

TIME-VARYING COEFFICIENT MODELS FOR JOINT MODELING BINARY AND CONTINUOUS OUTCOMES IN LONGITUDINAL DATA

Esra Kürüm¹, Runze Li², Saul Shiffman³, and Weixin Yao⁴

¹*Istanbul Medeniyet University*, ²*The Pennsylvania State University*,

³*University of Pittsburgh* and ⁴*University of California, Riverside*

Abstract: Motivated by an empirical analysis of ecological momentary assessment data (EMA) collected in a smoking cessation study, we propose a joint modeling technique for estimating the time-varying association between two intensively measured longitudinal responses: a continuous one and a binary one. A major challenge in joint modeling these 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. We develop a two-stage estimation procedure and establish the asymptotic normality of the resulting estimators. We also derived the standard error formulas for estimated coefficients. We conduct a Monte Carlo simulation study to assess the finite sample performance of our procedure. The proposed method is illustrated by an empirical analysis of smoking cessation data, in which the important question of interest is to investigate the association between urge to smoke, continuous response, and the status of alcohol use, the binary response, and how this association varies over time.

Key words and phrases: Generalized linear models; Local linear regression; Varying coefficient models.

1. Introduction

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 was motivated by an empirical analysis in Section 3.2, in which the data were collected intensively during a smoking cessation study (Shiffman et al., 1996) and 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), and both have been associated with an increased risk of lapsing back to smoking (Kahler et al., 2010; Shiffman et al., 2007). Moreover, there is some hint that the relationship between these stimuli and smoking (and therefore perhaps urge to smoke) may vary over time, particularly weakening after the initial few days of abstinence. In this paper 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. Hence, we develop a new joint modeling method for longitudinal binary and continuous responses, along with a corresponding estimation procedure.

The major difficulty in modeling binary and continuous responses jointly is the lack of a natural multivariate distribution. To overcome this difficulty, many authors (Catalano and Ryan, 1992; Cox and Wermuth, 1992; Dunson, 2000; Fitzmaurice and Laird, 1995; Gueorguieva and Agresti, 2001; Liu et al., 2009; Regan and Catalano, 1999; Sammel et al., 1997) have employed what is now a well-known method, namely, introducing a continuous latent variable underlying the binary response, and assuming 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.

Motivated by our empirical analysis in Section 3, we proposed time-varying coefficient models for jointly modeling binary and continuous response in this paper. Time-varying coefficient models have been introduced to model continuous response in both independent and identically distributed (iid) data and longitudinal data (Hastie and Tibshirani, 1993; Brumback and Rice, 1998; Hoover et al., 1998; Wu et al., 1998; Zhang and Lee, 2000), and have been proposed for iid data

with binary response (Cai et al., 2000). To our best knowledge, time-varying coefficient models have not been applied for jointly modeling binary and continuous responses in longitudinal data setting. This work will fill this gap in the literature. In this article we focus on estimating the time-varying association between longitudinal binary and continuous responses measured at the same time point within a subject.

We proposed an estimation procedure to time-varying coefficient model for jointly modeling binary and responses outcomes in longitudinal data. Adapting from the existing literature, 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. In the second stage we use generalized time-varying coefficient models (Cai et al. (2000) for iid data) to fit the conditional model of the binary response. We systematically study the sampling property of the newly proposed estimation procedure. We establish the asymptotic normality of the proposed estimators. The efficacy of our methodology is demonstrated by a simulation study.

The remainder of the paper is organized as follows. In Section 2, we first propose a joint model for longitudinal binary and continuous responses, and further develop our two-stage estimation procedure by using local linear regression techniques. In this section, we also study asymptotic properties of the resulting estimators. In Section 3 we conduct an extensive simulation study to investigate the finite sample behavior of our estimators, and further illustrate the proposed methodology by a real data example. Regularity conditions and proofs are given in the supplementary material.

2. Joint Models for Binary and Continuous Responses

In this section, we first propose time-varying coefficient models for joint modeling binary and continuous responses in Section 2.1. We propose an estimation procedure in Section 2.2, and study the sampling property of the proposed estimate in Section 2.3.

2.1. Joint Models

Let us begin with introduction of notation. For the i th subject, $i = 1, \dots, n$, denote the binary response measured at time point t_{ij} by $Q_i(t_{ij})$, the continuous response by $W_i(t_{ij})$, where $j = 1, \dots, n_i$. In addition, define the latent variable underlying $Q_i(t_{ij})$ by $Y_i(t_{ij})$. Let $\mathbf{X}_i(t_{ij}) = (X_{i1}(t_{ij}), \dots, X_{ip}(t_{ij}))^\top$ be the vector of predictors with $X_{i1} \equiv 1$ to include an intercept term, $\boldsymbol{\beta}(t_{ij}) = (\beta_1(t_{ij}), \dots, \beta_p(t_{ij}))^\top$, and $\boldsymbol{\alpha}(t_{ij}) = (\alpha_1(t_{ij}), \dots, \alpha_p(t_{ij}))^\top$ be the vectors of regression coefficients. Consider the following bivariate model:

$$\begin{aligned} W_i(t_{ij}) &= \mathbf{X}_i^\top(t_{ij})\boldsymbol{\beta}(t_{ij}) + \varepsilon_{1i}(t_{ij}), \\ Y_i(t_{ij}) &= \mathbf{X}_i^\top(t_{ij})\boldsymbol{\alpha}(t_{ij}) + \varepsilon_{2i}(t_{ij}), \end{aligned} \quad (2.1)$$

where both $\varepsilon_{1i}(t)$ and $\varepsilon_{2i}(t)$ follow normal distributions with mean zero and time-varying variances $\sigma_1^2(t)$ and $\sigma_2^2(t)$, respectively, **which are strictly positive**. Furthermore, we define $\boldsymbol{\varepsilon}_i(t_{ij}) = (\varepsilon_{1i}(t_{ij}), \varepsilon_{2i}(t_{ij}))^\top$, and assume that $\boldsymbol{\varepsilon}_i(t_{ij})$ follows a bivariate normal distribution with $\text{corr}\{\varepsilon_{1i}(t_{ij}), \varepsilon_{2i}(t_{ij})\} = \tau(t_{ij})$. 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

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

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 $j = 1, \dots, n_i$. The marginal model for the continuous response is defined in (2.1). 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,

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

where

$$\mu_i(t_{ij}) = \mathbf{X}_i^T(t_{ij})\boldsymbol{\alpha}(t_{ij}) + \frac{\sigma_2(t_{ij})}{\sigma_1(t_{ij})}\tau(t_{ij})\varepsilon_{1i}(t_{ij}), \quad (2.2)$$

and

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

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

$$P \{Q_i(t_{ij}) = 1 \mid W_i(t_{ij})\} = \Phi \left[\frac{\mu_i(t_{ij})}{\sqrt{\sigma_2^2(t_{ij}) \{1 - \tau^2(t_{ij})\}}} \right], \quad (2.3)$$

where $\mu_i(t_{ij})$ is defined in (2.2). **Note that the bivariate normal assumption for $\varepsilon_i(t_{ij})$ is only necessary to obtain the conditional distribution of the binary response given the continuous response as described in (2.3); it is not necessary for the asymptotic results shown in Section 2.3.**

