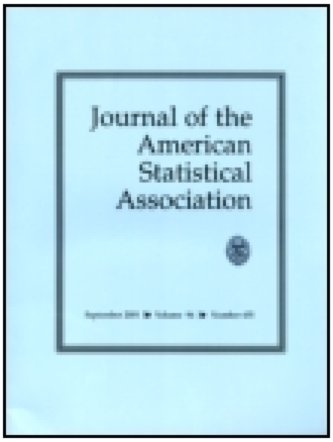

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Partial Linear Regression Models for Clustered Data

Kani Chen[a] & Zhezhen Jin[a]

[a] Kani Chen is Associate Professor, Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong . Zhezhen Jin is Assistant Professor, Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY 10032 . The authors thank R. Carroll, M. Paik, and N. Wang for their helpful comments and encouragement. This research was supported in part by a Hong Kong RGC grant, the National Science Foundation career award, and the New York City Council Speaker's Fund for Public Health Research. The authors also thank the editor, the associate editor, and the referees for their constructive comments.
Published online: 01 Jan 2012.

To cite this article: Kani Chen & Zhezhen Jin (2006) Partial Linear Regression Models for Clustered Data, Journal of the American Statistical Association, 101:473, 195-204, DOI: 10.1198/016214505000000592

To link to this article: http://dx.doi.org/10.1198/016214505000000592

Partial Linear Regression Models for Clustered Data

Kani CHEN and Zhezhen JIN

This article considers the analysis of clustered data via partial linear regression models. Adopting the idea of modeling the within-cluster correlation from the method of generalized estimating equations, a least squares type estimate of the slope parameter is obtained through piecewise local polynomial approximation of the nonparametric component. This slope estimate has several advantages: (a) It attains $n^{1/2}$-consistency without undersmoothing; (b) it is efficient when correct within-cluster correlation is used, assuming multivariate normality of the error; (c) the preceding properties hold regardless of whether or not the nonparametric component is of cluster level; and (d) this estimation method naturally extends to deal with generalized partial linear models. Simulation studies and a real example are presented in support of the theory.

KEY WORDS: Asymptotic bias; Clustered data; Generalized partial linear regression model; Mean squared error; Nonparametric curve estimation; Piecewise local polynomial method.

1. INTRODUCTION

The partial linear model is a popular tool in statistical data analysis. Its application to the analysis of nonclustered data has been well explored. The aim of this article is to propose an estimation method through least squares criterion and piecewise polynomial function approximation and to demonstrate this method's superiority.

The partial linear regression model is of a semiparametric nature in that it relates the response variable with key covariates linearly and with the rest of the covariates nonparametrically. A comprehensive theory about this model for nonclustered data has been well established (see, e.g., Chen 1988; Speckman 1988; Hastie and Tibshirani 1990; Severini and Staniswalis 1994; Carroll, Fan, Gijbels, and Wand 1997). The same issues then naturally arise in the analysis of clustered data using marginal partial linear models. A major problem involved is the modeling and incorporation of the within-cluster correlation. This subject has recently attracted considerable attention (see, e.g., Zhang, Lin, Raz, and Sower 1998; Lin and Carroll 2001a,b; He, Zhu, and Fung 2002). In particular, Lin and Carroll (2001a,b) developed a general profile-kernel methodology for the analysis of clustered data with various models along the lines of generalized estimating equations. Lin and Carroll (2001a) dealt with generalized partial linear models with the nonparametric component not of cluster level. They showed that the $n^{1/2}$-consistency of the derived estimate of the slope parameter requires either working independence and, more seriously, equal marginal density of covariates or, when the correct within-cluster correlation is used, artificial undersmoothing. Following Lin and Carroll (2001a), by undersmoothing we mean the bandwidth $h = o(n^{-1/(2p+3)})$ for odd p, where p is the order of local polynomial smoothers or the piecewise polynomial approximation. Lin and Carroll also argued that, assuming normality for errors, a semiparametrically efficient estimate would still be difficult to obtain, because in theory it involves the inverse of a Fredholm operator, even when correct within-cluster correlation is used. "Our main conclusion is that, unlike for the independent data, the conventional profile-kern method is not semiparametric efficient and must be modified in ad hoc ways (undersmoothing) or to be made less efficient (working independence) to even be made $\sqrt{n}$ consistent" (Lin and Carroll 2001a, p. 1046). In contrast, He et al. (2002) focused on an extension of M-estimation with B-spline smoothing. Although the slope estimators of He et al. (2002) are valid, they do not incorporate within-cluster correlation and fail to achieve semiparametric efficiency in general.

