson (2000) and Gueorguieva and Agresti (2001) are among the authors who used this approach, i.e. introducing a latent variable, to derive the joint distribution of a continuous response and a binary response. Catalano and Ryan (1992) and Cox and Wermuth (1992) factorized the joint model into two components: a marginal model for the continuous response and a correlated probit model for the binary response. Here, we briefly summarize the factorization approach proposed by Catalano and Ryan (1992). As mentioned above, latent variable models assume that an unobserved continuous variable exists and that the binary event occurs only if this continuous variable exceeds a certain level. As to be seen below, one advantage of the latent variable approach is that it provides us with a convenient form for the binary event under the normality assumption. For the i^{th} subject, let W_i be the continuous variable and Y_i be the unobserved latent variable corresponding to the binary variable

$$\begin{aligned} W_i &= \alpha_0 + \alpha_1 X_i + \varepsilon_{1i}, \\ Y_i &= \beta_0 + \beta_1 X_i + \varepsilon_{2i}, \end{aligned}$$

where $i = 1, \ldots, n$. Let

$$\boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim \mathcal{N}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \tau\sigma_1\sigma_2 \\ \tau\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right\},$$

where τ denotes the correlation between W_i and Y_i. Denote the binary variable by $Q_i = 1$ if $Y_i > 0$ and $Q_i = 0$ if $Y_i \leq 0$. The probit model for Q_i is defined as

$$P(Q_i = 1) = \Phi\left(\frac{\beta_0 + \beta_1 X_i}{\sigma_2}\right).$$

The joint distribution of W_i and Q_i is not easy to derive; however, the marginal distributions of these variables are readily obtained. Hence, we decompose the joint distribution into two components: a marginal distribution and a conditional distribution,

$$f_{W_i,Q_i}(w, q) = f_{W_i}(w) f_{Q_i|W_i}(q|w).$$

To derive the conditional distribution $Q_i|W_i$, Catalano and Ryan (1992) started with obtaining the conditional distribution $Y_i|W_i$. As standard normal theory shows, like the marginal distribution, the conditional distribution $Y_i|W_i$

also follows a Gaussian distribution. The mean of this distribution depends on the residuals from the marginal model for the continuous response,

$$Y_i|W_i \sim N\left\{\mu_1, \sigma_2^2(1-\tau^2)\right\},$$

where $\mu_1 = \beta_0 + \beta_1 X_i + \left(\frac{\sigma_2}{\sigma_1}\right)\tau e_{1i}$ and $e_{1i} = W_i - (\alpha_0 + \alpha_1 X_i)$ is the residual from the marginal model. Hence, the model for conditional distribution of $Q_i|W_i$ is probit,

$$P(Q_i = 1|W_i) = \Phi\left(\frac{\mu_1}{\sqrt{\sigma_2^2(1-\tau^2)}}\right).$$

Not all parameters in this model are estimable, hence Catalano and Ryan (1992) reparameterized and obtained the following:

$$P(Q_i = 1|W_i) = \Phi(\beta_0^* + \beta_1^* X_i + \beta_2^* e_{1i}). \tag{2.27}$$

Model (2.27) links the continuous response with the binary response in a probit regression model using the residuals from the marginal model as a covariate. From the conditional model, it can be seen that $\beta_2^* = (\sigma_2/\sigma_1)\tau$, and assuming that $\sigma_1^2 = \sigma_2^2$, then τ would be estimable. Catalano and Ryan (1992) proposed a two-step algorithm for fitting the joint distribution on observed data. The first step is to apply generalized estimating equations (GEE) (Liang and Zeger, 1986; Zeger and Liang, 1986) to fit the continuous response on the predictor(s). The second step involves applying GEE to fit the correlated probit regression of $Q|W$ using residuals from the marginal model as one of the covariates. The benefit of using GEE approach is that it allows to correct for the mismodeling of the covariance structure without losing much efficiency.

Fitzmaurice and Laird (1995) also proposed a two-stage method; however, they reversed the conditioning order. They specified a marginal model for the binary response and a conditional model for the continuous response given the binary response. The joint density of the binary and continuous response is given as

$$f_{W_i,Q_i}(w_i, q_i) = f_{Q_i}(q_i) f_{W_i|Q_i}(w_i|q_i).$$

To describe their general approach and introduce their notation, we first review the case when the units within clusters are assumed to be independent. The marginal distribution of the binary response is assumed to be Bernoulli,

$$f(q_i|\mathbf{X}_i) = \exp\left[q_i\theta_i - \log\left\{1 + \exp(\theta_i)\right\}\right], \tag{2.28}$$

where $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})$, $\theta_i = \log\left\{\frac{\mu_{1i}}{(1-\mu_{1i})}\right\} = \mathbf{X}_i\boldsymbol{\beta}_1$, $\mu_{1i} = E(Q_i) = P(Q_i = 1|\mathbf{X}_i, \boldsymbol{\beta}_1)$ and $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{1p})^{\mathrm{T}}$. Another assumption is that the conditional distribution $W_i|Q_i$ follows a Gaussian distribution:

$$f_{W_i|Q_i}(w_i|q_i) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}\left\{w_i - \mu_{2i} - \gamma(q_i - \mu_{1i})\right\}^2\right], \tag{2.29}$$

where $\mu_{2i} = \mathbf{X}_i\boldsymbol{\beta}_2$ with $\boldsymbol{\beta}_2 = (\beta_{21}, \ldots, \beta_{2p})^{\mathrm{T}}$ and γ is a parameter from the regression of the continuous response on the binary response. As model (2.29) shows the mean of the continuous response depends on the binary response, whereas in model (2.27) it is the reverse. Let $\boldsymbol{\alpha} = (\boldsymbol{\beta}_2, \gamma)$, a $(p+1) \times 1$ vector of parameters and $\mathbf{Z}_i = (\mathbf{X}_i, (Q_i - \mu_{1i}))$, a $1 \times (p+1)$ vector, then

$$E(W_i|Q_i) = \mathbf{Z}_i\boldsymbol{\alpha},$$

and by construction $E(Q_i) = \mathbf{X}_i\boldsymbol{\beta}_2$. Hence, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are regression parameters with marginal interpretations. Fitzmaurice and Laird (1995) developed a likelihood-based approach to estimate $(\boldsymbol{\alpha}, \sigma^2)$ and $\boldsymbol{\beta}_1$ jointly. Regardless of whether or not the likelihood has been correctly specified, the estimate $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$ are consistent. However, when the likelihood has been misspecified, we may obtain inconsistent estimates of the asymptotic variance of the estimate $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$.

Fitzmaurice and Laird (1995) extended the approach described above to allow for clustering. For the i^{th} cluster the responses are denoted by $(\mathbf{W}_i, \mathbf{Q}_i)$ with $\mathbf{W}_i = (W_{i1}, \ldots, W_{in_i})^{\mathrm{T}}$ and $\mathbf{Q}_i = (Q_{i1}, \ldots, Q_{in_i})^{\mathrm{T}}$; the covariate matrix is denoted by $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i})^{\mathrm{T}}$ with $\mathbf{X}_{ij}$ as the $1 \times p$ covariate vector for each unit within a cluster and $i = 1, \ldots, n$. Assume that the model for the mean is

$$\begin{aligned} \operatorname{logit}\{E(\mathbf{Q}_i)\} &= \mathbf{X}_i\boldsymbol{\beta}_1, \\ E(W_{ij}|\mathbf{Q}_i) &= \mathbf{X}_{ij}\boldsymbol{\beta}_2 + \gamma_1(Q_{ij} - \mu_{1_{ij}}) + \gamma_2 S_i, \end{aligned}$$

where $\operatorname{logit}\{E(\mathbf{Q}_i)\} = \big(\operatorname{logit}(\mu_{1_{i1}}), \ldots, \operatorname{logit}(\mu_{1_{in_i}})\big)^{\mathrm{T}}$ and $S_i = \sum_{j=1}^{n_i}(Q_{ij} - \mu_{1_{ij}})$. In this model γ_2 characterizes the association between the members of the same cluster (intracluster) and $(\gamma_1 + \gamma_2)$ characterizes the association between the binary and continuous responses of the same unit within a cluster. Hence, $\boldsymbol{\Gamma} = (\gamma_1, \gamma_2)$ denotes the parameter vector for the correlation between $\mathbf{Q}_i$ and $\mathbf{W}_i$. To simplify the notation, let $\boldsymbol{\alpha} = (\boldsymbol{\beta}_2, \boldsymbol{\Gamma})$ be a $(p+2) \times 1$ vector of parameters. Maximum likelihood approach is complicated for this model, since the model structure in (2.28) and (2.29) is not the same for clusters with different sizes and thus can only be used when all clusters have the same size. One can find other joint model representations that can handle different cluster sizes; however, the likelihood equations for regression parameters in these representations are not robust to misspecification of the distributional assumptions. Another problem with a full likelihood-based approach is that the number of nuisance parameters we need to estimate increases as the cluster size increases. Hence, Fitzmaurice and Laird (1995) proposed using GEE approach (Liang and Zeger, 1986; Zeger and Liang, 1986), which is a computationally convenient method, to estimate $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. They pointed out that the estimate $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$ is consistent and asymptotically normally distributed, as long as the model for the marginal expectation of the responses has been correctly specified.

In addition to the two factorization-based methodologies described above, Sammel et al. (1997) proposed another method for joint modeling, which was not based on factorization. They introduced the latent variable mixed effects model, which is a generalization of the previous work on latent variable models for continuous (Sammel and Ryan, 1996) and discrete responses (Legler and Ryan, 1997). Assume that there are M outcomes for each individual. Let $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iM})^{\mathrm{T}}$ be the data observed on individual i and $i = 1, \ldots, n$. Sammel et al. (1997) assumed that each manifest variable $\mathbf{Z}_i$ is from a one-parameter exponential family and the manifest variables are connected to the subject-specific latent variable Y_i and covariates $\mathbf{x}_i$ via a generalized linear model,

$$f_m(Z_{im}|Y_i) = \exp\left[\left\{Z_{im}\eta_m - d(\eta_m)\right\}/\phi_m + c_m(Z_{im}, \phi_m)\right] \text{ for } m = 1, \ldots, M, \tag{2.30}$$

with

$$\begin{aligned} \mu_m &= E(Z_{im}) = d'(\eta_m), \\ V_m &= \mathrm{var}(Z_{im}) = d''(\eta_m), \end{aligned}$$

where ϕ_m is the scale factor and may not be known depending on the assumed distribution (McCullagh and Nelder, 1989); and $\eta_m = g_m(\mu_m)$ is a linear function of y_i and $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$, p fixed covariates, such that

$$\eta_m = \beta_{0m} + \beta_{1m}y_i + \beta_{2m}x_{i1} + \ldots + \beta_{(1+p)m}x_{ip} = \mathcal{X}_i^{\mathrm{T}}\boldsymbol{\beta}_m,$$

where $\mathcal{X}_i = (1, Y_i, x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ and $\boldsymbol{\beta}_m = \left(\beta_{0m}, \ldots, \beta_{(p+1)m}\right)^{\mathrm{T}}$. Model (2.30) is flexible enough to use different link functions. Of interest is to estimate the parameters $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_M)$ that describes the relationship between the response, the latent variable and the fixed covariates for each individual. These parameters are estimated via the marginal distribution of $\mathbf{Z}_i$,

$$f(\mathbf{z}_i) = \int f(\mathbf{z}_i|y_i)f(y_i)dy_i. \tag{2.31}$$

Sammel et al. (1997) proposed a modified EM algorithm to fit the model. In the expectation step Sammel et al. (1997) aimed to compute the expectations with respect to the posterior distribution of the missing data, conditionally on the observed data $h(y_i|\mathbf{z}_i)$. The challenge in this step is that; even though, Sammel et al. (1997) assume a single latent variable, the integration in (2.31) does not have a closed form for many choices of $f(\mathbf{z}|y)$. Moreover, we need to evaluate the same integration to obtain the posterior distribution of y_i. Hence, Sammel et al. (1997) proposed using approximations whenever necessary to be able to compute the necessary expectations with respect to the posterior distribution. They suggested two approaches. One of them is the Monte Carlo approximation (Wei and Tanner, 1990). The expectation step of this approach can be considered as a multiple imputation of the missing data at each step. In this step, if we could generate a sample of ys from the posterior distribution, then the expectation could be estimated easily,

$$E_y g(y) = \int g(\mathbf{z}_i, y)h(y|\mathbf{z}_i)dy \approx \frac{1}{N}\sum_{t=1}^{T} g(\mathbf{z}_i, b_t), \tag{2.32}$$

where N is the sample size (Tanner, 1991). However, as we mentioned, the posterior distribution of $h(y_i|\mathbf{z}_i)$ does not have a closed form expression and thus we have to write the posterior distribution in terms of the marginal distribution of the latent variable $f(y)$,

$$h(y|\mathbf{z}_i) = \frac{f(\mathbf{z}_i|y)f(y)}{\int f(\mathbf{z}_i|y)f(y)dy}. \tag{2.33}$$

Substituting (2.33) in (2.32), we obtain

$$E_y g(y) = \frac{\int g(\mathbf{z}_i, y)f(y|\mathbf{z}_i)f(y)dy}{\int f(y|\mathbf{z}_i)f(y)dy}.$$

Now, we can generate a sample of ys from the distribution $f(y)$ and compute the expectation via Monte Carlo approximation. However, the Monte Carlo approximation is based on the weak law of large numbers and it requires a large sample of ys. Hence, this approach is quite slow. An alternative approach is to apply a numerical approximation via Gauss-Hermite quadrature (Abramowitz and Stegun, 1987). This approximation has a faster convergence. The expectation obtained using either approach is the input of the next maximization step. The authors pointed out that for the expectations step, importance sampling (Tanner, 1991) or Gibbs sampling (Zeger and Karim, 1991) can also be used to obtain the expectations; however these methods are computationally more intensive.

Another extension to the correlated probit model was proposed by Regan and Catalano (1999). They also followed the latent variable approach and assumed that the latent variable and the continuous variable follow a multivariate normal distribution. The marginal parameters for both binary and continuous responses and the correlation parameter were estimated through a maximum likelihood approach. The drawback in their estimation procedure is that it was assumed that the correlation structure between the continuous and the latent response is exchangeable. This assumption can be restrictive for some data sets.

A similar correlated probit model to Catalano and Ryan (1992) was studied by Gueorguieva and Agresti (2001). For the i^{th} cluster, denote continuous response at the j^{th} observation by W_{ij} and the latent response by Y_{ij}, $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. They defined the following linear mixed model:

$$\begin{aligned} W_{ij} &= \mathbf{X}_{i1j}^{\mathrm{T}}\boldsymbol{\beta}_1 + \mathbf{Z}_{i1j}\mathbf{b}_{i1} + \varepsilon_{i1j}, \\ Y_{ij} &= \mathbf{X}_{i2j}^{\mathrm{T}}\boldsymbol{\beta}_2 + \mathbf{Z}_{i2j}\mathbf{b}_{i2} + \varepsilon_{i2j}, \end{aligned}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are unknown $p_1\times 1$ and $p_2\times 1$ parameter vectors and $\mathbf{X}_{i1j}, \mathbf{X}_{i2j}, \mathbf{Z}_{i1j}$

and $\mathbf{Z}_{i2j}$ are known $p_1 \times 1$, $p_2 \times 1$, $q_1 \times 1$ and $q_2 \times 1$ vectors. The random effects $\mathbf{b}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2})^{\mathrm{T}}$ and the random errors $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{i1j}, \varepsilon_{i2j})^{\mathrm{T}}$ are assumed to be normally distributed. Hence, conditional on the random effects, the model for the binary and continuous response was obtained as

$$\mu_{i1j} = \mathbf{X}_{i1j}^{\mathrm{T}}\boldsymbol{\beta}_1 + \mathbf{Z}_{i1j}^{\mathrm{T}}\mathbf{b}_{i1} \quad \text{and} \quad \Phi^{-1}(\mu_{i2j}) = \mathbf{X}_{i2j}^{\mathrm{T}}\boldsymbol{\beta}_2 + \mathbf{Z}_{i2j}^{\mathrm{T}}\mathbf{b}_{i2},$$

where μ_{i1j} and μ_{i2j} are conditional means for the continuous and binary response, respectively. There are several special cases. For the continuous response alone, the model is a general linear model with equicorrelated structure and for the binary response alone, the model is a multivariate probit model with equicorrelated structure. The model is a special case of Regan and Catalano (1999) with no modeling of covariance parameters, if the covariance between ε_{i1j} and ε_{i2j} is zero and the random effects structure consists only of the intercepts for each variable.

Gueorguieva and Agresti (2001) proposed a modified EM algorithm to obtain the maximum likelihood estimates for the correlated probit model. The first approach is the extension of the work by Chan and Kuk (1997) called Monte Carlo expectation-conditional maximization (ECM). The second approach is the parameter expanded (PX-EM) algorithm, which is another extension of the EM algorithm proposed by Liu et al. (1998). Let us present a summary of the advantages and disadvantages of these algorithms. First of all, ECM can take more iterations to converge than the EM algorithm. However, it can be faster computationally because it has a simpler maximization step. Similar to ECM, PX-EM also aims to simplify the maximization step. The challenge with these algorithms is the expectation step for the correlated probit model is computationally intensive. Even when the authors used both ECM and PX-EM, the algorithm was still slow. Hence, they proposed adapting the Monte Carlo EM algorithm proposed by Liao (1999) and Delyon et al. (1999), which is called stochastic approximation expectation-conditional maximization (SAECM) algorithm. This algorithm replaced the expectation step in the EM algorithm by one iteration of a stochastic approximation procedure. In their real-life application the authors showed that the SAEM algorithm using Gibbs sampler is much faster than the ECM and both algorithms experienced convergence problems when initial estimates were far from the true values.

In addition to the models introduced for the clustered data setting, Liu et al. (2009) proposed a method for joint modeling of longitudinal binary and continuous responses using the factorization approach. The main interest in Liu et al. (2009) was to model the association between a longitudinal binary response and a longitudinal continuous response unlike the previous studies. We briefly introduce the model proposed by Liu et al. (2009). Denote the binary response by Q_{ij} for subject i at time j and denote the continuous response by W_{ij} with $i = 1, \cdots, n$ and$j = 1, \cdots, J$. Define the vector of binary response as $\mathbf{Q}_i=(Q_{i1}, \cdots, Q_{iJ})^{\mathrm{T}}$ and the vector of continuous response as $\mathbf{W}_i=(W_{i1}, \cdots, W_{iJ})^{\mathrm{T}}$. In addition, define the vector of latent variables underlying $\mathbf{Q}_i$ as $\mathbf{Y}_i=(Y_{i1}, \cdots, Y_{iJ})^{\mathrm{T}}$. Let $\mathbf{V}_i$ be the vector of joint processes $\mathbf{V}_i = \left(\mathbf{Y}_i^{\mathrm{T}}, \mathbf{W}_i^{\mathrm{T}}\right)$. The joint distribution of binary and continuous variables over time was modeled using multivariate normal distribution $\left(\mathbf{Y}_i^{\mathrm{T}}, \mathbf{W}_i^{\mathrm{T}}\right)^{\mathrm{T}} \sim N\left(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i\right)$,

$$\mathbf{V}_i = \begin{pmatrix} \mathbf{Y}_i \\ \mathbf{W}_i \end{pmatrix} = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,$$

where

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{X}_{1i} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{2i} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \text{ and } \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i) \text{ with } \boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_{i,11} & \boldsymbol{\Sigma}_{i,12} \\ \boldsymbol{\Sigma}_{i,21} & \boldsymbol{\Sigma}_{i,22} \end{pmatrix}.$$

Liu et al. (2009) defined $Q_{ij} = I\{Y_{ij} > 0\}$ by using the probit formulation for the binary process.

Of interest is to model $\mathbf{\Sigma}_{i,12}$ as a function of the covariate(s). However, both $\mathbf{\Sigma}_{i,12}$ and $\mathbf{\Sigma}_i$ are difficult to model, because of positive definiteness constraints (Daniels and Kass, 2001; Pourahmadi and Daniels, 2002). Hence, Liu et al. (2009) proposed to decompose the joint distribution of the continuous variable and the latent variable into two components: a marginal model for the latent variable $\mathbf{Y}_i$ and a correlated regression model for $\mathbf{W}_i$ given $\mathbf{Y}_i$,

$$\begin{aligned}\mathbf{Y}_i &= \mathbf{X}_{1i}\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_{1i},\\ \mathbf{W}_i &= \mathbf{X}_{2i}\boldsymbol{\beta}_2 + \mathbf{B_i}\left(\mathbf{Y}_i - \mathbf{X}_{1i}\boldsymbol{\beta}_1\right) + \boldsymbol{\varepsilon}_{2i},\end{aligned}$$

where $\boldsymbol{\varepsilon}_{1i} \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}_{i,11})$, $\boldsymbol{\varepsilon}_{2i} \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}^*_{i,22})$ with $\mathbf{\Sigma}^*_{i,22} = \mathbf{\Sigma}_{i,22} - \mathbf{\Sigma}_{i,21}\mathbf{\Sigma}^{-1}_{i,11}\mathbf{\Sigma}_{i,12}$ and $\mathbf{B}_i = \mathbf{\Sigma}_{i,21}\mathbf{\Sigma}^{-1}_{i,11}$ is the matrix that shows the association between the $\mathbf{Y}_i$ and $\mathbf{W}_i$. The decomposition of the covariance matrix $\mathbf{\Sigma}_i$ into $\left(\mathbf{\Sigma}_{i,11}, \mathbf{B}_i, \mathbf{\Sigma}^*_{i,22}\right)$ is known as the Bartlett decomposition. The advantage of the Bartlett decomposition is that the components of the association matrix $\mathbf{B_i}$ are directly related to $\mathbf{\Sigma}_{i,12}$. Furthermore, since the components of $\mathbf{B}_i$ are unconstrained, a better reparameterization for examining the association between the latent variable and the continuous variable is obtained. However, the association matrix is high dimensional and expected to be sparse. Hence, Liu et al. (2009) proposed a Bayesian variable selection method to reduce the number of parameters. This method is an extension of the work by Smith and Kohn (2002). Furthermore, Liu et al. (2009) proposed an MCMC sampling algorithm that they address several issues such as efficiently sampling a correlation matrix from its full distribution. See Liu et al. (2009) for details of this algorithm.

Chapter 3

Time–varying Coefficient Models for Longitudinal Mixed Responses

In this chapter we present our nonparametric joint modeling approach. In Section 3.1 we develop our joint model for longitudinal binary and continuous responses. In Section 3.2 we describe our two-stage estimation procedure by using local linear regression techniques. In Section 3.3 we study asymptotic properties of the resulting estimators. In Section 3.4 we conduct a Monte Carlo simulation study to investigate the finite sample behavior of our estimators. In Section 3.5 we further illustrate the proposed methodology by analyzing the smoking cessation study (Shiffman et al., 1996). In Section 3.6 we present the technical conditions and proofs.

3.1 Joint Models

In this section we introduce the joint models for estimating the association of a longitudinal binary and a continuous response. For the i^{th} subject, $i = 1, \ldots, n$, denote the binary response measured at time point t_{ij} by $Q_i(t_{ij})$, the continuous response by $W_i(t_{ij})$, where $j = 1, \ldots, n_i$. In addition, define the latent variable underlying $Q_i(t_{ij})$ by $Y_i(t_{ij})$. Let $\mathbf{X}_i(t_{ij}) = (X_{i1}(t_{ij}), \ldots, X_{ip}(t_{ij}))^{\mathrm{T}}$ be the vector of predictors with $X_{i1} \equiv 1$ to include an intercept term, $\boldsymbol{\beta}(t_{ij}) = (\beta_1(t_{ij}), \ldots, \beta_p(t_{ij}))^{\mathrm{T}}$, and $\boldsymbol{\alpha}(t_{ij}) = (\alpha_1(t_{ij}), \ldots, \alpha_p(t_{ij}))^{\mathrm{T}}$ be the vectors of regression coefficients. Consider the following bivariate model:

$$\begin{aligned} W_i(t_{ij}) &= \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\beta}(t_{ij}) + \varepsilon_{1i}(t_{ij}), \\ Y_i(t_{ij}) &= \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\alpha}(t_{ij}) + \varepsilon_{2i}(t_{ij}), \end{aligned} \tag{3.1}$$

where $\varepsilon_{1i}(t_{ij}) \sim \mathcal{N}\{0, \sigma_1^2(t_{ij})\}$ with $\text{corr}\{\varepsilon_{1i}(t_{ij}), \varepsilon_{1i}(t_{ik})\} = \rho_1(t_{ij}, t_{ik})$ for $j \neq k$, $\varepsilon_{2i}(t_{ij}) \sim \mathcal{N}\{0, \sigma_2^2(t_{ij})\}$ with $\text{corr}\{\varepsilon_{2i}(t_{ij}), \varepsilon_{2i}(t_{ik})\} = \rho_2(t_{ij}, t_{ik})$ for $j \neq k$, and $j, k = 1, \ldots, n_i$. Furthermore, we define $\boldsymbol{\varepsilon}_i(t_{ij}) = (\boldsymbol{\varepsilon}_{1i}(t_{ij}), \boldsymbol{\varepsilon}_{2i}(t_{ij}))^{\mathrm{T}}$, and assume that $\boldsymbol{\varepsilon}_i(t_{ij})$ follows a bivariate normal distribution with $\text{corr}\{\varepsilon_{1i}(t_{ij}), \varepsilon_{2i}(t_{ij})\} = \tau(t_{ij})$, and $\text{corr}\{\varepsilon_{1i}(t_{ij}), \varepsilon_{2i}(t_{ik})\} = \rho_{12}(t_{ij}, t_{ik})$ for $j \neq k$. The relation between the latent variable and the binary variable is defined as $Q_i(t_{ij}) = 1$ if $Y_i(t_{ij}) > 0$, and $Q_i(t_{ij}) = 0$ if $Y_i(t_{ij}) \leq 0$. Thus, the probit model for the binary response $Q_i(t_{ij})$ is defined as

$$P\{Q_i(t_{ij}) = 1 \mid \mathbf{X}_i(t_{ij})\} = \Phi\left\{\frac{\mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\alpha}(t_{ij})}{\sigma_2(t_{ij})}\right\}.$$

The joint distribution of $W_i(t_{ij})$ and $Q_i(t_{ij})$ is challenging to derive; however, the marginal distributions are readily obtained. Thus, we factorize the joint distribution of the continuous variable and the binary variable into two components: a marginal model for the continuous variable $W_i(t_{ij})$ and a conditional model for $Q_i(t_{ij})$ given $W_i(t_{ij})$,

$$f\{q_i(t_{ij}), w_i(t_{ij})\} = f_W\{w_i(t_{ij})\} f\{q_i(t_{ij}) \mid w_i(t_{ij})\},$$

where $j = 1, \ldots, n_i$. The marginal model for the continuous response is defined in (3.1). To derive the conditional model for $Q_i(t_{ij})$ given $W_i(t_{ij})$, we start by obtaining the conditional model $Y_i(t_{ij}) \mid W_i(t_{ij})$. As standard normal theory shows the conditional distribution $Y_i(t_{ij}) \mid W_i(t_{ij})$ follows a Gaussian distribution. The mean of this distribution depends on the error from the marginal model of the continuous response,

$$Y_i(t_{ij}) \mid W_i(t_{ij}) \sim \mathcal{N}\left[\mu_i(t_{ij}), \sigma_2^2(t_{ij})\left\{1-\tau(t_{ij})\right\}\right],$$

where

$$\mu_i(t_{ij}) = \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\alpha}(t_{ij}) + \frac{\sigma_2(t_{ij})}{\sigma_1(t_{ij})}\tau(t_{ij})\varepsilon_{1i}(t_{ij}), \tag{3.2}$$

and

$$\varepsilon_{1i}(t_{ij}) = W_i(t_{ij}) - \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\beta}(t_{ij})$$

is the error from the marginal model of the continuous response. Thus,

$$P\left\{Q_i(t_{ij}) = 1 \mid W_i(t_{ij})\right\} = \Phi\left[\frac{\mu_i(t_{ij})}{\sqrt{\sigma_2^2(t_{ij})\left\{1-\tau(t_{ij})\right\}}}\right], \tag{3.3}$$

where $\mu_i(t_{ij})$ is defined in (3.2). Not all parameters in model (3.3) are estimable, hence we reparameterize (3.3) to a more parsimonious and fully estimable form:

$$P\left\{Q_i(t_{ij}) = 1 \mid W_i(t_{ij})\right\} = \Phi\left\{\mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\alpha}^*(t_{ij}) + \alpha_{p+1}^*(t_{ij})\varepsilon_{1i}(t_{ij})\right\}, \tag{3.4}$$

where $\boldsymbol{\alpha}^*(t_{ij}) = \left(\alpha_1^*(t_{ij}), \ldots, \alpha_p^*(t_{ij})\right)^{\mathrm{T}}$. Model (3.4) links the continuous response with the binary response in a probit regression model using the error from the marginal model as a covariate. From the conditional model, it can be seen that $\alpha_{p+1}^*(t_{ij}) = \left\{\sigma_2(t_{ij})/\sigma_1(t_{ij})\right\}\tau(t_{ij})$, and assuming that $\sigma_1^2(t_{ij}) = \sigma_2^2(t_{ij})$, then $\tau(t_{ij})$ would be estimable. Hence assuming the equality of the variances enables us to estimate the time-varying association, $\tau(t_{ij})$. However, even if the equality of variances is not assumed, the proposed methodology can reveal the significance and the direction of the time-varying association, and its estimate up to the positive factor $\left\{\sigma_2(t_{ij})/\sigma_1(t_{ij})\right\}$.

