Asymptotic Confidence Regions for Kernel Smoothing of a Varying-Coefficient Model with Longitudinal Data

Author(s): Colin O. Wu, Chin-Tsang Chiang and Donald R. Hoover

Source: *Journal of the American Statistical Association*, Vol. 93, No. 444 (Dec., 1998), pp. 1388-1402
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2670054
Accessed: 23-08-2016 17:51 UTC

Asymptotic Confidence Regions for Kernel Smoothing of a Varying-Coefficient Model With Longitudinal Data

Colin O. WU, Chin-Tsang CHIANG, and Donald R. HOOVER

We consider the estimation of the $k+1$-dimensional nonparametric component $\boldsymbol{\beta}(t)$ of the varying-coefficient model $Y(t) = \mathbf{X}^T(t)\boldsymbol{\beta}(t) + \varepsilon(t)$ based on longitudinal observations $(Y_{ij}, \mathbf{X}_i(t_{ij}), t_{ij}), i = 1, \ldots, n, j = 1, \ldots, n_i$, where t_{ij} is the jth observed design time point t of the ith subject and Y_{ij} and $\mathbf{X}_i(t_{ij})$ are the real-valued outcome and R^{k+1} valued covariate vectors of the ith subject at t_{ij}. The subjects are independently selected, but the repeated measurements within subject are possibly correlated. Asymptotic distributions are established for a kernel estimate of $\boldsymbol{\beta}(t)$ that minimizes a local least squares criterion. These asymptotic distributions are used to construct a class of approximate pointwise and simultaneous confidence regions for $\boldsymbol{\beta}(t)$. Applying these methods to an epidemiological study, we show that our procedures are useful for predicting CD4 (T-helper lymphocytes) cell changes among HIV (human immunodeficiency virus)-infected persons. The finite-sample properties of our procedures are studied through Monte Carlo simulations.

KEY WORDS: Asymptotic normality; Confidence interval; Kernel estimate; Nonparametric regression.

1. INTRODUCTION

In a longitudinal study, outcomes and covariates are observed from different subjects each repeatedly measured at a set of distinct time points. This type of data is common in medical and epidemiological studies. For example, longitudinal studies are needed to evaluate progression of disease and effects of treatment on health status. Let $Y(t)$ and $\mathbf{X}(t)$ be the real- and the R^{k+1}-valued outcome and covariate vectors at time t. Let t_{ij} be the time of the jth measurement of the ith subject, $Y_{ij} = Y_i(t_{ij})$, and $\mathbf{X}_i(t_{ij})$ be the real- and R^{k+1}-valued outcome and covariate of the ith subject measured at time t_{ij}. Although the longitudinal measurements given by $(Y_{ij}, \mathbf{X}_i(t_{ij}), t_{ij}), i = 1, \ldots, n$ and $j = 1, \ldots, n_i$, are independent between different subjects, they are likely to be correlated within each subject. For generality, we do not need to restrict t_{ij} to be nonnegative, that is, $t_{ij} \in R$.

Most regression analyses with longitudinal data have concentrated on parametric approaches, such as multivariate linear regression, analysis of variance and generalized linear models (see, e.g., Diggle 1988; Diggle, Liang and Zeger 1994; Liang and Zeger 1986). Longitudinal analyses with nonlinear models have been discussed by Davidian and Giltinan (1995), among others. Nonparametric regression models have the advantage of being more flexible. Smoothing methods with longitudinal nonparametric regression models have been investigated by authors including Altman (1990), Hart (1991), Hart and Wehrly (1986), and Rice and Silverman (1991). These authors only discussed smoothing methods for $Y(t)$ based on t and did not extend to the possible effects of a covariate vector $\mathbf{X}(t)$. Combining parametric and nonparametric regression models, Moyeed and Diggle (1994) and Zeger and Diggle (1994) considered the following semiparametric model:

$$Y_{ij} = \mu_0(t_{ij}) + \sum_{r=0}^{k} \theta_r X_{ir}(t_{ij}) + \varepsilon_i(t_{ij}), \tag{1}$$

where $\mathbf{X}_i(t_{ij}) = (1, X_{i1}(t_{ij}), \ldots, X_{ik}(t_{ij}))^T, X_{il}(t)$ are real-valued covariates at time $t, \mu_0(t)$ is a smooth function of t, θ_r are unknown constants, and the $\varepsilon_i(t)$ are mean 0 stochastic processes. However, because the covariate effects are assumed to be partially linear, (1) may still be restrictive in some settings, particularly for initial exploration of the data.

Perhaps the most general approach is to model $Y(t)$ as a multivariate smooth nonparametric function of $(\mathbf{X}(t), t)$. Unfortunately, smooth estimation of a general multivariate nonparametric regression may require excessively large sample sizes when the dimensionality of the covariate is high, and the smoothing results may be difficult to interpret. A useful alternative is to impose some specific structure between $Y(t)$ and $(\mathbf{X}(t), t)$, so that the proposed model has meaningful interpretations and yet retains certain general nonparametric characteristics. As a promising dimensionality-reduction approach, we consider here the linear time-varying coefficient model of the form

$$Y_{ij} = \mathbf{X}_i^T(t_{ij})\boldsymbol{\beta}(t_{ij}) + \varepsilon_i(t_{ij}), \tag{2}$$

where the $\varepsilon_i(t)$ are as defined in (1), $\mathbf{X}_i(t) = (1, X_{i1}(t), \ldots, X_{ik}(t))^T$ and $\varepsilon_i(t)$ are independent, $t_{ij} \in R, \boldsymbol{\beta}(t) = (\beta_0(t), \ldots, \beta_k(t))^T$, and $\beta_l(t) \in R$ for all $l = 0, \ldots, k$. Because (2) is a special case of the functional linear models of Ramsay and Dalzell (1991) or the general varying-coefficient models of Hastie and Tibshirani (1993), the co-

Colin O. Wu is Associate Professor, Department of Mathematical Sciences, G. W. C. Whiting School of Engineering, The Johns Hopkins University, Baltimore, MD 21218 (E-mail: colinwu@columbo.mts.jhu.edu). Chin-Tsang Chiang is Assistant Professor, Department of Statistics, Tunghai University, Taiwan. Donald R. Hoover is Associate Professor, Department of Epidemiology, School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, MD 21205. This research was supported by the National Institute on Drug Abuse through grant R01 DA10184-01. Data was provided by the MACS Public Use Data Set Release PO4 (1984-1991). The authors thank an associate editor and two referees for their careful reading of the original manuscript and insightful suggestions that have greatly improved the presentation of the article and Alfred Saah for many valuable comments on HIV epidemiology.

Journal of the American Statistical Association
December 1998, Vol. 93, No. 444, Theory and Methods

efficient curves $\beta(t)$ can be estimated by the functional data analysis procedures of Ramsay and Dalzell or the smoothing splines of Hastie and Tibshirani.

Under (2) for repeated measurements $\{(Y_{ij}, \mathbf{X}_i(t_{ij}), t_{ij}): 1 \le i \le n, 1 \le j \le n_i\}$, Hoover, Rice, Wu, and Yang (1998) studied a class of linear smoothers including smoothing splines and local polynomials for the estimation of $\beta(t)$ and investigated the large-sample mean squared risks of a kernel estimator when the number of subjects n approaches infinity. However, large-sample risks of smoothing splines and local polynomials and asymptotic distributions of any of the previously described smoothing estimators of $\beta(t)$ in this model have not been investigated. Thus currently there are no practical confidence procedures for $\beta(t)$. Although a "resampling subject" bootstrap procedure has been used by Hoover et al. (1998) to construct pointwise standard error bars for $\beta(t)$, because no bias adjustment has been considered, this procedure may not lead to adequate confidence regions for $\beta(t)$. Furthermore, theoretical properties of the "resampling subject" bootstrapping have not yet been studied for this setting.

The aim of this article is to derive asymptotic distributions of the kernel estimator $\hat{\beta}(t;h)$ of Hoover et al. (1998) and propose a class of asymptotic procedures for constructing pointwise and simultaneous confidence regions for $\beta(t)$. Because the repeated measurements within each subject may be correlated, our asymptotic distributions are obtained by letting n go to infinity and the maximum proportion of observations contributed by any given subject go to 0; that is, $\lim_{n\to\infty} \max_{1\le i\le n}(n_i/N) = 0$, where $N = \sum_{i=1}^n n_i$. Unlike nonparametric regression with independent cross-sectional data, the convergence rates of $\hat{\beta}(t;h)$ also depend on how fast $\sum_{i=1}^n n_i^2$ and N tend to infinity relative to n.

A motivating example for this article is modeling the changes in CD4 cells (also known as T-helper lymphocytes) over time following seroconversion to human immunodeficiency virus (HIV). It is well known that loss of CD4 cells due to HIV is what leads to acquired immune deficiency syndrome (AIDS) and HIV-related death (Kaslow et al. 1987). Therefore, models of the attenuation in CD4 percent (CD4 percentage of lymphocyte cells) may shed light on HIV pathogenesis. Two important covariates that should be considered in such models are preseroconversion CD4 percent and cigarette smoking behavior. It has been shown in one study (Park, Margolick, Giorgi, Ferbas, Bauer, Kaslow, Muñoz, and the Multicenter AIDS Cohort Study 1992) that cigarette smokers had higher CD4 levels than nonsmokers, but this difference vanished over the years after seroconversion. Similarly, one may expect that those with higher CD4 levels prior to seroconversion may continue to have higher CD4 levels immediately following seroconversion. But this positive effect of preseroconversion CD4 level may also vanish with time. Due to the complicated relationship between CD4 level following seroconversion and smoking and preinfection CD4 level, there is no currently known parametric or semiparametric model that has been justified for modeling the effects of these covariates. Thus nonparametric varying-coefficient models and our confidence procedures appear to be a natural exploratory tool for at least initial modeling of this relationship.

The rest of the article is organized as follows. Section 2 summarizes the derivation of $\hat{\beta}(t;h)$ and a cross-validation criterion for selecting bandwidths. Section 3 establishes the asymptotic distributions of $\hat{\beta}(t;h)$. Section 4 uses these asymptotic distributions to develop approximate pointwise and simultaneous confidence regions for $\beta(\cdot)$. Section 5 presents the application of our procedures to the HIV/CD4 study, and Section 6 investigates the finite-sample properties of our procedures using Monte Carlo simulations. Finally, two appendixes provide proofs of the main technical results.

2. LOCAL ESTIMATION METHODS

2.1 Estimation Based on Local Least Squares

For each given $t \in R$, an equivalent form of (2) is

$$Y(t) = \mathbf{X}^T(t)\beta(t) + \varepsilon(t), \tag{3}$$

where $\varepsilon(t)$ is a mean 0 stochastic process with variance $\sigma^2(t)$ and covariance $\rho_\varepsilon(t_1,t_2)$ for any $t_1 \ne t_2$ and $\mathbf{X}(\cdot)$ and $\varepsilon(\cdot)$ are independent. Assume for the rest of the article that the conditional expectations given $t \in R$,

$$E_{\mathbf{XX}^T}(t) = E[\mathbf{X}(t)\mathbf{X}^T(t)]$$

and

$$E_{\mathbf{X}Y}(t) = E[\mathbf{X}(t)Y(t)],$$

exist and that $E_{\mathbf{XX}^T}(t)$ is invertible. Then for any $t \in R, \beta(t)$ uniquely minimizes $E\{[Y(t) - \mathbf{X}^T(t)\mathbf{b}(t)]^2\}$ with respect to $\mathbf{b}(t)$ and is given by

$$\beta(t) = E^{-1}_{\mathbf{XX}^T}(t)E_{\mathbf{X}Y}(t). \tag{4}$$

Because $\beta(t)$ is a unique minimizer of the second moment of $[Y(t) - \mathbf{X}^T(t)\beta(t)]$, it is natural to estimate $\beta(t)$ through a local least squares criterion. Let $K(\cdot)$ be a Borel-measurable kernel function and let h be a positive bandwidth that may depend on n and $n_i, i = 1, \ldots, n$. Then a local kernel estimator $\hat{\beta}(t;h)$ of $\beta(t)$ can be obtained by minimizing

$$L_N(t) = \sum_{i=1}^n \sum_{j=1}^{n_i} [Y_{ij} - \mathbf{X}_i^T(t_{ij})\mathbf{b}(t)]^2 K\left(\frac{t - t_{ij}}{h}\right)$$

with respect to $\mathbf{b}(t)$. Equivalently, we can write $L_N(t)$ in the matrix form,

$$L_N(t) = \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{X}_i\beta(t))^T \mathbf{K}_i(t;h)(\mathbf{Y}_i - \mathbf{X}_i\beta(t)), \tag{5}$$

where

$$\mathbf{X}_i = \begin{pmatrix} 1 & X_{i1}(t_{i1}) & \cdots & X_{ik}(t_{i1}) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{i1}(t_{in_i}) & \cdots & X_{ik}(t_{in_i}) \end{pmatrix}, \quad \mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{pmatrix},$$

and $\mathbf{K}_i(\cdot;\cdot)$ is a diagonal kernel matrix such that

$$\mathbf{K}_i(t;h) = \mathrm{diag}\left(K\left(\frac{t - t_{i1}}{h}\right), \ldots, K\left(\frac{t - t_{in_i}}{h}\right)\right).$$

Assuming that $(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{K}_i(t;h)\mathbf{X}_i)$ is also invertible, $\hat{\boldsymbol{\beta}}(t;h)$, as the unique minimizer of (5), is given by the following $k+1$ column vector:

$$\hat{\boldsymbol{\beta}}(t;h) = \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{K}_i(t;h)\mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{K}_i(t;h)\mathbf{Y}_i\right). \quad (6)$$

Remark 1. The main advantage of using $\hat{\boldsymbol{\beta}}(t;h)$ is that, because of its simple and explicit mathematical expression, it is easy to implement in practice and has tractable asymptotic properties. However, because $\hat{\boldsymbol{\beta}}(t;h)$ involves only one bandwidth, it may not be well equipped to provide adequate smoothing to all components of $\boldsymbol{\beta}(t)$ simultaneously when $\beta_0(t), \ldots, \beta_k(t)$ belong to different smoothness families. Further study is needed to investigate the statistical properties of other least squares–type smoothing methods, such as smoothing splines and local polynomials, that use multiple smoothing parameters to adapt for the different smoothing needs of $\beta_0(t), \ldots, \beta_k(t)$. But the main results of this article are concentrated on the asymptotic distributions and inference procedures of the kernel estimator $\hat{\boldsymbol{\beta}}(t;h)$. These asymptotic properties of $\hat{\boldsymbol{\beta}}(t;h)$ also provide useful insights into the statistical behaviors of other natural smoothing estimators of $\boldsymbol{\beta}(t)$.