In summary, the key issues of both theoretical and practical significance are (a) necessity/redundancy of artificial undersmoothing, (b) appropriate/inappropriate use of within-cluster correlation and particularly working independence, and (c) efficiency/inefficiency of the slope estimators in the aforementioned sense. Recently, Wang, Carroll, and Lin (2005) tackled the issue of efficiency along the line of the Wang's (2003) seemingly unrelated regression technique by using nonlocal observations in a cluster via residuals to incorporate within-cluster correlation. In this article we propose an alternative estimation method that has the desired properties. This method is based on a hybrid of the ideas of the piecewise polynomial approximation and generalized estimating equations, and offers several appealing advantages. First, the $n^{1/2}$-consistency of the slope estimate is ensured without artificial undersmoothing, without working independence, and without assuming equal marginal densities for the covariates. Second, assuming normality of the errors, this estimate is automatically semiparametrically efficient when correct within-correlation is used. Finally, the method is equally effective regardless of whether or not the nonparametric component is of cluster level, and it can be naturally extended to cover generalized partial linear models.

The next section introduces the partial linear model and presents the proposed slope estimate and its inferences. Section 3 reports on numerical studies. Section 4 contains a local polynomial estimate of $\theta(\cdot)$. Section 5 generalizes the estimation method to generalized partial linear models. Section 6 presents some concluding remarks.

2. ESTIMATION OF THE SLOPE PARAMETER

For the jth observation of the ith cluster, let Y_{ij} be the response variable, let $\mathbf{X}_{ij}$ be a d-vector covariate, and let T_{ij} be a scalar covariate, $j = 1, \dots, J_i$ and $i = 1, \dots, n$. The partial linear regression model under study assumes that

$$Y_{ij} = \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \theta(T_{ij}) + \epsilon_{ij}, \tag{1}$$

Kani Chen is Associate Professor, Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (E-mail: *makchen@ust.hk*). Zhezhen Jin is Assistant Professor, Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY 10032 (E-mail: *zjin@biostat.columbia.edu*). The authors thank R. Carroll, M. Paik, and N. Wang for their helpful comments and encouragement. This research was supported in part by a Hong Kong RGC grant, the National Science Foundation career award, and the New York City Council Speaker's Fund for Public Health Research. The authors also thank the editor, the associate editor, and the referees for their constructive comments.

Journal of the American Statistical Association
March 2006, Vol. 101, No. 473, Theory and Methods
DOI 10.1198/016214505000000592

where $\boldsymbol{\beta}$ is the d-dimensional parameter of primary interest, $\theta(\cdot)$ is the unknown function, and ϵ_{ij} is the error. Let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iJ_i})^{\mathrm{T}}$, $\mathbf{X}_i = (\mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{iJ_i})^{\mathrm{T}}$, $\mathbf{T}_i = (T_{i1}, T_{i2}, \ldots, T_{iJ_i})^{\mathrm{T}}$, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{iJ_i})^{\mathrm{T}}$. Throughout the article, a univariate function acting on a vector is set to be the vector of the function of each component, for example, $\boldsymbol{\theta}(\mathbf{T}_i) = (\theta(T_{i1}), \theta(T_{i2}), \ldots, \theta(T_{iJ_i}))^{\mathrm{T}}$. With this notation, model (1) can be rewritten as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\theta}(\mathbf{T}_i) + \boldsymbol{\epsilon}_i, \qquad (1')$$

where $(\mathbf{Y}_i, \mathbf{X}_i, \mathbf{T}_i)$ are assumed to be independent, $E[\boldsymbol{\epsilon}_i|\mathbf{X}_i, \mathbf{T}_i] = \mathbf{0}$, and $\mathrm{var}(\boldsymbol{\epsilon}_i|\mathbf{X}_i, \mathbf{T}_i) = \boldsymbol{\Sigma}_i$. We also assume the boundedness of $\mathbf{X}_i$ and $\mathbf{T}_i$ and that the eigenvalues of $\boldsymbol{\Sigma}_i$ are uniformly bounded above and bounded below away from 0. For simplicity, we assume throughout the article that $J_i \equiv J$, $(\mathbf{Y}_i, \mathbf{X}_i, \mathbf{T}_i)$ are iid, the densities of $(\mathbf{X}_{ij}, T_{ij})$ are smooth, and the marginal density of T_{ij}, denoted by $f_j(\cdot)$, has support on [0, 1].

We apply the least squares criterion to obtain the estimate of $\boldsymbol{\beta}$, using piecewise polynomial functions of degree p ($p \geq 0$) to estimate $\theta(\cdot)$. Let $1 > h > 0$ be the bandwidth and let $N = \max\{n \geq 1 : n \leq 1/h\}$ be the largest integer less than or equal to $1/h$. Set $I_k = [h(k-1), hk)$, $k = 1, \ldots, N-1$, and $I_N = [h(N-1), 1]$. Then the interval [0, 1] is partitioned into subintervals of I_k, $k = 1, \ldots, N$. Let $\boldsymbol{\psi}(t) = (1, t, t^2, \ldots, t^p)^{\mathrm{T}}$ and $I_k(t) = 1$ if $t \in I_k$ and 0 otherwise. Let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^{\mathrm{T}}, \ldots, \boldsymbol{\alpha}_N^{\mathrm{T}})$, where $\boldsymbol{\alpha}_k$ is a $(p+1)$-vector, and $\mathbf{D}_i = (\mathbf{D}_{i1}, \mathbf{D}_{i2}, \ldots, \mathbf{D}_{iN})$, where $\mathbf{D}_{ik}$ is the $J \times (p+1)$ matrix with the jth row being $\boldsymbol{\psi}(T_{ij})^{\mathrm{T}} I_k(T_{ij})$, $j = 1, \ldots, J$. Consider functions

$$g(t) = \sum_{k=1}^{N} \boldsymbol{\alpha}_k^{\mathrm{T}} \boldsymbol{\psi}(t) I_k(t) = \sum_{k=1}^{N} \left(\sum_{l=0}^{p} \alpha_{kl} t^l \right) I_k(t). \qquad (2)$$

