Statistica Sinica **14**(2004), 763-788

POLYNOMIAL SPLINE ESTIMATION AND INFERENCE FOR VARYING COEFFICIENT MODELS WITH LONGITUDINAL DATA

Jianhua Z. Huang[1], Colin O. Wu[2] and Lan Zhou[1]

[1] *University of Pennsylvania and* [2] *National Heart, Lung and Blood Institute*

Abstract: We consider nonparametric estimation of coefficient functions in a varying coefficient model of the form $Y_{ij} = X_i^T(t_{ij})\boldsymbol{\beta}(t_{ij}) + \epsilon_i(t_{ij})$ based on longitudinal observations $\{(Y_{ij}, X_i(t_{ij}), t_{ij}),\ i = 1,\ldots,n,\ j = 1,\ldots,n_i\}$, where t_{ij} and n_i are the time of the jth measurement and the number of repeated measurements for the ith subject, and Y_{ij} and $X_i(t_{ij}) = (X_{i0}(t_{ij}),\ldots,X_{iL}(t_{ij}))^T$ for $L \ge 0$ are the ith subject's observed outcome and covariates at t_{ij}. We approximate each coefficient function by a polynomial spline and employ the least squares method to do the estimation. An asymptotic theory for the resulting estimates is established, including consistency, rate of convergence and asymptotic distribution. The asymptotic distribution results are used as a guideline to construct approximate confidence intervals and confidence bands for components of $\boldsymbol{\beta}(t)$. We also propose a polynomial spline estimate of the covariance structure of $\epsilon(t)$, which is used to estimate the variance of the spline estimate $\hat{\boldsymbol{\beta}}(t)$. A data example in epidemiology and a simulation study are used to demonstrate our methods.

Key words and phrases: Asymptotic normality, confidence intervals, nonparametric regression, repeated measurements, varying coefficient models.

1. Introduction

Longitudinal data occur frequently in medical and epidemiological studies. A convenient setup for such data is that the observed sequence of measurements on an individual is sampled from a realization of a continuous-time stochastic process $\{(Y(t), \boldsymbol{X}(t)), t \in \mathcal{T}\}$, where $Y(t)$ and $\boldsymbol{X}(t) = (X_0(t),\ldots,X_L(t))'$ denote, respectively, the real valued outcome of interest and the $\mathbb{R}^{L+1}$, $L \ge 1$, valued covariate, and $\mathcal{T}$ denotes the time interval on which the measurements are taken. Suppose there are n randomly selected subjects and let t_{ij}, $j = 1,\ldots,n_i$, be the observation times of the ith individual, $i = 1,\ldots,n$. The observed measurements for the ith individual are $(Y_{ij} = Y_i(t_{ij}), \boldsymbol{X}_{ij} = (X_{ij0},\ldots,X_{ijL})' = \boldsymbol{X}_i(t_{ij}), t_{ij})$, where $\{(Y_i(t), \boldsymbol{X}_i(t))\}$ are independent copies of the stochastic process $\{(Y(t), \boldsymbol{X}(t))\}$.

Statistical analyses with this type of data are usually concerned with modeling the mean curves of $Y(t)$ and the effects of the covariates on $Y(t)$, and developing the corresponding estimation and inference procedures. Theory and methods for estimation and inferences based on parametric models have been extensively studied and summarized in Diggle, Liang and Zeger (1994), Davidian and Giltnan (1995), Vonesh and Chinchilli (1997) and Verbeke and Mollenberghs (2000), among others. There has been substantial interest recently in extending the parametric models to allow for nonparametric covariate effects; see, for example, Hart and Wehrly (1986), Rice and Silverman (1991), Zeger and Diggle (1994), Moyeed and Diggle (1994), Besse, Cardot and Ferraty (1997), Brumback and Rice (1998), Staniswalis and Lee (1998), Cheng and Wei (2000) and Lin and Carroll (2000).

It is well known that nonparametric methods suffer from the "curse of dimensionality" when there are many covariates, hence dimensionality reduction techniques are desired in practice. A useful dimensionality reduction approach for longitudinal data is the time-varying coefficient model:

$$Y(t) = \boldsymbol{X}'(t)\boldsymbol{\beta}(t) + \epsilon(t), \qquad t \in \mathcal{T}, \tag{1.1}$$

where $\boldsymbol{X}(t) = (X_0(t), \ldots, X_L(t))'$, $\boldsymbol{\beta}(t) = (\beta_0(t), \ldots, \beta_L(t))'$, $X_0(t) \equiv 1$, $\beta_0(t)$ represents the baseline effect, $\epsilon(t)$ is a mean 0 stochastic process with variance function $\sigma_\epsilon^2(t)$ and covariance function $C_\epsilon(t_1, t_2)$ for $t_1 \neq t_2$, and $\boldsymbol{X}(t)$ and $\epsilon(t)$ are independent. Model (1.1) assumes a linear model for each fixed time t but allows the coefficients to vary with time. This model is attractive because it has a meaningful interpretation and still retains certain general nonparametric characteristics. Estimation of this model using the local polynomial method and smoothing splines has been studied in Hoover, Rice, Wu and Yang (1998), Fan and Zhang (2000), Wu and Chiang (2000), Wu, Yu and Chiang (2000), Chiang, Rice and Wu (2001) and others.

Recently, Huang, Wu and Zhou (2002) proposed a class of global estimation methods for the varying coefficient model based on basis approximations. The idea is to approximate the coefficient functions by a basis expansion, such as an expansion with B-splines, and employ the least squares method. This approach provides a simple universal solution to estimation and inference for the varying coefficient model with longitudinal data: It can handle both time-invariant and time-dependent covariates; no data binning is needed when observations are sparse at distinct observation time points; flexibility can be obtained by using different basis approximations when approximating different coefficient functions. Huang, Wu and Zhou (2002) established the consistency and convergence rates for a general class of basis choices including polynomials, trigonometric polynomials and B-splines. Rice and Wu (2001) also proposed a B-spline method for a different class of nonparametric models with time-invariant covariates, but have not investigated the theoretical properties of their estimation procedures.

The aim of this paper is to derive the asymptotic distributions of the polynomial spline estimators of the coefficient functions $\beta_l(t)$ in model (1.1), and to investigate their applications in statistical inferences. We show that the spline estimators are asymptotically normal, and use this result to develop approximate pointwise and simultaneous confidence intervals. Comparing with the asymptotic results for the local polynomial kernel methods of Wu, Chiang and Hoover (1998), our polynomial spline method can adjust the individual smoothing needs desired by different components of $\boldsymbol{\beta}(t)$ through the use of multiple smoothing parameters. The nonparametric inference procedures developed from our method are also much simpler than those proposed in Wu et al. (1998), as we do not rely on estimating the underlying density of the measurement time and the joint moments $E[X_l(t)X_{l'}(t)]$ of the covariates. Unlike Huang et al. (2002), which considers general basis approximations, we restrict our attention to B-splines and establish rates of convergence for estimators under weaker conditions.

Section 2 describes the estimation method. Section 3 presents the asymptotic theory including consistency, rate of convergence and asymptotic distribution of the spline estimates. Section 4 develops approximate pointwise confidence intervals and simultaneous confidence bands. Section 5 presents the application of our procedures to a CD4 depletion dataset from the Multicenter AIDS Cohort Study and reports the results from a small simulation study. The Appendix contains proofs of all theoretical results.

2. The Estimation Method

2.1. Spline approximation and least squares estimation

Polynomial splines are piecewise polynomials with the polynomial pieces jointing together smoothly at a set of interior knot points. A (polynomial) spline of degree $d \geq 0$ on $\mathcal{T}$ with knot sequence $\xi_0 < \xi_1 < \cdots < \xi_{M+1}$, where ξ_0 and ξ_{M+1} are the two end points of the interval $\mathcal{T}$, is a function that is a polynomial of degree d on each of the intervals $[\xi_m, \xi_{m+1})$, $0 \leq m \leq M-1$, and $[\xi_M, \xi_{M+1}]$, and globally has continuous $d-1$ continuous derivatives for $d \geq 1$. A piecewise constant function, linear spline, quadratic spline and cubic spline corresponds to $d = 0, 1, 2, 3$, respectively. The collection of spline functions of a particular degree and knot sequence form a linear space. The books by de Boor (1978) and Schumaker (1980) are good references for spline functions.

Suppose that each $\beta_l(t)$, $l = 1, \ldots, L$, can be approximated by some spline function, that is,

$$\beta_l(t) \approx \sum_{k=1}^{K_l} \gamma_{lk} B_{lk}(t), \qquad l = 0, \ldots, L, \tag{2.1}$$

where, for each $l = 0, \ldots, L$, $\{B_{lk}(\cdot), k = 1, \ldots, K_l\}$ is a basis for a linear space $\mathbb{G}_l$ of spline functions on $\mathcal{T}$ with a fixed degree and knot sequence. In our applications we use the B-spline basis for its good numerical properties. The approximation sign in (2.1) will be replaced by a strict equality with a fixed and known K_l when $\beta_l(t)$ belongs to the linear space $\mathbb{G}_l$. For the general case that $\beta_l(t)$ may not be restricted to $\mathbb{G}_l$, it is natural to allow K_l to increase with the sample size, allowing a more accurate approximation when the sample size increases. Following (1.1) and (2.1), we have

$$Y_{ij} \approx \sum_{l=0}^{L} \sum_{k=1}^{K_l} X_{ijl} B_{lk}(t_{ij}) \gamma_{lk} + \epsilon_{ij},$$

and can estimate γ_{lk}, hence $\beta_l(t)$, based on (2.1) with any given K_l by minimizing

$$\ell = \sum_{i=1}^{n} w_i \sum_{j=1}^{n_i} \Big(Y_{ij} - \sum_{l=0}^{L} \sum_{k=1}^{K_l} X_{ijl} B_{lk}(t_{ij}) \gamma_{lk} \Big)^2$$

with respect to γ_{lk}. Usual choices of w_i include $w_i \equiv 1$ and $w_i \equiv 1/n_i$, which correspond to providing equal weight to each single observation and equal weight to each subject, respectively. For $i = 1, \ldots, n$, $j = 1, \ldots, n_i$ and $l = 0, \ldots, L$, set $\boldsymbol{\gamma}_l = (\gamma_{l0}, \ldots, \gamma_{lK_l})'$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_0', \ldots, \boldsymbol{\gamma}_L')'$,

$$\mathbf{B}(t) = \left(\begin{array}{ccccccccc} B_{01}(t) & \cdots & B_{0K_0}(t) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & B_{L1}(t) & \cdots & B_{LK_L}(t) \end{array} \right\},$$

$\mathbf{U}_{ij}' = \boldsymbol{X}_i'(t_{ij})\mathbf{B}(t_{ij})$, $\mathbf{U}_i = (\mathbf{U}_{i1}, \ldots, \mathbf{U}_{in_i})'$, $\mathbf{W}_i = \mathrm{diag}(w_i, \ldots, w_i)$, and $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$. We have $\ell = \ell(\boldsymbol{\gamma}) = \sum_{i=1}^{n} (\mathbf{Y}_i - \mathbf{U}_i\boldsymbol{\gamma})'\mathbf{W}_i(\mathbf{Y}_i - \mathbf{U}_i\boldsymbol{\gamma})$. If $\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i$ is invertible, a condition that is satisfied under mild conditions (see Lemma A.3), then $\ell(\boldsymbol{\gamma})$ has a unique minimizer

$$\hat{\boldsymbol{\gamma}} = \Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i \Big)^{-1} \sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{Y}_i. \tag{2.2}$$

Write $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_0', \ldots, \hat{\boldsymbol{\gamma}}_L')'$ with $\hat{\boldsymbol{\gamma}}_l = (\hat{\gamma}_{l0}, \ldots, \hat{\gamma}_{lK_l})'$ for $l = 0, \ldots, L$. The spline estimate of $\boldsymbol{\beta}(t)$ is $\hat{\boldsymbol{\beta}}(t) = \mathbf{B}(t)\hat{\boldsymbol{\gamma}} = (\hat{\beta}_0(t), \ldots, \hat{\beta}_L(t))'$, where $\hat{\beta}_l(t) = \sum_k \hat{\gamma}_{lk} B_{lk}(t)$.