Remark 2. Two other types of smoothing methods found to be useful for independent cross-sectional data are the local likelihood and penalized likelihood estimators (e.g. Green and Silverman 1994; Hastie and Tibshirani 1990). These smoothing methods can be extended to the current longitudinal context if additional information about the random process $\varepsilon(\cdot)$, such as the distribution of $\varepsilon(t)$ and the covariance structure, is available. In most applications, the distribution and the covariance structure of $\varepsilon(t)$ are unknown, so that appropriate likelihood functions are difficult to specify. Thus in this case, the local least squares are generally preferable to the likelihood-based approaches.

2.2 Automatic Bandwidth Choices

It is well known that in kernel regression with independent cross-sectional data, bandwidth choice strongly influences the adequacy of the estimator, whereas kernel choices often have little effect (cf. Härdle 1990; Silverman 1986). It is reasonable to expect that bandwidth choice has a similar effect in the current longitudinal data case. A suitable bandwidth for $\hat{\boldsymbol{\beta}}(t;h)$ can be selected subjectively by examining fitted curves. However, an automatic bandwidth procedure is of both theoretical and practical interest and is usually needed to provide a preliminary idea of a suitable bandwidth range that is suggested by the data.

Because subjects are independent, an intuitive bandwidth selection procedure suggested by Rice and Silverman (1991) is to use a "leave-one-subject-out" cross-validation. Suppose that we would like to measure the risk of $\hat{\boldsymbol{\beta}}(t;h)$ by its average prediction squared error,

$$\mathrm{APSE}(\hat{\boldsymbol{\beta}}) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} E\{[Y_{ij}^* - \mathbf{X}_i^T(t_{ij})\hat{\boldsymbol{\beta}}(t_{ij};h)]^2\}$$

with Y_{ij}^* being a new observation at $(\mathbf{X}_i(t_{ij}), t_{ij})$. Then the corresponding "leave-one-subject-out" cross-validation criterion is given by

$$\mathrm{CV}(h) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} [Y_{ij} - \mathbf{X}_i^T(t_{ij})\hat{\boldsymbol{\beta}}^{(-i)}(t_{ij};h)]^2, \quad (7)$$

where $\hat{\boldsymbol{\beta}}^{(-i)}(t_{ij};h)$ is the kernel estimator of $\boldsymbol{\beta}(t)$ computed with all measurements of the ith subject deleted. A cross-validation bandwidth h_{CV} is then obtained by minimizing $\mathrm{CV}(h)$ with respect to h; that is, $h_{\mathrm{CV}} = \inf_{h>0} \mathrm{CV}(h)$. This cross-validation criterion can be easily extended to other smoothing estimators, such as smoothing splines and local polynomials.

Remark 3. Using the decomposition

$$\begin{aligned}\mathrm{CV}(h) = {} & \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} [Y_{ij} - \mathbf{X}_i^T(t_{ij})\boldsymbol{\beta}(t_{ij})]^2 \\ & + \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \{\mathbf{X}_i^T(t_{ij})[\boldsymbol{\beta}(t_{ij}) - \hat{\boldsymbol{\beta}}^{(-i)}(t_{ij};h)]\}^2 \\ & - \frac{2}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \{[Y_{ij} - \mathbf{X}_i^T(t_{ij})\boldsymbol{\beta}(t_{ij})] \\ & \times [\mathbf{X}_i^T(t_{ij})(\boldsymbol{\beta}(t_{ij}) - \hat{\boldsymbol{\beta}}^{(-i)}(t_{ij};h))]\}, \end{aligned} \quad (8)$$

it can be shown by straightforward algebra that when n is large, h_{CV} approximately minimizes $\mathrm{APSE}(\hat{\boldsymbol{\beta}})$. Hart and Wehrly (1993) proved the consistency of a similar cross-validation criterion based on a different, but simpler, regression model. However, further theoretical properties of h_{CV} still need to be investigated.

3. ASYMPTOTIC NORMALITY OF THE LOCAL KERNEL ESTIMATOR

In this section we derive the asymptotic distributions of $\hat{\boldsymbol{\beta}}(t_0;h)$ at a fixed point $t_0 \in R$. We assume throughout the rest of the article that the design points $t_{ij}, i = 1, \ldots, n, j = 1, \ldots, n_i$, are random and iid according to an underlying density f with respect to the Lebesgue measure. Let $\mathcal{S}(f)$ be the support of f. We assume that t_0 is an interior point of $\mathcal{S}(f)$ and, as in Section 2, both $E_{\mathbf{X}\mathbf{X}^T}(t_0)$ and $\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{K}_i(t_0;h)\mathbf{X}_i$ are invertible. Although the iid assumption of the time design points is technically only suitable for the random design case, the results and proofs of this article can be modified to accommodate the fixed-design case.

To simplify the notation, we denote

$$\sigma^2(t_0) = E[\varepsilon^2(t_0)], \qquad \rho_\varepsilon(t_0) = \lim_{\Delta \to 0} E[\varepsilon(t_0+\Delta)\varepsilon(t_0)],$$

and

$$\eta_{lr}(t_0) = E[X_{il}(t_{ij})X_{ir}(t_{ij}) | t_{ij} = t_0], \qquad l, r = 0, \ldots, k.$$

In general, the variance term $\sigma^2(t_0)$ may not equal $\rho_\varepsilon(t_0)$. Such inequalities between $\sigma^2(t_0)$ and $\rho_\varepsilon(t_0)$ arise, for example, when the error process $\varepsilon(t)$ includes a stationary Gaussian process and some measurement errors that are independent at different time points t (see, e.g., Zeger and Diggle 1994). The following regularity conditions are also assumed throughout the rest of the article.

a. The bandwidth satisfies $h = N^{-1/5}h_0$ for some constant $h_0 > 0$.
b. $\lim_{n\to\infty} N^{-6/5}\sum_{i=1}^n n_i^2 = \lambda$ for some $0 \le \lambda < \infty$.
c. The kernel function $K(\cdot)$ has a compact support on R and satisfies

$$\int K(u)\,du = 1, \quad \int K^2(u)\,du < \infty,$$

$$\int u^2K(u)\,du < \infty,$$

and

$$\int uK(u)\,du = 0.$$

d. There exists a constant $\delta > 0$ such that $E(|\varepsilon(t)|^{2+\delta}) < \infty$ and $E(|X_{il}(t_{ij})|^{4+\delta}) < \infty$ for all $i = 1,\ldots,n, j = 1,\ldots,n_i, l = 0,\ldots,k$ and $t \in \mathcal{S}(f)$.
e. For all $l, r = 0,\ldots,k, \beta_r(t), \eta_{lr}(t)$ and $f(t)$ have continuous second derivatives at t_0.
f. The variance and covariance functions, $\sigma^2(t)$ and $\rho_\varepsilon(t)$, are continuous at t_0.

In practice, the intracorrelation structures are usually unknown. Thus $\hat\beta(t;h)$ is derived based on the ordinary least squares approach of (6), which completely ignores any intracorrelation. Otherwise, a generalized least squares approach might seem preferable. However, because $\sigma^2(t)$ and $\rho_\varepsilon(t)$ are assumed to be continuous only at t_0, it is unknown what kind of generalized least squares would have to be used. For the asymptotic properties of $\hat\beta(t;h)$, it is certainly natural to allow $n \to \infty$. But in most medical and epidemiological studies, the numbers of repeated measurements n_i may vary significantly. By assumption b, we consider only those cases in which $n_i, i = 1,\ldots,n$, do not converge to ∞ too fast relative to N. This is realistic for many practical situations and is more general than the situations considered by Moyeed and Diggle (1994). The effects of the sizes of $n, \sum_{i=1}^n n_i^2$ and N on the asymptotic properties of $\hat{\boldsymbol\beta}(t;h)$ are discussed further in Remark 5. The compactness of the support of $K(\cdot)$ is only a technical assumption that can usually be relaxed in real applications by using kernels with small tails; for example, the standard Gaussian density.

Denote, for all $l, r = 0,\ldots,k$,

$$\mu_1(K) = \int u^2K(u)\,du, \quad \mu_2(K) = \int K^2(u)\,du;$$

$$\begin{aligned} b_l(t_0) = h_0^{3/2}\sum_{c=0}^k \{&\mu_1(K)[\beta_c'(t_0)\eta_{lc}'(t_0)f(t_0) \\ &+ \beta_c'(t_0)\eta_{lc}(t_0)f'(t_0) \\ &+ (1/2)\beta_c''(t_0)\eta_{lc}(t_0)f(t_0)]\}, \quad \text{for} \quad l = 0,\ldots,k; \end{aligned} \tag{9}$$

$$\mathbf{B}(t_0) = (f(t_0))^{-1}E^{-1}_{\mathbf{XX}^T}(t_0)(b_0(t_0),\ldots,b_k(t_0))^T; \tag{10}$$

$$D_{lr}(t_0) = \sigma^2(t_0)\eta_{lr}(t_0)f(t_0)\mu_2(K) + \lambda h_0\rho_\varepsilon(t_0)\eta_{lr}(t_0)f^2(t_0); \tag{11}$$

$$\mathbf{D}(t_0) = \begin{pmatrix} D_{00}(t_0) & D_{01}(t_0) & \cdots & D_{0k}(t_0) \\ \vdots & \vdots & \vdots & \vdots \\ D_{k0}(t_0) & D_{k1}(t_0) & \cdots & D_{kk}(t_0) \end{pmatrix};$$

and

$$\mathbf{D}^*(t_0) = (f(t_0))^{-2}E^{-1}_{\mathbf{XX}^T}(t_0)\mathbf{D}(t_0)E^{-1}_{\mathbf{XX}^T}(t_0). \tag{12}$$

We summarize the main result in the next theorem. A more general version of this theorem, which describes the asymptotic normality of $\hat{\boldsymbol\beta}(\cdot;h)$ at a set of distinct time points, is given in Appendix A.

Theorem 1. Suppose that Assumptions a–f are satisfied. With proper normalization, $\hat{\boldsymbol\beta}(t_0;h) - \boldsymbol\beta(t_0)$ has asymptotically a multivariate Gaussian distribution; that is,

$$(Nh)^{1/2}(\hat{\boldsymbol\beta}(t_0;h) - \boldsymbol\beta(t_0)) \to \mathcal{N}(\mathbf{B}(t_0), \mathbf{D}^*(t_0)) \quad \text{in distribution as} \quad n \to \infty,$$

where $\mathbf{B}(t_0)$ and $\mathbf{D}^*(t_0)$ are defined in (10) and (12).

Proof. See Appendix A.

Remark 4. The main influence of the intrasubject correlations on the asymptotic distribution of $\hat\beta(t_0;h)$ appears at the second term on the right side of (11). The asymptotic bias of $\hat{\boldsymbol\beta}(t_0;h)$ is not affected by the intracorrelations. When $\lambda = 0$ or $\rho_\varepsilon(t_0) = 0$, (11) is similar to the asymptotic variance of kernel regression estimators with independent cross-sectional data (cf. Härdle 1990). From the proof of Theorem 1, we can also derive that if $h = o(N^{-1/5})$ but assumptions b–f are satisfied, then the asymptotic bias term disappears and

$$(Nh)^{1/2}(\hat{\boldsymbol\beta}(t_0;h) - \boldsymbol\beta(t_0)) \to \mathcal{N}(\mathbf{0}, \mathbf{D}^*(t_0)) \quad \text{in distribution as} \quad n \to \infty,$$

where $\mathbf{0}$ is the $k+1$ column vector of 0's.

Remark 5. A direct implication of Theorem 1 is that, to ensure good asymptotic properties of $\hat\beta(t_0;h)$, the numbers of repeated measurements $n_1,\ldots,n_n$ must be small relative to the overall sample size N. It was shown by Hoover et al. (1998) that $\hat{\boldsymbol\beta}(t_0;h)$ is a consistent estimator of

$\beta(t_0)$ if and only if $\sum_{i=1}^{n} n_i^2 = o(N^2)$, which is equivalent to $\max_{1\le i\le n}(n_i/N) = o(1)$. Here we assume a somewhat stronger condition, $\sum_{i=1}^{n} n_i^2 = O(N^{6/5})$, which ensures that $\hat{\beta}(t_0; h)$ has an attainable rate of $N^{-2/5}$. If $\sum_{i=1}^{n} n_i^2$ converges to infinity faster than $N^{6/5}$, then it can be shown with a slight modification of the method of this article that the attainable rate for $\hat{\beta}(t_0; h)$ is slower than $N^{-2/5}$, and the asymptotic distribution of $\hat{\beta}(t_0; h)$ may be similarly derived, but at the expense of more tedious computations.

Remark 6. Another important assumption in this section is that t_0 is an interior point of the support $\mathcal{S}(f)$. It is well known in the independent cross-sectional data case that kernel estimators suffer from increased biases at the boundary of the design intervals. Methods for improving the theoretical and practical performance of kernel estimators have been extensively studied in the literature (e.g. Hall and Wehrly 1991; Müller 1993; Rice 1984). Here it is natural to expect $\hat{\beta}(t_0; h)$ to have a relatively larger bias when t_0 is near the boundary of $\mathcal{S}(f)$. Further study of the properties of $\hat{\beta}(t_0; h)$ near the boundary and possible modifications to improve these properties is warranted.

4. ASYMPTOTIC CONFIDENCE REGIONS

Pointwise confidence intervals for $\beta(t)$ may be constructed using bootstrap procedures by randomly sampling the entire repeated measurements of subjects with replacement. But statistical properties of bootstrapping with longitudinal data are still not well understood. In this section we first propose a class of asymptotic pointwise confidence intervals for $\beta(t_0)$ based on Theorem 1 and a class of kernel estimators of $\mathbf{B}(t_0)$ and $\mathbf{D}^*(t_0)$ and then, as a direct extension, suggest a Bonferroni-type method for constructing simultaneous confidence bands for $\beta(\cdot)$ over an interval $[a, b]$ within $\mathcal{S}(f)$.