For notational simplicity, the indices n or N may be suppressed in some of the variables and functions defined here and in the sequel. The weighted least squares estimates of $\boldsymbol{\beta}$ and $\theta(\cdot)$ can be obtained by minimizing

$$\sum_{i=1}^{n} \{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{g}(\mathbf{T}_i)\}^{\mathrm{T}} \mathbf{V}_i^{-1} \{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{g}(\mathbf{T}_i)\}$$
$$= \sum_{i=1}^{n} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{D}_i\boldsymbol{\alpha})^{\mathrm{T}} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{D}_i\boldsymbol{\alpha})$$

with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, where $\mathbf{V}_i$ is the working covariance matrix with subscript "n" suppressed. The minimization can be solved algebraically in a straightforward fashion. Denote the resulting minimizers by $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}} = (\hat{\boldsymbol{\alpha}}_1^{\mathrm{T}}, \ldots, \hat{\boldsymbol{\alpha}}_N^{\mathrm{T}})$. Set

$$\begin{aligned}
\mathbf{B}_n &= \sum_{i=1}^{n} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{D}_i, \\
\mathbf{C}_n &= \sum_{i=1}^{n} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{X}_i, \\
\mathbf{A}_n &= \sum_{i=1}^{n} (\mathbf{X}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{X}_i) - \mathbf{C}_n^{\mathrm{T}} \mathbf{B}_n^{-1} \mathbf{C}_n, \qquad \text{and} \\
\mathbf{Z}_i &= \mathbf{V}_i^{-1} (\mathbf{X}_i - \mathbf{D}_i \mathbf{B}_n^{-1} \mathbf{C}_n).
\end{aligned} \qquad (3)$$

We have

$$\hat{\boldsymbol{\beta}} = \mathbf{A}_n^{-1} \sum_{i=1}^{n} \mathbf{Z}_i^{\mathrm{T}} \mathbf{Y}_i \qquad \text{and}$$
$$\hat{\boldsymbol{\alpha}} = \mathbf{B}_n^{-1} \sum_{i=1}^{n} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}). \qquad (4)$$

Then $\hat{\boldsymbol{\beta}}$ is the weighted least squares estimate of $\boldsymbol{\beta}$, and $\sum_{k=1}^{N} \hat{\boldsymbol{\alpha}}_k^{\mathrm{T}} \boldsymbol{\psi}(t) I_k(t)$ is that of $\theta(t)$, $t \in [0, 1]$.

Remark 1. We note that although the piecewise polynomial approximation does not give a smooth estimate of $\theta(\cdot)$, it only plays the role of a pilot or intermediate estimate for the purpose of obtaining the best estimate of $\boldsymbol{\beta}$. For this purpose, it greatly outperforms its counterpart, the local polynomial estimate, in the profile-kernel method, as seen from the $n^{1/2}$-consistency/inconsistency and efficiency/inefficiency of the resulting slope estimates. In Section 4 we also present a local polynomial estimate of $\theta(\cdot)$ along the lines of that of Chen and Jin (2005), which improves the piecewise polynomial estimate in terms of smoothness of the estimate.

The following regularity conditions are needed for the asymptotic properties of the estimate $\hat{\boldsymbol{\beta}}$:

(C1) There exists a $\lambda \in (0, 1)$ such that as $n \to \infty$, $n^{\lambda} h \to \infty$ and $n^{1/2} h^{2(p+1)} \to 0$.

(C2) The function $\theta(\cdot)$ has continuous derivatives of order $p+1$. The eigenvalues of $\mathbf{V}_i$ and $\mathbf{A}_n/n$ are uniformly bounded above and bounded below away from 0 for all large n in probability. Moreover,

$$\sup\{\|\mathbf{V}_i^{-1} - \tilde{\mathbf{V}}_i^{-1}\| : i = 1, \ldots, n\} = O_P(n^{-1/4}), \qquad (5)$$

where $\tilde{\mathbf{V}}_i^{-1} = \{f_{k,l}^n(\mathbf{X}_i, \mathbf{T}_i)\}_{1 \leq k,l \leq J}$ and $f_{k,l}^n$ is a smooth and deterministic function. Here and throughout, $\|\cdot\|$ for a vector is its Euclidean norm, and for any square matrix $\mathbf{A}$ of $K \times K$ dimension, $\|\mathbf{A}\| = \max\{\|\mathbf{A}\mathbf{x}\|/\|\mathbf{x}\| : \|\mathbf{x}\| = 1, \mathbf{x} \in \mathbb{R}^K\}$. The functions $f_{k,l}^n$ converge uniformly as $n \to \infty$.

