DEPARTMENT OF STATISTICS
The Pennsylvania State University
University Park, PA 16802 U.S.A.

TECHNICAL REPORTS AND PREPRINTS
Number 14-02: August 2014

Time-Varying Copula Models for Longitudinal Data

Esra Kürüm\textsuperscript{a}, John Hughes\textsuperscript{b,\ast}, Runze Li\textsuperscript{c}, Saul Shiffman\textsuperscript{d}

\textsuperscript{a}Division of Biostatistics and Department of Statistics, Istanbul Medeniyet University, Istanbul, Turkey
\textsuperscript{b}Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455, USA
\textsuperscript{c}Department of Statistics and The Methodology Center, The Pennsylvania State University, University Park, PA 16802, USA
\textsuperscript{d}Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260, USA

\ast Corresponding author. Drs. Kürüm and Hughes contributed equally to this work.
Time-varying copula models for longitudinal data

Esra Kürüm\textsuperscript{a}, John Hughes\textsuperscript{b,c}, Runze Li\textsuperscript{d}, Saul Shiffman\textsuperscript{d}

\textsuperscript{a}Department of Biostatistics and Department of Statistics, Istanbul Medipol University, Istanbul, Turkey
\textsuperscript{b}Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455, USA
\textsuperscript{c}Department of Statistics and The Methodology Center, The Pennsylvania State University, University Park, PA 16802, USA
\textsuperscript{d}Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260, USA

Abstract

We propose a copula-based joint modeling framework for mixed longitudinal responses. Our approach permits all model parameters to vary with time, and thus will enable researchers to reveal dynamic response-predictor relationships and response-response associations. We call the new class of models TIMECOP because we model dependence using a time-varying copula. We develop a one-step estimation procedure for the TIMECOP parameter vector, and also describe how to estimate standard errors. We investigate the finite sample performance of our procedure via three simulation studies, one of which shows that our procedure performs well under ignorable missingness. We also illustrate the applicability of our approach by analyzing binary and continuous responses from the Women's Intimacy HIV Study and a smoking cessation program.

Keywords: Bivariad kernel; HIV; Joint model; Local regression; Varying coefficient model

1. Introduction

Analyses of multivariate longitudinal outcomes are now common, but current models for such data cannot reveal the nature of time-varying dependence among the coordinates of a $d$-dimensional response unless $d = 2$ and the two processes in question are Bernoulli and Gaussian [58]. What is needed is a modeling approach that is flexible enough to permit the estimation of time-varying parameters for a response of higher dimension, having coordinate processes of practically any type. In this paper we develop a flexible modeling framework that addresses these concerns.

A number of joint models for longitudinal binary and continuous responses are well established [see, for example, 6, 9, 24, 53, 49, 14, 27, 42, 38]. The

\textsuperscript{a}Corresponding author. Esra Kürüm and Hughes contributed equally to this work.

\textsuperscript{b}Email addresses: esra.kurum@medipol.edu.tr (Esra Kürüm), hughes@umn.edu (John Hughes), rzl@pitt.edu (Runze Li), Shiffman@pitt.edu (Saul Shiffman)

\textit{Preprint submitted to Elsevier} August 30, 2014
main challenge in developing such models is that there is no natural multivariate distribution for responses of mixed type. One way to overcome this problem is to introduce a latent variable underlying the binary response, and assume that the continuous response and the latent variable are jointly Gaussian. The resulting joint distribution can then be factorized, leading to one of two formulations: (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. Some approaches of this sort are not limited to binary and continuous outcomes, but can also handle other types of outcomes and $d > 2$ [see, for example, 53, 14].

A second solution to the above mentioned problem is the joint mixed-effects model [26, 27]. In this approach a random effect is assumed for each outcome, and the outcomes are associated via a joint distribution for the random effects. As pointed out by Verbeke et al. [65], fitting these models becomes computationally burdensome as the number of outcomes increases, and maximum likelihood estimation is possible only when the dimension is low or strong assumptions are made. An example of the latter can be found in Roy and Lin [59], where corresponding random effects for various outcomes are assumed to be perfectly correlated. Another potential drawback of the mixed-effects approach is confounding [31], which may inflate the variance of fixed-effects estimators, preventing the discovery of important response-predictor relationships.

A third solution is to join a set of marginal distributions using a copula. (See Nielsen [45] for an introduction to copulas, and Song et al. [63] and/or Madsen and Fang [43] for information on copula-based regression models.) This is the approach we adopt for the remainder of this article.

In a longitudinal study the relationship between a response and predictors, or the association between a pair of responses, may change over time. The inability of ordinary models to capture these dynamic patterns led Kürüm et al. [38] to develop time-varying models [2] for longitudinal binary and continuous responses. Kürüm et al. [38] adopted the latent variable approach and first type of factorization described above. This implies a two-step estimation procedure. In the second stage of the procedure, the association between the responses takes the form of a time-varying regression coefficient. The advantage of this method is that it allows all parameters, including association parameters, to be time varying. But the method of Kürüm et al. [38] relies on the assumption that the latent variable and the continuous response follow a joint normal distribution. Moreover, their method is limited to estimating the time-varying association between only two longitudinal outcomes.

In this paper we develop a flexible and intuitive framework for modeling multivariate longitudinal data. Although we focus on revealing time-varying dependence relationships, our framework can easily accommodate all manner of time-varying parameters for the coordinate processes: regression coefficients, variances, etc. To achieve this goal, we exploit the modularity of copula-based modeling, which allows us to model the marginal distributions and dependence structure separately before joining them by way of the probability integral transform.
form [35]. In our view, this is perhaps the simplest and most natural way to
construct a multivariate distribution for discrete or mixed responses, and is even
more intuitive when the responses are continuous.

Our main contribution is twofold. First, our approach brings time-varying
parameters to multivariate longitudinal modeling. This will allow researchers to
uncover complex dynamic patterns of dependence and response predictor rela-
tionships. Second, our approach brings arbitrary response type and dimension
greater than two to time-varying joint modeling, i.e., our approach, unlike the
approach of Kürüm et al. [38], is not limited to a binary-continuous response.
Moreover, our model requires neither latent variables nor factorization, and so
does not rely on the assumptions required by Kürüm et al. [38] to estimate
time-varying dependence parameters.

The remainder of the paper is organized as follows. Section 2 describes
our time-varying copula model for longitudinal mixed outcomes; our estimation
procedure, which uses local regression techniques; and bandwidth selection. Sec-
tion 3 assesses the finite sample behavior of our approach via simulation, and
shows that our procedure can handle ignorable missingness. Sections 4 and 5
apply our method to a subset of the WIHS data and to smoking cessation data,
respectively. And Section 6 contains concluding remarks.