Not all parameters in model (2.3) are estimable, hence we reparameterize (2.3) to a more parsimonious and fully estimable form:

$$P \{Q_i(t_{ij}) = 1 \mid W_i(t_{ij})\} = \Phi \left\{ \sum_{r=1}^p X_{ir}(t_{ij})\alpha_r^*(t_{ij}) + \alpha_{p+1}^*(t_{ij})\varepsilon_{1i}(t_{ij}) \right\}, \quad (2.4)$$

where $\alpha_r^*(t_{ij}) = \alpha_r(t_{ij})/\sqrt{\sigma_2^2(t_{ij}) \{1 - \tau^2(t_{ij})\}}$ with $r = 1, \dots, p$. Model (2.4) 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

$$\alpha_{p+1}^*(t_{ij}) = \frac{1}{\sigma_1(t_{ij})} \cdot \frac{\tau(t_{ij})}{\sqrt{1 - \tau^2(t_{ij})}}.$$

Therefore, $\alpha_{p+1}^*(t_{ij})$ has the same sign as $\tau(t_{ij})$ and increases when $\tau(t_{ij})$ increases. Let $b(t_{ij}) = \alpha_{p+1}^*(t_{ij})\sigma_1(t_{ij})$. Then

$$\tau(t_{ij}) = \frac{b(t_{ij})}{\sqrt{1 + b^2(t_{ij})}}. \quad (2.5)$$

This joint modeling approach introduces time-varying effects to the well-known joint modeling framework suggested by Catalano and Ryan (1992), that

is, introducing a normally distributed latent variable and then decomposing the joint distribution of the continuous variable and the binary variable. This decomposition can be done one of two ways: (1) a marginal distribution for the continuous response along with a conditional distribution for the binary response given the continuous response, or (2) a marginal distribution for the binary response along with a conditional distribution for the continuous response given the binary response. Although we followed the first formulation, one can easily employ the second option, and extend the work by Fitzmaurice and Laird (1995) and include time-varying effects. Another approach would be to incorporate time-varying effects to the joint mixed-effects model (Gueorguieva, 2001; Gueorguieva and Agresti, 2001). However, as pointed out by Verbeke et al. (2010), maximum likelihood estimation in this approach is possible only when strong assumptions are made. An example of this is demonstrated by Roy and Lin (2000), where corresponding random effects for various outcomes are assumed to be perfectly correlated. Moreover, confounding can also be a problem with the mixed-effects approach (Hodges and Reich, 2010). This may lead to an increase in the variance of fixed-effects estimators; hence, preventing the discovery of important response–predictor relationships.

2.2. Estimation Procedure

We propose a two-stage estimation procedure to estimate the time-varying association 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) to the marginal model of the continuous response (2.1). Here we employ the local linear fitting (Fan and Gijbels, 1996) to estimate the nonparametric coefficient functions. 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 $r = 1, \dots, p$. Let $\mathbf{a} = (a_1, \dots, a_p)^\top$, and $\mathbf{b} = (b_1, \dots, b_p)^\top$. We minimize the

following equation:

$$\ell(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^n \sum_{j=1}^{n_i} \{W_i(t_{ij}) - \mathbf{X}_i^T(t_{ij})\mathbf{a} - \mathbf{X}_i^T(t_{ij})\mathbf{b}(t_{ij} - t_0)\}^2 K_{h_1}(t_{ij} - t_0), \quad (2.6)$$

with respect to (\mathbf{a}, \mathbf{b}) , where $K_h(\cdot) = h^{-1}K(\cdot/h)$ with $K(\cdot)$ as the kernel function and h_1 is the bandwidth at the first stage. Let $\mathbf{W} = (\mathbf{W}_1^T, \dots, \mathbf{W}_n^T)^T$ be the vector of continuous responses for all subjects with $\mathbf{W}_i = (W_{i1}, \dots, W_{in_i})^T$ and $i = 1, \dots, n$. The solution to the least squares algorithm is,

$$\hat{\mathbf{a}} = \hat{\boldsymbol{\beta}}(t_0) = (I_p, 0_p)(\mathcal{X}^T \kappa \mathcal{X})^{-1} \mathcal{X}^T \kappa \mathbf{W}, \quad (2.7)$$

where I_p is the identity matrix with size p , 0_p is a size p matrix with each entry equal to zero, $\mathcal{X} = (\mathcal{X}_1, \dots, \mathcal{X}_n)^T$, $\mathcal{X}_i = ((1, t_{i1} - t_0) \otimes \mathbf{X}_{i1}, \dots, (1, t_{in_i} - t_0) \otimes \mathbf{X}_{in_i})$ and κ is an $N \times N$ diagonal matrix with each entry equal to $K_{h_1}(t_{ij} - t_0)$ for $i = 1, \dots, n$ and $j = 1, \dots, n_i$.

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:

$$\widehat{\text{cov}}\{\hat{\boldsymbol{\beta}}(t_0)\} \approx (I_p, 0_p)(\mathcal{X}^T \kappa \mathcal{X})^{-1} (\mathcal{X}^T \kappa \mathcal{Q} \kappa \mathcal{X}) (\mathcal{X}^T \kappa \mathcal{X})^{-1} (I_p, 0_p)^T, \quad (2.8)$$

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

In the second stage we fit a generalized time-varying coefficient model to the conditional model (2.4) with longitudinal data. The generalized varying coefficient models for independent and identically distributed data were introduced by Cai et al. (2000). We adapt the related techniques to a longitudinal setting in this paper. We locally approximate the functions in a neighborhood of a fixed point t_0 via the Taylor expansion,

$$\alpha_r^*(t) \approx \alpha_r^*(t_0) + \alpha_r^{*'}(t_0)(t - t_0) \equiv a_r^* + b_r^*(t - t_0),$$

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

likelihood,

$$\ell_n(\mathbf{a}^*, \mathbf{b}^*) = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{n_i} \ell \left(g^{-1} \left[\sum_{r=1}^{p+1} \{a_r^* + b_r^*(t_{ij} - t_0)\} X_{ir}^*(t_{ij}) \right], Q_i(t_{ij}) \right) K_{h_2}(t_{ij} - t_0), \quad (2.9)$$

where $g(\cdot)$ is the link function and h_2 is the bandwidth at the second stage. For our model (2.4), the link function is probit. Hence, the local likelihood with probit link becomes

$$\begin{aligned} \ell_n(\mathbf{a}^*, \mathbf{b}^*) &= \frac{1}{N} \sum_{Q_i(t_{ij})=1} \log \left(\Phi \left[\sum_{r=1}^{p+1} \{a_r^* + b_r^*(t_{ij} - t_0)\} X_{ir}^*(t_{ij}) \right] \right) K_{h_2}(t_{ij} - t_0) \\ &\quad + \frac{1}{N} \sum_{Q_i(t_{ij})=0} \log \left(1 - \Phi \left[\sum_{r=1}^{p+1} \{a_r^* + b_r^*(t_{ij} - t_0)\} X_{ir}^*(t_{ij}) \right] \right) K_{h_2}(t_{ij} - t_0), \end{aligned} \quad (2.10)$$

where $\Phi(\cdot)$ is the cumulative distribution function (cdf) 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 (2.10). 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 (2.10). Then, we update $(\mathbf{a}^*, \mathbf{b}^*)$ according to