4.1 General Formulation

4.1.1 Pointwise Confidence Intervals. Let t_0 be a fixed interior point in the support $\mathcal{S}(f)$ and let $\mathbf{A} = (a_0, \ldots, a_k)^T$ be any $k+1$ column vector of constants $a_r \in R, r = 0, \ldots, k$. We consider the approximate intervals for $\mathbf{A}^T\beta(t_0)$. If $\mathbf{B}(t_0)$ and $\mathbf{D}^*(t_0)$ were known, then, by Assumptions a–f and Theorem 1,

$$\lim_{n\to\infty} P(L_\alpha(t_0) \le \mathbf{A}^T\beta(t_0) \le U_\alpha(t_0)) = 1 - \alpha, \qquad (13)$$

where the lower and upper bounds $L_\alpha(t_0)$ and $U_\alpha(t_0)$ are given by

$$[\mathbf{A}^T\hat{\beta}(t_0; h) - (Nh)^{-1/2}\mathbf{A}^T\mathbf{B}(t_0)] \pm z_{\alpha/2}(Nh)^{-1/2}(\mathbf{A}^T\mathbf{D}^*(t_0)\mathbf{A})^{1/2},$$

and $z_{\alpha/2}$ is the $1-\alpha/2$ quantile value of the standard Gaussian distribution.

Similarly, suppose that Assumptions b–f hold but the bandwidth satisfies $h = o(N^{-1/5})$. By Remark 4, the bias of $\hat{\beta}(t_0; h)$ is of negligible magnitude compared to the variance of $\hat{\beta}(t_0; h)$. Then

$$\lim_{n\to\infty} P(L_\alpha^*(t_0) \le \mathbf{A}^T\beta(t_0) \le U_\alpha^*(t_0)) = 1 - \alpha, \qquad (14)$$

with $L_\alpha^*(t_0)$ and $U_\alpha^*(t_0)$ given by

$$\mathbf{A}^T\hat{\beta}(t_0; h) \pm z_{\alpha/2}(Nh)^{-1/2}(\mathbf{A}^T\mathbf{D}^*(t_0)\mathbf{A})^{1/2}.$$

Because $\mathbf{B}(t_0)$ and $\mathbf{D}^*(t_0)$ are unknown in practice, approximate $1-\alpha$ confidence intervals of $\mathbf{A}^T\beta(t_0)$ can be straightforwardly constructed by substituting $(\mathbf{B}(t_0), \mathbf{D}^*(t_0))$ of (13) or $\mathbf{D}^*(t_0)$ of (14) with their appropriate estimates. Let $\hat{\mathbf{B}}(t_0)$ and $\hat{\mathbf{D}}^*(t_0)$ be any consistent estimators of $\mathbf{B}(t_0)$ and $\mathbf{D}^*(t_0)$. Then, under Assumption a, an approximate $1-\alpha$ confidence interval of $\mathbf{A}^T\beta(t_0)$ with bias correction can be given by $(\hat{L}_\alpha(t_0), \hat{U}_\alpha(t_0))$, with $\hat{L}_\alpha(t_0)$ and $\hat{U}_\alpha(t_0)$ satisfying

$$[\mathbf{A}^T\hat{\beta}(t_0; h) - (Nh)^{-1/2}\mathbf{A}^T\hat{\mathbf{B}}(t_0)] \pm z_{\alpha/2}(Nh)^{-1/2}(\mathbf{A}^T\hat{\mathbf{D}}^*(t_0)\mathbf{A})^{1/2}. \qquad (15)$$

When the bias of $\hat{\beta}(t_0; h)$ is negligible, a computationally simpler approximate interval can be constructed without bias correction. In this case we select $h = o(N^{-1/5})$ and construct a $1-\alpha$ approximate interval $(\hat{L}_\alpha^*(t_0), \hat{U}_\alpha^*(t_0))$ based on (14) and a consistent estimator $\hat{\mathbf{D}}^*(t_0)$, such that $\hat{L}_\alpha^*(t_0)$ and $\hat{U}_\alpha^*(t_0)$ satisfy

$$\mathbf{A}^T\hat{\beta}(t_0; h) \pm z_{\alpha/2}(Nh)^{-1/2}(\mathbf{A}^T\hat{\mathbf{D}}^*(t_0)\mathbf{A})^{1/2}. \qquad (16)$$

Remark 7. The approximate intervals (15) and (16) depend on the adequacy of $\hat{\mathbf{B}}(t_0)$ and $\hat{\mathbf{D}}^*(t_0)$ as estimators of $\mathbf{B}(t_0)$ and $\mathbf{D}^*(t_0)$. A class of consistent kernel-type estimators of $\mathbf{B}(t_0)$ and $\mathbf{D}^*(t_0)$ is proposed in Section 4.2, which in general requires bandwidth sequences that may be different from h. Although $(\hat{L}_\alpha^*(t_0), \hat{U}_\alpha^*(t_0))$ relies on the bandwidth choice $h = o(N^{-1/5})$, which asymptotically provides a slower convergence rate for $\hat{\beta}(t_0; h)$ than the attainable rate of $N^{-2/5}$, it has the advantage of being computationally simple, because no estimator of $\mathbf{B}(t_0)$ is needed. Thus in practice, $(\hat{L}_\alpha^*(t_0), \hat{U}_\alpha^*(t_0))$ can be a very competitive procedure compared to $(\hat{L}_\alpha(t_0), \hat{U}_\alpha(t_0))$.

4.1.2 Bonferroni-Type Variability Bands. Construction of appropriate, and yet computationally simple, simultaneous confidence bands for regression curves is known to be difficult even with independent cross-sectional data. One popular approach, which has been investigated by Bickel and Rosenblatt (1973), Eubank and Speckman (1993), Härdle (1989), and Johnston (1982), among others, is to develop the level of the confidence bands based on the extreme value theory of Gaussian processes. These bands have the advantage of being simple to implement (cf. Eubank and Speckman 1993). But under the current context, the theory of Gaussian approximation with the presence of possible intrasubject correlations has not been developed. Another approach, which has been considered by Hall and Titterington (1988), Härdle and Marron (1991), Knafl, Sacks, and Ylvisaker (1985), and others, is to construct variability bands by first establishing simultaneous confidence intervals for a set of grid points and then bridging the gaps between grid points via smoothness conditions of the regression curves.

Let $[a, b]$ be an interval of the interior points of $\mathcal{S}(f)$. Following Knafl et al. (1985), we suggest here a simple Bonferroni-type method to construct variability bands for $\mathbf{A}^T\boldsymbol{\beta}(t)$ over $t \in [a, b]$. Let M be a positive integer, $\gamma = (b-a)/M$, and let $a = \xi_1 < \cdots < \xi_{M+1} = b$ be $M+1$ grid points such that $\xi_{j+1} - \xi_j = \gamma, j = 1, \ldots, M$. A set of approximate simultaneous confidence intervals for $\mathbf{A}^T\boldsymbol{\beta}(\xi_j), j = 1, \ldots, M+1$, with level greater or equal to $1-\alpha$, is then denoted by $\{(\hat{l}_\alpha(\xi_j), \hat{u}_\alpha(\xi_j)); j = 1, \ldots, M+1\}$, so that

$$\lim_{n\to\infty} P[\hat{l}_\alpha(\xi_j) \le \mathbf{A}^T\boldsymbol{\beta}(\xi_j) \le \hat{u}_\alpha(\xi_j) \quad \forall\, j = 1, \ldots, M+1] \ge 1-\alpha. \tag{17}$$

Using (15) and (16), Bonferroni-type choices of $(\hat{l}_\alpha(\xi_j), \hat{u}_\alpha(\xi_j))$ may include $(\hat{L}_{\alpha/(M+1)}(\xi_j), \hat{U}_{\alpha/(M+1)}(\xi_j))$ and $(\hat{L}^*_{\alpha/(M+1)}(\xi_j), \hat{U}^*_{\alpha/(M+1)}(\xi_j))$.

For $t \in [\xi_j, \xi_{j+1}]$, let $(\mathbf{A}^T\boldsymbol{\beta})^{(I)}(t)$, $\hat{l}_\alpha^{(I)}(t)$, $\hat{u}_\alpha^{(I)}(t)$, $\hat{L}^{(I)}_{\alpha/(M+1)}(t)$, $\hat{U}^{(I)}_{\alpha/(M+1)}(t)$, $\hat{L}^{*(I)}_{\alpha/(M+1)}(t)$, and $\hat{U}^{*(I)}_{\alpha/(M+1)}(t)$ be the linear interpolations of $\mathbf{A}^T\boldsymbol{\beta}$, $\hat{l}_\alpha$, $\hat{u}_\alpha$, $\hat{L}_{\alpha/(M+1)}$, $\hat{U}_{\alpha/(M+1)}$, $\hat{L}^*_{\alpha/(M+1)}$, and $\hat{U}^*_{\alpha/(M+1)}$, based on their values at ξ_j and ξ_{j+1}. For example,

$$\begin{aligned}(\mathbf{A}^T\boldsymbol{\beta})^{(I)}(t) &= (\xi_{j+1}-t)\gamma^{-1}\mathbf{A}^T\boldsymbol{\beta}(\xi_j) \\ &\quad + (t-\xi_j)\gamma^{-1}\mathbf{A}^T\boldsymbol{\beta}(\xi_{j+1}), \quad \text{for} \quad t \in [\xi_j, \xi_{j+1}].\end{aligned}$$

Then, it easily follows from (17) that

$$\lim_{n\to\infty} P[\hat{l}_\alpha^{(I)}(t) \le (\mathbf{A}^T\boldsymbol{\beta})^{(I)}(t) \le \hat{u}_\alpha^{(I)}(t), \quad \forall\, t \in [a, b] \ge 1-\alpha. \tag{18}$$

To bridge the gaps between $\mathbf{A}^T\boldsymbol{\beta}(t)$ and $(\mathbf{A}^T\boldsymbol{\beta})^{(I)}(t)$, smoothness conditions of $\mathbf{A}^T\boldsymbol{\beta}(t)$ must be specified. We consider the following two conditions, which have been considered by Knafl et al. (1985):

$$\sup_{t\in[a,b]} |(\mathbf{A}^T\boldsymbol{\beta})'(t)| \le c_1, \quad \text{for a known constant} \quad c_1 > 0, \tag{19}$$

and

$$\sup_{t\in[a,b]} |(\mathbf{A}^T\boldsymbol{\beta})''(t)| \le c_2, \quad \text{for a known constant} \quad c_2 > 0. \tag{20}$$

If (19) is satisfied, then it can be shown by direct calculation that

$$|\mathbf{A}^T\boldsymbol{\beta}(t) - (\mathbf{A}^T\boldsymbol{\beta})^{(I)}(t)| \le c_1\gamma^{-1}(\xi_{j+t}-t)(t-\xi_j), \quad \forall\, t \in [\xi_j, \xi_{j+1}]. \tag{21}$$

Similarly, if (20) is satisfied, then we can show that

$$|\mathbf{A}^T\boldsymbol{\beta}(t) - (\mathbf{A}^T\boldsymbol{\beta})^{(I)}(t)| \le \frac{1}{2}c_2(\xi_{j+t}-t)(t-\xi_j), \quad \forall\, t \in [\xi_j, \xi_{j+1}]. \tag{22}$$

For other smoothness families, such as those considered by Hall and Titterington (1988), further analyses are required to evaluate the upper bounds of $|\mathbf{A}^T\boldsymbol{\beta}(t) - (\mathbf{A}^T\boldsymbol{\beta})^{(I)}(t)|$.

Thus if (19) or (20) are satisfied, then one can use

$$\begin{aligned}&(\hat{l}_\alpha^{(I)}(t) - c_1\gamma^{-1}(\xi_{j+1}-t)(t-\xi_j), \\ &\qquad \hat{u}_\alpha^{(I)}(t) + c_1\gamma^{-1}(\xi_{j+1}-t)(t-\xi_j))\end{aligned} \tag{23}$$

or

$$\begin{aligned}&\Big(\hat{l}_\alpha^{(I)}(t) - \frac{1}{2}c_2(\xi_{j+1}-t)(t-\xi_j), \\ &\qquad \hat{u}_\alpha^{(I)}(t) + \frac{1}{2}c_2(\xi_{j+1}-t)(t-\xi_j)\Big),\end{aligned} \tag{24}$$

respectively, for all $t \in [\xi_j, \xi_{j+1}]$ and $j = 1, \ldots, M$, as the level $1-\alpha$ confidence bands of $\mathbf{A}^T\boldsymbol{\beta}(t)$ over $t \in [a, b]$. Specific choices of $(\hat{l}_\alpha^{(I)}(t), \hat{u}_\alpha^{(I)}(t))$ include

$$(\hat{L}^{(I)}_{\alpha/(M+1)}(t), \hat{U}^{(I)}_{\alpha/(M+1)}(t)) \quad \text{or} \quad (\hat{L}^{*(I)}_{\alpha/(M+1)}(t), \hat{U}^{*(I)}_{\alpha/(M+1)}(t)). \tag{25}$$

Remark 8. The choice of M, the number of grids, is an important factor affecting the overall widths of the bands of (23) and (24). One possible choice suggested by Hall and Titterington (1988) and Knafl et al. (1985) is to take M to be the smallest integer such that $M > (b-a)/h$. Under the framework of a single nonparametric regression curve with fixed design points and independent cross-sectional data, Hall and Titterington (1988) established the best rate for the widths of confidence bands converging to 0 and showed that their best rate is attained if the number of grids is chosen correctly. However, analogous rates for model (2) have not been developed, and further study is also needed for the optimal selection of M. It is also interesting to note that Bonferroni bands are usually very conservative, and methods for improvement have been extensively studied in the literature (see, e.g., Hoover 1990). Further study for extending these improved bands to the current setting may be warranted.

4.2 Estimation of Bias and Variance

By (9)–(12), $\mathbf{B}(t_0)$ and $\mathbf{D}^*(t_0)$ depend on $f(t_0)$, $f'(t_0)$, $\beta_r'(t_0)$, $\beta_r''(t_0)$, $\eta_{lr}(t_0)$, $\eta_{lr}'(t_0)$, $\rho_\varepsilon(t_0)$, and $\sigma^2(t_0)$. We now propose a class of intuitive and computationally straightforward kernel smoothers for the estimation of these quantities.