Remark 2. We note that, as with some other notation, subscript "n" is omitted from $\mathbf{V}_i$ and $\tilde{\mathbf{V}}_i$. Condition (C2) is assumed here to avoid overly erratic modeling/estimation of the covariance matrices. It can be relaxed for establishing the following proposition with increasing technicalities. The elements of $\mathbf{V}_i$ are usually based on averages of iid random variables, and thus (5) is usually satisfied by the empirical approximation. This is the case for nonrandom working covariance matrices (including working independence), exchangeable correlations, autoregressive correlations, and spatial correlations.

Proposition 1. Under conditions (C1) and (C2), we have the following results:

(a) The estimate $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normal, that is,

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}^*), \qquad (6)$$

where $\boldsymbol{\Sigma}^*$ can be consistently estimated by

$$n\mathbf{A}_n^{-1} \sum_{i=1}^{n} (\mathbf{Z}_i^{\mathrm{T}} \boldsymbol{\Sigma}_i \mathbf{Z}_i) \mathbf{A}_n^{-1}. \qquad (7)$$

In particular, the foregoing result holds if $h \sim n^{-1/(2p+3)}$.

(b) Suppose, moreover, that condition (C2) holds with $\tilde{\mathbf{V}}_i^{-1}$ replaced by $\boldsymbol{\Sigma}_i^{-1}$ and the conditional distribution of $\boldsymbol{\epsilon}_i$, given $\mathbf{X}_i$ and $\mathbf{T}_i$, is normal for all $i \geq 1$. Then (a) holds, and $\hat{\boldsymbol{\beta}}$ is semiparametrically efficient.

Remark 3. Part (a) of the proposition establishes the key consistency and asymptotic normality of the slope estimate and thus justifies the use of this estimate. Because the asymptotic variance also has a simple closed form, its inference is easy to obtain as well. Part (a) also indicates that the proposed estimation method, unlike some existing methods, is valid under a quite broad range of bandwidth selection. Moreover, it does not rely on any kind of artificial undersmoothing. Such a virtue renders simplicity in computation involving bandwidth selection. Many well-established criteria for bandwidth selection available for nonparametric curve estimation can be directly adopted. Moreover, implementation of the usual and generalized cross-validation criteria for the selection of h can be simplified considerably by using the algebra of deletion in classic linear regression (Atkinson 1985, pp. 19–21). Part (b) of the proposition shows that if $\boldsymbol{\epsilon}_i$ is conditionally Gaussian, then the estimate of $\boldsymbol{\beta}$ is semiparametrically efficient when the working covariance matrix is close to the true covariance matrix $\boldsymbol{\Sigma}_i$. The analytic expression of the asymptotic variance involves a projection, and the projected function [e.g., $g_k^*(\cdot)$ defined in the App.], in general can be expressed only through the inverse of a Fredholm operator. But this does not prevent a closed-form variance estimate, as given in (7).

Remark 4. The underlying reason for the success of the piecewise polynomial method lies in the fact that $\mathbf{Z}_i$ is orthogonal/uncorrelated to $\mathbf{D}_i\boldsymbol{\alpha}$ for all $\boldsymbol{\alpha}$; see (A.15) in the Appendix. The bias of $\hat{\boldsymbol{\beta}}$ originates from the terms $\mathbf{Z}_i'\boldsymbol{\theta}(\mathbf{T}_i)$. Because of this orthogonality, this term is approximately mean 0 and thereby makes the bias negligible without artificially forcing h to be too small (undersmoothing). In contrast, for the profile-kernel estimate of the slope $\boldsymbol{\beta}$, the bias also comes from the terms $[\mathbf{V}_i^{-1}\{\mathbf{X}_i - (\partial/\partial\boldsymbol{\beta})\hat{\boldsymbol{\theta}}(\mathbf{T}_i, \boldsymbol{\beta})\}]^{\mathrm{T}}\{\hat{\boldsymbol{\theta}}(\mathbf{T}_i) - \boldsymbol{\theta}(\mathbf{T}_i)\}$, where $\hat{\boldsymbol{\theta}}(\mathbf{T}_i) - \boldsymbol{\theta}(\mathbf{T}_i)$ is approximately a function of $\mathbf{T}_i$ and is of order h^{p+1}. This bias term cannot be negligible, because $\mathbf{V}_i^{-1}\{\mathbf{X}_i - (\partial/\partial\boldsymbol{\beta})\hat{\boldsymbol{\theta}}(\mathbf{T}_i, \boldsymbol{\beta})\}$ is not orthogonal to the space of $\mathbf{f}(\mathbf{T}_i)$ for all functions $f(\cdot)$, unless the marginal densities of $(\mathbf{X}_i, \mathbf{T}_i)$ are all equal and $\mathbf{V}_i$ is identity matrix (see the discussion after thm. 1 in Lin and Carroll 2001a).