3.2 Estimation Procedure

We propose a two-stage estimation procedure to estimate the time-varying correlation between a longitudinal binary and a continuous response. This technique also allows us to estimate the regression coefficients in the marginal model of the continuous response. In the first stage we fit a time-varying coefficient model (Brumback and Rice, 1998; Hoover et al., 1998; Fan and Zhang, 2000; Wu et al., 1998) to the marginal model of the continuous response (3.1). In this stage we adapt the varying coefficient model (Zhang and Lee, 2000) estimation techniques to a longitudinal setting. We employ the local linear fitting (Fan and Gijbels, 1996), because of its nice properties such as its minimax efficiency and being design-adaptive (Fan, 1993). We locally approximate the regression coefficient functions in a neighborhood of a fixed point t_0 via the Taylor expansion,

$$\beta_r(t) \approx \beta_r(t_0) + \beta_r'(t_0)(t - t_0) \equiv a_r + b_r(t - t_0),$$

for $r = 1, \ldots, p$. Let $\mathbf{a} = (a_1, \ldots, a_p)^{\mathrm{T}}$, and $\mathbf{b} = (b_1, \ldots, b_p)^{\mathrm{T}}$. We minimize the following equation:

$$\ell(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{ W_i(t_{ij}) - \mathbf{X}_i^{\mathrm{T}}(t_{ij})\mathbf{a} - \mathbf{X}_i^{\mathrm{T}}(t_{ij})\mathbf{b}(t_{ij} - t_0) \right\}^2 K_h(t_{ij} - t_0), \quad (3.5)$$

with respect to $(\mathbf{a}, \mathbf{b})$, where $K_h(\cdot) = h^{-1}K(\cdot/h)$ with $K(\cdot)$ as the kernel function. Let $\mathbf{W} = (\mathbf{W}_1, \ldots, \mathbf{W}_n)^{\mathrm{T}}$ be the vector of continuous responses for all subjects with $\mathbf{W}_i = (W_i(t_{i1}), \ldots, W_i(t_{in_i}))^{\mathrm{T}}$ and $i = 1, \ldots, n$. The solution to the least squares algorithm is,

$$\hat{\mathbf{a}} = \hat{\boldsymbol{\beta}}(t_0) = (\mathbf{I}_p, \mathbf{0}_p)(\boldsymbol{\Gamma}_t^{\mathrm{T}} \boldsymbol{\kappa} \boldsymbol{\Gamma}_t)^{-1} \boldsymbol{\Gamma}_t^{\mathrm{T}} \boldsymbol{\kappa} \mathbf{W}, \quad (3.6)$$

where $\mathbf{I}_p$ is the identity matrix with size p, $\mathbf{0}_p$ is a size p matrix with each entry equal to zero, $\boldsymbol{\Gamma}_t = (\mathbf{X}, \mathbf{T}_t\mathbf{X})$ with $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^{\mathrm{T}}$, $\mathbf{X}_i = (\mathbf{X}_i(t_{i1}), \ldots, \mathbf{X}_i(t_{i\,n_i}))^{\mathrm{T}}$, and $\mathbf{T}_t$ an N-dimensional vector with each element being $(t_{ij} - t)$, and $\boldsymbol{\kappa}_t$ is an $N \times N$ diagonal matrix with each entry equal to $K_h(t_{ij} - t)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$.

In order to construct pointwise confidence intervals, we need an estimator for the asymptotic covariance matrix. Following the conventional techniques, we propose to estimate the asymptotic covariance matrix by using the following sandwich formula:

$$\widehat{\text{cov}}\{\hat{\boldsymbol{\beta}}(t_0)\} \approx (\mathbf{I}_p, \mathbf{0}_p)(\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}\boldsymbol{\Gamma}_t)^{-1}\left(\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}\mathcal{Q}\boldsymbol{\kappa}\boldsymbol{\Gamma}_t\right)(\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}\boldsymbol{\Gamma}_t)^{-1}(\mathbf{I}_p, \mathbf{0}_p)^{\mathrm{T}}, \tag{3.7}$$

where $\mathcal{Q} = \text{diag}(\mathcal{E}_1, \ldots, \mathcal{E}_n)$ with $\mathcal{E}_i = (e_i^2(t_{i1}), \ldots, e_i^2(t_{i\,n_i}))$ and $e_i(t_{ij}) = W_i(t_{ij}) - \mathbf{X}_i^{\mathrm{T}}(t_{ij})\hat{\boldsymbol{\beta}}(t_{ij})$, $i = 1, \ldots, n$, and $j = 1, \ldots, n_i$.

In the second stage we fit a generalized time-varying coefficient model to the conditional model (3.4). The generalized varying coefficient models for independent and identically distributed data were introduced by Cai et al. (2000), in this chapter we adapt these models to a longitudinal setting. We locally approximate the functions in a neighborhood of a fixed point t_0 via the Taylor expansion,

$$\alpha_r^*(t) \approx \alpha_r^*(t_0) + \alpha_r^{*\prime}(t_0)(t - t_0) \equiv a_r^* + b_r^*(t - t_0),$$

for $r = 1, \ldots, p+1$. Let $\mathbf{a}^* = (a_1^*, \ldots, a_p^*, a_{p+1}^*)^{\mathrm{T}}$, and $\mathbf{b}^* = (b_1^*, \ldots, b_p^*, b_{p+1}^*)^{\mathrm{T}}$. For the i^{th} subject, denote $\mathbf{X}_i^*(t_{ij}) = \left(\mathbf{X}_i^{\mathrm{T}}(t_{ij}), e_i(t_{ij})\right)^{\mathrm{T}}$ to be the design matrix with $e_i(t_{ij})$ as the residual from the marginal model. We maximize the local likelihood,

$$\ell_n(\mathbf{a}^*, \mathbf{b}^*) = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i} \ell\left(g^{-1}\left[\sum_{r=1}^{p+1}\{a_r^* + b_r^*(t_{ij} - t_0)\}\, X_{ir}^*(t_{ij})\right], Q_i(t_{ij})\right) K_h(t_{ij} - t_0), \tag{3.8}$$

where $g(\cdot)$ is the link function. For our model (3.4), the link function is probit. Hence, the local likelihood with probit link becomes

$$\begin{aligned} \ell_n(\mathbf{a}^*, \mathbf{b}^*) = {} & \frac{1}{N}\sum_{Q_i(t_{ij})=1} \log\left(\phi\left[\sum_{r=1}^{p+1}\{a_r^* + b_r^*(t_{ij} - t_0)\}\, X_{ir}^*(t_{ij})\right]\right) K_h(t_{ij} - t_0) \\ & + \frac{1}{N}\sum_{Q_i(t_{ij})=0} \log\left(1 - \phi\left[\sum_{r=1}^{p+1}\{a_r^* + b_r^*(t_{ij} - t_0)\}\, X_{ir}^*(t_{ij})\right]\right) K_h(t_{ij} - t_0), \end{aligned} \tag{3.9}$$

where $\phi(\cdot)$ is the probability density function for the standard normal distribution. We extend the iterative local maximum likelihood algorithm described in Cai et al. (2000) in order to find solutions to (3.9). Denote the value of a_r^* and b_r^* at the k^{th} iteration as $a_r^{*(k)}$ and $b_r^{*(k)}$. Let $\ell_n'(\mathbf{a}^*, \mathbf{b}^*)$ and $\ell_n''(\mathbf{a}^*, \mathbf{b}^*)$ be the gradient and Hessian matrix for the local likelihood (3.9),

$$\ell_n'(\mathbf{a}^*, \mathbf{b}^*) = \begin{pmatrix} v_{n,0} \\ v_{n,1} \end{pmatrix} \text{ and } \ell_n''(\mathbf{a}^*, \mathbf{b}^*) = \begin{pmatrix} H_{n,0} & H_{n,1} \\ H_{n,1} & H_{n,2} \end{pmatrix},$$

where

$$\begin{aligned} v_{n,s} = & \quad \frac{1}{N} \sum_{Q_i(t_{ij})=1} \phi(z_i)\Phi^{-1}(z_i)\mathbf{X}_i^*(t_{ij})K_h(t_{ij}-t_0)(t_{ij}-t_0)^s \\ & - \frac{1}{N} \sum_{Q_i(t_{ij})=0} \phi(z_i)\left\{1-\Phi(z_i)\right\}^{-1}\mathbf{X}_i^*(t_{ij})K_h(t_{ij}-t_0)(t_{ij}-t_0)^s, \\ H_{n,d} = & - \frac{1}{N}\Big(\sum_{Q_i(t_{ij})=1} \left\{z_i\phi(z_i)\Phi^{-1}(z_i) + \phi^2(z_i)\Phi^{-2}(z_i)\right\} \mathbf{X}_i^*(t_{ij})\mathbf{X}_i^{*\mathrm{T}}(t_{ij})K_h(t_{ij}-t_0)(t_{ij}-t_0)^d \\ & + \sum_{Q_i(t_{ij})=0} \left[\phi^2(z_i)\left\{1-\Phi(z_i)\right\}^{-2} - z_i\phi(z_i)\left\{1-\Phi(z_i)\right\}^{-1}\right] \mathbf{X}_i^*(t_{ij})\mathbf{X}_i^{*\mathrm{T}}(t_{ij})K_h(t_{ij}-t_0)(t_{ij}-t_0)^d\Big), \\ z_i = & \quad \sum_{r=1}^{p+1} \left\{a_r^{*(k)} + b_r^{*(k)}(t_{ij}-t_0)\right\} X_{ir}^*(t_{ij}), \; s=0,1, \; d=0,1,2, \end{aligned}$$

and $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution. Then, we update $(\mathbf{a}^*, \mathbf{b}^*)$ according to

$$\begin{pmatrix} \mathbf{a}^{*(k+1)} \\ \mathbf{b}^{*(k+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{a}^{*(k)} \\ \mathbf{b}^{*(k)} \end{pmatrix} - \left\{\ell_n''(\mathbf{a}^*, \mathbf{b}^*)\right\}^{-1}\ell_n'(\mathbf{a}^*, \mathbf{b}^*).$$

The solution of this iterative regression algorithm satisfies $\ell(\mathbf{a}^*, \mathbf{b}^*) = 0$ and the estimators are given by $\hat{\mathbf{a}}^* = \hat{\boldsymbol{\alpha}}^*(t_0) = \left(\hat{\alpha}_1^*(t_0), \ldots, \hat{\alpha}_p^*(t_0), \hat{\alpha}_{p+1}^*(t_0)\right)^{\mathrm{T}}$, and $\hat{\boldsymbol{\alpha}}^{*\prime}(t_0) = \hat{\mathbf{b}}^*$. The asymptotic covariance matrix of these estimators can be estimated as follows:

$$\widehat{\text{cov}}\{\hat{\boldsymbol{\alpha}}^*(t_0)\} = (\mathbf{I}_p, \mathbf{0}_p)\widehat{\boldsymbol{\Gamma}}(t_0)^{-1}\widehat{\boldsymbol{\Lambda}}(t_0)\widehat{\boldsymbol{\Gamma}}(t_0)^{-1}(\mathbf{I}_p, \mathbf{0}_p)^{\mathrm{T}}, \tag{3.10}$$

where

$$\widehat{\boldsymbol{\Gamma}}(t_0) = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\varpi_2\left[\sum_{r=1}^{p+1}\left\{\hat{a}_r^* + \hat{b}_r^*(t_{ij}-t_0)\right\}X_{ir}^*(t_{ij}), Q_i(t_{ij})\right]K_h(t_{ij}-t_0)\begin{pmatrix}\mathbf{X}_i^*(t_{ij})\\ \mathbf{X}_i^*(t_{ij})(t_{ij}-t_0)\end{pmatrix}^{\bigotimes 2},$$

$$\widehat{\boldsymbol{\Lambda}}(t_0) = \frac{h}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\varpi_1^2\left[\sum_{r=1}^{p+1}\left\{\hat{a}_r^* + \hat{b}_r^*(t_{ij}-t_0)\right\}X_{ir}^*(t_{ij}), Q_i(t_{ij})\right]K_h^2(t_{ij}-t_0)\begin{pmatrix}\mathbf{X}_i^*(t_{ij})\\ \mathbf{X}_i^*(t_{ij})(t_{ij}-t_0)\end{pmatrix}^{\bigotimes 2}$$

with $\varpi_d(\mathcal{Z}, q) = (\partial^d/\partial\mathcal{Z}^d)l\{g^{-1}(\mathcal{Z}), q\}$, and $A^{\bigotimes 2}$ denotes AA^T for a matrix or vector A.

The estimator of a_{p+1}^* gives us $\hat{\alpha}_{p+1}^*(t_0)$ and if the equality of $\sigma_1^2(t_0)$ and $\sigma_2^2(t_0)$ is assumed, then $\hat{\tau}(t_0) = \hat{\alpha}_{p+1}^*(t_0)$. If the equality of variances assumption is not employed, $\hat{\alpha}_{p+1}^*(t_0)$ gives us the estimator of the time-varying correlation coefficient, $\hat{\tau}(t_0)$, up to the factor $\{\sigma_2(t_0)/\sigma_1(t_0)\}$ along with its sign. Furthermore, the pointwise asymptotic confidence intervals of $\alpha_{p+1}^*(t_0)$ give us information on the significance of the correlation coefficient.

3.3 Asymptotic Results

In this section we study the asymptotic properties of the estimators in both stages of the estimation procedure. The proofs of the theorems presented in this section are provided in Section 3.6. Define $\mu_k = \int t^k K(t)dt$, $\nu_k = \int t^k K^2(t)dt$, and $\mathbf{H} = \text{diag}\{1, h\} \otimes \mathbf{I}_{(p\times J)}$ with $\otimes$ denoting the Kronecker product and $\mathbf{I}_{(p\times J)}$ the $(p\times J)\times(p\times J)$ identity matrix, and let $f(t)$ denote the marginal density of T, the time covariate.

The first theorem shows the asymptotic properties of the estimators in the first stage of the estimation procedure, which are obtained by minimizing the local likelihood (3.5). Define $\mathbf{s} = (t_1, \ldots, t_J)$, $J \geq 1$, $\boldsymbol{\theta}(\mathbf{s}) = (\boldsymbol{\theta}(t_1), \ldots, \boldsymbol{\theta}(t_J))^{\mathrm{T}}$ with $\boldsymbol{\theta}(t_0) = (a_1, \ldots, a_p, b_1, \ldots, b_p)^{\mathrm{T}}$, $\hat{\boldsymbol{\theta}}(\mathbf{s}) = \big(\hat{\boldsymbol{\theta}}(t_1), \ldots, \hat{\boldsymbol{\theta}}(t_J)\big)^{\mathrm{T}}$ with $\hat{\boldsymbol{\theta}}(t_0) = (\hat{\mathbf{a}}^{\mathrm{T}}, \hat{\mathbf{b}}^{\mathrm{T}})^{\mathrm{T}}$, and $t_0, t_{01}, t_{02} \in \mathbf{s}$. Let

$$\begin{aligned}
\Gamma_1(t_0) &= E\{\mathbf{X}_i^{\mathrm{T}}(t_{ij})\mathbf{X}_i(t_{ij}) \mid t_{ij} = t_0\}, \qquad (3.11)\\
\Upsilon_1(t_0) &= E\{\mathbf{X}_i^{\mathrm{T}}(t_{ij})W_i(t_{ij}) \mid t_{ij} = t_0\},\\
\eta_{lr}(t_0) &= E\{X_{il}(t_{ij})X_{ir}(t_{ij}) \mid t_{ij} = t_0\},\\
\eta_{lr}(t_{01}, t_{02}) &= E\{X_{il}(t_{ij_1})X_{ir}(t_{ij_2}) \mid t_{ij_1} = t_{01}, t_{ij_2} = t_{02}\},\\
\rho_1(t_{01}, t_{02}) &= \mathrm{cov}\{\varepsilon_{1i}(t_{01}), \varepsilon_{1i}(t_{02})\} = E\{\varepsilon_{1i}(t_{01})\varepsilon_{1i}(t_{02})\},
\end{aligned}$$

where $\varepsilon_{1i}(t_{ij})$ is the error term in (3.1), $l, r = 1, \ldots, p$, $i = 1, \ldots, n$, and $j = 1, \ldots, n_i$.

Theorem 3.1. *Under the regularity conditions (A)—(I) given in Section 3.6, we have the following result for* $\hat{\boldsymbol{\theta}}(\mathbf{s})$*:*

$$\sqrt{Nh}\left[H\{\hat{\boldsymbol{\theta}}(\mathbf{s}) - \boldsymbol{\theta}(\mathbf{s})\} - h^{1/2}\boldsymbol{\Delta}_\theta^{-1}\boldsymbol{\Lambda}\right] \xrightarrow{D} \mathcal{N}\left(0, \boldsymbol{\Delta}_\theta^{-1}\boldsymbol{\Sigma}_\theta(\mathbf{s})\boldsymbol{\Delta}_\theta^{-1}\right),$$

where

$$\boldsymbol{\Lambda} = \big(\boldsymbol{\Lambda}_1(t_1), \ldots, \boldsymbol{\Lambda}_p(t_1), \ldots, \boldsymbol{\Lambda}_1(t_J), \ldots, \boldsymbol{\Lambda}_p(t_J)\big)^T \text{ with}$$

$$\begin{aligned}
\boldsymbol{\Lambda}_l(t_0) = \sum_{r=1}^{p}\Bigg[& h^{3/2}\beta_r'(t_0)\eta_{lr}(t_0)f(t_0)\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}\\
&+ h^{5/2}\begin{pmatrix}\mu_2\\ \mu_3\end{pmatrix}\{\beta_r'(t_0)\eta_{lr}'(t_0)f(t_0) + \beta_r'(t_0)\eta_{lr}(t_0)f'(t_0) + \frac{1}{2}\beta_r''(t_0)\eta_{lr}(t_0)f(t_0)\}\Bigg] + o(h^2),
\end{aligned}$$

$$\boldsymbol{\Delta}_\theta = diag\big(\mathcal{D}(t_1), \ldots, \mathcal{D}(t_J)\big)^T \text{ with } \mathcal{D}(t_0) = f(t_0)\begin{pmatrix}1 & \mu_1\\ \mu_1 & \mu_2\end{pmatrix} \otimes \Gamma_1(t_0),$$

$$\mathbf{\Sigma}_\theta(\mathbf{s}) = h\mathbf{H}^{-1}\mathbf{\Sigma}\mathbf{H}^{-1} \ \ with\ \mathbf{\Sigma}(\mathbf{s}) = \begin{pmatrix} \Sigma_\theta(t_1,t_1) & \dots & \Sigma_\theta(t_1,t_J) \\ \vdots & \vdots & \vdots \\ \Sigma_\theta(t_J,t_1) & \dots & \Sigma_\theta(t_J,t_J) \end{pmatrix},$$

$$\Sigma_\theta(t_{01},t_{02}) = \begin{pmatrix} \mathfrak{d}_{11}(t_{01})\mathfrak{d}_{11}(t_{02}) & \dots & \mathfrak{d}_{11}(t_{01})\mathfrak{d}_{2p}(t_{02}) \\ \vdots & \vdots & \vdots \\ \mathfrak{d}_{1l}(t_{01})\mathfrak{d}_{11}(t_{02}) & \dots & \mathfrak{d}_{1l}(t_{01})\mathfrak{d}_{2p}(t_{02}) \\ \mathfrak{d}_{2l}(t_{01})\mathfrak{d}_{11}(t_{02}) & \dots & \mathfrak{d}_{2l}(t_{01})\mathfrak{d}_{2p}(t_{02}) \\ \vdots & \vdots & \vdots \\ \mathfrak{d}_{2p}(t_{01})\mathfrak{d}_{11}(t_{02}) & \dots & \mathfrak{d}_{2p}(t_{01})\mathfrak{d}_{2p}(t_{02}) \end{pmatrix},$$

$$\mathfrak{d}_{1l}(t_{01})\mathfrak{d}_{1r}(t_{02}) = \begin{cases} \{\sigma_1^2(t_{01})\eta_{lr}(t_{01})f(t_{01})\nu_0\} & \\ +\lambda h_0\left\{\rho_1(t_{01},t_{01})\eta_{lr}(t_{01},t_{01})f^2(t_{01})\right\}, & if\ t_{01}=t_{02} \\ \lambda h_0\left\{\rho_1(t_{01},t_{02})\eta_{lr}(t_{01},t_{02})f(t_{01})f(t_{02})\right\}, & if\ t_{01}\neq t_{02}, \end{cases}$$

$$\begin{aligned} \mathfrak{d}_{1l}(t_{01})\mathfrak{d}_{2r}(t_{02}) &= \mathfrak{d}_{2l}(t_{01})\mathfrak{d}_{1r}(t_{02}) \\ &= \begin{cases} \{\sigma_1^2(t_{01})\eta_{lr}(t_{01})f(t_{01})\nu_1\} & \\ +\lambda h_0\left\{\rho_1(t_{01},t_{01})\eta_{lr}(t_{01},t_{01})f^2(t_{01})\mu_1\right\}, & if\ t_{01}=t_{02} \\ \lambda h_0\left\{\rho_1(t_{01},t_{02})\eta_{lr}(t_{01},t_{02})f(t_{01})f(t_{02})\mu_1\right\}, & if\ t_{01}\neq t_{02}, \end{cases} \\ \mathfrak{d}_{2l}(t_{01})\mathfrak{d}_{2r}(t_{02}) &= \begin{cases} \{\sigma_1^2(t_{01})\eta_{lr}(t_{01})f(t_{01})\nu_2\} & \\ +\lambda h_0\left\{\rho_1(t_{01},t_{01})\eta_{lr}(t_{01},t_{01})f^2(t_{01})\mu_1^2\right\}, & if\ t_{01}=t_{02} \\ \lambda h_0\left\{\rho_1(t_{01},t_{02})\eta_{lr}(t_{01},t_{02})f(t_{01})f(t_{02})\mu_1^2\right\}, & if\ t_{01}\neq t_{02}, \end{cases} \end{aligned}$$

$t_0,\ t_{01}, t_{02} \in \mathbf{s}, \quad and \quad l = 1,\dots,p.$

The following lemma demonstrates how we link the first and the second stages of the estimation procedure. In theory we need to use the errors from the first stage to fit the model in the second stage. However, these errors are unobservable, and so we use the residuals in place of the errors. The following result holds for all $i = 1, \ldots, n$, so we drop the subscript i. Let $e(t)$ and $\varepsilon_1(t)$ denote the residual and error, respectively, at time point t.

Lemma 3.1. *Under regularity conditions (A)—(I) given in Section 3.6, and that $nh^8 \to 0$ and $nh^2/(logn)^2 \to \infty$,*

$$\sup_t |\varepsilon_1(t) - e(t)| \to 0 \quad as\ n \to \infty.$$

The third theorem shows the asymptotic properties of the estimators in the second stage of the estimation procedure, which are obtained by minimizing the local likelihood (3.8). Note that, although, in Section 3.2 we used $\boldsymbol{\alpha}^*(t_{ij})$ and $\alpha^*_{p+1}(t_{ij})$ to denote the regression coefficients in this stage, for the sake of simplicity, we demonstrate the asymptotic normality using a different notation. In order words, for this stage we define the regression coefficients at time point t_{ij} as $\boldsymbol{\alpha}(t_{ij}) = (\alpha_1(t_{ij}), \ldots, \alpha_p(t_{ij}))^{\mathrm{T}}$. Define $\boldsymbol{\vartheta}(\mathbf{s}) = (\boldsymbol{\vartheta}(t_1), \ldots, \boldsymbol{\vartheta}(t_J))^{\mathrm{T}}$ with $\boldsymbol{\vartheta}(t_0) = (a_1, \ldots, a_p, b_1, \ldots, b_p)^{\mathrm{T}}$, and $\hat{\boldsymbol{\vartheta}}(\mathbf{s}) = \left(\hat{\boldsymbol{\vartheta}}(t_1), \ldots, \hat{\boldsymbol{\vartheta}}(t_J)\right)^{\mathrm{T}}$ with $\hat{\boldsymbol{\vartheta}}(t_0) = (\hat{\mathbf{a}}^{\mathrm{T}}, \hat{\mathbf{b}}^{\mathrm{T}})^{\mathrm{T}}$. Let

$$\begin{aligned}
&\Gamma_2(t_0) = E\left[\rho\{t_0, \mathbf{X}_i(t_{ij})\}\mathbf{X}_i(t_{ij})\mathbf{X}_i^{\mathrm{T}}(t_{ij}) \mid t_{ij} = t_0\right], \\
&\Gamma_3(t_{01}, t_{02}) = E\left[\rho\{t_{01}, t_{02}, \mathbf{X}_i(t_{ij_1}), \mathbf{X}_i(t_{ij_2})\}\mathbf{X}_i^{\mathrm{T}}(t_{ij_1})\mathbf{X}_i(t_{ij_2}) \mid t_{ij_1} = t_{01}, t_{ij_2} = t_{02}\right], \\
&\rho\{t_0, \mathbf{x}_i(t_{ij})\} = (g\left[m\{t_0, \mathbf{x}_i(t_{ij})\}\right])^2 \operatorname{cov}\{Q_i(t_{ij}) \mid t_{ij} = t_0, \mathbf{X}_i(t_{ij}) = \mathbf{x}_i(t_{ij})\}, \\
&\rho\{t_{01}, t_{02}, \mathbf{x}_i(t_{ij_1}), \mathbf{x}_i(t_{ij_2})\} = \left\{g\left(m\left[(t_{01}, t_{02})^{\mathrm{T}}, \{\mathbf{x}_i(t_{ij_1}), \mathbf{x}_i(t_{ij_2})\}^{\mathrm{T}}\right]\right)\right\}^2 \\
&\qquad\qquad \times \operatorname{cov}\{Q_i(t_{ij_1}), Q_i(t_{ij_2}) \mid t_{ij_1} = t_{01}, t_{ij_2} = t_{02}, \mathbf{x}_i(t_{ij_1}), \mathbf{x}_i(t_{ij_2})\}, \\
&m\{t_0, \mathbf{x}_i(t_{ij})\} = E\{Q_i(t_{ij}) \mid t_{ij} = t_0, \mathbf{X}_i(t_{ij}) = \mathbf{x}_i(t_{ij})\},
\end{aligned}$$

with $t_0, t_{01}, t_{02} \in \mathbf{s}$, $i = 1, \ldots, n$, and $j = 1, \ldots, n_i$.

Theorem 3.2. *Under regularity conditions (C)—(F) and (J)—(N) given in Section 3.6, and that* $h_0 \to 0$ *and* $nh \to \infty$ *as* $n \to \infty$ *with* $h_0 = N^{-1/5}h$ *and* $N = \sum_{i=1}^{n} n_i$*, we have the following result:*

$$\sqrt{Nh}\left[\mathbf{H}\{\hat{\boldsymbol{\vartheta}}(\mathbf{s}) - \boldsymbol{\vartheta}(\mathbf{s})\} - \boldsymbol{\Xi}\,\boldsymbol{\Lambda}_{\vartheta}\right] \xrightarrow{D} \mathcal{N}\left(0, \boldsymbol{\Xi}\,\boldsymbol{\Sigma}_{\vartheta}(\mathbf{s})\,\boldsymbol{\Xi}\right),$$

where

$$\boldsymbol{\Lambda}_{\vartheta} = \big(\boldsymbol{\Lambda}_{\vartheta}(t_1), \ldots, \boldsymbol{\Lambda}_{\vartheta}(t_J)\big)^T \text{ with } \boldsymbol{\Lambda}_{\vartheta}(t_0) = \frac{h^2 f(t_0)}{2}\begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} \otimes \Gamma_2(t_0)\boldsymbol{\alpha}''(t_0) + o(1),$$

$$\boldsymbol{\Xi} = diag\big(\Delta_{\vartheta}^{-1}(t_1), \ldots, \Delta_{\vartheta}^{-1}(t_J)\big)^T \text{ with } \Delta_{\vartheta}(t_0) = f(t_0)\begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \Gamma_2(t_0),$$

$$\boldsymbol{\Sigma}_{\vartheta}(\mathbf{s}) = \begin{pmatrix} \boldsymbol{\Sigma}_{\vartheta}(t_1, t_1) & \ldots & \boldsymbol{\Sigma}_{\vartheta}(t_1, t_J) \\ \vdots & \vdots & \vdots \\ \boldsymbol{\Sigma}_{\vartheta}(t_J, t_1) & \ldots & \boldsymbol{\Sigma}_{\vartheta}(t_J, t_J) \end{pmatrix} \text{ with}$$

$$\boldsymbol{\Sigma}_{\vartheta}(t_{01}, t_{02}) = \begin{cases} f(t_{01})\begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix} \otimes \Gamma_2(t_{01}) \\ +\lambda h_0 f^2(t_{01})\begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_1^2 \end{pmatrix} \otimes \Gamma_3(t_{01}, t_{01}), & \text{if } t_{01} = t_{02} \\ \lambda h_0 f(t_{01}) f(t_{02})\begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_1^2 \end{pmatrix} \otimes \Gamma_3(t_{01}, t_{02}), & \text{if } t_{01} \neq t_{02}, \end{cases}$$

and t_0, t_{01}, $t_{02} \in \mathbf{s}$.

