Smoothing Spline Estimation for Varying Coefficient Models With Repeatedly Measured Dependent Variables

Chin-Tsang CHIANG, John A. RICE, and Colin O. WU

Longitudinal samples, i.e., datasets with repeated measurements over time, are common in biomedical and epidemiological studies such as clinical trials and cohort observational studies. An exploratory tool for the analyses of such data is the varying coefficient model $Y(t) = \mathbf{X}^T(t)\boldsymbol{\beta}(t)+\epsilon(t)$, where $Y(t)$ and $\mathbf{X}(t) = (X^{(0)}(t), \ldots, X^{(k)}(t))^T$ are the response and covariates at time t, $\beta(t) = (\beta_0(t), \ldots, \beta_k(t))^T$ are smooth coefficient curves of t and $\epsilon(t)$ is a mean zero stochastic process. A special case that is of particular interest in many situations is data with time-dependent response and time-independent covariates. We propose in this article a componentwise smoothing spline method for estimating $\beta_0(t), \ldots, \beta_k(t)$ nonparametrically based on the previous varying coefficient model and a longitudinal sample of $(t, Y(t), \mathbf{X})$ with time-independent covariates $\mathbf{X} = (X^{(0)}, \ldots, X^{(k)})^T$ from n independent subjects. A "leave-one-subject-out" cross-validation is suggested to choose the smoothing parameters. Asymptotic properties of our spline estimators are developed through the explicit expressions of their asymptotic normality and risk representations, which provide useful insights for inferences. Applications and finite sample properties of our procedures are demonstrated through a longitudinal sample of opioid detoxification and a simulation study.

KEY WORDS: Asymptotic normality, Clinical trials, Confidence bands, Longitudinal data, Mean squared errors, Smoothing parameters

1. INTRODUCTION

Longitudinal data, that is, samples with n independent subjects, each measured repeatedly over a time period, are common in biomedicine, epidemiology, and other fields of natural and social sciences. Examples of such samples can be easily found in clinical trials, follow-up studies for monitoring disease progression, and observational cohort studies. For a general setup of the data, we consider a real valued time variable t, and define $Y(t)$ to be the real valued response or outcome variable at time t and $\mathbf{X}(t) = (X^{(0)}(t), \ldots, X^{(k)}(t))^T$, with $k \geq 0$ and $X^{(l)}(t)$ being a real valued covariate, to be R^{k+1} valued independent covariate vector at time t. Depending on the choice of the time origin, t is not necessarily non-negative and is generally assumed to be on the real line. For a longitudinal sample of n randomly selected subjects, the jth measurement of $(t, Y(t), \mathbf{X}(t))$ for the ith subject is $(t_{ij}, Y_{ij}, \mathbf{X}_{ij})$, where $1 \leq i \leq n$, $1 \leq j \leq m_i$, m_i is the number of repeated measurements of the ith subject, t_{ij} is the measurement time, Y_{ij} is the observed response variable at t_{ij} and $\mathbf{X}_{ij} = (X_{ij}^{(0)}, \ldots, X_{ij}^{(k)})^T$ is the observed covariate vector. The total number of observations across all the subjects is $N = \sum_{i=1}^{n} m_i$.

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. Under the framework of parametric models, such as the marginal linear and nonlinear models and the mixed effects models, theory and methods for parameter estimation and inferences have been studied extensively and can be found, for example, in Laird and Ware (1982), Liang and Zeger (1986), Diggle (1988), Diggle, Liang, and Zeger (1994), Davidian and Giltinan (1995), and Vonesh and Chinchilli (1997), among others. Nonparametric smoothing methods, such as kernel estimators and smoothing splines, for estimating the mean response curve without the presence of covariates other than time have been proposed and studied by Hart and Wehrly (1986), Müller (1988), Altman (1990), Rice and Silverman (1991), among others. These authors also investigated a number of procedures for selecting the smoothing parameters that are uniquely tailored to the structures of longitudinal data.

Motivated by an epidemiological example of predicting the depletion of T-lymphocytes (CD4) cell counts among Human Immunodeficiency Virus infected persons, Zeger and Diggle (1994) and Moyeed and Diggle (1994) considered the following partially linear model:

$$Y(t) = \beta_0(t) + \sum_{l=1}^{k}\{\beta_l X^{(l)}(t)\} + \epsilon(t), \tag{1}$$

where $\beta_0(t)$ is a smooth function of t, β_l are unknown Euclidean parameters, and $\epsilon(t)$ is a mean zero stochastic process, and investigated iterative procedures for the estimation of $(\beta_0(t), \beta_1, \ldots, \beta_k)$. Recently, Cheng and Wei (2000) proposed an alternative method for estimating $(\beta_1, \ldots, \beta_k)$ of (1) without relying on iteration.

Although (1) is more flexible than the classical linear models, it requires the covariate effects β_l, $l = 1, \ldots, k$, to be constants over time, a restriction that may not be realistic for many situations. On the other hand, the actual sample sizes in most longitudinal studies may not be large enough to support a completely general nonparametric model when the dimensionality of $\mathbf{X}(t)$ is high. Thus for a more practical generalization

Chin-Tsang Chiang is Assistant Professor, Department of Statistics, Tunghai University, Taiwan (E-mail: Chiangct@mail.thu.edu.tw). John A. Rice is Professor, Department of Statistics, University of California, Berkeley, CA 94720 (E-mail: rice@stat.berkeley.edu). Colin O. Wu is Associate Professor, Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD 21218 (E-mail: colin@mts.jhu.ed). Partial support of this research was provided by the National Science Council of Taiwan, 88-2118-M-029-007, the National Institute on Drug Abuse, R01 DA10184-01, the National Science Foundation, DMS-9213353, and the Acheson J. Duncan Fund of the Johns Hopkins University. The authors thank Professor Shih-Ping Han, Dr. A. Umbricht-Schneiter, and the reviewers for insightful comments and suggestions.

Journal of the American Statistical Association
June 2001, Vol. 96, No. 454, Theory and Methods

of (1), Hoover, Rice, Wu, and Yang (1998) considered the following varying coefficient model:

$$Y(t) = \mathbf{X}^T(t)\boldsymbol{\beta}(t) + \epsilon(t), \tag{2}$$

and suggested a class of smoothing splines and local polynomials for the estimation of $\boldsymbol{\beta}(t)$, where $\boldsymbol{\beta}(t) = (\beta_0(t), \ldots, \beta_k(t))^T$ is a vector of smooth functions of t, $\epsilon(t)$ is the mean zero stochastic process as defined in (1), and for all t, $\mathbf{X}(t)$ and $\epsilon(t)$ are independent. Intuitively, (2) gives a linear model between $Y(t)$ and $\mathbf{X}(t)$ at each fixed time t, and reduces to a projection of $Y(t)$ to the linear space spanned by the components of $\mathbf{X}(t)$ when the linear relationship does not hold. Large sample properties for the estimation procedures of Hoover et al. (1998) have only been developed for the special case of kernel estimators. It has been noted by these authors that the smoothing splines proposed in Hoover et al. can be computationally very extensive.

Because the local polynomial estimates of Hoover et al. rely on only one bandwidth to smooth all $(k+1)$ coefficient curves, they may not lead to adequate smoothing for all the coefficient curves simultaneously. As a remedy, Fan and Zhang (2000) studied a two-step smoothing alternative for estimating $\boldsymbol{\beta}(t)$ of (2), which first computes raw estimates of $\boldsymbol{\beta}(t)$ at distinct time points and then smooths each component of these raw estimates. They further suggested a modification of the two-step smoothing based by binning the adjacent observations to treat the data that have sparse observations at some distinct time points. It is expected that, through appropriate binning, the two-step smoothings will have satisfactory theoretical and practical performances under the data framework of this article.

A special case of (2) that is of particular interest is longitudinal data with time-independent covariates, $\mathbf{X}(t) \equiv \mathbf{X}$. In practice, time-independent covariates such as treatment, dosage, and gender are very common in longitudinal studies. Based on (2) with $\mathbf{X}(t) \equiv \mathbf{X}$, Wu and Chiang (2000) showed that a componentwise kernel method for the estimation of $\boldsymbol{\beta}(t)$ may be more flexible in practice than the kernel method of Hoover et al. (1998).

In this article, we focus on Model (2) with time-independent covariates $\mathbf{X}$ and develop a componentwise smoothing spline procedure for the estimation and inferences of $\boldsymbol{\beta}(t)$ based on the longitudinal sample $\{(t_{ij}, Y_{ij}, \mathbf{X}_i)\}; 1 \le i \le n, 1 \le j \le m_i\}$. Comparing with the results that have already been established in the literature, our estimation and inference procedure has a number of interesting features. First, in contrast to the penalized least squares of Hoover et al. (1998), our spline estimators are based on componentwise penalized least squares. This approach allows us to significantly simplify the computation by solving $k+1$ separate linear equations, instead of a large linear system involving all $k+1$ components. Second, the asymptotic properties of our estimators are developed through their asymptotic distributions, which are useful for constructing confidence regions and other inference procedures, yet comparable asymptotic properties for the smoothing splines of Hoover et al. (1998) have not been developed. Third, similar to Rice and Silverman (1991), we select our smoothing parameters by a cross-validation that deletes the entire observations of each subject one at a time. The computation involved in our cross-validation is much simpler than that required by the cross-validation of Hoover et al. (1998). For the finite sample properties of our cross-validation and smoothing splines, we demonstrate through a Monte Carlo simulation that, at least for the special case of time-independent covariates, our procedures are either superior to or comparable with the ones currently available in the literature. Applications of our procedures are demonstrated through a biomedical example of evaluating the treatment effects in an opioid detoxification study.

In section 2, we present our componentwise smoothing splines and the cross-validation method. In section 3, we develop the asymptotic distributions and mean squared risks of our spline estimators. Application of our procedures to the opioid detoxification study is discussed in section 4. Monte Carlo simulations for our procedures and their comparisons with the existing approaches in the literature are presented in section 5. Finally, proofs of the main results are provided in the Appendix.

2. ESTIMATION PROCEDURES

2.1 Componentwise Smoothing Splines

We assume throughout that

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

with $\boldsymbol{\beta}(t)$ and $\epsilon(t)$ defined in (2), $E(\mathbf{XX}^T)$ is invertible and its inverse is denoted by $E^{-1}_{\mathbf{XX}^T}$. Multiplying both sides of (3) by $\mathbf{X}$ and taking expectations, it is easy to verify that $\boldsymbol{\beta}(t)$ is well-defined and is given by $\boldsymbol{\beta}(t) = (\beta_0(t), \ldots, \beta_k(t))^T = (E^{-1}_{\mathbf{XX}^T})E[\mathbf{X}Y(t)]$. For each $r = 0, \ldots, k$, the $(r+1)$th component $\beta_r(t)$ of $\boldsymbol{\beta}(t)$ is

$$\beta_r(t) = E\left[\left(\sum_{l=0}^{k} e_{r+1\ l+1} X^{(l)}\right) Y(t)\right], \tag{4}$$

where e_{pq} is the (p, q)th element of $E^{-1}_{\mathbf{XX}^T}$.

Based on (4), a natural approach for the estimation of $\boldsymbol{\beta}(t)$ is to construct smoothing estimators for each corresponding $\beta_r(t)$. Because $E(\mathbf{XX}^T)$ is unknown and does not depend on t, it can be estimated by the sample mean

$$\widehat{E(\mathbf{XX}^T)} = n^{-1}\sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^T).$$

We assume further that $\widehat{E(\mathbf{XX}^T)}$ is also invertible and denote its inverse by $\widehat{E}^{-1}_{\mathbf{XX}^T}$. Then, a natural estimator of e_{pq} is $\hat{e}_{pq}$, the (p, q)th element of $\widehat{E}^{-1}_{\mathbf{XX}^T}$.

Suppose that the support of the design time points is contained in a compact set $[a, b]$ and $\beta_r(t)$ are twice differentiable for all $t \in [a, b]$. We can obtain a penalized least squares estimator $\hat{\beta}_r(t; \mathbf{w})$ of $\beta_r(t)$ by minimizing

$$J_{\mathbf{w}}(\beta_r, \lambda_r) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{w_i\left[\widehat{Z}_{ijr} - \beta_r(t_{ij})\right]^2\right\} + \lambda_r\int_a^b\left[\beta_r''(s)\right]^2 ds, \tag{5}$$

where λ_r is a non-negative smoothing parameter, $\mathbf{w} = (w_1, \ldots, w_n)$ with w_i being non-negative weights and $\widehat{Z}_{ijr} = \sum_{l=0}^{k}(\hat{e}_{r+1,l+1} X_i^{(l)} Y_{ij})$. Then our estimator of $\boldsymbol{\beta}(t)$ is $\hat{\boldsymbol{\beta}}(t; \mathbf{w}) = (\hat{\beta}_0(t; \mathbf{w}), \ldots, \hat{\beta}_k(t; \mathbf{w}))^T$. Usual choices of w_i may include $w_i \equiv 1/N$ and $w_i \equiv 1/(nm_i)$.