2.2. Expression of the conditional variance of the spline estimators

Let $\mathcal{X} = \{(\boldsymbol{X}_i(t_{ij}), t_{ij});\ i = 1, \ldots, n,\ j = 1, \ldots, n_i\}$. It is easily seen from (2.2) that the variance-covariance matrix of $\hat{\boldsymbol{\gamma}}$ conditioning on $\mathcal{X}$ is

$$\text{var}(\hat{\boldsymbol{\gamma}}) = \Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1}\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{V}_i\mathbf{W}_i\mathbf{U}_i\Big)\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1},$$

where $\mathbf{V}_i = \text{var}(\mathbf{Y}_i) = (C_\epsilon(t_{ij}, t_{ij'}))$ and $C_\epsilon(t,s)$ is the variance-covariance function of $\epsilon(t)$. For simplicity, we omit symbols for conditioning on $\mathcal{X}$ in our notation. The variance-covariance matrix of $\hat{\boldsymbol{\beta}}(t)$ conditioning on $\mathcal{X}$ is

$$\text{var}(\hat{\boldsymbol{\beta}}(t)) = \mathbf{B}(t)\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1}\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{V}_i\mathbf{W}_i\mathbf{U}_i\Big)\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1}\mathbf{B}'(t). \tag{2.3}$$

Let e_{l+1} be the $(L+1)$-dimensional vector with the $(l+1)$th element taken to be 1 and zero elsewhere. The conditional variance of $\hat{\beta}_l(t)$ is

$$\text{var}(\hat{\beta}_l(t)) = e_{l+1}'\text{var}(\hat{\boldsymbol{\beta}}(t))e_{l+1}, \qquad l = 0, \ldots, L. \tag{2.4}$$

Note that the only unknown quantity in (2.4) is $\mathbf{V}_i = (C_\epsilon(t_{ij}, t_{ij'}))$.

2.3. Automatic selection of smoothing parameters

Because of computational complexity, it is often impractical to automatically select all three components involved in the smoothing parameters: the degrees of splines and the numbers and locations of knots. Similar to Rice and Wu (2001), we use splines with equally spaced knots and fixed degrees and select only K_l, the numbers of knots, using the data. Here K_l is the dimension of $\mathbb{G}_l$ and is related to the number M_l of interior knots through $K_l = M_l + 1 + d$, where d is the degree of the spline. We use "leave-one-subject-out" cross-validation (Rice and Silverman (1991), Hart and Wehrly (1993) and Hoover et al. (1998)). Specifically, let $\hat{\boldsymbol{\beta}}^{(-i)}(t)$ be the spline estimator obtained by deleting the measurements of the ith subject and

$$\text{CV} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\Big\{w_i\Big(Y_{ij} - \boldsymbol{X}_i^T(t_{ij})\hat{\boldsymbol{\beta}}^{(-i)}(t_{ij})\Big)^2\Big\} \tag{2.5}$$

be the cross-validation score. We select $(K_0, \ldots, K_L)$ by minimizing this cross-validation score. One advantage of this approach is that, by deleting the entire measurements of the subject one at a time, it is expected to preserve the intra-subject correlation.

When there are a large number of subjects, calculating the "leave-one-subject-out" cross-validation score can be computationally intensive. In such a case, we can use the "leave-subjects-out" K-fold cross-validation by splitting the subjects into K roughly equal-sized parts. Let $k[i]$ be the part containing subject i and denote by $\hat{\boldsymbol{\beta}}^{-k[i]}$ the estimate of $\boldsymbol{\beta}$ with the measurements of the $k[i]$th part of the subjects removed. Then the K-fold cross-validation score is

$$\mathrm{CV}^* = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{w_i\left(Y_{ij} - \boldsymbol{X}_i^T(t_{ij})\hat{\boldsymbol{\beta}}^{(-k[i])}(t_{ij})\right)^2\right\},$$

and we select $(K_0, \ldots, K_L)$ by minimizing this K-fold cross-validation score. The difference of the "delete-subjects-out" K-fold cross-validation and the ordinary K-fold cross-validation is that the measurements corresponding to the same subjects are deleted altogether.

Remark 2.1. In this paper we restrict our attention to splines with equally spaced knots. This worked well for the applications we considered. It might be worthwhile to investigate using the data to decide the knot positions (free-knot splines). There has been considerable work on free-knot splines for i.i.d. data; see Stone, Hansen, Kooperberg and Truong (1997), Hansen and Kooperberg (2002) and Stone and Huang (2002). Extension of the methodology and theory of free-knot splines to longitudinal data is beyond the scope of this paper.

3. Asymptotic Theory

We now describe the asymptotic properties of the spline estimates $\hat{\beta}_l$ when the number of subjects n tends to infinity while, for each subject i, the number of observations n_i may or may not tend to infinity and, for $l = 0, \ldots, L$, the dimensionality $K_l = K_{ln}$ of the spline space $\mathbb{G}_l$ may or may not tend to infinity. We only present results for the weight scheme $w_i \equiv 1/n_i$. Results for other choices of w_i can be obtained using the same arguments.

We first introduce some technical conditions.

(C1) The observation times t_{ij}, $j = 1, \ldots, n_i$, $i = 1, \ldots, n$, are chosen independently according to a distribution F_T on $\mathcal{T}$; moreover, they are independent of the response and covariate processes $\{(Y_i(t), \boldsymbol{X}_i(t))\}$, $i = 1, \ldots, n$. The distribution F_T has a Lebesgue density $f_T(t)$ which is bounded away from 0 and infinity uniformly over $t \in \mathcal{T}$.

(C2) The eigenvalues $\lambda_0(t) \le \cdots \le \lambda_L(t)$ of $\Sigma(t) = E[\boldsymbol{X}(t)\boldsymbol{X}'(t)]$ are bounded away from 0 and infinity uniformly in $t \in \mathcal{T}$; that is, there are positive constants M_1 and M_2 such that $M_1 \le \lambda_0(t) \le \cdots \le \lambda_L(t) \le M_2$ for $t \in \mathcal{T}$.

(C3) There is a positive constant M_3 such that $|X_l(t)| \le M_3$ for $t \in \mathcal{T}$ and $l = 0, \ldots, L$.

(C4) There is a constant M_4 such that $E[\epsilon(t)^2] \le M_4 < \infty$ for $t \in \mathcal{T}$.

(C5) $\limsup_n(\max_l K_l / \min_l K_l) < \infty$.

(C6) The process $\epsilon(t)$ can be decomposed as the sum of two independent stochastic processes, $\epsilon^{(1)}(t)$ and $\epsilon^{(2)}(t)$, where $\epsilon^{(1)}$ is an arbitrary mean zero process and $\epsilon^{(2)}$ is a process of measurement errors that are independent at different time points and have mean zero and constant variance σ^2.

These are mild conditions that are satisfied in many practical situations. Condition (C1) guarantees that the observation times are randomly scattered and can be modified or weakened (Remarks 3.1 and 3.2). Let $\|a\|_{L_2}$ denote the L_2 norm of a square integrable function $a(t)$ on $\mathcal{T}$. We define $\hat{\beta}_l(\cdot)$ to be a consistent estimator of $\beta_l(\cdot)$ if $\lim_{n\to\infty}\|\hat{\beta}_l - \beta_l\|_{L_2} = 0$ holds in probability. Let $K_n = \max_{0\le l\le L} K_l$ and $\mathrm{dist}(\beta_l, \mathbb{G}_l) = \inf_{g\in\mathbb{G}_l}\sup_{t\in\mathcal{T}}|\beta_l(t) - g(t)|$ be the L_∞ distance between $\beta_l(\cdot)$ and $\mathbb{G}_l$.

Theorem 1.(Consistency) *Suppose conditions* (C1)−(C5) *hold,* $\lim_n \mathrm{dist}(\beta_l, \mathbb{G}_l) = 0$*,* $l = 0,\ldots,L$*, and* $\lim_n K_n \log K_n/n = 0$*. Then* $\hat{\beta}_l$*,* $l = 0,\ldots,L$*, are uniquely defined with probability tending to one. Moreover,* $\hat{\beta}_l$*,* $l = 0,\ldots,L$*, are consistent.*

Let $\tilde{\beta}_l(t) = E[\hat{\beta}_l(t)]$ be the mean of $\hat{\beta}_l(t)$ conditioning on $\mathcal{X}$. It is useful to consider the decomposition $\hat{\beta}_l(t) - \beta_l(t) = \hat{\beta}_l(t) - \tilde{\beta}_l(t) + \tilde{\beta}_l(t) - \beta_l(t)$, where $\hat{\beta}_l(t) - \tilde{\beta}_l(t)$ and $\tilde{\beta}_l(t) - \beta_l(t)$ contribute to the variance and bias terms respectively. Let $\rho_n = \max_{0\le l\le L}\mathrm{dist}(\beta_l, \mathbb{G}_l)$.

Theorem 2.(Rates of Convergence) *Suppose conditions* (C1)−(C5) *hold. If* $\lim_n K_n \log K_n/n = 0$*, then* $\|\tilde{\beta}_l - \beta_l\|_{L_2} = O_P(\rho_n)$ *and* $\|\hat{\beta}_l - \tilde{\beta}_l\|^2_{L_2} = O_P(1/n + K_n n^{-2}\sum_i n_i^{-1})$*; consequently,* $\|\hat{\beta}_l - \beta_l\|^2_{L_2} = O_P(1/n + K_n n^{-2}\sum_i n_i^{-1} + \rho_n^2)$.

This theorem implies that the magnitude of the bias term is bounded in probability by the best approximation rates obtainable by the spaces $\mathbb{G}_l$. When the number of observations for each subject is bounded, that is, $n_i \le C$, $1 \le i \le n$, for some constant C, the rate of convergence of $\|\tilde{\beta}_l - \beta_l\|^2_{L_2}$ reduces to $O_P(K_n/n + \rho_n^2)$, the same rate for i.i.d. data (Huang (1998) and (2001)).

By condition (C5), the approximation rate ρ_n can be determined in terms of K_n under commonly used smoothness conditions on the β_l. When the β_l have bounded second derivatives, we have $\rho_n = O(K_n^{-2})$ (Schumaker (1981, Theorem 6.27)), and the rates of convergence in Theorem 2 become $\|\hat{\beta}_l - \beta_l\|^2_{L_2} = O_P(K_n n^{-2}\sum_i n_i^{-1} + K_n^{-4})$. Choosing $K_n \sim (n^{-2}\sum_i n_i^{-1})^{-1/5}$, we find $\|\hat{\beta}_l - \beta_l\|^2_{L_2} = O_P((n^{-2}\sum_i n_i^{-1})^{4/5})$. When the number of observations for each subject is bounded, we get $\|\hat{\beta}_l - \beta_l\|^2_{L_2} = O_P(n^{-4/5})$, the same optimal rate as for i.i.d. data (Stone (1982)). When n_i is bounded by a fixed constant, the requirement on K_n to achieve the $n^{-4/5}$ rate reduces to $K_n \sim n^{1/5}$. Huang et al. (2002) established consistency and rates of convergence of general basis estimators. The conditions on K_n required in Theorems 1 and 2 are less stringent than those in Huang et al. (2002).

For positive definite matrices A and B, let $B^{1/2}$ denote the unique square root of B and let $A^{-1/2} = (A^{-1})^{1/2}$.

Theorem 3.(Asymptotic Normality) *Suppose conditions* (C1)−(C6) *hold. If* $\lim_n K_n \log K_n/n = 0$ *and* $\lim_n K_n \max_i n_i/n = 0$*, then* $\{\text{var}(\hat{\boldsymbol{\beta}}(t))\}^{-1/2}(\hat{\boldsymbol{\beta}}(t) - \tilde{\boldsymbol{\beta}}(t)) \longrightarrow N(0, I)$ *in distribution, where* $\tilde{\boldsymbol{\beta}}(t) = (\tilde{\beta}_0(t), \ldots, \tilde{\beta}_L(t))'$*, and in particular, for* $l = 0, \ldots, L$*,* $\{\text{var}(\hat{\beta}_l(t))\}^{-1/2}(\hat{\beta}_l(t) - \tilde{\beta}_l(t)) \longrightarrow N(0, 1)$ *in distribution.*

The above result extends similar result of Huang (2003) for i.i.d. data. It can be used to construct asymptotic confidence intervals; see, for example, Section 3.5 of Hart (1997). One sensible approach is to think of $\tilde{\beta}_l(t)$ as the estimable part of $\beta_l(t)$ and construct an asymptotic confidence interval for $\tilde{\beta}_l(t)$. Note that $\tilde{\beta}_l(t)$ can be interpreted as the best approximation in the estimation space $\mathbb{G}_l$ to $\beta_l(t)$. Another approach is to undersmooth so that the squared bias term $(\tilde{\beta}_l(t) - \beta_l(t))^2$ is asymptotically negligible relative to the variance.