4.2.1 Kernel Estimators. In addition to assumption c, we assume that the kernels used here are twice continuously differentiable with respect to u. If $h_{(f,0)}$ and $h_{(\eta_{lr},0)}$ are bandwidths satisfying $\lim_{n\to\infty} h_{(\cdot,0)} = 0$ and $\lim_{n\to\infty} Nh_{(\cdot,0)} = \infty$, then $f(t_0)$ and $\eta_{lr}(t_0)$ can be estimated by

$$\hat{f}(t_0; h_{(f,0)}) = (Nh_{(f,0)})^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i} K\left(\frac{t_0 - t_{ij}}{h_{(f,0)}}\right) \tag{26}$$

and

$$\hat{\eta}_{lr}(t_0; h_{(\eta_{lr},0)})$$
$$= (\hat{f}(t_0; h_{(\eta_{lr},0)}))^{-1}(Nh_{(\eta_{lr},0)})^{-1}$$
$$\times \sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{il}(t_{ij})X_{ir}(t_{ij})K\left(\frac{t_0 - t_{ij}}{h_{(\eta_{lr},0)}}\right). \quad (27)$$

Let $\hat{E}_{\mathbf{XX}^T}(t_0)$ be the $(k+1)\times(k+1)$ matrix whose (l,r)th element is $\hat{\eta}_{lr}(t_0; h_{(\eta_{lr},0)})$. Suppose that $\hat{E}_{\mathbf{XX}^T}(t_0)$ is invertible. Then $E_{\mathbf{XX}^T}(t_0)$ and $E^{-1}_{\mathbf{XX}^T}(t_0)$ can be estimated by $\hat{E}_{\mathbf{XX}^T}(t_0)$ and $\hat{E}^{-1}_{\mathbf{XX}^T}(t_0)$.

Let $f^{(d)}, \beta_r^{(d)}$, and $\eta_{lr}^{(d)}, d = 1, 2$, be the dth derivatives of $f(t_0), \beta_r(t_0)$, and $\eta_{lr}(t_0)$ and let $h_{(f,d)}, h_{(\beta_r,d)}$, and $h_{(\eta_{lr},d)}$ be the bandwidths satisfying $\lim_{n\to\infty} h_{(\cdot,d)} = 0$ and $\lim_{n\to\infty} Nh^{2d+1}_{(\cdot,d)} = \infty$. Then, following kernel derivative estimators in the independent cross-sectional data case (see, e.g., Härdle 1990, chap. 3), $f^{(d)}, \beta_r^{(d)}$, and $\eta_{lr}^{(d)}$ can be estimated by the corresponding dth derivatives $\hat{f}^{(d)}(t_0; h_{(f,d)}), \hat{\beta}_r^{(d)}(t_0; h_{(\beta_r,d)})$, and $\hat{\eta}_{lr}^{(d)}(t_0; h_{(\eta_{lr},d)})$ of $\hat{f}(t_0; h_{(f,d)}), \hat{\beta}_r(t_0; h_{(\beta_r,d)})$, and $\hat{\eta}_{lr}(t_0; h_{(\eta_{lr},d)})$, where $\hat{\beta}_r(t_0; h_{(\beta_r,d)})$ is the rth component of (6).

For estimating the variance $\sigma^2(t_0)$ and covariance $\rho_\varepsilon(t_0)$, we rely on smoothing functionals of the residuals $\hat{\varepsilon}_i(t_{ij}; h) = Y_{ij} - \mathbf{X}_i^T(t_{ij})\hat{\beta}(t_{ij}; h)$. Let $h_{(\sigma)}$ be a bandwidth satisfying $\lim_{n\to\infty} h_{(\sigma)} = 0$ and $\lim_{n\to\infty} Nh_{(\sigma)} = \infty$. The variance $\sigma^2(t_0)$ can be simply estimated by

$$\hat{\sigma}^2(t_0; h_{(\sigma)}) = \frac{1}{Nh_{(\sigma)}\hat{f}(t_0; h_{(\sigma)})} \times \sum_{i=1}^{n}\sum_{j=1}^{n_i} \hat{\varepsilon}_i^2(t_{ij}; h)K\left(\frac{t_0 - t_{ij}}{h_{(\sigma)}}\right). \quad (28)$$

Estimation of $\rho_\varepsilon(t_0)$ is usually more difficult. Because $\rho_\varepsilon(s_1, s_2) = E(\varepsilon_i(s_1)\varepsilon_i(s_2))$ for $s_1 \neq s_2$ and $\rho_\varepsilon(t_0) = \lim_{s\to t_0} \rho_\varepsilon(s, t_0)$, we can estimate $\rho_\varepsilon(t_0)$ by smoothing $\hat{\varepsilon}_i(t_{ij_1}; h)\hat{\varepsilon}_i(t_{ij_2}; h)$ for $j_1 \neq j_2$ when $n_i, i = 1, \ldots, n$, are large; that is,

$$\hat{\rho}_\varepsilon(t_0; h_{(\rho)}) = \frac{\sum_{i=1}^n \sum_{j_1\neq j_2} \hat{\varepsilon}_i(t_{ij_1}; h)\hat{\varepsilon}_i(t_{ij_2}; h) \times K\left(\frac{t_0 - t_{ij_1}}{h_{(\rho)}}\right)K\left(\frac{t_0 - t_{ij_2}}{h_{(\rho)}}\right)}{\sum_{i=1}^n \sum_{j_1\neq j_2} K\left(\frac{t_0 - t_{ij_1}}{h_{(\rho)}}\right)K\left(\frac{t_0 - t_{ij_2}}{h_{(\rho)}}\right)}, \quad (29)$$

where $h_{(\rho)}$ satisfies $\lim_{n\to\infty} h_{(\rho)} = 0$ and $\lim_{n\to\infty} Nh_{(\rho)} = \infty$.

Finally, we obtain the kernel estimator $\hat{\mathbf{B}}(t_0)$ by substituting $f^{(d)}(t_0), \beta_c^{(d)}(t_0)$, and $\eta_{lc}^{(d)}(t_0), d = 0, 1, 2$, of (10) with $\hat{f}^{(d)}(t_0; h_{(f,d)}), \hat{\beta}_c^{(d)}(t_0; h_{(\beta_c,d)})$, and $\hat{\eta}_{lc}^{(d)}(t_0; h_{(\eta_{lc},d)})$, and obtain $\hat{\mathbf{D}}^*(t_0)$ by substituting $f(t_0), \eta_{lr}(t_0), \sigma^2(t_0)$, and $\rho_\varepsilon(t_0)$ of (12) with $\hat{f}(t_0; h_{(f,0)}), \hat{\eta}_{lr}(t_0; h_{(\eta_{lr},0)}), \hat{\sigma}^2(t_0; h_{(\sigma)})$, and $\hat{\rho}_\varepsilon(t_0; h_{(\rho)})$.

Remark 9. Because $\hat{\rho}_\varepsilon(t_0; h_{(\rho)})$ is obtained by smoothing the adjacent residuals for each subject, it works well asymptotically only when the numbers of repeated measurements $n_i, i = 1, \ldots, n$, are large, so that one can actually smooth $\hat{\varepsilon}_i(t_{ij_1}; h)\hat{\varepsilon}_i(t_{ij_2}; h)$ in the vicinity of t_0. When there is no measurement error—that is, $\sigma^2(t_0) = \rho_\varepsilon(t_0)$—$\hat{\sigma}^2(t_0; h_{(\sigma)})$ is practically a better estimator of $\rho_\varepsilon(t_0)$ than $\hat{\rho}_\varepsilon(t_0; h_{(\rho)})$. But when $\sigma^2(t_0) \neq \rho_\varepsilon(t_0)$ and $n_i, i = 1, \ldots, n$, are small, both $\hat{\sigma}^2(t_0; h_{(\sigma)})$ and $\hat{\rho}_\varepsilon(t_0; h_{(\rho)})$ are subject to large biases for the estimation of $\rho_\varepsilon(t_0)$. It is also interesting to note from (28) and (29) that the adequacy of $\hat{\sigma}^2(t_0; h_{(\sigma)})$ and $\hat{\rho}_\varepsilon(t_0; h_{(\rho)})$ also depends on the bandwidth h of $\hat{\beta}(\cdot; h)$.

4.2.2 *Consistency of the Estimators.*

Lemma 1. Suppose that, with the exception of assumption a, the conditions of Theorem 1 are satisfied, and $K(u)$ is continuously twice differentiable with a compact support on the real line.

a. If $\lim_{n\to\infty} h_{(\cdot;d)} = 0$ and $\lim_{n\to\infty} Nh^{2d+1}_{(\cdot;d)} = \infty$ for $d = 0, 1, 2$, then $\hat{f}^{(d)}(t_0; h_{(f,d)}) \to f^{(d)}(t_0)$, $\hat{\beta}_r^{(d)}(t_0; h_{(\beta_r,d)}) \to \beta_r^{(d)}(t_0)$, and $\hat{\eta}_{lr}^{(d)}(t_0; h_{(\eta_{lr},d)}) \to \eta_{lr}(t_0), l, r = 0, \ldots, k$, in probability as $n \to \infty$.
b. If $\lim_{n\to\infty} h_{(\cdot)} = 0$ and $\lim_{n\to\infty} Nh_{(\cdot)} = \infty$, then $\hat{\sigma}^2(t_0; h_{(\sigma)}) \to \sigma^2(t_0)$ and $\hat{\rho}_\varepsilon(t_0; h_{(\rho)}) \to \rho_\varepsilon(t_0)$ in probability as $n \to \infty$.

Proof. See Appendix B.

4.2.3 *Bandwidth Choices.* Similar to the estimation of $\beta(t_0)$, selection of appropriate bandwidths $h_{(\cdot,d)}$ and $h_{(\cdot)}$ are crucial for obtaining adequate estimators $\hat{\mathbf{B}}(t_0)$ and $\hat{\mathbf{D}}^*(t_0)$. Asymptotically, we require $h_{(\cdot,d)}$ and $h_{(\cdot)}$ to satisfy the conditions of Lemmas 1a and 1b. In practice, it is possible to choose these bandwidths subjectively by examining the plots of the corresponding fitted curves. Alternatively, the following data-driven procedures may provide some useful guidelines for some practical bandwidth choices. But statistical properties of these procedures have not been developed.

Following the heuristic suggestions of Rice and Silverman (1991), $h_{(f,0)}, h_{(\eta_{lr},0)}$, and $h_{(\sigma)}$ can be chosen by the "leave-one-subject-out" cross-validation bandwidths $h_{\mathrm{CV}(f,0)}, h_{\mathrm{CV}(\eta_{lr},0)}$, and $h_{\mathrm{CV}(\sigma)}$, such that $h_{\mathrm{CV}(f,0)}$ minimizes

$$\mathrm{CV}[\hat{f}(\cdot; h_{(f,0)})] = \int [\hat{f}(t; h_{(f,0)})]^2\, dt - \frac{2}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i} \hat{f}^{(-i)}(t_{ij}; h_{(f,0)}), \quad (30)$$

$h_{\mathrm{CV}(\eta_{lr},0)}$ minimizes

$$\mathrm{CV}[\hat{\eta}_{lr}(\cdot; h_{(\eta_{lr},0)})] = \sum_{i=1}^{n}\sum_{j=1}^{n_i} [X_{il}(t_{ij})X_{ir}(t_{ij}) - \hat{\eta}_{lr}^{(-i)}(t_{ij}; h_{(\eta_{lr},0)})]^2, \quad (31)$$

and $h_{\mathrm{CV}(\sigma)}$ minimizes

$$\mathrm{CV}[\hat{\sigma}^2(\cdot; h_{(\sigma)})] = \sum_{i=1}^{n}\sum_{j=1}^{n_i} [\hat{\varepsilon}_i^2(t_{ij}; h) - \hat{\sigma}^2_{(-i)}(t_{ij}; h_{(\sigma)})]^2, \quad (32)$$

where $\hat{f}^{(-i)}(t_{ij}; h_{(f,0)}), \hat{\eta}_{lr}^{(-i)}(t_{ij}; h_{(\eta_{lr},0)})$, and $\hat{\sigma}^2_{(-i)}(t_{ij}; h_{(\sigma)})$ are kernel estimators of $f(t_{ij}), \eta_{lr}(t_{ij})$, and $\sigma^2(t_{ij})$ as defined in (27), (28), and (29) based on the data with the observations of the ith subject left out.

An appropriate bandwidth $h_{(\rho)}$ of $\hat{\rho}_\varepsilon(t_0; h_{(\rho)})$ is generally more difficult to determine than that of $\hat{\sigma}^2(t_0; h_{(\sigma)})$. We do not yet have a satisfactory method for computing a suitable data-driven $h_{(\rho)}$. But by Lemma 1b, one possible choice is to use $h_{\mathrm{CV}(\sigma)}$ for $h_{(\rho)}$.

Bandwidth choices for derivative curve estimation with independent cross-sectional data are usually obtained by the procedures described by Härdle (1990) and Rice (1985), among others. Because of possible intrasubject correlations, these procedures may not be directly extended to the derivative estimators of Section 4.2.1. In general, Lemma 1a, (30), (31), and (32) imply that $h_{(f,d)}, h_{(\beta_r,d)}$, and $h_{(\eta_{lr},d)}$ may be chosen as

$$h_{(f,d)} = h_{\mathrm{CV}(f,0)}^{1/(2d+1)} c(f,d), \qquad h_{(\beta_r,d)} = h_{\mathrm{CV}}^{1/(2d+1)} c(\boldsymbol{\beta}, d),$$

and

$$h_{(\eta_{lr},d)} = h_{\mathrm{CV}(\eta_{lr},d)}^{1/(2d+1)} c(\eta_{lr}, d)$$

for some constants $c(f,d), c(\boldsymbol{\beta},d)$, and $c(\eta_{lr},d)$ that do not depend on the sample sizes. In this article we choose $h_{(f,d)}, h_{(\beta_r,d)}$, and $h_{(\eta_{lr},d)}$ for $d = 1, 2$ subjectively.

Remark 10. Because the cross-validation procedures (30), (31), and (32) are very computationally intensive, the potential benefits of deriving cross-validated bandwidths may be outweighed by the subsequent computational complexity. Thus for simplicity, one may use h_{CV} of (7) in place of $h_{(\cdot,d)}, d = 0, 1, 2$, and $h_{(\cdot)}$. In Section 6, this simple bandwidth choice produced satisfactory confidence regions for the simulated example.

4.3 Conservative Confidence Regions for Special Cases

As discussed in Remark 9, because of the difficulty in estimating $\rho_\varepsilon(t_0)$, we may not be able to estimate $\mathbf{D}^*(t_0)$ accurately when $\sigma^2(t_0) \neq \rho_\varepsilon(t_0)$ and $n_i, i = 1, \ldots, n$, are small. Consequently, confidence regions as given in (15), (16), (23), and (24) may not be easily computed. But in many situations, it is possible to construct conservative confidence regions by substituting the widths of (15), (16), (23), and (24) with estimates of their upper bounds. Here we construct conservative asymptotic confidence intervals and Bonferroni-type variability bands for $\beta_r(t)$; that is, $\mathbf{A}$ is selected such that $a_r = 1$ and $a_j = 0$ for all $j \neq r$.