3.4 Simulation Studies

In this section we study the performance of the proposed methodology via a Monte Carlo simulation study. The Epanechnikov kernel, $K(t) = 0.75(1-t^2)_+$, on an equidistant set of grid points $\{t_k, k = 1, \ldots, n_{grid}\}$ between 0 and 1 with $n_{grid} = 200$ is used in our simulation. In this study we generate 500 intensive longitudinal data sets, in which for each unit the number of measurements is randomly selected using a discrete uniform distribution on $[20, 40]$ and the measurement times $T_i = (t_{i1}, \ldots, t_{in_i})$ are generated from a uniform distribution on $[0, 1]$ with $n_i \in [20, 40]$. We use sample size $n = 75$. The continuous and latent variables are generated from the following models:

$$\begin{aligned} W_i(t_{ij}) &= \beta_1(t_{ij}) + \beta_2(t_{ij})X_i(t_{ij}) + \varepsilon_{1i}(t_{ij}), \\ Y_i(t_{ij}) &= \alpha_1(t_{ij}) + \alpha_2(t_{ij})X_i(t_{ij}) + \varepsilon_{2i}(t_{ij}), \end{aligned} \tag{3.12}$$

where $\beta_1(t_{ij}) = \sin(0.5\pi t_{ij})$, $\beta_2(t_{ij}) = \cos(\pi t_{ij} - 1/8)$, $\alpha_1(t_{ij}) = \sin(\pi t_{ij}) - 0.5$, $\alpha_2(t_{ij}) = 0.5\cos(2\pi t_{ij})$, $i = 1, \ldots, 75$ and $j \in [20, 40]$. In Section 3.1 we defined the binary variable as $Q_i(t_{ij}) = 1$ if $Y_i(t_{ij}) > 0$, and $Q_i(t_{ij}) = 0$ if $Y_i(t_{ij}) \leq 0$. However, we would like to demonstrate that decreasing the percentage of successes in the binary response does not decrease the efficacy of our method. Hence, we define the relation between the latent variable and the binary variable as $Q_i(t_{ij}) = 1$ if $Y_i(t_{ij}) > 0.3$, and $Q_i(t_{ij}) = 0$ if $Y_i(t_{ij}) \leq 0.3$. So, each of our 500 simulated data sets has approximately 45% failure. The predictor variable $X_i(t_{ij})$ is generated from the standard Gaussian distribution. The error variable for the continuous response $\varepsilon_{1i}(t_{ij})$ follows a Gaussian distribution with mean zero, variance $0.5 + 0.5\sin^2(2\pi t_{ij})$, and $\text{corr}\{\varepsilon_{1i}(t_{ij}), \varepsilon_{1i}(t_{ij'})\} = \rho_1(t_{ij}, t_{ij'}) = 0.3^{|t_{ij} - t_{ij'}|}$ for $j \neq j'$. In addition, $\varepsilon_{2i}(t_{ij})$ follows a Gaussian distribution with mean zero, variance $0.5 + 0.5\sin^2(2\pi t_{ij})$, and $\text{corr}\{\varepsilon_{2i}(t_{ij}), \varepsilon_{2i}(t_{ij'})\} = \rho_2(t_{ij}, t_{ij'}) = 0.4^{|t_{ij} - t_{ij'}|}$ for $j \neq j'$. Hence, $\boldsymbol{\varepsilon}_i(t_{ij}) = (\varepsilon_{1i}(t_{ij}), \varepsilon_{2i}(t_{ij}))^T$ follows a bivariate Gaussian distribution with $\text{corr}\{\varepsilon_{1i}(t_{ij}), \varepsilon_{2i}(t_{ij})\} = \tau(t_{ij}) = 0.2\sin(\pi t_{ij})$ and $\text{corr}\{\varepsilon_{1i}(t_{ij}), \varepsilon_{2i}(t_{ij'})\} = \rho_{12}(t_{ij}, t_{ij'}) = 0.2\sqrt{\sin(\pi t_{ij})\sin(\pi t_{ij'})}$ for $j \neq j'$.

In the first stage we fit the time-varying coefficient model to the marginal model of the continuous response (3.12). To evaluate the performance of the estimators in this stage, we use root average squared error (RASE),

$$\text{RASE} = \left[\frac{1}{200} \sum_{r=1}^{2} \sum_{k=1}^{200} \{\beta_r(t_k) - \hat{\beta}_r(t_k)\}^2 \right]^{1/2}.$$

In our simulation, we generate several pilot simulation data sets, and use a cross-validation bandwidth selector to get an overall picture about the optimal bandwidth. To save computing time, we fix the bandwidth to be close to the optimal ones from the pilot simulation data sets. Specifically, we set the bandwidth to be h = 0.20. It is of interest to examine the performance of the proposed procedure with a wide range of the bandwidth. Thus, we also set the bandwidth to be h/2=0.10 and 2h=0.40, corresponding to undersmoothing and oversmoothing simulations. Table 3.1 shows the sample means and the sample standard deviations of the RASE values, based on 500 replications, computed at three bandwidths, 0.10, 0.20 and 0.40. According to the mean RASE values in Table 3.1, the bandwidth $h = 0.20$ gives the minimum value.

Table 3.1: Summary of simulation results for the first stage

h	RASE	t	$\hat{\beta}_1(t)$		$\hat{\beta}_2(t)$	
	Mean(SD)		SD	SE (SD_{se})	SD	SE (SD_{se})
0.10	0.078 (0.033)	0.30	0.055	0.050 (0.005)	0.058	0.050 (0.005)
		0.50	0.044	0.038 (0.003)	0.038	0.037 (0.004)
		0.70	0.055	0.050 (0.004)	0.052	0.050 (0.005)
0.20	0.062 (0.016)	0.30	0.037	0.034 (0.003)	0.034	0.034 (0.003)
		0.50	0.032	0.028 (0.002)	0.029	0.028 (0.003)
		0.70	0.038	0.034 (0.003)	0.035	0.034 (0.003)
0.40	0.112 (0.042)	0.30	0.027	0.024 (0.002)	0.024	0.023 (0.002)
		0.50	0.023	0.022 (0.002)	0.022	0.021 (0.002)
		0.70	0.027	0.024 (0.002)	0.024	0.023 (0.002)

Figures 3.1(a) and 3.1(b) depict the typical estimates of the parameter functions along with the 2.5 and 97.5 percentiles based on 500 Monte Carlo runs at $h = 0.2$. We see that the typical estimated coefficient functions are close to the true functions.

We also test the accuracy of the proposed standard error formula (3.7). The standard deviation of 500 $\hat{\beta}_r(t)$, based on 500 simulations, denoted by SD in Table 3.1, can be viewed as the true standard error. The sample average and the sample standard deviation of the 500 estimated standard errors of $\hat{\beta}_r(t)$ are denoted by SE and SD_{se} in Table 3.1, respectively. They summarize the overall performance of the standard error formula (3.7). Table 3.1 presents the results at the points $t = 0.30, 0.50$ and 0.70. From Table 3.1 we can see that our standard error formula slightly underestimates the true standard error, but

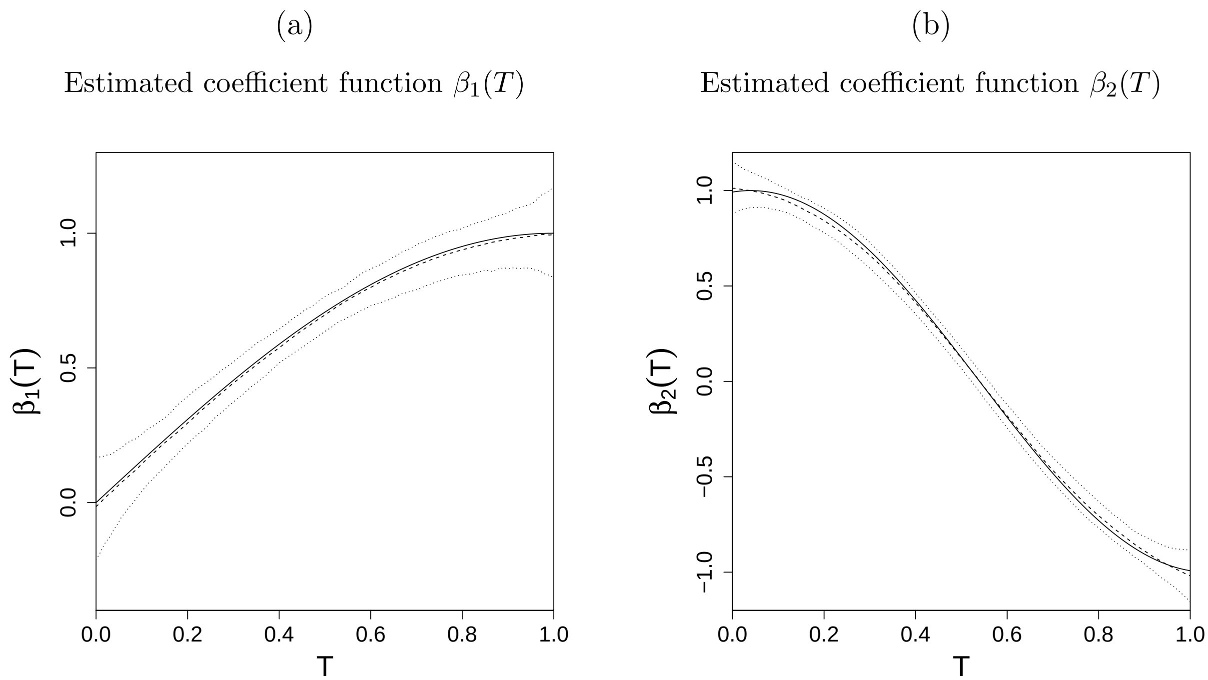

Figure 3.1: Estimated varying coefficient functions (dashed) of the time-varying coefficient model fit to the continuous response overlaying the true coefficient functions (solid) along with the 2.5 and 97.5 percentiles based on 500 Monte Carlo runs (dotted).

their difference is less than two times the SD_{se}.

Next, in the second stage we fit a generalized time-varying coefficient model to the conditional model of the binary response given the continuous response.

$$P\left\{Q_i(t_{ij}) = 1 \mid W_i(t_{ij})\right\} = \Phi\left\{\alpha_1^*(t_{ij}) + \alpha_2^*(t_{ij})X_i(t_{ij}) + \alpha_3^*(t_{ij})e_i(t_{ij})\right\},$$

where $e_i(t_{ij}) = W_i(t_{ij}) - \{\hat{\beta}_1(t_{ij}) + X_i(t_{ij})\hat{\beta}_2(t_{ij})\}$ is the residual from the first stage, $\alpha_3^*(t_{ij}) = \{\sigma_2(t_{ij})/\sigma_1(t_{ij})\}\,\tau(t_{ij})$. Since we assume the equality of $\sigma_1^2(t_{ij})$ and $\sigma_2^2(t_{ij})$, the estimate of $\alpha_3^*(t_{ij})$ is equal to the estimate of the correlation coefficient, $\hat{\tau}(t_{ij})$. Thus, our main focus in this stage is the performance of $\hat{\alpha}_3^*(t_{ij})$, which demonstrates how the association between the continuous response and the binary response changes over time. Similar to the first stage, we evaluate the performance of the estimator in the second stage using RASE,

$$\text{RASE} = \left[\frac{1}{200} \sum_{k=1}^{200} \{\alpha_3^*(t_k) - \hat{\alpha}_3^*(t_k)\}^2 \right]^{1/2}.$$

Table 3.2 gives the sample means and the sample standard deviations

Table 3.2: Summary of simulation results for the second stage

h	RASE	t	$\hat{\alpha}_3^*(t)$	
	Mean(SD)		SD	SE (SD_{se})
		0.30	0.046	0.041 (0.003)
0.40	0.073 (0.035)	0.50	0.041	0.038 (0.003)
		0.70	0.045	0.041 (0.003)
		0.30	0.045	0.040 (0.003)
0.50	0.069 (0.036)	0.50	0.034	0.034 (0.002)
		0.70	0.043	0.039 (0.003)
		0.30	0.043	0.039 (0.003)
0.60	0.071 (0.032)	0.50	0.033	0.032 (0.002)
		0.70	0.042	0.039 (0.003)

of the RASE values based on 500 replications, computed at three bandwidths, $0.40, 0.50$ and 0.60. According to the mean RASE values in Table 3.2, we set the bandwidth to be $h = 0.50$.

Estimated correlation coefficient function $\tau(T)$

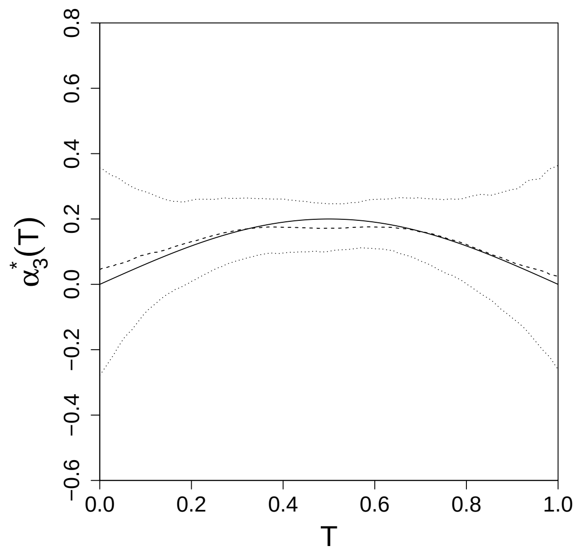

Figure 3.2: Estimated correlation coefficient function (dashed) overlaying the true correlation function (solid) along with the 2.5 and 97.5 percentiles based on 500 Monte Carlo runs (dotted).

Figure 3.2 depicts the typical estimate of the correlation coefficient function along with the 2.5 and 97.5 percentiles based on 500 Monte Carlo runs at $h = 0.50$. Figure 3.2 indicates that the typical estimated correlation coefficient function is close to the underlying true correlation coefficient function.

Table 3.2 presents the SD, the SE and SD_{se} at the points $t = 0.30, 0.50$ and 0.70 for $\hat{\alpha}_3^*(t)$. In Table 3.2 SE and SD_{se} summarize the overall performance of the standard error formula (3.10). The SE is slightly less than the true standard error, but their difference is less than two times the SD_{se}.

3.5 Application to the Smoking Cessation Study

In this section we apply our proposed joint modeling methodology to the Ecological Momentary Assessment (EMA) data described in the introduction. Shiffman et al. (1996) collected data on 304 smokers using palm-top computers that beeped at random times. At each random assessment prompt, participants recorded their answers to a series of questions about their current activities and setting, such as their alcohol use and the presence of other smokers. Current mood and urge to smoke were also recorded. The data collection process is described below.

First, the participants were monitored for a two-week interval during which they were engaged in their normal activities. During this period they were asked to record all their smoking occasions and to respond to the random assessment prompts. Subjects were then instructed to stop smoking on day 17, called the wait day. When the electronic diary records showed that the participant had abstained for 24 hours, that day was recorded as the subject's quit day. Once the participants quit, they were asked to keep responding to the random assessment prompts and to record any episodes of smoking (lapses) or strong temptations. During this observation period, 149 subjects lapsed. Our goal is to analyze the data for the lapsed participants.

In our analysis we are mainly interested in the randomly scheduled assessment data recorded two weeks before and after each subject's quit day. Subjects with missing values on wait day or quit day were excluded from the analysis. Data alignment was necessary because different subjects may have different quit days.

Previous research regarding smoking cessation suggests that the mood variables—affect, arousal, and attention—are important factors on smoking (Shiffman et al., 2002). It has been shown that both positive and negative affect are associated with smoking through urge to smoke. One question of interest is how these predictors (the mood variables) affect urge to smoke, and how this impact changes over time. However, as mentioned in the introduction, our main interest is the association between drinking alcohol and smoking. Alcohol and tobacco researchers are interested in explaining this association in order to improve the treatments and prevention techniques for both smokers and drinkers. Previous studies have shown that the relationship between smoking and drinking alcohol is positive by fitting a regression model of urge to smoke on alcohol use. In our method we are estimating the partial correlation between urge to smoke and alcohol use, that is, we are estimating the correlation between drinking alcohol and urge to smoke given the mood variables—affect, arousal, and attention. Moreover, we are allowing this partial correlation to be time-varying. We investigate how the association between urge to smoke and alcohol use changes from two weeks before to two weeks after the quit day in order to advance our knowledge about the relationship between drinking and smoking. Note that urge to smoke was recorded on a scale ranging from 0 to 11.

In our analysis we use the Epanechnikov kernel function and bandwidth $h = 5.0$, which is obtained using a leave-one-out cross validation procedure. In the first stage of our estimation procedure, we answer the question of how the relationship between urge to smoke and mood variables changes over time. We fit the following time-varying coefficient model to urge to smoke:

$$W_i(t_{ij}) = \beta_0(t_{ij}) + \beta_1(t_{ij})X_{i1}(t_{ij}) + \beta_2(t_{ij})X_{i2}(t_{ij}) + \beta_3(t_{ij})X_{i3}(t_{ij}) + \varepsilon_{1i}(t_{ij}), \tag{3.13}$$

where

$W_i(t_{ij})$: the score of urge to smoke of the i^{th} subject at time t_{ij},

$X_{i1}(t_{ij})$: the centered score of negative affect of the i^{th} subject at time t_{ij},

$X_{i2}(t_{ij})$: the centered score of arousal of the i^{th} subject at time t_{ij},

$X_{i3}(t_{ij})$: the centered score of attention of the i^{th} subject at time t_{ij}.

The estimated time-varying regression coefficients are depicted in Figure 3.3. From Figures 3.3 (a) and (b), we can see that before the quit day, the coefficient functions for the intercept and negative affect are close to being time-invariant. According to Figure 3.3 (a), the intercept function starts to decrease at the quit day. From Figure 3.3 (b), we see that the coefficient for negative affect is always positive, i.e., as negative affect increases, urge to smoke also increases. At the quit day, we see a sudden increase in this coefficient, i.e., the effect of negative affect on urge to smoke increases. Figure 3.3 (c) shows that the coefficient for arousal is positive before the quit day and becomes negative after the quit day. Figure 3.3 (d) shows that the coefficient for attention is time-varying and is positive until approximately day 13 after the quit day, and so we conclude that as the difficulty in concentrating increases, urge to smoke also increases. At day 13 after the quit day, the coefficient starts to decrease and becomes negative on day 15 after the quit day, indicating that the effect of attention on urge to smoke is decreasing.

Next, we use the second stage of our estimation procedure to determine how the association between alcohol use and urge to smoke changes over time. By obtaining the residuals from the marginal model (3.13), $e(t_{ij})$, we fit the

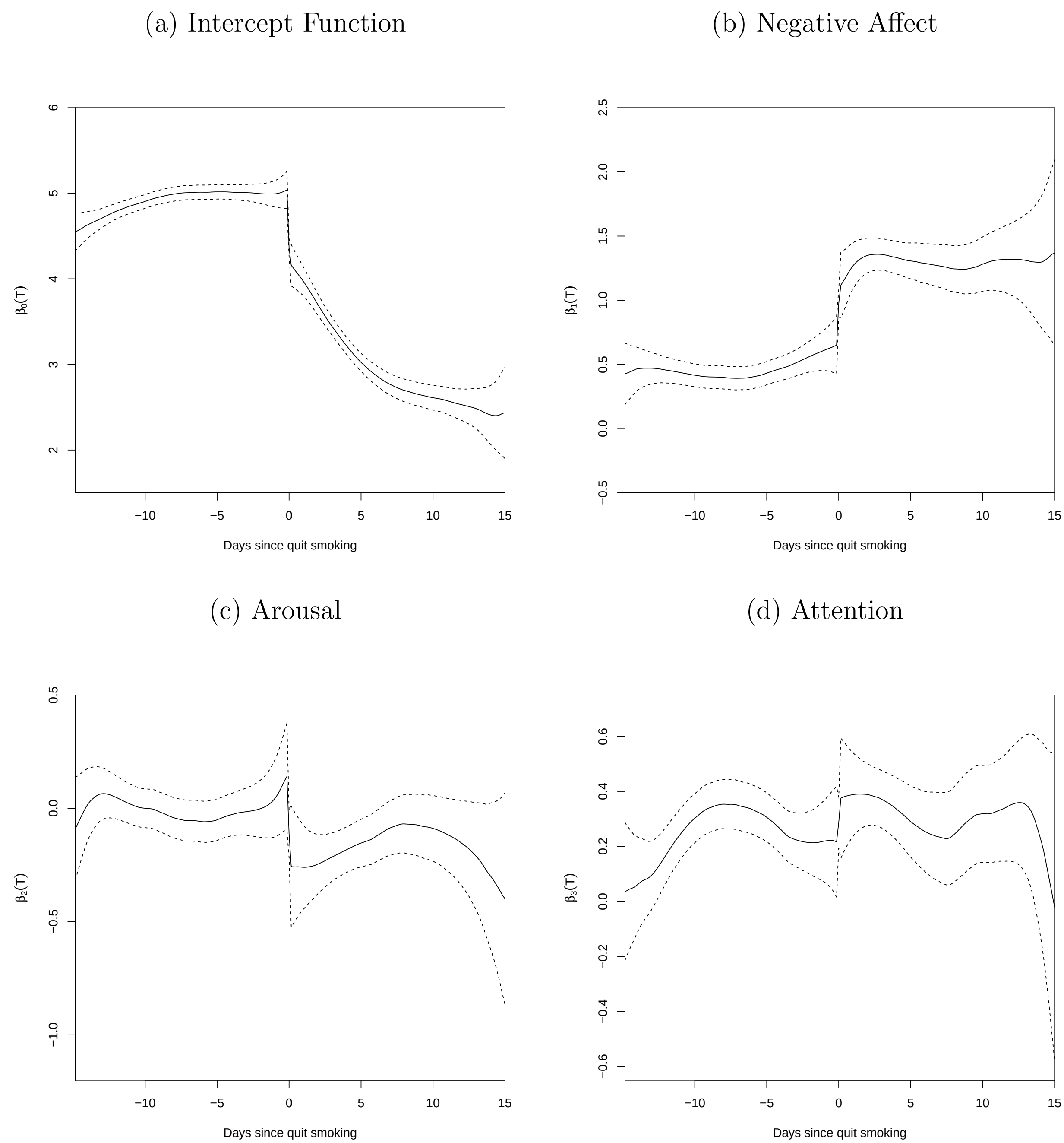

Figure 3.3: Plots of estimated coefficient functions (solid) of the time-varying coefficient model fit to urge to smoke response along with the 95% pointwise asymptotic confidence intervals before and after quitting smoking (dashed). We aligned the data so that all subjects have quit day at day zero. (a) Intercept function, (b) negative affect, (c) arousal, and (d) attention.

following generalized time-varying coefficient model:

$$\begin{aligned} P\{Q_{1i}(t_{ij}) = 1 \mid W_i(t_{ij})\} &= \Phi\big\{\alpha_0^*(t_{ij}) + \alpha_1^*(t_{ij})X_{i1}(t_{ij}) + \alpha_2^*(t_{ij})X_{i2}(t_{ij}) + \alpha_3^*(t_{ij})X_{i3}(t_{ij}) \\ &+ \alpha_4^*(t_{ij})e_i(t_{ij})\big\}, \end{aligned}$$

where $Q_{1i}(t_{ij})$ is the alcohol use of the i^{th} subject at time t_{ij}, and $X_{i1}(t_{ij}), X_{i2}(t_{ij})$ and $X_{i3}(t_{ij})$ are defined in (3.13). As mentioned in Section 3.1, $\alpha_4^*(t_{ij}) = \{\sigma_2(t_{ij})/\sigma_1(t_{ij})\}\,\tau_1(t_{ij})$ with $\sigma_1^2(t_{ij})$ and $\sigma_2^2(t_{ij})$ as the variances of urge to smoke and the latent variable underlying the alcohol use at time t_{ij}, respectively, and $\tau_1(t_{ij})$ as the correlation between urge to smoke and alcohol use at time t_{ij}. Figure 3.4 (a) depicts the estimated correlation coefficient, $\hat{\alpha}_4^*(t_{ij})$. Figure 3.4 (a) demonstrates that the correlation between alcohol use and urge to smoke is time-varying. We observe that before the quit day urge to smoke and alcohol use have a positive association, but after the quit day the association becomes negative. In other words, before the quit day increased urge to smoke is associated with alcohol usage, whereas after the quit day reduced urge to smoke is associated with alcohol usage. We can also test the significance of the association by using the confidence intervals. The association is significant before the quit day but becomes insignificant after the quit day, which may be due to lack of enough data.

Also of interest is association between urge to smoke and the presence of other smokers. Shiffman and Balabanis (1995), McDermut and Haaga (1998) and Warren and McDonough (1999) observed that the sight of other smokers tends to provoke a craving to smoke. By employing our joint modeling technique, we can study how this association changes over time. First we fit the time-varying coefficient model to urge to smoke. This model is the same as the one in the alcohol usage analysis (3.13). We observe the same trends in the regression

(a) Estimated coefficient function $\alpha_4^*(T)$

(b) Estimated coefficient function $\gamma_4^*(T)$

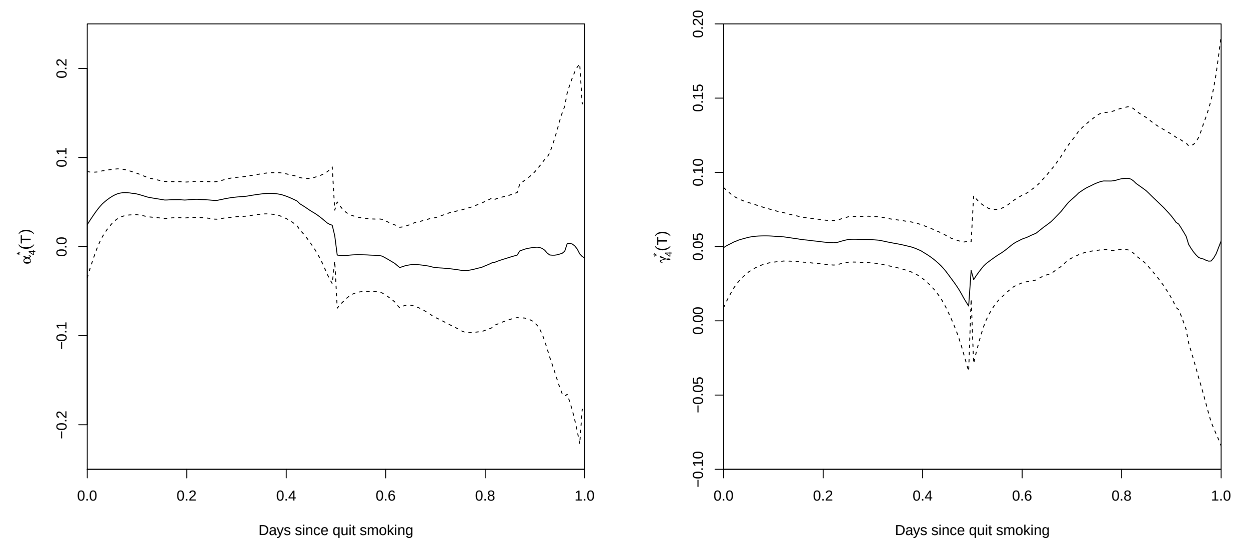

Figure 3.4: Estimated coefficient functions (solid) of the generalized time-varying coefficient model fit along with the 95% pointwise asymptotic confidence intervals (dashed) (a) alcohol versus urge to smoke and (b) presence of other smokers versus urge to smoke.

coefficients as in Figure 3.3, since the only difference between these two analyses is that we removed subjects with missing values for this binary response. The plots are omitted.

In the second stage of our estimation procedure, we use the residuals from the marginal model (3.13) and fit the following generalized time-varying coefficient model:

$$\begin{aligned} P\{Q_{2i}(t_{ij}) = 1 \mid W_i(t_{ij})\} &= \Phi\big\{\gamma_0^*(t_{ij}) + \gamma_1^*(t_{ij})X_{i1}(t_{ij}) + \gamma_2^*(t_{ij})X_{i2}(t_{ij}) + \gamma_3^*(t_{ij})X_{i3}(t_{ij}) \\ &+ \gamma_4^*(t_{ij})e_i(t_{ij})\big\}, \end{aligned}$$

where $Q_{2i}(t_{ij})$ is the response of the i^{th} subject at time t_{ij} on presence of other smokers, and $X_{i1}(t_{ij}), X_{i2}(t_{ij})$ and $X_{i3}(t_{ij})$ are defined in (3.13). In this model, $\gamma_4^*(t_{ij}) = \{\sigma_3(t_{ij})/\sigma_1(t_{ij})\}\,\tau_2(t_{ij})$ with $\sigma_3^2(t_{ij})$ as the variance of the latent variable underlying the binary variable, presence of other smokers, and $\tau_2(t_{ij})$ as the correlation between urge to smoke and presence of other smokers at time t_{ij}. Figure 3.4 (b) demonstrates the estimated correlation coefficient, $\hat{\gamma}_4^*(t_{ij})$. Figure 3.4 (b) indicates that the correlation between the presence of other smokers and urge to smoke is time-varying and is always positive. In other words, the presence of other smokers is always associated with increased urge to smoke. The correlation starts to decrease approximately 3 days before the quit day and starts to increase right after the quit day. Around day 13 after the quit day, the correlation appears to become insignificant; however, this might be due to not having sufficient number of observations.

In summary, our procedure will allow researchers to investigate the time-varying association between urge to smoke and a binary variable of interest, which may lead to more efficacious smoking interventions. Moreover, by allowing the regression coefficients to vary over time, our method shows how the relationship between urge to smoke and the mood variables changes over time. Our methods can also be applied in other prevention/intervention settings.