Theorem 4.(Bias) *Suppose conditions* (C1)−(C5) *hold and* $\lim_n K_n \log K_n/n = 0$*. Then* $\sup_{t\in\mathcal{T}} |\tilde{\beta}_l(t) - \beta_l(t)| = O_P(\rho_n)$*,* $l = 0, \ldots, L$*.*

We now give a sufficient condition for the bias term to be negligible relative to the variance term.

Corollary 1. *Suppose assumptions in Theorem* 4 *hold and condition* (C6) *holds. In addition, suppose* $\beta_l(t)$*,* $0 \le l \le L$*, have bounded second derivatives. If* $\lim_n K_n^5/(n \max_i n_i) = \infty$*, then* $\sup_{t\in\mathcal{T}} |\{\text{var}(\hat{\beta}_l(t))\}^{-1/2}(\tilde{\beta}_l(t) - \beta_l(t))| = o_P(1)$*,* $l = 0, \ldots, L$*.*

Remark 3.1. The results in this section still hold when the observation times $\{t_{ij}\}$ are deterministic. In this case we need to replace condition (C1) by the following.

(i′) There are constants M_1 and M_2 such that

$$M_1\|g\|_{L_2}^2 \le \frac{1}{n}\sum_i \frac{1}{n_i}\sum_j g^2(t_{ij}) \le M_2\|g\|_{L_2}^2, \qquad g \in G_l,\ l = 0, \ldots, L. \tag{3.1}$$

See Appendix A.9 for some technical details. A sufficient condition for (i′) to hold is that

$$\sup_{t\in\mathcal{T}} |F_n(t) - F_T(t)| = o(1/K_n) \tag{3.2}$$

for some distribution function $F_T(t)$ which has a Lebesgue density $f_T(t)$ that is bounded away from 0 and infinity uniformly over $t \in \mathcal{T}$, where $F_n(t) = (1/n)\sum_i(1/n_i)\sum_j 1_{\{t_{ij}\le t\}}$ and $1_{\{\cdot\}}$ is the indicator function; see Appendix A.10 for a proof.

Remark 3.2. The requirement in condition (C1) that the t_{ij} are independent of each other can be relaxed. In light of Remark 3.1, condition (C1) can be replaced by the requirement that (3.1) or, sufficiently, (3.2) holds with probability tending to one.

4. Asymptotic Confidence Intervals and Confidence Bands

4.1. Pointwise confidence intervals

Under regularity conditions, for $0 \le l \le L$ and $t \in \mathcal{T}$,

$$\{\mathrm{var}(\hat\beta_l(t))\}^{-1/2}(\hat\beta_l(t) - E[\hat\beta_l(t)]) \longrightarrow N(0,1) \qquad \text{in distribution} \tag{4.1}$$

as n tends to infinity, where $E[\hat\beta_l(t)]$ and $\mathrm{var}(\hat\beta_l(t))$ are the mean and variance of $\hat\beta_l(t)$ conditioning on $\mathcal{X}$. Suppose that there is an estimate $\widehat{\mathrm{var}}(\hat\beta_l(t))$ of $\mathrm{var}(\hat\beta_l(t))$ such that $\widehat{\mathrm{var}}(\hat\beta_l(t))/\mathrm{var}(\hat\beta_l(t)) \to 1$ in probability as $n \to \infty$. It follows from (4.1) and Slutzky's Theorem that, as $n \to \infty$, $\{\widehat{\mathrm{var}}(\hat\beta_l(t))\}^{-1/2}(\hat\beta_l(t) - E[\hat\beta_l(t)]) \longrightarrow N(0,1)$ in distribution, so that an approximate $(1-\alpha)$ asymptotic confidence interval for $E[\hat\beta_l(t)]$ has end points

$$\hat\beta_l(t) \pm z_{\alpha/2}(\widehat{\mathrm{var}}(\hat\beta_l(t)))^{1/2}, \tag{4.2}$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)$th quantile value of the standard Gaussian distribution. If the bias $E[\hat\beta_l(t)] - \beta_l(t)$ is asymptotically negligible relative to the variance of $\hat\beta_l(t)$ (see Section 3 for specific conditions), then $\hat\beta_l(t) \pm z_{\alpha/2}(\widehat{\mathrm{var}}(\hat\beta_l(t)))^{1/2}$ is also a $(1-\alpha)$ asymptotic confidence interval for $\beta_l(t)$.

The procedure for constructing confidence intervals given here is simpler than the kernel based method of Wu et al. (1998), since the construction of estimates of $\mathrm{var}(\hat\beta_l)$ requires only estimation of the variance and covariance functions of $\epsilon(t)$. This is in contrast to the asymptotic normality result for the kernel method (Theorem 1 of Wu et al. (1998)), where the asymptotic variance of the estimate depends not only on the variance-covariance structure of $\epsilon(t)$ but also on the design density $f_T(t)$ of the observation time and the joint moments $E(X_l(t)X_{l'}(t))$ of the covariate process. The plug-in type approximate confidence intervals suggested by Wu et al. (1998) rely on estimating the extra quantities involving $f_T(t)$ and $E(X_l(t)X_{l'}(t))$ through kernel smoothing methods.

4.2. Simultaneous variability bands

We present here a simple approach that extends the above pointwise confidence intervals to simultaneous bands for $E[\hat\beta_l(t)]$ and $\beta_l(t)$ over a given subinterval $[a,b]$ of $\mathcal{T}$. Our approach is similar to the ones used in Knafl, Sacks and Ylvisaker (1985), Hall and Titterington (1988) and Wu et al. (1998). Partitioning $[a,b]$ according to $M+1$ equally spaced grid points $a = \xi_1 < \cdots < \xi_{M+1} = b$ for some integer $M \ge 1$, we get a set of approximate $(1-\alpha)$ simultaneous confidence intervals $(l_{l,\alpha}(\xi_r), u_{l,\alpha}(\xi_r))$ for $E[\hat\beta_l(\xi_r)]$, such that $\lim_{n\to\infty} P(l_{l,\alpha}(\xi_r) \le$

$E[\hat{\beta}_l(\xi_r)] \le u_{l,\alpha}(\xi_r),$ for all $r = 1, \ldots, M+1) \ge 1-\alpha$. A simple approach based on the Bonferroni adjustment is to choose $(l_{l,\alpha}(\xi_r), u_{l,\alpha}(\xi_r))$ to be

$$\hat{\beta}_l(\xi_r) \pm z_{\alpha/[2(M+1)]}(\widehat{\mathrm{var}}(\hat{\beta}_l(\xi_r)))^{1/2}. \tag{4.3}$$

Let $E^{(I)}[\hat{\beta}_l(t)]$ be the linear interpolation of $E[\hat{\beta}_l(\xi_r)]$ and $E[\hat{\beta}_l(\xi_{r+1})]$, $\xi_r \le t \le \xi_{r+1}$:

$$E^{(I)}[\hat{\beta}_l(t)] = M\Big(\frac{\xi_{r+1}-t}{b-a}\Big)E[\hat{\beta}_l(\xi_r)] + M\Big(\frac{t-\xi_r}{b-a}\Big)E[\hat{\beta}_l(\xi_{r+1})].$$

Similarly, let $l^{(I)}_{l,\alpha}(t)$ and $u^{(I)}_{l,\alpha}(t)$ be the linear interpolations of $l_{l,\alpha}(\xi_r)$ and $u_{l,\alpha}(\xi_r)$, respectively. Then, $(l^{(I)}_{l,\alpha}(t), u^{(I)}_{l,\alpha}(t))$ is an approximate $(1-\alpha)$ confidence band for $E^{(I)}[\hat{\beta}_l(t)]$ in the sense that $\lim_{n\to\infty} P(l^{(I)}_{l,\alpha}(t) \le E^{(I)}[\hat{\beta}_l(t)] \le u^{(I)}_{l,\alpha}(t),$ for all $t \in [a,b]) \ge 1-\alpha$.

To construct the bands for $E[\hat{\beta}_l(t)]$, we assume one of

$$\sup_{t\in[a,b]}\Big|\{E[\hat{\beta}_l(t)]\}'\Big| \le c_1, \qquad \text{for a known constant } c_1 > 0, \tag{4.4}$$

$$\sup_{t\in[a,b]}\Big|\{E[\hat{\beta}_l(t)]\}''\Big| \le c_2, \qquad \text{for a known constant } c_2 > 0. \tag{4.5}$$

Direct calculation using Taylor's expansions shows that, for $\xi_r \le t \le \xi_{r+1}$,

$$\Big|E[\hat{\beta}_l(t)] - E^{(I)}[\hat{\beta}_l(t)]\Big| \le \begin{cases} 2c_1 M\left(\frac{(\xi_{r+1}-t)(t-\xi_r)}{b-a}\right), & \text{if (4.4) holds;} \\ \frac{1}{2}c_2\left(\xi_{r+1}-t\right)\left(t-\xi_r\right), & \text{if (4.5) holds.} \end{cases}$$

Adjusting the bands for $E^{(I)}[\hat{\beta}_l(t)]$, our approximate $(1-\alpha)$ confidence bands for $E[\hat{\beta}_l(t)]$ are

$$\Big(l^{(I)}_{l,\alpha}(t) - 2c_1M\Big(\frac{(\xi_{r+1}-t)(t-\xi_r)}{b-a}\Big), u^{(I)}_{l,\alpha}(t) + 2c_1M\Big(\frac{(\xi_{r+1}-t)(t-\xi_r)}{b-a}\Big)\Big) \tag{4.6}$$

under (4.4) or, under (4.5),

$$\Big(l^{(I)}_{l,\alpha}(t) - \frac{1}{2}c_2(\xi_{r+1}-t)(t-\xi_r), u^{(I)}_{l,\alpha}(t) + \frac{1}{2}c_2(\xi_{r+1}-t)(t-\xi_r)\Big). \tag{4.7}$$

When the bias $E[\hat{\beta}_l(t)] - \beta_l(t)$ is asymptotically negligible (a sufficient condition is given in Corollary 1 in Section 3) and either $\sup_{t\in[a,b]}|\beta_l'(t)| \le c_1$ or $\sup_{t\in[a,b]}|\beta_l''(t)| \le c_2$ holds for known positive constants c_1 or c_2, then (4.6) or (4.7) are the corresponding asymptotic confidence bands for $\beta_l(t)$, $0 \le l \le L$.

Remark 4.1. The Bonferroni adjustment (4.3), although simple, often leads to conservative bands. For refinements, one may use the inclusion-exclusion identities to calculate $(l_{l,\alpha}(\xi_r), u_{l,\alpha}(\xi_r))$ with more accurate coverage probabilities; see, for example, Naiman and Wynn (1997). These refinements, however, usually involve extensive computations and may not be practical for large longitudinal studies. Another related issue is the choice of M. Although some heuristic suggestions for the simple case of kernel regression with independent identically distributed samples have been provided by Hall and Titterington (1988), theoretical guidelines for the choice of M under the current situation is still unknown.

4.3. Estimation of the covariance structure

Since the conditional variance of $\hat{\beta}_l(t)$ is determined by the covariance structure of the process $\epsilon(t)$ (see (2.3) and (2.4)), a crucial step in estimating the conditional variance of $\hat{\beta}_l(t)$ is to estimate the covariance function $C_\epsilon(t,s)$ of $\epsilon(t)$. Diggle and Verbyla (1998) developed a local smoothing method to estimate the covariance structure. However, local smoothing could be computationally expensive in the current context, since the estimated covariance function needs to be evaluated at each distinct pair of observation times. We propose here a spline based estimate of the covariance function.