2. Time-varying copula models

In longitudinal and ecological studies, response predictor associations may
deviate from time, temperature, or geographical location. Varying coefficient
models, which allow regression parameters to change with some underlying co-

variate(s), were developed to address the inability of ordinary regression models
to capture these dynamic relationships.

Varying coefficient models were introduced by Cleveland et al. [8] and pop-
arIALIZED by Hastie and Tibshirani [38]. A linear varying coefficient model takes the
form

\[ Y = x' \beta(U) + \varepsilon, \]  

where \( Y \) is the response variable, \( x = (x_1, \ldots, x_p)' \) is a vector of predictors,
\( U \) is a scalar covariate, \( \beta(U) = (\beta_1(U), \ldots, \beta_p(U))' \) are unknown coefficient
functions, and \( \varepsilon \) is an error such that \( \text{E}(\varepsilon | x, U) = 0 \). Since varying coefficient
models are local linear models [20], kernel smoothing is a natural approach to
estimation for model (1), and so we use kernel smoothing for the class of models
developed below: time-varying copula models for longitudinal data, TIMECOP
short. The TIMECOP framework permits not only regression coefficients but
all parameters, including dependence parameters, to be time varying.

Suppose we have \( n \) independent subjects. For subject \( i \) we observe the \( d \) -

variate process \( Y_i(t) = (Y_{i1}(t), \ldots, Y_{id}(t))' \) at random times \( t = (t_1, \ldots, t_m)' \).
That is, we observe \( Y_{ij} \equiv Y_i(t_{ij}) = (Y_{ij1}, \ldots, Y_{ijd})' \) \( (j = 1, \ldots, n_i) \).
The number of observations and the observation times may vary from subject
to subject.
We assume that coordinate $k$ of the response has marginal distribution function $F_{k,t}$ and density/mass function $f_{k,t}$, both of which may depend on time-varying parameters $\theta_k(t)$, some of which may be regression coefficients $\beta_k(t)$. For example, in the oft-cited bivariate continuous-binary case we might adopt a Gaussian linear model for one coordinate:

\[ Y_{i1}(t) \sim \mathcal{N}(x_{i1}(t)\beta_1(t), \sigma^2(t)) \]

and a logistic model for the other:

\[ Y_{i2}(t) \sim \mathcal{B}\left(\exp\{-x_{i2}(t)\beta_2(t)\}\right)^{-1}, \]

where $\mathcal{B}(p)$ denotes a Bernoulli random variable with mean $p$, and $x_{i1}(t)$ and $x_{i2}(t)$ are vectors of predictors for subject $i$, measured at time $t$. Although it is often convenient to work within the familiar generalized linear model (GLM) framework, in which case our local model (Section 2.2) is reminiscent of the vector GLM [53], there are of course many other options for the marginal specifications: extreme value distributions, beta regression models, zero-inflated models, skew-normal models, heavy-tailed distributions, etc.

We model dependence using a time-varying $d$-copula $C_{\gamma(t)} \{ u_1(t), \ldots, u_d(t) \}$, where $\gamma(t)$ are copula parameters [45]. A convenient choice is the Gaussian copula [35, 61]

\[ \Phi_{\mathbf{R}(t)}[\Phi^{-1}{\{u_1(t)\}}, \ldots, \Phi^{-1}{\{u_d(t)\}}], \]

where $\Phi_{\mathbf{R}(t)}$ is the cdf of a $d$-variate multinormal random variable with mean vector $\theta$ and correlation matrix $\mathbf{R}(t)$, and $\Phi^{-1}$ is the univariate standard normal quantile function. Other attractive choices are the $t$ copula [13], which can accommodate tail dependence, or the skew $t$ copula [60], which can accommodate tail dependence and asymmetric dependence.

We use a one-step estimation procedure based on optimization of an approximation to the local kernel-weighted log likelihood of $\theta(t) = (\theta_1'(t), \ldots, \theta_d'(t), \gamma'(t))'$. The approximation is based on the distributional transform (explained below) and was first proposed by Kaischke and Pflü [36] for fitting Gaussian copula geostatistical models.

2.1. Likelihood inference for Gaussian copula models

In this subsection we revert temporarily to an ordinary likelihood setting. This will ease notation and allow us to motivate our likelihood approximation as simply and clearly as possible. In Section 2.2 we will reintroduce time varyingness and describe our approach to local likelihood inference for TIMECOP.

Likelihood inference is fairly straightforward for Gaussian copula models with continuous marginals. To see this, first note that the density for the Gaussian $d$-copula is

\[ c_{\mathbf{R}}(u) = \frac{\phi_{\mathbf{R}}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d))}{\prod_{i=1}^d \phi(\Phi^{-1}(u_i))} \propto |\mathbf{R}|^{-1/2} \exp \left\{ -\frac{1}{2} z' (\mathbf{R}^{-1} - 1) z \right\}, \]
where $z = (z_1, \ldots, z_d)' = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d))'$ and $I$ is the $d \times d$ identity matrix. If the desired marginal distributions $F_1, \ldots, F_d$ are continuous also, the likelihood of the parameters $\theta$ given the data $y$ has the form

$$L(\theta \mid y) \propto \prod_{i=1}^{d} f_i(y_i) \prod_{i=1}^{d} F_i(y_i),$$

where $f_i$ is the density function corresponding to $F_i$. This implies the log likelihood

$$\ell(\theta \mid y) = -\frac{1}{2} \log |R| - \frac{1}{2} z'(R^{-1} - I)z + \sum_{i=1}^{d} \log f_i(y_i). \quad (2)$$

where $z_i = \Phi^{-1}(F_i(y_i))$. This log likelihood can be optimized to arrive at the maximum likelihood estimate of $\theta$.

When some of the marginal distributions are discrete, the likelihood does not have the simple form given above because $z_i = \Phi^{-1}(F_i(y_i))$ is not standard normal (since $F_i(y_i)$ is not standard uniform if $F_i$ has jumps). In this case the true likelihood is more complicated [39, 63] and becomes unwieldy as the number of discrete coordinates increases. An appealing alternative to the true likelihood is an approximation based on the distributional transform.

It is well known that if $Y \sim F$ is continuous, $F(Y)$ has a standard uniform distribution. But if $Y$ is discrete, $F(Y)$ tends to be stochastically larger, and $F(Y^-) = \min_{x < y} F(x)$ tends to be stochastically smaller, than a standard uniform random variable. This can be remedied by stochastically “smoothing” $F$ at its jumps. This technique goes at least as far back as Ferguson [22], who used it in connection with hypothesis tests. More recently, the distributional transform has been applied to stochastic ordering [51], conditional value at risk [3], and the extension of limit theorems for the empirical copula process to general distributions [52].