3.6 Technical Conditions and Proofs

The following regularity conditions are needed to facilitate proofs of the theorems presented in this chapter.

Regularity Conditions:

A. The observed sample $\{t_{ij}, X_i(t_{ij}), W_i(t_{ij})\}$, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$ is an independent and identically distributed realization of (T, X, W). The $\varepsilon_{1i}(t_{ij})$ are iid from a distribution with mean zero and finite variance $\sigma_1^2(t_{ij})$. The covariate T has finite support $\mathcal{T} = [\mathcal{L},\mathcal{U}]$. The support for X is a closed and bounded interval in $\mathbb{R}^p$, denoted by $\mathcal{X}$.

B. The varying coefficient functions $\beta_r(t_{ij})$, $r = 1, \ldots, p$, have continuous second order derivatives over $\mathcal{T}$.

C. Without loss of generality the kernel density function $K(\cdot)$ has bounded support and satisfies

$$\int K(t)dt = 1, \quad \int |t|^3 K(t)dt < \infty, \quad \int t^2 K^2(t)dt < \infty.$$

Note that for theorem 1, the condition $\int |t|^2 K(t)dt < \infty$ is sufficient.

D. The marginal density function $f(t_{ij})$ of T is continuous and positive for all $t_{ij} \in \mathcal{T}$.

E. The bandwidth h satisfies $h = N^{-1/5}h_0$ for some constant $h_0 > 0$, and $N = \sum_{i=1}^n n_i$.

F. $\lim_{n\to\infty} N^{-6/5} \sum_{i=1}^n n_i^2 = \lambda$ for some $0 \le \lambda < \infty$.

G. There exists a constant $\delta > 0$ such that $E\{|\varepsilon_{1i}(t_{ij})\ |^{2+\delta}\} < \infty$, and $E\{|X_{il}(t_{ij})\ |^{4+\delta}\} < \infty$ for all $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, $l = 1, \ldots, p$, and $t_{ij} \in \mathcal{T}$.

H. For all $l, r = 1, \ldots, p$, $\beta_r(t_{ij}), \eta_{lr}(t_{ij})$ and $f(t_{ij})$ have continuous second derivatives at $t_{ij} = t_0$.

I. The variance and covariance functions, $\sigma_1^2(t_{ij})$ and $\rho_1(t_{ij})$ are continuous at $t_{ij} = t_0$.

J. The function $\varpi_2(\mathcal{Z}, q) < 0$ for $\mathcal{Z} \in \mathcal{R}$, and q in the range of the binary response.

K. The varying coefficient functions $\alpha_r(t_{ij})$, $r = 1, \ldots, p$ has continuous second order derivatives in the neighborhood of the point $t_{ij} = t_0$.

L. The functions $\Gamma_2(t_{ij}), \Gamma_3(t_{ij}), V\left[m\{t_{ij}, \mathbf{x}(t_{ij})\}\right], V'\left[m\{t_{ij}, \mathbf{x}(t_{ij})\}\right]$, and $g'''\left[m\{t_{ij}, \mathbf{x}(t_{ij})\}\right]$ are continuous at the point $t_{ij} = t_0$. Moreover, assume that $f(t_0) > 0$, $\Gamma_2(t_0) > 0$, and $\Gamma_3(t_0) > 0$.

M. $E\{|\mathbf{X}_i(t_{ij})|^3 \mid t_{ij} = t_0\}$ is continuous.

N. $E\{Q_i^4(t_{ij}) \mid t_{ij} = t_0, \mathbf{X}_i(t_{ij}) = \mathbf{x}_i(t_{ij})\}$ is bounded.

By Condition (B), we assume that the parameter space for $\boldsymbol{\theta}(\mathbf{s})$, namely, Θ_1, is a closed and bounded subset of $\mathbb{R}^{2d}$. Conditions (D) and (L) implies that $\varpi_1(\cdot,\cdot)$, $\varpi_2(\cdot,\cdot)$, $\varpi_3(\cdot,\cdot)$, $\rho'(\cdot,\cdot)$, and $m'(\cdot,\cdot)$ are continuous. By condition (F), we only consider cases in which n_i, do not converge to ∞ faster than $N = \sum_{i=1}^{n} n_i$. Condition (J) guarantees that the local likelihood function (3.8) is concave.

Proof of Theorem 3.1

Proof. Let $\mathbf{s} = (t_1, \ldots, t_J)$, $J \geq 1$, be a set of distinct interior points in $\mathcal{T}$. By Taylor expanding the first derivative of the least squares function $\ell_\theta\{\hat{\boldsymbol{\theta}}(\mathbf{s})\}$ around $\boldsymbol{\theta}(\mathbf{s})$, we obtain

$$\mathbf{H}\left\{\hat{\boldsymbol{\theta}}(\mathbf{s}) - \boldsymbol{\theta}(\mathbf{s})\right\} = -N\mathbf{H}\left[\ell_{\theta\theta^T}\left\{\boldsymbol{\theta}(\mathbf{s})\right\}\right]^{-1}\mathbf{H}N^{-1}\mathbf{H}^{-1}\ell_\theta\left\{\boldsymbol{\theta}(\mathbf{s})\right\},$$

where $\ell_{\theta\theta^T}\left\{\boldsymbol{\theta}(\mathbf{s})\right\}$ denotes the second derivative of the likelihood $\ell\left\{\boldsymbol{\theta}(\mathbf{s})\right\}$ with respect to θ.

We will show that $N^{-1}\mathbf{H}^{-1}\left[\ell_{\theta\theta^T}\left\{\boldsymbol{\theta}(\mathbf{s})\right\}\right]^{-1}\mathbf{H}^{-1}$ converges in probability and that $N^{-1}\mathbf{H}^{-1}l_\theta\left\{\boldsymbol{\theta}(\mathbf{s})\right\}$ converges in distribution. Thus, Theorem 3.1 follows by using the Slutsky's Theorem. Let

$$\begin{aligned} \ell_\theta\left\{\boldsymbol{\theta}(\mathbf{s})\right\} &= \left[\ell_\theta\left\{\boldsymbol{\theta}(t_1)\right\}, \ldots, \ell_\theta\left\{\boldsymbol{\theta}(t_J)\right\}\right]^{\mathrm{T}} \quad \text{with} \qquad (3.14)\\ \ell_\theta\left\{\boldsymbol{\theta}(t_0)\right\} &= \left[\ell_{1\,\theta}\left\{\boldsymbol{\theta}(t_0)\right\}, \ldots, \ell_{l\,\theta}\left\{\boldsymbol{\theta}(t_0)\right\}, \ldots, \ell_{p\,\theta}\left\{\boldsymbol{\theta}(t_0)\right\}\right]^{\mathrm{T}}, \end{aligned}$$

where $t_0 \in \mathbf{s}$, and $\ell_{l\,\theta}\left\{\boldsymbol{\theta}(t_0)\right\} = (\partial\ell/\partial a, \partial\ell/\partial b)^{\mathrm{T}} \equiv \left\{\ell_{1l}(t_0), \ell_{2l}(t_0)\right\}^{\mathrm{T}}$, with
$\ell_{1l}(t_0) = -\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{W_i(t_{ij}) - \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\beta}(t_0)\right\}\mathbf{X}_i^{\mathrm{T}}(t_{ij})K_h(t_{ij} - t_0)$,
$\ell_{2l}(t_0) = -\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{W_i(t_{ij}) - \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\beta}(t_0)\right\}\mathbf{X}_i^{\mathrm{T}}(t_{ij})(t_{ij} - t_0)K_h(t_{ij} - t_0)$,
$\boldsymbol{\beta}(t_0) = \mathbf{a} + \mathbf{b}(t_{ij} - t_0)$ with $\mathbf{a} = (a_1, \ldots, a_p)$ and $\mathbf{b} = (b_1, \ldots, b_p)$, and $l = 1, \ldots, p$. To simplify notation, let's rewrite $\ell_{1l}(t_0)$, and $\ell_{2l}(t_0)$ as follows:

$$\begin{aligned} \ell_{1l}(t_0) &= -\sum_{i=1}^{n}\psi_{il}(t_0; h),\\ \ell_{2l}(t_0) &= -\sum_{i=1}^{n}(t_{ij} - t_0)\psi_{il}(t_0; h), \end{aligned}$$

where $\psi_{il}(t_0; h) = \sum_{j=1}^{n_i}\left\{\xi_{il}(t_0, t_{ij})K_h(t_{ij} - t_0)\right\}$, $\xi_{il}(t_0, t_{ij}) = \sum_{r=1}^{p}\big[X_{il}(t_{ij})X_{ir}(t_{ij})\{\beta_r(t_{ij}) - \beta_r(t_0)\}\big] + X_{il}(t_{ij})\varepsilon_{1i}(t_{ij})$. We start with finding the expected values of $\ell_{1l}(t_0)$

and $\ell_{2l}(t_0)$.

$$
\begin{aligned}
E\{\ell_{1l}(t_0)\} &= -\sum_{i=1}^{n} E\left\{\psi_{il}(t_0;h)\right\} \\
E\left\{\psi_{il}(t_0;h)\right\} &= E\Big\{\sum_{j=1}^{n_i} K_h(t_{ij}-t_0)\xi_{il}(t_0,t_{ij})\Big\} \\
&= n_i \int E\left\{K_h(t_{ij}-t_0)\xi_{il}(t_0,t_{ij}) \mid t_{ij}=\varsigma\right\} f(\varsigma)d\varsigma \\
&= n_i \int hK_h(u)\sum_{r=1}^{p}\left\{\beta_r(t_0+hu)-\beta_r(t_0)\right\}\eta_{lr}(t_0+hu)f(t_0+hu)du \\
&\quad \text{with } \eta_{lr}(t_0+hu) = E\big\{X_{il}(t_{ij})X_{ir}(t_{ij}) \mid t_{ij}=t_0+hu\big\}.
\end{aligned}
$$

By Taylor expanding $\beta_r(t_0+hu)$, $\eta_{lr}(t_0+hu)$, and $f(t_0+hu)$, we get

$$
\begin{aligned}
E\left\{\psi_{il}(t_0;h)\right\} &= n_i\sum_{r=1}^{p}\big[h\beta_r'(t_0)\eta_{lr}(t_0)f(t_0)\mu_1 \\
&+ h^2\mu_2\{\beta_r'(t_0)\eta_{lr}'(t_0)f(t_0)+\beta_r'(t_0)\eta_{lr}(t_0)f'(t_0)+\frac{1}{2}\beta_r''(t_0)\eta_{lr}(t_0)f(t_0)\}\big]+o(n_ih^2).
\end{aligned}
$$

Then,

$$
\begin{aligned}
E\{\ell_{1l}(t_0)\} &= -N\sum_{r=1}^{p}\big[h\beta_r'(t_0)\eta_{lr}(t_0)f(t_0)\mu_1 \\
&+ h^2\mu_2\{\beta_r'(t_0)\eta_{lr}'(t_0)f(t_0)+\beta_r'(t_0)\eta_{lr}(t_0)f'(t_0)+\frac{1}{2}\beta_r''(t_0)\eta_{lr}(t_0)f(t_0)\}\big]+o(Nh^2), \\
E\{\ell_{2l}(t_0)\} &= -N\sum_{r=1}^{p}\big[h^2\beta_r'(t_0)\eta_{lr}(t_0)f(t_0)\mu_2 \\
&+ h^3\mu_3\{\beta_r'(t_0)\eta_{lr}'(t_0)f(t_0)+\beta_r'(t_0)\eta_{lr}(t_0)f'(t_0)+\frac{1}{2}\beta_r''(t_0)\eta_{lr}(t_0)f(t_0)\}\big]+o(Nh^3).
\end{aligned}
$$

Let

$$
\begin{aligned}
\mathbf{\Lambda} &= (N\mathbf{H})^{-1}E[\ell_\theta\{\boldsymbol{\theta}(\mathbf{s})\}] = (N\mathbf{H})^{-1}\left(E\left[\ell_\theta\left\{\boldsymbol{\theta}(t_1)\right\}\right], \ldots, E\left[\ell_\theta\left\{\boldsymbol{\theta}(t_J)\right\}\right]\right)^{\mathrm{T}}, \\
E\left[\ell_\theta\left\{\boldsymbol{\theta}(t_0)\right\}\right] &= \big(e_{1\theta}(t_0), \ldots, e_{p\theta}(t_0)\big)^{\mathrm{T}}, \text{and} \\
e_{l\theta}(t_0) &= -N\,\mathrm{diag}\,(1,h)\begin{pmatrix} E[\ell_{1l}\left\{\boldsymbol{\theta}(t_0)\right\}] \\ E[\ell_{2l}\left\{\boldsymbol{\theta}(t_0)\right\}] \end{pmatrix} \\
&= -N\,\mathrm{diag}\,(1,h)\,\left(\sum_{r=1}^{p}\left[h\beta_r'(t_0)\eta_{lr}(t_0)f(t_0)\begin{pmatrix}\mu_1 \\ \mu_2\end{pmatrix}\right.\right. \\
&+ \left.\left. h^2\begin{pmatrix}\mu_2 \\ \mu_3\end{pmatrix}\{\beta_r'(t_0)\eta_{lr}'(t_0)f(t_0) + \beta_r'(t_0)\eta_{lr}(t_0)f'(t_0) + \frac{1}{2}\beta_r''(t_0)\eta_{lr}(t_0)f(t_0)\}\right]\right) \\
&+ o(Nh^2),
\end{aligned}
$$

where $l = 1, \ldots, p$. As a result, by the Law of Large Numbers,

$$(N\mathbf{H})^{-1}\ell_\theta\big\{\boldsymbol{\theta}(\mathbf{s})\big\} \xrightarrow{p} \mathbf{\Lambda}.$$

Next, we compute the covariance of $\ell_\theta\big\{\boldsymbol{\theta}(\mathbf{s})\big\}$ as follows:

$$\mathbf{\Sigma}(\mathbf{s}) = \mathrm{cov}\big[\ell_\theta\big\{\boldsymbol{\theta}(\mathbf{s})\big\}, \ell_\theta\big\{\boldsymbol{\theta}(\mathbf{s})\big\}\big] = E\big[\ell_\theta\big\{\boldsymbol{\theta}(\mathbf{s})\big\}\ell_\theta\big\{\boldsymbol{\theta}(\mathbf{s})\big\}^{\mathrm{T}}\big] - E\big[\ell_\theta\big\{\boldsymbol{\theta}(\mathbf{s})\big\}\big]E\big[\ell_\theta\big\{\boldsymbol{\theta}(\mathbf{s})\big\}\big]^{\mathrm{T}},$$

where $\ell_\theta\big\{\boldsymbol{\theta}(\mathbf{s})\big\}$ is defined in (3.14). Thus, we need to calculate the following expectations for both $t_{01} = t_{02}$, and $t_{01} \neq t_{02}$ with $l, r = 1, \ldots, p$, and t_{01}, $t_{02} \in \mathbf{s}$:

$$
\begin{aligned}
E\left\{\ell_{1l}(t_{01})\ell_{1r}(t_{02})\right\} - E\left\{\ell_{1l}(t_{01})\right\}E\left\{\ell_{1r}(t_{02})\right\} &= E\left[\left\{\sum_{i=1}^{n}\psi_{il}(t_{01};h)\right\}\left\{\sum_{i=1}^{n}\psi_{ir}(t_{02};h)\right\}\right] \\
&- E\left\{\ell_{1l}(t_{01})\right\}E\left\{\ell_{1r}(t_{02})\right\}, \\
E\left\{\ell_{1l}(t_{01})\ell_{2r}(t_{02})\right\} - E\left\{\ell_{1l}(t_{01})\right\}E\left\{\ell_{2r}(t_{02})\right\} &= E\left[\left\{\sum_{i=1}^{n}\psi_{il}(t_{01};h)\right\}\left\{\sum_{i=1}^{n}\psi_{ir}(t_{02};h)(t_{ij}-t_{02})\right\}\right] \\
&- E\left\{\ell_{1l}(t_{01})\right\}E\left\{\ell_{2r}(t_{02})\right\}, \\
E\left\{\ell_{2l}(t_{01})\ell_{1r}(t_{02})\right\} - E\left\{\ell_{2l}(t_{01})\right\}E\left\{\ell_{1r}(t_{02})\right\} &= E\left[\left\{\sum_{i=1}^{n}\psi_{il}(t_{01};h)(t_{ij}-t_{01})\right\}\left\{\sum_{i=1}^{n}\psi_{ir}(t_{02};h)\right\}\right] \\
&- E\left\{\ell_{2l}(t_{01})\right\}E\left\{\ell_{1r}(t_{02})\right\}, \\
E\left\{\ell_{2r}(t_{01})\ell_{2r}(t_{02})\right\} - E\left\{\ell_{2l}(t_{01})\right\}E\left\{\ell_{2r}(t_{02})\right\} &= E\left[\left\{\sum_{i=1}^{n}\psi_{il}(t_{01};h)(t_{ij}-t_{01})\right\}\left\{\sum_{i=1}^{n}\psi_{ir}(t_{02};h)(t_{ij}-t_{02})\right\}\right] \\
&- E\left\{\ell_{2l}(t_{01})\right\}E\left\{\ell_{2r}(t_{02})\right\}.
\end{aligned}
\tag{3.15}
$$

Let's start with the first expectation in (3.15).

$$
\begin{aligned}
E &\left\{\ell_{1l}(t_{01})\ell_{1r}(t_{02})\right\} - E\left\{\ell_{1l}(t_{01})\right\}E\left\{\ell_{1r}(t_{02})\right\} \\
= & E\left[\left\{\sum_{i=1}^{n}\psi_{il}(t_{01};h)\right\}\left\{\sum_{i=1}^{n}\psi_{ir}(t_{02};h)\right\}\right] \\
= & \sum_{i=1}^{n}E\left\{\psi_{il}(t_{01};h)\psi_{ir}(t_{02};h)\right\} + \sum_{i_1\neq i_2}E\left\{\psi_{i_1l}(t_{01};h)\psi_{i_2r}(t_{02};h)\right\} \\
- & E\left\{\ell_{1l}(t_{01})\right\}E\left\{\ell_{1r}(t_{02})\right\}.
\end{aligned}
\tag{3.16}
$$

We start with the first term on the right-side (3.16).

$$
\begin{aligned}
E &\left\{\psi_{il}(t_{01};h)\psi_{ir}(t_{02};h)\right\} \\
= & E\left[\left\{\sum_{j=1}^{n_i}K_h(t_{ij}-t_{01})\xi_{il}(t_{01},t_{ij})\right\}\left\{\sum_{j=1}^{n_i}K_h(t_{ij}-t_{02})\xi_{ir}(t_{02},t_{ij})\right\}\right] \\
= & E\Big\{\sum_{j=1}^{n_i}\xi_{il}(t_{01},t_{ij})\xi_{ir}(t_{02},t_{ij})K_h(t_{ij}-t_{01})K_h(t_{ij}-t_{02}) \\
+ & \sum_{j_1\neq j_2}\xi_{il}(t_{01},t_{ij_1})\xi_{ir}(t_{02},t_{ij_2})K_h(t_{ij_1}-t_{01})K_h(t_{ij_2}-t_{02})\Big\}.
\end{aligned}
\tag{3.17}
$$

First, we deal with the first term in (3.17).

$$\begin{aligned}\sum_{j=1}^{n_i} & E \quad \Big\{\xi_{il}(t_{01},t_{ij})\xi_{ir}(t_{02},t_{ij})K_h(t_{ij}-t_{01})K_h(t_{ij}-t_{02})\Big\} \\ & = \quad n_i \int K_h(\varsigma - t_{01})K_h(\varsigma - t_{02})E\left\{\xi_{il}(t_{01},t_{ij})\xi_{ir}(t_{02},t_{ij}) \mid t_{ij}=\varsigma\right\} f(\varsigma)d\varsigma,\end{aligned}$$

where

$$\begin{aligned} E & \quad \{\xi_{il}(t_{01},t_{ij})\xi_{ir}(t_{02},t_{ij}) \mid t_{ij}=\varsigma\} \\ = & \quad \sum_{c=1}^{p}\{\beta_c(\varsigma)-\beta_c(t_{01})\}\{\beta_c(\varsigma)-\beta_c(t_{02})\} E\big\{X_{il}(t_{ij})X_{ic}^2(t_{ij})X_{ir}(t_{ij}) \mid t_{ij}=\varsigma\big\} \\ + & \quad E\big\{X_{il}(t_{ij})X_{ir}(t_{ij})\varepsilon_{1i}^2(t_{ij}) \mid t_{ij}=\varsigma\big\} \\ + & \quad \sum_{c_1\neq c_2}\{\beta_{c_1}(\varsigma)-\beta_{c_1}(t_{01})\}\{\beta_{c_2}(\varsigma)-\beta_{c_2}(t_{02})\} E\big\{X_{il}(t_{ij})X_{ic_1}(t_{ij})X_{ic_2}(t_{ij})X_{ir}(t_{ij}) \mid t_{ij}=\varsigma\big\} \\ \to & \quad \sigma_1^2(t_{0c})\eta_{lr}(t_{0c}) \text{ as } n\to\infty \text{ if } \varsigma\to t_{0c},\ c=1,2.\end{aligned}$$

Then,

$$\begin{aligned}\sum_{j=1}^{n_i} & E \quad \Big\{\xi_{il}(t_{01},t_{ij})\xi_{ir}(t_{02},t_{ij})K_h(t_{ij}-t_{01})K_h(t_{ij}-t_{02})\Big\} \\ & = \begin{cases} \frac{n_i}{h}\left\{\sigma_1^2(t_{01})\eta_{lr}(t_{01})f(t_{01})\nu_0\right\} + o(n_i/h), & \text{if } t_{01}=t_{02} \\ \\ o(n_i/h), & \text{if } t_{01}\neq t_{02}.\end{cases}\end{aligned}$$

Next, we deal with the second term in (3.17).

$$\begin{aligned} E & \quad \Big\{\sum_{j_1\neq j_2}\xi_{il}(t_{01},t_{ij_1})\xi_{ir}(t_{02},t_{ij_2})K_h(t_{ij_1}-t_{01})K_h(t_{ij_2}-t_{02})\Big\} \\ = & \quad n_i(n_i-1)\int\int K_h(t_{ij}-t_{01})K_h(t_{ij}-t_{02})E\left\{\xi_{il}(t_{01},t_{ij_1})\xi_{ir}(t_{02},t_{ij_2}) \mid t_{ij_1}=\varsigma_1, t_{ij_2}=\varsigma_2\right\} f(\varsigma_1)f(\varsigma_2)d\varsigma_1 d\varsigma_2,\end{aligned}$$

where

$$
\begin{aligned}
E \quad & \{\xi_{il}(t_{01}, t_{ij_1})\xi_{ir}(t_{02}, t_{ij_2}) \mid t_{ij_1} = \varsigma_1, t_{ij_2} = \varsigma_2\} \\
= \quad & \sum_{c=1}^{p} \{\beta_c(\varsigma_1) - \beta_c(t_{01})\} \{\beta_c(\varsigma_2) - \beta_c(t_{02})\} E\big\{X_{il}(t_{ij_1})X_{ic}(t_{ij_1})X_{ic}(t_{ij_2})X_{ir}(t_{ij_2}) \mid t_{ij_1} = \varsigma_1, t_{ij_2} = \varsigma_2\big\} \\
+ \quad & \rho_1(\varsigma_1, \varsigma_2)\eta_{lr}(\varsigma_1, \varsigma_2) + \sum_{c_1 \neq c_2} \{\beta_{c_1}(\varsigma_1) - \beta_{c_1}(t_{01})\} \{\beta_{c_2}(\varsigma_2) - \beta_{c_2}(t_{02})\} \\
\times \quad & E\big\{X_{il}(t_{ij_1})X_{ic_1}(t_{ij_1})X_{ic_2}(t_{ij_2})X_{ir}(t_{ij_2}) \mid t_{ij_1} = \varsigma_1, t_{ij_2} = \varsigma_2\big\} \rightarrow \\
& \begin{cases} \rho_1(t_{01}, t_{01})\eta_{lr}(t_{01}, t_{01}), \text{ if } t_{01} = t_{02} \\ \\ \rho_1(t_{01}, t_{02})\eta_{lr}(t_{01}, t_{02}), \text{ if } t_{01} \neq t_{02}. \end{cases}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
E \quad & \Big\{\sum_{j_1 \neq j_2} \xi_{il}(t_{01}, t_{ij_1})\xi_{ir}(t_{02}, t_{ij_2}) K_h(t_{ij_1} - t_{01}) K_h(t_{ij_2} - t_{02})\Big\} \\
= \quad & \begin{cases} n_i(n_i - 1)\left\{\rho_1(t_{01}, t_{01})\eta_{lr}(t_{01}, t_{01}) f^2(t_{01})\right\} + o\{n_i(n_i - 1)\}, & \text{if } t_{01} = t_{02} \\ \\ n_i(n_i - 1)\left\{\rho_1(t_{01}, t_{02})\eta_{lr}(t_{01}, t_{02}) f(t_{01}) f(t_{02})\right\} + o\{n_i(n_i - 1)\}, & \text{if } t_{01} \neq t_{02}. \end{cases}
\end{aligned}
$$

Thus, (3.17) is equal to

$$
\begin{aligned}
E \quad & \{\psi_{il}(t_{01}; h)\psi_{ir}(t_{02}; h)\} \\
= \quad & \begin{cases} \frac{n_i}{h}\left\{\sigma_1^2(t_0)\eta_{lr}(t_0) f(t_0)\nu_0\right\} + n_i(n_i - 1)\left\{\rho_1(t_{01}, t_{01})\eta_{lr}(t_{01}, t_{01}) f^2(t_{01})\right\} & \\ \\ +o\{n_i(n_i - 1)\}, & \text{if } t_{01} = t_{02} \\ \\ n_i(n_i - 1)\left\{\rho_1(t_{01}, t_{02})\eta_{lr}(t_{01}, t_{02}) f(t_{01}) f(t_{02})\right\} + o\{n_i(n_i - 1)\}, & \text{if } t_{01} \neq t_{02}. \end{cases}
\end{aligned}
$$

Now, what is left to show is the second part in (3.16).

$$
N^{-1} h \left[\sum_{i_1 \neq i_2} E\left\{\psi_{i_1 l}(t_{01}; h)\psi_{i_2 r}(t_{02}; h)\right\} - E\left\{\ell_{1l}(t_{01})\right\} E\left\{\ell_{1r}(t_{02})\right\}\right].
$$

There exists a positive constant $M < \infty$ such that

$$\begin{aligned} N^{-1}h \quad & \Big|\sum_{i_1 \neq i_2} E\left\{\psi_{i_1 l}(t_{01};h)\psi_{i_2 r}(t_{02};h)\right\} - E\left\{\ell_{1l}(t_{01})\right\} E\left\{\ell_{1r}(t_{02})\right\}\Big| \quad (3.18) \\ \leq \quad & Nh\left\{1 - N^{-2}\sum_{i=1}^{n}\left(n_i \sum_{i' \neq i} n_{i'}\right)\right\} M \to 0, \text{ as } n \to \infty. \end{aligned}$$

The limit of (3.18) holds because, by regularity condition (F),

$$1 - N^{-2}\sum_{i=1}^{n}\left(n_i \sum_{i' \neq i} n_{i'}\right) = N^{-2}\sum_{i=1} n_i^2 = o\big\{(Nh)^{-1}\big\}.$$

Hence,

$$\begin{aligned} E\left\{\ell_{1l}(t_{01})\ell_{1r}(t_{02})\right\} &= \sum_{i=1}^{n} E\left\{\psi_{il}(t_{01};h)\psi_{ir}(t_{02};h)\right\} \\ &= \begin{cases} \frac{N}{h}\left\{\sigma_1^2(t_0)\eta_{lr}(t_0)f(t_0)\nu_0\right\} \\ +\big(\sum_{i=1}^{n} n_i^2 - N\big)\big\{\rho_1(t_{01},t_{01})\eta_{lr}(t_{01},t_{01})f^2(t_{01})\big\} \\ +o\big(\sum_{i=1}^{n} n_i^2 - N\big), \text{ if } t_{01} = t_{02} \\ \big(\sum_{i=1}^{n} n_i^2 - N\big)\{\rho_1(t_{01},t_{02})\eta_{lr}(t_{01},t_{02})f(t_{01})f(t_{02})\} \\ +o\big(\sum_{i=1}^{n} n_i^2 - N\big), \text{ if } t_{01} \neq t_{02}. \end{cases} \end{aligned}$$