To estimate $C_\epsilon(t,s)$, we approximate it by a tensor product spline on $\mathcal{T}\times\mathcal{T}$, that is,

$$C_\epsilon(t,s) \approx \sum_k \sum_l u_{kl} B_k(t) B_l(s), \qquad t,s \in \mathcal{T}, t \neq s,$$

where $\{B_k\}$ is a spline basis on $\mathcal{T}$ with a fixed knot sequence. The above approximation is only required to hold when $t \neq s$, since, in most practical longitudinal settings, the correlation function $C_\epsilon(t,s)$ is not necessarily continuous at $t = s$, that is, $\lim_{s\to t} C_\epsilon(t,s) \neq C_\epsilon(t,t)$; see, for example, Diggle (1988) and Diggle and Verbyla (1998). Note that $E[\epsilon(t_{ij})\epsilon(t_{ij'})] = C_\epsilon(t_{ij}, t_{ij'})$ for $j \neq j'$ and $C(t,s) = C(s,t)$. If $\{\epsilon_i(t_{ij}),\ i = 1,\ldots,n,\ j = 1,\ldots,n_i\}$ were observed, $C_\epsilon(t,s)$, $t \neq s$, could be estimated by finding $\{u_{kl} : u_{kl} = u_{lk}\}$ which minimize

$$\sum_{i=1}^{n} \sum_{j,j'=1,j<j'}^{n_i} \Big(\epsilon_i(t_{ij})\epsilon_i(t_{ij'}) - \sum_k \sum_l u_{kl} B_k(t_{ij}) B_l(t_{ij'})\Big)^2. \tag{4.8}$$

Since $\epsilon_i(t_{ij})$ are not observed, we minimize (4.8) with $\epsilon_i(t_{ij})$ replaced by the residuals $\hat{\epsilon}_i(t_{ij}) = Y_{ij} - \boldsymbol{X}_i'\hat{\boldsymbol{\beta}}(t_{ij})$. Denoting the minimizers as $\hat{u}_{kl}$, the spline estimate of $C_\epsilon(t,s)$ for $t \neq s$ is

$$\hat{C}_\epsilon(t,s) = \sum_k \sum_l \hat{u}_{kl} B_k(t) B_l(s), \qquad t,s \in \mathcal{T},\ t \neq s.$$

For the estimation of $\sigma_\epsilon^2(t) = C_\epsilon(t,t)$, we use the spline approximation $C_\epsilon(t,t) \approx \sum_k v_k B_k(t)$ and define $\hat{\sigma}^2(t) = \sum_k \hat{v}_k B_k(t)$ to be the spline estimate, where the

$\hat{v}_k$ minimize

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\Big(\hat{\epsilon}_i^2(t_{ij}) - \sum_k v_k B_k(t_{ij})\Big)^2.$$

Our spline estimates of $\mathrm{var}(\boldsymbol{\beta}(t))$ and $\mathrm{var}(\beta_l(t))$ are then obtained by substituting $C_\epsilon(t,s)$ and $\sigma_\epsilon(t)$ with $\hat{C}_\epsilon(t,s)$ and $\hat{\sigma}_\epsilon(t)$ in (2.3) and (2.4), respectively.

The estimation of $C_\epsilon(t,s)$ and $\sigma_\epsilon(t)$ relies on choosing the appropriate spline spaces. In practice, we can use equally spaced knot sequences and select the numbers of knots either subjectively or through the cross-validation procedures described in Section 2.3. However, such data-driven choices are often very computationally intensive. In our simulation study, we found that a number of knots between 5 and 10 gave satisfactory results.

Remark 4.2. Our estimator of the covariance function is motivated by a moment condition similar to the ones used in Section 2. The same arguments as in the proofs of Theorems 1 and 2 can be used to show the consistency of the proposed covariance function estimator under mild regularity conditions. In fact, one can view $Z_{ijj'} = \epsilon_i(t_{ij})\epsilon_i(t_{ij'})$ as a longitudinal observation on the product domain $\mathcal{T} \times \mathcal{T}$, with the mean function $C_\epsilon(t_{ij}, t_{ij'})$. This is a setup analogous to that in Theorems 1 and 2, except there is no covariate.

Remark 4.3. Similar to the local polynomial estimator of Diggle and Verbyla (1998), the proposed spline estimator of the covariance function need not be positive definite for a given finite sample, although, by its consistency, it is asymptotically positive definite. So far, there is no satisfactory solution to the problem of constructing a nonparametric covariance function estimator that is positive definite under finite longitudinal samples. How to impose the finite sample positive definiteness constraint to the current spline estimator is an important problem that deserves futher investigation.

5. Numerical Results

5.1. Application to CD4 depletion in HIV infection

We analyze a subset of the Multicenter AIDS Cohort Study, which includes cigarette smoking status (smoking versus non-smoking), age at HIV infection, pre-HIV infection CD4 cell percent (CD4 cell count divided by the total number of lymphocites) and repeatedly measured post-infection CD4 cell percent of 283 homosexual men who were infected by HIV during the study period between 1984 and 1991. Details about the design, methods and medical implications of the Multicenter AIDS Cohort Study can be found in Kaslow et al. (1987). Although all the individuals were scheduled to have their measurements made at semi-annual visits, due to the missing visits and the fact that HIV infections occurred randomly during the study, not all the individuals were observed at a common set of time points. The number of repeated measurements per subject ranged from 1 to 14, with a median of 6 and a mean of 6.57. The number of distinct measurement time points was 59.

Define t_{ij} to be the time (in years) of the jth measurement of the ith individual after HIV infection; Y_{ij} the ith individual's CD4 percent measured at time t_{ij}; $X_i^{(1)}$ the ith individual's smoking status, taken to be 1 or 0 if the ith individual ever or never smoked cigarettes, respectively, after his infection; $X_i^{(2)}$ the ith individual's centered age at HIV infection, computed by subtracting the sample average age at the infection from the ith individual's age at the infection; and $X_i^{(3)}$ the ith individual's centered pre-infection CD4 percent, computed by subtracting the average pre-infection CD4 percent of the sample from the ith individual's observed pre-infection CD4 percent. We consider the time-varying coefficient model

$$Y_{ij} = \beta_0(t_{ij}) + X_i^{(1)}\beta_1(t_{ij}) + X_i^{(2)}\beta_2(t_{ij}) + X_i^{(3)}\beta_3(t_{ij}) + \epsilon_{ij}, \tag{5.1}$$

where the baseline CD4 percent $\beta_0(t)$ represents the mean CD4 percent at t years after the infection for a non-smoker with average pre-infection CD4 percent and average age at the infection, $\beta_1(t)$, $\beta_2(t)$ and $\beta_3(t)$ describe the time-varying effects for cigarette smoking, age at HIV infection and pre-infection CD4 percent, respectively, on the post-infection CD4 percent at time t. The centered covariates, $X_i^{(2)}$ and $X_i^{(3)}$, are used in (5.1) to ensure a clear biological interpretation of the baseline coefficient function $\beta_0(t)$. This model is appropriate for an initial exploration analysis because there have been no known parametric models that have been justified for this situation, while a high dimensional nonparametric fitting would be unrealistic for the given sample size.

We fitted (5.1) using cubic splines with equally spaced knots and $w_i = 1/n_i$. Using the cross-validation of Section 2.3, the numbers of interior knots for $\hat{\beta}_0(\cdot)$, $\hat{\beta}_1(\cdot)$, $\hat{\beta}_2(\cdot)$ and $\hat{\beta}_3(\cdot)$ were chosen to be 0, 5, 1 and 3, respectively. Pointwise confidence intervals and simultaneous confidence bands were constructed using the procedures of Section 4. The covariance structure of $\epsilon(t)$ was estimated using the method of Section 4.3 with cubic splines and five equally spaced knots. The Bonferroni bands were computed using (4.6) with $c_1 = 3$ and $M = 60$.

Figure 1 shows the fitted coefficient functions (solid curves), their 95% pointwise confidence intervals (dotted curves) and Bonferroni bands (dashed curves). These curves imply that (a) the baseline CD4 percent of the population depletes with time, but the rate of depletion appears to be gradually slowing down; (b) cigarette smoking and age of HIV infection do not show any significant effect on the post-infection CD4 percent; (c) pre-infection CD4 percent appears to be positively associated with high post-infection CD4 percent. These findings agree with the ones obtained in Wu and Chiang (2000) and Fan and Zhang (2000). Note that the asymptotic confidence intervals in Figure 1 are similar to the boostrap confidence intervals in Huang et al. (2002).

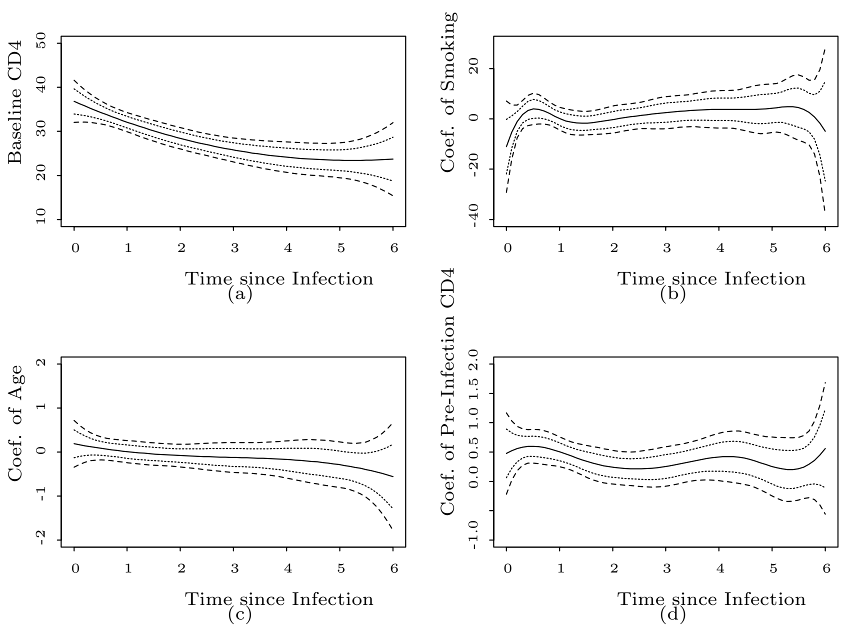

Figure 1. CD4 Cell Data. Estimated coefficient curves (solid), their 95% pointwise confidence intervals (dotted), and 95% conservative Bonferroni-type simultaneous bands (dashed). (a) baseline CD4 percentage, (b) smoking effect, (c) age effect, (d) pre-infection CD4 percentage effect.

5.2. Monte Carlo simulation

In each simulation run, we generated a simple random sample of 200 subjects according to the model

$$Y_{ij} = \beta_0(t_{ij}) + X_i(t_{ij})\beta_1(t_{ij}) + \epsilon_i(t_{ij}), \qquad j = 1, \ldots, n_i, \quad i = 1, \ldots, 200,$$

where $X_i(t)$ was the ith subject's realization of the random variable $X(t)$ from the Gaussian distribution with mean $3\exp(t/30)$ and variance 1,

$$\beta_0(t) = 15 + 20\sin\left(\frac{t\pi}{60}\right) \qquad \text{and} \qquad \beta_1(t) = 2 - 3\cos\left(\frac{(t-25)\pi}{15}\right). \quad (5.2)$$

Each individual was assigned a set of "scheduled" time points $\{0, 1, \ldots, 30\}$, and each "scheduled" time, except time 0, had a probability of 60% being skipped, so that the actual observation time points were the non-skipped "scheduled" ones. This led to unequal numbers of repeated measurements n_i and different observed time points t_{ij} per subject. The random errors $\epsilon_{ij} = \epsilon_i(t_{ij})$ were independent from the covariates and given by $\epsilon_{ij} = Z_i(t_{ij}) + E_{ij}$, where $Z_i(t_{ij})$ were generated from a stationary Gaussian process with zero mean and a covariance function

$$\mathrm{cov}(Z_{i_1}(t_{i_1j_1}), Z_{i_2}(t_{i_1j_2})) = \begin{cases} 4\exp\left(-|t_{i_1j_1} - t_{i_2j_2}|\right), & \text{if } i_1 = i_2, \\ 0, & \text{if } i_1 \neq i_2, \end{cases}$$

and the E_{ij} were independent measurement errors from a $N(0, 4)$ distribution.

We repeated this simulation process 500 times. For each simulated data set, we computed the spline estimators using cubic splines with five equally spaced interior knots and weights $w_i \equiv 1/n_i$. The procedure in Section 4.3 was used to estimate the covariance structure, also using cubic splines with five equally spaced interior knots. Asymptotic 95% pointwise confidence intervals were constructed at 61 equally spaced points on the interval $[0, 30]$ according to the procedure in Section 4.1. The empirical coverage probabilities of these pointwise intervals are close to the nominal level, as shown in Figures 2 (the standard errors of the empirical coverage probabilities were approximately 0.01). Asymptotic 95% confidence bands (4.6) were constructed with $M = 60$ and $c_1 = 3$. The simultaneous coverage probabilities were 99.2% and 98.2% for $\beta_0(t)$ and $\beta_1(t)$ respectively. As expected, the Bonferroni-type bands were conservative.