Let $W \sim U(0, 1)$, and suppose that $Y \sim F$ and is independent of $W$. Then the distributional transform

$$G(W, Y) = WF(Y^-) + (1 - W)F(Y)$$

follows a standard uniform distribution, and $F^{-1}\{G(W, Y)\}$ follows the same distribution as $Y$. Taking the expectation of $G(W, Y)$ with respect to $W$ yields the approximation

$$\frac{F(Y^-) + F(Y)}{2} \approx WF(Y^-) + (1 - W)F(Y).$$

To construct the approximate log likelihood, we replace $F_i(y_i)$ in (2) with

$$\frac{F_i(y_i^-) + F_i(y_i)}{2}$$

for each discrete coordinate of the response. Although seemingly crude, this approximation performs quite well for NICECOR—even for binary data, as the simulation results in Section 3 show—and allows us to work with a simple meta-Gaussian likelihood instead of the more cumbersome true likelihood.
2.2. Local likelihood inference for TIMECOP

Now we return to TIMECOP, for which we recommend local likelihood inference. That is, we estimate \( \theta(t) \) at time \( t_0 \) by maximizing the local kernel-weighted log likelihood

\[
\ell(\theta(t_0) \mid T) = \sum_{i=1}^{n} \sum_{j=1}^{m_j} \ell(\theta(t_0) \mid y_{ij}) K\{(t_0 - t_{ij})/h\}/h,
\]

where \( T = (t_1, \ldots, t_n) \), \( \ell(\theta(t_0) \mid y_{ij}) \) is the log likelihood of \( \theta(t_0) \) given the outcomes for subject \( i \) at time point \( t_{ij} \), \( K \) is a kernel, and \( h \) is a bandwidth. For TIMECOP we use the approximate log likelihood described in the previous section. Specifically, if we partition the response vector so that the first \( d_1 \) coordinates are continuous and the remaining coordinates are discrete, we have

\[
\ell(\theta(t_0) \mid y_{ij}) = -\frac{1}{2} \log |R| \quad \text{(3)}
\]

\[
-\frac{1}{2} (z_{ij}^t z_{ij}^t) (R^{-1} - I) (z_{ij}^t z_{ij}^t)',
\]

\[
+ \sum_{k=1}^{d} \log f_{ik}(y_{ik}(t_{ij})),
\]

where

\[
z_{ij} = (\Phi^{-1} [F_{i1}(y_{11}(t_{ij}))], \ldots, \Phi^{-1} [F_{i(k-1)+1}(y_{i(k-1)+1}(t_{ij}))])',
\]

\[
z_{ij}' = (\Phi^{-1} [u_{i(k-1)+1}(t_{ij})], \ldots, \Phi^{-1} [u_{i(k-1)+1}(t_{ij})])'.
\]

The distributional transform approximation enters through computation of the \( u_{ik}(t_{ij}) \) \((k = d_1 + 1, \ldots, d)\):

\[
u_{ik}(t_{ij}) = \frac{F_{ik}(y_{ij}^{-}(t_{ij})) + F_{ik}(y_{ij}(t_{ij}))}{2},
\]

We obtain \( \hat{\theta}(t_0) \) using the quasi-Newton method of Byrd et al. [5] so that estimated dependence and scale parameters can be appropriately constrained.

We used local constant estimation in our simulation studies, i.e., we assumed that \( \theta(t_0) \) is constant on a neighborhood of \( t_0 \). It is straightforward to use a higher-order polynomial approximation to \( \theta(t_0) \), but even a linear approximation—in which one assumes that

\[
\theta(t_0) = \theta(t_0) + \hat{\theta}(t_0)(t_0 - t)
\]

for \( t \) on a neighborhood of \( t_0 \)—increases the computational burden quite a bit while reducing bias only slightly.

In the final step of our procedure we estimate the variance of \( \hat{\theta}(t_0) \). Here we use results obtained by Fan et al. [16]. We begin with the approximate
conditional variance (conditional on $T$), which has a sandwich form:

$$\Sigma(t_0) = \nabla \{ \theta(t_0) \mid T \}$$

$$= \kappa(t_0) \mathcal{H}^{-1}(t_0) \mathcal{J}(t_0) \mathcal{H}^{-1}(t_0)$$

$$= n(t_0) [\mathcal{E}(\theta(t_0) \mid T)]^{-1} \nabla \{ \mathcal{E}(\theta(t_0) \mid T) \} [\mathcal{E}(\theta(t_0) \mid T)]^{-1},$$

where $\kappa(t_0) = \sum_{i=1}^{n} \sum_{j=1}^{n} K^2 \{(t_0 - t_{ij})/h\}/h^2$ and $\mathcal{H}(t_0)$ is the Hessian matrix, which can be estimated by $\overline{\mathcal{E}}(\theta(t_0) \mid T)$ as a side effect of optimization. The variance of the score, $\mathcal{J}(t_0)$, can be estimated by

$$\nabla \{ \mathcal{E}(\theta(t_0) \mid y_{ij}) K \{(t_0 - t_{ij})/h\}/h \} / \sum_{i=1}^{n} \sum_{j=1}^{n} K \{(t_0 - t_{ij})/h\}/h,$$

where $\nabla$ denotes the gradient.

In view of asymptotic normality [28, 12] and using $\hat{\Sigma}(t_0)$, we construct a pointwise $(1 - \alpha)100\%$ confidence interval for the $\alpha$th element of $\theta(t_0)$ as

$$\hat{\theta}_\alpha(t_0) + \Phi^{-1}(1 - \alpha/2) \sqrt{\hat{\Sigma}_\alpha(t_0)},$$

where $\hat{\Sigma}_\alpha(t_0)$ is the $\alpha$th diagonal element of $\hat{\Sigma}(t_0)$.

Note that our procedure does not account for intra-subject dependence, but theory suggests that intra-subject dependence can be ignored (for estimation of $\theta(t_0)$) when the number of subjects is sufficiently large. As shown in Lin and Carroll [30], the method of kernel generalized estimation equations (kernel GEE) yields a root-$n$ consistent estimator regardless of the working correlation structure. Furthermore, kernel GEE with working independence correlation matrix yields the most efficient estimator for the nonparametric regression function in a longitudinal setting. The procedure proposed here shares the spirit of kernel GEE, and so we suspect that our estimator is root-$n$ consistent and is likely the most efficient. Theoretical justification is beyond the scope of this paper and must be left to a future investigation.