Similarly,

$$E\left\{\ell_{1l}(t_{01})\ell_{2r}(t_{02})\right\} = \sum_{i=1}^{n} E\left\{\psi_{il}(t_{01};h)\psi_{ir}(t_{02};h)(t_{ij}-t_{02})\right\}$$

$$= \begin{cases} N\left\{\sigma_1^2(t_0)\eta_{lr}(t_0)f(t_0)\nu_1\right\} \\ +h\left(\sum_{i=1}^n n_i^2 - N\right)\left\{\rho_1(t_{01},t_{01})\eta_{lr}(t_{01},t_{01})f^2(t_{01})\mu_1\right\} \\ +o\left\{h\left(\sum_{i=1}^n n_i^2 - N\right)\right\}, \text{ if } t_{01}=t_{02} \\ h\left(\sum_{i=1}^n n_i^2 - N\right)\{\rho_1(t_{01},t_{02})\eta_{lr}(t_{01},t_{02})f(t_{01})f(t_{02})\mu_1\} \\ +o\left\{h\left(\sum_{i=1}^n n_i^2 - N\right)\right\}, \text{ if } t_{01}\neq t_{02}, \end{cases}$$

$$E\left\{\ell_{2l}(t_{01})\ell_{1r}(t_{02})\right\} = \sum_{i=1}^{n} E\left\{\psi_{il}(t_{01};h)\psi_{ir}(t_{02};h)(t_{ij}-t_{01})\right\}$$

$$= \begin{cases} N\left\{\sigma_1^2(t_0)\eta_{lr}(t_0)f(t_0)\nu_1\right\} \\ +h\left(\sum_{i=1}^n n_i^2 - N\right)\left\{\rho_1(t_{01},t_{01})\eta_{lr}(t_{01},t_{01})f^2(t_{01})\mu_1\right\} \\ +o\left\{h\left(\sum_{i=1}^n n_i^2 - N\right)\right\}, \text{ if } t_{01}=t_{02} \\ h\left(\sum_{i=1}^n n_i^2 - N\right)\{\rho_1(t_{01},t_{02})\eta_{lr}(t_{01},t_{02})f(t_{01})f(t_{02})\mu_1\} \\ +o\left\{h\left(\sum_{i=1}^n n_i^2 - N\right)\right\}, \text{ if } t_{01}\neq t_{02}, \end{cases}$$

and

$$E\left\{\ell_{2l}(t_{01})\ell_{2r}(t_{02})\right\} = \sum_{i=1}^{n} E\left\{\psi_{il}(t_{01};h)\psi_{ir}(t_{02};h)(t_{ij}-t_{01})(t_{ij}-t_{02})\right\}$$

$$= \begin{cases} Nh\left\{\sigma_1^2(t_0)\eta_{lr}(t_0)f(t_0)\nu_2\right\} \\ +h^2\left(\sum_{i=1}^n n_i^2 - N\right)\left\{\rho_1(t_{01},t_{01})\eta_{lr}(t_{01},t_{01})f^2(t_{01})\mu_1^2\right\} \\ +o\left\{h^2\left(\sum_{i=1}^n n_i^2 - N\right)\right\}, \text{ if } t_{01}=t_{02} \\ h^2\left(\sum_{i=1}^n n_i^2 - N\right)\left\{\rho_1(t_{01},t_{02})\eta_{lr}(t_{01},t_{02})f(t_{01})f(t_{02})\mu_1^2\right\} \\ +o\left\{h^2\left(\sum_{i=1}^n n_i^2 - N\right)\right\}, \text{ if } t_{01}\neq t_{02}. \end{cases}$$

Now, we can write the following:

$$N^{-1}h\,\mathrm{diag}\,(1,1/h)\ \mathrm{cov}\{\ell_{l\,\theta}(t_{01})\ell_{r\theta}(t_{01})\}\,\mathrm{diag}\,(1,1/h)$$

$$= \begin{cases} \sigma_1^2(t_{01})\eta_{lr}(t_{01})f(t_{01})\begin{pmatrix}\nu_0 & \nu_1\\ \nu_1 & \nu_2\end{pmatrix} \\ +h\,N^{-1}\left(\sum_{i=1}^n n_i^2 - N\right)\rho_1(t_{01},t_{01})\eta_{lr}(t_{01},t_{01})f^2(t_{01})\begin{pmatrix}1 & \mu_1\\ \mu_1 & \mu_1^2\end{pmatrix}, & \text{if } t_{01}=t_{02} \\ h\,N^{-1}\left(\sum_{i=1}^n n_i^2 - N\right)\rho_1(t_{01},t_{02})\eta_{lr}(t_{01},t_{02})f(t_{01})f(t_{02})\begin{pmatrix}1 & \mu_1\\ \mu_1 & \mu_1^2\end{pmatrix}, & \text{if } t_{01}\neq t_{02}. \end{cases}$$

By regularity conditions (E) and (F), it is easy to see that as $n\to\infty$,

$$hN^{-1}\left(\sum_{i=1}^n n_i^2 - N\right) = N^{-6/5}\left(\sum_{i=1}^n n_i^2 - N\right)h_0 \to \lambda h_0.$$

Let $\boldsymbol{\Sigma}_\theta(\mathbf{s}) = h\mathbf{H}^{-1}\boldsymbol{\Sigma}(\mathbf{s})\mathbf{H}^{-1}$ and by using the Central Limit Theorem,

$$h^{1/2}(N\mathbf{H})^{-1}\ell_\theta\{\boldsymbol{\theta}(\mathbf{s})\} \xrightarrow{p} \mathcal{N}\left(h^{1/2}\boldsymbol{\Lambda}, \boldsymbol{\Sigma}_\theta(\mathbf{s})\right)$$

as $n \to \infty$.

Next consider

$$\ell_{\theta\theta^{\mathrm{T}}}\{\boldsymbol{\theta}(\mathbf{s})\} = \left[\ell_{\theta\theta^{\mathrm{T}}}\left\{\boldsymbol{\theta}(t_1)\right\}, \ldots, \ell_{\theta\theta^{\mathrm{T}}}\left\{\boldsymbol{\theta}(t_J)\right\}\right]^{\mathrm{T}} \text{ with } \ell_{\theta\theta^{\mathrm{T}}}\left\{\boldsymbol{\theta}(t_0)\right\} = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix},$$

where $t_0 \in \mathbf{s}$, and

$$\begin{aligned} L_{11} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} \mathbf{X}_i^{\mathrm{T}}(t_{ij})\mathbf{X}_i(t_{ij})K_h(t_{ij}-t_0), \\ L_{12} &= L_{21} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} \mathbf{X}_i^{\mathrm{T}}(t_{ij})\mathbf{X}_i(t_{ij})(t_{ij}-t_0)K_h(t_{ij}-t_0), \\ L_{22} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} \mathbf{X}_i^{\mathrm{T}}(t_{ij})\mathbf{X}_i(t_{ij})(t_{ij}-t_0)^2K_h(t_{ij}-t_0). \end{aligned}$$

The expectation of $\ell_{\theta\theta^{\mathrm{T}}}\{\boldsymbol{\theta}(t_0)\}$ can be given as follows:

$$\begin{aligned} E(L_{11}) &= E\Bigg\{\sum_{i=1}^{n}\sum_{j=1}^{n_i} \mathbf{X}_i^{\mathrm{T}}(t_{ij})\mathbf{X}_i(t_{ij})K_h(t_{ij}-t_0)\Bigg\} \\ &= N\int E\Big\{\mathbf{X}_i^{\mathrm{T}}(t_{ij})\mathbf{X}_i(t_{ij})K_h(t_{ij}-t_0) \mid t_{ij}=\varsigma\Big\}f(\varsigma)d\varsigma \\ &= Nf(t_0)\otimes\Gamma_1(t_0). \end{aligned}$$

Similarly, $E(L_{12}) = E(L_{21}) = Nhf(t_0)\otimes\Gamma_1(t_0)$, and $E(L_{22}) = Nh^2f(t_0)\otimes\Gamma_1(t_0)$. Then,

$$E\left[\ell_{\theta\theta^{\mathrm{T}}}\{\boldsymbol{\theta}(t_0)\}\right] = N\begin{pmatrix} 1 & h \\ h & h^2 \end{pmatrix} f(t_0) \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \Gamma_1(t_0).$$

Hence, by the Weak Law of Large Numbers,

$$N^{-1}\mathcal{H}\ell_{\theta\theta^{\mathrm{T}}}\{\boldsymbol{\theta}(t_0)\}\mathcal{H}^{-1} \xrightarrow{p} E\left[\ell_{\theta\theta^{\mathrm{T}}}\{\boldsymbol{\theta}(t_0)\}\right],$$

where $\mathcal{H} = \mathrm{diag}(1,h)\otimes\mathbf{I}_p$. Finally, by Slutsky's theorem,

$$h^{1/2}\mathbf{H}\ell_{\theta\theta^{\mathrm{T}}}\{\boldsymbol{\theta}(\mathbf{s})\}\mathbf{H}\mathbf{H}^{-1}\ell_{\theta}\{\boldsymbol{\theta}(\mathbf{s})\} \xrightarrow{D} \mathcal{N}\left(h^{1/2}\boldsymbol{\Delta}_\theta^{-1}\boldsymbol{\Lambda}, \boldsymbol{\Delta}_\theta^{-1}\boldsymbol{\Sigma}_\theta(\mathbf{s})\boldsymbol{\Delta}_\theta^{-1}\right),$$

where $\boldsymbol{\Delta}_\theta = (N\mathbf{H})^{-1}\mathrm{diag}(E\left[\ell_{\theta\theta^{\mathrm{T}}}\{\boldsymbol{\theta}(t_1)\}\right], \ldots, E\left[\ell_{\theta\theta^{\mathrm{T}}}\{\boldsymbol{\theta}(t_J)\}\right])\mathbf{H}^{-1}$. □

Proof of Lemma 3.1

We first present two lemmas taken from (Yao and Li, 2012).

Lemma 3.2. *Let* $(X_1, W_1), \ldots, (X_n, W_n)$ *be independent and identically distributed random vectors, where the* W_i *are scalar random variables. Further assume that for some* $k > 2$ *and interval* $[\mathcal{A}, \mathcal{B}]$*.*

$$E|W|^k < \infty \text{ and } \sup_{x \in [\mathcal{A},\mathcal{B}]} \int |w|^k \varrho(x, w) dw < \infty,$$

where $\varrho(\cdot, \cdot)$ *denotes the joint density of* (X, W)*. Let* $K(\cdot)$ *be a bounded positive function with a bounded support satisfying the Lipschitz condition. Then*

$$\sup_{x \in [\mathcal{A},\mathcal{B}]} \left| \frac{1}{n} \sum_{i=1}^{n} K_h(X_i - x) W_i - E\{K_h(X_i - x) W_i\} \right| = O_p \left[\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} \right]$$

provided that $h \to 0$*, for some* $\delta > 0$*,* $n^{1-2k^{-1}-2\delta} h \to \infty$*.*

Lemma 3.3. *Let* $\{x_1(t_{1j}), w_1(t_{1j}), \ldots, x_n(t_{nj}), w_n(t_{nj})\}$ *be independent and identically distributed random vectors for each* $j = 1, \ldots, \mathcal{J}$*, where* $w_i(t_{ij})$ *are scalar random variables. Further assume that for some* $k > 2$ *and interval* $[\mathcal{A}, \mathcal{B}]$

$$E|W(t_{.j})|^k < \infty \text{ and } \sup_{x \in [\mathcal{A},\mathcal{B}]} \int |w|^k \varrho_j(x, w) dw < \infty,$$

where $\varrho_j(\cdot, \cdot)$ *denotes the joint density of* $\{X_i(t_{ij}), W_i(t_{ij})\}$*. Let* $K(\cdot)$ *be a bounded positive function with a bounded support satisfying the Lipschitz condition. Then*

$$\sup_{x \in [\mathcal{A},\mathcal{B}]} \left| \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{\mathcal{J}} K_h\{x_i(t_{ij}) - x\} w_i(t_{ij}) - E\left[K_h\{x_i(t_{ij}) - x\} w_i(t_{ij})\right] \right| = O_p \left[\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} \right]$$

provided that $h \to 0$*, for some* $\delta > 0$*,* $n^{1-2k^{-1}-2\delta} h \to \infty$*.*

Proof. Let $e(t)$ and $\widehat{W}(t)$ be the residual and fitted value, respectively, observed at time t. Then

$$\sup_t |\varepsilon_1(t) - e(t)| \leq \sup_t |\widehat{W}(t) - E\{W(t)\}|.$$

We need to show that (†) $\sup_t |\widehat{W}(t) - E\{W(t)\}| \to 0$ as $n \to \infty$. By the Cauchy-Schwarz inequality,

$$\begin{aligned}\sup_t |\widehat{W}(t) - E\{W(t)\}| &= \sup_t |\mathbf{X}^{\mathrm{T}}(t)\hat{\boldsymbol{\beta}}(t) - \mathbf{X}^{\mathrm{T}}(t)\boldsymbol{\beta}(t)| \\ &\le \sup_t \|\mathbf{X}(t)\| \sup_t \|\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\|,\end{aligned}$$

where $\hat{\boldsymbol{\beta}}(t)$ is defined in (3.6). Using Lemma 3.3 and the same techniques used in Yao and Li (2012), we can show that $\sup_t \|\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\| = O_p(c_n) + O_p(h^2)$, where $c_n = \left\{\frac{\log(1/h)}{nh}\right\}^{1/2}$. Under the assumptions that $nh^8 \to 0$ and $nh^2/(logn)^2 \to \infty$, $\sup_t \|\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\| \to 0$ as $n \to \infty$. This result and condition (A) imply (†), which establishes the result. □

Proof of Theorem 3.2

Note that $\varpi_d(\mathcal{Z}, q) = (\partial^d/\partial\mathcal{Z}^d) l\{g^{-1}(\mathcal{Z}), q\}$ is linear in q for fixed $\mathcal{Z}$ such that

$$\begin{aligned}&\varpi_1\left(g\left[m\{t_0, \mathbf{x}_i(t_{ij})\}\right], m\{t_0, \mathbf{x}_i(t_{ij})\}\right) = 0 \text{ and} \\ &\varpi_2\left(g\left[m\{t_0, \mathbf{x}_i(t_{ij})\}\right], m\{t_0, \mathbf{x}_i(t_{ij})\}\right) = -\rho\{t_0, \mathbf{x}_i(t_{ij})\}.\end{aligned} \tag{3.19}$$

Proof. Let $\mathbf{s} = (t_1, \ldots, t_J)$, $J \ge 1$, be a set of distinct interior points in $\mathcal{T}$ and $\boldsymbol{\vartheta}^*(\mathbf{s}) = (\boldsymbol{\vartheta}^*(t_1), \ldots, \boldsymbol{\vartheta}^*(t_J))^{\mathrm{T}}$ with

$$\boldsymbol{\vartheta}^*(t_0) = \gamma_n^{-1}\left(a_1 - \alpha_1(t_0), \ldots, a_p - \alpha_p(t_0), h\{b_1 - \alpha_1'(t_0)\}, \ldots, h\{b_p - \alpha_p'(t_0)\}\right)^{\mathrm{T}},$$

where $t_0 \in \mathbf{s}$, and $\gamma_n = (Nh)^{-1/2}$. Let $\bar{\eta}\{t_0, t_{ij}, \mathbf{x}_i(t_{ij})\} = \bar{\eta}\{t_0, t_{ij}, x_{i1}(t_{ij}), \ldots, x_{ip}(t_{ij})\} = \sum_{r=1}^{p}\{\alpha_r(t_0) + \alpha_r'(t_0)(t_{ij} - t_0)x_{ir}(t_{ij})\}$. Hence,

$$\sum_{r=1}^{p}\{a_r + b_r(t_{ij} - t_0)\}X_{ir}(t_{ij}) = \bar{\eta}\{t_0, t_{ij}, \mathbf{X}_i(t_{ij})\} + \gamma_n \boldsymbol{\vartheta}^{*\mathrm{T}}(t_0)\mathbf{Z}_i(t_{ij}),$$

where $\mathbf{Z}_i(t_{ij}) = \left(\mathbf{X}_i^{\mathrm{T}}(t_{ij}), (t_{ij} - t_0)/h\, \mathbf{X}_i^{\mathrm{T}}(t_{ij})\right)^{\mathrm{T}}$. Hence the local log likelihood function (3.8) can be written as

$$\ell(\mathbf{a}, \mathbf{b}) = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i} \ell\left(g^{-1}\left[\bar{\eta}\{t_0, t_{ij}, \mathbf{X}_i(t_{ij})\} + \gamma_n \boldsymbol{\vartheta}^{*\mathrm{T}}(t_0)\mathbf{Z}_i(t_{ij})\right], Q_i(t_{ij})\right) K_h(t_{ij} - t_0),$$

which is a function of $\boldsymbol{\vartheta}^*(t_0)$, denoted by $\ell\{\boldsymbol{\vartheta}^*(t_0)\}$. Let

$$\hat{\boldsymbol{\vartheta}}^*(t_0) = \gamma_n^{-1}\left(\hat{a}_1 - \alpha_1(t_0), \ldots, \hat{a}_p - \alpha_p(t_0), h\{\hat{b}_1 - \alpha_1'(t_0)\}, \ldots, h\{\hat{b}_p - \alpha_p'(t_0)\}\right)^{\mathrm{T}}.$$

Since $(\hat{\mathbf{a}}, \hat{\mathbf{b}})^{\mathrm{T}}$ maximizes (3.8), then $\hat{\boldsymbol{\vartheta}}^*(t_0)$ maximizes $\ell\{\boldsymbol{\vartheta}^*(t_0)\}$, and $\hat{\boldsymbol{\vartheta}}^*(t_0)$ also maximizes the following function

$$\ell\{\boldsymbol{\vartheta}^*(t_0)\} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left(\ell\left[g^{-1}\left\{\bar{\eta}_{ij}(t_0) + \gamma_n \boldsymbol{\vartheta}^{*\mathrm{T}}(t_0)\mathbf{Z}_i(t_{ij})\right\}, Q_i(t_{ij})\right] - \ell\left[g^{-1}\{\bar{\eta}_{ij}(t_0)\}, Q_i(t_{ij})\right]\right) K_h(t_{ij} - t_0),$$

where $\bar{\eta}_{ij}(t_0) = \bar{\eta}\{t_0, t_{ij}, \mathbf{X}_i(t_{ij})\}$.

According to regularity condition (J), $\ell_n(\cdot)$ is concave in $\boldsymbol{\vartheta}^*(t_0)$. We locally approximate $\ell\{g^{-1}(\cdot), Q\}$ via the Taylor expansion and we obtain

$$\begin{aligned} \ell\{\boldsymbol{\vartheta}^*(t_0)\} = {} & \gamma_n \sum_{i=1}^{n}\sum_{j=1}^{n_i} \varpi_1\{\bar{\eta}_{ij}(t_0), Q_i(t_{ij})\}\, \boldsymbol{\vartheta}^{*\mathrm{T}}(t_0)\mathbf{Z}_i(t_{ij}) K\{(t_{ij} - t_0)/h\} \qquad (3.20) \\ & + \frac{\gamma_n^2}{2}\sum_{i=1}^{n}\sum_{j=1}^{n_i} \varpi_2\{\bar{\eta}_{ij}(t_0), Q_i(t_{ij})\}\{\boldsymbol{\vartheta}^{*\mathrm{T}}(t_0)\mathbf{Z}_i(t_{ij})\}^2 K\{(t_{ij} - t_0)/h\} \\ & + \frac{\gamma_n^3}{6}\sum_{i=1}^{n}\sum_{j=1}^{n_i} \varpi_3\{\eta_{ij}(t_0), Q_i(t_{ij})\}\{\boldsymbol{\vartheta}^{*\mathrm{T}}(t_0)\mathbf{Z}_i(t_{ij})\}^3 K\{(t_{ij} - t_0)/h\}, \end{aligned}$$

where $\eta_{ij}(t_0)$ is between $\bar{\eta}_{ij}(t_0)$ and $\bar{\eta}_{ij}(t_0) + \gamma_n \boldsymbol{\vartheta}^{*\mathrm{T}}(t_0)\mathbf{Z}_i(t_{ij})$. Let

$$R_n(t_0) = \gamma_n \sum_{i=1}^{n}\sum_{j=1}^{n_i} \varpi_1\{\bar{\eta}_{ij}(t_0), Q_i(t_{ij})\}\, \boldsymbol{\vartheta}^{*\mathrm{T}}(t_0)\mathbf{Z}_i(t_{ij}) K\{(t_{ij} - t_0)/h\},$$

and

$$\Delta_n(t_0) = \gamma_n^2 \sum_{i=1}^{n}\sum_{j=1}^{n_i} \varpi_2\{\bar{\eta}_{ij}(t_0), Q_i(t_{ij})\}\, \mathbf{Z}_i(t_{ij})\mathbf{Z}_i^{\mathrm{T}}(t_{ij}) K\{(t_{ij} - t_0)/h\}.$$

Hence, (3.20) becomes

$$\ell_n\{\boldsymbol{\vartheta}^*(t_0)\} = R_n^{\mathrm{T}}(t_0)\boldsymbol{\vartheta}^*(t_0) + \frac{1}{2}\boldsymbol{\vartheta}^{*\mathrm{T}}(t_0)\Delta_n(t_0)\boldsymbol{\vartheta}^*(t_0) + \frac{\gamma_n^3}{6}\sum_{i=1}^{n}\sum_{j=1}^{n_i} \varpi_3\{\eta_{ij}, Q_i(t_{ij})\}\{\boldsymbol{\vartheta}^{*\mathrm{T}}(t_0)\mathbf{Z}_i(t_{ij})\}^3 K\{(t_{ij} - t_0)/h\}. \qquad (3.21)$$

It is known that

$$\{\Delta_n(t_0)\}_{ij} = \{E\Delta_n(t_0)\}_{ij} + O_p\left([Var\{\Delta_n(t_0)\}_{ij}]^{1/2}\right).$$

The expected value of $\Delta_n(t_0)$ is equal to

$$E\{\Delta_n(t_0)\} = N\gamma_n E\left(\varpi_2\left[\bar{\eta}\{t_0, t_{ij}, \mathbf{X}_i(t_{ij})\}, m\{t_{ij}, \mathbf{X}_i(t_{ij})\}\right] K\{(t_{ij}-t_0)/h\}\mathbf{Z}_i(t_{ij})\mathbf{Z}_i^{\mathrm{T}}(t_{ij})\right).$$

Let

$$\eta\{t_{ij}, \mathbf{x}_i(t_{ij})\} = g\left[m\{t_{ij}, \mathbf{x}_i(t_{ij})\}\right] = \sum_{r=1}^{p} \alpha_r(t_{ij}) x_{ir}(t_{ij}).$$

Using Taylor expansion of $\eta\{t_{ij}, \mathbf{x}_i(t_{ij})\}$ around t_0 with $|t_{ij} - t_0| < h$ and the first result in (3.19), we have the following:

$$\eta\{t_{ij}, \mathbf{x}_i(t_{ij})\} = \bar{\eta}\{t_0, t_{ij}, \mathbf{x}_i(t_{ij})\} + \frac{(t_{ij}-t_0)^2}{2}\eta''\{t_0, \mathbf{x}_i(t_{ij})\} + o(h^2),$$

where $\eta''\{t_{ij}, \mathbf{x}_i(t_{ij})\} = (\partial/\partial t_{ij}^2)\eta\{t_{ij}, \mathbf{x}_i(t_{ij})\} = \sum_{r=1}^{p} \alpha_r''(t_{ij}) x_{ir}(t_{ij})$. Furthermore, we have the following results:

$$\varpi_1\left[\bar{\eta}\{t_0, t_{ij}, \mathbf{x}_i(t_{ij})\}, m\{t_{ij}, \mathbf{x}_i(t_{ij})\}\right] = \rho\{t_0, \mathbf{x}_i(t_{ij})\}\frac{(t_{ij}-t_0)^2}{2}\eta''\{t_0, \mathbf{x}_i(t_{ij})\} + o(h^2), \quad (3.22)$$

and similarly,

$$\varpi_2\left[\bar{\eta}\{t_0, t_{ij}, \mathbf{x}_i(t_{ij})\}, m\{t_{ij}, \mathbf{x}_i(t_{ij})\}\right] = -\rho\{t_0, \mathbf{x}_i(t_{ij})\} + o(1). \quad (3.23)$$

Using (3.19) and (3.23), we obtain the following:

$$E\{\Delta_n(t_0)\} \to -f(t_0)\begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \Gamma_2(t_0) = -\Delta_\vartheta(t_0). \quad (3.24)$$

Similarly, $Var\left[\{\Delta_n(t_0)\}_{ij}\right] = O\{(Nh)^{-1}\}$. Thus,

$$\Delta_n(t_0) = -\Delta_\vartheta(t_0) + o_p(1). \quad (3.25)$$

For the last term in (3.21), we have the following result:

$$O\left(n\gamma_n^3 E\left|\varpi_3\{\eta_{1j}, Q_1(t_{1j})\}\mathbf{X}_1^3(t_{1j})K\{(t_{1j}-t_0)/h\}\right|\right) = O(\gamma_n), \tag{3.26}$$

which follows from $K(\cdot)$ being bounded, $\varpi_3(.,.)$ being linear in $Q_1(t_{1j})$, $E\{|Q_1(t_{1j})| \mid t_{1j}, \mathbf{X}_1(t_{1j})\} < \infty$ and regularity condition (M). Combining (3.21), (3.24), (3.25) and (3.26), we obtain the following:

$$\ell_n\{\boldsymbol{\vartheta}^*(t_0)\} = R_n^{\mathrm{T}}(t_0)\boldsymbol{\vartheta}^*(t_0) - \frac{1}{2}\boldsymbol{\vartheta}^{*\mathrm{T}}(t_0)\Delta_\vartheta(t_0)\boldsymbol{\vartheta}^*(t_0) + o_p(1).$$

Using the quadratic approximation lemma (see Fan and Gijbels, 1996, p.210),

$$\hat{\boldsymbol{\vartheta}}^*(t_0) = \Delta_\vartheta^{-1}(t_0)R_n(t_0) + o_p(1),$$

if $R_n(t_0)$ is a sequence of stochastically bounded random vectors. Furthermore, define $\mathbf{R}_n(\mathbf{s}) = (R_n(t_1), \ldots, R_n(t_J))^{\mathrm{T}}$ and $\boldsymbol{\Xi} = \mathrm{diag}(\Delta_\vartheta(t_1)^{-1}, \ldots, \Delta_\vartheta(t_J)^{-1})$. Hence,

$$\hat{\boldsymbol{\vartheta}}^*(\mathbf{s}) = \boldsymbol{\Xi}\,\mathbf{R}_n(\mathbf{s}) + o_p(1).$$

The above equation implies that asymptotic normality of $\hat{\boldsymbol{\vartheta}}^*(\mathbf{s})$ follows from that of $\mathbf{R}_n(\mathbf{s})$. We first establish the asymptotic normality of $\mathbf{R}_n(\mathbf{s})$. Note that $R_n(t_0)$ can be written as follows:

$$R_n(t_0) = \gamma_n \sum_{i=1}^{n} \psi_i(t_0), \tag{3.27}$$

where $\psi_i(t_0) = \sum_{j=1}^{n_i} \varpi_1\{\bar{\eta}_{ij}(t_0), Q_i(t_{ij})\}\,\mathbf{Z}_i(t_{ij})K\{(t_{ij}-t_0)/h\}$. This implies that $R_n(t_0)$ is a sum of independent random vectors. To prove the asymptotic normality of $\mathbf{R}_n(\mathbf{s})$, we first find the mean and the covariance matrix of $\mathbf{R}_n(\mathbf{s})$ and then we check the Lyapunov condition. Let $\boldsymbol{\Lambda}_\vartheta = E\{\mathbf{R}_n(\mathbf{s})\} = (\boldsymbol{\Lambda}_\vartheta(t_1), \ldots, \boldsymbol{\Lambda}_\vartheta(t_J))^{\mathrm{T}}$ with $\boldsymbol{\Lambda}_\vartheta(t_0) = E\{\mathbf{R}_n(t_0)\}$ and $t_0 \in \mathbf{s}$. Using (3.22), the mean of $R_n(t_0)$ is calculated as follows:

$$
\begin{aligned}
E\{R_n(t_0)\} &= E\left(N\gamma_n \varpi_1\left[\bar{\eta}_{ij}(t_0), m\{t_{ij}, \mathbf{X}_i(t_{ij})\}\right] \mathbf{Z}_i(t_{ij}) K\{(t_{ij}-t_0)/h\}\right) \qquad (3.28)\\
&= \frac{h^2 f(t_0)}{2\gamma_n} \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} \otimes \Gamma_2(t_0)\boldsymbol{\alpha}''(t_0) + o(1).
\end{aligned}
$$

Define $\boldsymbol{\Sigma}_R(\mathbf{s}) = \mathrm{cov}\{\mathbf{R}_n(\mathbf{s}), \mathbf{R}_n(\mathbf{s})\}$. Using (3.28) and the definition of $\varpi_1(\cdot,\cdot)$, the covariance matrix is calculated as follows:

$$
\mathrm{cov}\{R_n(t_{01}), R_n(t_{02})\} = E\{R_n(t_{01}), R_n(t_{02})\} - E\{R_n(t_{01})\}E\{R_n(t_{02})\},
$$

where $t_{01}, t_{02} \in \mathbf{s}$ and by using (3.27)

$$
\begin{aligned}
E\{R_n(t_{01}), R_n(t_{02})\} &= E\Big(\left[N\gamma_n\varpi_1\left\{\bar{\eta}_{ij}(t_{01}), Q_i(t_{ij})\right\}\mathbf{Z}_i(t_{ij})K\{t_{ij}-t_{01}/h\}\right]\\
&\times \left[N\gamma_n\varpi_1\left\{\bar{\eta}_{ij}(t_{02}), Q_i(t_{ij})\right\}\mathbf{Z}_i(t_{ij})K\{(t_{ij}-t_{02})/h\}\right]\Big)\\
&= (Nh)^{-1}\left[\sum_{i=1}^{n} E\left\{\psi_i(t_{01})\psi_i(t_{02})\right\} + \sum_{i_1\neq i_2} E\left\{\psi_{i_1}(t_{01})\psi_{i_2}(t_{02})\right\}\right].
\end{aligned}
$$