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

Let $\boldsymbol{\alpha}^*(t) = (\alpha_1^*(t), \dots, \alpha_p^*(t), \alpha_{p+1}^*(t))^T$. The solution of this iterative regression algorithm satisfies $\ell(\mathbf{a}^*, \mathbf{b}^*) = 0$ and the estimators are given by $\hat{\mathbf{a}}^* = \hat{\boldsymbol{\alpha}}^*(t_0) = (\hat{\alpha}_1^*(t_0), \dots, \hat{\alpha}_p^*(t_0), \hat{\alpha}_{p+1}^*(t_0))^T$. The asymptotic covariance matrix of these estimators can be estimated as follows:

$$\widehat{\text{cov}}\{\hat{\boldsymbol{\alpha}}^*(t_0)\} = (I_p, \theta_p) \widehat{\Gamma}(t_0)^{-1} \widehat{\Lambda}(t_0) \widehat{\Gamma}(t_0)^{-1} (I_p, \theta_p)^T, \quad (2.11)$$

where

$$\begin{aligned} \widehat{\Gamma}(t_0) &= \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{n_i} \varpi_2 \left[\sum_{r=1}^{p+1} \{a_r^* + b_r^*(t_{ij} - t_0)\} X_{ir}^*(t_{ij}), Q_i(t_{ij}) \right] K_{h_2}(t_{ij} - t_0) \begin{pmatrix} \mathbf{X}_i^*(t_{ij}) \\ \mathbf{X}_i^*(t_{ij})(t_{ij} - t_0) \end{pmatrix}^{\otimes 2}, \\ \widehat{\Lambda}(t_0) &= \frac{h}{N} \sum_{i=1}^n \sum_{j=1}^{n_i} \varpi_1^2 \left[\sum_{r=1}^{p+1} \{a_r^* + b_r^*(t_{ij} - t_0)\} X_{ir}^*(t_{ij}), Q_i(t_{ij}) \right] K_{h_2}^2(t_{ij} - t_0) \begin{pmatrix} \mathbf{X}_i^*(t_{ij}) \\ \mathbf{X}_i^*(t_{ij})(t_{ij} - t_0) \end{pmatrix}^{\otimes 2} \end{aligned}$$

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

The estimator of a_{p+1}^* gives us $\hat{\alpha}_{p+1}^*(t_0)$ and the pointwise asymptotic confidence intervals of $\alpha_{p+1}^*(t_0)$ give us information on the significance of the association. To obtain $\hat{\tau}(t_0)$, we also need to find the estimate for $\sigma_1^2(t_0)$. We propose using the kernel estimator

$$\hat{\sigma}_1^2(t_0) = \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} e_i^2(t_{ij}) K_h(t_{ij} - t_0)}{\sum_{i=1}^n \sum_{j=1}^{n_i} K_h(t_{ij} - t_0)}. \quad (2.12)$$

Plugging $\hat{\sigma}_1(t_0)$ and $\hat{\alpha}_{p+1}^*(t_0)$ into (2.5) gives the estimate for $\tau(t_0)$.

2.3. Asymptotic Results

In this section we study the asymptotic properties of the estimators in both stages of the estimation procedure. It is assumed throughout the rest of this paper that $n_i = J$ and thus $N = nJ$ in order to simplify the presentation of the asymptotic results. The proofs of the theorems are provided in the supplementary material.

Let $f(t)$ denote the marginal density of t_{ij} and $f(t_{ij}, t_{ik})$ denote the joint density of t_{ij} and t_{ik} for $j \neq k$. Let $\mu_k = \int t^k K(t) dt$, $\nu_k = \int t^k K^2(t) dt$, and

$$\begin{aligned} \Gamma_1(t_0) &= E\{\mathbf{X}_i(t_{ij})\mathbf{X}_i^\top(t_{ij}) \mid t_{ij} = t_0\}, \\ \eta_{lr}(t_1, t_2) &= E\{X_{il}(t_{ij})X_{ir}(t_{ik}) \mid t_{ij} = t_1, t_{ik} = t_2\}, \\ \rho_1(t_1, t_2) &= E\{\varepsilon_{1i}(t_{ij})\varepsilon_{1i}(t_{ik}) \mid t_{ij} = t_1, t_{ik} = t_2\}, \end{aligned}$$

where $\varepsilon_{1i}(t_{ij})$ is the error term in (2.1), $l, r = 1, \dots, p$, $i = 1, \dots, n$, and $j = 1, \dots, J$.

The following regularity conditions are needed to state the first main result.

- A. The observed sample $\{t_{ij}, \mathbf{X}_i(t_{ij}), W_i(t_{ij}), i = 1, \dots, n\}$ are an independent and identically distributed (iid) realization of (T, X, W) for all $j = 1, \dots, J$. The $\{\varepsilon_{1i}(t_{ij}), i = 1, \dots, n\}$ are iid from a distribution with mean zero and finite variance $\sigma_1^2(t_{ij})$ for $j = 1, \dots, J$. The covariate T has finite support $\mathcal{T} = [\mathcal{L}, \mathcal{U}]$. The support for \mathbf{X} is a closed and bounded interval in \mathbb{R}^p , denoted by Ω .
- B. $\beta_r(t)$ has continuous second order derivatives for $r = 1, \dots, p$.
- C. $\Gamma_1(t), \eta_{lr}(t_1, t_2), \rho_1(t_1, t_2), \sigma_1(t), f(t)$, and $f(t_1, t_2)$ are continuous for $l, r = 1, \dots, p$,

D. The kernel density function $K(\cdot)$ is symmetric about 0 with bounded support and satisfies the Lipschitz condition and

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

E. $E\{|\varepsilon_{1i}(t_{ij})|^3 \mid t_{ij}\} < \infty$ and is continuously differentiable.

By Condition B, we assume that the parameter space for $(\beta(t_0), \beta'(t_0))$ is a closed and bounded subset of \mathbb{R}^{2p} for any given t_0 . The continuous condition of $\rho_1(t_1, t_2)$ and $\eta_{lr}(t_1, t_2)$ when t_1 and t_2 converges to the same time point might not hold if the predictors and error process contain some measurement errors that are independent at different time points t . However our proofs are still valid after some slight modifications of notations. For example, we can replace $\rho_1(t_0, t_0) = \sigma_1^2(t_0)$ by $\lim_{t_1 \rightarrow t_0, t_2 \rightarrow t_0} \rho_1(t_1, t_2)$. The bounded support condition in D about kernel function is imposed for the simplicity of proof and can be relaxed.