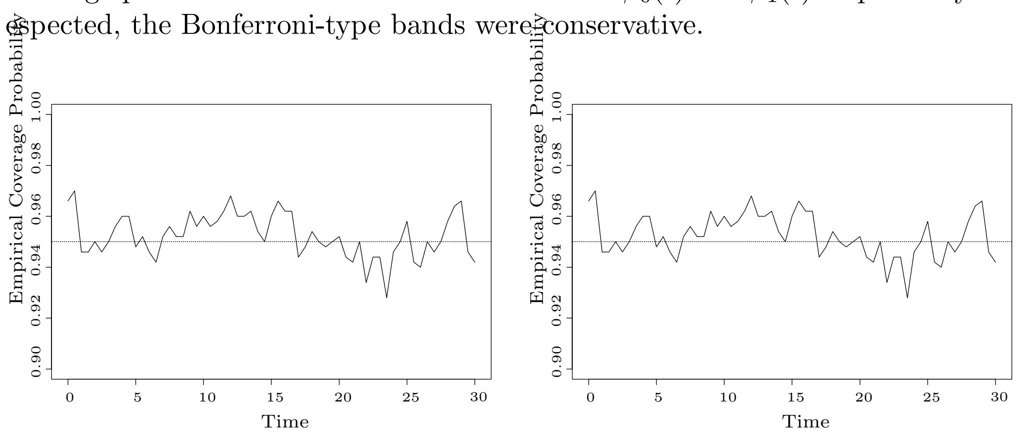

Figure 2. Empirical Coverage Probabilities of asymptotic pointwise confidence intervals for $\beta_0(t)$ (left) and $\beta_1(t)$ (right).

Acknowledgements

Jianhua Z. Huang's work was partially supported by National Science Foundation grant DMS-0204556. Partial support for Colin O. Wu was provided by the

National Institute on Drug Abuse, R01 DA10184-01, the National Science Foundation, DMS-0103832, and the Acheson J. Duncan Fund while he was with the Johns Hopkins University. The authors are grateful to Prof. Joseph Margolick for providing the MACS Public Use Data Set Release PO4 (1984-1991).

Appendix. Proofs

A.1. Notation

Let $|\boldsymbol{a}|$ denote the Euclidean norm of a real valued vector $\boldsymbol{a}$. For a matrix $A = (a_{ij})$, $\|A\|_\infty = \max_i \sum_j |a_{ij}|$. For a real valued function g on $\mathcal{T}$, $\|g\|_\infty = \sup_{t\in\mathcal{T}} |g(t)|$ denotes its supreme norm. For a vector valued function $\boldsymbol{g} = (g_0, \ldots, g_L)$, denote $\|\boldsymbol{g}\|_{L_2} = \{\sum_{0\le l\le L} \|g_l\|_{L_2}^2\}^{1/2}$ and $\|\boldsymbol{g}\|_\infty = \max_{0\le l\le L} \|g_l\|_\infty$. Given sequences of positive numbers a_n and b_n, $a_n \lesssim b_n$ and $b_n \gtrsim a_n$ mean a_n/b_n is bounded, and $a_n \asymp b_n$ means both $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ hold.

A.2. Properties of splines

We can currently choose a convenient basis system in our technical arguments and the results for the function estimates hold true for other basis choices of the same function space. We use B-splines in our proofs. For each $l = 0, \ldots, L$, $B_{lk}, k = 1, \ldots, K_l$, are the B-spline basis functions that span $\mathbb{G}_l$. The B-splines have the following properties (de Boor (1978)): $B_{lk}(t) \ge 0$, $\sum_{k=1}^{K_l} B_{lk}(t) = 1$, $t \in \mathcal{T}$;

$$\frac{M_1}{K_l}\sum_k \gamma_{lk}^2\,dt \le \int_{\mathcal{T}}\Big(\sum_k \gamma_{lk}B_{lk}(t)\Big)^2 \le \frac{M_2}{K_l}\sum_k \gamma_{lk}^2, \qquad \gamma_{lk}\in\mathbb{R}, k = 1,\ldots,K_l.$$

Moreover, there is a constant M such that $\|g\|_\infty \le M\sqrt{K_l}\|g\|$ for $g \in \mathbb{G}_l$, $l = 0, \ldots, L$.

A.3. Inner products

For $\boldsymbol{g}^{(1)}(t) = (g_0^{(1)}(t), \ldots, g_L^{(1)}(t))'$ and $\boldsymbol{g}^{(2)}(t) = (g_0^{(2)}(t), \ldots, g_L^{(2)}(t))'$, define the empirical inner product as

$$\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)}\rangle_n = \frac{1}{n}\sum_i \frac{1}{n_i}\sum_j \Big(\sum_l X_{il}(t_{ij})g_l^{(1)}(t_{ij})\Big)\Big(\sum_l X_{il}(t_{ij})g_l^{(2)}(t_{ij})\Big),$$

the theoretical inner product as

$$\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)}\rangle = E\Big[\Big(\sum_l X_l(T)g_l^{(1)}(T)\Big)\Big(\sum_l X_l(T)g_l^{(2)}(T)\Big)\Big],$$

where T is the random observation time with distribution $F_T(\cdot)$. Denote the corresponding norms by $\|\cdot\|_n$ and $\|\cdot\|$. Note that $E[\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)}\rangle_n] = \langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)}\rangle$.

Lemma A.1. *Let* $g_l(t) = \sum_k \gamma_{lk} B_{lk}(t)$ *and* $\boldsymbol{\gamma}_l = (\gamma_{l0}, \ldots, \gamma_{lK_l})'$ *for* $l = 0, \ldots, L$, *and* $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_0', \ldots, \boldsymbol{\gamma}_L')'$. *Set* $\boldsymbol{g}(t) = (g_0(t), \ldots, g_L(t))'$. *Then* $\|\boldsymbol{g}\|^2 \asymp \sum_{l=0}^L \|g_l\|_{L_2}^2 \asymp |\boldsymbol{\gamma}|^2/K_n$.

Proof. Since T is independent of $\{\boldsymbol{X}(t)\}$,

$$E\Big[\Big(\sum_l X_l(T) g_l(T)\Big)^2\Big] = \int_{\mathcal{T}} E\Big(\sum_l X_l(t) g_l(t)\Big)^2 f_T(t)dt = \int_{\mathcal{T}} \boldsymbol{g}'(t)\Sigma(t)\boldsymbol{g}(t) f_T(t)dt.$$

Thus, it follows from conditions (C1) and (C2) that,

$$\|\boldsymbol{g}\|^2 = E\Big[\Big(\sum_l X_l(T) g_l(T)\Big)^2\Big] \asymp \int_{\mathcal{T}} \boldsymbol{g}'(t)\boldsymbol{g}(t)\, dt \asymp \sum_{l=0}^L \|g_l\|_{L_2}^2,$$

uniformly in $g_l \in \mathbb{G}_l$, $l = 0, \ldots, L$. By the properties of B-spline basis functions, $\|g_l\|_{L_2}^2 \asymp |\boldsymbol{\gamma}_l|^2/K_l$, $l = 0, \ldots, L$. The conclusion then follows by condition (C5).

Lemma A.2. *Let* $\mathbb{G}$ *denote the collection of vectors of functions* $\boldsymbol{g} = (g_0, \ldots, g_L)'$ *with* $g_l \in \mathbb{G}_l$ *for* $l = 0, \ldots, L$. *Then*

$$P\Big(\sup_{\boldsymbol{g}_1, \boldsymbol{g}_2 \in \mathbb{G}} \frac{|\langle \boldsymbol{g}_1, \boldsymbol{g}_2\rangle_n - \langle \boldsymbol{g}_1, \boldsymbol{g}_2\rangle|}{\|\boldsymbol{g}_1\|\|\boldsymbol{g}_2\|} > s\Big) \le C_1 K_n^2 \exp\Big(-C_2 \frac{n}{K_n}\frac{s^2}{1+s}\Big), \qquad s > 0.$$

Consequently, if $\lim_n K_n \log K_n/n = 0$, *then* $\sup_{\boldsymbol{g} \in \mathbb{G}} |\|\boldsymbol{g}\|_n^2/\|\boldsymbol{g}\|^2 - 1| = o_P(1)$; *that is,*

$$\sup_{g_l \in G_l, l=0,\ldots,L} \left| \frac{\frac{1}{n}\sum_i \frac{1}{n_i}\sum_j \Big(\sum_l X_{il}(t_{ij}) g_l(t_{ij})\Big)^2}{E\Big(\sum_l X(T) g_l(T)\Big)^2} - 1 \right| = o_P(1).$$

Proof. Let $\mathbf{B}_{lk}(t)$ be the $(L+1)$-dimensional vector with the $(l+1)$th entry being $B_{lk}(t)$ and all other entries zero. Then $\mathbf{B}_{lk}(\cdot)$, $k = 1, \ldots, K_l$, $l = 0, \ldots, L$, constitute a basis of $\mathbb{G}$. Note that

$$\langle \mathbf{B}_{lk}, \mathbf{B}_{l'k'}\rangle_n = \frac{1}{n}\sum_i \frac{1}{n_i}\sum_j [X_{il}(t_{ij}) B_{lk}(t_{ij}) X_{il'}(t_{ij}) B_{l'k'}(t_{ij})].$$

It follows from condition (C3) and the boundedness of B-spline basis functions that $\langle \mathbf{B}_{lk}, \mathbf{B}_{l'k'}\rangle_n$ is bounded. Moreover, $\mathrm{var}(\langle \mathbf{B}_{lk}, \mathbf{B}_{l'k'}\rangle_n) \le (1/n^2)\sum_i \max_j E[X_{il}^2(t_{ij}) B_{lk}^2(t_{ij}) X_{il'}^2(t_{ij}) B_{l'k'}^2(t_{ij})]$. By conditioning on t_{ij} and using Condition (C3), we see that the right-hand side of the inequality is bounded above by a constant multiple of $\sup_{l,l',k,k'} E[B_{lk}^2(t_{ij}) B_{l'k'}^2(t_{ij})] \lesssim 1/K_n$. Consequently, $\mathrm{var}(\langle \mathbf{B}_{jk}, \mathbf{B}_{j'k'}\rangle_n) \lesssim 1/(nK_n)$. Note that $|B_{jk}(t)| \le 1$ for all j, k. Applying Bernstein's inequality, we obtain that, for some constants C_1, C_2 and C_3,

$$P(|\langle \mathbf{B}_{lk}, \mathbf{B}_{l'k'}\rangle_n - \langle \mathbf{B}_{lk}, \mathbf{B}_{l'k'}\rangle| > s) \le C_1 \exp\Big(-\frac{(ns)^2}{C_2(n/K_n) + C_3 ns}\Big), \qquad s > 0.$$

Hence, there is an event Ω_n with $P(\Omega_n^c) \le C_4 K_n^2 \exp(-C_5(n/K_n)s^2/(1+s))$ such that on Ω_n, $|\langle \mathbf{B}_{lk}, \mathbf{B}_{l'k'}\rangle_n - \langle \mathbf{B}_{lk}, \mathbf{B}_{l'k'}\rangle_n| \le s/K_n$ for all $k = 1, \ldots, K_l$, $k' = 1, \ldots, K_{l'}$ and $l, l' = 0, \ldots, L$.