The minimizer $\hat{\beta}_r(t; \mathbf{w})$ of (5) is a cubic spline and a linear statistic of $\widehat{Z}_{ijr}$. To see the linearity of $\hat{\beta}_r(t; \mathbf{w})$, we define $\mathcal{H}_{[a,b]}$ to be the set of compactly supported functions such that

$$\mathcal{H}_{[a,b]} = \Big\{ g(\cdot) : g \text{ and } g' \text{ are absolutely continuous on } [a,b], \text{ and } \int_a^b [g''(s)]^2 ds < \infty \Big\}.$$

Setting the Gateaux derivative of $J_{\mathbf{w}}(\beta_r; \lambda_r)$ to zero, $\hat{\beta}_r(t; \mathbf{w})$ uniquely minimizes (5) if and only if it satisfies the normal equation

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{ w_i[\widehat{Z}_{ijr} - \hat{\beta}_r(t_{ij}; \mathbf{w})] g(t_{ij}) \right\} = \lambda_r \int_a^b \hat{\beta}_r''(s; \mathbf{w}) g''(s)\, ds, \quad (6)$$

for all g in a dense subset of $\mathcal{H}_{[a,b]}$. The same argument as in Wahba (1975) then shows that there is a symmetric function $S_{\lambda_r}(t, s)$, which belongs to $\mathcal{H}_{[a,b]}$ when either t or s is fixed, so that $\hat{\beta}_r(t; \mathbf{w})$ is a cubic spline estimator and given by

$$\hat{\beta}_r(t; \mathbf{w}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[w_i S_{\lambda_r}(t, t_{ij}) \widehat{Z}_{ijr} \right]. \quad (7)$$

The explicit expression of $S_{\lambda_r}(t, s)$ is unknown. For the theoretical development of $\hat{\beta}_r(t; \mathbf{w})$, we will approximate $S_{\lambda_r}(t, s)$ by an equivalent kernel function whose explicit expression is available.

Remark 1. The penalized least squares approach of minimizing (5) is different from first estimating $E[\mathbf{X}Y(t)]$ by $E[\widetilde{\mathbf{X}Y}(t)]$, a consistent estimator obtained by smoothing each component of $E[\mathbf{X}Y(t)]$, and then estimating $\boldsymbol{\beta}(t)$ by

$$\tilde{\boldsymbol{\beta}}(t) = \widehat{E}_{\mathbf{XX}^T}^{-1} E[\widetilde{\mathbf{X}Y}(t)].$$

To see this, consider the simple case of (3) with $\mathbf{X} = (1, X)^T$, $E(X) = 1$ and $\mathrm{var}(X) = 1$. Then, we have $\beta(t) = (\beta_0(t), \beta_1(t))^T$,

$$\mathbf{XX}^T = \begin{pmatrix} 1 & X \\ X & X^2 \end{pmatrix}$$

and

$$E[\mathbf{X}Y(t)] = \begin{pmatrix} \beta_0(t) + \beta_1(t) \\ \beta_0(t) + 2\beta_1(t) \end{pmatrix}.$$

The first and second components of $E[\widetilde{\mathbf{X}Y}(t)]$ are then consistent estimators of $[\beta_0(t) + \beta_1(t)]$ and $[\beta_0(t) + 2\beta_1(t)]$, respectively. Thus in contrast to $\hat{\boldsymbol{\beta}}(t; \mathbf{w})$, $\tilde{\boldsymbol{\beta}}(t)$ is constructed by a linear combination of consistent smoothing estimators of $[\beta_0(t) + \beta_1(t)]$ and $[\beta_0(t) + 2\beta_1(t)]$ with random weights that depend on $\mathbf{X}_i$. When $\beta_0(t)$ and $\beta_1(t)$ satisfy different smoothness conditions, larger mean squared errors may result from estimating $[\beta_0(t) + \beta_1(t)]$ and $[\beta_0(t) + 2\beta_1(t)]$ than estimating $\beta_0(t)$ and $\beta_1(t)$ separately. Thus in general, $\tilde{\boldsymbol{\beta}}(t)$ is less desirable than $\hat{\boldsymbol{\beta}}(t; \mathbf{w})$. Similar phenomena will evidently hold for the general $\mathbf{X}$ with $k \geq 1$. In addition, we note that because our estimators rely on calculating the inverse of $E(\widehat{\mathbf{XX}^T})$, they may be unstable when $E(\widehat{\mathbf{XX}^T})$ is nearly singular.

Remark 2. The smoothing splines of Hoover et al. (1998) are obtained by minimizing

$$J^*(\boldsymbol{\beta}; \boldsymbol{\lambda}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{ Y_{ij} - \left[\sum_{l=0}^{k} X_i^{(l)} \beta_l(t_{ij})\right]\right\}^2 + \sum_{l=0}^{k} \lambda_l \int_a^b [\beta_l''(s)]^2 ds, \quad (8)$$

with $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_k)^T$ being non-negative smoothing parameters. From (5) and (8), we see that $J_{\mathbf{w}}(\beta_r; \lambda_r)$ and $J^*(\boldsymbol{\beta}; \boldsymbol{\lambda})$ use different square and penalty terms. Computationally, minimizing (8) requires solving a linear system that involves all the components of $\boldsymbol{\beta}(t)$ simultaneously. On the other hand, (5) is minimized with respect to $\beta_r(t)$ only. Thus particularly when k is large, the computation involved in (5) is much simpler than the computation involved in (8). Because $\hat{\beta}_r(t; \mathbf{w})$ has a simple linear expression, its asymptotic properties can be developed by methods similar to that with independent identically distributed data. Theoretical properties of the spline estimators obtained by minimizing (8) have not been developed.

Remark 3. Different choices of $\mathbf{w}$ generally lead to different finite sample and asymptotic properties for $\hat{\beta}_r(t; \mathbf{w})$. The intuitive choices of $w_i \equiv 1/N$ and $w_i \equiv 1/(nm_i)$ essentially correspond to providing equal weight to each single observation and equal weight to each subject, respectively. Ideally the optimal choice of $\mathbf{w}$ may depend on the correlation structures of the data. But, because the correlation structures are usually unknown and may be difficult to estimate, we do not have a uniformly optimal choice of $\mathbf{w}$. In practice, $w_i \equiv 1/N$ and $w_i \equiv 1/(nm_i)$ generally give satisfactory estimators. For the special case of n tends to infinity while m_i, $i = 1, \ldots, n$, remain bounded. Lin and Carroll (2000) suggested that $w_i \equiv 1/N$ leads to asymptotically optimal kernel smoothers for the generalized estimating equations.

2.2 Cross-Validation

Adequate smoothing parameters for $\hat{\beta}_r(t; \mathbf{w})$ may depend on the structures of the possible intracorrelations. When the correlation structures are completely unknown, a useful approach suggested by Rice and Silverman (1991) is to select the smoothing parameters by a "leave one subject out" cross-validation procedure. Extending their approach to the current setting, we define

$$\mathrm{CV}(\boldsymbol{\lambda}, \mathbf{w}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{ w_i\left[Y_{ij} - \sum_{l=0}^{k} X_i^{(l)} \hat{\beta}_l^{(-i)}(t_{ij}; \mathbf{w})\right]^2\right\} \quad (9)$$

to be the cross-validation score of $\hat{\beta}(t;\mathbf{w})$, where $\hat{\beta}_r^{(-i)}(t;\mathbf{w})$ is the smoothing spline estimator computed from (5) using the remaining data with all the observations of the ith subject deleted. The cross-validation smoothing parameters $\boldsymbol{\lambda}_{cv} = (\lambda_{0,cv}, \ldots, \lambda_{k,cv})^T$ are then defined to be the minimizer of $\mathrm{CV}(\boldsymbol{\lambda};\mathbf{w})$.

Theoretical properties of $\boldsymbol{\lambda}_{cv}$ have not yet been developed. For a heuristic justification, it can be shown by the same argument as in Remark 3 of Wu, Chiang, and Hoover (1998) that $\boldsymbol{\lambda}_{cv}$ approximately minimizes an average prediction error of $\hat{\beta}_r(t;\mathbf{w})$. However, a systematic optimization algorithm for $\boldsymbol{\lambda}_{cv}$ is currently still unavailable.

3. ASYMPTOTIC PROPERTIES

We present in this section the asymptotic properties of $\hat{\beta}_r(t;\mathbf{w}_0)$ with fixed time designs and the uniform weight $\mathbf{w}_0 = (1/N, \ldots, 1/N)$. Without loss of generality, we assume that $a=0$ and $b=1$. Extension to general $[a,b]$ can be obtained using the affine transformation $u=(t-a)/(b-a)$ for $t\in[a,b]$.

3.1 Assumptions

The following technical conditions, which will be used to establish the asymptotic properties of $\hat{\beta}_r(t;\mathbf{w}_0)$, are imposed mainly for mathematical simplicity and may be modified if necessary.

A1. The time design points $\{t_{ij}\}$ are nonrandom and satisfy

$$D_N = \sup_{t\in[0,1]} |F_N(t) - F(t)| \to 0, \text{ as } n\to\infty,$$

for some distribution function F with strictly positive density f on $[0,1]$, where $F_N(t) = N^{-1}\sum_{i=1}^n \sum_{j=1}^{m_i} 1_{[t_{ij}\le t]}$ and $1_{[\,]}$ is the indicator function. The density f is three times differentiable and uniformly continuous on $[0,1]$. The νth derivative $f^{(\nu)}(t)$ of $f(t)$ satisfies $f^{(\nu)}(0) = f^{(\nu)}(1) = 0$ for $\nu = 1,2$.

A2. The coefficient curves $\beta_r(t)$, $r=0,\ldots,k$, are four times differentiable and satisfy the boundary conditions $\beta_r^{(\nu)}(0) = \beta_r^{(\nu)}(1) = 0$ for $\nu=2,3$. The fourth derivatives $\beta_r^{(4)}(t)$, $r=0,\ldots,k$, are Lipschitz continuous in the sense that $|\beta_r^{(4)}(s_1) - \beta_r^{(4)}(s_2)| \le c_{1r}|s_1 - s_2|^{c_{2r}}$ for all $s_1, s_2\in[0,1]$ and some positive constants c_{1r} and c_{2r}.

A3. There exists a positive constant $\delta>0$ such that $E(|\epsilon(t)|^{2+\delta})<\infty$ and $E(X_r^{4+\delta})<\infty$ for all $r=0,\ldots,k$.

A4. The smoothing parameters λ_r, $r=0,\ldots,k$, are nonrandom and satisfy $\lambda_r\to 0$, $N\lambda_r^{1/4}\to\infty$ and $\lambda_r^{-5/4}D_N\to 0$ as $n\to\infty$.

A5. Define $\sigma^2(t) = E[\epsilon^2(t)]$ and $\rho_\epsilon(t) = \lim_{t'\to t} E[\epsilon(t)\epsilon(t')]$. Both $\sigma^2(t)$ and $\rho_\epsilon(t)$ are continuous at t.

In general, $\sigma^2(t)$ may not equal $\rho_\epsilon(t)$. Strict inequality between $\sigma^2(t)$ and $\rho_\epsilon(t)$ appears, for example, when $\epsilon(t)$ is the sum of a stationary process of t and an independent measurement error (see Zeger and Diggle 1994). Because in most applications $\sigma^2(t)$ and $\rho_\epsilon(t)$ are unknown, we do not require further specific structures for $\sigma^2(t)$ and $\rho_\epsilon(t)$, except for their continuity in A5. When $\{t_{ij}\}$ are from random designs, we would require almost sure convergence of D_N to 0, as suggested in Nychka (1995, section 2).

3.2 Equivalent Kernel Estimator

Because $S_{\lambda_r}(t,s)$ does not have an explicit expression, we would like to approximate it by an explicit equivalent kernel function. Substituting $S_{\lambda_r}(t,s)$ of (7) with the equivalent kernel, the asymptotic properties of $\hat{\beta}_r(t;\mathbf{w}_0)$ can be established through the equivalent kernel function. For the independent identically distributed data, an equivalent kernel is usually obtained by approximating the Green's function of a differential equation; see, for example, Silverman (1984), Messer (1991), Messer and Goldstein (1993), and Nychka (1995). Under the current context, we consider the following fourth order differential equation

$$\lambda_r g_r^{(4)}(t) + f(t)g_r(t) = f(t)\beta_r(t), \quad t\in[0,1], \tag{10}$$

with $g_r^{(\nu)}(0) = g_r^{(\nu)}(1) = 0$ for $\nu=2,3$. Let $G_{\lambda_r}(t,s)$ be the Green's function associated with (10). Then, any solution $g_r(t)$ of (10) satisfies $g_r(t) = \int_0^1 G_{\lambda_r}(t,s)\beta_r(s)f(s)\,ds$.