The first theorem shows the asymptotic properties of the estimators $\hat{\beta}(t_0)$ in the first stage of the estimation procedure, which are obtained by minimizing the weighted least squares (2.6).

Theorem 1 *Under the regularity conditions (A)–(E), if $Jh_1 \rightarrow 0$ and $Nh_1 \rightarrow \infty$, we have the following result for $\hat{\beta}(t_0)$:*

$$\sqrt{Nh_1} \left\{ \hat{\beta}(t_0) - \beta(t_0) - \frac{1}{2} h_1^2 \mu_2 \beta''(t_0) + o_p(h_1^2) \right\} \xrightarrow{L} N_p(0, V_1),$$

where $\beta(t_0)$ is the true value and

$$V_1 = f(t_0)^{-1} \nu_0 \sigma_1^2(t_0) \Gamma_1^{-1}(t_0).$$

Note that the above results require the condition $Jh_1 \rightarrow 0$, which holds if the number of observations per subject, J , is finite or goes to infinity in a slower rate compared to h_1^{-1} . Based on the above theorem, we can see that the asymptotic bias of $\hat{\beta}(t_0)$ is $\frac{1}{2} h_1^2 \mu_2 \beta''(t_0)$ and the asymptotic variance of $\hat{\beta}(t_0)$ is $(Nh_1)^{-1} f(t_0)^{-1} \nu_0 \sigma_1^2(t_0) \Gamma_1^{-1}(t_0)$. Therefore, the asymptotic bias and variance are the same as the independent data. This is meaningful, since the condition $Jh_1 \rightarrow 0$ guarantees that asymptotically there is no more than one effective observation per subject in the local area around t_0 .

Theorem 1 can be proved by noting that

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = (\mathcal{X}^T \kappa \mathcal{X})^{-1} \mathcal{X}^T \kappa (\mathbf{W} - \mathcal{X} \boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta} = (\boldsymbol{\beta}(t_0), h_1 \boldsymbol{\beta}'(t_0))$. We can show that $N^{-1} \mathcal{X}^T \kappa \mathcal{X}$ converges in probability to

$$f(t_0) \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \Gamma_1(t_0)$$

and that $\sqrt{N h_1} \{N^{-1} \mathcal{X}^T \kappa (\mathbf{W} - \mathcal{X} \boldsymbol{\theta}_0) - \text{bias}(t_0)\}$ converges to the normal distribution. Thus, Theorem 1 follows by using the Slutsky's Theorem. For more detail about the proof, please see our supplementary file.

Next, we give the asymptotic properties of the estimators, $\hat{\boldsymbol{\alpha}}^*(t_0)$, in the second stage of the estimation procedure, which are obtained by maximizing the local likelihood (2.9).

Let $\tilde{\mathbf{X}}_{ij} = (\mathbf{X}_i^T(t_{ij}), \varepsilon_{1i}(t_{ij}))^T$, $\theta(t_{ij}) = \{\alpha_{p+1}^*(t_0) + \alpha_{p+1}'^*(t_0)(t_{ij} - t_0)\}$ and

$$m(t_{ij}, \tilde{\mathbf{x}}_{ij}) = E\{Q_i(t_{ij}) \mid t_{ij}, \tilde{\mathbf{x}}_{ij}\} = \tilde{\mathbf{x}}_{ij}^T \boldsymbol{\alpha}^*(t_{ij}) = \sum_{r=1}^p \alpha_r^*(t_{ij}) x_{ijr} + \alpha_{p+1}^*(t_{ij}) \varepsilon_{1i}(t_{ij}),$$

$$\rho(t_{ij}, \tilde{\mathbf{x}}_{ij}) = -\varpi_2[g\{m(t_{ij}, \tilde{\mathbf{x}}_{ij})\}, m(t_{ij}, \tilde{\mathbf{x}}_{ij})],$$

$$\Gamma_2(t_0) = E\{\rho(t_{ij}, \tilde{\mathbf{X}}_{ij}) \tilde{\mathbf{X}}_{ij} \tilde{\mathbf{X}}_{ij}^T \mid t_{ij} = t_0\},$$

$$\boldsymbol{\omega}(t_{ij}) = \varpi_1[g\{m(t_{ij}, \tilde{\mathbf{X}}_{ij})\}, Q_i(t_{ij})] \tilde{\mathbf{X}}_{ij} - \theta(t_{ij}) \Gamma_2(t_{ij}) \Gamma_1^{-1}(t_{ij}) \mathbf{X}_i(t_{ij}) \varepsilon_{1i}(t_{ij}),$$

$$\Gamma_3(t_1, t_2) = E\{\boldsymbol{\omega}(t_{ij}) \boldsymbol{\omega}(t_{ik})^T \mid t_{ij} = t_1, t_{ik} = t_2\}.$$

We need several more conditions for the second result.

- F. The function $\varpi_2(\mathcal{Z}, q) < 0$ for $\mathcal{Z} \in \mathcal{R}$, and q in the range of the binary response.
- G. The varying coefficient functions $\alpha_r^*(t_{ij})$, $r = 1, \dots, p + 1$ has continuous second order derivatives.
- H. The functions $\Gamma_2(t)$, $\Gamma_3(t_1, t_2)$, $\varpi_1(\cdot, \cdot)$, $\varpi_2(\cdot, \cdot)$, and $\varpi_3(\cdot, \cdot)$ are continuous.

Condition (F) guarantees that the local likelihood function (2.9) is concave.

Theorem 2 *Under regularity conditions (A)—(H), if $Nh_1^5 \rightarrow 0$, $nh_1^{1+\delta} \rightarrow \infty$ for some $\delta > 0$, $Jh_2 \rightarrow 0$, and $Nh_2 \rightarrow \infty$, we have the following result:*

$$\sqrt{Nh_2} \left\{ \hat{\boldsymbol{\alpha}}^*(t_0) - \frac{1}{2}h_2^2\mu_2\boldsymbol{\alpha}^{*''}(t_0) + o_p(h_2^2) \right\} \xrightarrow{L} N(0, V_2),$$

where

$$V_2 = f(t_0)^{-1}\nu_0\Gamma_2^{-1}(t_0)\Gamma_3(t_0, t_0)\Gamma_2^{-1}(t_0).$$

Based on the above theorem, we can see that the asymptotic normality of $\hat{\boldsymbol{\alpha}}^*(t_0)$ requires the under-smoothing condition of $Nh_1^5 \rightarrow 0$ in the first stage. In addition, we can see that the asymptotic bias of $\hat{\boldsymbol{\alpha}}^*(t_0)$ is $\frac{1}{2}h_2^2\mu_2\boldsymbol{\alpha}^{*''}(t_0)$ and the asymptotic variance of $\hat{\boldsymbol{\alpha}}^*(t_0)$ is $(Nh_2)^{-1}f(t_0)^{-1}\nu_0\Gamma_2^{-1}(t_0)\Gamma_3(t_0, t_0)\Gamma_2^{-1}(t_0)$.