For $\boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)} \in \mathbb{G}$, write $\boldsymbol{g}^{(1)} = \sum_l \sum_k \gamma_{lk}^{(1)} \mathbf{B}_{lk}$ and $\boldsymbol{g}^{(2)} = \sum_l \sum_k \gamma_{lk}^{(2)} \mathbf{B}_{lk}$. Then

$$|\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)}\rangle_n - \langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)}\rangle| = \Big|\sum_{l,k}\sum_{l',k'} \gamma_{lk}^{(1)}\gamma_{l'k'}^{(2)}(\langle \mathbf{B}_{lk}, \mathbf{B}_{l'k'}\rangle_n - \langle \mathbf{B}_{lk}, \mathbf{B}_{l'k'}\rangle)\Big|.$$

Let $(l', k') \in A(l, k)$ if the intersection of the supports of $B_{l'k'}$ and B_{lk} contains an open interval. Then $\langle \mathbf{B}_{lk}, \mathbf{B}_{l'k'}\rangle_n = \langle \mathbf{B}_{lk}, \mathbf{B}_{l'k'}\rangle = 0$ if $(l', k') \notin A(l, k)$. Moreover, $\#\{A(l, k)\} \le C$ for some constant C for all l, k, and, on Ω_n,

$$|\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)}\rangle_n - \langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)}\rangle| \le \sum_{l,k}\sum_{l',k'} |\gamma_{lk}^{(1)}||\gamma_{l'k'}^{(2)}| \frac{s}{K_n} \operatorname{ind}\{(l, k) \in A(l', k')\}. \quad \text{(A.1)}$$

Applying the Cauchy-Schwarz inequality twice, we see that the right-hand side of the above display is bounded above by

$$\begin{aligned} &\frac{s}{K_n}\sum_{l,k}|\gamma_{lk}^{(1)}|\Big\{\sum_{l',k'}|\gamma_{l'k'}^{(2)}|^2 \operatorname{ind}\{(l,k)\in A(l',k')\}\Big\}^{1/2} C^{1/2} \\ &\le \frac{s}{K_n}\Big(\sum_{l,k}|\gamma_{lk}^{(1)}|^2\Big)^{1/2}\Big(\sum_{l,k}\sum_{l',k'}|\gamma_{l'k'}^{(2)}|^2 \operatorname{ind}\{(l,k)\in A(l',k')\}\Big)^{1/2} C^{1/2} \\ &\le \frac{s}{K_n} C |\boldsymbol{\gamma}^{(1)}||\boldsymbol{\gamma}^{(2)}|, \end{aligned}$$

where $\boldsymbol{\gamma}^{(1)}$ and $\boldsymbol{\gamma}^{(2)}$ denote respectively the vectors with entries $\gamma_{lk}^{(1)}$ and $\gamma_{lk}^{(2)}$. It follows from Lemma A.1 that $\|\boldsymbol{g}^{(i)}\|^2 \asymp |\boldsymbol{\gamma}^{(i)}|^2/K_n$ for $i = 1, 2$. Hence, on Ω_n, $|\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)}\rangle_n - \langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)}\rangle| \le Cs\|\boldsymbol{g}^{(1)}\|\|\boldsymbol{g}^{(2)}\|$ for some constant C. The conclusions of the lemma follow.

A.4. Proof of Theorem 1: Consistency

The existence of $\boldsymbol{\gamma}$ and thus of $\hat{\beta}_l$, $l = 0, \ldots, L$, follows from (2.2) and the following Lemma. The consistency of $\hat{\beta}_l$ is a consequence of Theorem 2.

Lemma A.3. *There are positive constants M_1 and M_2 such that, except on an event whose probability tends to zero, all the eigenvalues of $(K_n/n)\mathbf{U}'\mathbf{W}\mathbf{U}$ fall between M_1 and M_2, and consequently, $\mathbf{U}'\mathbf{W}\mathbf{U} = \sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i$ is invertible.*

Proof. Set $\boldsymbol{g}_{\boldsymbol{\gamma}} = \sum_{l,k} \gamma_{lk}\mathbf{B}_{lk}$ for $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_0', \ldots, \boldsymbol{\gamma}_L')'$ with $\boldsymbol{\gamma}_l = (\gamma_{l1}, \ldots, \gamma_{lK_l})'$. By Lemmas A.1 and A.2, except on an event whose probability tends to zero, $\|\boldsymbol{g}_{\boldsymbol{\gamma}}\|_n^2 \asymp \|\boldsymbol{g}_{\boldsymbol{\gamma}}\|^2 \asymp |\boldsymbol{\gamma}|^2/K_n$. Note that $\boldsymbol{\gamma}'(\mathbf{U}'\mathbf{W}\mathbf{U}/n)\boldsymbol{\gamma} = \|\boldsymbol{g}_{\boldsymbol{\gamma}}\|_n^2$. The desired result follows.

A.5. Proof of Theorem 2: Rates of convergence

Set $\tilde{Y}_{ij} = \boldsymbol{X}_i(t_{ij})'\boldsymbol{\beta}(t_{ij})$, $\tilde{\mathbf{Y}}_i = (\tilde{Y}_{i1}, \ldots, \tilde{Y}_{in_i})'$, $\tilde{\mathbf{Y}} = (\tilde{\mathbf{Y}}_i, \ldots, \tilde{\mathbf{Y}}_n)'$ and

$$\tilde{\boldsymbol{\gamma}} = \Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1} \sum_i \mathbf{U}_i'\mathbf{W}_i\tilde{\mathbf{Y}}_i. \tag{A.2}$$

Then $E(\hat{\boldsymbol{\gamma}}) = \tilde{\boldsymbol{\gamma}}$ and $E(\hat{\boldsymbol{\beta}}(t)) = \tilde{\boldsymbol{\beta}}(t) = \mathbf{B}(t)\tilde{\boldsymbol{\gamma}}$, $t \in \mathcal{T}$, where the expectation is taken conditioning on $\mathcal{X}$. By the triangle inequality, $|\hat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}(t)| \le |\hat{\boldsymbol{\beta}}(t)-\tilde{\boldsymbol{\beta}}(t)|+|\tilde{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}(t)|$. In Lemmas A.5 and A.7 we give upper bounds of the L_2-norms of $\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}$, respectively, from which Theorem 2 follows.

Lemma A.4. $|(\mathbf{U}'\mathbf{W}\mathbf{U})^{-1}\mathbf{U}'\mathbf{W}\boldsymbol{\epsilon}|^2 = O_P((K_n^2/n^2)\sum_i\{(1/n_i) + (1/K_n)(1 - (1/n_i))\})$.

Proof. Note that

$$|(\mathbf{U}'\mathbf{W}\mathbf{U})^{-1}\mathbf{U}'\mathbf{W}\boldsymbol{\epsilon}|^2 = \frac{K_n^2}{n^2}\boldsymbol{\epsilon}'\mathbf{W}\mathbf{U}\Big(\frac{K_n}{n}\mathbf{U}'\mathbf{W}\mathbf{U}\Big)^{-1}\Big(\frac{K_n}{n}\mathbf{U}'\mathbf{W}\mathbf{U}\Big)^{-1}\mathbf{U}'\mathbf{W}\boldsymbol{\epsilon}.$$

It follows from Lemma A.3 that

$$|(\mathbf{U}'\mathbf{W}\mathbf{U})^{-1}\mathbf{U}'\mathbf{W}\boldsymbol{\epsilon}|^2 \asymp \frac{K_n^2}{n^2}\boldsymbol{\epsilon}'\mathbf{W}\mathbf{U}\mathbf{U}'\mathbf{W}\boldsymbol{\epsilon} = \frac{K_n^2}{n^2}\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\boldsymbol{\epsilon}_i\Big)'\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\boldsymbol{\epsilon}_i\Big),$$

except on an event whose probability tends to zero. Recall that $\mathbf{U}_i = (\mathbf{U}_{i1}, \ldots, \mathbf{U}_{in_i})'$, $\mathbf{W}_i = \mathrm{diag}(1/n_i, \ldots, 1/n_i)$ and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})'$. So, $\mathbf{U}_i'\mathbf{W}_i\boldsymbol{\epsilon}_i = (1/n_i)\sum_j \mathbf{U}_{ij}\epsilon_{ij}$. Since

$$\mathbf{U}_{ij} = (X_{i0}(t_{ij})B_{01}(t_{ij}) \cdots X_{i0}(t_{ij})B_{0k_0}(t_{ij}) \cdots X_{iL}(t_{ij})B_{L1}(t_{ij}) \cdots \\ X_{iL}(t_{ij})B_{Lk_L}(t_{ij}))',$$

we have that $|\mathbf{U}_i'\mathbf{W}_i\boldsymbol{\epsilon}_i|^2 = (1/n_i^2)\sum_{l,k}(\sum_j X_{il}(t_{ij})B_{lk}(t_{ij})\epsilon_{ij})^2$. By conditions (C3) and (C4), and properties of B-splines, $E[X_{il}^2(t_{ij})B_{lk}^2(t_{ij})\epsilon_{ij}^2] \le C/K_n$, and

$$E|X_{il}(t_{ij})B_{lk}(t_{ij})\epsilon_{ij}X_{il}(t_{ij'})B_{lk}(t_{ij'})\epsilon_{ij'}| \le CE[B_{lk}(t_{ij})]E[B_{lk}(t_{ij'})] \\ \le C\frac{1}{K_n^2}, \qquad j \ne j'.$$

Thus, $E(|\mathbf{U}_i'\mathbf{W}_i\boldsymbol{\epsilon}_i|^2) \lesssim 1/n_i + (1/K_n)(1 - 1/n_i)$. Consequently, $E[\boldsymbol{\epsilon}'\mathbf{W}\mathbf{U}\mathbf{U}'\mathbf{W}\boldsymbol{\epsilon}] = \sum_i E(|\mathbf{U}_i'\mathbf{W}_i\boldsymbol{\epsilon}_i|^2) \lesssim \sum_i\{1/n_i + (1/K_n)(1 - 1/n_i)\}$. Hence, $\boldsymbol{\epsilon}'\mathbf{W}\mathbf{U}\mathbf{U}'\mathbf{W}\boldsymbol{\epsilon} = O_P(\sum_i\{1/n_i + (1/K_n)(1 - 1/n_i)\})$. The conclusion of the lemma follows.

Lemma A.5. $\|\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\|_{L_2}^2 = O_P((K_n/n^2)\sum_i\{(1/n_i) + (1/K_n)(1 - (1/n_i))\})$.

Proof. Since $\hat{\boldsymbol{\beta}}(t) = \mathbf{B}(t)\hat{\boldsymbol{\gamma}}$ and $\tilde{\boldsymbol{\beta}}(t) = \mathbf{B}(t)\tilde{\boldsymbol{\gamma}}$, $\|\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\|_{L_2}^2 \asymp |\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}|^2/K_n$. On the other hand,

$$\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}} = \Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1}\sum_i \mathbf{U}_i'\mathbf{W}_i(\mathbf{Y}_i - \tilde{\mathbf{Y}}_i) = \Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1}\sum_i \mathbf{U}_i'\mathbf{W}_i\boldsymbol{\epsilon}_i.$$

Thus, the result follows from Lemma A.4.

Lemma A.6. *If* $\lim_n K_n \log K_n/n = 0$*, then there is a constant* C *such that, except on an event whose probability tends to zero,* $\sup_{l,k}(1/n)\sum_i(1/n_i)\sum_j B_{lk}(t_{ij}) \le (C/K_n)$.

Proof. Observe that

$$\operatorname{var}\Big(\frac{1}{n}\sum_i \frac{1}{n_i}\sum_j B_{lk}(t_{ij})\Big) = \frac{1}{n^2}\sum_i \operatorname{var}\Big(\frac{1}{n_i}\sum_j B_{lk}(t_{ij})\Big) \le \frac{1}{n}E[B_{lk}(T)] \lesssim \frac{1}{nK_n}.$$

Using the Bernstein inequality and the argument as in Lemma A.2, we obtain that

$$P\Big(\sup_{l,k}\Big|\frac{1}{n}\sum_i\frac{1}{n_i}\sum_j B_{lk}(t_{ij}) - E[B_{lk}(T)]\Big| > \frac{s}{K_n}\Big) \le C_1 K_n^2 \exp\Big(-C_2\frac{n}{K_n}\frac{s^2}{1+s}\Big),\ s>0.$$

Noting that $E[B_{lk}(T)] \asymp 1/K_n$, the result follows.