3. NUMERICAL STUDIES

In the numerical examples that follow, data are generated from the model

$$Y_{ij} = x_{ij1}\beta_1 + x_{ij2}\beta_2 + \theta(T_{ij}) + \epsilon_{ij}, \qquad j = 1, \ldots, J; i = 1, \ldots, n,$$

where $\theta(t) = \sin(2t) + \cos(t)$, $\beta_1 = 1$, $\beta_2 = 0$, $J = 3$, and $n = 100$. The x_{ij1} and x_{ij2} are time-varying covariates. The covariate $x_{ij1} = T_{ij} + \delta_{ij}$, where δ_{ij} follows $\mathrm{N}(0, (.5j)^2)$ and the covariate x_{ij2} follows a Bernoulli distribution with success probability .5. The error $(\epsilon_{i1}, \ldots, \epsilon_{iJ})^{\mathrm{T}}$ follows a multivariate normal distribution with mean 0, exchangeable correlation .6, and marginal variances $.3^2$, $.5^2$, and $.4^2$. The T_{ij}'s, $j = 1, \ldots, J$, are independently generated from the uniform distribution on [0, 1]. Estimation of the slope $\boldsymbol{\beta}$ was done by three methods: the proposed method, Lin and Carroll's profile-kernel method (2001a,b), and Wang et al.'s method (2005).

To be comparable with Lin and Carroll's (2001a,b) method, which is based on local linear smoothing, the piecewise local polynomial approximation is taken with degree $p = 1$. The Epanechnikov kernel $K(t) = .75(1 - t^2)I\{|t| \leq 1\}$ is used in Lin and Carroll's method. Three different working correlation matrices are used: the identity matrix, the true correlation matrix, and the estimated correlation matrix. The estimated correlation matrix was obtained with the moment estimation method after fitting the model with the identity correlation matrix. In all tables, $\boldsymbol{\Sigma}_0$ is the true covariance matrix, $\hat{\boldsymbol{\Sigma}}_0$ is the estimated covariance matrix, $\mathbf{I}$ is the identity matrix, and h is the bandwidth. The bandwidth h_0 is selected by adopting the idea of cross-validation by deleting one cluster at a time in each simulated dataset. In Table 1, bias is the empirical bias, SE is the empirical standard error, SEE is the average of the estimated standard error, MSE is the empirical mean squared error of the estimates with the new method, RMSE_1 is the ratio of the empirical MSE of the estimates with Lin and Carroll's method to the MSE, and RMSE_2 is the ratio of the empirical MSE of the estimates obtained with the method of Wang et al. to the MSE. Simulation results are all based on 1,000 replications.

The results show that the new method performs well in finite samples. The parameter estimates are virtually unbiased, the estimated standard errors are close to the empirical standard errors, and the parameter estimates with the true or estimated correlation matrix have smaller bias, smaller standard error, and smaller MSE compared with the estimates with the identity correlation matrix, which is the quintessence of generalized estimating equations that the correct specification of correlation matrix leads to efficient estimates. The ratios RMSE_1 in the table demonstrate that the estimates obtained with the new method are much more efficient than the estimates obtained with Lin and Carroll's method when the true correlation or estimated correlation matrix is used. Moreover, the estimates obtained with the new method using the true correlation or estimated correlation matrix have smaller SE and MSE compared with the estimates obtained with Lin and Carroll's method using the identity correlation matrix. When the true correlation matrix is used, except for the case with bandwidth $h = .05$, the ratios RMSE_2 are very close to 1, indicating that the two methods have similar efficiency. When the estimated correlation matrix is used, the ratios RMSE_2 are around value 1.1, indicating that our method performs slightly better than Wang et al.'s method.

In summary, all of the foregoing simulation results support the theory that the proposed new method with piecewise local polynomial approximation yields the best slope estimate when the within-cluster correlation is correctly specified, regardless of whether or not the nonparametric component is of cluster level.

In what follows, we apply the proposed estimation method to a real example of a longitudinal hormone study previously analyzed by Zhang et al. (1998) and He et al. (2002). In this study,

Table 1. Summary of Simulation Results for Slope Estimates With Independent T_{ij}