Theoretical guarantee to accurately estimate $\tau(t)$ depends on the accuracy of the kernel estimate of the variance $\sigma_1^2(t)$ and the estimate for $\alpha_{p+1}^*(t)$ in (2.4). We proved the latter in Theorem 2 and the former was shown in Fan et al. (2007).

Based on the quadratic approximation lemma (see Fan and Gijbels, 1996, p.210), we can get a simple expansion for the estimator $(\hat{\mathbf{a}}^*, \hat{\mathbf{b}}^*)$ with the following form

$$\hat{\boldsymbol{\vartheta}}^* = \Delta^{-1}D_n + o_p(1),$$

where

$$\hat{\boldsymbol{\vartheta}}^* = \sqrt{Nh_2} \left(\hat{\mathbf{a}}^* - \boldsymbol{\alpha}^*(t_0), h_2\{\hat{\mathbf{b}}^* - \boldsymbol{\alpha}^{*'}(t_0)\} \right),$$

$$\Delta = f(t_0) \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \Gamma_2(t_0),$$

$$D_n = (Nh_2)^{-1/2}h_2 \sum_{i=1}^n \sum_{j=1}^J \varpi_1 \left[\mathbf{x}_{ij}^{*T} \{ \boldsymbol{\alpha}^*(t_0) + (t_{ij} - t_0)\boldsymbol{\alpha}^{*'}(t_0) \}, Q_{ij} \right] \mathbf{z}_{ij}^* K_{h_2}(t_{ij} - t_0),$$

$$\mathbf{z}_{ij}^* = \left(\mathbf{x}_{ij}^{*T}, (t_{ij} - t_0)/h_2\mathbf{x}_{ij}^{*T} \right)^T,$$

and \mathbf{x}_{ij}^* is the estimate of $\tilde{\mathbf{x}}_{ij}$ by replacing $\varepsilon_1(t_{ij})$ with $e_i(t_{ij})$, the residual from the marginal model. Then, the result of Theorem 2 follows if we can establish the asymptotic normality of D_n . The main difficulty dealing with D_n lies in the the residual $e_i(t_{ij})$ used in \mathbf{x}_{ij}^* . Based on the Taylor expansion of D_n around $\tilde{\mathbf{x}}_{ij}$ and the asymptotic results for the residual $e_i(t_{ij})$ in Theorem 1, we can prove the asymptotic normality of D_n . The under-smoothing condition of $Nh_1^5 \rightarrow 0$ in the

first stage allow the bias of $e_i(t_{ij})$ to be asymptotically negligible for the second stage estimator. For more detail about the proof, please see our supplementary file.

3. Numerical Studies

In this section we examine the finite sample performance of the proposed methodology via a Monte Carlo simulation study, and illustrate the proposed methodology by a real data example. In this section we set $K(t) = 0.75(1 - t^2)_+$, the Epanechnikov kernel.

3.1. Simulation Studies

In this study we generate 500 intensive longitudinal data sets, in which for each unit the number of measurements, n_i , is randomly selected using a discrete uniform distribution on [10, 20] and the measurement times $\mathbf{T}_i = (t_{i1}, \dots, t_{in_i})$ are generated from a uniform distribution on [0, 1]. We use sample size $n = 150$. The continuous and latent variables are generated from the following models:

$$\begin{aligned} W_i(t_{ij}) &= \beta_1(t_{ij}) + \beta_2(t_{ij})X_i(t_{ij}) + \varepsilon_{1i}(t_{ij}), \\ Y_i(t_{ij}) &= \alpha_1(t_{ij}) + \alpha_2(t_{ij})X_i(t_{ij}) + \varepsilon_{2i}(t_{ij}), \end{aligned} \quad (3.1)$$

where $\beta_1(t_{ij}) = \sin(0.5\pi t_{ij})$, $\beta_2(t_{ij}) = \cos(\pi t_{ij} - 1/8)$, $\alpha_1(t_{ij}) = \sin(\pi t_{ij}) - 0.5$, $\alpha_2(t_{ij}) = 0.5 \cos(2\pi t_{ij})$, $i = 1, \dots, 150$ and $j = 1, \dots, n_i$. In Section 2.1 we defined the binary variable as $Q_i(t_{ij}) = 1$ if $Y_i(t_{ij}) > 0$, and $Q_i(t_{ij}) = 0$ if $Y_i(t_{ij}) \leq 0$. It is interesting 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% of response being 0. 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, $\varepsilon_i(t_{ij}) =$

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})
.10	.078 (.016)	.30	.051	.049	(.005)	.049	.049	(.005)
		.50	.042	.037	(.003)	.038	.037	(.003)
		.70	.054	.049	(.004)	.051	.049	(.005)
.20	.062 (.014)	.30	.035	.034	(.002)	.035	.034	(.003)
		.50	.032	.028	(.002)	.028	.028	(.002)
		.70	.038	.034	(.002)	.033	.033	(.003)
.40	.072 (.014)	.30	.025	.024	(.001)	.023	.023	(.002)
		.50	.022	.022	(.001)	.022	.022	(.001)
		.70	.026	.024	(.001)	.023	.023	(.002)

$(\varepsilon_{1i}(t_{ij}), \varepsilon_{2i}(t_{ij}))^T$ 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 (3.1). To evaluate the performance of the estimators in this stage, we use root average squared error (RASE),

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

where $\{t_k, k = 1, \dots, n_{grid}\}$ is an equidistant set of grid points between 0 and 1 with $n_{grid} = 200$ used in our simulation.

We are interested in examining the performance of the proposed procedure with a range of bandwidths. 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, we set the bandwidth to be $h_1 = 0.20$ for the first stage.

Figures (a) and (b) depict the typical estimates of the parameter functions along with the empirical and the mean theoretical pointwise 95% confidence bands based on 500 Monte Carlo simulations at $h_1 = 0.20$. We see that the typical estimated coefficient functions are close to the true functions. The standard errors for the selected bandwidth of $h = 0.20$ are very accurate for these parameters.