As for inference, Fan et al. [16] noted that the intervals given by (5) might be too narrow for some datasets (due to intra-subject dependence), in which case one can get better coverage by using the wider intervals obtained from an undersmoothed fit. Undersmoothing is effective here because any two kernel-weighted intra-subject observations are nearly uncorrelated when $h$ is sufficiently small and the serial dependence is short- or medium-range [17]. See Section 3 for details regarding appropriate undersmoothing for TIMEGEE.

Practical application of our approach depends on a selection of a suitable bandwidth. For this we recommend a form of cross-validation proposed by Fan and Zhang [20]. We leave out a single subject at a time rather than a single observation, since the latter approach is inappropriate when there is intra-subject dependence [32]. After removing the $i$th subject, we estimate $\theta(\cdot)$ based on the remaining subjects. After doing this for each of the $m$ subjects, we combine the
results to form the cross-validation score

\[ CV(h) = - \sum_{i=1}^{m} \sum_{j=1}^{n_i} \ell \{ \hat{\theta}^i \{ t_{ij} \} \mid y_{ij}, h \} , \]

where \( \hat{\theta}^i \{ t_{ij} \} \) is the leave-i-out estimate for time \( t_{ij} \). We compute the cross-validation score for a range of bandwidths and select the bandwidth that minimizes the score.

We recommend that a bimodal kernel \([1]\) be used because doing so leads to a more accurate estimate in the presence of intra-subject dependence. Informally, a bimodal kernel removes serial dependence by down weighting observations that are very close to \( t_0 \). This prevents undersmoothing by preventing the estimation procedure from “mistaking” local similarity for structure that should be fitted. We use the member of the so-called \( \varepsilon \)-optimal class of bimodal kernels recommended by De Bruijne et al. \([1]\). Specifically, we use

\[ K_\varepsilon(u) = \frac{4}{4 - 3\varepsilon - \varepsilon^2} \begin{cases} \frac{3}{4} (1 - \varepsilon^2) 1 \{|u| \leq 1\} & \text{if } |u| \geq \varepsilon \\ \frac{3}{4} \left( 1 + \varepsilon^2 \right) |u| & \text{if } |u| < \varepsilon \end{cases} \]

with \( \varepsilon = 0.1 \), where \( 1 \{-\} \) denotes the indicator function.

3. Simulated application

We investigated the finite sample performance of our estimator using a simulation study designed to mimic the WHIS data that we analyze in Section 4. The response in our study was binary–continuous. Specifically, we let \( Y_1(t) \) be a Gaussian process with mean \( x_{1i}(t) \beta_1(t) \) and variance \( \sigma_1^2(t) \), and \( Y_2(t) \) be a Bernoulli process with mean

\[ 1 + \exp \left\{ -x_{2i}(t) \beta_2(t) \right\} \]

as described above in Section 2. For the time-varying coefficients \( \beta_1(t) \) and \( \beta_2(t) \), we used

\[ \beta_1(t) = (\beta_{10}(t), \beta_{12}(t), \beta_{13}(t))' = 0.2 \left( \cos 2\pi t, \sin 2\pi t, -\sin 2\pi t, 1 + \sin 2\pi t \right)' \]
\[ \beta_2(t) = (\beta_{20}(t), \beta_{21}(t))' = 0.2 \left( \sin 2\pi t, 1 + \cos 2\pi t \right)' \]

with \( \beta_{41}(t) \) an intercept, and \( \beta_{41}(t), \beta_{42}(t) \) and \( \beta_{43}(t) \) slopes. We simulated the predictors from the standard normal distribution. \( Y_1(t) \)’s time-varying standard deviation was \( \sigma(t) = 0.8 + 0.2 \sin 2\pi t \).

Although it does not enter into our estimation procedure, we simulated both processes with first-order continuous autoregressive, i.e., CAR(1), dependence:

\[ \rho_1(s, t) = 0.7^{\mid s-t \mid} \]
\[ \rho_2(s, t) = 0.7^{\mid s-t \mid} , \]
for any two times \( s \) and \( t \) in the unit interval. These functions correspond to consequential but short-range dependence. Specifically, if we define the effective range to be the distance \(|s - t|\) at which the correlation between two observations has dropped to 0.05, these functions have an effective range of approximately 0.2.

We simulated a single dataset as follows. For subject \( i \):

1. let \( n_i = 10 \);
2. simulate \( n_i \) measurement times \( t_i = (t_{i1}, \ldots, t_{in_i})' \) from the standard uniform distribution;
3. for \( j = 1, \ldots, n_i \), construct the \( 2 \times 2 \) correlation matrix with off-diagonal entries \( \rho(t_{ij}) \), and use the Cholesky root to impose the correlation on \((W_{1i}(t_{ij}), W_{2i}(t_{ij}))' \sim N(0, I)\);
4. construct the \( n_i \times n_i \) correlation matrices \( R_1(t_i) \) and \( R_2(t_i) \) according to the CAR(1) specification given above, and use the corresponding Cholesky roots to impose the correlation structures on \((W_{1i}(t_{i1}), \ldots, W_{1i}(t_{in_i}))' \text{ and } (W_{2i}(t_{i1}), \ldots, W_{2i}(t_{in_i}))' \);
5. form \( Z = (W_{1i}(t_{i1}), W_{1i}(t_{i1}), \ldots, W_{1i}(t_{in_i}), W_{2i}(t_{in_i}))' \);
6. apply the probability integral transform (PIT) to each element of \( Z \) to arrive at:

\[
U = (U_{1i}(t_{i1}), U_{2i}(t_{i1}), \ldots, U_{1i}(t_{in_i}), U_{2i}(t_{in_i}))'
\]

\[
= (\Phi^{-1}\{Z_{1i}(t_{i1})\}, \Phi^{-1}\{Z_{1i}(t_{i1})\}, \ldots, \Phi^{-1}\{Z_{1i}(t_{in_i})\}, \Phi^{-1}\{Z_{2i}(t_{in_i})\})'.
\]

the elements of which are uniformly distributed; and
7. apply the inverse PIT to \( U \) to produce the response:

\[
(Y_{1i}(t_{ij}), Y_{2i}(t_{ij}))' = (F_{1i}^{-1}\{U_{1i}(t_{ij})\}, F_{2i}^{-1}\{U_{2i}(t_{ij})\})' \quad (j = 1, \ldots, n_i).
\]

where \( F_{1i}^{-1} \) and \( F_{2i}^{-1} \) are the inverse cdfs corresponding to the desired Gaussian and Bernoulli marginal distributions described above.