We first calculate $E\left\{\psi_i(t_{01})\psi_i(t_{02})\right\}$ as follows:

$$
\begin{aligned}
E\left\{\psi_i(t_{01})\psi_i(t_{02})\right\} &= \sum_{j=1}^{n_i} E\Big[\varpi_1\left\{\bar{\eta}_{ij}(t_{01}), Q_i(t_{ij})\right\}\varpi_1\left\{\bar{\eta}_{ij}(t_{02}), Q_i(t_{ij})\right\}\mathbf{Z}_i(t_{ij})\mathbf{Z}_i^{\mathrm{T}}(t_{ij})\\
&\times K\{(t_{ij}-t_{01})/h\}K\{(t_{ij}-t_{02})/h\}\Big]\\
&+ \sum_{j_1\neq j_2}^{n_i} E\Big[\varpi_1\left\{\bar{\eta}_{ij_1}(t_{01}), Q_i(t_{ij_1})\right\}\varpi_1\left\{\bar{\eta}_{ij_2}(t_{02}), Q_i(t_{ij_2})\right\}\mathbf{Z}_i(t_{ij_1})\mathbf{Z}_i^{\mathrm{T}}(t_{ij_2})\\
&\times K\{(t_{ij_1}-t_{01})/h\}K\{(t_{ij_2}-t_{02})/h\}\Big].
\end{aligned} \qquad (3.29)
$$

Let's deal with the first term in (3.29).

$$\sum_{j=1}^{n_i} E\Big[\varpi_1\left\{\bar{\eta}_{ij}(t_{01}), Q_i(t_{ij})\right\}\varpi_1\left\{\bar{\eta}_{ij}(t_{02}), Q_i(t_{ij})\right\}\mathbf{Z}_i(t_{ij})\mathbf{Z}_i^{\mathrm{T}}(t_{ij})K\{(t_{ij}-t_{01})/h\}K\{(t_{ij}-t_{02})/h\}\Big]$$

$$= \sum_{j=1}^{n_i} E\Big[\varpi_1^2\left\{\bar{\eta}_{ij}(t_{01}), Q_i(t_{ij})\right\}\mathbf{Z}_i(t_{ij})\mathbf{Z}_i^{\mathrm{T}}(t_{ij})K^2\{(t_{ij}-t_{01})/h\}\Big]$$

$$= \begin{cases} n_i\, hf(t_{01})\begin{pmatrix}\nu_0 & \nu_1\\ \nu_1 & \nu_2\end{pmatrix}\otimes\Gamma_2(t_{01})+o(n_ih), & \text{if } t_{01}=t_{02}\\ o(n_ih), & \text{if } t_{01}\neq t_{02}.\end{cases}$$

Let's deal with the second term in (3.29).

$$\sum_{j_1\neq j_2}^{n_i} E\Big[\varpi_1\left\{\bar{\eta}_{ij_1}(t_{01}), Q_i(t_{ij_1})\right\}\varpi_1\left\{\bar{\eta}_{ij_2}(t_{02}), Q_i(t_{ij_2})\right\}\mathbf{Z}_i(t_{ij_1})\mathbf{Z}_i^{\mathrm{T}}(t_{ij_2})K\{(t_{ij_1}-t_{01})/h\}K\{(t_{ij_2}-t_{02})/h\}\Big]$$

$$= \sum_{j_1\neq j_2}^{n_i}\int\int E\Big[\varpi_1\left\{\bar{\eta}_{ij_1}(t_{01}), Q_i(t_{ij_1})\right\}\varpi_1\left\{\bar{\eta}_{ij_2}(t_{02}), Q_i(t_{ij_2})\right\}\mathbf{Z}_i(t_{ij_1})\mathbf{Z}_i^{\mathrm{T}}(t_{ij_2})$$

$$\times\ K\{(t_{ij_1}-t_{01})/h\}K\{(t_{ij_2}-t_{02})/h\}\mid t_{ij_1}=\varsigma_1, t_{ij_2}=\varsigma_2\Big]f(\varsigma_1)f(\varsigma_2)d\varsigma_1 d\varsigma_2$$

$$= \begin{cases} n_i(n_i-1)h^2f(t_{01})f(t_{02})\begin{pmatrix}1 & \mu_1\\ \mu_1 & \mu_1^2\end{pmatrix}\otimes\Gamma_3(t_{01},t_{02})+o\{n_i(n_i-1)h^2\}, & \text{if } t_{01}=t_{02}\\ n_i(n_i-1)h^2f^2(t_{01})\begin{pmatrix}1 & \mu_1\\ \mu_1 & \mu_1^2\end{pmatrix}\otimes\Gamma_3(t_{01},t_{01})+o\{n_i(n_i-1)h^2\}, & \text{if } t_{01}\neq t_{02}.\end{cases}$$

Finally, there exists a positive constant $M<\infty$ such that

$$\left|(Nh)^{-1}\sum_{i_1\neq i_2}E\left\{\psi_{i_1}(t_{01})\psi_{i_2}(t_{02})\right\}-E\left\{(Nh)^{-1/2}\sum_{i=1}^{n}\psi_i(t_{01})\right\}E\left\{(Nh)^{-1/2}\sum_{i=1}^{n}\psi_i(t_{02})\right\}\right| \tag{3.30}$$

$$\leq Nh\left\{1-N^{-2}\sum_{i=1}^{n}\left(n_i\sum_{i'\neq i}n_{i'}\right)\right\}M\to 0,\ \text{ as } n\to\infty.$$

The inequality in (3.30) holds because, by the regularity condition (F),

$$1-N^{-2}\sum_{i=1}^{n}\left(n_i\sum_{i'\neq i}n_{i'}\right)=N^{-2}\sum_{i=1}^{n}n_i^2=o\{(Nh)^{-1}\}.$$

Hence, $\text{cov}\{R_n(t_{01}), R_n(t_{02})\}$ is equal to

$$\text{cov}\{R_n(t_{01}), R_n(t_{02})\} = \begin{cases} f(t_{01}) \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix} \otimes \Gamma_2(t_{01}) \\ +(Nh)^{-1}(\sum_{i=1}^n n_i^2 - N)h^2 f^2(t_{01}) \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_1^2 \end{pmatrix} \otimes \Gamma_3(t_{01}, t_{01}) \\ +o\{(Nh)^{-1}(\sum_{i=1}^n n_i^2 - N)h^2\}, \text{ if } t_{01} = t_{02} \\ (Nh)^{-1}(\sum_{i=1}^n n_i^2 - N)h^2 f(t_{01}) f(t_{02}) \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_1^2 \end{pmatrix} \otimes \Gamma_3(t_{01}, t_{02}) \\ +o\{(Nh)^{-1}(\sum_{i=1}^n n_i^2 - N)h^2\}, \text{ if } t_{01} \neq t_{02}. \end{cases} \tag{3.31}$$

Since $h = N^{-1/5}h_0$, and $\lim_{n\to\infty} N^{-6/5}\left(\sum_{i=1}^n n_i^2\right) = \lambda$. It is easy to see that

$$N^{-1}h\left(\sum_{i=1}^n n_i^2 - N\right) = N^{-6/5}\left(\sum_{i=1}^n n_i^2 - N\right) h_0 \to \lambda h_0 \text{ as } n \to \infty.$$

Rewriting (3.31) using the equation above,

$$\text{cov}\{R_n(t_{01}), R_n(t_{02})\} = \begin{cases} f(t_{01}) \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix} \otimes \Gamma_2(t_{01}) \\ +\lambda h_0 f^2(t_{01}) \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_1^2 \end{pmatrix} \otimes \Gamma_3(t_{01}, t_{01}), & \text{if } t_{01} = t_{02} \\ \lambda h_0 f(t_{01}) f(t_{02}) \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_1^2 \end{pmatrix} \otimes \Gamma_3(t_{01}, t_{02}), & \text{if } t_{01} \neq t_{02}. \end{cases}$$

We now employ the Cramer-Wold theorem to show the asymptotic normality of $\mathbf{R}_n(\mathbf{s})$. For any unit vector $\mathbf{v} \in \mathcal{R}^{2pJ}$, if

$$\left\{\mathbf{v}^{\mathrm{T}}\, \boldsymbol{\Sigma}_R(\mathbf{s})\, \mathbf{v}\right\}^{-1/2} \left[\mathbf{v}^{\mathrm{T}}\mathbf{R}_n(\mathbf{s}) - \mathbf{v}^{\mathrm{T}} E\{\mathbf{R}(\mathbf{s})\}\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{3.32}$$

then

$$\{\mathbf{\Sigma}_R(\mathbf{s})\}^{-1/2}\left[\mathbf{R}_n(\mathbf{s}) - E\{\mathbf{R}(\mathbf{s})\}\right] \stackrel{\mathcal{D}}{\rightarrow} \mathcal{N}(0, \mathbf{I}_{2pJ}). \tag{3.33}$$

Since (3.32) satisfies the Lyapunov condition, (3.33) follows. Finally, we have the following result:

$$\hat{\boldsymbol{\vartheta}}^*(\mathbf{s}) - \mathbf{\Xi}\,\mathbf{\Lambda}_\vartheta \stackrel{\mathcal{D}}{\rightarrow} \mathcal{N}\left(0, \mathbf{\Xi}\,\mathbf{\Sigma}_\vartheta(\mathbf{s})\mathbf{\Xi}\right),$$

where $\mathbf{\Lambda}_\vartheta = E\{\mathbf{R}(\mathbf{s})\}$ and $\mathbf{\Sigma}_\vartheta(\mathbf{s}) = \mathbf{\Sigma}_R(\mathbf{s})$. This completes the proof of theorem 3. □

Chapter 4

A Semiparametric Approach for Modeling Longitudinal Binary and Continuous Responses Jointly

In this chapter we describe our semiparametric framework for joint modeling of longitudinal binary and continuous responses. Notation in this chapter is similar to that of Chapter 3. In Section 4.1 we propose our joint model for longitudinal binary and continuous responses. We describe our two-stage estimation procedure and the asymptotic properties of the resulting estimators in Sections 4.2 and 4.3, respectively. Section 4.4 presents the results of a simulation study designed to show the finite sample behavior of our estimators. In Section 4.5 we further illustrate the proposed methodology by analyzing a data from the Women's Interagency HIV Study. In Section 4.6 we present the technical conditions and proofs.

4.1 Joint Models

In this section we introduce the joint models for estimating the association of a longitudinal binary and a continuous response. For the i^{th} subject, $i = 1, \ldots, n$, denote the binary response measured at time point t_{ij} by $Q_i(t_{ij})$, the continuous response by $W_i(t_{ij})$, where $j = 1, \ldots, n_i$. Furthermore, define the latent variable underlying the binary response by $Y_i(t_{ij})$. Let $\mathbf{X}_i(t_{ij}) =$

$(X_{i1}(t_{ij}), \ldots, X_{ip}(t_{ij}))^{\mathrm{T}}$ be the vector of predictors, $\boldsymbol{\beta}_W = (\beta_{W1}, \ldots, \beta_{Wp})^{\mathrm{T}}$, and $\boldsymbol{\beta}_Y = (\beta_{Y1} \ldots, \beta_{Yp})^{\mathrm{T}}$ be the unknown regression coefficient vectors for the continuous variable and the latent variable, respectively. In addition, let $\alpha_W(t_{ij})$, and $\alpha_Y(t_{ij})$ be the nonparametric smooth baseline functions for the continuous variable and the latent variable, respectively. Consider the bivariate model

$$\begin{aligned} W_i(t_{ij}) &= \alpha_W(t_{ij}) + \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\beta}_W + \varepsilon_{1i}(t_{ij}), \qquad (4.1) \\ Y_i(t_{ij}) &= \alpha_Y(t_{ij}) + \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\beta}_Y + \varepsilon_{2i}(t_{ij}), \end{aligned}$$

where $\varepsilon_{1i}(t_{ij}) \sim \mathcal{N}\left\{0, \sigma_1^2(t_{ij})\right\}$ with $\mathrm{corr}\left\{\varepsilon_{1i}(t_{ij}), \varepsilon_{1i}(t_{ij'})\right\} = \rho_1(t_{ij}, t_{ij'})$ for $j \neq j'$, $\varepsilon_{2i}(t_{ij}) \sim \mathcal{N}\left\{0, \sigma_2^2(t_{ij})\right\}$ with $\mathrm{corr}\left\{\varepsilon_{2i}(t_{ij}), \varepsilon_{2i}(t_{ij'})\right\} = \rho_2(t_{ij}, t_{ij'})$ for $j \neq j'$, and $j, j' = 1, \ldots, n_i$. Moreover, we define $\boldsymbol{\varepsilon}_i(t_{ij}) = (\boldsymbol{\varepsilon}_{1i}(t_{ij}), \boldsymbol{\varepsilon}_{2i}(t_{ij}))^{\mathrm{T}}$, and assume that $\boldsymbol{\varepsilon}_i(t_{ij})$ follows a bivariate normal distribution with $\mathrm{corr}\left\{\varepsilon_{1i}(t_{ij}), \varepsilon_{2i}(t_{ij})\right\} = \tau(t_{ij})$, and $\mathrm{corr}\left\{\varepsilon_{1i}(t_{ij}), \varepsilon_{2i}(t_{ij'})\right\} = \rho_{12}(t_{ij}, t_{ij'})$ for $j \neq j'$. The relation between the binary variable and the latent variable is defined as $Q_i(t_{ij}) = 1$ if $Y_i(t_{ij}) > 0$, and $Q_i(t_{ij}) = 0$ if $Y_i(t_{ij}) \leq 0$. Hence, the probit model for the binary response $Q_i(t_{ij})$ is defined as follows:

$$P\left\{Q_i(t_{ij}) = 1 | \mathbf{X}_i(t_{ij})\right\} = \Phi\left\{\frac{\alpha_Y(t_{ij}) + \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\beta}_Y}{\sigma_2(t_{ij})}\right\}.$$

The joint distribution of continuous and the binary responses is challenging to obtain; however, the marginal distributions are readily derived. Hence, we decompose the joint distribution of the continuous variable and the binary variable into two components: a marginal model for the continuous variable $W_i(t_{ij})$ and a conditional model for $Q_i(t_{ij})$ given $W_i(t_{ij})$,

$$f\left\{q_i(t_{ij}), w_i(t_{ij})\right\} = f_W\left\{w_i(t_{ij})\right\} f\left\{q_i(t_{ij}) | w_i(t_{ij})\right\},$$

where $j = 1, \ldots, n_i$. The marginal model for the continuous response is defined as in (4.1). To derive the conditional model for $Q_i(t_{ij})$ given $W_i(t_{ij})$, we first obtain the conditional model $Y_i(t_{ij})|W_i(t_{ij})$. The standard normal theory shows that the conditional distribution $Y_i(t_{ij})|W_i(t_{ij})$ follows a Gaussian distribution, and the mean of this conditional distribution depends on the error from the marginal model of the continuous response,

$$Y_i(t_{ij})|W_i(t_{ij}) \sim \mathcal{N}\left[\mu_i(t_{ij}), \sigma_2^2(t_{ij})\left\{1-\tau(t_{ij})\right\}\right],$$

where

$$\mu_i(t_{ij}) = \alpha_Y(t_{ij}) + \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\beta}_Y + \frac{\sigma_2(t_{ij})}{\sigma_1(t_{ij})}\tau(t_{ij})\varepsilon_{1i}(t_{ij}), \tag{4.2}$$

and

$$\varepsilon_{1i}(t_{ij}) = W_i(t_{ij}) - \left\{\alpha_W(t_{ij}) + \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\beta}_W\right\}$$

is the error from the marginal model of the continuous response. Hence,

$$P\left\{Q_i(t_{ij}) = 1|W_i(t_{ij})\right\} = \Phi\left[\frac{\mu_i(t_{ij})}{\sqrt{\sigma_2^2(t_{ij})\left\{1-\tau(t_{ij})\right\}}}\right], \tag{4.3}$$

where $\mu_i(t_{ij})$ is defined in (4.2). Note that not all of the parameters in model (4.3) are estimable. We reparameterize (4.3) to a more parsimonious and fully estimable form:

$$P\left\{Q_i(t_{ij}) = 1|W_i(t_{ij})\right\} = \Phi\left\{\beta_0^*(t_{ij}) + \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\beta}^*(t_{ij}) + \beta_{p+1}^*(t_{ij})\varepsilon_{1i}(t_{ij})\right\}, \tag{4.4}$$

where $\boldsymbol{\beta}^*(t_{ij}) = \left(\beta_1^*(t_{ij}), \ldots, \beta_p^*(t_{ij})\right)^{\mathrm{T}}$. We link the continuous response with the binary response in a probit regression model (4.4) using the error from the marginal model as a covariate. From model (4.4), we see that $\beta_{p+1}^*(t_{ij}) = \left\{\sigma_2(t_{ij})/\sigma_1(t_{ij})\right\}\tau(t_{ij})$ and assuming that $\sigma_1^2(t_{ij})$ is equal to $\sigma_2^2(t_{ij})$, then $\tau(t_{ij})$ would be estimable. Hence, assuming the equality of the variances, $\sigma_1^2(t_{ij})$ and $\sigma_2^2(t_{ij})$, allows us to obtain the estimate of the time-varying association $\tau(t_{ij})$. However, even if the equality of variances assumption is invalid, the proposed semiparametric method can show the significance and the direction of the time-varying association along with its estimate up to the positive factor $\left\{\sigma_2(t_{ij})/\sigma_1(t_{ij})\right\}$.

4.2 Estimation Procedure

We propose a two-stage estimation procedure to estimate the time-varying correlation between a longitudinal binary and a continuous response. This procedure also allows us to estimate the regression coefficients in the marginal model of the continuous response. In the first stage we fit a partially linear model (Fan and Li, 2004; Lin and Ying, 2001; Martinussen and Scheike, 1999; Moyeed and Diggle, 1994; Zeger and Diggle, 1994) to the marginal model for the continuous response (4.1). In this stage we employ the back-fitting algorithm (Zeger and Diggle, 1994). Each iteration of the algorithm has the following steps:

1. Given the generalized least squares estimate $\hat{\boldsymbol{\beta}}_W^{(k)}$ at the k^{th} iteration, we calculate $W_i^*(t_{ij}) = W_i(t_{ij}) - \mathbf{X}_i^{\mathrm{T}}(t_{ij})\hat{\boldsymbol{\beta}}_W$. We fit a nonparametric model to solve the regression problem $W_i^*(t_{ij}) = \alpha_W(t_{ij}) + \varepsilon_{1i}(t_{ij})$, and obtain $\hat{\alpha}_W^{(k)}(t_{ij})$. In our procedure local linear fitting (Fan and Gijbels, 1996) is used, because of its desirable properties such as its minimax efficiency and being design-adaptive (Fan, 1993). Using the Taylor expansion, we locally approximate the regression coefficient function in a neighborhood of a fixed point t_0,

$$\alpha_W(t) \approx \alpha_W(t_0) + \alpha_W'(t_0)(t - t_0) \equiv a_W + b_W(t - t_0).$$

We minimize the following equation:

$$\ell(a_W, b_W) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{W_i^*(t_{ij}) - a_W - b_W(t_{ij} - t_0)\right\}^2 K_h(t_{ij} - t_0),$$

with respect to (a_W, b_W), where $K_h(\cdot) = h^{-1}K(\cdot/h)$ with $K(\cdot)$ as the kernel function. Define $\mathbf{W}^* = (\mathbf{W}_1^*, \ldots, \mathbf{W}_n^*)^{\mathrm{T}}$ with $\mathbf{W}_i^* = (W_i^*(t_{i1}), \ldots, W_i^*(t_{in_i}))^{\mathrm{T}}$, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, and

$N = \sum_{i=1}^{n} n_i$. The solution to the least squares algorithm is calculated as follows:

$$\hat{a}_W = (1,0)(\mathbf{\Gamma}_{t_0}^{\mathrm{T}} \boldsymbol{\kappa}_{t_0} \mathbf{\Gamma}_{t_0})^{-1} \mathbf{\Gamma}_{t_0}^{\mathrm{T}} \boldsymbol{\kappa}_{t_0} \mathbf{W}^{*}, \tag{4.5}$$

where $\mathbf{\Gamma}_t = (\mathbf{1}, \mathbf{T}_t \mathbf{1})$, $\mathbf{1}$ is an N-dimensional column vector with each entry being one, $\mathbf{T}_t$ an N-dimensional vector with each element equal to $(t_{ij} - t)$, and $\boldsymbol{\kappa}_t$ is an $N \times N$ diagonal matrix with each entry being $K_h(t_{ij} - t)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. The estimate of the baseline function at the k^{th} iteration is equal to $\hat{a}_W$.

2. Given $\hat{\alpha}_W^{(k)}(t_{ij})$, we calculate $\widetilde{W}_i(t_{ij}) = W_i(t_{ij}) - \hat{\alpha}_W^{(k)}(t_{ij})$, and obtain the generalized least squares estimate $\hat{\boldsymbol{\beta}}_W^{(k+1)}$ by fitting the model $\widetilde{W}_i(t_{ij}) = \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\beta}_W + \varepsilon_{1i}(t_{ij})$ with $i = 1, \ldots, n$, and $j = 1, \ldots, n_i$.

We iterate both steps until the algorithm converges. When the algorithm converges, we obtain the estimates for $\boldsymbol{\beta}_W$ and $\alpha_W(\cdot)$.

The next step is to derive pointwise confidence intervals for the baseline function, $\alpha_W(\cdot)$. This requires an estimator for the asymptotic covariance matrix. We propose the following sandwich formula that is obtained using conventional techniques:

$$\widehat{\mathrm{cov}}\{\hat{\alpha}_W(t_0)\} \approx (1,0)(\mathbf{\Gamma}_{t_0}^{\mathrm{T}} \boldsymbol{\kappa}_{t_0} \mathbf{\Gamma}_{t_0})^{-1} \left(\mathbf{\Gamma}_{t_0}^{\mathrm{T}} \boldsymbol{\kappa}_{t_0} \mathcal{Q} \boldsymbol{\kappa}_{t_0} \mathbf{\Gamma}_{t_0}\right) (\mathbf{\Gamma}_{t_0}^{\mathrm{T}} \boldsymbol{\kappa}_{t_0} \mathbf{\Gamma}_{t_0})^{-1} (1,0)^{\mathrm{T}}, \tag{4.6}$$

where $\mathcal{Q} = \mathrm{diag}(\mathcal{E}_1, \ldots, \mathcal{E}_n)$ with $\mathcal{E}_i = (e_i^2(t_{i1}), \ldots, e_i^2(t_{i\,n_i}))$ and $e_i(t_{ij}) = W_i(t_{ij}) - \hat{\alpha}_W(t_{ij}) - \mathbf{X}_i^{\mathrm{T}}(t_{ij})\hat{\boldsymbol{\beta}}_W$, $i = 1, \ldots, n$, and $j = 1, \ldots, n_i$.

In the second stage we fit a generalized time-varying coefficient model to the conditional model (4.4). Cai et al. (2000) introduced the generalized varying coefficient models for independent and identically distributed data, in this chapter, as in Chapter 3, we adapt these models to a longitudinal setting. We start by locally approximating the functions in a neighborhood of a fixed point t_0 via the Taylor expansion,

$$\beta_r^*(t) \approx \beta_r^*(t_0) + \beta_r^{*\prime}(t - t_0) \equiv a_r^* + b_r^*(t - t_0),$$

for $r = 0, \ldots, p+1$. Let $\mathbf{a}^* = (a_0^*, \ldots, a_{p+1}^*)^{\mathrm{T}}$ and $\mathbf{b}^* = (b_0^*, \ldots, b_{p+1}^*)^{\mathrm{T}}$. For the i^{th} subject, let $\mathbf{X}_i^*(t_{ij}) = \left(1, \mathbf{X}_i^{\mathrm{T}}(t_{ij}), e_i(t_{ij})\right)^{\mathrm{T}}$ denote the design matrix with $e_i(t_{ij})$ as the residual from the marginal model. We maximize the local likelihood,

$$\ell_n(\mathbf{a}^*, \mathbf{b}^*) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \ell\left(g^{-1}\left[\sum_{r=0}^{p+1} \left\{a_r^* + b_r^*(t_{ij} - t_0)\right\} X_{ir}^*(t_{ij})\right], Q_i(t_{ij})\right) K_h(t_{ij} - t_0),$$

where $g(\cdot)$ is the link function. For our model (4.4), the link function is probit. The local likelihood with probit link is defined as

$$\begin{aligned} \ell_n(\mathbf{a}^*, \mathbf{b}^*) = & \frac{1}{N} \sum_{Q_i(t_{ij})=1} \log\left(\phi\left[\sum_{r=0}^{p+1} \left\{a_r^* + b_r^*(t_{ij} - t_0)\right\} X_{ir}^*(t_{ij})\right]\right) K_h(t_{ij} - t_0) \\ & + \frac{1}{N} \sum_{Q_i(t_{ij})=0} \log\left(1 - \phi\left[\sum_{r=0}^{p+1} \left\{a_r^* + b_r^*(t_{ij} - t_0)\right\} X_{ir}^*(t_{ij})\right]\right) K_h(t_{ij} - t_0), \end{aligned} \tag{4.7}$$

where $\phi(\cdot)$ is the probability density function for the standard normal distribution. In order to find solutions to (4.7), we adapt the iterative local maximum likelihood algorithm described in Cai et al. (2000) to a longitudinal setting. Let $a_r^{*(k)}$ and $b_r^{*(k)}$ be the values of a_r^* and b_r^* at the k^{th} iteration, respectively. Let $\ell_n'(\mathbf{a}^*, \mathbf{b}^*)$ and $\ell_n''(\mathbf{a}^*, \mathbf{b}^*)$ be the gradient and Hessian matrix for the local likelihood (4.7),

$$\ell_n'(\mathbf{a}^*, \mathbf{b}^*) = \begin{pmatrix} v_{n,0} \\ v_{n,1} \end{pmatrix} \text{ and } \ell_n''(\mathbf{a}^*, \mathbf{b}^*) = \begin{pmatrix} H_{n,0} & H_{n,1} \\ H_{n,1} & H_{n,2} \end{pmatrix},$$

where

$$
\begin{aligned}
v_{n,s} = & \quad \frac{1}{N} \sum_{Q_i(t_{ij})=1} \phi(z_i)\Phi^{-1}(z_i)\mathbf{X}_i^*(t_{ij})K_h(t_{ij}-t_0)(t_{ij}-t_0)^s \\
& - \frac{1}{N} \sum_{Q_i(t_{ij})=0} \phi(z_i)\left\{1-\Phi(z_i)\right\}^{-1}\mathbf{X}_i^*(t_{ij})K_h(t_{ij}-t_0)(t_{ij}-t_0)^s, \\
H_{n,d} = & - \frac{1}{N}\Big(\sum_{Q_i(t_{ij})=1} \left\{z_i\phi(z_i)\Phi^{-1}(z_i)+\phi^2(z_i)\Phi^{-2}(z_i)\right\}\mathbf{X}_i^*(t_{ij})\mathbf{X}_i^{*\mathrm{T}}(t_{ij})K_h(t_{ij}-t_0)(t_{ij}-t_0)^d \\
& + \sum_{Q_i(t_{ij})=0} \left[\phi^2(z_i)\left\{1-\Phi(z_i)\right\}^{-2} - z_i\phi(z_i)\left\{1-\Phi(z_i)\right\}^{-1}\right]\mathbf{X}_i^*(t_{ij})\mathbf{X}_i^{*\mathrm{T}}(t_{ij})K_h(t_{ij}-t_0)(t_{ij}-t_0)^d\Big), \\
z_i = & \quad \sum_{r=0}^{p+1}\left\{a_r^{*(k)}+b_r^{*(k)}(t_{ij}-t_0)\right\}X_{ir}^*(t_{ij}),\ s=0,1,\ d=0,1,2,
\end{aligned}
$$

and $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution. We update $(\mathbf{a}^*, \mathbf{b}^*)$ according to

$$
\begin{pmatrix} \mathbf{a}^{*(k+1)} \\ \mathbf{b}^{*(k+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{a}^{*(k)} \\ \mathbf{b}^{*(k)} \end{pmatrix} - \left\{\ell_n''(\mathbf{a}^*, \mathbf{b}^*)\right\}^{-1}\ell_n'(\mathbf{a}^*, \mathbf{b}^*).
$$

The solution of this iterative regression algorithm satisfies $\ell(\mathbf{a}^*, \mathbf{b}^*) = 0$, and the estimators are given by $\hat{\mathbf{a}}^* = \hat{\boldsymbol{\beta}}^*(t_0) = (\hat{\beta}_0^*(t_0), \ldots, \hat{\beta}_{p+1}^*(t_0))^{\mathrm{T}}$, and $\hat{\boldsymbol{\beta}}^{*\prime}(t_0) = \hat{\mathbf{b}}^*$. Let $\mathbf{I}_{p+2}$ be the identity matrix with size $p+2$, $\mathbf{0}_{p+2}$ be a size $p+2$ matrix with each entry equal to zero. The asymptotic covariance matrix of the estimators $\hat{\boldsymbol{\beta}}^*(t_0)$ can be estimated using the following sandwich formula:

$$
\widehat{\mathrm{cov}}\{\hat{\boldsymbol{\beta}}^*(t_0)\} = (\mathbf{I}_{p+2}, \mathbf{0}_{p+2})\widehat{\mathbf{\Gamma}}(t_0)^{-1}\widehat{\mathbf{\Lambda}}(t_0)\widehat{\mathbf{\Gamma}}(t_0)^{-1}(\mathbf{I}_{p+2}, \mathbf{0}_{p+2})^{\mathrm{T}}, \qquad (4.8)
$$

where

$$
\begin{aligned}
\widehat{\mathbf{\Gamma}}(t_0) &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\varpi_2\left[\sum_{r=0}^{p+1}\left\{\hat{a}_r^* + \hat{b}_r^*(t_{ij}-t_0)\right\}X_{ir}^*(t_{ij}), Q_i(t_{ij})\right]K_h(t_{ij}-t_0)\begin{pmatrix} \mathbf{X}_i^*(t_{ij}) \\ \mathbf{X}_i^*(t_{ij})(t_{ij}-t_0) \end{pmatrix}^{\otimes 2}, \\
\widehat{\mathbf{\Lambda}}(t_0) &= \frac{h}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\varpi_1^2\left[\sum_{r=0}^{p+1}\left\{\hat{a}_r^* + \hat{b}_r^*(t_{ij}-t_0)\right\}X_{ir}^*(t_{ij}), Q_i(t_{ij})\right]K_h^2(t_{ij}-t_0)\begin{pmatrix} \mathbf{X}_i^*(t_{ij}) \\ \mathbf{X}_i^*(t_{ij})(t_{ij}-t_0) \end{pmatrix}^{\otimes 2}
\end{aligned}
$$

with $\varpi_d(\mathcal{Z}, q) = (\partial^d/\partial \mathcal{Z}^d) l\{g^{-1}(\mathcal{Z}), q\}$, and $A^{\bigotimes 2}$ denotes AA^T for a matrix or vector A.