We also test the accuracy of the proposed standard error formula (2.8). The

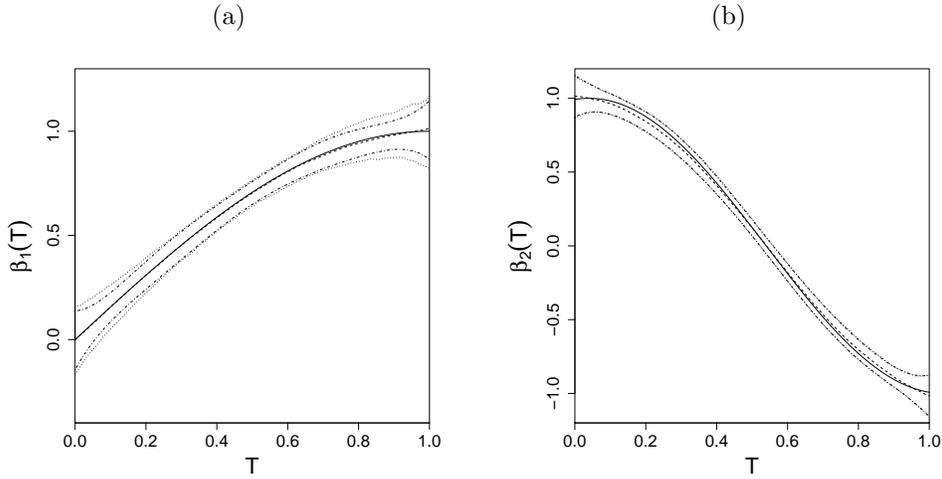


Figure 3.1: Estimated varying coefficient functions (dashed) of the time-varying coefficient model fit to the continuous response overlaying the true coefficient functions (solid) along with the empirical (dotted) and the mean theoretical (dashed-dotted) pointwise 95% confidence bands based on 500 Monte Carlo replications.

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 (2.8). 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 their difference is less than two times the SD_{se} . This is typical for standard error estimation by using sandwich formula.

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.

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

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. At this stage we obtain $\hat{\alpha}_k^*(t)$ ($k = 1, 2, 3$) and the kernel estimate of $\sigma_1^2(t)$ at the optimal bandwidth for the first stage $h_1 = 0.20$ using (2.12), then

Table 3.2: Summary of simulation results for the second stage

h	RASE	t	$\hat{\alpha}_1^*(t)$		$\hat{\alpha}_2^*(t)$		$\hat{\alpha}_3^*(t)$	
	Mean(SD)		SD	SE (SD _{se})	SD	SE (SD _{se})	SD	SE (SD _{se})
.10	.080(.020)	.30	.069	.066(.002)	.068	.066(.003)	.071	.071(.005)
		.50	.073	.069(.002)	.077	.074(.005)	.100	.095(.007)
		.70	.070	.066(.002)	.067	.066(.003)	.077	.071(.005)
.20	.052(.019)	.30	.048	.046(.001)	.047	.047(.002)	.056	.052(.003)
		.50	.048	.048(.001)	.052	.050(.002)	.064	.060(.004)
		.70	.048	.046(.001)	.046	.046(.002)	.057	.053(.003)
.40	.064(.020)	.30	.035	.034(.001)	.035	.034(.001)	.043	.041(.002)
		.50	.034	.033(.001)	.034	.033(.001)	.041	.037(.002)
		.70	.034	.034(.001)	.034	.034(.001)	.044	.041(.002)

we estimate $\tau(t)$ using (2.5). We evaluate the performance of the correlation coefficient estimator using RASE,

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

Table 3.2 gives 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 3.2, we set the bandwidth to be $h_2 = 0.20$.

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 $h_2 = 0.20$. Figure 3.2 indicates that the typical estimated correlation coefficient function is close to the underlying true correlation coefficient function.

We now test the accuracy of our standard error formula (2.11). Table 3.2 presents the SD, SE and SD_{se} values at the points $t = 0.30, 0.50$ and 0.70 for $\hat{\alpha}_k^*(t)$ with $k = 1, 2, 3$. This table suggests that our formula somewhat underestimates the true standard error but their difference is within two standard deviations of the estimated standard errors. This is typical for standard error estimation by using sandwich formula.

3.2. Application to the Smoking Cessation Study

In this section we apply our proposed joint modeling methodology to the

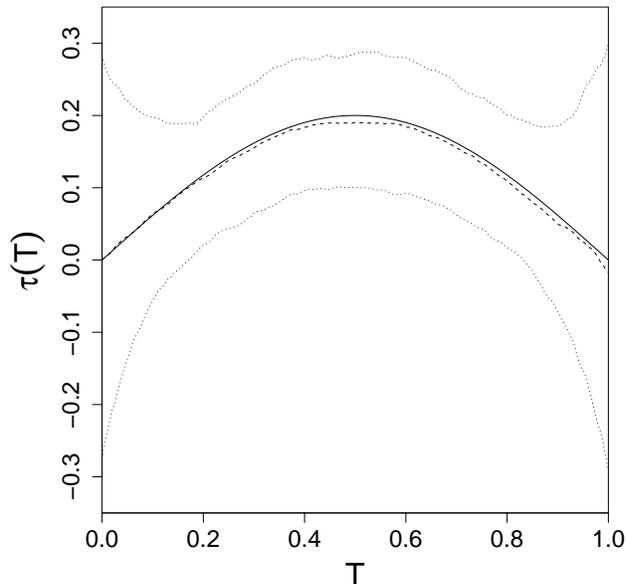


Figure 3.2: Estimated correlation coefficient function (dashed) overlaying the true correlation function (solid) along with the 2.5 and 97.5 percentiles based on 500 Monte Carlo runs (dotted).

EMA data described in the introduction. Shiffman et al. (1996, 2002) 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 target quit 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. **The number of observations for each subject varies from 23 to 197.**

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 target quit 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. Although previous studies have shown that the relationship between smoking and drinking alcohol is positive, it has been a concern that this association becomes negative during smoking cessation programs—increased drinking might be associated with reduced smoking. Hence, 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 following leave-one-subject-out cross validation score for both stages of our estimation procedure:

$$CV(h) = \sum_i \|V_i - \hat{V}_{-i}\|^2, \quad (3.2)$$

where V_i denotes the observed value of the response V for subject i and \hat{V}_{-i} is the fitted value of this response with subject i excluded. V stands for the continuous and binary responses, while choosing the bandwidth for the first and second stages, respectively. We compute this cross validation score for a range

of bandwidths and select the bandwidth that minimizes this score. The selected bandwidth $h = 5.0$ for both stages. We evaluated the performance of this cross validation formula using our simulation study presented in Section 3.1. The mean absolute difference between the bandwidths that minimizes the RASE and the cross validation score are 0.023 and 0.022, for the first and second stages, respectively, based on 500 replications. Therefore, the bandwidths chosen based on cross validation are very close to the ones that minimize the RASE. Note that the initial values for the regression coefficients were obtained using ordinary regression and generalized linear regression models for the first and second stages, respectively.