We used a pilot study to choose several bandwidths. For each bandwidth we simulated and fitted 500 datasets, each having \( m = 300 \) subjects, and estimated \( \beta(t) = (\beta_1(t), \beta_2(t))' \) at 200 grid points equally spaced over the unit interval. Figures 1 and 2 show the results for \( h_0 = 0.1 \), the bandwidth that minimized the cross-validation score described in Section 2.

Our procedure performed well overall for this scenario with respect to bias, as the biases are generally small. The standard errors for the selected bandwidth of \( h_0 = 0.1 \) were very accurate for the slope functions \( \beta_1(t) \), \( \beta_2(t) \), \( \beta_3(t) \), and \( \beta_4(t) \), and so the coverage rates for these functions were very close to the desired 95%. For the other parameters, especially \( \sigma(t) \) and \( \rho(t) \), the procedure tended to yield optimistic confidence intervals. We remedied this by using a smaller bandwidth for variance estimation. Specifically, we used \( h_1 \) in \( O(n^{-1/4}) \), where \( n = \sum n_i \), since the asymptotically optimal bandwidth is in \( O(n^{-1/5}) \). For our simulation scenario, this leads to \( h_1 = 0.067 \). We see from the plots in Figures 1 and 2 that this bandwidth yielded accurate confidence intervals.
Figure 1: Part I of the results from our simulation study. Each row shows three plots for a given time-varying parameter. The first plot shows the true function (solid) and the empirical line of our estimator (dotted). The second plot shows the empirical pointwise 95% confidence band (solid) and the union theoretical pointwise 95% confidence band (dotted). The third plot shows the desired coverage rate (solid) and the empirical pointwise coverage rates (dotted).
Figure 2: Part 2 of the results from our simulation study. Each row shows three plots for a given time-varying parameter. The first plot shows the true function (solid) and the empirical bias of our estimator (dotted). The second plot shows the empirical pointwise 95% confidence band (solid) and the mean theoretical pointwise 95% confidence band (dotted). The third plot shows the desired coverage rate (solid) and the empirical pointwise coverage rates (dotted).
It is of interest to observe the performance of our methodology in a missing-at-random (MAR) scenario [41] since (1) missingness of this type is common in longitudinal studies, and (2) our procedure employs a likelihood approximation. As Little and Rubin [41] suggest, an MAR mechanism should depend only on predictors. Therefore, we designed our MAR study as follows: for any subject and any time point, if the value of the second predictor for the continuous response is greater than some cutoff value, delete that observation. Our goal was to create approximately 15% missingness in each simulated data set. Since the second predictor is standard normal, a cutoff value of 1.03 allowed us to achieve our target rate. Figure 3 shows selected results under this missingness scenario for the same bandwidth that was used in the first study. The plots show that our procedure performed comparably for the two studies, which suggests that our approach can handle ignorable missingness.

Our approach performs well with considerably fewer subjects if the binary data are replaced by, say, count data. To demonstrate this we present selected results from a second simulation study. For the second study we simulated a bivariate process for 100 subjects. The first coordinate was Gaussian and identical to the process used in the first study. The second coordinate was Poisson with mean $\exp(x_2(t)\beta_2(t))$ and the same serial dependence as the Bernoulli process from the first study. And the third coordinate was Beta$(\alpha(t), 2)$ with $\alpha(t) = 5 + 0.2 \sin 2\pi t$ and serial dependence $\rho_3(s, t) = 1 - \exp(-d)$. The three dependence functions for the joint process at time $t$ were

\begin{align*}
\rho_{12}(t) &= 0.2 + 0.15 \sin 2\pi t & \text{Gaussian-Poisson} \\
\rho_{13}(t) &= 0.2 + 0.15 \cos 2\pi t & \text{Gaussian-Beta} \\
\rho_{23}(t) &= 0.3 + 0.15 \sin 2\pi t & \text{Poisson-Beta.}
\end{align*}

The results are shown in Figure 4. Note that we once again undersmoothed to obtain accurate confidence intervals for the non-slope parameter functions.

Additional simulation studies suggest that TIMECOP performs well for several, or even many, outcomes in realistic scenarios similar to those considered above, i.e., for at least 100 subjects and at least ten observations per subject, say.

4. Application to HIV data

In this section we apply our proposed methodology to data from the Women's Interagency HIV Study (WHIS). These data contain information on 372 women recruited between 1994 and 1995 from HIV primary care clinics, research programs, community outreach sites, women's support groups, drug rehabilitation programs, and HIV testing sites in Chicago, Los Angeles, New York City, San Francisco, and Washington, DC. Participants were evaluated at WHIS sites every six months with an extensive interview that included physical and oral examinations, blood and gynecological specimen collection, and collection of information regarding participants' daily activities (such as their sexual behaviors and tobacco use). Our analysis is restricted to 292 participants who were HIV
Figure 3: Selected results from our simulation study with MAR. Each row shows three plots for a given time-varying parameter. The first plot shows the true function (solid) and the empirical bias of our estimator (dotted). The second plot shows the empirical pointwise 95% confidence band (solid) and the mean theoretical pointwise 95% confidence band (dotted). The third plot shows the desired coverage rate (solid) and the empirical pointwise coverage rates (dotted).
Figure 4: Selected results from our trivariate simulation study. Each row shows three plots for a given time-varying parameter. The first plot shows the true function (solid) and the empirical bias of our estimator (dotted). The second plot shows the empirical pointwise 95% confidence band (solid) and the mean theoretical pointwise 95% confidence band (dotted). The third plot shows the desired coverage rate (solid) and the empirical pointwise coverage rates (dotted).
positive. Among these subjects, 26% self-identified as Latina or Hispanic, 45% of the women were of African-American non-Hispanic origin, and 12% were of white non-Hispanic origin. Sixty-six percent of the participants were smokers, and, while taking part in the study, 8.3% of the smokers quit smoking while 8.1% of the non-smokers started smoking. Although our data set contains follow-up information on women aged 25–55 until 2006, many participants failed to attend some of their scheduled visits, which led to unequal numbers of measurements and different measurement times. The number of observations for each subject varies from one to eight.