The estimator of a^*_{p+1} is equal to $\hat{\beta}^*_{p+1}(t_0)$, and if the equality of variances assumption is employed, then $\hat{\beta}^*_{p+1}(t_0)$ gives us $\hat{\tau}(t_0)$. If the equality of variances, $\sigma_1^2(t_0)$ and $\sigma_2^2(t_0)$, is not assumed, the estimator of $\beta^*_{p+1}(t_0)$ gives us the estimator of the time-varying correlation coefficient, $\hat{\tau}(t_0)$, up to the positive factor $\{\sigma_2(t_0)/\sigma_1(t_0)\}$ along with its sign. In addition, the pointwise asymptotic confidence intervals of $\beta^*_{p+1}(t_0)$ reveal the significance of the correlation coefficient.

4.3 Asymptotic Results

In this section we study the asymptotic properties of the estimators in both stages of the estimation procedure. The proof of the theorem presented in this section is provided in Sections 4.6. Define $\mu_k = \int t^k K(t)dt$ and $f(t)$ denotes the marginal density of T, the time covariate.

The first theorem shows the asymptotic properties of the backfitting estimator, $\hat{\boldsymbol{\beta}}_W$, obtained in the first stage of the estimation procedure. This theorem is given in Hu et al. (2004). In this theorem the number of observations for each subject, J, is assumed to be fixed. We assume that the kernel function, $K(\cdot)$ is a bounded symmetric density function with bounded support and satisfies the Lipschitz condition. Hu et al. (2004) also assumed that the second moment of the kernel function is one; however, we do not employ this assumption. The result below is based on using a local linear fit with working independence in the nonparametric estimation. Define $\mathbf{W}_i = (W_i(t_{i1}), \ldots, W_i(t_{iJ}))^{\mathrm{T}}$, $\mathbf{X}_i = (\mathbf{X}_i(t_{i1}), \ldots, \mathbf{X}_i(t_{i\,J}))^{\mathrm{T}}$, and $\mathbf{T}_i = (t_{i1}, \ldots, t_{iJ})^{\mathrm{T}}$. Let $\mathbf{V} = \mathrm{diag}(V_1, \ldots, V_n)$ denote the working covariance matrix for this stage, and $\widetilde{\mathbf{X}}_i = \dfrac{\partial \hat{\alpha}_W(\mathbf{T}_i, \boldsymbol{\beta}_W)}{\partial \boldsymbol{\beta}_W}$ with $\hat{\alpha}_W(\mathbf{T}_i, \boldsymbol{\beta}_W) = E(\mathbf{W}_i|\mathbf{T}_i) - E(\mathbf{X}_i|\mathbf{T}_i)\boldsymbol{\beta}_W$ as the estimator of $\boldsymbol{\alpha}_W(\mathbf{T}_i) = (\alpha(t_{i1}), \ldots, \alpha(t_{iJ}))^{\mathrm{T}}$. Assume that $\mathbf{V}$ is invertible and $h \propto n^{d-}$, where $1/5 \leq d \leq 1/3$ (Lin and Carroll, 2001).

Theorem 4.1. *Under the regularity conditions given in Section 4.6, and that $n_i \leq \infty$, the following result for the backfitting estimator $\hat{\boldsymbol{\beta}}_W$ holds:*

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_W - \boldsymbol{\beta}_W - bias_{BF}\right) \to \mathcal{N}\left(0, cov_{BF}\right),$$

where

$$\begin{aligned} bias_{BF} &= \frac{h^2}{2}\mu_2\, E(\widetilde{\mathbf{X}}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\widetilde{\mathbf{X}}_i)^{-1} E\left\{\mathbf{X}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\alpha}''_W(\mathbf{T}_i)\right\}, \\ cov_{BF} &= E(\widetilde{\mathbf{X}}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\widetilde{\mathbf{X}}_i)^{-1} E\left\{(\mathbf{Z}_1 - \mathbf{Z}_2)\boldsymbol{\Sigma}_{1i}(\mathbf{Z}_1 - \mathbf{Z}_2)\right\} E(\widetilde{\mathbf{X}}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\widetilde{\mathbf{X}})^{-1}, \end{aligned}$$

with $\mathbf{Z}_1 = \mathbf{V}_i^{-1}\mathbf{X}_i$, *$j^{\text{th}}$ row of $\mathbf{Z}_2$ is* $\mathbf{Z}_2^j = \dfrac{\left[\sum_{k=1}^J \sum_{l=1}^J E\{\mathbf{X}_i(t_{ik})V_i^{kl}|t_{il} = t_{ij}\}\right]}{f(t_{ij})}$, $\boldsymbol{\Sigma}_{1i} = cov(\mathbf{W}_i|\mathbf{X}_i, \mathbf{T}_i)$, *$V_i^{kl}$ is the (k,l) entry of $\mathbf{V}_i^{-1}$, and $\widetilde{\mathbf{X}}_i$ becomes $\{\mathbf{X}_i - E(\mathbf{X}_i|\mathbf{T}_i)\}$.*

As in the nonparametric approach described in Chapter 3, we need to show how the first and the second stages of the estimation procedure are linked. In theory the errors from the first stage are used to fit the model in the second stage. However, we do not observe these errors, so we use the residuals in place of the errors. The result given in Lemma 3.1 shows that the residuals obtained in the first stage can be used in the second stage of the estimation procedure. This result also holds here.

Next, we need to describe the asymptotic properties of the estimators obtained in the second stage of the estimation procedure. At this stage, we have a generalized varying coefficient model for a longitudinal setting. Define $\mathbf{s} = (t_1, \ldots, t_J)$, $J \geq 1$, $\boldsymbol{\vartheta}^*(\mathbf{s}) = (\boldsymbol{\vartheta}^*(t_1), \ldots, \boldsymbol{\vartheta}^*(t_J))^{\mathrm{T}}$ with $\boldsymbol{\vartheta}^*(t_0) = \left(a_0^*, \ldots, a_p^*, a_{p+1}^*, b_0^*, \ldots, b_p^*, b_{p+1}^*\right)^{\mathrm{T}}$, and $\hat{\boldsymbol{\vartheta}}^*(\mathbf{s}) = \left(\hat{\boldsymbol{\vartheta}}^*(t_1), \ldots, \hat{\boldsymbol{\vartheta}}^*(t_J)\right)^{\mathrm{T}}$ with $\hat{\boldsymbol{\vartheta}}^*(t_0) = (\hat{\mathbf{a}}^{*\mathrm{T}}, \hat{\mathbf{b}}^{*\mathrm{T}})^{\mathrm{T}}$. In Theorem 3.2, we establish the asymptotic normality of the regression coefficients in a generalized varying coefficient model for a longitudinal setting. Hence, the asymptotic normality of $\hat{\boldsymbol{\vartheta}}^*(\mathbf{s})$ follows from this theorem.

4.4 Simulation Studies

In this section we demonstrate the performance of the proposed procedure via a Monte Carlo simulation study. In this simulation study, we use the Epanechnikov kernel, $K(t) = 0.75(1 - t^2)_+$, on an equidistant set of grid points $\{t_k, k = 1, \ldots, n_{grid}\}$ between 0 and 1 with $n_{grid} = 200$. We generate 500 intensive longitudinal data sets, in which for each unit the number of measurements is randomly selected using a discrete uniform distribution on $[20, 40]$ and the measurement times $T_i = (t_{i1}, \ldots, t_{in_i})$ are generated from a uniform distribution on $[0, 1]$ with $n_i \in [20, 40]$. We use sample size $n = 100$. We generate the continuous and latent variables from the following models:

$$\begin{aligned} W_i(t_{ij}) &= \alpha_W(t_{ij}) + X_i^{\mathrm{T}}(t_{ij})\beta_W + \varepsilon_{1i}(t_{ij}), \quad (4.9) \\ Y_i(t_{ij}) &= \alpha_Y(t_{ij}) + X_i^{\mathrm{T}}(t_{ij})\beta_Y + \varepsilon_{2i}(t_{ij}), \end{aligned}$$

where $\alpha_W(t_{ij}) = \cos(2\pi t_{ij}) + 1$, $\beta_W = 1.1$, $\alpha_Y(t_{ij}) = \sin(2\pi t_{ij})$, $\beta_W = 0.3$, $i = 1, \ldots, 100$ and $j \in [20, 40]$. The predictor variable $X_i(t_{ij})$ is generated from the standard Gaussian distribution. The error variable for the continuous response $\varepsilon_{1i}(t_{ij})$ follows a Gaussian distribution with mean zero, variance $0.4 + 0.4\sin^2(2\pi t_{ij})$, and $\mathrm{corr}\{\varepsilon_{1i}(t_{ij}), \varepsilon_{1i}(t_{ij'})\} = \rho_1(t_{ij}, t_{ij'}) = 0.3^{|t_{ij} - t_{ij'}|}$ for $j \neq j'$. Moreover, the error variable for the latent variable $\varepsilon_{2i}(t_{ij})$ follows a Gaussian distribution with mean zero, variance $0.4{+}0.4\sin^2(2\pi t_{ij})$, and $\mathrm{corr}\{\varepsilon_{2i}(t_{ij}), \varepsilon_{2i}(t_{ij'})\} = \rho_2(t_{ij}, t_{ij'}) = 0.4^{|t_{ij} - t_{ij'}|}$ for $j \neq j'$. Thus, $\boldsymbol{\varepsilon}_i(t_{ij}) = (\varepsilon_{1i}(t_{ij}), \varepsilon_{2i}(t_{ij}))^{\mathrm{T}}$ follows a bivariate Gaussian distribution with $\mathrm{corr}\{\varepsilon_{1i}(t_{ij}), \varepsilon_{2i}(t_{ij})\} = \tau(t_{ij}) = 0.2\cos(0.5\pi t_{ij})$, and corr $\{\varepsilon_{1i}(t_{ij}), \varepsilon_{2i}(t_{ij'})\} = \rho_{12}(t_{ij}, t_{ij'}) = 0.2\sqrt{\cos(0.5\pi t_{ij})\cos(0.5\pi t_{ij'})}$ for $j \neq j'$. In Section 4.1 the binary variable was defined as $Q_i(t_{ij}) = 1$ if $Y_i(t_{ij}) > 0$, and $Q_i(t_{ij}) = 0$ if $Y_i(t_{ij}) \leq 0$. However, in this simulation study we would like to show that decreasing the percentage of successes in the binary response does not decrease the efficacy of our procedure. Hence, the relation between the latent variable and the binary variable is defined as $Q_i(t_{ij}) = 1$ if $Y_i(t_{ij}) > 0.25$, and $Q_i(t_{ij}) = 0$ if $Y_i(t_{ij}) \leq 0.25$. Therefore, each of our 500 simulated data sets has approximately 40% failure.

In the first stage we fit the partially linear model to the marginal model of the continuous response (4.9) using the backfitting algorithm. In this stage for the parametric estimation, the working covariance matrix is set to be the true within-subject covariance of $\mathbf{W}_i$. To assess the performance of $\hat{\beta}_W$, we use mean square error (MSE),

$$MSE = \frac{1}{500}\sum_{l=1}^{500} \parallel \hat{\beta}_{W_l} - \beta_W \parallel^2,$$

where $\hat{\beta}_{W_l}$ is the value of $\hat{\beta}_W$ obtained at the l^{th} simulation run. To evaluate the performance of $\hat{\alpha}_W(\cdot)$, we use the root average squared error (RASE),

$$RASE = \left[\frac{1}{200}\sum_{k=1}^{200}\{\alpha_W(t_k) - \hat{\alpha}_W(t_k)\}^2\right]^{1/2}.$$

In our simulation, we generate several pilot simulation data sets, and use a cross-validation bandwidth selector to get an overall picture about the optimal bandwidth. To save computing time, we fix the bandwidth to be close to the optimal ones from the pilot simulation data sets. Specifically, we set the bandwidth to be h = 0.10. It is of interest to examine the performance of the proposed procedure with a wide range of the bandwidth. Thus, we also set the bandwidth to be h/2=0.05 and 2h=0.20, corresponding to undersmoothing and oversmoothing simulations. The MSE values for $\hat{\beta}_W$, the sample average and the sample standard deviations of the RASE values for $\hat{\alpha}_W(\cdot)$, computed at three different bandwidths, 0.05, 0.10, and 0.20, are listed in Table 4.1. These results are based on 500 replications. According to the MSE and mean RASE values in Table 4.1, the bandwidth $h = 0.10$ gives the minimum value.

Table 4.1: Summary of simulation results for the first stage

h	MSE	RASE	t	$\hat{\alpha}_W(t)$	
	$\hat{\beta}_W$	Mean(SD)		SD	SE (SD_{se})
			0.30	0.064	0.055 (0.005)
0.05	8.964×10^{-7}	0.072 (0.020)	0.50	0.045	0.040 (0.004)
			0.70	0.063	0.054 (0.005)
			0.30	0.042	0.038 (0.003)
0.10	8.958×10^{-7}	0.056 (0.017)	0.50	0.032	0.029 (0.002)
			0.70	0.042	0.038 (0.003)
			0.30	0.029	0.026 (0.002)
0.20	8.981×10^{-7}	0.091 (0.013)	0.50	0.024	0.022 (0.002)
			0.70	0.028	0.026 (0.002)

(a) Estimated baseline function $\alpha_W(T)$

(b) Estimated correlation coefficient function $\tau(T)$

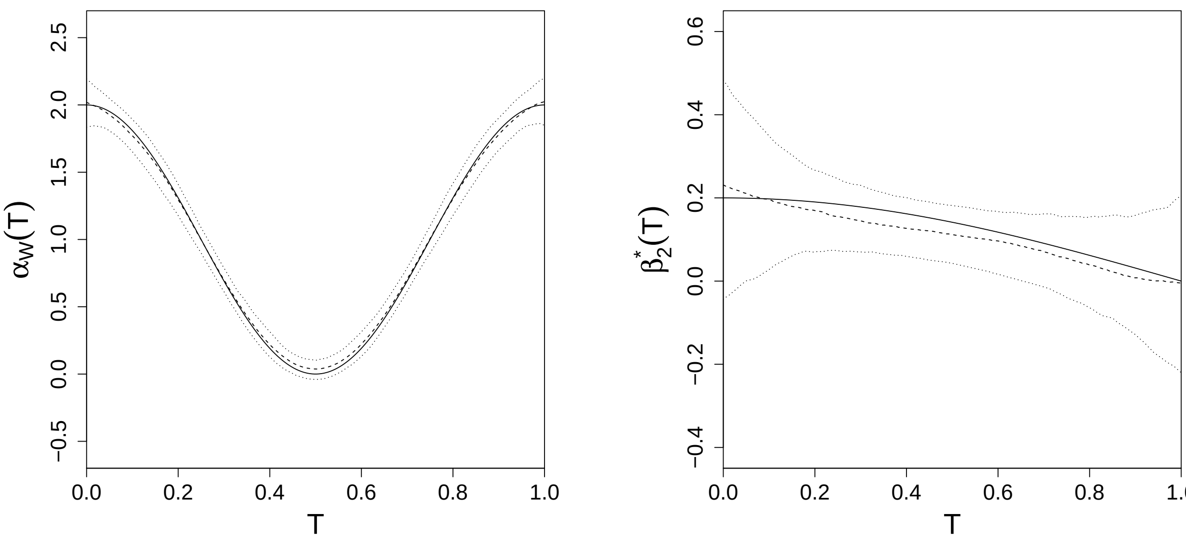

Figure 4.1: (a) Estimated baseline function (dashed) of the partially linear model fit to the continuos response overlaying the true coefficient functions (solid) along with the 2.5 and 97.5 percentiles based on 500 Monte Carlo runs (dotted). (b) Estimated correlation coefficient function (dashed) overlaying the true correlation function (solid) along with the 2.5 and 97.5 percentiles based on 500 Monte Carlo runs (dotted).

The typical estimate of the parameter function $\alpha_W(\cdot)$ along with the 2.5 and 97.5 percentiles based on 500 Monte Carlo runs at $h = 0.1$ are depicted in Figure 4.1(a). According to Figure 4.1(a), the typical estimated coefficient function is very close to the true function.

The accuracy of the proposed standard error formula (4.6) is also evaluated. The standard deviation of 500 $\hat{\alpha}_W(\cdot)$, based on 500 simulations is regarded as the true standard error and is denoted by SD in Table 4.1. In Table 4.1 the sample average and the sample standard deviation of the 500 estimated standard errors are denoted by SE and SD_{se}, respectively. They summarize the overall performance of the standard error formula (4.6). Table 4.1 shows the results at points $t = 0.30, 0.50$ and 0.70. According to Table 4.1, the SE is slightly less than the true standard error, but their difference is less than two times the SD_{se}.

In the second stage we fit a generalized time-varying coefficient model to the conditional model of the binary response given the continuous response.

$$P\left\{Q_i(t_{ij}) = 1|W_i(t_{ij})\right\} = \Phi\left\{\beta_0^*(t_{ij}) + X_i^{\mathrm{T}}(t_{ij})\beta_1^*(t_{ij}) + \beta_2^*(t_{ij})e_i(t_{ij})\right\},$$

where $e_i(t_{ij}) = W_i(t_{ij}) - \{\hat{\alpha}_W(t_{ij}) + X_i^{\mathrm{T}}(t_{ij})\hat{\beta}_W\}$ is the residual from the first stage, $\beta_2^*(t_{ij}) = \{\sigma_2(t_{ij})/\sigma_1(t_{ij})\}\,\tau(t_{ij})$. Since the equality of $\sigma_1^2(t_{ij})$ and $\sigma_2^2(t_{ij})$ is assumed, the estimate of $\beta_2^*(t_{ij})$ gives us the estimate of the correlation coefficient, $\hat{\tau}(t_{ij})$. Hence, our main goal in this stage is to assess the performance of $\hat{\beta}_2^*(t_{ij})$, which shows how the association between the continuous response and the binary response changes over time. Similar to the first stage, the performance of the estimator in the second stage is evaluated using RASE,

$$\mathrm{RASE} = \left[\frac{1}{200}\sum_{k=1}^{200}\{\beta_2^*(t_k) - \hat{\beta}_2^*(t_k)\}^2\right]^{1/2}.$$

The sample means and the sample standard deviations of the RASE values based on 500 replications, computed at three bandwidths, $0.40, 0.50$ and 0.60 are given in Table 4.2. According to the mean RASE values in Table 4.2, the bandwidth for the second stage is to be $h = 0.50$.

The typical estimate of the correlation coefficient function along with the 2.5 and 97.5 percentiles based on 500 Monte Carlo runs at $h = 0.50$ is depicted in Figure 4.1(b). From Figure 4.1(b) we can see that the typical estimated correlation coefficient function is close to the underlying true correlation coefficient function.

Table 4.2: Summary of simulation results for the second stage

h	RASE	t	$\hat{\beta}_2^*(t)$	
	Mean(SD)		SD	SE (SD_{se})
0.40	0.070 (0.029)	0.30	0.041	0.038 (0.002)
		0.50	0.037	0.036 (0.002)
		0.70	0.043	0.039 (0.003)
0.50	0.058 (0.024)	0.30	0.040	0.037 (0.002)
		0.50	0.034	0.033 (0.002)
		0.70	0.041	0.037 (0.003)
0.60	0.060 (0.027)	0.30	0.039	0.036 (0.002)
		0.50	0.031	0.031 (0.002)
		0.70	0.040	0.036 (0.003)

Next, we assess the performance of the standard error formula (4.8) at points $t = 0.30, 0.50$ and 0.70 via SE and SD_{se} given in Table 4.2. Table 4.2 demonstrates that our standard error formula slightly underestimates the true standard error, but the difference is less than two times the SD_{se}.

4.5 Application to the Women's Interagency HIV Study

We now illustrate the proposed joint modeling methodology via an analysis of a subset from the Women's Interagency HIV Study (WIHS) Cohort I. This subset contains 205 women from Chicago, San Francisco, Los Angeles, Washington, DC, and New York City (Bronx and Brooklyn), who were recruited in the study between 1994-1995. Among these participants 50 (24%) were HIV negative, 155 (76%) were infected with HIV. In our subset women were between ages 25 and 55. Participants were scheduled to have their socio-demographic, medical, obstetric, gynecological history recorded at six-month intervals along with their sexual behaviors, and alcohol, tobacco and other drug use. In addition, plasma, serum, lymphocyte and urine samples were also taken. During this study, since many participants missed some of their scheduled visits, there were unequal numbers of repeated measurements and different measurement times for each individual.

It is known that cigarette smoking has effects on the immune system (Ferson et al., 1979; Galai et al., 1997; Halonen et al., 1982; Hughes et al., 1985), however it is not yet clear whether any of these effects would influence the progression of HIV. Burns et al. (1996) reported that in a cohort of 131 homosexual men, there is no difference between smokers and nonsmokers in the risk of developing AIDS. Similarly, Galai et al. (1997) used Kaplan-Meier analysis and multivariate Cox regression models to investigate the effect of cigarette smoking on development of AIDS in Multicenter AIDS Cohort Study of homosexual men. Their analysis revealed that smoking was not significantly associated with progression to AIDS. However, Nieman et al. (1993) found that in a case series of 84 individuals, smokers progressed to AIDS more rapidly than nonsmokers.

Given these controversial arguments, our main interest is to examine how the association between HIV progression and smoking changes over time among the women enrolled in WIHS. It is known that HIV reduces the number of CD4 cells, hence, CD4 cell count is used to assess the progression of the disease. Therefore, our main goal becomes observing the association between CD4 cell count, a continuous response, and smoking status, a binary response, over time. Our method differs from a regression fit of CD4 cell count on smoking status by allowing us to estimate the partial correlation between CD4 and smoking given a set of predictors. In addition, our procedure also allows us to investigate the relationship between CD4 cell count and a set of predictors. Based on some preliminary analysis and findings in the literature (Zeger and Diggle, 1994; Obirikorang and Yeboah, 2009), the following predictors are used in our analysis: baseline CD4 cell count (measured at the first visit), number of sexual partners, hematocrit value (the volume percentage of red blood cells in blood), mean corpuscular volume (measure of the average red blood cell size), platelet count, and the Center for Epidemiologic Studies Depression (CESD) scale score. All of these predictors are continuous variables. After accounting for these predictor variables in the first stage of the estimation procedure, in the second stage, we observe how the relationship between the continuous variable, CD4 cell count, and the binary variable, smoking status, varies over time.

In our analysis Epanechnikov kernel function is used in both stages. We use a multifold cross validation and choose $h = 12.5$ and $h = 13.0$ as the bandwidths for the first and second stages, respectively. In the first stage of our estimation procedure, we fit the following partially linear model to CD4 cell count:

$$
\begin{aligned}
W_i(t_{ij}) = \quad & \alpha_W(t_{ij}) + \beta_{1W} X_{i1}(t_{i1}) + \beta_{2W} X_{i2}(t_{ij}) + \beta_{3W} X_{i3}(t_{ij}) + \beta_{4W} X_{i4}(t_{ij}) \\
+ \quad & \beta_{5W} X_{i5}(t_{ij}) + \beta_{6W} X_{i6}(t_{ij}) + \varepsilon_{1i}(t_{ij}),
\end{aligned}
$$

where

$W_i(t_{ij})$: the CD4 cell count of the i^{th}subject at time t_{ij},

$X_{i1}(t_{i1})$: the baseline CD4 cell count of the i^{th}subject at the first visit (t_{i1}),

$X_{i2}(t_{ij})$: the number of sexual partners (PART) of the i^{th}subject at time t_{ij},

$X_{i3}(t_{ij})$: the hematocrit (HCV) value of the i^{th}subject at time t_{ij},

$X_{i4}(t_{ij})$: the mean corpuscular volume (MCV) of the i^{th}subject at time t_{ij},

$X_{i5}(t_{ij})$: the platelet count (PLAT) of the i^{th}subject at time t_{ij},

$X_{i6}(t_{ij})$: the CESD scale score of the i^{th}subject at time t_{ij},

and t_{ij} stands for the age of the i^{th} subject at the j^{th} visit. For our analysis, all the predictors are centered. In this stage, we need to incorporate the correlation structure for the parametric estimation. We found that the heterogenous exponential covariance matrix fitted the data reasonably well. The estimated time-varying intercept function is depicted in Figure 4.2 (a). We can see that the baseline function seems to increase until approximately age 42, and then starts to decrease. The confidence intervals indicate that between ages 25 and 55, the intercept function appears to be significant. Table 4.3 demonstrates the estimates and the standard error values for the regression coefficients of the predictors. From Table 4.3 we can see that all of the predictors, except number of sexual partners, are significantly associated with CD4 cell count for this data set. The response, CD4 cell count, has a positive relationship with the baseline CD4 cell count and the platelet count. However, the number of sexual partners, the volume percentage of red blood cells in blood (HCT), average red blood cell size (MCV) and the CESD scale are associated with decreased CD4 cell counts.

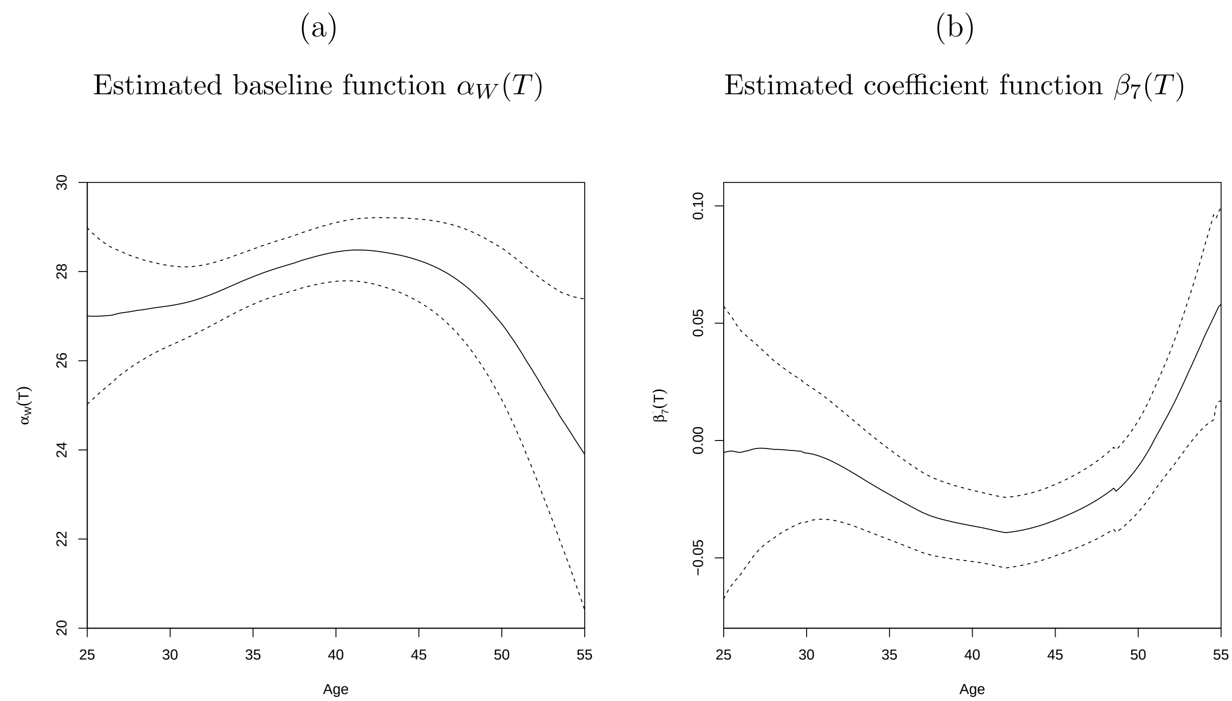

Figure 4.2: (a) Estimated time-varying intercept function (solid) of the partially linear model fit to CD4 cell count along with the 95% pointwise asymptotic confidence intervals (dashed). (b) Estimated coefficient function (solid) of the residuals from the first stage in the generalized time-varying coefficient model fit along with the 95% pointwise asymptotic confidence intervals (dashed).