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:

$$W_i(t_{ij}) = \beta_0(t_{ij}) + \beta_1(t_{ij})X_{i1}(t_{ij}) + \beta_2(t_{ij})X_{i2}(t_{ij}) + \beta_3(t_{ij})X_{i3}(t_{ij}) + \varepsilon_{1i}(t_{ij}), \quad (3.3)$$

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 nearly zero 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

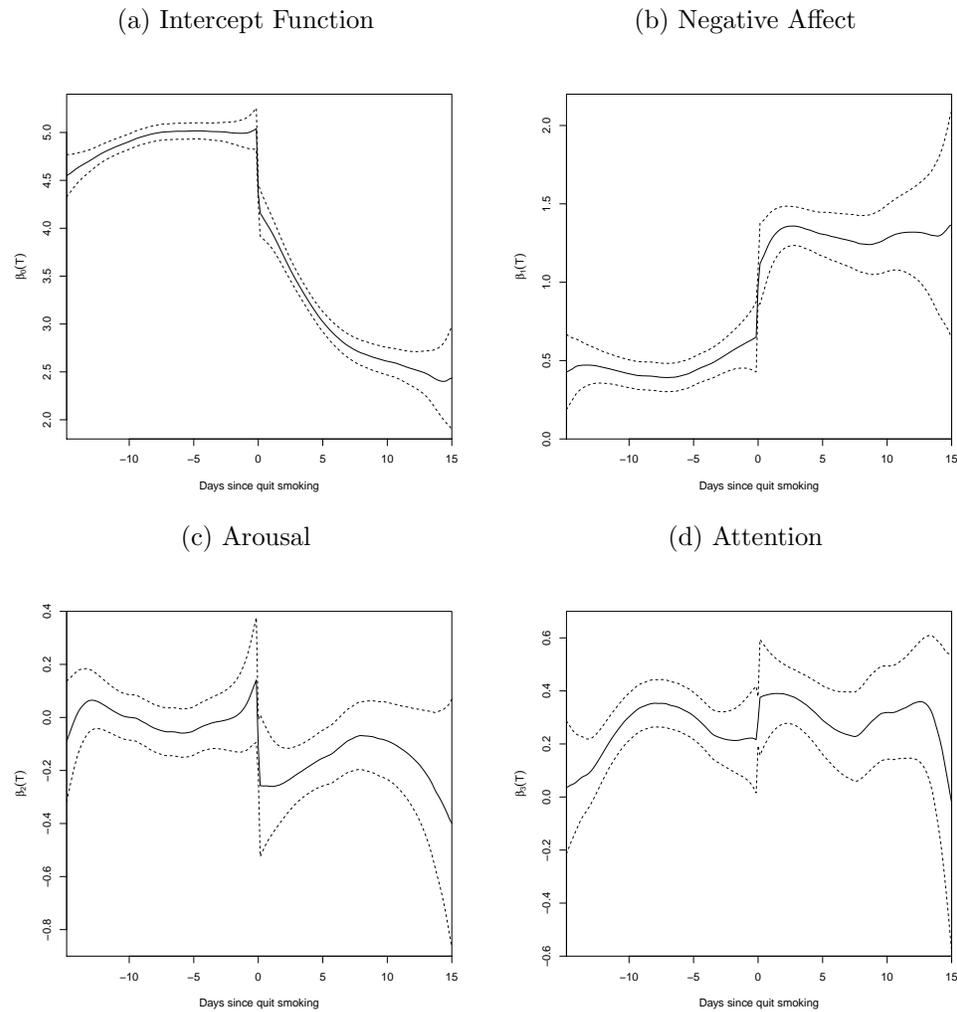


Figure 3.3: Plots of estimated coefficient functions (solid) of the time-varying coefficient model fit to urge to smoke response along with the 95% pointwise asymptotic confidence intervals before and after quitting smoking (dashed). We aligned the data so that all subjects have quit day at day zero.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

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 (3.3), $e_i(t_{ij})$, we fit the following generalized time-varying coefficient model:

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

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 (3.3). As mentioned in Section 2.1, $\tau_1(t) = b(t)/\sqrt{1 + b^2(t)}$ with $b(t) = \alpha_4^*(t)\sigma_1(t)$, $\sigma_1^2(t)$ is the variance of urge to smoke at time t , which is estimated using kernel estimator (2.12) at the bandwidth for the first stage $h = 5.0$, and $\tau_1(t)$ shows the association between urge to smoke and alcohol use at time t . We obtained bootstrap samples by resampling from independent subjects, and repeating our estimation procedure 500 times. Figure 3.4(a) presents the estimated association $\hat{\tau}_1(t)$ along with 2.5 and 97.5 percentiles of 500 bootstrap samples. According to Figure 3.4(a), before the quit day urge to smoke and alcohol use have a positive relationship but after the quit day the relationship 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. Based on the relationship between $\tau_1(t)$ and $\alpha_4^*(t)$, we investigate the significance of the association using the confidence intervals for $\alpha_4^*(t)$. Figure 3.4(b) depicts the estimated regression coefficient $\hat{\alpha}_4^*(t)$ along with its confidence intervals. Figure 3.4(b) demonstrates that the association between alcohol use and urge to smoke is time-varying and 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. In this analysis, the selected bandwidth $h = 5.0$ for both stages. 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