Let $\gamma = \int_0^1 (f(s))^{1/4}\,ds$ and $\Gamma(t) = \gamma^{-1}\int_0^t (f(s))^{1/4}\,ds$. We define

$$H_{\lambda_r}(t,s) = H^U_{\lambda_r/\gamma^4}(\Gamma(t),\Gamma(s))\,\Gamma^{(1)}(s)(f(s))^{-1} \tag{11}$$

to be the equivalent kernel of $S_{\lambda_r}(t,s)$, where

$$H^U_{\lambda_r}(t,s) = \frac{\lambda_r^{-1/4}}{2\sqrt{2}}\left[\sin\left(\frac{\lambda_r^{-1/4}}{\sqrt{2}}|t-s|\right) + \cos\left(\frac{\lambda_r^{-1/4}}{\sqrt{2}}|t-s|\right)\right]\exp\left(-\frac{\lambda_r^{-1/4}}{\sqrt{2}}|t-s|\right). \tag{12}$$

It is straightforward to verify that $H_{\lambda_r}(t,s)$ reduces to $H^U_{\lambda_r}(t,s)$ when $f(\cdot)$ is the uniform density. Substituting $S_{\lambda_r}(t,t_{ij})$ in (7) by $H_{\lambda_r}(t,t_{ij})$, our equivalent kernel estimator of $\beta_r(t)$ with the uniform weight $\mathbf{w}_0$ is

$$\tilde{\beta}_r(t;\mathbf{w}_0) = \frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{m_i}\left[H_{\lambda_r}(t,t_{ij})\widehat{Z}_{ijr}\right]. \tag{13}$$

The next lemma shows that $H_{\lambda_r}(t,s)$ is the dominating term of $G_{\lambda_r}(t,s)$, which in turn approximates $S_{\lambda_r}(t,s)$.

Lemma 3.1. Assume that conditions A1 and A4 are satisfied. When n is sufficiently large, there are positive constants α_1, α_2, κ_1, and κ_2 so that

$$|G_{\lambda_r}(t,s) - H_{\lambda_r}(t,s)| \le \kappa_1\exp\left(-\alpha_1\lambda_r^{-1/4}|t-s|\right), \tag{14}$$

$$\left|\frac{\partial^\nu}{\partial t^\nu}G_{\lambda_r}(t,s)\right| \le \kappa_1\lambda_r^{-(\nu+1)/4}\exp\left(-\alpha_2\lambda_r^{-1/4}|t-s|\right), \tag{15}$$

$$|S_{\lambda_r}(t,s) - G_{\lambda_r}(t,s)| \le \kappa_2\lambda_r^{-1/2}D_N \times\exp\left(-\alpha_1\lambda_r^{-1/4}|t-s|\right), \tag{16}$$

and

$$\left|\frac{\partial^\nu}{\partial t^\nu}S_{\lambda_r}(t,s)\right| \le \kappa_2\lambda_r^{-(\nu+1)/4}D_N\exp\left(-\alpha_2\lambda_r^{-1/4}|t-s|\right) \tag{17}$$

hold uniformly for $t,s\in[0,1]$ and $0\le\nu\le 3$.

Remark 4. Note that $H_{\lambda_r}(t, s)$ is not the only equivalent kernel that could be considered. Another possibility is to use the equivalent kernels suggested by Messer and Goldstein (1993). However, because (11) has a sharper exponential bound than the bounds given in Theorem 4.1 of Messer and Goldstein (1993), we use $H_{\lambda_r}(t, s)$ in this article

3.3 Asymptotic Distributions and Risks

We now summarize the main results of this section. Define $U_r = \sum_{l=0}^{k}(e_{r+1\, l+1} X^{(l)})$.

$$M_r^{(0)}(t_1, t_2) = \sum_{r_1, r_2=0}^{k} \{\beta_{r_1}(t_1)\beta_{r_2}(t_2)E[X^{(r_1)}X^{(r_2)}U_r^2]\} - \beta_r(t_1)\beta_r(t_2),$$

$$M_r^{(1)}(t) = M_r^{(0)}(t) + \sigma^2(t)e_{rr},$$

$$M_r^{(2)}(t_1, t_2) = M_r^{(0)}(t_1, t_2) + \rho_\epsilon(t_1, t_2)e_{rr},$$

$M_r^{(0)}(t) = M_r^{(0)}(t, t)$ and $M_r^{(2)}(t) = M_r^{(2)}(t, t)$.

Theorem 3.1. Suppose that A1 through A5 are satisfied, t is an interior point of $[0, 1]$, and there are constants $\lambda_{r,0} \geq 0$ and $a_0 \geq 0$ such that $\lim_{n\to\infty} N^{1/2}\lambda_r^{9/8} = \lambda_{r,0}$, $\lim_{n\to\infty}\{N^{-1}[\sum_{i=1}^{n} m_i^2]\lambda_r^{1/4}\} = a_0$ and $\lim_{n\to\infty} Nn^{-9/8} = 0$. When $n \to \infty$, $\hat{\beta}_r(t; \mathbf{w}_0)$ is asymptotically normal in the sense that

$$(N\lambda_r^{1/4})^{1/2}[\hat{\beta}_r(t; \mathbf{w}_0) - \beta_r(t)] \longrightarrow \mathrm{N}(\lambda_{r,0}b_r(t), [\tau_r(t)]^2) \quad \text{in distribution}, \tag{18}$$

where

$$b_r(t) = -[f(t)]^{-1}\beta_r^{(4)}(t) \tag{19}$$

and

$$\tau_r(t) = \left\{\frac{1}{4\sqrt{2}}[f(t)]^{-3/4}M_r^{(1)}(t) + a_0 M_r^{(2)}(t)\right\}^{1/2}. \tag{20}$$

The prior theorem implies that the asymptotic distributions of $\hat{\beta}_r(t; \mathbf{w}_0)$ are affected by n, m_i, and the intrasubject correlations of the data. These correlations affect the asymptotic variance term $[\tau_r(t)]^2$ if a_0 is strictly positive, that is, $\sum_{i=1}^{n} m_i^2$ tends to infinity sufficiently fast. In the interesting case that m_i are bounded, the probability that there are at least two data points from the same subject in a shrinking neighborhood tends to be 0, hence, the intrasubject correlation does not play a role in local smoothing.

Risks of spline estimators are usually measured by their asymptotic mean squared errors. However, because $\widehat{Z}_{ij}$ involves the inverse of $E[\widehat{\mathbf{XX}^T}]$, the first and second moments, hence the mean squared errors, of $\hat{\beta}_r(t; \mathbf{w}_0)$ may not exist. An alternative measure of risks that has been shown to be appropriate for the local polynomial estimators is the mean squared errors conditioning on the observed covariates (see Fan 1992, Fan and Gijbels 1996 and Ruppert and Wand 1994). Let $\mathcal{X}_n = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$. Conditioning on $\mathcal{X}_n$, we measure the risk of $\hat{\beta}_r(t; \mathbf{w}_0)$ by the asymptotic representation of the mean squared error

$$\mathrm{MSE}[\hat{\beta}_r(t; \mathbf{w}_0)|\mathcal{X}_n] = E\left\{[\hat{\beta}_r(t; \mathbf{w}_0) - \beta_r(t)]^2|\mathcal{X}_n\right\}. \tag{21}$$

More generally, we measure the risk of $\hat{\boldsymbol{\beta}}(t; \mathbf{w}_0)$ by the asymptotic representation of

$$\mathrm{MSE}_{\mathbf{p}}[\hat{\boldsymbol{\beta}}(t; \mathbf{w}_0)|\mathcal{X}_n] = \sum_{r=0}^{k}\left\{p_r \mathrm{MSE}[\hat{\boldsymbol{\beta}}(t; \mathbf{w}_0)|\mathcal{X}_n]\right\}, \tag{22}$$

where $\mathbf{p} = (p_0, \ldots, p_k)^T$, $p_r \geq 0$.

Theorem 3.2. Suppose that A1 through A5 are satisfied and t is an interior point of $[0, 1]$. When n is sufficiently large,

$$\begin{aligned}\mathrm{MSE}[l\hat{\beta}_r(t; \mathbf{w}_0)|\mathcal{X}_n] &= \lambda_r^2[b_r(t)]^2 + V_r(t) \\ &\quad + o_p\left(N^{-1}\lambda^{-1/4} + \sum_{i=1}^{n}\left(\frac{m_i}{N}\right)^2\right) \\ &\quad + O_p(n^{-1/2}\lambda_r) + O_p(n^{-1}) + o_p(\lambda_r^2),\end{aligned} \tag{23}$$

where $b_r(t)$ is defined in (19) and

$$V_r(t) = \frac{1}{4\sqrt{2}}N^{-1}\lambda_r^{-1/4}[f(t)]^{-3/4}M_r^{(1)}(t) + \left[\sum_{i=1}^{n}\left(\frac{m_i}{N}\right)^2\right]M_r^{(2)}(t). \tag{24}$$

Furthermore, $\lim_{n\to\infty} V_r(t) = 0$ if and only if $\lim_{n\to\infty} \max_{1\leq i\leq n}(m_i/N) = 0$.

3.3.1 Remark 5. We note that, unlike Theorem 3.1, the previous theorem does not require any further rate condition on λ_r other than (A4) and allows for any choice of nonrandom m_i. Thus under the conditions of Theorem 3.2, $\hat{\beta}_r(t; \mathbf{w}_0)$ is consistent in the sense that $\mathrm{MSE}[\hat{\beta}_r(t; \mathbf{w}_0)|\mathcal{X}_n] \to 0$ in probability as $n \to \infty$. The rate of $V_r(t)$ tending to 0 depends on n, m_i, $i = 1, \ldots, n$, λ_r and the intrasubject correlations. If $\lambda_r^{-1/4}N^{-1}$ converges to 0 in a rate slower than $\sum_{i=1}^{n}(m_i/N)^2$, then the second term of the right side of (24) becomes negligible, so that the effect of the intrasubject correlations disappears from the asymptotic representation of $\mathrm{MSE}[\hat{\beta}_r(t; \mathbf{w}_0)|\mathcal{X}_n]$. This occurs, for example, when the m_i are bounded, which is a case of practical interest. However, in general, the contributions of the intrasubject correlations are not negligible. If $m_i \to \infty$ sufficiently fast as $n \to \infty$, then the second term of the right side of (24) may dominate. This occurs, for example, when $m_i = n^{\alpha}$ for some $\alpha > 0$

3.3.2 Remark 6 The derivations of Theorem 3.2 can be extended to random designs and other weight choices. Suppose that t_{ij} are independent identically distributed with distribution function F and density f For the uniform weight $\mathbf{w}_0 = (1/N, \ldots, 1/N)$, we would require the almost sure convergence of $\lambda_r^{-5/4}D_N$ to 0 as $n \to \infty$ and consider the same equivalent kernel estimator as defined in (13). For the weight choice of $\mathbf{w}_1 = (w_1, \ldots, w_n)$ with $w_i = 1/(nm_i)$, we would replace F_N and D_N of A1 by $F_N^*(t) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}(nm_i)^{-1}1_{[t_{ij}\leq t]}$ and $D_N^* = \sup_{t\in[0,1]}|F_N^*(t) - F(t)|$, respectively, and under the almost sure convergence of $\lambda_r^{-5/4} D_N^*$ to 0 as $n \to \infty$, consider the equivalent kernel estimator

$$\tilde{\beta}_r(t; \mathbf{w}_1) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \big[(nm_i)^{-1} H_{\lambda_r}(t, t_{ij}) \widehat{Z}_{ijr}\big].$$

The asymptotic distributions and the asymptotic conditional mean squared errors of these equivalent kernel estimators can be derived explicitly. However, as noted by Nychka (1995, section 7), because the exponential bound of (16) may not be sharp enough to establish the asymptotic equivalence between the smoothing spline and the equivalent kernel estimators, improved error bounds have to be developed under these situations.

3.3.3 Remark 7. Suppose that the time design points t_{ij} are nonrandom, $\mathbf{w}_1 = (w_1, \ldots, w_n)$ with $w_i = 1/(nm_i)$ is used and A1 holds for $F_N^*(t)$ and D_N^*. By Lemma 3.1, we can show that the variance of $\hat{\beta}_r(t; \mathbf{w}_1)$ conditioning on $\mathcal{X}_n$ can be approximated by

$$\sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ \left(\frac{1}{nm_i} \right)^2 G_{\lambda_r}^2(t; t_{ij}) \operatorname{var}\left(\widehat{Z}_{ijr} \middle| \mathbf{X}_i \right) \right\}$$
$$+ \sum_{(i_1, j_1) \neq (i_2, j_2)} \left\{ \left(\frac{1}{n^2 m_{i_1} m_{i_2}} \right) G_{\lambda_r}(t, t_{i_1 j_1}) G_{\lambda_r}(t, t_{i_2 j_2}) \right.$$
$$\left. \times \operatorname{cov}\left(\widehat{Z}_{i_1 j_1 r}, \widehat{Z}_{i_2 j_2 r} \middle| \mathbf{X}_{i_1}, \mathbf{X}_{i_2} \right) \right\}.$$

Unfortunately, the two previous summations cannot be easily approximated by some straightforward integrals without further assumptions on m_i. Similarly, we do not have an explicit asymptotic risk representation for $\hat{\beta}_r(t; \mathbf{w})$ with general $\mathbf{w}$.