			New method				*Lin and Carroll's method*			*Wang et al.'s method*		
h	$\boldsymbol{\Sigma}_i$	β	*Bias* ($\times 10^3$)	*SE* ($\times 10^2$)	*SEE* ($\times 10^2$)	*MSE* ($\times 10^2$)	*Bias* ($\times 10^3$)	*SE* ($\times 10^2$)	*RMSE*$_1$	*Bias* ($\times 10^3$)	*SE* ($\times 10^2$)	*RMSE*$_2$
.05	$\boldsymbol{\Sigma}_0$	β_1	.31	.88	.89	.01	−1.36	1.30	2.22	.25	.84	.91
		β_2	.34	3.60	3.56	.13	−.11	4.73	1.72	−.21	3.41	.90
	$\hat{\boldsymbol{\Sigma}}_0$	β_1	.23	.91	.87	.01	−1.36	1.30	2.09	.29	.95	1.11
		β_2	.51	3.71	3.47	.14	−.28	4.76	1.65	1.39	3.93	1.12
	I	β_1	.17	1.18	1.08	.01	.41	1.27	1.15	.02	1.12	.91
		β_2	2.22	4.65	4.29	.22	2.81	5.62	1.46	1.87	4.42	.90
.1	$\boldsymbol{\Sigma}_0$	β_1	.22	.85	.85	.01	−1.31	1.29	2.33	.15	.82	.94
		β_2	−.20	3.41	3.37	.12	.32	4.69	1.89	−.18	3.36	.97
	$\hat{\boldsymbol{\Sigma}}_0$	β_1	.17	.87	.84	.01	−1.36	1.29	2.26	.24	.91	1.10
		β_2	−.08	3.48	3.31	.12	.23	4.71	1.83	.83	3.71	1.14
	I	β_1	.03	1.13	1.08	.01	.34	1.26	1.23	−.01	1.11	.96
		β_2	2.06	4.44	4.28	.20	2.39	5.56	1.57	2.09	4.35	.96
.2	$\boldsymbol{\Sigma}_0$	β_1	.09	.82	.83	.01	−1.26	1.30	2.50	.16	.82	1.00
		β_2	−.46	3.36	3.29	.11	.31	4.66	1.93	−.11	3.32	.98
	$\hat{\boldsymbol{\Sigma}}_0$	β_1	.07	.84	.82	.01	−1.33	1.30	2.44	.26	.89	1.14
		β_2	−.30	3.41	3.23	.12	.17	4.67	1.88	.72	3.60	1.12
	I	β_1	−.08	1.11	1.09	.01	.31	1.25	1.26	−.03	1.11	.98
		β_2	1.83	4.37	4.28	.19	2.21	5.54	1.61	1.97	4.32	.98
h_0	$\boldsymbol{\Sigma}_0$	β_1	.22	.83	.84	.01	−1.35	1.29	2.45	.14	.82	.97
		β_2	.05	3.39	3.33	.11	.43	4.68	1.91	−.09	3.35	1.00
	$\hat{\boldsymbol{\Sigma}}_0$	β_1	.18	.85	.83	.01	−1.37	1.30	2.36	.25	.90	1.12
		β_2	.23	3.46	3.27	.12	.36	4.69	1.84	.89	3.66	1.12
	I	β_1	−.01	1.12	1.08	.01	.30	1.25	1.25	−.04	1.11	.97
		β_2	2.32	4.40	4.28	.19	2.27	5.56	1.59	2.06	4.35	.97

34 healthy women had urine samples collected in a menstrual cycle and urinary progesterone assayed on alternative days. We consider the model

$$Y_{ij} = \beta_1 \mathrm{Age}_i + \beta_2 \mathrm{BMI}_i + \theta(t_{ij}) + \epsilon_{ij},$$

where Y is the logarithm of progesterone hormone level, BMI is the body mass index, and $\theta(\cdot)$ is a completely unspecified smooth function. In our analysis, Age is centered at a median of 36 years and divided by 100, BMI is centered at a median of 26 kg/m^2 and divided by 100, and t_{ij} is centered at a median of 14 days and divided by 10, all as in the study of Zhang et al. (1998). We consider two types of covariance matrices, the identity matrix and the estimated covariance matrix obtained by Zhang et al. (1998), with ϵ_{ij} decomposed as

$$\epsilon_{ij} = b_i + U_i(t_{ij}) + \delta_{ij}$$

with $b_i \sim \mathrm{N}(0, \phi)$, $\delta_{ij} \sim \mathrm{N}(0, \sigma^2)$, and $U_i(t)$ following the nonhomogeneous Ornstein–Uhlenbeck process with $\mathrm{var}(U_i(t)) = \exp\{\xi_0 + \xi_1 t + \xi_2 t^2\}$ and $\mathrm{corr}(U_i(t), U_i(s)) = \rho^{|t-s|}$, where $\phi = .2617$, $\sigma^2 = .1366$, $\rho = .0963$, $\xi_0 = -2.1563$, $\xi_1 = 3.6710$, and $\xi_2 = -2.1598$. Table 2 summarizes the results of data analysis with different choices of bandwidth h and $p = 1$.

The table shows that both β_1 and β_2 are not significantly different from 0 at significance level .05. These findings are consistent with the results obtained by Zhang et al. (1998) and He et al. (2002). It also indicates that the standard errors obtained with the estimated covariance matrix are smaller than those obtained with the identity matrix.