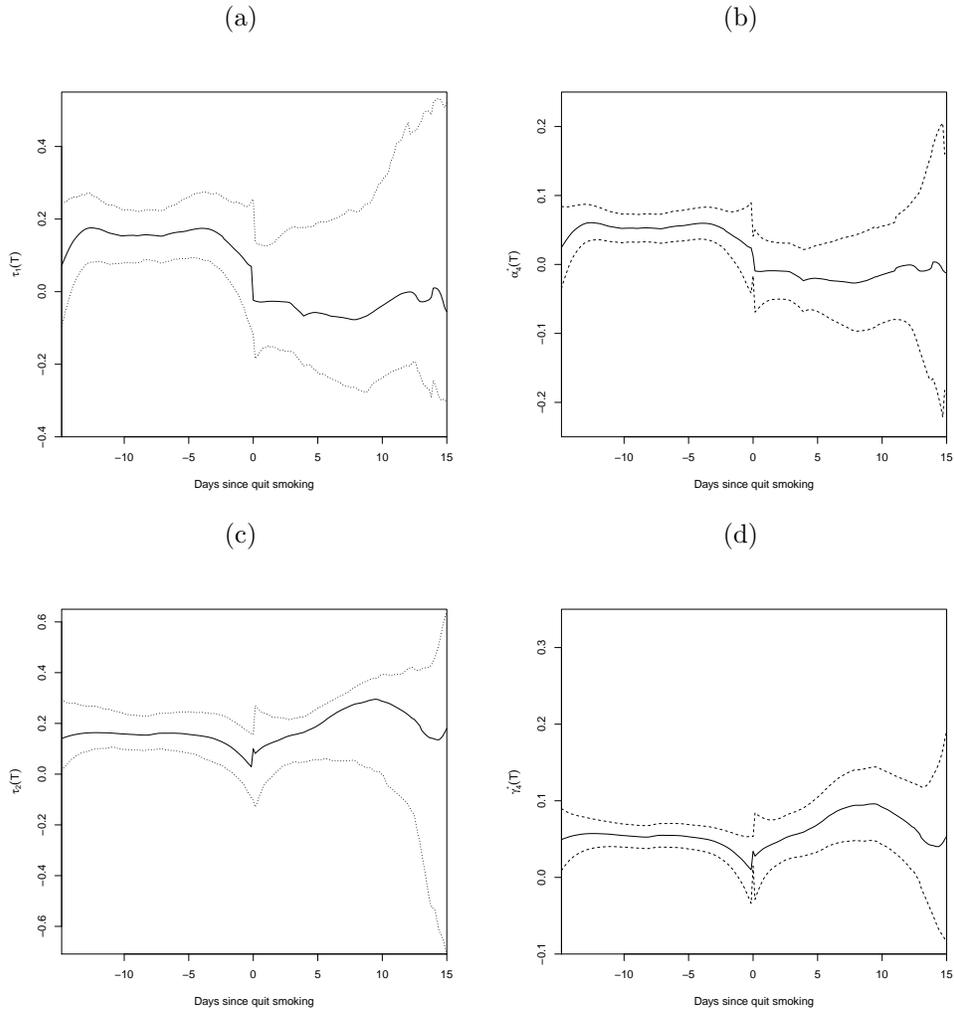


Figure 3.4: Estimated time-varying associations along with 2.5 and 97.5 percentiles of 500 bootstrap samples (a) alcohol versus urge to smoke and (c) presence of other smokers versus urge to smoke. Estimated coefficient functions (solid) of the generalized time-varying coefficient model fit along with the 95% pointwise asymptotic confidence intervals (dashed) (b) alcohol use analysis and (d) presence of other smokers analysis.

(3.3). We observe the same trends in the regression 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 (3.3) and fit the following generalized time-varying coefficient model:

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

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 (3.3). In this model, $\tau_2(t) = b(t)/\sqrt{1 + b^2(t)}$ with $b(t) = \gamma_4^*(t)\sigma_1(t)$, and $\tau_2(t)$ shows the association between urge to smoke and presence of other smokers at time t . Similar to our analysis for estimating $\tau_1(t)$, we estimate $\sigma_1(t)$ using the kernel estimator (2.12) at the bandwidth for the first stage $h = 5.0$. Figure 3.4(c) demonstrates the estimated association $\hat{\tau}_2(t)$ along with 2.5 and 97.5 percentiles of 500 bootstrap samples. Figure 3.4(c) indicates that the association between the presence of other smokers and urge to smoke is almost always positive. In other words, the presence of other smokers is always associated with increased urge to smoke. As indicated by the relationship between $\tau_2(t)$ and $\gamma_4^*(t)$, we can investigate the significance of the association using the confidence intervals for $\gamma_4^*(t)$. Figure 3.4(d) shows the estimated regression coefficient $\hat{\gamma}_4^*(t)$ along with its confidence intervals. According to Figure 3.4(d) the relationship between urge to smoke and presence of other smokers is time-invariant and significant until around day 13 after the quit day. After this day, it appears to become insignificant, which might be due to not having sufficient number of observations.

4. Discussion

In this article, we proposed a joint modeling methodology for estimating the time-varying association between longitudinal binary and continuous responses. We also developed a two-stage estimation procedure based on local linear regression, and derived standard error formulas for our estimators in both stages. By using a simulation study, we showed that our procedure works well on estimating

both the time-varying relationship between longitudinal binary and continuous responses and the true standard errors of the estimators. We applied our method to a dataset of lapsed participants in a smoking cessation study and gained insight regarding two relationships: urge to smoke and alcohol use, and urge to smoke and presence of other smokers.

Note that we are aware that in practice an association between the binary and continuous responses measured at different time points may exist; however, we did not model this association and showed in Section 2.3 that this association does not affect the asymptotic behavior of the proposed estimates in either stages of the estimation procedure.

Acknowledgement

Li's research was supported by National Institute on Drug Abuse (NIDA) grant P50-DA10075, and National Cancer Institute (NCI) grant R01 CA168676. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA or the NIH.

References

- 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**, 961–976.
- Cai, Z., Fan, J., and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association* **95**, 888–902.
- 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**, 651–658.
- Cox, D. R. and Wermuth, N. (1992). Response models for mixed binary and quantitative variables. *Biometrika* **79**, 441–461.
- Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society. Series B* **62**, 355–366.

- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
- Fan, J., Huang, T., and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of American Statistical Association* **102**, 632–641.
- 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**, 845–852.
- Gueorguieva, R. (2001). A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling: An International Journal* **1**, 177–193.
- 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**, 1102–1112.
- Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B* **55**, 757–796.
- Hodges, J. and Reich, B. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician* **64**, 325–334.
- 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**, 809–822.
- 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.
- Kahler, C. W., Spillane, N. S., and Metrik, J. (2010). Alcohol use and initial smoking lapses among heavy drinkers in smoking cessation treatment. *Nicotine & Tobacco Research* **12**, 781–785.
- 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**, 429–438.

- McDermut, W. and Haaga, D. A. (1998). Effect of stage of change on cue reactivity in continuing smokers. *Experimental and Clinical Psychopharmacology* **6**, 316–324.
- Regan, M. M. and Catalano, P. J. (1999). Likelihood models for clustered binary and continuous outcomes: Application to developmental toxicology. *Biometrics* **55**, 760–768.
- Roy, J. and Lin, X. (2000). Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics* **56**, 1047–1054.
- 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**, 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., Balabanis, M. H., Gwaltney, C. J., Paty, J. A., Gnys, M., Kassel, J. D., Hickcox, M., and Paton, S. M. (2007). Prediction of lapse from associations between smoking and situational antecedents assessed by ecological momentary assessment. *Drug and Alcohol Dependence* **91**, 159–168.
- 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**, 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**, 993–1002.
- Verbeke, G., Molenberghs, G., and Rizopoulos, D. (2010). Random effects models for longitudinal data. In van Montfort, K., Oud, J., and Satorra, A., editors, *Longitudinal Research with Latent Variables*, pages 37–96. Springer-Verlag, Berlin.

- Warren, C. A. and McDonough, B. E. (1999). Event-related brain potentials as indicators of smoking cue-reactivity. *Clinical Neurophysiology* **110**, 1570–1584.
- 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**, 1388–1402.
- Yao, W. and Li, R. (2013). New local estimation procedure for nonparametric regression function of longitudinal data. *Journal of the Royal Statistical Society: Series B* **75**, 123–138.
- Zhang, W. and Lee, S.-Y. (2000). Variable bandwidth selection in varying-coefficient models. *Journal of Multivariate Analysis* **74**, 116–134.

Division of Biostatistics and Department of Statistics, Istanbul Medeniyet University, Istanbul 34700, Turkey.

E-mail: esra.kurum@medeniyet.edu.tr

Phone: 090 533 513 5400

Department of Statistics and The Methodology Center, The Pennsylvania State University, University Park, PA 16802-2111, U.S.A.

E-mail: rzli@psu.edu

Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A.

E-mail: Shiffman@pitt.edu

Department of Statistics, University of California, Riverside, California 92521, U.S.A.

E-mail: weixin.yao@ucr.edu