3.4 Inferences

The asymptotic distribution of Theorem 3.1 is potentially useful for making approximate inferences for $\hat{\beta}_r(t; \mathbf{w}_0)$. In particular, if n, m_i, $i = 1, \ldots, n$, and λ_r satisfy the conditions stated in the theorem and there are consistent estimators $(\hat{b}_r(t), \hat{\tau}_r(t))$ of $(b_r(t), \tau_r(t))$, then an approximate $100(1-\alpha)\%$ confidence interval for $\beta_r(t)$ can be given by

$$\big[\hat{\beta}_r(t; \mathbf{w}_0) - \lambda_r \hat{b}_r(t)\big] \pm u_{1-\alpha/2} N^{-1/2} \lambda_r^{-1/8} \hat{\tau}_r(t), \qquad (25)$$

where $0 < \alpha < 1$, u_p is the pth quantile of the standard normal distribution. In principle, it is possible to construct the consistent estimators $\hat{b}_r(t)$ and $\hat{\tau}_r(t)$ by substituting the unknown quantities of (19) and (20) with their consistent estimators. But, in practice, $b_r(t)$ is difficult to estimate because it depends on the fourth derivative of $\beta_r(t)$. One possible approach to circumvent the difficulty of estimating $b_r(t)$ is to select a small smoothing parameter λ_r so that the asymptotic bias of (18) is negligible. For the estimation of $\tau_r(t)$, one has to construct adequate smoothing estimators for the variance and covariance processes $\sigma^2(t)$ and $\rho_\epsilon(t)$. Wu, Chiang, and Hoover (1998) investigated some plug-in type asymptotic confidence procedures based on the kernel estimator of Hoover et al. (1998). However, a practical procedure based on (25) requires substantial further development.

Another approach that may be useful for the inferences of $\boldsymbol{\beta}(t)$ is the bootstrap. Hoover et al. (1998) suggested a bootstrap procedure that resamples subjects from the original data with replacement. Here, a bootstrap spline estimator $\beta_r^{\mathrm{boot}}(t; \mathbf{w}_0)$ of $\beta_r(t)$ can be computed based on (7) and the bootstrap sample. Thus to estimate the variance of $\hat{\beta}_r(t; \mathbf{w}_0)$, we first obtain B independent bootstrap samples and then use the sample variance of the B bootstrap spline estimators $\hat{\beta}_r^{\mathrm{boot}}(t; \mathbf{w}_0)$. Let $\widehat{V}_r^{\mathrm{boot}}(t)$ be the bootstrap estimator of the variance of $\hat{\beta}_r(t; \mathbf{w}_0)$. A bootstrap approximate $100(1-\alpha)\%$ interval for $\beta_r(t)$ can be constructed by

$$\hat{\beta}_r(t; \mathbf{w}_0) \pm u_{1-\alpha/2} \sqrt{\widehat{V}_r^{\mathrm{boot}}(t)}. \qquad (26)$$

We note that (26) is an approximate $100(1-\alpha)\%$ confidence interval for $\beta_r(t)$ only if the bias term of (18) is negligible, which occurs, for example, when $\beta_r(t)$ is time-invariant. However, in general, because no bias-adjustment has been provided, (26) may not be a rigorous $100(1-\alpha)\%$ confidence interval for $\beta_r(t)$. Comparing with the plug-in type asymptotic interval, (26) avoids the complications of estimating the components of $\tau_r(t)$. We present in section 5 some empirical coverage probabilities of (26) obtained through a simulation. However, the theoretical properties of this bootstrap procedure have not been investigated.

An important issue of model diagnostic based on (3) is to determine whether some or all of the coefficient curves $\beta_r(t)$, $r = 0, \ldots, k$, can be approximated by some parametric families. Further theoretical and practical investigations on the following two inferential extensions are warranted. The first is to examine simultaneous confidence bands for $\beta_r(t)$. The second is to test the hypotheses that, for some or all $r = 0, \ldots, k$, $\beta_r(t)$ belongs to a parametric family. A case that is of particular interest in practice is to test the hypothesis that the covariate effects are time-invariant, so that the model reduces to (1).

4. APPLICATION TO OPIOID DETOXIFICATION STUDY

The main objective of this randomized clinical trial is to evaluate the effect of a combination treatment of naltrexone with buprenorphine for opioid detoxification. As one of the two pharmacologic treatments (methadone maintenance and opioid detoxification) for opioid dependence, opioid detoxification has the advantage of being relatively inexpensive and accessible to vast majority of opioid-dependent individuals (e.g., Gold, Redmond, and Kleber 1978). Among 60 opioid-dependent individuals who were not using methadone, 32 were randomly assigned to the NB (naltrexone–buprenorphine) treatment group and 28 were assigned to the PB (placebo–buprenorphine) group. During the 8-day impatient clinical period, each patient received OOW (observer-rated opioid withdrawal scale) measurements and NB/PB interventions at 9 scheduled trial times per day. Because some individuals randomly missed their scheduled measurements or quit the treatment process, the intervention times were different per individual. Further details of the design, medical implications, and an initial analysis of the study can be found in Umbricht-Schneiter, Montoya, Hoover, Demuch, Chiang, and Preston (1999).

Because the usual chronological time does not provide a meaningful scale to measure the treatment effects, we follow Umbricht-Schneiter et al. (1999) and define the trial time to be the trial number of an intervention. Specifically, with m_i repeated measurements, t_{ij} have integer values between 1 and 72.

For the purpose of demonstration, we evaluate here the effects of the treatment and the baseline OOW scores, defined as the OOW scores at the entry of the study, on the OOW scores over trial time. Here, Y_{ij} is the ith patient's OOW score at the jth trial, $X_i^{(0)} \equiv 1$,

$$X_i^{(1)} = \begin{cases} 1 & \text{if the individual belongs to the NB group,} \\ 0 & \text{if he/she belongs to the PB group,} \end{cases}$$

and $X_i^{(2)}$ represents the individual's centered baseline OOW score, computed by subtracting the sample average from the individual's actual baseline OOW score.

Although parametric analyses or the descriptive statistics as in Umbricht-Schneiter et al. (1999) can be used, it is of interest to explore whether these approaches reasonably fit the data, as misspecified models may lead to erroneous conclusions. Under (3), $\beta_0(t)$ represents the mean OOW curve for individuals in the PB group with average baseline OOW score; $\beta_1(t)$ represents the treatment difference at t; and $\beta_2(t)$ represent the change of mean OOW score at t associated with the unit change of baseline OOW score.

We computed the smoothing spline estimators $\hat{\beta}_l(t; \mathbf{w})$, $l = 0, 1, 2$, for both $w_i \equiv 1/N$ and $w_i \equiv 1/(nm_i)$ weights based on (7) with $\boldsymbol{\lambda}_{cv}$ and a subjective smoothing vector $\boldsymbol{\lambda} = (.001, .01, .1)^T$. As precise global minimum of the cross-validation scores may be hard to find in some cases, the subjective smoothing parameter $\boldsymbol{\lambda}$ was considered because its corresponding cross-validation score was very close to that for $\boldsymbol{\lambda}_{cv}$. Both $w_i \equiv 1/N$ and $w_i \equiv 1/(nm_i)$ weights give similar results for this dataset, so we only present the results for the $w_i \equiv 1/N$ weight. Based on the "resampling-subject" bootstrap described in the previous section, we also computed the 95% bootstrap intervals (26) for $\beta_0(t)$, $\beta_1(t)$, and $\beta_2(t)$. As discussed in section 3.4, because no bias adjustment has been considered, these intervals may not accurately approximate the 95% confidence intervals. Figure 1 shows the estimates and their corresponding 95% bootstrap intervals computed based on the cross-validated smoothing parameters $\boldsymbol{\lambda}_{cv}$. The estimated mean curve for the PB group appears to be slightly undersmoothed. But the estimates for the NB treatment effect and the baseline OOW effect seem to have adequate smoothing. Figure 2 shows the estimates and their 95% bootstrap intervals computed based on the subjective smoothing vector $\boldsymbol{\lambda}$. Comparing these figures, we see that the estimates in (1b), (1c), (2b), and (2c) are virtually identical, yet the estimated curve in (2a) seems to be more appropriate than the one in (1a).

Qualitatively, we can see from these figures that the mean OOW curve $\beta_0(t)$ for the PB group generally stays around 1.0 throughout the trial. The estimated treatment effect $\beta_1(t)$ appears to imply that the NB treatment is positively associated with higher OOW scores at the beginning of the trial, but it generally leads to lower OOW scores for the later half of the trial. The baseline OOW scores appear to be positively associated with higher OOW scores throughout the trial, but this positive association has a decreasing trend. These assessments are generally consistent with that of Umbricht-Schneiter et al. (1999), who presented the mean OOW scores over time by their daily averages. The reader is referred to their article for a detailed discussion of the biomedical implications of the time-varying treatment effect and the potential advantages of the naltrexone–buprenorphine treatment. Although we have not developed inference tools as formally as would ultimately be desirable, we note that, if the treatment effects were in fact constant, the pointwise intervals would not be affected by bias, according to our asymptotic results. The envelopes traced by the intervals in the figures, or by wider intervals produced by a Bonferroni adjustment, are not inconsistent with constant treatment effects. Thus there is not strong evidence for nonconstant coefficients, although the steady declining trend in Figure (1c) or Figure (2c) would seem plausible.

5. SIMULATION

To evaluate the practical performance of our procedures, we consider the Model (3) with

$$\beta_0(t) = 3.5 + 6.5 \sin\left(\frac{t\pi}{60}\right),$$

$$\beta_1(t) = -.2 - 1.6 \cos\left(\frac{(t-30)\pi}{60}\right),$$

$$\beta_2(t) = .25 - 0.0074\left(\frac{30-t}{10}\right)^3$$

and $\mathbf{X} = (1, X^{(1)}, X^{(2)})^T$, where $X^{(1)}$ and $X^{(2)}$ are independent random variables with Bernoulli(.5) and N(0, 4) distributions, respectively. The simulation was independently repeated 500 times. Within each repetition, we randomly generated 400 subjects and their corresponding covariate vectors $\mathbf{X}_i$, $i = 1, \ldots, 400$. Each subject was assigned a probability of 40% to be observed at the integer design points $\{0, 1, \ldots, 30\}$, so that the average number of repeated measurements per subject was approximately 12. Because the time design points were different from subject to subject, we obtained uneven t_{ij} with $i = 1, \ldots, 400$ and $j = 1, \ldots, m_i$. The errors ϵ_{ij} were generated from the mean zero Gaussian process with covariance function

$$\text{cov}\,(\epsilon_{i_1 j_1}, \epsilon_{i_2 j_2}) = \begin{cases} .0625 \exp(-|t_{i_1 j_1} - t_{i_2 j_2}|), & \text{if } i_1 = i_2, \\ 0, & \text{if } i_1 \neq i_2. \end{cases}$$

The time-dependent responses Y_{ij} were obtained by substituting the available $\mathbf{X}_i$, t_{ij}, and ϵ_{ij} into (3).

For each generated sample, we computed the componentwise spline estimators $\hat{\beta}_l(t; \mathbf{w}_0)$ of $\beta_l(t)$, $l = 0, 1, 2$, using the cross-validated smoothing parameters $\boldsymbol{\lambda}_{cv}$ and the subjec-

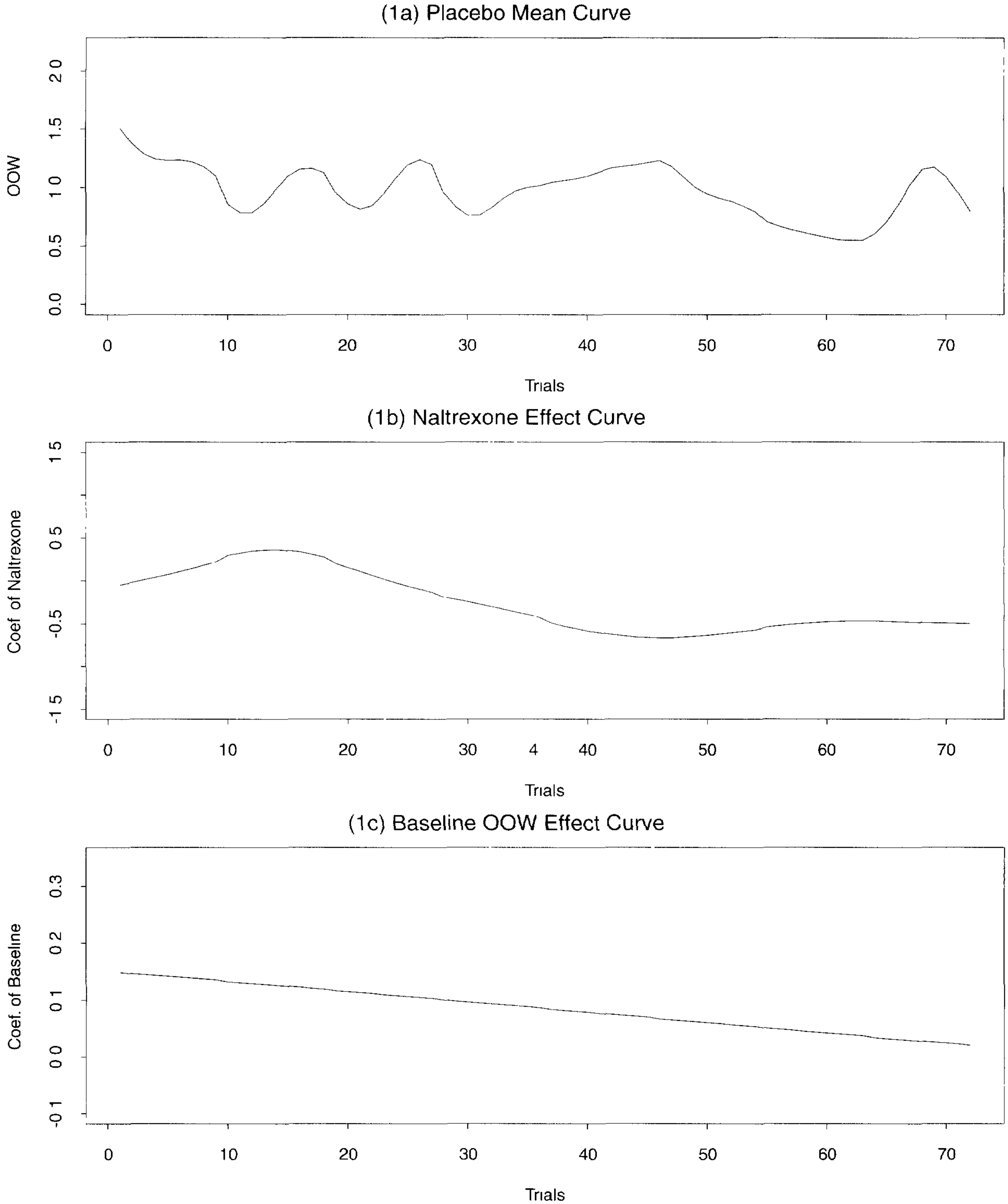

Figure 1. The Solid Curves Represent the Estimated Effects and the Dotted Curves Show the Corresponding 95% Bootstrap Intervals When the Cross Validated Smoothing Parameters $\boldsymbol{\lambda}_{cv} = (.0001, .01, 10)^T$ and $\mathbf{w}_0 = (1/N, \ldots, 1/N)$ Are Used. (a) The estimated PB mean OOW curve $\hat{\beta}_0(t; \mathbf{w}_0)$. (b) The estimated NB treatment effect $\hat{\beta}_1(t, \mathbf{w}_0)$ (c) The estimated baseline OOW effect $\hat{\beta}_2(t; \mathbf{w}_0)$.