It is known that cigarette smoking has effects on the immune system [23, 29, 33, 25], but it is not yet clear whether any of these effects influence the progression of HIV. Burns et al. [4] analyzed data on a cohort of 3,221 HIV-seropositive men and women enrolled in the Terry Beirn Community Programs for Clinical Research on AIDS. They used proportional hazards regression analysis to assess the differences between never, former, and current cigarette smokers in terms of clinical outcomes, and found no association between cigarette smoking and the overall risk of disease progression or death. Similarly, Gali et al. [25] used Kaplan-Meier analysis and multivariate Cox regression models to investigate the effect of cigarette smoking on the development of AIDS in the Multicenter AIDS Cohort Study of homosexual men. Their analysis revealed that smoking was not significantly associated with progression to AIDS. However, Nieman et al. [47] found that in a case series of 84 individuals, smokers progressed to AIDS more rapidly than nonsmokers. Their analysis used life tables and compared the median time to develop AIDS for smokers and nonsmokers.

Given these inconsistent findings, our primary interest is to investigate this matter further by applying our model to two WIHS variables: CD4 cell percentage (a measure of HIV progression) and smoking status. Our analysis will differ from the above mentioned analyses in two important ways. First, our analysis will not employ survival methods, for our data are not censored, and we do not wish to define a patient lifetime that ends when the patient has progressed to AIDS. We wish to explore the dynamics of the relationship between CD4 cell percentage and smoking status throughout the study. Second, some of the above mentioned analyses excluded patients who altered their smoking behavior during the study. It is not unusual for behaviors to change during a longitudinal study, and ignoring these changes may lead to biased results.

We will treat CD4 percentage and smoking status as response variables so that we can estimate the time-varying partial association [7] between them, i.e., we aim to reveal the association between CD4 percentage and smoking status conditional on predictors of interest. We could regress one of CD4 percentage and smoking status on the other, but it is not obvious how to assign the role of response or predictor to either variable. (In HIV studies it is customary to regress CD4 cell percentage on smoking status, assuming that smoking leads to a change in CD4 cell percentage. But Mamara et al. [44] showed that the percentage of smokers among HIV patients is higher than the percentage of smokers in the general adult population. In addition, HIV patients tend to have a pessimistic view of their survival, which might result in lack of motivation to
quit smoking or even an increase in cigarette consumption. These results suggest that being an HIV patient could be predictive of smoking, which implies that one should regress smoking status on CD4 cell percentage.) And a regression model would not allow us to reveal the nature of this association having controlled for the predictors, such as depression, nor would it allow us to reveal response-predictor relationships for both variables. We could of course achieve the latter goal by fitting two univariate regression models, but this would provide no information about the association, which is of interest to us. Moreover, joint modeling can lead to considerably more precise estimators [26, 62]. This gain in efficiency grows as the strength of dependence increases, especially for smaller samples.

For our analysis we chose predictors based on the HIV literature [66, 48] and exploratory analyses. For the continuous response we used baseline CD4 cell percentage (measured at the first visit), number of sexual partners, hematoctrit value (the volume percentage of red cells in the blood), mean corpuscular volume (a measure of average red blood cell size), platelet count, and Center for Epidemiologic Studies Depression (CESD) scale score. For the binary response we used the CESD scale score and race. All predictors save race are continuous, and we centered them. The race variable originally had five levels: African American, white, Asian/Pacific Islander, native American/Alaskan native, and other. However, our subset of the data had just two participants in each of the Asian/Pacific Islander and native American/Alaskan native categories. Hence, we recategorized race into three levels: African American, white, and other.

To minimize modeling bias we assumed a maximally flexible model, which is to say that we permitted all parameters, including the dependence parameter \( \rho(t) \), to be time varying. If the confidence bands for some parameters suggest that these parameters may be constant with respect to time, our methodology can be used to fit a semivarying model [67].

We assume that CD4 cell percentage, \( Y_{12}(t) \), is a Gaussian process with mean \( x_{11}(t) \beta(t) \) and variance \( \sigma^2(t) \), where

\[
\beta(t) = (\beta_{10}(t), \beta_{11}(t), \beta_{12}(t), \beta_{13}(t), \beta_{14}(t), \beta_{15}(t), \beta_{16}(t))^T
\]

and \( x_{11}(t) = (1, x_{111}(t), x_{112}(t), x_{113}(t), x_{114}(t), x_{115}(t), x_{116}(t)) \) with

- \( x_{111}(t) \): the baseline CD4 (BaseCD4) cell percentage of subject \( i \) at the first visit,
- \( x_{112}(t) \): the number of sexual partners (PART) of subject \( i \) at time \( t \),
- \( x_{113}(t) \): the hematoctrit (HCV) value of subject \( i \) at time \( t \),
- \( x_{114}(t) \): the mean corpuscular volume (MCV) of subject \( i \) at time \( t \),
- \( x_{115}(t) \): the platelet count (PLAT) of subject \( i \) at time \( t \),
- \( x_{116}(t) \): the CESD scale score of subject \( i \) at time \( t \).

We assume that smoking status, \( Y_{13}(t) \), is a Bernoulli process with mean

\[
[1 + \exp(-x_{13}(t) \beta(t))]^{-1},
\]
where \( \beta_2(t) = (\beta_{21}(t), \beta_{22}(t), \beta_{23}(t))^\top \) and \( x_{i2}(t) = (1, x_{i21}(t), x_{i22}(t), x_{i23}(t))^\top \) with

\( x_{i21}(t) \): the CESD scale score of subject \( i \) at time \( t \),
\( x_{i22}(t) \): the first dummy variable for race (RACE 1) \( (x_{i22}(t) = 1 \text{ if subject } i \text{ is African American}) \),
\( x_{i23}(t) \): the second dummy variable for race (RACE 2) \( (x_{i23}(t) = 1 \text{ if subject } i \text{ is white}) \).

We used the \( K_{0.4} \) bimodal kernel and chose a bandwidth of \( h = 14 \) using the cross-validation procedure described in Section 2. Note that the time covariate in this study is the age of the participant. The estimated time-varying regression coefficient functions and variance for the continuous response (CD4 cell percentage) are shown in Figure 5.

- From panel (a) we see that the intercept function is time varying and increases with age.
- The plot in panel (b) suggests that the effect for baseline CD4 is time varying and decreases with age. Moreover, the effect is always significant and positive for ages between 25 and 55.
- The confidence band in panel (c) suggests that the coefficient for number of sexual partners may be time invariant. And the effect is significantly different from zero only between ages 43 and 55.
- The plot in panel (d) suggests that the effect of hematocrit may be time invariant, but the effect is significant and positive for ages between 28 and 50.
- According to panel (e), the effect of mean corpuscular volume may be time invariant, but the effect is significant and positive after age 26.
- Panel (f) shows that the effect of platelet count is significant and positive after age 29. But the confidence band is too wide to support the conclusion that the effect is time varying.
- From panel (g) we see that the effect of CESD score is always significant and negative, i.e., depression is associated with a lower CD4 cell percentage. The effect may be constant with respect to time, however.
- The final panel (just barely) allows us to conclude that the variance of CD4 percentage is time varying and increases with age.