Let $\boldsymbol{g}^* = (g_0^*,\ldots,g_L^*) \in \mathbb{G}$ be such that $\|\boldsymbol{g}^* - \boldsymbol{\beta}\|_\infty = \rho_n = \inf_{\boldsymbol{g}\in\mathbb{G}}\|\boldsymbol{g}-\boldsymbol{\beta}\|_\infty$. Then there exists $\boldsymbol{\gamma}^* = (\boldsymbol{\gamma}_0^{*\prime},\ldots,\boldsymbol{\gamma}_L^{*\prime})'$ with $\boldsymbol{\gamma}_l^* = (\gamma_{l1}^*,\ldots,\gamma_{lK_l}^*)'$ such that $\boldsymbol{g}^*(t) = \mathbf{B}(t)\boldsymbol{\gamma}^* = \sum_{l,k}\gamma_{lk}^* B_{lk}(t)$. Recall that $\tilde{\boldsymbol{\beta}}(t) = \mathbf{B}(t)\tilde{\boldsymbol{\gamma}}$ where $\tilde{\boldsymbol{\gamma}}$ is given in (A.2). By the triangle inequality, $|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}| \le |\tilde{\boldsymbol{\beta}} - \boldsymbol{g}^*| + |\boldsymbol{g}^* - \boldsymbol{\beta}|$.

Lemma A.7. $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{g}^*\|_{L_2} = O_P(\rho_n)$. *Consequently,* $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_{L_2} = O_P(\rho_n)$.

Proof. By the properties of B-spline basis functions, $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{g}^*\|_{L_2} = \|\mathbf{B}\tilde{\boldsymbol{\gamma}} - \mathbf{B}\boldsymbol{\gamma}^*\|_{L_2} \asymp |\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*|/\sqrt{K_n}$. Note that

$$\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^* = \Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1}\sum_i \mathbf{U}_i'\mathbf{W}_i\tilde{\mathbf{Y}}_i - \Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1}\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\boldsymbol{\gamma}^*.$$

Arguing as in the proof of Lemma A.4, we obtain that

$$\begin{aligned} |\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*|^2 &\asymp \frac{K_n^2}{n^2}\Big|\sum_i \mathbf{U}_i'\mathbf{W}_i(\tilde{\mathbf{Y}}_i - \mathbf{U}_i\boldsymbol{\gamma}^*)\Big|^2 \\ &= \frac{K_n^2}{n^2}\sum_{l,k}\Big(\sum_i\frac{1}{n_i}\sum_j X_{il}(t_{ij})B_{lk}(t_{ij})(\tilde{\mathbf{Y}}_i - \mathbf{U}_i\boldsymbol{\gamma}^*)_j\Big)^2. \end{aligned}$$

Since $(\tilde{\mathbf{Y}}_i - \mathbf{U}_i\boldsymbol{\gamma}^*)_j = \boldsymbol{X}_i'(t_{ij})\boldsymbol{\beta}(t_{ij}) - \mathbf{U}_{ij}'\boldsymbol{\gamma}^* = \boldsymbol{X}_i'(t_{ij})\boldsymbol{\beta}(t_{ij}) - \boldsymbol{X}_i'(t_{ij})\mathbf{B}(t_{ij})\boldsymbol{\gamma}^*$, it follows from condition (C3) that $|(\tilde{\mathbf{Y}}_i - \mathbf{U}_i\boldsymbol{\gamma}^*)_j| \lesssim \|\boldsymbol{\beta} - \boldsymbol{g}^*\|_\infty = \rho_n$. Thus, by the Cauchy–Schwarz inequality, Lemma A.6 and condition (C3),

$$|\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*|^2 \le K_n^2\rho_n^2\sum_{l,k}\Big(\frac{1}{n}\sum_i\frac{1}{n_i}\sum_j X_{il}^2(t_{ij})B_{lk}(t_{ij})\Big)\Big(\frac{1}{n}\sum_i\frac{1}{n_i}\sum_j B_{lk}(t_{ij})\Big) \le K_n\rho_n^2.$$

The desired result follows.

A.6. Proof of Theorem 3: Asymptotic normality

Lemma A.8. *If* $\lim_n K_n \max_i n_i/n = 0$*, then for any* c_n *whose components are not all zero,* $c_n'(\hat{\gamma}-\tilde{\gamma})/\mathrm{SD}\{c_n'(\hat{\gamma}-\tilde{\gamma})\} \longrightarrow N(0,1)$ *in distribution, where*

$$\mathrm{SD}\{c_n'(\hat{\gamma}-\tilde{\gamma})\} = c_n'\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1}\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{V}_i\mathbf{W}_i\mathbf{U}_i\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1}c_n.$$

Proof. Note that $c_n'(\hat{\gamma}-\tilde{\gamma}) = \sum_i c_n'(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i)^{-1}\mathbf{U}_i'\mathbf{W}_i\boldsymbol{\epsilon}_i = \sum_i a_i\xi_i$, where $a_i^2 = c_n'(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i)^{-1}\ U_i'\mathbf{W}_i\mathbf{V}_i\mathbf{W}_i\mathbf{U}_i(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i)^{-1}c_n$ and, conditioning on $\{\boldsymbol{X}_i(t),\ i=1,\ldots,n\}$, ξ_i are independent with mean zero and variance one. It follows easily by checking Lindeberg condition that if

$$\frac{\max_i a_i^2}{\sum_i a_i^2} \to_P 0, \tag{A.3}$$

then $\sum_i a_i\xi_i/\sqrt{\sum_i a_i^2}$ is asymptotically $N(0,1)$.

We only need to show that (A.3) holds. Since $E[\epsilon^2(t)] \le C$, it follows that, for $\boldsymbol{\theta}_i = (\theta_{i1},\ldots,\theta_{in_i})'$,

$$\boldsymbol{\theta}_i V_i \boldsymbol{\theta}_i = E\Big[\Big(\sum_j \theta_{ij}\epsilon_i(t_{ij})\Big)^2\Big] \le |\boldsymbol{\theta}_i|^2 \sum_j E[\epsilon_i^2(t_{ij})] \le Cn_i|\boldsymbol{\theta}_i|^2.$$

Hence, for any $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_L)'$ with $\boldsymbol{\lambda}_l = (\lambda_{l1},\ldots,\lambda_{l,K_l})'$,

$$\begin{aligned}\boldsymbol{\lambda}'\mathbf{U}_i'\mathbf{W}_i\mathbf{V}_i\mathbf{W}_i\mathbf{U}_i\boldsymbol{\lambda} \lesssim n_i\boldsymbol{\lambda}'\mathbf{U}_i'\mathbf{W}_i\mathbf{W}_i\mathbf{U}_i\boldsymbol{\lambda} &= \frac{1}{n_i}\sum_j \boldsymbol{\lambda}'\mathbf{U}_{ij}\mathbf{U}_{ij}'\boldsymbol{\lambda}\\ &= \frac{1}{n_i}\sum_j\Big(\sum_l X_{il}(t_{ij})g_{\boldsymbol{\lambda},l}(t_{ij})\Big)^2,\end{aligned}$$

where $g_{\boldsymbol{\lambda},l}(t) = \sum_k \lambda_{lk}B_{lk}(t)$ for $l=0,\ldots,L$. Observe that

$$\begin{aligned}\Big(\sum_l X_{il}(t_{ij})g_{\boldsymbol{\lambda},l}(t_{ij})\Big)^2 &\le \sum_l X_{il}^2(t_{ij})\sum_l g_{\boldsymbol{\lambda},l}^2(t_{ij}) \lesssim \sum_l \|g_{\boldsymbol{\lambda},l}\|_\infty^2\\ &\lesssim K_n\sum_l \|g_{\boldsymbol{\lambda},l}\|^2 \lesssim |\boldsymbol{\lambda}|^2.\end{aligned}$$

Consequently,

$$\boldsymbol{\lambda}'\mathbf{U}_i'\mathbf{W}_i\mathbf{V}_i\mathbf{W}_i\mathbf{U}_i\boldsymbol{\lambda} \lesssim |\boldsymbol{\lambda}|^2. \tag{A.4}$$

Under condition (C6), we have that $\mathbf{V}_i = \sigma^2 I_i + \widetilde{\mathbf{V}}_i$, where I_i is the $n_i\times n_i$ identity matrix and $\widetilde{\mathbf{V}}_i$ is the $n_i\times n_i$ matrix with (j,j') entry $E[\epsilon^{(2)}(t_{ij})\epsilon^{(2)}(t_{ij'})]$.

Note that $\widetilde{\mathbf{V}}_i$ is nonnegative definite,

$$\begin{aligned}\boldsymbol{\lambda}'\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{V}_i\mathbf{W}_i\mathbf{U}_i\Big)\boldsymbol{\lambda} &\geq \sigma^2\boldsymbol{\lambda}'\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{W}_i\mathbf{U}_i\Big)\boldsymbol{\lambda} \\ &= n\sigma^2\Big\{\frac{1}{n}\sum_i \frac{1}{n_i^2}\sum_j\Big(\sum_l X_{il}(t_{ij})g_{\boldsymbol{\lambda},l}(t_{ij})\Big)^2\Big\} \\ &\geq \sigma^2 \min_i \frac{n}{n_i}\|\boldsymbol{g}_{\boldsymbol{\lambda}}\|_n^2,\end{aligned}$$

where $\boldsymbol{g}_{\boldsymbol{\lambda}} = (g_{\boldsymbol{\lambda},0},\ldots,g_{\boldsymbol{\lambda},L})'$. By Lemmas A.1 and A.2, there is an event Ω_n with $P(\Omega_n)\to 1$ such that on Ω_n, $\|\boldsymbol{g}_{\boldsymbol{\lambda}}\|_n^2 \asymp \|\boldsymbol{g}_{\boldsymbol{\lambda}}\|^2 \asymp |\boldsymbol{\lambda}|^2/K_n$. Thus, on Ω_n,

$$\boldsymbol{\lambda}'\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{V}_i\mathbf{W}_i\mathbf{U}_i\Big)\boldsymbol{\lambda} \gtrsim \frac{n}{\max_i n_i}\frac{1}{K_n}|\boldsymbol{\lambda}|^2. \tag{A.5}$$

Combining (A.4) and (A.5), we obtain that

$$\frac{\max_i \boldsymbol{\lambda}'\mathbf{U}_i'\mathbf{W}_i\mathbf{V}_i\mathbf{W}_i\mathbf{U}_i\boldsymbol{\lambda}}{\boldsymbol{\lambda}'\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{V}_i\mathbf{W}_i\mathbf{U}_i\Big)\boldsymbol{\lambda}} \lesssim \max_i n_i \frac{K_n}{n},$$

except on an event whose probability tends to zero. Hence (A.3) follows from the requirement that $\lim_n K_n \max_i n_i/n = 0$. The proof of Lemma A.8 is complete.

A.7. Proof of Theorem 4: Bound in L_∞-norm of the bias term

The proof of Theorem 4 is broken up into three lemmas: Lemmas A.9−A.11. Note the following identity

$$\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^* = K_n\Big(\frac{K_n}{n}\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1}\frac{1}{n}\sum_i \mathbf{U}_i'\mathbf{W}_i(\tilde{\mathbf{Y}}_i - \mathbf{U}_i\boldsymbol{\gamma}^*). \tag{A.6}$$

Lemma A.9. *There is an absolute constant C that does not depend on n such that*

$$\Big\|\Big(\frac{K_n}{n}\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1}\Big\|_\infty \leq C,$$

except on an event whose probability tends to zero as $n\to\infty$.

Proof. We use the following result of Demko (1977). If A is a symmetric matrix such that A^s has no more than Ms nonzero entries in each row for every positive integer s, then $\|A^{-1}\|_\infty \leq 33\sqrt{M}\|A^{-1}\|_2^{5/4}\|A\|_2^{1/4}$; here $\|A\|_2 = (\sup_u\{|Au|/|u|\})^{1/2}$ is a matrix norm.

Set $A = (K_n/n)\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i$. It follows from Lemma A.3 that both $\|A\|_2$ and $\|A^{-1}\|_2$ are bounded. We say that two B-splines B_{lk} and $B_{l'k'}$ overlap if the collection of t such that $B_{lk}(t)B_{l'k'}(t) \neq 0$ contains an open interval. Since $\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i = \sum_i(1/n_i)\sum_j \mathbf{U}_{ij}\mathbf{U}_{ij}'$, the $(lk, l'k')$-element of A is $(K_n/n)\sum_i(1/n_i)$ $\sum_j X_{il}(t_{ij})B_{lk}(t_{ij})X_{il'}(t_{ij})B_{l'k'}(t_{ij})$, which is non-zero only when B_{lk} and $B_{l'k'}$ overlap. Fix any positive integer s. The $(lk, l'k')$-element of A^s is non-zero only when there is a sequence of B-splines $B^{(0)} = B_{lk}, B^{(1)}, \ldots, B^{(s-1)}, B^{(s)} = B_{l'k'}$ (not necessarily different) chosen from $\{B_{lk}, k = 1, \ldots, K_l, l = 0, \ldots, L\}$ such that $B^{(i)}$ and $B^{(i+1)}$ overlap for $i = 0, \ldots, s-1$. Hence, it follows from the properties of B-splines that there is a positive constant M such that the number of non-zero elements in each row of A^s is bounded above by Ms for every s. The desired result now follows from the cited result of Demko.