tive smoothing parameters $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2)^T$ with $\lambda_l = 1, 10$, and 100 for $l = 0, 1, 2$. As a comparison, we also estimated $\beta_l(t)$, $l = 0, 1, 2$, using the two-step smoothing method of Fan and Zhang (2000) with both local linear and smoothing spline smoothers. The cross-validation of Rice and Silverman (1991) was used to select the smoothing parameters for the two-step smoothing spline estimators. For the two-step local linear estimators of $\beta_l(t)$, we used the Epanechnikov kernel and the bandwidths suggested in Fan and Zhang (2000). However, because similar results were obtained by the local linear and the spline smoothers, we only present here the two-step smoothing spline estimators.

As an intuitive measure of the bias and variability of the estimators of $\beta_l(t)$, we computed the averages and the

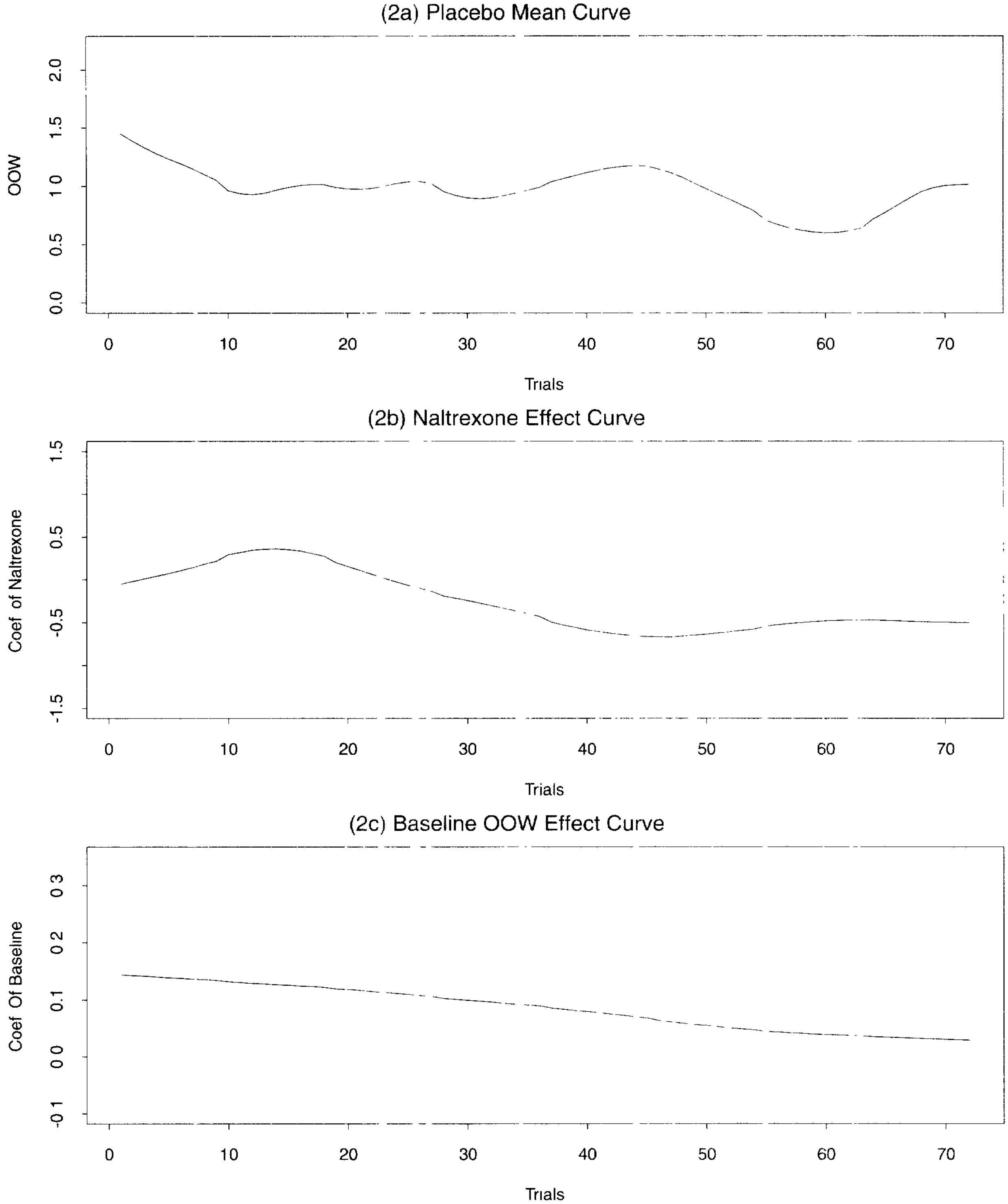

Figure 2. The Solid Curves Represent the Estimated Effects and the Dotted Curves Show the Corresponding 95% Bootstrap Intervals When $\boldsymbol{\lambda} = (.001, .01, .1)^T$ and $\hat{\beta}_0(t;\mathbf{w}_0)$ Are Used. (a) The estimated PB mean OOW curve $\hat{\beta}_0(t;\mathbf{w}_0)$. (b) The estimated NB treatment effect $\hat{\beta}_1(t;\mathbf{w}_0)$. (c) The estimated baseline OOW effect $\hat{\beta}_2(t;\mathbf{w}_0)$.

standard errors, over the 500 simulated samples, of the componentwise and two-step smoothing spline estimators. Figure 3 shows the true $\beta_l(t)$, $l = 0, 1, 2$, curves, the sample averages of $\hat{\beta}_l(t; \mathbf{w}_0)$ using the cross-validated smoothing parameters, the sample averages of the two-step smoothing spline estimators of $\beta_l(t)$, and the corresponding ± 2 sample standard error bars of these estimators at several time points. Possibly due to the fact that the cross-validation scores have little variation for a wide range of λ_2 values, we see from this figure that the sample average of $\hat{\beta}_2(t; \mathbf{w}_0)$ with the cross-validated smoothing parameters is slightly oversmoothed. The averages for the other estimators are reasonably close to their corresponding true curves. Thus in terms of bias, it seems that both the componentwise and the two-step smoothing methods are comparable for $l = 0, 1$, and the componentwise smoothing has a slightly larger bias than its counterpart produced by the

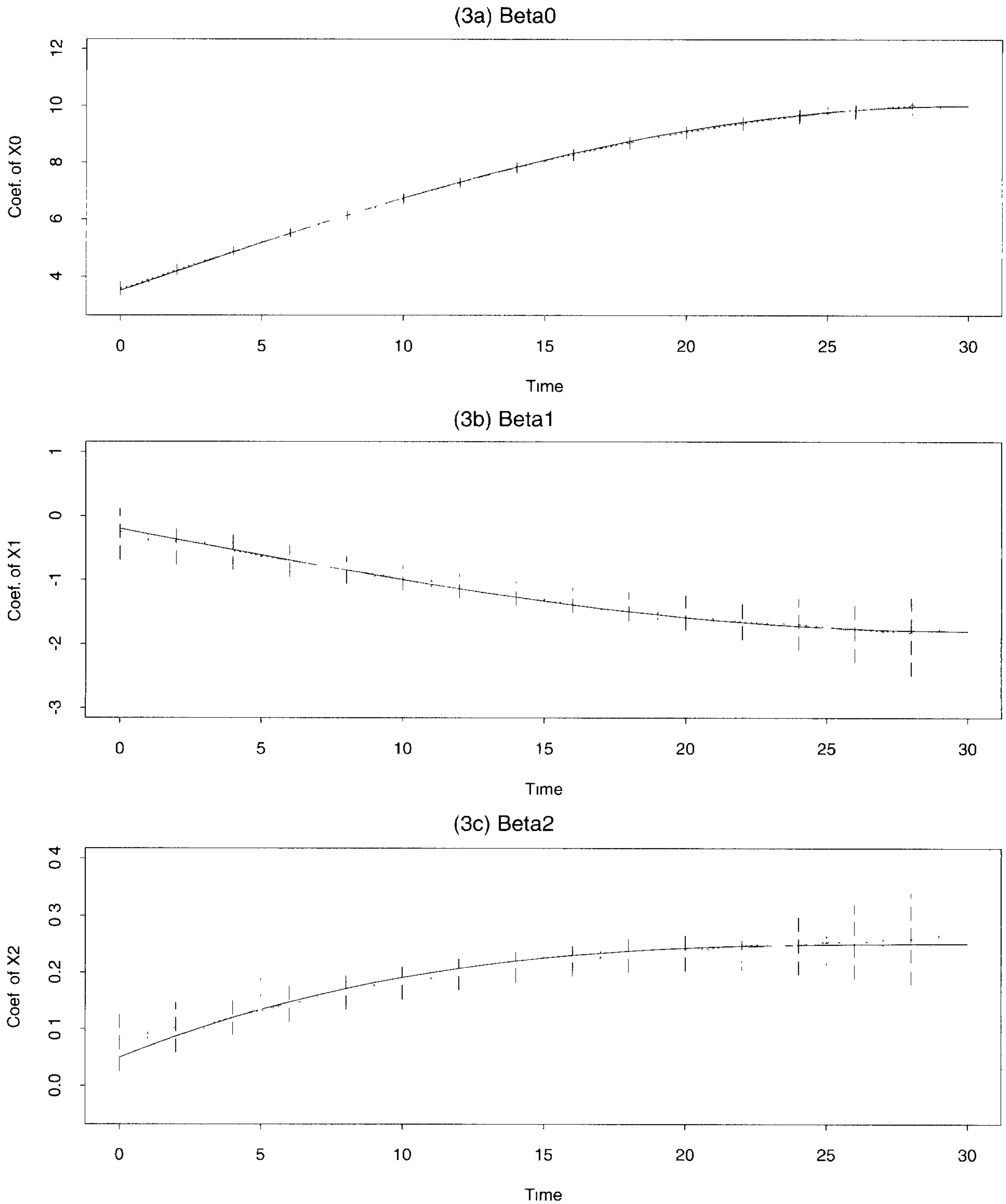

Figure 3. The Real $\beta_l(t)$ Curves, the 500 Simulation Averages of $\hat{\beta}_l(t, \mathbf{w}_0)$ (dashed curves) with $\lambda_{l,cv}$ and the Two-Step Smoothing Spline Estimators (dotted curves), and the Corresponding ±2 Sample Standard Error Bars of the Estimators. (a) $l = 0$. (b) $l = 1$. (c) $l = 2$.

two-step method. Yet, the sample variances of the two-step estimators are generally larger than those of the componentwise estimators.