Figure 6 shows the estimated time-varying regression coefficient functions for the binary response (smoking status) along with the estimated time-varying association, \( \rho(t) \), between CD4 cell percentage and smoking status.

- Panel (a) shows that the intercept function is time varying and increases with age.
Figure 5: The results of our data analysis for the continuous response, CD4 cell percentage. For each panel, the solid curve shows the estimate, the dashed curves show the estimated 95% pointwise confidence band, and the dotted line marks zero.
• Panel (b) suggests that the coefficient for CESD score is time varying. The coefficient is significant until age 50 and decreases with age. We see that the effect is positive, which implies an association between depression and smoking. The association is evidently weaker for older patients.

• In panel (c) the upper and lower three-curve groups are the estimates and confidence bands for the RACE 1 and RACE 2 variables, respectively. The confidence bands reveal that the coefficient for RACE 1 is time varying while the coefficient for RACE 2 may be time invariant. We see that the coefficient for RACE 1 is always significant and positive, and until age 45 is greater than the coefficient for RACE 2. That is, African Americans have higher odds of smoking than do patients of other races. After age 45 the confidence bands for RACE 1 and RACE 2 overlap, which indicates that the difference between the two groups becomes insignificant. We also observe that RACE 2 is not a significant predictor of smoking.

Above, we mentioned that joint modeling can result in more precise estimation of marginal parameters. To demonstrate this for the WHS data, we fitted a univariate time-varying model for each of the outcomes. In each univariate model, the other response was included as an additional predictor. Figure 7 compares standard errors for the joint model to those for the univariate models, for selected parameters. We see that joint modeling leads to considerable efficiency gains near the endpoints. The difference is negligible between ages 35 and 45, which is not surprising: approximately 55% of our data were observed in this time interval, and, as Gueorguieva and Agresti [27] argue, for larger sample sizes the efficiency gained through joint modeling is less pronounced.

Finally, let us interpret the last panel of Figure 6, which shows the estimated time-varying partial association. Judging from the confidence band in this plot, we do not have sufficient evidence to conclude that ρ(t) is time varying. But we can conclude that the partial association is always significant and negative, i.e., for women enrolled in the WHS, decreased CD4 cell percentage is partially associated with smoking. Although we studied women only, this finding provides evidence that smokers progress to AIDS more rapidly than nonsmokers. In addition to its negative association with CD4 percentage, smoking is known to decrease the adherence to highly active antiretroviral therapy [21]. Smoking also poses additional threats to HIV-positive patients, such as pulmonary-related complications (pneumonia, asthma, and chronic obstructive pulmonary disease) and increased incidence of opportunistic infections [1, 10, 37]. Therefore, the findings of our study and others suggest that smoking cessation counseling is a necessary component of any program that seeks to enhance quality of life and disease management for HIV patients. Nwana et al. [46] provided a review of existing cessation techniques for HIV patients, and also suggested ways to improve research studies so that more effective cessation treatments can be discovered.

Since some of the parameters may be constant with respect to time, a semivarying model is probably the most appropriate model for these data. Although
Figure 6: The results of our data analysis for the binary response (smoking status), and the estimated time-varying association between smoking status and CD4 cell percentage. For each panel, the solid curve shows the estimate, the dashed (or dashed-dotted) curves show the estimated 95% pointwise confidence band, and the dotted line marks zero.
Figure 7: Comparison of standard errors for our joint analysis and separate univariate analyses of CD4 cell percentage and smoking status. In each univariate analysis, the other outcome was included as an additional predictor. Panels (a), (b), and (c) show results for CD4 cell percentage, and panels (d) and (e) show results for smoking status. For each panel, the solid curve and dashed-dotted curve show the standard errors for the joint and univariate models, respectively.
our methods can be adapted to the semivarying setting, it seems clear that fitting such a model to the WHIS data would not result in substantive changes to our conclusions.

5. Application to smoking cessation data

We will also apply TIMECP to binary and continuous responses from a smoking cessation study [56, 57]. The data were collected using hand-held palm-top computers that prompted each participant at random times. When prompted, the subjects recorded their answers to a series of questions about their current setting and activities as well as current mood and urge to smoke. The data collection process is described below.

First, the subjects were monitored for a two-week interval during which they engaged in their ordinary smoking behavior. They were asked to record all their smoking occasions during this period, and to respond to the random assessment prompts. Patients were instructed to quit smoking at the end of this two-week period. When a patient had abstained for 24 hours, the current day was recorded as that patient’s quit day. After the subjects quit, they were required to continue responding to the random assessment prompts and to record any episodes of smoking (lapses) or strong temptations. Although all subjects were instructed to quit on a certain date, different subjects had different quit days, and the prompts were random. Thus the subjects have unequal numbers of measurements and different measurement times.

According to a 2004 report by the U.S. Department of Health and Human Services, cigarette smoking is one of the leading preventable causes of several diseases including coronary heart disease, acute myeloid leukemia, and bladder, esophageal, laryngeal, lung, oral, and throat cancers. Therefore, prevention scientists have designed studies to explore the motivation behind smoking and factors that might promote smoking. These studies revealed that alcohol, coffee, food, and presence of others smoking increase the odds of smoking [15, 55, 34, 57].

Urges to smoke is another variable that is often of interest in smoking cessation research because of (1) the intuitive link between urge to smoke and smoking, and (2) the importance of urges in some theories of smoking (for instance, many theories posit that the influence of emotional states on smoking is impacted through the urge to smoke). Shiffman et al. [57] investigated the association between urge to smoke and smoking and concluded that there is a strong and positive association between these variables, especially for lower levels of urge to smoke.

A drawback of the previous analyses was that they did not explore the dynamics of the association between these factors and smoking. However, we suspect that the relationship between these stimuli and smoking (and therefore, perhaps, urge to smoke) might vary over the course of a cessation study, especially before and after a subject quits smoking. Thus we created a binary response by combining the factors that might lead to smoking (specifically, the
binary outcome is zero iff none of the factors is present), and applied TIMECOP with the new binary outcome and urge to smoke as response variables.

In our analysis we focus on the randomly scheduled assessment data collected two weeks before and after the quit day, so that we can study the differences between these periods. We analyzed the data for 206 smokers, each of which had from 46 to 222 observations.