4.3.1 Upper Bounds of Variance. Suppose that the error term of (2) is given by

$$\varepsilon_i(t_{ij}) = Z(t_{ij}) + W_{ij}, \tag{33}$$

where $W_{ij}, i = 1, \ldots, n, j = 1, \ldots, n_i$, are iid measurement errors satisfying $E(W_{ij}) = 0$ and $E(W_{ij}^2) \leq \infty, Z(t_{ij})$ are realizations of a mean 0 stochastic process $Z(t)$, and $\mathrm{cov}(Z(t_1), Z(t_2))$ is continuous in R^2. Because we do not require $Z(t)$ to be a stationary process, (33) is a generalization of the error processes considered by Moyeed and Diggle (1994) and Zeger and Diggle (1994). A direct implication of (33) is that $|\rho_\varepsilon(t)| \leq \sigma^2(t)$ for all $t \in \mathcal{S}(f)$. Then, by (11) and (12),

$$\mathbf{A}^T \mathbf{D}^*(t_0) \mathbf{A} \leq V_r(t_0),$$

where

$$V_r(t_0) = e_r^2(t_0) \eta_{rr}(t_0) \sigma^2(t_0) [\mu_2(K)(f(t_0))^{-1} + \lambda h_0]$$

and $e_r(t_0)$ is the rth diagonal element of $E_{\mathbf{XX}^T}(t_0)$. A kernel estimator $\hat{V}_r(t_0)$ of $V_r(t_0)$ can be obtained by substituting $e_r(t_0), \eta_{rr}(t_0), \sigma^2(t_0)$, and $f(t_0)$ with their corresponding kernel estimators of Section 4.2.1.

4.3.2 Conservative Confidence Intervals. If $h = O(N^{-1/5})$, then a $1 - \alpha$ conservative confidence interval of $\beta_r(t_0)$ with bias correction is given by $(\tilde{L}_{(\alpha,r)}(t_0), \tilde{U}_{(\alpha,r)}(t_0))$ such that $\tilde{L}_{(\alpha,r)}(t_0)$ and $\tilde{U}_{(\alpha,r)}(t_0)$ satisfy

$$[\hat{\beta}_r(t_0; h) - (Nh)^{-1/2} \hat{B}_r(t_0)] \pm z_{\alpha/2} (Nh)^{-1/2} [\hat{V}_r(t_0)]^{1/2}, \tag{34}$$

where $\hat{\beta}_r(t_0; h)$ and $\hat{B}_r(t_0)$ are the rth components of $\boldsymbol{\beta}(t_0; h)$ and $\hat{\mathbf{B}}(t_0)$. Similarly, if $h = o(N^{-1/5})$, then a $1 - \alpha$ conservative confidence interval of $\beta_r(t_0)$ without bias correction is given by $(\tilde{L}^*_{(\alpha,r)}(t_0), \tilde{U}^*_{(\alpha,r)}(t_0))$ such that $\tilde{L}^*_{(\alpha,r)}(t_0)$ and $\tilde{U}^*_{(\alpha,r)}(t_0)$ satisfy

$$\hat{\beta}_r(t_0; h) \pm z_{\alpha/2} (Nh)^{-1/2} [\hat{V}_r(t_0)]^{1/2}. \tag{35}$$

4.3.3 Conservative Bonferroni-Type Bands. Let $\xi_1, \ldots, \xi_{M+1}$ be the grid points as defined in (17) and let $(\tilde{L}^{(I)}_{(\alpha/(M+1),r)}(t), \tilde{U}^{(I)}_{(\alpha/(M+1),r)}(t))$ and $(\tilde{L}^{*(I)}_{(\alpha/(M+1),r)}(t), \tilde{U}^{*(I)}_{(\alpha/(M+1),r)}(t))$, with $\xi_j \leq t \leq \xi_{j+1}$, be the linear interpolations of $(\tilde{L}_{(\alpha/(M+1),r)}(t), \tilde{U}_{(\alpha/(M+1),r)}(t))$ and $(\tilde{L}^*_{(\alpha/(M+1),r)}(t), \tilde{U}^*_{(\alpha/(M+1),r)}(t))$ based on ξ_j and ξ_{j+1}. If $\beta_r(t)$ satisfies (19), then our $1 - \alpha$ conservative Bonferroni-type confidence bands are given by (23), with $(\hat{l}^{(I)}_\alpha(t), \hat{u}^{(I)}_\alpha(t))$ replaced by

$$(\hat{l}^{(B)}_\alpha(t), \hat{u}^{(B)}_\alpha(t)) = \begin{cases} (\tilde{L}^{(I)}_{(\alpha/(M+1),r)}(t), \tilde{U}^{(I)}_{(\alpha/(M+1),r)}(t)), \\ \quad \text{if } h = O(N^{-1/5}) \\ (\tilde{L}^{*(I)}_{(\alpha/(M+1),r)}(t), \tilde{U}^{*(I)}_{(\alpha/(M+1),r)}(t)), \\ \quad \text{if } h = o(N^{-1/5}). \end{cases}$$

Similarly, if $\beta_r(t)$ satisfies (20), then $1 - \alpha$ conservative bands can be constructed using (24), with $(\hat{l}^{(I)}_\alpha(t), \hat{u}^{(I)}_\alpha(t))$ replaced by $(\hat{l}^{(B)}_\alpha(t), \hat{u}^{(B)}_\alpha(t))$.

5. APPLICATION TO CD4 DEPLETION IN HIV INFECTION

Because higher levels of CD4 cell percentages are typically beneficial to HIV-infected persons, the main objective of our analysis is to determine the trend of CD4 percentage depletion since HIV infection and evaluate the effects of preinfection CD4 percentage and smoking status on the mean depletion of CD4 percentages over time. For the purpose of demonstration and simplicity, the possible effects of other available covariates are omitted.

The Multicenter AIDS Cohort Study (MACS) evaluated 400 homosexual men who became infected by HIV virus while being followed between 1984 and 1991. Details of the design and methods of the study have been described by Kaslow et al. (1987). Observations including CD4 cell percent, cigarette smoking status, scores of psychological tests, and laboratory results were recorded from each individual at 6-month scheduled visits. Here $t_{ij}, j = 1, \ldots, n_i$, denotes the time length in years between seroconversion and the jth measurement of the ith individual after the infection. For various reasons, some individuals missed scheduled visits. Each person was infected randomly during the study. This leads to unequal numbers of repeated measurements n_i and different time points t_{ij} for each individual. The number of repeated measurements per subject ranges from 1 to 14, with a median of 6 and a mean of 6.57.

To ensure a simple biological interpretation, we denote X_{i1} to be the centered preinfection CD4 percentage of the ith individual. This value is computed by subtracting the mean preinfection CD4 percentage of the entire sample from the ith individual's actual preinfection CD4 percentage. This type of centering, commonly used in classical linear models, does not affect the practical applicability of the procedures of Sections 2, 3, and 4. We denote $X_{i2}(t_{ij})$ to be 1 or 0 if the ith person is a smoker or a nonsmoker at time t_{ij}, and Y_{ij} to be the CD4 percentage of the ith individual at time t_{ij}. Fitting model (2) to the data, $\beta_0(t), \beta_1(t)$, and $\beta_2(t)$ can be viewed as the baseline CD4 effect and the effects of preinfection CD4 percentage and cigarette smoking on the postinfection CD4 percentage. We computed the kernel estimators of $\beta_0(t), \beta_1(t)$, and $\beta_2(t)$ using the Epanechnikov kernel and the cross-validated bandwidth $h_{\text{CV}} = .46$. Other choices of kernel functions, such as the standard Gaussian kernel, gave similar results. For computational simplicity, we used the Epanechnikov kernel and $h_{\text{CV}} = .46$ for the kernel estimators of $f(t), f'(t), \beta_r'(t), \beta_r''(t), \eta_{lr}(t), \eta_{lr}'(t)$, and $\sigma^2(t)$ as described in Section 4.2.1. Estimates of λ and h_0 of (11) were computed by $N^{-6/5}\sum_{i=1}^{n} n_i^2$ and $h_{\text{CV}}N^{1/5}$.

The connected lines of Figure 1(a) show the repeatedly measured CD4 cell percentages versus time since HIV infection (in years). Figure 1, (b)–(d), gives the kernel estimated curves of $\beta_r(t), r = 0, 1, 2$, and their corresponding simultaneous 95% confidence bands, which were constructed using the bias-corrected conservative Bonferroni-type bands of Section 4.3.3 with $M = 138$ and assuming that all of the curves satisfy (19) with $c_1 = 3$. Despite the conservativeness of our confidence bands, these figures still give a clear indication that the population CD4 cell percentage generally decreases over time. The preinfection CD4 cell percentage has obvious positive effect on the postinfection CD4 cell percentages, at least for the first 3 years

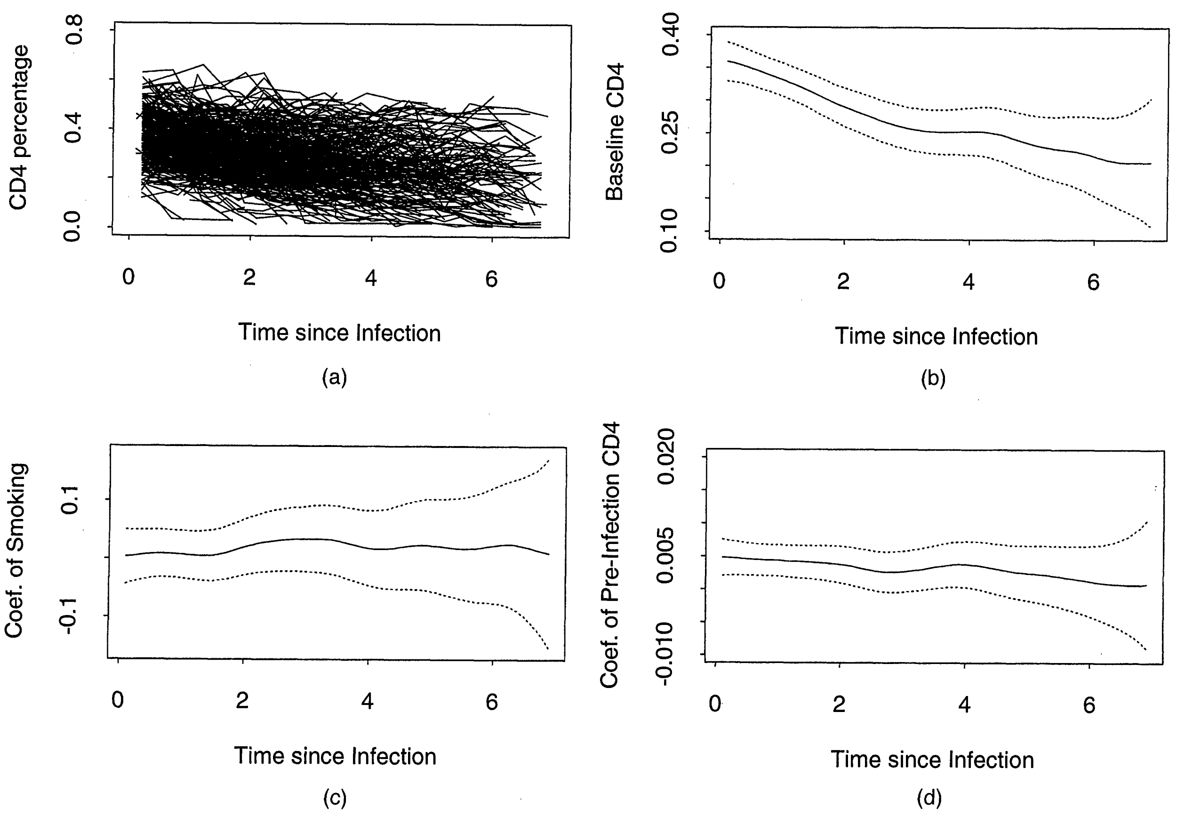

Figure 1. CD4 Cell Data. (a) Plot of repeatedly measured CD4 percentages versus time (in years) since HIV infection, (b) estimated baseline CD4 percentage curve, $\hat{\beta}_0(t; h_{CV})$, and its 95% conservative Bonferroni-type simultaneous band; (c) estimated smoking effect, $\hat{\beta}_2(t; h_{CV})$, and its 95% conservative Bonferroni-type simultaneous band; (d) estimated effect of preinfection CD4 percentage, $\hat{\beta}_1(t; h_{CV})$, and its 95% conservative Bonferroni-type simultaneous band. All of the confidence bands were computed with bias correction.

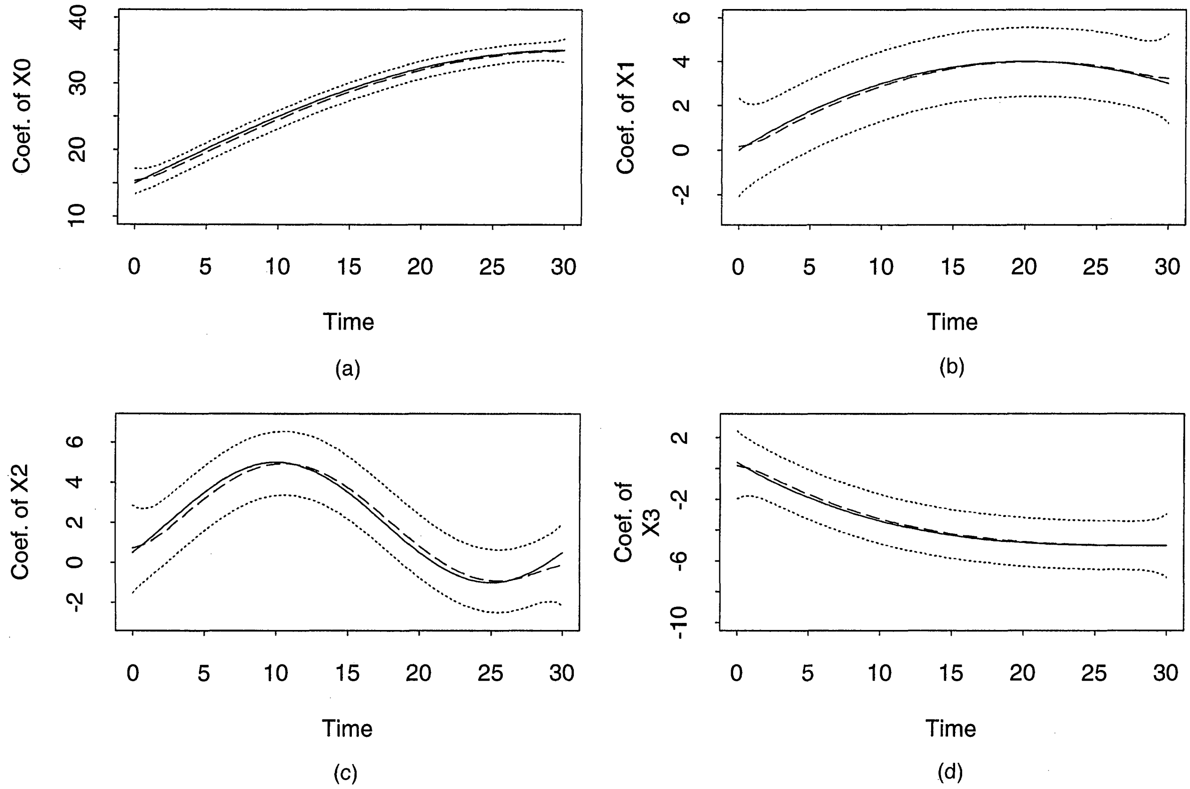

Figure 2. The Actual Functions $\beta_l(t)$, $l = 0, \ldots, 3$ (——), the Mean Estimated Curves of $\beta_l(t)$ (- - -), and the Averages of Their 95% Conservative Bonferroni-Type Simultaneous Bands ($\cdots$) Over all of the Simulated Datasets. All the estimated curves were computed based on the Epanechnikov kernel and their cross-validated bandwidths. The confidence bands were computed with bias correction. (a) β_0; (b) β_1; (c) β_2; (d) β_3

since infection. But this positive effect appears to be weakened over time, suggesting that higher preinfection CD4 percentage may not provide a lasting benefit. Based on the confidence band of Figure 1(c), we could not detect any significant effect of cigarette smoking, although a slight positive association is shown by the estimated curve of $\beta_2(t)$. This may imply that cigarette smoking has a minor effect on the postinfection CD4 percentage or may simply indicate that the effects of cigarette smoking can not be detected due to the use of conservative Bonferroni bands.