To assess the effects of the smoothing parameters and the practical performance of the bootstrap intervals (26), we computed the 95% bootstrap intervals for $\beta_l(t)$ based on $\hat{\beta}_l(t; \mathbf{w}_0)$ with the cross-validated smoothing parameters $\lambda_{l,cv}$ and the subjective smoothing parameters $\lambda_l = 1, 10, 100$ for $l = 0, 1, 2$. Here the bootstrap procedure was carried out with 200 bootstrap repetitions. Tables 1–3 show the empirical coverage probabilities, computed over the 500 simulated samples, of the 95% bootstrap intervals for $\beta_l(t)$. The entries of these tables show that, although no bias adjustment was used, the coverage probabilities for $\hat{\beta}_l(t; \mathbf{w}_0)$ with $\lambda_{l,cv}$ and the subjective smoothing parameters $\lambda_l = 1$ and 10 are generally close to their nominal level, except for a few exceptions. These cov-

Table 1. Empirical Coverage Probabilities of 95% Bootstrap Intervals for $\beta_0(t)$ at Nine Time Points Based on $\hat{\beta}_0(t;\mathbf{w}_0)$ With the Cross-Validated Smoothing Parameters and $\lambda_0 = 1.0, 10.0, 100.0$

Time point	3.0	6.0	9.0	12.0	15.0	18.0	21.0	24 0	27.0
$\lambda_{0,cv}$	90	.93	.95	94	93	.94	.90	92	.95
$\lambda_0 = 1.0$	.92	93	.95	93	.92	.93	94	93	.93
$\lambda_0 = 10\ 0$	.89	.93	.93	90	.88	.88	.86	90	.93
$\lambda_0 = 100.0$	.18	.86	44	01	00	00	17	89	.61

Table 3. Empirical Coverage Probabilities of 95% Bootstrap Intervals for $\beta_2(t)$ at Nine Time Points Based on $\hat{\beta}_2(t;\mathbf{w}_0)$ With the Cross-Validated Smoothing Parameters and $\lambda_2 = 1.0, 10.0, 100.0$

Time Point	3.0	6 0	9 0	12.0	15.0	18.0	21.0	24 0	27 0
$\lambda_{2,cv}$	.90	.92	87	82	82	91	95	.93	.94
$\lambda_2 = 1.0$	.94	.96	.94	.92	.94	94	92	.94	.92
$\lambda_2 = 10.0$	94	93	.92	.93	.94	.92	93	93	92
$\lambda_2 = 100.0$	.92	92	79	.66	.69	.84	.93	.93	92

erage probabilities may be further improved if proper bias adjustment could be used. However, the coverage probabilities for $\hat{\beta}_0(t;\mathbf{w}_0)$ with $\lambda_l = 100$ are frequently significantly lower than their nominal level, indicating an inappropriate choice for the smoothing parameter.

APPENDIX A: GREEN'S FUNCTION FOR UNIFORM DENSITY

This special case serves an important linkage between $G_{\lambda_r}(t,s)$ and $H_{\lambda_r}(t,s)$. Its dominating term will be used to establish Lemma 3.1. Direct derivation shows that, for the uniform density $f(t) = 1_{[0,1]}(t)$, the Green's function $G^U_{\lambda_r}(t,s)$ of (10) is the solution of

$$\lambda_r \frac{\partial^4}{\partial t^4} G^U_{\lambda_r}(t,s) + G^U_{\lambda_r}(t,s) = 0, \quad \text{for} \quad t \neq s \tag{A.1}$$

subject to the following conditions:

(a) $G^U_{\lambda_r}(t,s) = G^U_{\lambda_r}(s,t) = G^U_{\lambda_r}(1-t, 1-s)$;

(b) $(\partial^\nu/\partial t^\nu) G^U_{\lambda_r}(0,t) = (\partial^\nu/\partial t^\nu) G^U_{\lambda_r}(1,t) = 0$ for $\nu = 2, 3$;

(c) $(\partial^\nu/\partial t^\nu) G^U_{\lambda_r}(t,s)|_{s=t^-} - (\partial^\nu/\partial t^\nu) G^U_{\lambda_r}(t,s)|_{s=t^+} = 0$ for $\nu = 0, 1, 2$;

(d) $(\partial^3/\partial t^3) G^U_{\lambda_r}(t,s)|_{s=t^-} - (\partial^3/\partial t^3) G^U_{\lambda_r}(t,s)|_{s=t^+} = \lambda_r^{-1}$

Lemma A. Suppose that $G^U_{\lambda_r}(t,s)$ is the Green's function of the differential equation (10) with $f(t) = 1_{[0,1]}(t)$. When $\lambda_r \to 0$, the solution $G^U_{\lambda_r}(t,s)$ of (A.1) is given by

$$G^U_{\lambda_r}(t,s) = H^U_{\lambda_r}(t,s)\{1 + O[\exp(-\lambda_r^{-1/4}/\sqrt{2})]\}, \tag{A.2}$$

where $H^U_{\lambda_r}(t,s)$ is defined in (12).

Proof of Lemma A. Because the proof involves tedious algebra, we only sketch the main steps. By the well-known result in differential equations, for example, Brauer and Nohel (1973), a general solution $G^U_{\lambda_r}(t,s)$ of (A.1) can be expressed as

$$\begin{aligned} G^U_{\lambda_r}(t,s) = \sum_{l=1,3,5,7} \{ & [C_{jl}\sin(\lambda_r^{-1/4}\xi_l(t,s)/\sqrt{2}) \\ & + C_{j,l+1}\cos(\lambda_r^{-1/4}\xi_{l+1}(t,s)/\sqrt{2})] \\ & \times \exp(\lambda_r^{-1/4}\zeta_l(t,s)/\sqrt{2})\}, \end{aligned}$$

where $j = 1$ or 2, when $t \le s$ or $t > s$, $\xi_1(t,s) = \zeta_1(t,s) = t - s$, $\xi_3(t,s) = \zeta_3(t,s) = t + s$, $\xi_5(t,s) = -\zeta_5(t,s) = t - s$ and $\xi_7(t,s) = -\zeta_7(t,s) = t + s$.

By $G^U_{\lambda_r}(t,s) = G^U_{\lambda_r}(s,t)$ in condition (a) of (A.1), we can obtain that $C_{11} = -C_{25}$, $C_{12} = C_{26}$, $C_{13} = C_{23}$, $C_{14} = C_{24}$, $C_{15} = -C_{21}$, $C_{16} = C_{22}$, $C_{17} = C_{27}$ and $C_{18} = C_{28}$. Furthermore, $G^U_{\lambda_r}(t,s) = G^U_{\lambda_r}(1-t, 1-s)$ implies that

$$\begin{aligned} C_{13} = [&-\cos(\sqrt{2}\lambda_r^{-1/4})C_{17} + \sin(\sqrt{2}\lambda_r^{-1/4})C_{18}] \\ & \times \exp(-\sqrt{2}\lambda_r^{-1/4}) \end{aligned} \tag{A.3}$$

and

$$\begin{aligned} C_{14} = [&\sin(\sqrt{2}\lambda_r^{-1/4})C_{17} + \cos(\sqrt{2}\lambda_r^{-1/4})C_{18}] \\ & \times \exp(-\sqrt{2}\lambda_r^{-1/4}). \end{aligned} \tag{A.4}$$

Let $\lambda^*_{r1} = 2^{-1/2}\lambda_r^{-1/4}$, $\lambda^*_{r2} = 2^{-1/2}\lambda_r^{-1/4} - (\pi/4)$ and $\lambda^*_{r3} = 2^{-1/2} \times \lambda_r^{-1/4} + (\pi/4)$. Taking derivatives of $G^U_{\lambda_r}(t,s)$ with respect to t, we can derive from condition (b) of (A.1) that

$$C_{11} + C_{13} - C_{15} - C_{17} = \sum_{l=1}^{4}(-1)^{l-1}C_{1l} + \sum_{l=5}^{8} C_{1l} = 0, \tag{A.5}$$

$$\begin{aligned} & [\cos(\lambda^*_{r1})(C_{11} - C_{17}) + \sin(\lambda^*_{r1})(C_{12} + C_{18})] \\ & \times \exp(-2\lambda^*_{r1}) + \cos(\lambda^*_{r1})(C_{13} - C_{15}) \\ & - \sin(\lambda^*_{r1})(C_{14} + C_{16}) = 0 \end{aligned} \tag{A.6}$$

and

$$\begin{aligned} & [\sin(\lambda^*_{r3})(C_{11} - C_{17}) + \sin(\lambda^*_{r2})(C_{12} + C_{18})] \\ & \times \exp(-2\lambda^*_{r1}) + \sin(\lambda^*_{r2})(C_{13} - C_{15}) \\ & - \sin(\lambda^*_{r3})(C_{14} + C_{16}) = 0. \end{aligned} \tag{A.7}$$

From conditions (c) and (d) of (A.1), we get

$$C_{11} + C_{12} + C_{15} - C_{16} = 0 \tag{A.8}$$

and

$$C_{11} - C_{12} + C_{15} + C_{16} = -\lambda^*_{r1}. \tag{A.9}$$

Table 2. Empirical Coverage Probabilities of 95% Bootstrap Intervals for $\beta_1(t)$ at Nine Time Points Based on $\hat{\beta}_1(t;\mathbf{w}_0)$ With the Cross-Validated Smoothing Parameters and $\lambda_1 = 1.0, 10.0, 100.0$

Time point	3 0	6.0	9.0	12.0	15.0	18 0	21 0	24 0	27.0
$\lambda_{1,cv}$	.92	.94	93	.86	.87	.90	92	.92	.95
$\lambda_1 = 1.0$	.92	.93	.93	.93	.92	93	93	.93	.93
$\lambda_1 = 10.0$	.93	.94	.93	92	.92	.93	93	.93	.95
$\lambda_1 = 100.0$	.92	.94	.93	.85	83	.85	.90	.93	94

By (A.6) and (A.9), we can express C_{15} through C_{18} as linear combinations of C_{1j}, $j = 1, \dots, 4$, and get $C_{15} = -C_{11} - 2^{-3/2}\lambda_i^{-1/4}$, $C_{16} = C_{12} - 2^{-3/2}\lambda_i^{-1/4}$, $C_{17} = 2C_{11} + C_{13} + 2^{-3/2}\lambda_i^{-1/4}$ and $C_{18} = -2(C_{11} + C_{13}) + C_{14} + 2^{-3/2}\lambda_i^{-1/4}$. Substituting C_{15} through C_{18} with their corresponding linear combinations of C_{1j}, $j = 1, \dots, 4$, we can derive from (A.3), (A.4), (A.6), and (A.7) that

$$\begin{aligned} 2\exp(-2\lambda^*_{i1})C_{11} &+ [\exp(-2\lambda^*_{i1}) + \cos(2\lambda^*_{i1})]C_{13} \\ &- \sin(2\lambda^*_{i1})C_{14} + 2^{-1}\lambda^*_{i1}\exp(-2\lambda^*_{i1}) = 0, \end{aligned} \quad (A.10)$$

$$\begin{aligned} &2\exp(-2\lambda^*_{i1})C_{11} + [2\exp(-2\lambda^*_{i1}) + \sin(2\lambda^*_{i1})]C_{13} \\ &\quad -[\exp(-2\lambda^*_{i1}) - \cos(2\lambda^*_{i1})] \\ &\quad \times C_{14} - 2^{-1}\lambda^*_{i1}\exp(-2\lambda^*_{i1}) = 0. \end{aligned} \quad (A.11)$$

$$\begin{aligned} &\{\cos(\lambda^*_{i1}) - [\cos(\lambda^*_{i1}) + 2\sin(\lambda^*_{i1})]\exp(-2\lambda^*_{i1})\} \\ &\quad \times (C_{11} + C_{13}) - \sin(\lambda^*_{i1})[1 - \exp(-2\lambda^*_{i1})] \\ &\quad \times (C_{12} + C_{14}) + 2^{-1}\lambda^*_{i1}\cos(\lambda^*_{i1})[1 - \exp(-2\lambda^*_{i1})] \\ &\quad -2^{-1}\lambda^*_{i1}\sin(\lambda^*_{i1})[1 + \exp(-2\lambda^*_{i1})] = 0 \end{aligned} \quad (A.12)$$

and

$$\begin{aligned} &\{-\sin(\lambda^*_{i2}) + [\sin(\lambda^*_{i3}) + 2\sin(\lambda^*_{i2})]\exp(-2\lambda^*_{i1})\} \\ &\quad \times (C_{11} + C_{13}) - [\sin(\lambda^*_{i3}) + \sin(\lambda^*_{i2})\exp(-2\lambda^*_{i1})] \\ &\quad \times (C_{12} + C_{14}) - 2^{-1}\lambda^*_{i1}[\sin(\lambda^*_{i2}) - \sin(\lambda^*_{i3})] \\ &\quad \times [1 + \exp(-2\lambda^*_{i1})] = 0. \end{aligned} \quad (A.13)$$

Suppose first that $\lambda_i \neq 2^{-2}[(k + 2^{-1})\pi]^{-4}$ and $\lambda_i \neq 2^{-2}(k\pi)^{-4}$ for any positive integer k. When $\lambda_i \to 0$, it can be derived from (A.10) to (A.13) that

$$C_{1l} = (-1)^l\left(\lambda_i^{-1/4}/(2\sqrt{2})\right)\left\{1 + O\left(\exp(-\lambda_i^{-1/4}/\sqrt{2})\right)\right\}, \quad l = 1, 2. \quad (A.14)$$

and

$$C_{1l} = O\left(\lambda_i^{-1/4}\exp(-\lambda_i^{-1/4}/\sqrt{2})\right), \quad l = 3, 4. \quad (A.15)$$

Finally, C_{15} through C_{18} can be directly calculated by using (A.14) and (A.15), so that

$$C_{1l} = O\left(\lambda_i^{-1/4}\exp(-\lambda_i^{-1/4}/\sqrt{2})\right), \quad l = 5, 6, \quad (A.16)$$

$$C_{17} = -\left(\lambda_i^{-1/4}/(2\sqrt{2})\right)\left[1 + O\left(\exp(-\lambda_i^{-1/4}/\sqrt{2})\right)\right] \quad (A.17)$$

and

$$C_{18} = \left(3\lambda_i^{-1/4}/(2\sqrt{2})\right)\left[1 + O\left(\exp(-\lambda_i^{-1/4}/\sqrt{2})\right)\right] \quad (A.18)$$

Then (A.2) is obtained by substituting (A.14) through (A.18) into the general expression of $G^U_{\lambda_i}(t, s)$. When $\lambda_i = 2^{-2}[(k + 2^{-1})\pi]^{-4}$ or $2^{-2}(k\pi)^{-4}$, the same argument as before shows that the coefficients in (A.14) through (A.18) also hold. This completes the proof.