Based on previous analyses of smoking cessation data [54, 58, 57, 38], we used the mood variables negative affect, arousal, and attention disturbance as predictors for urge to smoke (see [57] for more information regarding these scores). We used the same set of predictors for the binary response. All of these predictors are continuous.

We assume that urge to smoke, $Y_{1i}(t)$, is a Gaussian process with mean $\mu_{i1}(t)\beta_1(t)$ and variance $\sigma^2(t)$, where $\beta_1(t) = (\beta_{10}(t), \beta_{11}(t), \beta_{12}(t), \beta_{13}(t))^\prime$ and $\mu_{i1}(t) = (\mu_{i11}(t), \mu_{i12}(t), \mu_{i13}(t))^\prime$ with

- $\mu_{i11}(t)$: the centered score of negative affect of subject $i$ at time $t$,
- $\mu_{i12}(t)$: the centered score of arousal of subject $i$ at time $t$,
- $\mu_{i13}(t)$: the centered score of attention disturbance of subject $i$ at time $t$.

We assume that smoking triggers, $Y_{2i}(t)$, is a Bernoulli process with mean

$$[1 + \exp(-\mu_{i1}(t)\beta_2(t))]^{-1},$$

where $\beta_2(t) = (\beta_{20}(t), \beta_{21}(t), \beta_{22}(t), \beta_{23}(t))^\prime$.

As in the first data application, we used the $K_{0.1}$ bimodal kernel in our estimation procedure and chose a bandwidth of $h = 7$ using the cross-validation procedure described in Section 2. The estimated time-varying regression coefficient functions and variance for the continuous response (urge to smoke) are shown in Figure 8.

- The plot in panel (a) suggests that the intercept function is time varying, and decreases after the quit day.
- The confidence band in panel (b) shows that the coefficient for negative affect is time varying. The effect is always significant and positive, that is, as negative affect increases, urge to smoke also increases. This effect increases until five days after the quit day and then starts to decrease.
- According to panel (c) we observe that the effect of attention disturbance may be time invariant, but it is significantly different from zero and always positive.
- Panel (d) allows us to conclude that the coefficient for attention disturbance is time varying. The effect becomes significant and negative just prior to quit day. Approximately at the same date that the coefficient for negative affect starts to decrease, the effect of attention disturbance starts to increase.
• The final panel shows that the variance of urge to smoke varies over time. The variance is almost constant prior to the quit day, and then it starts to decrease.

Figure 9 shows the estimated time-varying regression coefficient functions for the binary response (factors that may lead to smoking) along with the estimated time-varying association, $\hat{\rho}(t)$, between this binary response and urge to smoke.

• From panel (a) we see that the intercept function is time varying and significant.

• Panel (b) shows that the coefficient for negative affect score might be time invariant. The effect is significant and negative until quit day, after which the effect becomes insignificant.

• In panel (c) we observe that the effect of arousal may be time invariant but is significant and positive.

• The confidence band in panel (d) suggests that the effect of attention disturbance is significant and just barely time varying. The effect is positive and decreases until the quit day. Then the effect becomes nearly constant.

According to the confidence band in the last panel of Figure 9, which shows the estimated time-varying partial association, we do not have sufficient evidence to conclude that $\rho(t)$ is time varying. However, we observe that this partial association is always significant and positive, i.e., for smokers enrolled in this study, exhibiting at least one of the factors is associated with an increased urge to smoke. This result could help prevention scientists to design better cessation programs. For instance, before someone tries to quit smoking, it may be wise of him/her to decrease the urge to smoke by mitigating some or all of the factors.

As in the application to HIV data, these results suggest that some of the parameters may be time invariant, in which case a semivarying model would be more appropriate. But fitting a semivarying model to these data would not lead to substantive changes to our conclusions.

6. Conclusion

In this article we developed a new class of joint models for longitudinal responses, and an estimation procedure based on the local likelihood approach. This new class of models can accommodate (1) responses of mixed type, (2) time-varying association parameters, and (3) all manner of time-varying marginal parameters. We demonstrated the efficacy of our approach via three simulation studies, one of which showed that our approach performs well under ignorable missingness. Then we used our methodology to analyze a bivariate response taken from the Women's Interagency HIV Study. This analysis revealed a significant negative partial association between CD4 cell percentage and smoking status. We also applied our approach to smoking cessation data. Our analysis
Figure 8: The results of our data analysis for the continuous response, urge to smoke. For each panel, the solid curve shows the estimate, the dashed curves show the estimated 95% pointwise confidence band, and the dotted line marks zero.
Figure 3: The results of our data analysis for the binary response (potential factors leading to smoking) and the estimated time-varying association between this binary response and urge to smoke. For each panel, the solid curve shows the estimate, the dashed (or dashed-dotted) curves show the estimated 95% pointwise confidence band, and the dotted line marks zero.
revealed a significant and positive partial association between urge to smoke and factors that may lead to smoking.

Our proposed methodology assumes that all parameters have the same degree of smoothness. However, it may be of interest to develop an estimation procedure capable of handling multiple degrees of smoothness. Perhaps the two-step estimation procedure proposed by Fan and Zhang [19] can be adapted for this purpose.

In this paper we used confidence bands to assess features of parameter functions. It may be desirable to develop hypothesis testing procedures, in which case results presented in [18] may prove useful.

Acknowledgments

This work was partially supported by a grant from the Simons Foundation (#246677 to John Hughes) and grants from NIH (P50 DA16875 and P50 DA36167 to Runze Li).

Data in this manuscript were collected by the Women's Interagency HIV Study (WIHS) Collaborative Study Group with centers (PrINCIPal Investigators) at New York City/Bronx Consortium (Kathryn Anastos); Brooklyn, NY (Howard Minkoff); Washington, DC, Metropolitan Consortium (Mary Young); The Connie Wofsy Study Consortium of Northern California (Ruth Greenblatt); Los Angeles County/Southern California Consortium (Alexandra Levine); Chicago Consortium (Marjorie Cohen); Data Coordinating Center (Stephen Gange). The WIHS is funded by the National Institute of Allergy and Infectious Diseases (UO1-AI-35084, U01-AI-31834, U01-AI-34984, U01-AI-34989, U01-AI-34993, and U01-AI-42590) and by the Enrico Kennedy Shriver National Institute of Child Health and Human Development (U01-HD-32632). The study is co-funded by the National Cancer Institute, the National Institute on Drug Abuse, and the National Institute on Deafness and Other Communication Disorders. Funding is also provided by the National Center for Research Resources (UCSF-CTSI Grant Number ULI RR024131). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.