4. ESTIMATION OF $\theta(\cdot)$

The piecewise polynomial approximation serves as an intermediate estimate for obtaining an efficient estimate of $\boldsymbol{\beta}$ in an easy and straightforward fashion. It may not be appealing to use it directly as the estimate of the function $\theta(\cdot)$, because it is only piecewise smooth. But with the efficient estimate of $\boldsymbol{\beta}$, an ideal estimate of $\theta(\cdot)$ can be obtained via conventional local polynomial regression smoothers. Specifically, let t_0 be a fixed time point. Let $K(\cdot)$ be a smooth symmetric kernel function with support on $[-1, 1]$, and define $K_h(t) = K(t/h)/h$. Then $\theta(t_0)$ can be estimated by the first component of

$$\left[\sum_{i=1}^{n} \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i \mathbf{T}_i(t_0)\right]^{-1} \sum_{i=1}^{n} \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i(\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}), \qquad (8)$$

Table 2. Estimates of the Regression Coefficients for the Hormone Data

		Identity matrix **I**		*Estimated covariance matrix* $\hat{\boldsymbol{\Sigma}}_i$	
h	*Parameter*	*Estimate*	*SE*	*Estimate*	*SE*
.1	β_1	1.2009	2.1864	.9840	1.9122
	β_2	−2.4408	2.6287	−2.7183	2.3562
.2	β_1	1.5179	2.0751	1.0252	1.9070
	β_2	−2.6882	2.5001	−2.7502	2.3532
.3	β_1	1.6404	2.0179	1.0652	1.9040
	β_2	−2.9303	2.4404	−2.9062	2.3513
.4	β_1	1.7165	2.0057	1.1243	1.9034
	β_2	−2.7863	2.4455	−2.8150	2.3514
.5	β_1	1.7898	2.0006	1.1631	1.9032
	β_2	−2.8327	2.4303	−2.8577	2.3511

where

$$\mathbf{T}_i(t_0) = \begin{pmatrix} 1 & (T_{i1}-t_0) & \cdots & (T_{i1}-t_0)^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (T_{iJ}-t_0) & \cdots & (T_{iJ}-t_0)^p \end{pmatrix},$$

$$\mathbf{w}_i = \{\mathbf{K}_i^{-1/2}\mathbf{V}_i\mathbf{K}_i^{-1/2}\}^{-1},$$

and

$$\mathbf{K}_i = \mathrm{diag}\{K_h(T_{i1}-t_0), \ldots, K_h(T_{iJ}-t_0)\}.$$

Here the matrix inverse is the Moore–Penrose generalized inverse (see also Chen and Jin 2005). For any symmetric $J \times J$ matrix $\mathbf{A}$, its generalized inverse, denoted still by $\mathbf{A}^{-1}$, is defined as a symmetric matrix such that $\mathbf{A}\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}$ and $\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}$. Specifically, let $\mathbf{A} = \boldsymbol{\Gamma}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_J)\boldsymbol{\Gamma}^{\mathrm{T}}$ by the decomposition of symmetric matrices, where $\boldsymbol{\Gamma}$ is an orthonormal matrix, that is, $\boldsymbol{\Gamma}^{\mathrm{T}} = \boldsymbol{\Gamma}^{-1}$. Then $\mathbf{A}^{-1} = \boldsymbol{\Gamma}\,\mathrm{diag}(1/\lambda_1, \ldots, 1/\lambda_J)\boldsymbol{\Gamma}^{\mathrm{T}}$, where $1/0$ denotes 0. Obviously, the generalized inverse of the 0 matrix is the 0 matrix.

Under the very general conditions of the "existence of partial density" (as defined in Chen and Jin 2005), the asymptotic bias and variances of $\hat{\theta}(t_0)$ can be derived in an entirely analogous way to that of Chen and Jin (2005). But the notation is complex, and the proof becomes very lengthy. For simplicity and clarity, we present the properties of $\hat{\theta}(t_0)$ under a relatively restrictive condition that the joint density of $(T_{i1}, \ldots, T_{iJ})^{\mathrm{T}}$ exists (see Chen and Jin 2005 for the general conditions).

Some more notation is needed. Let $\mathbf{e}_1 = (1, 0, \ldots, 0)^{\mathrm{T}}$ be the $(p+1)$-vector with first element 1 and the remaining elements 0. Set $\mu_i = \int t^i K(t)\,dt$, $\nu_i = \int t^i K^2(t)\,dt$, $\mathbf{c}_p = (\mu_{p+1}, \ldots, \mu_{2p+1})^{\mathrm{T}}$, $\mathbf{S} = (\mu_{i+j})_{0\le i,j\le p}$, and $\mathbf{S}^* = (\nu_{i+j})_{0\le i,j\le p}$. It is easy to see that $\mathbf{S}$ and $\mathbf{S}^*$ are $(p+1)\times(p+1)$ matrices. Let $\mathcal{H}_n$ be the σ-algebra generated by $(\mathbf{X}_i, \mathbf{T}_i)$, $i = 1, \ldots, n$.

Proposition 2. Under conditions (C1) and (C2), if the joint density of $(T_{i1}, \ldots, T_{iJ})^{\mathrm{T}}$ exists with the marginal density of T_{ij} being $f_j(\cdot)$ for $j = 1, \ldots, J$, then the following hold:

(a) The conditional bias of $\hat{\theta}(t_0)$ is

$$Bias\{\hat{\theta}(t_0)|\mathcal{H}_n\} = \mathbf{e}_1^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{c}_p\frac{\theta^{(p+1)}(t_0)}{(p+1)!}h^{p+1} + o_P(h^{p+1}). \quad (9)$$

For even p, the main term is 0.