APPENDIX B: PROOF OF LEMMA 3.1

Before establishing the relationship between $G^U_{\lambda_i}(t, s)$ and $G_{\lambda_i}(t, s)$, we first consider a transformation $Q_i(t, s)$ such that

$$Q_i(\Gamma(t), \Gamma(s))\Gamma^{(1)}(s) = G_{\lambda_i}(t, s)f(s) \quad (B.1)$$

Let

$$\begin{aligned} q_i(u) &= \int_0^1 Q_i(u, v)\beta_i(\Gamma^{-1}(v))\,dv, \\ \phi_1(t) &= \{6[\Gamma^{(1)}(t)]^2\Gamma^{(2)}(t)\}[f(t)]^{-1}, \\ \phi_2(t) &= \{3[\Gamma^{(2)}(t)]^2 + 4\Gamma^{(1)}(t)\Gamma^{(3)}(t)\}[f(t)]^{-1} \end{aligned}$$

and $\phi_3(t) = \Gamma^{(4)}(t)[f(t)]^{-1}$. It is straightforward to verify that $g_i(t) = q_i(\Gamma(t))$ and $q_i(u)$ is the solution of the following fourth order differential equation

$$\left[\left(\frac{\lambda_i}{\gamma^4}\right)q_i^{(4)}(u) + q_i(u)\right] + \lambda_i\sum_{l=1}^{3}\phi_l(u)q_i^{(4-l)}(u) = \beta_i(\Gamma^{-1}(u)), \quad (B.2)$$

subject to the boundary conditions $q_i^{(\nu)}(0) = q_i^{(\nu)}(1) = 0$ for $\nu = 2, 3$.

Let $\mathcal{D}$ and $\mathcal{I}$ be the operators for differentiation and identity, and $\mathcal{M}_\phi$ be the multiplication operator $\mathcal{M}_\phi g = \phi \cdot g$. Then (B.2) can be expressed as

$$(\mathcal{I} + \mathcal{A}_i)\mathcal{L}_i q_i(u) = \beta_i(\Gamma^{-1}(u)), \quad (B.3)$$

where $\mathcal{L}_i = [(\lambda_i/\gamma^4)\mathcal{D}^4 + \mathcal{I}]$ and $\mathcal{A}_i = \lambda_i(\sum_{l=1}^{3}\mathcal{M}_{\phi_l}\mathcal{D}^{4-l})\mathcal{L}_i^{-1}$. Let $A_i^\nu(u, v)$ be the kernel associated with the integral operator $\mathcal{A}_i^\nu$. We can verify by the induction argument in the proof of (A.1) of Nychka (1995) that, when n is large, there are constants $\alpha_0 > 0$ and $\kappa_0 > 0$ so that

$$\left|A_i^\nu(u, v)\right| \le \kappa_0 W^\nu \exp\left(-\alpha_0\lambda_i^{-1/4}|u - v|\right), \qquad \nu \ge 1$$

where W is some positive constant $W < 1$. Because $|A_i^\nu(u, v)| < 1$ for sufficiently small λ_i, the integral operator $\mathcal{L}_i^{-1}(\mathcal{I} + \mathcal{A}_i)^{-1}$ has the expansion

$$\mathcal{L}_i^{-1}(\mathcal{I} + \mathcal{A}_i)^{-1} = \mathcal{L}_i^{-1}\left[\mathcal{I} + \sum_{\nu=1}^{\infty}(-\mathcal{A}_i)^\nu\right]. \quad (B.4)$$

Thus by interchanging the integration and summation signs, (B.4) implies that

$$\begin{aligned} Q_i(u, v) &= G^U_{\lambda_i/\gamma^4}(u, v) \\ &\quad + \sum_{\nu=1}^{\infty}(-1)^\nu\int_0^1 G^U_{\lambda_i/\gamma^4}(u, s)A_i^\nu(s, v)\,ds. \end{aligned} \quad (B.5)$$

Applying Lemma 4.2 of Nychka (1995) and Lemma A with $u = \Gamma(t)$ and $v = \Gamma(s)$ to (B.5), there are positive constants α_0, α_0', κ_0', and κ_0^{**} so that, uniformly for $t, s \in [0,1]$,

$$\begin{aligned}
|Q_r(u,v) - G^t_{\lambda_r\gamma^4}(u,v)| &\le \left|\sum_{\nu=1}^{\infty}(-1)^\nu \int_0^1 G^t_{\lambda_r,\gamma^4}(u,s)A_r^\nu(s,v)\,ds\right| \\
&\le \kappa_0^*\lambda_r^{-1/4}\left(\sum_{\nu=1}^{\infty} W^\nu\right)\int_0^1 \exp\left(-\frac{\lambda_r^{-1/4}}{\sqrt{2}}|u-s| - \alpha_0\lambda_r^{-1/4}|s-v|\right)ds \\
&\le \kappa_0^{**}\exp(-\alpha_0^*\lambda_r^{-1/4}|u-v|) \\
&\le \kappa_0^{**}\exp\left(-\alpha_0^*\lambda_r^{-1/4}|t-s|\inf_{s\le u\le t}|\Gamma^{(1)}(u)|\right) \\
&\le \kappa_0^{**}\exp(-\alpha_0^{**}\lambda_r^{-1/4}|t-s|). \qquad \text{(B.6)}
\end{aligned}$$

From (B.1), we also have that

$$Q_r(u,v) - G^{t'}_{\lambda_r/\gamma^4}(u,v) = \frac{f(s)}{\Gamma^{(1)}(s)}\left[G_{\lambda_r}(t,s) - G^{t'}_{\lambda_r/\gamma^4}(\Gamma(t),\Gamma(s))\frac{\Gamma^{(1)}(s)}{f(s)}\right] \qquad \text{(B.7)}$$

Then, (14) is a direct consequence of Lemma A, (B.6), (B.7) and (11). The exponential bounds of (15) can be obtained using the same method.

For the proofs of (16) and (17), we can show from (6) and (7) that

$$\int_0^1 S_{\lambda_r}(t_{ij},s)g_r(s)\,dF_N(s) + \lambda_r\int_0^1 \frac{\partial^2}{\partial s^2}S_{\lambda_r}(t_{ij},s)g_r^{(2)}(s)\,ds = g_r(t_{ij}) \qquad \text{(B.8)}$$

Let $\mathcal{R}_r$ be the integral operator such that

$$\mathcal{R}_r[g_r(\cdot)](t) = \int_0^1 G_{\lambda_r}(t,s)g_r(s)\,d(F-F_N)(s)$$

By (14), (15), and the induction argument in the proof of Nychka (1995), there are positive constants κ_1^*, κ_1^{**} and α_1, such that, uniformly for $t,s\in[0,1]$ and $0\le\mu\le 3$,

$$\left|\frac{\partial^\mu}{\partial t^\mu}\mathcal{R}_r^\nu[G_{\lambda_r}(\cdot,s)](t)\right| \le \kappa_1^*(\kappa_1^{**}D_N\lambda_r^{-1/4})^\nu\lambda_r^{-(\mu+1)/4}\exp[-\alpha_1\lambda_r^{-1/4}|t-s|] \qquad \text{(B.9)}$$

Also, by Cox (1983) (cf. Lemma 3.1 of Nychka 1995), a solution of (B.8) satisfies

$$S_{\lambda_r}(t,t_{ij}) = G_{\lambda_r}(t,t_{ij}) + \mathcal{R}_r[S_{\lambda_r}(\cdot,t_{ij})](t)$$

and, when n is sufficiently large,

$$S_{\lambda_r}(t,t_{ij}) = G_{\lambda_r}(t,t_{ij}) + \sum_{\nu=1}^{\infty}\mathcal{R}_r^\nu[G_{\lambda_r}(\cdot,t_{ij})](t) \qquad \text{(B.10)}$$

Taking $\kappa_2 \ge \kappa_1^*\kappa_1^{**}/(1-\kappa_1^{**}D_N\lambda_r^{-1/4})$, we can derive from (B.9), (B.10) and condition A4 that, uniformly for $t,s\in[0,1]$,

$$\begin{aligned}
|S_{\lambda_r}(t,s) - G_{\lambda_r}(t,s)| &\le \sum_{\nu=1}^{\infty}|\mathcal{R}_r^\nu[G_{\lambda_r}(\cdot,s)](t)| \\
&\le \kappa_1^*\lambda_r^{-1/4}\left(\frac{\kappa_1^{**}D_N\lambda_r^{-1/4}}{1-\kappa_1^{**}D_N\lambda_r^{-1/4}}\right)\exp[-\alpha_1\lambda_r^{-1/4}|t-s|] \\
&\le \kappa_2\lambda_r^{-1/2}D_N\exp[-\alpha_1\lambda_r^{-1/4}|t-s|]
\end{aligned}$$

This completes the proof of (16). Again, (17) can be shown by similar derivations.

APPENDIX C: THREE TECHNICAL LEMMAS

Lemma C.1. If assumptions A1 and A4 of Section 3.1 are satisfied, then, when λ_r is sufficiently small,

$$\int_0^1 G^2_{\lambda_r}(t,s)M_r^{(1)}(s)f(s)\,ds = \frac{1}{4\sqrt{2}}f^{-3/4}(t)\lambda_r^{-1/4}M_r^{(1)}(t)(1+o(1)) \qquad \text{(C.1)}$$

and

$$\int_0^1 G_{\lambda_r}(t,s)M_r^{(2)}(t,s)f(s)\,ds = M_r^{(2)}(t)(1+o(1)) \qquad \text{(C.2)}$$

hold for all $t\in[\tau, 1-\tau]$ with some $\tau>0$.

Proof of Lemma C.1. By Lemma 3.1, the properties of double exponential distributions and straightforward algebra, we can show that, for some positive constants κ, α, and c,

$$\begin{aligned}
&\left|\int_0^1[G^2_{\lambda_r}(t,s) - H^2_{\lambda_r}(t,s)]M_r^{(1)}(s)f(s)\,ds\right| \\
&\le \int_0^1|G_{\lambda_r}(t,s) - H_{\lambda_r}(t,s)|(|G_{\lambda_r}(t,s)| + |H_{\lambda_r}(t,s)|)\left|M_r^{(1)}(s)\right|f(s)\,ds \\
&\le \int_0^1 \kappa^2\lambda_r^{-1/4}\exp(-\alpha\lambda_r^{-1/4}|t-s|)\left|M_r^{(1)}(s)\right|f(s)\,ds \\
&\to c\left|M_r^{(1)}(t)\right|f(t), \quad \text{as} \quad \lambda_r\to 0. \qquad \text{(C.3)}
\end{aligned}$$

Similarly, denoting $u=\Gamma(t)$ and $v=\Gamma(s)$, we can show from (12) and the properties of double exponential distributions that, for λ_r sufficiently small,

$$\begin{aligned}
&\int_0^1 H^2_{\lambda_r}(t,s)M_r^{(1)}(s)f(s)\,ds \\
&= \int_0^1\left[H^t_{\lambda_r/\gamma^4}(u,v)\right]^2 M_r^{(1)}(\Gamma^{-1}(v))\left[\frac{f^{1/4}(\Gamma^{-1}(v))}{\gamma f(\Gamma^{-1}(v))}\right]dv \\
&= \frac{1}{4\sqrt{2}}f^{-3/4}(t)\lambda_r^{-1/4}M_r^{(1)}(t)(1+o(1)). \qquad \text{(C.4)}
\end{aligned}$$

Thus, (C.1) follows from (C.3) and (C.4), and (C.2) can be shown by similar calculations.

Lemma C.2. If $\beta_r(t)$ satisfies assumption A2 of section 3.1 and $g_r(t)$ is a solution of (10), then $g_r^{(4)}(t)\to\beta_r^{(4)}(t)$ uniformly for $t\in[0,1]$ as $\lambda_r\to 0$.

Proof of Lemma C.2. This lemma is a special case of Lemma 6.1 of Nychka (1995).

Lemma C.3. Let $\tilde\beta_r^*(t,\mathbf{w}_0) = N^{-1}\sum_{i=1}^n\sum_{j=1}^{m_i}\{H_{\lambda_r}(t,t_{ij})Z_{ijr}\}$. If the conditions in Theorem 3.1 are satisfied, then $\tilde\beta_r^*(t;\mathbf{w}_0)$ is asymptotically normal in the sense that (18) holds with $\hat\beta_r(t;\mathbf{w}_0)$ replaced by $\tilde\beta_r^*(t,\mathbf{w}_0)$.