6. MONTE CARLO SIMULATION

For simplicity, we consider model (2) with a time-independent covariate $\mathbf{X} = (1, X_1, X_2, X_3)^T$, where X_1 and X_2 are two Bernoulli random variables and X_3 is a N(0, .25) random variable. The coefficient curves are given by

$$\beta_0(t) = 15 + 20\sin\left(\frac{t\pi}{60}\right), \qquad \beta_1(t) = 4 - \left(\frac{t-20}{10}\right)^2,$$

Table 1. Estimated Coverage Probabilities of 95% Pointwise Intervals (With Bias Correction) for $\beta_r(t)$, $r = 0, \ldots, 3$, at Eight Different Time Points

	Time point							
	1.0	5.0	9.0	13.0	17.0	21.0	25.0	27.0
$\beta_0(t)$	.999	.939	.946	.965	.983	.989	.998	.999
$\beta_1(t)$	.998	.996	.998	.999	1.000	1.000	1.000	. 998
$\beta_2(t)$	.998	.989	.996	1.000	.993	.995	.999	1.000
$\beta_3(t)$	.996	.994	1.000	.999	1.000	1.000	1.000	1.000

$$\beta_2(t) = 2 - 3\cos\left(\frac{(t-25)\pi}{15}\right)^2,$$

and

$$\beta_3(t) = -5 + \frac{(30-t)^3}{5{,}000}.$$

A simple random sample of 200 subjects, $\mathbf{X}_i, i = 1, \ldots, 200$, was generated for (X_1, X_2, X_3) based on the joint density

$$f(x_1, x_2, x_3) = \frac{.5}{(2\pi)^{1/2}} \exp(-2x_3^2) 1_{\{0,1\}}(x_1) \times 1_{\{0,1\}}(x_2) 1_{(-\infty,\infty)}(x_3).$$

To create design time points that are similar to the MACS of Section 5, we generated 30 equally spaced "scheduled" time points and 200 random displacement points s_{i1} from the $U(0,1)$ distribution such that $s_{il} = s_{i1} + (l-1), l = 1, \ldots, 30$. In addition, each "scheduled" time point s_{il} had a probability of 60% of being randomly missing. The remaining observed time points were denoted by t_{ij}. This led to unequal numbers of repeated measurements n_i and different observed time points t_{ij} per subject. The random errors $\varepsilon_i(t_{ij})$ were generated according to the mean 0 Gaussian process with covariance matrix

$$\operatorname{cov}[\varepsilon_{i_1}(t_{i_1 j_1}), \varepsilon_{i_2}(t_{i_2 j_2})] = \begin{cases} 4\exp(-|t_{i_1 j_1} - t_{i_2 j_2}|) & \text{if } i_1 = i_2 \\ 0 & \text{if } i_1 \neq i_2. \end{cases}$$

Table 2. Estimated Coverage Probabilities of 95% Bonferroni Bands (With Bias Correction) for $\beta_r(\cdot)$, r = 0, ..., 3, Based on Different Bandwidth Choices

	Bandwidth								
	.05	.10	.50	h_{CV}	1.50	2.50	3.50	4.50	5.50
$\beta_0(\cdot)$	.382	.120	1.000	1.000	1.000	.980	.371	.001	0
$\beta_1(\cdot)$	.405	.158	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\beta_2(\cdot)$	.383	.145	1.000	1.000	1.000	1.000	.999	.981	.780
$\beta_3(\cdot)$	.381	.154	1.000	1.000	1.000	1.000	1.000	1.000	1.000

NOTE: The cross-validated bandwidths ranged from .50 to 1.50.

The outcomes Y_{ij} were obtained by substituting the observed $(t_{ij}, \mathbf{X}_i, \varepsilon_i(t_{ij}))$ and the foregoing coefficient curves into (2).

This simulation process was repeated 1,000 times, so that independent replications of $\{(Y_{ij}, \mathbf{X}_i(t_{ij}), t_{ij})\colon 1 \le i \le n, 1 \le j \le n_i\}$ were generated. For each simulated sample, kernel estimators of $\beta_l(t), l = 0, \ldots, 3$, and their 95% bias corrected confidence regions were computed using the Epanechnikov kernel, the confidence procedures of Section 4.3, the corresponding cross-validated bandwidth h_{CV}, and a range of other bandwidth choices. Among the 1,000 simulated datasets, the cross-validated bandwidths have a range between .50 and 1.50 and a median of approximately 1.00. Estimators based on other kernel functions, such as the standard Gaussian kernel and the uniform kernel, gave similar results.

Figure 2 shows the actual functions $\beta_l(t), l = 0, \ldots, 3$, in solid lines and the mean estimated curves and the averages of their corresponding 95% Bonferroni-type bands with bias correction over all the simulated datasets in dashed and dotted lines. For each simulated dataset, the cross-validated bandwidth was used for computing the estimated curves. From these plots, we see that on average, the kernel estimators and cross-validation bandwidths give generally satisfactory results.

The entries of Table 1 show the observed coverage probabilities of 95% pointwise confidence intervals of $\beta_l(t), l = 0, \ldots, 3$, constructed using (34) and the cross-validated bandwidths at eight time points. As expected, high coverage probabilities are observed for $\beta_l(t), l = 0, \ldots, 3$, at all eight different time points. Assuming that $\beta_l(t)$ satisfy (19) with $c_1 = 3$, Table 2 shows the observed coverage probabilities of 95% bias-corrected Bonferroni bands of $\beta_r(t)$ constructed using the method of Section 4.3.3 with $M = 151$, the cross-validated bandwidths, and a range of other bandwidth choices. From this table, we see that when the cross-validated bandwidths are used, the conservative Bonferroni bands yield high coverage probabilities. However, poor coverage probabilities appear at extremely small or large bandwidths.

Table 3. Estimated Coverage Probabilities of 95% Bonferroni Bands (With Bias Correction) for $\beta_r(\cdot)$, r = 0, ..., 3, Based on Different Bandwidth Choices When the Intracorrelations Were Ignored; That is, $\lambda = 0$

	Bandwidth								
	.05	.10	.50	h_{CV}	1.50	2.50	3.50	4.50	5.50
$\beta_0(\cdot)$	.373	.084	1.000	.996	.987	.698	.001	0	0
$\beta_1(\cdot)$	.396	.125	1.000	1.000	.999	.999	.996	.988	.980
$\beta_2(\cdot)$	.376	.120	1.000	1.000	1.000	.969	.541	.007	0
$\beta_3(\cdot)$	.373	.122	1.000	1.000	1.000	1.000	1.000	.978	.873

NOTE: The cross-validated bandwidths ranged from .50 to 1.50.

Because the asymptotic properties of $\hat{\beta}_r(t; h)$ are equivalent to that with independent cross-sectional data when $\lambda = 0$, it is interesting to check whether the Bonferroni bands for $\beta_r(t)$ have the same high coverage probabilities when the intracorrelations are ignored. Table 3 shows the observed coverage probabilities of 95% Bonferroni bands for $\beta_r(t)$ constructed using the same method as Table 2 but assuming that $\lambda = 0$. These coverage probabilities stay the same as those given in Table 2 when the cross-validated bandwidths are used. But for bandwidths that are greater or equal to 2.50, some coverage probabilities of Table 3 are significantly lower than their counterparts in Table 2. Given that cross-validation bandwidths often lead to undersmoothness in practice, the intracorrelations may not be simply ignored in longitudinal estimation.

APPENDIX A: ASYMPTOTIC NORMALITY

We state and prove a general asymptotic normality result for $\hat{\beta}(t; h)$. For any interior point s of the support $\mathcal{S}(f)$, it can be shown by (6) and direct calculation (or eq. 4.2 of Hoover et al. 1998) that when n is sufficiently large,

$$(1 + o_p(1))[\hat{\beta}(s; h) - \beta(s)] = (f(s))^{-1} E_{\mathbf{X}\mathbf{X}^T}^{-1}(s) \hat{\mathbf{R}}(s; h), \quad \text{(A.1)}$$

where $\hat{\mathbf{R}}(s; h) = (\hat{R}_0(s; h), \ldots, \hat{R}_k(s; h))^T$ is a $k+1$ column vector such that

$$\hat{\mathbf{R}}(s; h) = (Nh)^{-1} \sum_{i=1}^{n} \{\mathbf{X}_i^T \mathbf{K}_i(s; h)[\mathbf{Y}_i - \mathbf{X}_i \beta(s)]\}. \quad \text{(A.2)}$$

Let $\mathbf{s} = (s_1, \ldots, s_J)$, $J \ge 1$, be a set of distinct interior points in the support of f and let $\hat{\mathbf{R}}(\mathbf{s}; h) = (\hat{\mathbf{R}}^T(s_1; h), \ldots, \hat{\mathbf{R}}^T(s_J; h))^T$. It suffices to study the asymptotic distribution of $\hat{\mathbf{R}}(\mathbf{s}; h)$.

We first state and prove a useful technical lemma.

Lemma A1. Suppose that assumptions a–f are satisfied, $\hat{\mathbf{R}}(\mathbf{s}; h)$ is defined in (A.2), and $\eta_{lr}(s_1, s_2)$ and $\rho_\varepsilon(s_1, s_2)$ are continuous in R^2. When n is sufficiently large,

$$E[(Nh)^{1/2} \hat{\mathbf{R}}(\mathbf{s}; h)] = \mathbf{b}(\mathbf{s}) + o_p(1)$$

and

$$\operatorname{cov}[(Nh)^{1/2} \hat{\mathbf{R}}(\mathbf{s}; h)] = \mathbf{D}(\mathbf{s}) + o_p(1), \quad \text{(A.3)}$$

where

$$\mathbf{b}(\mathbf{s}) = (b_0(s_1), \ldots, b_k(s_1), \ldots, b_0(s_J), \ldots, b_k(s_J))^T \quad \text{(A.4)}$$

with $b_l(s), l = 0, \ldots, k$ defined in (9).

$$\mathbf{D}(\mathbf{s}) = \begin{pmatrix} \mathbf{D}(s_1, s_1) & \cdots & \mathbf{D}(s_1, s_J) \\ \vdots & \vdots & \vdots \\ \mathbf{D}(s_J, s_1) & \cdots & \mathbf{D}(s_J, s_J) \end{pmatrix},$$

$$\mathbf{D}(s_1, s_2) = \begin{pmatrix} D_{00}(s_1, s_2) & \cdots & D_{0k}(s_1, s_2) \\ \vdots & \vdots & \vdots \\ D_{k0}(s_1, s_2) & \cdots & D_{kk}(s_1, s_2) \end{pmatrix},$$

and

$$D_{lr}(s_1, s_2) = \begin{cases} \sigma^2(s_1)\eta_{lr}(s_1)f(s_1)\mu_2(K) + \lambda h_0 \rho_\varepsilon(s_1)\eta_{lr}(s_1, s_2)f^2(s_1), \\ \quad \text{if } s_1 = s_2 \\ \lambda h_0 \rho_\varepsilon(s_1, s_2)\eta_{lr}(s_1, s_2)f(s_1)f(s_2), \\ \quad \text{if } s_1 \neq s_2, \end{cases} \tag{A.5}$$

with $\eta_{lr}(s_1, s_2) = E[X_{il}(t_{ij_1})X_{ir}(t_{ij_2})|t_{ij_1} = s_1, t_{ij_2} = s_2]$.

Proof. By (2) and (A.2), it can be shown by straightforward calculation that the lth element of $\hat{\mathbf{R}}(s; h)$ can be written as

$$\hat{R}_l(s; h) = (Nh)^{-1} \sum_{i=1}^{n} \psi_{il}(s; h), \qquad l = 0, \ldots, k, \tag{A.6}$$

where $\psi_{il}(s; h) = \sum_{j=1}^{n_i} [\xi_{il}(s, t_{ij})K((s - t_{ij})/h)]$ and

$$\xi_{il}(s, t_{ij}) = \sum_{r=0}^{k} \{X_{il}(t_{ij})X_{ir}(t_{ij})[\beta_r(t_{ij}) - \beta_r(s)]\} + X_{il}(t_{ij})\varepsilon_i(t_{ij}).$$

Then (A.6) implies that $\hat{\mathbf{R}}(\mathbf{s}; h)$ is a sum of independent vectors; that is,

$$\hat{\mathbf{R}}(\mathbf{s}; h) = (Nh)^{-1} \sum_{i=1}^{n} \boldsymbol{\Psi}_i(\mathbf{s}), \tag{A.7}$$

where $\boldsymbol{\Psi}_i(\mathbf{s})$ is a $J(k+1)$ column vector such that

$$\boldsymbol{\Psi}_i(\mathbf{s}) = (\psi_{i0}(s_1), \ldots, \psi_{ik}(s_1), \ldots, \psi_{i0}(s_J), \ldots, \psi_{ik}(s_J))^T.$$

Because the design points are independent, direct calculation and the change of variables show that

$$\begin{aligned} E(\psi_{il}(s)) &= \sum_{j=1}^{n_i} \int E(\xi_{il}(s, t_{ij})|t_{ij} = v)K\Big(\frac{s - v}{h}\Big) f(v)\, dv \\ &= n_i h \sum_{r=0}^{k} \int [\beta_r(s - hu) - \beta_r(s)] \\ &\quad \times \eta_{lr}(s - hu)f(s - hu)K(u)\, du. \end{aligned}$$

Then, by (A.6) and assumptions a, c, and f, and taking the Taylor expansions on the right side of the foregoing equation, we have $E[(Nh)^{1/2}\hat{\mathbf{R}}(\mathbf{s}; h)] = \mathbf{b}(\mathbf{s}) + o(1)$.