Lemma A.10. $|(1/n)\sum_i \mathbf{U}_i'\mathbf{W}_i(\tilde{\mathbf{Y}}_i - \mathbf{U}_i\boldsymbol{\gamma}^*)|_\infty = O_P(\rho_n/K_n)$.

Proof. Observe that $\sum_i \mathbf{U}_i'\mathbf{W}_i(\tilde{\mathbf{Y}}_i - \mathbf{U}_i\boldsymbol{\gamma}^*) = \sum_i(1/n_i)\sum_j \mathbf{U}_{ij}(\tilde{\mathbf{Y}}_i - \mathbf{U}_i\boldsymbol{\gamma}^*)_j$. Since $\mathbf{U}_{ij}' = \boldsymbol{X}_i'(t_{ij})\mathbf{B}(t_{ij})$,

$$\left|\frac{1}{n}\sum_i \mathbf{U}_i'\mathbf{W}_i(\tilde{\mathbf{Y}}_i - \mathbf{U}_i\boldsymbol{\gamma}^*)\right|_\infty \leq \max_{l,k}\left|\frac{1}{n}\sum_i \frac{1}{n_i}\sum_j X_{il}(t_{ij})B_{lk}(t_{ij})(\tilde{\mathbf{Y}}_i - \mathbf{U}_i\boldsymbol{\gamma}^*)_j\right|$$

$$\leq \max_l \sup_t |X_{il}(t)| \max_j |(\tilde{\mathbf{Y}}_i - \mathbf{U}_i\boldsymbol{\gamma}^*)_j| \max_{l,k}\left(\frac{1}{n}\sum_i \frac{1}{n_i}\sum_j B_{lk}(t_{ij})\right).$$

From the proof of Lemma A.7, $\max_j |(\tilde{\mathbf{Y}}_i - \mathbf{U}_i\boldsymbol{\gamma}^*)_j| = O_P(\rho_n)$. The conclusion then follows from Condition (C3) and Lemma A.6.

Lemma A.11. $|\tilde{\gamma} - \boldsymbol{\gamma}^*|_\infty = O_P(\rho_n)$. *Consequently,* $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty = O_P(\rho_n)$.

Proof. It follows from (A.6), Lemmas A.9 and A.10 that $|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*|_\infty = O_P(\rho_n)$. By properties of B-splines, $\|\tilde{\boldsymbol{\beta}} - g^*\|_\infty = \max_l \sup_t |\sum_k(\tilde{\gamma}_{lk} - \gamma_{lk}^*)B_{lk}(t)| \leq \max_l \sup_t \sum_k B_{lk}(t)|\tilde{\gamma} - \boldsymbol{\gamma}^*|_\infty \lesssim |\tilde{\gamma} - \boldsymbol{\gamma}^*|_\infty$. Thus, by the triangle inequality, $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty \leq \|\tilde{\boldsymbol{\beta}} - g^*\|_\infty + \|g^* - \boldsymbol{\beta}\|_\infty = O_P(\rho_n)$.

A.8. Proof of Corollary 1

It follows from (A.5), Lemma A.3, and the properties of B-splines that

$$\begin{aligned}
&\operatorname{var}(\hat{\beta}_l(t)) \\
&= e_{l+1}'\mathbf{B}(t)\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1}\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{V}_i\mathbf{W}_i\mathbf{U}_i\Big)\Big(\sum_i \mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i\Big)^{-1}\mathbf{B}'(t)e_{l+1} \\
&\gtrsim \frac{K_n}{n\max_i n_i}\sum_{k=1}^{K_l} B_{lk}^2(t) \gtrsim \frac{K_n}{n\max_i n_i}.
\end{aligned}$$

On the other hand, $\rho_n \asymp K_n^{-2}$. The conclusion thus follows from Theorem 4.

A.9. Technical supplement for deterministic observation times

The main argument can be modified to handle the case when the observation times $\{t_{ij}\}$ are deterministic. For $\boldsymbol{g}^{(1)}(t) = (g_0^{(1)}(t),\ldots,g_L^{(1)}(t))'$ and $\boldsymbol{g}^{(2)}(t) = (g_0^{(2)}(t),\ldots,g_L^{(2)}(t))'$, redefine the theoretical inner product as

$$\begin{aligned}\langle \boldsymbol{g}^{(1)},\boldsymbol{g}^{(2)}\rangle &= \frac{1}{n}\sum_i \frac{1}{n_i}\sum_j E\Big[\Big(\sum_l X_{il}(t_{ij})g_l^{(1)}(t_{ij})\Big)\Big(\sum_l X_{il}(t_{ij})g_l^{(2)}(t_{ij})\Big)\Big]\\ &= \frac{1}{n}\sum_i \frac{1}{n_i}\sum_j \boldsymbol{g}^{(1)\prime}(t_{ij})E[\boldsymbol{X}(t_{ij})\boldsymbol{X}'(t_{ij})]\boldsymbol{g}^{(2)}(t_{ij}).\end{aligned}$$

Lemma A.1 now follows from Condition (C2) and (3.1). In fact, for $\boldsymbol{g} = (g_0,\ldots, g_L)'$,

$$\|\boldsymbol{g}\|^2 \asymp \frac{1}{n}\sum_i \frac{1}{n_i}\sum_j \boldsymbol{g}'(t_{ij})\boldsymbol{g}(t_{ij}) = \frac{1}{n}\sum_i \frac{1}{n_i}\sum_j\sum_l g_l^2(t_{ij}) \asymp \sum_{l=0}^{L}\|g_l\|_{L_2}^2.$$

Lemma A.6 is also a consequence of (3.1).

A.10. Proof of the claim in Remark 3.1

Integrating by parts and using the fact $F_n(1)-F_T(1) = F_n(0)-F_T(0) = 1$, we obtain that

$$\int_{\mathcal{T}} g^2(t)\,dF_n(t) - \int_{\mathcal{T}} g^2\,dF_T(t) = -\int_{\mathcal{T}}[F_n(t)-F_T(t)]g(t)g'(t)\,dt, \qquad g\in G_l.$$

Note that $\|g'\|_{L_2} \lesssim K_n\|g\|_{L_2}$ by Theorem 5.1.2 of DeVore and Lorentz 1993. Thus,

$$\begin{aligned}\Big|\int_{\mathcal{T}} g^2(t)\,dF_n(t) - \int_{\mathcal{T}} g^2\,dF_T(t)\Big| &\le \sup_{t\in\mathcal{T}}|F_n(t)-F_T(t)|\|g\|_{L_2}\|g'\|_{L_2}\\ &\lesssim o(1/K_n)K_n\|g\|_{L_2}^2 = o(\|g\|_{L_2}^2).\end{aligned}$$

On the other hand, since the density $f_T(t)$ of F_T is bounded away from 0 and infinity uniformly over $t\in\mathcal{T}$, $\int_{\mathcal{T}} g^2\,dF_T(t) \asymp \|g\|_{L_2}^2$, uniformly in $g\in G_l$, $l = 0,\ldots,L$. Consequently, there are constants M_1 and M_2 such that

$$M_1\|g\|_{L_2}^2 \le \frac{1}{n}\sum_i\frac{1}{n_i}\sum_j g^2(t_{ij}) = \int_{\mathcal{T}} g^2(t)dF_n(t) \le M_2\|g\|_{L_2}^2, \quad g\in G_l, l=0,\ldots,L.$$

References

Besse, P., Cardot, H. and Ferraty, F. (1997). Simultaneous non-parametric regressions of unbalanced longitudinal data. *Comput. Statist. Data Anal.* **24**, 255-270.

de Boor, C. (1978). *A Practical Guide to Splines*. Springer, New York.

Brumback, B. A. and Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *J. Amer. Statist. Assoc.* **93**, 961-976.

Cheng, S. C. and Wei, L. J. (2000). Inferences for a semi-parametric model with panel data. *Biometrika* **87**, 89-97.

Chiang, C. T., Rice, J. A. and Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *J. Amer. Statist. Assoc.* **96**, 605-619.

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

Demko, S. (1977). Inverses of band matrices and local convergence of spline projections. *SIAM J. Numer. Anal.* **14**, 616-619.

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation*. Springer-Verlag, Berlin.

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 University Press, Oxford.

Diggle, P. J. and Verbyla, A. P. (1998). Nonparametric Estimation of Covariance Structure in Longitudinal Data. *Biometrics* **54**, 401-415.

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. 2nd edition. Marcel Dekker, New York.

Fan, J. and Marron, J. S. (1994). Fast implementations of nonparametric curve estimators. *J. Comput. Graph. Statist.* **3**, 35-56.

Fan, J. and Zhang, J.-T. (2000). Functional linear models for longitudinal data. *J. Roy. Statist. Soc. Ser. B* **62**, 303-322.

Hall, P. and Titterington, D. M. (1988). On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal.* **27**, 228-254.

Hansen, M. H. and Kooperberg, C. (2002). Spline adaptation in extended linear models (with discussion). *Statist. Sci.* **17**, 2-51.

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer-Verlag, New York.

Hart, J. D. and Wehrly, T. E. (1986). Kernel regression estimation using repeated measurements data. *J. Amer. Statist. Assoc.* **81**, 1080-1088.

Hart, J. D. and Wehrly, T. E. (1993). Consistency of cross-validation when the data are curves. *Stochastic Process. Appl.* **45**, 351-361.

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

Huang, J. Z. (1998). Projection estimation in multiple regression with applications to functional ANOVA models. *Ann. Statist.* **26**, 242-272.

Huang, J. Z. (2001). Concave extended linear modeling: A theoretical synthesis. *Statist. Sinica* **11**, 173-197.

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.* **31**, 1600-1635.

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

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. *Amer. J. Epidemiology* **126**, 310-318.

Knafl, G., Sacks, J. and Ylvisaker, D. (1985). Confidence bands for regression functions. *J. Amer. Statist. Assoc.* **80**, 683-691.

Lin, Z. and Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *J. Amer. Statist. Assoc.* **95**, 520-534.

Moyeed, R. A. and Diggle, P. J. (1994). Rates of convergence in semiparametric modeling of longitudinal data. *Austral. J. Statist.* **36**, 75-93.

Naiman, D. and Wynn, H. (1997). Abstract tubes, improved inclusion-exclusion identities and inequalities and importance sampling. *Ann. Statist.* **36**, 75-93.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53**, 233-243.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**, 253-259.

Schumaker, L. L. (1981). *Spline Functions: Basic Theory.* Wiley, New York.

Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *J. Econometrics* **75**, 317-343,

Staniswalis, J. G. and Lee, J. J. (1998). Nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **93**, 1403-1418.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 1348-1360.

Stone, C. J., Hansen, M., Kooperberg, C. and Truong, Y. (1997). Polynomial splines and their tensor products in extended linear modeling (with discussion). *Ann. Statist.* **25**, 1371-1470.

Stone, C. J. and Huang, J. Z. (2002). Free knot splines in concave extended linear modeling. *J. Statist. Plann. Inference* **108**, 219-253.

Verbeke, G. and Mollenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.

Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker, New York.

Wu, C. O. and Chiang, C.-T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statist. Sinica* **10**, 433-456.

Wu, C. O., Chiang, C.-T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **93**, 1388-1402.

Wu, C. O., Yu, K. F. and Chiang, C.-T. (2000). A two-step smoothing method for varying-coefficient models with repeated measurements. *Ann. Inst. Statist. Math.* **52**, 519-543.

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.

Department of Statistics, The Wharton School, 400 Jon M. Huntsman Hall, 3730 Walnut Street, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

E-mail: jianhua@wharton.upenn.edu

Office of Biostatistics Research, DECA, National Heart, Lung and Blood Institute, 2 Rockledge Center, Room 8218, 6701 Rockledge Drive, MSC 7938, Bethesda, MD 20892, U.S.A.

E-mail: wuc@nhlbi.nih.gov

Center of Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A.

E-mail: lzhou@cceb.upenn.edu

(Received December 2002; accepted May 2003)