Proof of Lemma C.3. Define $U_{ij} = \sum_{l=0}^k(e_{r+1}^T X_i^{(l)})$ and $Z_{ijr} = U_{ij}Y_{ij}$. It is easy to see that

$$E(Z_{ijr}) = E(U_{ij}\mathbf{X}_i^T\boldsymbol{\beta}(t_{ij})) = \beta_r(t_{ij})$$

By A1, A4, (11), (14), and Lemma C.2, we have

$$\begin{aligned}E[\tilde{\beta}_r^*(t,\mathbf{w}_0)]-\beta_r(t) &= \int_0^1 G_{\lambda_r}(t,s)\beta_r(s)f(s)\,ds-\beta_r(t)\\ &\quad+\int_0^1 (H_{\lambda_r}(t,s)-G_{\lambda_r}(t,s))\beta_r(s)f(s)\,ds\\ &\quad+\int_0^1 H_{\lambda_r}(t,s)\beta_r(s)\,d(F_N(s)-F(s))\\ &= -\lambda_r[f(t)]^{-1}g_r^{(4)}(t)(1+o(\lambda_r))\\ &= -\lambda_r b_r(t)(1+o(\lambda_r)). \qquad (C.5)\end{aligned}$$

For the variance of $\tilde{\beta}_r^*(t;\mathbf{w}_0)$, we consider $\mathrm{var}[\tilde{\beta}_r^*(t;\mathbf{w}_0)]=V_I+V_{II}+V_{III}$, where $V_I=N^{-2}\sum_{i=1}^n\sum_{j=1}^{m_i}[H_{\lambda_r}^2(t,t_{ij})\,\mathrm{var}(Z_{ijr})]$,

$$V_{II}=\frac{1}{N^2}\sum_{i=1}^n\sum_{j_1\neq j_2}[H_{\lambda_r}(t,t_{ij_1})H_{\lambda_r}(t,t_{ij_2})\,\mathrm{cov}(Z_{ij_1r},Z_{ij_2r})]$$

and, because the subjects are independent,

$$V_{III}=\frac{1}{N^2}\sum_{i_1\neq i_2}\sum_{j_1}\sum_{j_2}[H_{\lambda_r}(t,t_{i_1j_1})H_{\lambda_r}(t,t_{i_2j_2})\mathrm{cov}(Z_{i_1j_1r},Z_{i_2j_2r})]=0.$$

Because U_{ir} and ϵ_{ij} are independent, we have

$$\mathrm{var}(Z_{ijr})=\mathrm{var}\big[U_{ir}(\mathbf{X}_i^T\boldsymbol{\beta}(t_{ij}))\big]+\mathrm{var}(U_{ir}\epsilon_{ij})=M_r^{(1)}(t_{ij}),$$

hence, by A1 and (C.4),

$$V_I=\frac{1}{4\sqrt{2}}f^{-3/4}(t)N^{-1}\lambda_r^{-1/4}M_r^{(1)}(t)(1+o(1)).$$

Similar to the derivation in (C.3), because

$$\begin{aligned}\mathrm{cov}(Z_{ij_1r},Z_{ij_2r}) &= \mathrm{cov}\{U_{ir}[\mathbf{X}_i^T\boldsymbol{\beta}(t_{ij_1})],U_{ir}[\mathbf{X}_i^T\boldsymbol{\beta}(t_{ij_2})]\}\\ &\quad+\mathrm{cov}\{U_{ir}\epsilon_{ij_1},U_{ir}\epsilon_{ij_2}\}\\ &= M_r^{(2)}(t_{ij_1},t_{ij_2}),\end{aligned}$$

it is straightforward to compute that

$$\begin{aligned}V_{II} &= \left\{\sum_{i=1}^n\left(\frac{m_i}{N}\right)^2-\frac{1}{N}\right\}\iint H_{\lambda_r}(t,s_1)H_{\lambda_r}(t,s_2)M_r^{(2)}\\ &\quad\times(s_1,s_2)f(s_1)f(s_2)\,ds_1\,ds_2(1+o(1))\\ &= \left\{\sum_{i=1}^n\left(\frac{m_i}{N}\right)^2-\frac{1}{N}\right\}M_r^{(2)}(t)(1+o(1)).\end{aligned}$$

The previous equations and (20) imply that $var[\tilde{\beta}_r^*(t,\mathbf{w}_0)]=N^{-1}\lambda_r^{-1/4}\tau_r^2(t)(1+o(1))$. Finally, we can check from A3, (11), and (12) that $\tilde{\beta}^*(t;\mathbf{w}_0)$ satisfies the Lindeberg's condition for double arrays of random variables. The lemma follows from (C.5) and the central limit theorem for double arrays (e.g., Serfling 1980, section 1.9.3).

APPENDIX D: PROOFS OF MAIN THEOREMS

Proof of Theorem 3.1. By the definitions of U_{ir} and $\widehat{U}_{ir}$, conditions A1, A3, and A4 and Lemma 3.1, we have that, when n is sufficiently large,

$$\begin{aligned}\tilde{\beta}_r(t;\mathbf{w}_0)-\tilde{\beta}_r^*(t;\mathbf{w}_0) &= \frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{m_i}[H_{\lambda_r}(t;t_{ij})(\widehat{U}_{ir}-U_{ir})Y_{ij}]\\ &= O_p(n^{-1/2})\end{aligned}$$

and

$$\begin{aligned}\hat{\beta}_r(t;\mathbf{w}_0)-\tilde{\beta}_r(t,\mathbf{w}_0) &= \frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{m_i}\{[S_{\lambda_r}(t,t_{ij})-G_{\lambda_r}(t,t_{ij})]\widehat{U}_{ir}Y_{ij}\}\\ &= o_p(N^{-1/2}\lambda_r^{-1/8}).\end{aligned}$$

Then (18) is a direct consequence of Lemma C.3 and the prior equalities.

Proof of Theorem 3.2. Using the variance-bias squared decomposition on (21), we have

$$\begin{aligned}\mathrm{MSE}[\hat{\beta}_r(t,\mathbf{w}_0)|\mathcal{X}_n] &= \{E[\hat{\beta}_r(t;\mathbf{w}_0)|\mathcal{X}_n]-\beta_r(t)\}^2\\ &\quad+\mathrm{var}(\hat{\beta}_r(t;\mathbf{w}_0)|\mathcal{X}_n), \qquad (D.1)\end{aligned}$$

where, because $Y_{i_1j_1}$ and $Y_{i_2j_2}$ are independent when $i_1\neq i_2$, $\mathrm{var}(\hat{\beta}_r(t;\mathbf{w}_0)|\mathcal{X}_n)=V_I^*+V_{II}^*$,

$$V_I^*=\frac{1}{N^2}\sum_{i=1}^n\sum_{j=1}^{m_i}[S_{\lambda_r}^2(t,t_{ij})\widehat{U}_{ir}^2\,\mathrm{var}(Y_{ij})]$$

and

$$V_{II}^*=\frac{1}{N^2}\sum_{i=1}^n\sum_{j_1\neq j_2}[S_{\lambda_r}(t,t_{ij_1})S_{\lambda_r}(t,t_{ij_2})\widehat{U}_{ir}^2\,\mathrm{cov}(Y_{ij_1},Y_{ij_2})]$$

Using Lemma 3.1 and the derivation of (C.2), we can show that for sufficiently large n,

$$\begin{aligned}&\mathrm{var}(\hat{\beta}_r(t;\mathbf{w}_0)|\mathcal{X}_n)\\ &\quad= \frac{1}{4\sqrt{2}}N^{-1}\lambda_r^{-1/4}[f(t)]^{-3/4}\sigma^2(t)e_{rr}(1+o_p(1))\\ &\qquad+\left[\sum_{i=1}^n\left(\frac{m_i}{N}\right)^2-\frac{1}{N}\right]\rho_\epsilon(t)e_{rr}(1+o_p(1)). \qquad (D.2)\end{aligned}$$

For the conditional bias term of (D.1), we consider that for sufficiently large n,

$$\begin{aligned}&E[\hat{\beta}_r(t;\mathbf{w}_0)|\mathcal{X}_n]-\beta_r(t)\\ &\quad= \frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{m_i}\{S_{\lambda_r}(t,t_{ij})[U_{ir}\mathbf{X}_i^T\boldsymbol{\beta}(t_{ij})-E(U_{ir}\mathbf{X}_i^T\boldsymbol{\beta}(t_{ij}))]\}\\ &\qquad+E\left\{\frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{m_i}[S_{\lambda_r}(t,t_{ij})U_{ir}\mathbf{X}_i^T\boldsymbol{\beta}(t_{ij})]\right\}\\ &\qquad-\beta_r(t)+O_p(n^{-1/2}) \qquad (D.3)\end{aligned}$$

By similar quadratic expansions as V_I^* and V_{II}^*, Lemma 3.1, and the weak law of large numbers, we can show that for sufficiently large n,

$$\begin{aligned}&\left\{\frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{m_i}\{S_{\lambda_r}(t,t_{ij})[U_{ir}\mathbf{X}_i^T\boldsymbol{\beta}(t_{ij})-E(U_{ir}\mathbf{X}_i^T\boldsymbol{\beta}(t_{ij}))]\}\right\}^2\\ &\quad= \left\{\frac{1}{4\sqrt{2}}N^{-1}\lambda_r^{-1/4}[f(t)]^{-3/4}+\left[\sum_{i=1}^n\left(\frac{m_i}{N}\right)^2-\frac{1}{N}\right]\right\}\\ &\qquad\times M_r^{(0)}(t)(1+o_p(1)) \qquad (D.4)\end{aligned}$$

and, furthermore, by Lemma C.2,

$$\begin{aligned}&\left\{\frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{m_i}[S_{\lambda_r}(t,t_{ij})E(U_{ir}\mathbf{X}_i^T\boldsymbol{\beta}(t_{ij}))]-\beta_r(t)\right\}^2\\ &\quad= \lambda_r^2[b_r(t)]^2(1+o_p(1)). \qquad (D.5)\end{aligned}$$

The conclusion of the theorem (23) is then a direct consequence of (D.1) through (D.5).

[Received January 1999. Revised May 2000.]

REFERENCES

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

Brauer, F., and Nohel, J A (1973), *Ordinary Differential Equations: A First Course*, (2nd ed), Menlo Park, CA W. A. Benjamin.

Cheng, S C, and Wei, L J. (2000), *Inferences for a Semiparametric Model With Panel Data*, *Biometrika*, 87, 89–97

Cox, D. D (1983), Asymptotics for M-type Smoothing Splines, *The Annals of Statistics*, 11, 530–551.

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, England Oxford University Press

Efron, B., and Tibshirani, R. J (1993), *An Introduction to the Bootstrap*, New York. Chapman and Hall.

Eubank, R. L. (1988), *Spline Smoothing and Nonparametric Regression*, New York Marcel Dekker.

Fan, J (1992), Design-Adaptive Nonparametric Regression, *Journal of the American Statistical Association*, 87, 998–1004

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

Fan, J, and Zhang, J-T (2000), Functional Linear Models for Longitudinal Data *Journal of the Royal Statistical Society*, Ser B, 62, 303–322.

Gold, M S., Redmond, D. E, and Kleber, H D (1978), Clonidine Blocks Acute Opiate-Withdrawal Symptoms. *The Lancet*, 599–601

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

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

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

Laird, N M, and Ware, J H (1982), Random Effects Models For Longitudinal Data, *Biometrics*, 38, 963–974

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

Lin, X., and Carroll, R J (2000), Nonparametric Function Estimation for Clustered Data When the Predictor is Measured Without/With Error, *Journal of the American Statistical Association*, 95, 520–543

Messer, K. (1991), A Comparison of a Spline Estimate to its Equivalent Kernel Estimate, *The Annals of Statistics* 19, 817–829.

Messer, K, and Goldstein, L (1993), A New Class of Kernels for Nonparametric Curve Estimation, *The Annals of Statistics*, 21, 179–195

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

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

Nychka, D (1995), Splines as Local Smoothers, *The Annals of Statistics*, 23, 1175–1197

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

Ruppert, D., and Wand, M P (1994), Multivariate Weighted Least Squares Regression, *The Annals of Statistics*, 22, 1346–1370.

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

Silverman, B. W. (1984), Spline Smoothing The Equivalent Variable Kernel Method, *The Annals of Statistics*, 12, 898–916

Umbricht-Schneiter, A, Montoya, I D, Hoover, D R, Demuch, K. L., Chiang, C-T, and Preston, K L (1999), Naltrexone Shortened Opioid Detoxification With Buprenorphine, *Drug and Alcohol Dependence*, 56, 181–190

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

Wahba, G (1975), Smoothing Noisy Data With Spline Functions, *Numerische Mathematik*, 24, 383–393.

Wahba, G (1990), *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mathematics, 59, Philadelphia Society for Industrial and Applied Mathematics

Wu, C O, and Chiang, C-T (2000), Kernel Smoothing on Varying Coefficient Models With Longitudinal Dependent Variable, *Statistica Sinica*, 10, 433–456

Wu, C O, Chiang, C-T, and Hoover, D R (1998), Asymptotic Confidence Regions for Kernel Smoothing of a Varying-Coefficient Model With Longitudinal Data, *Journal of the American Statistical Association*, 93, 1388–1402

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