For the covariance of $(Nh)^{1/2}\hat{\mathbf{R}}(\mathbf{s}; h)$, because

$$\operatorname{cov}[\hat{R}_l(s_1; h), \hat{R}_r(s_2; h)] = E[\hat{R}_l(s_1; h)\hat{R}_r(s_2; h)] - E[\hat{R}_l(s_1; h)]E[\hat{R}_r(s_2; h)],$$

we need to compute the right terms of the following equation:

$$\begin{aligned} &E\left\{\left[(Nh)^{-1/2} \sum_{i=1}^{n} \psi_{il}(s_1)\right] \left[(Nh)^{-1/2} \sum_{i=1}^{n} \psi_{ir}(s_2)\right]\right\} \\ &= (Nh)^{-1} \left\{\sum_{i=1}^{n} E[\psi_{il}(s_1)\psi_{ir}(s_2)] + \sum_{i_1 \neq i_2} E[\psi_{i_1 l}(s_1)\psi_{i_2 r}(s_2)]\right\}. \end{aligned} \tag{A.8}$$

For the first term on the right side of (A.8), we consider the further decomposition

$$\begin{aligned} &\psi_{il}(s_1)\psi_{ir}(s_2) \\ &= \sum_{j=1}^{n_i} \xi_{il}(s_1, t_{ij})\xi_{ir}(s_2, t_{ij})K\Big(\frac{s_1 - t_{ij}}{h}\Big)K\Big(\frac{s_2 - t_{ij}}{h}\Big) \\ &\quad + \sum_{j_1 \neq j_2} \xi_{il}(s_1, t_{ij_1})\xi_{ir}(s_2, t_{ij_2})K\Big(\frac{s_1 - t_{ij_1}}{h}\Big)K\Big(\frac{s_2 - t_{ij_2}}{h}\Big). \end{aligned} \tag{A.9}$$

By assumption a, the change of variables, and the fact that $\varepsilon_i(\cdot)$ is a mean 0 stochastic process independent of $\mathbf{X}_i(t_{ij})$, it can be shown by direct calculation that, as $n \to \infty$,

$$\begin{aligned} &E[\xi_{il}(s_1, t_{ij})\xi_{ir}(s_2, t_{ij})|t_{ij} = v] \\ &= \sum_{c=0}^{k} \{[\beta_c(v) - \beta_c(s_1)][\beta_c(v) - \beta_c(s_2)] \\ &\quad \times E[X_{il}(t_{ij})X_{ic}^2(t_{ij})X_{ir}(t_{ij})|t_{ij} = v]\} \\ &\quad + \sigma^2(v)E[X_{il}(t_{ij})X_{ir}(t_{ij})|t_{ij} = v] \\ &\quad + \sum_{c_1 \neq c_2} \{[\beta_{c_1}(v) - \beta_{c_1}(s_1)][\beta_{c_2}(v) - \beta_{c_2}(s_2)] \\ &\quad \times E[X_{il}(t_{ij})X_{ic_1}(t_{ij})X_{ir}(t_{ij})X_{ic_2}(t_{ij})|t_{ij} = v]\} \to \\ &\quad \sigma^2(s_c)\eta_{lr}(s_c), \quad \text{if } v \to s_c, \qquad c = 1, 2. \end{aligned}$$

Then it follows from the foregoing equation that

$$\begin{aligned} &E\left[\sum_{j=1}^{n_i} \xi_{il}(s_1, t_{ij})\xi_{ir}(s_2, t_{ij})K\Big(\frac{s_1 - t_{ij}}{h}\Big) K\Big(\frac{s_2 - t_{ij}}{h}\Big)\right] \\ &= \sum_{j=1}^{n_i} \int E[\xi_{il}(s_1, t_{ij})\xi_{ir}(s_2, t_{ij})|t_{ij} = v] \\ &\quad \times K\Big(\frac{s_1 - v}{h}\Big)K\Big(\frac{s_2 - v}{h}\Big) f(v)\, dv \\ &= \begin{cases} n_i h \sigma^2(s_1)\eta_{lr}(s_1)f(s_1) \int K^2(u)\, du + o(n_i h), & \text{if } s_1 = s_2 \\ o(n_i h), & \text{if } s_1 \neq s_2. \end{cases} \end{aligned} \tag{A.10}$$

Similarly, it can be shown by direct calculation that as $n \to \infty$, $v_1 \to s_1$, and $v_2 \to s_2$,

$$\begin{aligned} &E[\xi_{il}(s_1, t_{ij_1})\xi_{ir}(s_2, t_{ij_2})|t_{ij_1} = v_1, t_{ij_2} = v_2] \\ &= \sum_{c=0}^{k} \{[\beta_c(v_1) - \beta_c(s_1)][\beta_c(v_2) - \beta_c(s_2)] \\ &\quad \times E[X_{il}(t_{ij_1})X_{ic}(t_{ij_1})X_{ir}(t_{ij_2}) \\ &\quad \times X_{ic}(t_{ij_2})|t_{ij_1} = v_1, t_{ij_2} = v_2]\} \\ &\quad + \rho_\varepsilon(v_1, v_2)E[X_{il}(t_{ij_1})X_{ir}(t_{ij_2})|t_{ij_1} = v_1, t_{ij_2} = v_2] \\ &\quad + \sum_{c_1 \neq c_2} \{[\beta_{c_1}(v_1) - \beta_{c_1}(s_1)][\beta_{c_2}(v_2) - \beta_{c_2}(s_2)] \\ &\quad \times E[X_{il}(t_{ij_1})X_{ic_1}(t_{ij_1})X_{ir}(t_{ij_2}) \\ &\quad \times X_{ic_2}(t_{ij_2})|t_{ij_1} = v_1, t_{ij_2} = v_2]\} \to \\ &\quad \begin{cases} \rho_\varepsilon(s_1, s_2)\eta_{lr}(s_1, s_2), & \text{if } s_1 \neq s_2 \\ \rho_\varepsilon(s_1)\eta_{lr}(s_1, s_1), & \text{if } s_1 = s_2, \end{cases} \end{aligned}$$

and, consequently, the expectation of the second term on the right side of (A.9) is

$$E\left[\sum_{j_1\neq j_2}\xi_{il}(s_1,t_{ij_1})\xi_{ir}(s_2,t_{ij_2})K\left(\frac{s_1-t_{ij_1}}{h}\right)K\left(\frac{s_2-t_{ij_2}}{h}\right)\right]$$

$$=\sum_{j_1\neq j_2}\left\{\int\int E[\xi_{il}(s_1,t_{ij_1})\xi_{ir}(s_2,t_{ij_2})|t_{ij_1}=v_1,t_{ij_2}=v_2]\right.$$

$$\left.\times K\left(\frac{s_1-v_1}{h}\right)K\left(\frac{s_2-v_2}{h}\right)f(v_1)f(v_2)\,dv_1\,dv_2\right\}$$

$$=\begin{cases} h^2n_i(n_i-1)\rho_\varepsilon(s_1,s_2)f(s_1)f(s_2)\eta_{lr}(s_1,s_2) \\ \quad +o(h^2n_i(n_i-1)), \\ \qquad \text{if } s_1\neq s_2 \\ h^2n_i(n_i-1)\rho_\varepsilon(s_1)f^2(s_1)\eta_{lr}(s_1,s_1)+o(h^2n_i(n_i-1)), \\ \qquad \text{if } s_1=s_2. \end{cases} \tag{A.11}$$

Combining (A.9), (A.10), and (A.11), it follows immediately that when n is sufficiently large,

$$(Nh)^{-1}\sum_{i=1}^{n}E[\psi_{il}(s_1)\psi_{ir}(s_2)]$$

$$=\begin{cases} \sigma^2(s_1)\eta_{lr}(s_1)f(s_1)\int K^2(u)\,du \\ \quad +o\left(hN^{-1}\left(\sum_{i=1}^n n_i^2-N\right)\right) \\ \quad +hN^{-1}\left(\sum_{i=1}^n n_i^2-N\right)\rho_\varepsilon(s_1)\eta_{lr}(s_1,s_1)f^2(s_1), \\ \qquad \text{if } s_1=s_2 \\ hN^{-1}\left(\sum_{i=1}^n n_i^2-N\right)\rho_\varepsilon(s_1,s_2)\eta_{lr}(s_1,s_2)f(s_1)f(s_2) \\ \quad +o\left(hN^{-1}\left(\sum_{i=1}^n n_i^2-N\right)\right), \\ \qquad \text{if } s_1\neq s_2. \end{cases} \tag{A.12}$$

Because $h=N^{-1/5}h_0$ and $\lim_{n\to\infty}N^{-6/5}\sum_{i=1}^n n_i^2=\lambda$, it is easy to see that as $n\to\infty$,

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

Similarly, we can show that for the second term on the right side of (A.8), there exists a positive constant $M<\infty$ such that

$$\left|(Nh)^{-1}\sum_{i_1\neq i_2}E[\psi_{i_1l}(s_1)\psi_{i_2r}(s_2)]-E\left[(Nh)^{-1/2}\sum_{i=1}^n\psi_{il}(s_1)\right]\right.$$

$$\left.\times\ E\left[(Nh)^{-1/2}\sum_{i=1}^n\psi_{ir}(s_2)\right]\right|$$

$$\leq Nh\left[N^{-2}\sum_{i=1}^n\left(n_i\sum_{i'\neq i}n_{i'}\right)-1\right]M\to 0,\quad \text{as}\quad n\to\infty. \tag{A.13}$$

The limit of (A.13) holds because, by assumption b,

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

Now, by assumption a and the calculations given in (A.8), (A.12), and (A.13), we have shown that for any interior points s_1 and s_2 in the support of f,

$$\operatorname{cov}[(Nh)^{1/2}\hat{R}_l(s_1;h),(Nh)^{1/2}\hat{R}_r(s_2;h)]$$
$$=D_{lr}(s_1,s_2)+o(1),\qquad l,r=0,\ldots,k.$$

This implies the assertion of the lemma.

Theorem A.1. Suppose that the conditions of Lemma A.1 are satisfied. With proper normalization, $\hat{\boldsymbol\beta}(\mathbf{s};h)-\boldsymbol\beta(\mathbf{s})$ has asymptotically a multivariate Gaussian distribution; that is,

$$(Nh)^{1/2}(\hat{\boldsymbol\beta}(\mathbf{s};h)-\boldsymbol\beta(\mathbf{s}))\to\mathcal{N}(\mathbf{B}(\mathbf{s}),\mathbf{D}^*(\mathbf{s}))$$
in distribution as $n\to\infty$,

where $\mathbf{B}(\mathbf{s})=(\mathbf{B}(s_1),\ldots,\mathbf{B}(s_J))^T$ with $\mathbf{B}(s)$ as defined in (10) and

$$\mathbf{D}^*(\mathbf{s})=\begin{pmatrix}\mathbf{D}^*(s_1,s_1) & \cdots & \mathbf{D}^*(s_1,s_J)\\ \vdots & \vdots & \vdots \\ \mathbf{D}^*(s_J,s_1) & \cdots & \mathbf{D}^*(s_J,s_J)\end{pmatrix}$$

with $\mathbf{D}^*(s_{r_1},\ s_{r_2})\ =\ f^{-1}(s_{r_1})f^{-1}(s_{r_2})E^{-1}_{\mathbf{XX}^T}(s_{r_1})\mathbf{D}(s_{r_1},s_{r_2})E^{-1}_{\mathbf{XX}^T}(s_{r_2})$ for $r_1,r_2=1,\ldots,J$.

Proof. By Assumption e, (A.6) and Lemma A.1, it is easy to verify that $\hat{\mathbf{R}}(\mathbf{s};h)$ satisfies the conditions of the Cramer–Wold theorem (cf. Serfling 1980, theorem of sec. 1.5.2) and the Lindeberg condition (cf. Serfling 1980, theorem A of sec. 1.9.2). Thus $(Nh)^{1/2}\hat{\mathbf{R}}(\mathbf{s};h)\to\mathcal{N}(\mathbf{b}(\mathbf{s}),\mathbf{D}(\mathbf{s}))$ in distribution as $n\to\infty$. The theorem then follows from (A.1) and the limiting distribution of $(Nh)^{1/2}\hat{\mathbf{R}}(\mathbf{s};h)$.

APPENDIX B: PROOF OF LEMMA 1

For brevity, we sketch the proofs for only $\hat{f}^{(1)}(t_0;\ h_{(f,1)})$, $\hat\eta_{lr}(t_0;h_{(\eta_{lr},0)})$, and $\hat\rho_\varepsilon(t_0;h_{(\rho)})$. The consistency of other estimators can be derived similarly with tedious but straightforward calculations.

Consistency of $\hat{f}^{(1)}(t_0;\ h_{(f,0)})$. Because subjects are independent, direct calculation, integration by parts and the change of variables show that as $n\to\infty$,

$$|E[\hat{f}^{(1)}(t_0;h_{(f,1)})]-f^{(1)}(t_0)|$$
$$=\left|\int K(u)[f^{(1)}(t_0-h_{(f,1)}u)-f^{(1)}(t_0)]\,du\right|\to 0$$

and

$$\operatorname{var}[\hat{f}^{(1)}(t_0;h_{(f,1)})]$$
$$=\ N^{-1}h^{-3}_{(f,1)}f(t_0)\left\{\int\ [K'(u)]^2\,du\right\}+o(N^{-1}h^{-3}_{(f,1)})\to 0.$$

Thus $\lim_{n\to\infty}E\{[\hat{f}^{(1)}(t_0;h_{(f,1)})-f^{(1)}(t_0)]^2\}=0$. Consequently, $\hat{f}^{(1)}(t_0;h_{(f,1)})\to f^{(1)}(t_0)$ in probability as $n\to\infty$.