Next, in the second stage of our estimation procedure, we investigate how the association between CD4 cell count and smoking status changes over time. We compute the residuals from the marginal model (4.10), and fit the following generalized time-varying coefficient model:

$$\begin{aligned} P\{Q_i(t_{ij}) = 1|W_i(t_{ij})\} = \; & \Phi \;\; \big\{\beta_0^*(t_{ij}) + \beta_1^*(t_{ij})X_{i1}(t_{i1}) + \beta_2^*(t_{ij})X_{i2}(t_{ij}) + \beta_3^*(t_{ij})X_{i3}(t_{ij}) \\ & + \;\; \beta_4^*(t_{ij})X_{i4}(t_{ij}) + \beta_5^*(t_{ij})X_{i5}(t_{ij}) + \beta_6^*(t_{ij})X_{i6}(t_{ij}) + \beta_7^*(t_{ij})e_i(t_{ij})\big\}, \end{aligned}$$

where $Q_i(t_{ij})$ is the smoking status of the i^{th}subject at time t_{ij}, and $X_{i1}(t_{i1})$, $X_{i2}(t_{ij})$, $X_{i3}(t_{ij})$, $X_{i4}(t_{ij}), X_{i5}(t_{ij})$, and $X_{i6}(t_{ij})$ are defined in (4.10). As men-

Table 4.3: Summary of results for the first stage

Variable	$\hat{\beta}$	$SE(\hat{\beta})$
Baseline CD4	0.857	0.026
PART	-0.043	0.054
HCT	-0.180	0.051
MCV	-0.189	0.019
PLAT	0.007	0.004
CESD scale	-0.029	0.019

tioned in Section 4.1, $\beta_7^*(t_{ij}) = \{\sigma_2(t_{ij})/\sigma_1(t_{ij})\}\,\tau(t_{ij})$ with $\sigma_1^2(t_{ij})$ and $\sigma_2^2(t_{ij})$ as the variances of CD4 cell count and the latent variable underlying the smoking status at time t_{ij}, respectively, and $\tau(t_{ij})$ as the correlation between CD4 cell count and smoking status at time t_{ij}. Figure 4.2 (b) depicts the estimated coefficient function, $\hat{\beta}_7^*(t_{ij})$. From Figure 4.2 (b), the correlation between smoking status and CD4 cell count is time-varying. It can be seen that approximately before the age of 51, the correlation between CD4 cell count and smoking status is negative. However, between ages 51 and 55, the relationship is positive, i.e, increased CD4 cell count is associated with smoking. We can also test the significance of the association by using the confidence intervals. Figure 4.2 (b) demonstrates that the association is significant between ages 34 and 49. We can also see that the relationship is insignificant between ages 25 and 34, and 49 and 55. However, this result need to be interpreted cautiously, since it might be due to lack of enough observations.

To summarize, our methodology allows researchers to examine the time-varying association between CD4 cell count and a binary variable of interest, such as smoking status. Investigating these relationships may lead to better understanding of the factors affecting the progress of the HIV disease. Our procedure can also be applied to other intensive longitudinal data sets, in which the question of interest is to observe the time-varying correlation between binary and continuous responses.

4.6 Technical Conditions and Proofs

The following regularity conditions are needed to facilitate proof of the theorem presented in this chapter.

Regularity Conditions:

A. The observed sample $\{t_{ij}, X_i(t_{ij}), W_i(t_{ij})\}$, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$ is an independent and identically distributed realization of (T, X, W). The $\varepsilon_{1i}(t_{ij})$ are iid from a distribution with mean zero and finite variance $\sigma_1^2(t_{ij})$. The covariate T has finite support $\mathcal{T} = [\mathcal{L}, \mathcal{U}]$. The support for X is a closed and bounded interval in $\mathbb{R}^p$, denoted by $\mathcal{X}$.

B. The varying coefficient function $\alpha_W(t_{ij})$, have continuous second order derivatives over $\mathcal{T}$.

C. Without loss of generality the kernel density function $K(\cdot)$ has bounded support and satisfies

$$\int K(t)dt = 1, \int |t|^3 K(t)dt < \infty, \int t^2 K^2(t)dt < \infty.$$

D. The marginal density function $f(t_{ij})$ of T is continuous and positive for all $t_{ij} \in \mathcal{T}$.

By Condition (B), we assume that the parameter space for $\boldsymbol{\alpha}_W(\cdot)$, namely, Ω_1, is a closed and bounded subset of $\mathbb{R}^{2d}$. Condition (C) is the usual regularity condition on the kernel function.

Proof of Theorem 4.1

Proof. Hu et al. (2004) gave the sketch of this proof. Here we present a more detailed proof.

At convergence, $\boldsymbol{\beta}_W = \left\{\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{S})\mathbf{X}\right\}^{-1}\left\{\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{S})\mathbf{W}\right\}$, where $\mathbf{S}$ is the smoother matrix, $\mathbf{W} = (\mathbf{W_1}, \ldots, \mathbf{W_n})^{\mathrm{T}}$, and $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^{\mathrm{T}}$. For local linear regression, $\mathbf{S} = (\mathbf{S}_1, \ldots, \mathbf{S}_n)^{\mathrm{T}}$, with $\mathbf{S}_i = (\mathbf{S}_{t_{i1}}, \ldots, \mathbf{S}_{t_{iJ}})^{\mathrm{T}}$, $\mathbf{S}_t^{\mathrm{T}} = \mathbf{d}_1^{\mathrm{T}}\left(\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}_t\boldsymbol{\Gamma}_t\right)^{-1}\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}_t$, $\mathbf{d}_1^{\mathrm{T}} = (1, 0)$, $t \in \mathbf{T}_i$, and $\boldsymbol{\Gamma}_t$ and $\boldsymbol{\kappa}_t$ defined in (4.5) (Opsomer and Ruppert, 1997). Let $\mathbf{I}$ denote the $(N \times N)$ identity matrix with $N = \sum_{i=1}^{n} J$, and $\mathbf{T} = (\mathbf{T_1}, \ldots, \mathbf{T_n})^{\mathrm{T}}$.

$$\begin{aligned}\hat{\boldsymbol{\beta}}_W - \boldsymbol{\beta}_W &= \left\{\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{S})\mathbf{X}\right\}^{-1}\left\{\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{S})\mathbf{W}\right\} - \boldsymbol{\beta}_W \qquad (4.10)\\ &= \left\{n^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{S})\mathbf{X}\right\}^{-1}\left[n^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{S})\left\{\boldsymbol{\alpha}_W(\mathbf{T}) + \boldsymbol{\varepsilon}_1(\mathbf{T})\right\}\right],\end{aligned}$$

where $\boldsymbol{\alpha}_W(\mathbf{T}) = (\boldsymbol{\alpha}_W(\mathbf{T}_1), \ldots, \boldsymbol{\alpha}_W(\mathbf{T}_n))^{\mathrm{T}}$ with $\boldsymbol{\alpha}_W(\mathbf{T}_i) = (\alpha_W(t_{i1}), \ldots, \alpha_W(t_{iJ}))^{\mathrm{T}}$, and $\boldsymbol{\varepsilon}_1(\mathbf{T}) = (\boldsymbol{\varepsilon}_{11}(\mathbf{T}_1), \ldots, \boldsymbol{\varepsilon}_{1n}(\mathbf{T}_n))^{\mathrm{T}}$ with $\boldsymbol{\varepsilon}_{1i}(\mathbf{T}_i) = \left(\varepsilon_{1i}(t_{i1}), \ldots, \varepsilon_{1i}(t_{iJ})\right)^{\mathrm{T}}$. Let us start with the first term in (4.10). Using the proof of Theorem 2.1 in Ruppert and Wand (1994), it can be seen that

$$n^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{S})\mathbf{X} = E\left[\mathbf{X}_i\mathbf{V}_i^{-1}\left\{\mathbf{X}_i - E(\mathbf{X}_i|\mathbf{T}_i)\right\}\right] + o_p(1), \qquad (4.11)$$

where $E\left[\mathbf{X}_i\mathbf{V}_i^{-1}\left\{\mathbf{X}_i - E(\mathbf{X}_i|\mathbf{T}_i)\right\}\right] = E\Big[\left\{\mathbf{X}_i - E(\mathbf{X}_i|\mathbf{T}_i)\right\}^{\mathrm{T}}\mathbf{V}_i^{-1}\left\{\mathbf{X}_i - E(\mathbf{X}_i|\mathbf{T}_i)\right\}\Big] = E\big(\widetilde{\mathbf{X}}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\widetilde{\mathbf{X}}_i\big)$. The second term in (4.10) can be written as

$$n^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{S})\left\{\boldsymbol{\alpha}_W(\mathbf{T}) + \boldsymbol{\varepsilon}_1(\mathbf{T})\right\} = n^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{S})\boldsymbol{\alpha}_W(\mathbf{T}) + n^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon}_1(\mathbf{T}). \qquad (4.12)$$

We start with the first component in (4.12). Define $\mathbf{C}_\alpha(t) = (\mathbf{C}_{1\alpha}(t), \ldots, \mathbf{C}_{n\alpha}(t))^{\mathrm{T}}$, $\mathbf{C}_{i\alpha}(t) = ((t_{i1}-t)^2, \ldots, (t_{iJ}-t)^2)^{\mathrm{T}}\alpha_W''(t)$, and $\mathbf{C}_1 = \left(\mathbf{S}_{t_{11}}^{\mathrm{T}}\mathbf{C}_\alpha(t_{11}), \ldots, \mathbf{S}_{t_{nJ}}^{\mathrm{T}}\mathbf{C}_\alpha(t_{nJ})\right)^{\mathrm{T}}$.

Applying the Taylor theorem as in the proof of Theorem 2.1 in Ruppert and Wand (1994), we obtain the following:

$$\mathbf{S}\boldsymbol{\alpha}_W(\mathbf{T}) \;\; = \;\; \boldsymbol{\alpha}_W(\mathbf{T}) + \frac{1}{2}\mathbf{C}_1 + o_P(h^2).$$

Let $t \in \mathbf{T}$, we compute each element of $\mathbf{C}_1$ as follows:

$$\mathbf{S}_t^{\mathrm{T}}\mathbf{C}_\alpha(t) = \mathbf{d}_1^{\mathrm{T}}\left(\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}_t\boldsymbol{\Gamma}_t\right)^{-1}\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}_t\,\mathbf{C}_\alpha(t),$$

where

$$\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}_t\boldsymbol{\Gamma}_t \;\; = \;\; \begin{pmatrix} \sum_{i=1}^n\sum_{j=1}^J K_h(t_{ij}-t) & \sum_{i=1}^n\sum_{j=1}^J K_h(t_{ij}-t)(t_{ij}-t) \\ \sum_{i=1}^n\sum_{j=1}^J K_h(t_{ij}-t)(t_{ij}-t) & \sum_{i=1}^n\sum_{j=1}^J K_h(t_{ij}-t)(t_{ij}-t)^2 \end{pmatrix}.$$

Let

$$\mathcal{B}_k = \frac{1}{N}\sum_{i=1}^n\sum_{j=1}^J K_h(t_{ij}-t)(t_{ij}-t),$$

where $k = 0, 1, 2$. Using the result $X = \mathrm{E}(X) + O_p\{\sqrt{\mathrm{Var}(X)}\}$, and change of variables,

$$\begin{aligned} \mathcal{B}_k \;\; &= \;\; \int u^k h^k K_h(u) f(hu+t)du + O_p\{\sqrt{log(1/h)/nh}\} \\ &= \;\; h^k\mu_k\, f(t) + O_p\{h^k + \sqrt{log(1/h)/nh}\}. \end{aligned}$$

By the assumption that the kernel function is symmetric and the equation above, we can see that

$$N^{-1}\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}_t\boldsymbol{\Gamma}_t = \begin{pmatrix} O_p\{1 + \sqrt{log(1/h)/nh}\} & O_p\{h + \sqrt{log(1/h)/nh}\} \\ O_p\{h + \sqrt{log(1/h)/nh}\} & h^2\mu_2 f(t) O_p\{h^2 + \sqrt{log(1/h)/nh}\} \end{pmatrix},$$

and

$$\left(N^{-1}\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}_t\boldsymbol{\Gamma}_t\right)^{-1} = \begin{pmatrix} \{f(t)\}^{-1}O_p\{1 + \sqrt{log(1/h)/nh}\} & O_p\{h + \sqrt{log(1/h)/nh}\} \\ O_p\{h + \sqrt{log(1/h)/nh}\} & \{h^2\mu_2 f(t)\}^{-1}O_p\{h^2 + \sqrt{log(1/h)/nh}\} \end{pmatrix}. \tag{4.13}$$

Similarly,

$$N^{-1}\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}_t\mathbf{C}_\alpha(t) = \begin{pmatrix} N^{-1}\sum_{i=1}^n\sum_{j=1}^J K_h(t_{ij}-t)(t_{ij}-t)^2\alpha''_W(t) \\ N^{-1}\sum_{i=1}^n\sum_{j=1}^J K_h(t_{ij}-t)(t_{ij}-t)^3\alpha''_W(t) \end{pmatrix}.$$

Hence,

$$\begin{aligned}\mathbf{S}_t^{\mathrm{T}}\mathbf{C}_\alpha(t) &= N^{-1}\sum_{i=1}^n\sum_{j=1}^J K_h(t_{ij}-t)(t_{ij}-t)^2\alpha''_W(t)\{f(t)\}^{-1} \\ &= N\int N^{-1}K_h(u)h^2u^2\alpha''_W(t)\{f(t)\}^{-1}f(hu+t)du + O_p\{h+\sqrt{log(1/h)/nh}\} \\ &= h^2\mu_2\alpha''_W(t).\end{aligned}$$

The first component in (4.12) becomes,

$$n^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{S})\boldsymbol{\alpha}_W(\mathbf{T}) = -\frac{h^2}{2}\mu_2 E\left\{\mathbf{X}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\alpha}''_W(\mathbf{T}_i)\right\} + o_P(h^2), \qquad (4.14)$$

where $\boldsymbol{\alpha}''_W(\mathbf{T}_i) = (\alpha''_W(t_{i1}),\ldots,\alpha''_W(t_{iJ}))^{\mathrm{T}}$. The equation (4.14) along with (4.11) establishes the bias of the estimator $\hat{\boldsymbol{\beta}}_W$. The second term in (4.12) determines the asymptotic distribution of the estimator (Hu et al., 2004), which can be written as follows:

$$n^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon}_1(\mathbf{T}) = n^{-1}\sum_{i=1}^n\mathbf{X}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\varepsilon}_{1i}(\mathbf{T}_i) - n^{-1}\sum_{i=1}^n\mathbf{X}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathcal{S}_i, \quad (4.15)$$

where $\mathcal{S}_i = \mathbf{S}_i\boldsymbol{\varepsilon}(\mathbf{T})$ with $\mathbf{S}_i = (\mathbf{S}_{t_{i1}},\ldots,\mathbf{S}_{t_{iJ}})^{\mathrm{T}}$. Let $t\in\mathbf{T}_i$, we calculate each element of $\mathcal{S}_i$ as follows:

$$\mathbf{S}_t^{\mathrm{T}}\boldsymbol{\varepsilon}_1(\mathbf{T}) = \mathbf{d}_1^{\mathrm{T}}\left(\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}_t\boldsymbol{\Gamma}_t\right)^{-1}\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}_t\boldsymbol{\varepsilon}_1(\mathbf{T}),$$

where $\left(N^{-1}\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}_t\boldsymbol{\Gamma}_t\right)^{-1}$ is given in (4.13), and

$$N^{-1}\boldsymbol{\Gamma}_t^{\mathrm{T}}\boldsymbol{\kappa}_t\boldsymbol{\varepsilon}_1(\mathbf{T}) = \begin{pmatrix} N^{-1}\sum_{i=1}^n\sum_{j=1}^J K_h(t_{ij}-t)\varepsilon_{1i}(t_{ij}) \\ N^{-1}\sum_{i=1}^n\sum_{j=1}^J K_h(t_{ij}-t)(t_{ij}-t)\varepsilon_{1i}(t_{ij}) \end{pmatrix}.$$

Therefore,

$$\mathbf{S}_t\boldsymbol{\varepsilon}_1(\mathbf{T}) = N^{-1}\{f(t)\}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J} K_h(t_{ij}-t)\varepsilon_{1i}(t_{ij}) + o_p(n^{-1/2}).$$

Then, $\mathcal{S}_i$ becomes

$$\mathcal{S}_i = \begin{pmatrix} N^{-1}\{f(t_{i1})\}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J} K_h(t_{ij}-t_{i1})\varepsilon_{1i}(t_{ij}) + o_p(n^{-1/2}) \\ \vdots \\ N^{-1}\{f(t_{iJ})\}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J} K_h(t_{ij}-t_{iJ})\varepsilon_{1i}(t_{ij}) + o_p(n^{-1/2}) \end{pmatrix}.$$

After obtaining $\mathcal{S}_i$, we plug (4.15) into (4.10). At this point, it is straight forward to calculate the covariance of $\hat{\beta}_W$, whose main components come from the expectation of the square of the terms in (4.15).

□

Chapter 5

Conclusions and Future Work

Motivated by a smoking cessation study (Shiffman et al., 1996), we proposed a nonparametric joint modeling methodology for estimating the association of a longitudinal binary and a continuous response. Our technique can determine both the significance and the direction of this relationship. Furthermore, if the variance of the latent variable is assumed to be equal to that of the continuous response, the estimate of the time-varying correlation can be computed. In addition to estimating the time-varying association, our method allows time-varying effects of predictors, hence researchers can investigate how the relationship between responses, binary and continuous, and predictors varies over time. We developed an estimation procedure via local linear regression and studied the large sample properties of the estimators. We proposed standard error estimates by using conventional techniques. The performance of our technique was examined via a Monte Carlo simulation study. Note that while our method was motivated by a smoking cessation study, it can be applied to a range of longitudinal data analysis problems.

In Chapter 4 we proposed a semiparametric approach to estimate the time-varying correlation between longitudinal binary and continuous responses. As in the nonparametric approach, this technique can also establish the significance and the direction of this association. We developed a two-stage estimation procedure and showed the asymptotic normality of the resulting estimators. The performance of our procedure was assessed via a simulation study and an application to the Women's Interagency HIV Study.

In addition to the work presented in this dissertation, some future research is needed in joint modeling of longitudinal responses:

1. In the present work estimation of the time-varying association requires introducing a latent variable and assuming that the variance of the latent variable is equal to the variance of the continuous response. However, a gaussian copula-based approach would not require these assumptions to estimate the time-varying association. Moreover, the gaussian copula-based approach can be extended to other types of copulas.

2. In this dissertation we investigated a method to estimate the relationship between continuous and binary responses. It would be of interest to extend this work to estimate the association between other types of mixed responses, such as the association between longitudinal ordinal and continuous responses.

3. It is of interest to extend our approach to cases where the relationship between the predictor and the responses, binary and continuous, is nonlinear. Such approach can be employed using nonlinear varying-coefficient models.

Bibliography

Abramowitz, M. and Stegun, I. (1987). *Handbook of Mathematical Functions*. New York: Dover Publications.

Brumback, B. A. and Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of American Statistical Association*, 93(443):961–976.

Burns, D. N., Hillman, D., Neaton, J. D., Sherer, R., Mitchell, T., Capps, L., Vallier, W. G., Thurnherr, M. D., and for the Terry Beirn Community Programs for Clinical Research on AIDS, F. M. G. (1996). Cigarette smoking, bacterial pneumonia, and other clinical outcomes in hiv-1 infection. *Journal of Acquired Immune Deficiency Syndromes*, 13(4):374–383.

Cai, Z., Fan, J., and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95(451):888–902.

Carroll, R. J., Rupert, D., and Welsh, A. H. (1998). Local estimating equations. *Journal of the American Statistical Association*, 93(441):214–227.

Catalano, P. J. and Ryan, L. M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association*, 87(419):651–658.

Chan, J. S. K. and Kuk, A. Y. C. (1997). Maximum likelihood estimation for probit-linear mixed models with correlated random effects. *Biometrics*, 53(1):86–97.

Chiang, C.-T., Rice, J. A., and Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *Journal of the American Statistical Association*, 96(454):605–619.

Cleveland, W., Grosse, E., and Shyu, W. (1992). *Local regression models*. Wadsworth & Brooks/Cole, Pacific Grove, CA.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.

Cox, D. R. and Wermuth, N. (1992). Response models for mixed binary and quantitative variables. *Biometrika*, 79(3):441–461.

Daniels, M. J. and Kass, R. E. (2001). Shrinkage estimators for covariance matrices. *Biometrics*, 57(4):1173–1184.

Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, 27(1):94–128.

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

Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society. Series B*, 62(2):355–366.

Engle, R. F., Granger, C. W. J., Rice, J., and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, 81(394):310–320.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87(420):998–1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, 21(1):196–216.

Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, 20(4):2008–2036.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, 99(467):710–723.

Fan, J. and Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society. Series B*, 62(2):303–322.

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics*, 27(5):1491–1518.

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface*, 1:179–195.

Ferson, M., Edwards, A., Lind, A., Milton, G. W., and Hersey, P. (1979). Low natural killer-cell activity and immunoglobulin levels associated with smoking in human subjects. *International Journal of Cancer*, 23(5):603–609.

Fitzmaurice, G. M. and Laird, N. M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American Statistical Association*, 90(431):845–852.

Galai, N., Park, L. P., Wesch, J., Visscher, B., Riddler, S., and Margolick, J. B. (1997). Effect of smoking on the clinical progression of hiv-1 infection. *Journal of Acquired Immune Deficiency Syndromes*, 14(5):451–8.

Gasser, T. and Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, 11(3):171–185.

Gueorguieva, R. V. and Agresti, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association*, 96(455):1102–1112.

Halonen, M., Barbee, R. A., Lebowitz, M. D., and Burrows, B. (1982). An epidemiologic study of the interrelationships of total serum immunoglobulin e, allergy skin-test reactivity, and eosinophilia. *Journal of Allergy and Clinical Immunology*, 69(2):221–228.

Härdle, W., Liang, H., and Gao, J. (2000). *Partially Linear Models*. New York: Springer-Verlag.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B*, 55(4):757–796.

Heckman, N. E. (1986). Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(2):244–248.

Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85(4):809–822.

Hu, Z., Wang, N., and Carroll, R. J. (2004). Profile-kernel versus backfitting in the partially linear models for longitudinal/clustered data. *Biometrika*, 91(2):251–262.

Huang, J. Z. and Shen, H. (2004). Functional coefficient regression models for non-linear time series: A polynomial spline approach. *Scandinavian Journal of Statistics*, 31(4):515–534.

Huang, J. Z., Wu, C. O., and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1):111–128.

Hughes, D. A., Haslam, P. L., Townsend, P. J., and Turner-Warwick, M. (1985). Numerical and functional alterations in circulatory lymphocytes in cigarette smokers. *Clinical & Experimental Immunology*, 61(2):459–466.

Hymowitz, N., Cummings, K. M., Hyland, A., Lynn, W. R., Pechacek, T. F., and Hartwell, T. D. (1997). Predictors of smoking cessation in a cohort of adult smokers followed for five years. *Tobacco Control*, 6:S57–S62.

Kauermann, G. and Tutz, G. (1999). On model diagnostics using varying coefficient models. *Biometrika*, 86(1):119–128.

Kurum, E., Li, R., Senturk, D., and Wang, Y. (2011). Nonlinear varying coefficient model and its applications. Manuscript sumbitted for publication.

Legler, J. M. and Ryan, L. M. (1997). Latent variable models for teratogenesis using multiple binary outcomes. *Journal of the American Statistical Association*, 92(437):13–20.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Liao, J. (1999). A simplified and accelerated monte carlo EM algorithm with application to a hierarchical mixture model. Technical report.

Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 96(453):103–113.

Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimationg equations. *Journal of American Statistical Association*, 96(455):1045–1056.

Liu, C., Rubin, D. B., and Wu, Y. N. (1998). Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika*, 85(4):755–770.

Liu, X., Daniels, M., and Marcus, B. (2009). Joint models for the association of longitudinal binary and continuous processes with application to a smoking cessation trial. *Journal of the American Statistical Association*, 104(486):429–438.

Martinussen, T. and Scheike, T. H. (1999). A semiparametric additive regression model for longitudinal data. *Biometrika*, 86(3):691–702.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*. Chapman and Hall, London.

McDermut, W. and Haaga, D. A. (1998). Effect of stage of change on cue reactivity in continuing smokers. *Experimental and Clinical Psychopharmacology*, 6(3):316–324.

Moyeed, R. and Diggle, P. (1994). Rates of convergence in semi-parametric modelling of longitudinal data. *Australian Journal of Statistics*, 36(1):75–93.

Müller, H.-G. (1988). *Nonparametric Analysis of Longitudinal Data*, volume 46 of *Lecture Notes in Statistics*. Springer-Verlag.

Nadaraya, E. (1964). On estimating regression. *Theory Probab. Appl.*, 9:141–142.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A*, 135(3):370–384.

Nieman, R. B., Fleming, J., Coker, R. J., William Harris, J. R., and Mitchell, D. M. (1993). The effect of cigarette smoking on the development of aids in hiv-1-seropositive individuals. *AIDS*, 7(5):705–710.

Obirikorang, C. and Yeboah, F. A. (2009). Blood haemoglobin measurement as a predictive indicator for the progression of hiv/aids in resource-limited setting. *Journal of Biomedical Science*, 16(1):102.

Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics*, 25(1):186–211.

Opsomer, J. D. and Ruppert, D. (1999). A root-n consistent backfitting estimator for semiparametric additive modeling. *Journal of Computational and Graphical Statistics*, 8(4):715–732.

Pourahmadi, M. and Daniels, M. J. (2002). Dynamic conditionally linear mixed models for longitudinal data. *Biometrics*, 58(1):225–231.

Regan, M. M. and Catalano, P. J. (1999). Likelihood models for clustered binary and continuous outcomes: Application to developmental toxicology. *Biometrics*, 55(3):760–768.

Rice, J. (1986). Convergence rates for partially splined models. *Statistics & Probability Letters*, 4(4):203–208.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics*, 22(3):1346–1370.

Sammel, M. D. and Ryan, L. M. (1996). Latent variable models with fixed effects. *Biometrics*, 52(2):650–663.

Sammel, M. D., Ryan, L. M., and Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society: Series B*, 59(3):667–678.

Shiffman, S. and Balabanis, M. (1995). Associations between alcohol and tobacco. In Fertig, J. B. and Allen, J. P., editors, *Alcohol and Tobacco: From basic science to clinical practice*, pages 17–36. Bethesda, MD: National Institutes of Health.

Shiffman, S., Gwaltney, C. J., Balabanis, M. H., Liu, K. S., Paty, J. A., Kassel, J. D., Hickcox, M., and Gnys, M. (2002). Immediate antecedents of cigarette smoking: An analysis from ecological momentary assessment. *Journal of Abnormal Psychology*, 111(4):531–545.

Shiffman, S., Hickcox, M., Paty, J. A., Gnys, M., Kassel, J. D., and Richards, T. J. (1996). Progression from a smoking lapse to relapse: Prediction from abstinence violation effects, nicotine dependence, and lapse characteristics. *Journal of Consulting and Clinical Psychology*, 64(5):993–1002.

Smith, M. and Kohn, R. (2002). Parsimonious covariance matrix estimation for longitudinal data. *Journal of the American Statistical Association*, 97(460):1141–1153.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(3):413–436.

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, 5(4):595–620.

Tanner, M. A. (1991). *Tools for Statistical Inference-Observed Data and Data Augmentation Methods*. New York: Springer.

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika*, 90(1):43–52.

Warren, C. A. and McDonough, B. E. (1999). Event-related brain potentials as indicators of smoking cue-reactivity. *Clinical Neurophysiology*, 110(9):1570–1584.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372.

Wei, G. C. G. and Tanner, M. A. (1990). A monte carlo implementation of the em algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, 85(411):699–704.

Wu, C. and Chiang, C. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statistica Sinica*, 10:433–456.

Wu, C. O., Chiang, C.-T., and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association*, 93(444):1388–1402.

Yao, W. and Li, R. (2012). New local estimation procedure for nonparametric regression function of longitudinal data. *Journal of the Royal Statistical Society: Series B. To appear.*

Zeger, S. L. and Diggle, P. (1994). Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv seroconverters. *Biometrics*, 50(3):689–699.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects; a gibbs sampling approach. *Journal of the American Statistical Association*, 86(413):79–86.

Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42(1):121–130.

Zhang, W. and Lee, S.-Y. (2000). Variable bandwidth selection in varying-coefficient models. *Journal of Multivariate Analysis*, 74(1):116–134.

Vita

Esra Kurum

Education

- Ph.D. Statistics, The Pennsylvania State University, 2012.

  Dissertation: Joint Modeling of Longitudinal Continuous and Binary Responses: New Models, Statistical Procedures and Applications.

  Advisor: Runze Li.

- M.S. Statistics, The Pennsylvania State University, 2009.

  Thesis: Estimation in Covariate-adjusted Nonlinear Regression.

  Advisor: Damla Şentürk.

- B.S. Statistics, Middle East Technical University (METU), 2007.

  GPA: 3.90 (over 4.00); *Ranked 1st at graduation in the Faculty of Arts and Science and Department of Statistics.*

  Minor: Computer Engineering-Information Systems.

Publications

- Kurum, E., Li, R., Şentürk., D., Wang, Y. (2012). Nonlinear Varying Coefficient Model and Its Applications. *Scandinavian Journal of Statistics*, under review.
- Hill, L. N., Kolanowski, A., Kurum, E. (2010). Agreeableness and Activity Engagement in Nursing Home Residents with Dementia. *Journal of Gerontological Nursing.*