Consistency of $\hat\eta_{lr}(t_0;\ h_{(\eta_{lr},0)})$. Write

$$\hat\nu_\eta(t_0;h_{(\eta_{lr},0)})=(Nh_{(\eta_{lr},0)})^{-1}$$
$$\times\sum_{i=1}^n\sum_{j=1}^{n_i}\left[X_{il}(t_{ij})X_{lr}(t_{ij})K\left(\frac{t_0-t_{ij}}{h_{(\eta_{lr},0)}}\right)\right].$$

Using the same method as in the proof of Lemma A, it is easy to show that $\hat{f}(t_0;h_{(\eta_{lr},0)})\to f(t_0)$ in probability as $n\to\infty$. Then, by (27), it suffices to show that $\hat\nu_\eta(t_0;h_{(\eta_{lr},0)})\to\eta_{lr}(t_0)f(t_0)$ in probability as $n\to\infty$.

By direct calculations, it can be verified that as $n \to \infty$,

$$\begin{aligned} &E[\hat{\nu}_\eta(t_0; h_{(\eta_{lr},0)})] \\ &= (N h_{\eta_{lr},0})^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{ \int E[X_{il}(t_{ij}) X_{ir}(t_{ij}) | t_{ij} = s] \right. \\ &\quad \left. \times K\left(\frac{t_0 - s}{h_{(\eta_{lr},0)}}\right) f(s)\, ds \right\} \to \eta_{lr}(t_0) f(t_0) \end{aligned}$$

and

$$\begin{aligned} &\operatorname{var}[\hat{\nu}_\eta(t_0, h_{(\eta_{lr},0)})] \\ &= (N h_{(\eta_{lr},0)})^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{ \int E[X_{il}^2(t_{ij}) X_{ir}^2(t_{ij}) | t_{ij} = s] \right. \\ &\quad \left. \times K^2\left(\frac{t_0 - s}{h_{(\eta_{lr},0)}}\right) f(s)\, ds \right\} + o(N^{-1} h_{(\eta_{lr},0)}^{-1}) \to 0. \end{aligned}$$

Then $E[\hat{\nu}_\eta(t_0; h_{(\eta_{lr},0)}) - \eta_{lr}(t_0) f(t_0)]^2 \to 0$ holds as $n \to \infty$, which further implies that $\hat{\nu}_\eta(t_0; h_{(\eta_{lr},0)}) \to \eta_{lr}(t_0) f(t_0)$ in probability as $n \to \infty$.

Consistency of $\hat{\rho}_\varepsilon(t_0; h_{(\rho)})$. Let

$$\begin{aligned} &\hat{\nu}_\rho(t_0; h_{(\rho)}) \\ &= \left(\sum_{i=1}^{n} n_i^2 - N\right)^{-1} h_{(\rho)}^{-2} \\ &\quad \times \left[\sum_{i=1}^{n} \sum_{j_1 \neq j_2} \hat{\varepsilon}_i(t_{ij_1}, h) \hat{\varepsilon}_i(t_{ij_2}, h) \right. \\ &\quad \left. \times K\left(\frac{t_0 - t_{ij_1}}{h_{(\rho)}}\right) K\left(\frac{t_0 - t_{ij_2}}{h_{(\rho)}}\right) \right] \end{aligned}$$

and

$$\begin{aligned} \hat{\nu}_\rho^*(t_0; h_{(\rho)}) &= \left(\sum_{i=1}^{n} n_i^2 - N\right)^{-1} h_{(\rho)}^{-2} \\ &\quad \times \left[\sum_{i=1}^{n} \sum_{j_1 \neq j_2} K\left(\frac{t_0 - t_{ij_1}}{h_{(\rho)}}\right) K\left(\frac{t_0 - t_{ij_2}}{h_{(\rho)}}\right) \right]. \end{aligned}$$

By (29), we have that $\hat{\rho}_\varepsilon(t_0; h_{(\rho)}) = \hat{\nu}_\rho(t_0; h_{(\rho)}) / \hat{\nu}_\rho^*(t_0; h_{(\rho)})$. It suffices to show that $\hat{\nu}_\rho^*(t_0; h_{(\rho)}) \to f^2(t_0)$ and $\hat{\nu}_\rho(t_0; h_{(\rho)}) \to \rho_\varepsilon(t_0) f^2(t_0)$ in probability as $n \to \infty$.

For the consistency of $\hat{\nu}_\rho^*(t_0; h_{(\rho)})$, we can show by direct calculations, such as those in the proof of Lemma A, that as $n \to \infty$,

$$\begin{aligned} &E[\hat{\nu}_\rho^*(t_0; h_{(\rho)})] \\ &= \left(\sum_{i=1}^{n} n_i^2 - N\right)^{-1} h_{(\rho)}^{-2} \\ &\quad \times \left[\sum_{i=1}^{n} \sum_{j_1 \neq j_2} \int\!\!\int K\left(\frac{t_0 - s_1}{h_{(\rho)}}\right) \right. \\ &\quad \left. \times K\left(\frac{t_0 - s_2}{h_{(\rho)}}\right) f(s_1) f(s_1)\, ds_1\, ds_2 \right] \to f^2(t_0) \end{aligned}$$

and

$$\begin{aligned} \operatorname{cov}[\hat{\nu}_\rho^*(t_0; h_{(\rho)})] &= \left(\sum_{i=1}^{n} n_i^2 - N\right)^{-1} h_{(\rho)}^{-2} \\ &\quad \times \left[\int K^2(u)\, du\right]^2 f^2(t_0) + o(1) \to 0. \end{aligned}$$

Thus as $n \to \infty$, $E[\hat{\nu}_\rho^*(t_0; h_{(\rho)}) - f^2(t_0)]^2 \to 0$, and, consequently, $\hat{\nu}_\rho^*(t_0; h_{(\rho)}) \to f^2(t_0)$ in probability.

To show the consistency of $\hat{\nu}_\rho(t_0; h_{(\rho)})$, we first define $\tilde{\varepsilon}_i(t_{ij}) = Y_{ij} - \mathbf{X}_i^T(t_{ij}) \boldsymbol{\beta}(t_{ij})$ and

$$\begin{aligned} &\tilde{\nu}_\rho(t_0; h_{(\rho)}) \\ &= \left(\sum_{i=1}^{n} n_i^2 - N\right)^{-1} h_{(\rho)}^{-2} \\ &\quad \times \sum_{i=1}^{n} \sum_{j_1 \neq j_2} \left[\tilde{\varepsilon}_i(t_{ij_1}) \tilde{\varepsilon}_i(t_{ij_2}) K\left(\frac{t_0 - t_{ij_1}}{h_{(\rho)}}\right) \right. \\ &\quad \left. \times K\left(\frac{t_0 - t_{ij_2}}{h_{(\rho)}}\right) \right]. \end{aligned}$$

By Theorem 1 and similar calculations in the consistency of $\hat{\nu}_\rho^*(t_0; h_{(\rho)})$, we can verify that there are constants $a_1 > 0$ and $a_2 > 0$ such that, as $n \to \infty$,

$$\begin{aligned} &\sup_{t_{ij_1} \in [t_0 - a_1, t_0 + a_1], t_{ij_2} \in [t_0 - a_2, t_0 + a_2]} |\hat{\varepsilon}_i(t_{ij_1}; h) \hat{\varepsilon}_i(t_{ij_2}; h) \\ &\quad - \tilde{\varepsilon}_i(t_{ij_1}) \tilde{\varepsilon}_i(t_{ij_2})| \to 0 \quad \text{in probability} \end{aligned}$$

and

$$\left(\sum_{i=1}^{n} n_i^2 - N\right)^{-1} h_{(\rho)}^{-2} \sum_{i=1}^{n} \sum_{j_1 \neq j_2} \left| K\left(\frac{t_0 - t_{ij_1}}{h_{(\rho)}}\right) K\left(\frac{t_0 - t_{ij_2}}{h_{(\rho)}}\right) \right|$$

is bounded in probability when n is sufficiently large. Thus

$$\begin{aligned} &|\hat{\nu}_\rho(t_0; h_{(\rho)}) - \tilde{\nu}_\rho(t_0; h_{(\rho)})| \\ &\leq \sup_{t_{ij_1} \in [t_0 - a_1, t_0 + a_1], t_{ij_2} \in [t_0 - a_2, t_0 + a_2]} |\hat{\varepsilon}_i(t_{ij_1}; h) \hat{\varepsilon}_i(t_{ij_2}; h) \\ &\quad - \tilde{\varepsilon}_i(t_{ij_1}) \tilde{\varepsilon}_i(t_{ij_2})| \left(\sum_{i=1}^{n} n_i^2 - N\right)^2 h_{(\rho)}^2 \\ &\quad \times \sum_{i=1}^{n} \sum_{j_1 \neq j_2} \left| K\left(\frac{t_0 - t_{ij_1}}{h_{(\rho)}}\right) K\left(\frac{t_0 - t_{ij_2}}{h_{(\rho)}}\right) \right| \to 0 \\ &\quad \text{in probability as} \quad n \to \infty. \end{aligned}$$

It suffices to show that $\tilde{\nu}_\rho(t_0; h_{(\rho)}) \to \rho_\varepsilon(t_0) f^2(t_0)$ in probability as $n \to \infty$.

Similar calculations as in the proof of the consistency of $\hat{\eta}_{lr}(t_0; h_{(\eta_{lr},0)})$ shows that, as $n \to \infty$, $E[\tilde{\nu}_\rho(t_0; h_{(\rho)})] \to \rho_\varepsilon(t_0) f^2(t_0)$ and $\operatorname{var}(\tilde{\nu}_\rho(t_0; h_{(\rho)})) \to 0$. Thus $\tilde{\nu}_\rho(t_0; h_{(\rho)}) \to \rho_\varepsilon(t_0) f^2(t_0)$ in probability as $n \to \infty$. This completes the proof.

[Received July 1996. Revised May 1998.]

REFERENCES

Altman, N. S. (1990), "Kernel Smoothing of Data With Correlated Errors," *Journal of the American Statistical Association*, 85, 749–759.

Bickel, P. J., and Rosenblatt, M. (1973), "On Some Global Measures of the Deviations of Density Function Estimates," *The Annals of Statistics*, 1, 1071–1091.

Davidian, M., and Giltinan, D. M. (1995), *Nonlinear Models for Repeated Measurement Data*, London: Chapman and Hall.

Diggle, P. J. (1988), "An Approach to the Analysis of Repeated Measurements," *Biometrics*, 44, 959–971.

Diggle, P. J., Liang, K. Y., and Zeger, S. L. (1994), *Analysis of Longitudinal Data*, Oxford, U.K.: Oxford University Press.

Green, P. J., and Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models—A Roughness Penalty Approach*, London: Chapman and Hall.

Eubank, R. L., and Speckman, P. L. (1993), "Confidence Bands in Nonparametric Regression," *Journal of the American Statistical Association*, 88, 1287–1301.

Hall, P., and Titterington, D. M. (1988), "On Confidence Bands in Nonparametric Density Estimation and Regression," *Journal of Multivariate Analysis*, 27, 228–254.

Hall, P., and Wehrly, T. E. (1991), "A Geometrical Method for Removing Edge Effects From Kernel-Type Nonparametric Regression Estimators," *Journal of the American Statistical Association*, 86, 665–672.

Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge, U.K.: Cambridge University Press.

Härdle, W., and Marron, J. S. (1991), "Bootstrap Simultaneous Error Bars for Nonparametric Regression," *The Annals of Statistics*, 19, 778–796.

Hart, J. D. (1991), "Kernel Regression Estimation With Time Series Errors," *Journal of the Royal Statistical Society*, Ser. B, 53, 173–187.

Hart, J. D., and Wehrly, T. E. (1986), "Kernel Regression Estimation Using Repeated Measurements Data," *Journal of the American Statistical Association*, 81, 1080–1088.

——— (1993), "Consistency of Cross-Validation When the Data Are Curves," *Stochastic Processes and Their Applications*, 45, 351–361.

Hastie, T. J., and Tibshirani, R. J. (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society*, Ser. B, 55, 757–796.

Hoover, D. R. (1990), "Subset Complement Addition Upper Bounds—An Improved Inclusion–Exclusion Method," *Journal of Statistical Planning and Inference*, 24, 195–202.

Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L. P. (in press), "Nonparametric Smoothing Estimates of Time-Varying Coefficient Models With Longitudinal Data," unpublished manuscript submitted to *Biometrika*.

Johnston, G. J. (1982), "Probabilities of Maximal Deviations for Nonparametric Regression Function Estimates," *Journal of Multivariate Analysis*, 12, 402–414.

Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F., and Rinaldo, C. R. (1987), "The Multicenter AIDS Cohort Study: Rationale, Organization and Selected Characteristics of the Participants," *American Journal of Epidemiology*, 126, 310–318.

Knafl, G., Sacks, J., and Ylvisaker, D. (1985), "Confidence Bands for Regression Functions," *Journal of the American Statistical Association*, 80, 683–691.

Liang, K. Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.

Moyeed, R. A., and Diggle, P. J. (1994), "Rates of Convergence in Semi-Parametric Modeling of Longitudinal Data," *Australian Journal of Statistics*, 36, 75–93.

Müller, H. G. (1993), "On the Boundary Kernel Method for Nonparametric Curve Estimation Near Endpoints," *Scandinavian Journal of Statistics*, 20, 313–328.

Park, L. P., Margolick, J. B., Giorgi, J. V., Ferbas, J., Bauer, K., Kaslow, R., Muñoz, A., and the Multicenter AIDS Cohort Study (1992), "Influence of HIV-1 Infection and Cigarette Smoking on Leukocyte Profile in Homosexual Men," *Journal of Acquired Immune Deficiency Syndromes*, 5, 1124–1130.

Ramsay, J. O., and Dalzell, C. J. (1991), "Some Tools for Functional Data Analysis," *Journal of the Royal Statistical Society*, Ser. B, 53, 539–572.

Rice, J. A. (1984), "Boundary Modification for Nonparametric Regression," *Communications in Statistics: Theory and Methods*, 13, 893–900.

——— (1985), "Bandwidth Choice for Differentiation," *Journal of Multivariate Analysis*, 20, 251–264.

Rice, J. A., and Silverman, B. W. (1991), "Estimating the Mean and Covariance Structure Nonparametrically When the Data Are Curves," *Journal of the Royal Statistical Society*, Ser. B, 53, 233–243.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Serfling, R. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

Zeger, S. L., and Diggle, P. J. (1994), "Semiparametric Models for Longitudinal Data With Application to CD4 Cell Numbers in HIV Seroconverters," *Biometrics*, 50, 689–699.