(b) The conditional variance of $\hat{\theta}(t_0)$ is

$$\mathrm{var}\{\hat{\theta}(t_0)|\mathcal{H}_n\} = \mathbf{e}_1^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\mathbf{e}_1\frac{\sum_{j=1}^J \sigma_j^2(t_0)f_j(t_0)/v_j^2(t_0)}{nh(\sum_{j=1}^J f_j(t_0)/v_j(t_0))^2} + o_P\left(\frac{1}{nh}\right), \quad (10)$$

where $v_j(t_0)$ and $\sigma_j^2(t_0), j = 1, \ldots, J$, are the limits of the diagonal elements of $\widetilde{\mathbf{V}}_1$ and that of $\boldsymbol{\Sigma}_1$ as $t \to t_0$. Expression (10) is minimized when $v_j(t_0) = \sigma_j^2(t_0)$ for $j = 1, \ldots, J$ with minimum

$$\frac{\mathbf{e}_1^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\mathbf{e}_1}{nh(\sum_{j=1}^J f_j(t_0)/\sigma_j^2(t_0))} + o_P\left(\frac{1}{nh}\right). \quad (11)$$

Remark 5. Part (a) of Proposition 2 suggests using a local polynomial with degree p an odd number to decrease the size of bias. Part (b) shows that the asymptotic variance is minimized when the true variances are used as the working variances.

Remark 6. Proposition 2 shows that the asymptotic conditional bias for $\hat{\theta}(t_0)$ is, at least to the leading term, the same as that of $\hat{\theta}(t_0)$ defined by replacing $\hat{\boldsymbol{\beta}}$ with the true $\boldsymbol{\beta}$. As a result, in deriving the asymptotic bias and variance of $\hat{\theta}(t_0)$, we can replace $\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}$ by $\boldsymbol{\theta}(\mathbf{T}_i) + \boldsymbol{\epsilon}_i$.

Remark 7. In conventional fashion, it is also possible to obtain the asymptotic conditional bias and variance of the estimators of $\theta^{(k)}(t_0)$, $k = 1, \ldots, p$ (see the results in Chen and Jin 2005; thm. 3.1 in Fan and Gijbels 1996).

5. EXTENSION TO GENERALIZED PARTIAL LINEAR MODELS

The estimation method derived in Section 2 has a straightforward extension to cover generalized partial linear models. Specifically, suppose that Y_{ij}, conditioning on $(\mathbf{X}_i, \mathbf{T}_i)$, has conditional mean $\mu\{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \theta(T_{ij})\}$ and conditional variance Φ_{ij}, where $\mu(\cdot)$ is a known function and Φ_{ij} may depend on $\mathbf{X}_i, \mathbf{T}_i, \boldsymbol{\beta}$, and $\theta(\cdot)$. Let $\dot{\mu}(\cdot)$ denote the derivative of $\mu(\cdot)$. Along the lines of the methodology of generalized estimating equations, an estimating equation is derived as

$$\sum_{i=1}^n \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\Delta}_i\mathbf{V}_i^{-1}\{\mathbf{Y}_i - \boldsymbol{\mu}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{D}_i\boldsymbol{\alpha})\} = \mathbf{0} \quad (12)$$

and

$$\sum_{i=1}^n \mathbf{D}_i^{\mathrm{T}}\boldsymbol{\Delta}_i\mathbf{V}_i^{-1}\{\mathbf{Y}_i - \boldsymbol{\mu}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{D}_i\boldsymbol{\alpha})\} = \mathbf{0}, \quad (13)$$

where $\boldsymbol{\Delta}_i = \mathrm{diag}(\dot{\boldsymbol{\mu}}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{D}_i\boldsymbol{\alpha}))$, $\mathbf{V}_i = \mathrm{diag}(\Phi_{i1}, \ldots, \Phi_{iJ})^{1/2} \times \mathbf{R}_i\,\mathrm{diag}(\Phi_{i1}, \ldots, \Phi_{iJ})^{1/2}$, and $\mathbf{R}_i$ is the modeled conditional correlation matrix of $(\epsilon_{i1}, \ldots, \epsilon_{iJ})$. Along the lines of the proof of Proposition 1 and using mainly the Taylor expansion, we can show that under conditions (C1) and (C2), $\hat{\boldsymbol{\beta}}$ is asymptotically normal with asymptotic variance as the limit of the upper-left $d \times d$ submatrix of $n\boldsymbol{\Omega}_n^{-1}\boldsymbol{\Psi}_n\boldsymbol{\Omega}_n^{-1}$, where

$$\boldsymbol{\Omega}_n = \sum_{i=1}^n \begin{pmatrix}\mathbf{X}_i^{\mathrm{T}} \\ \mathbf{D}_i^{\mathrm{T}}\end{pmatrix}\boldsymbol{\Delta}_i\mathbf{V}_i^{-1}\boldsymbol{\Delta}_i\,(\mathbf{X}_i \quad \mathbf{D}_i)$$

and

$$\boldsymbol{\Psi}_n = \sum_{i=1}^n \begin{pmatrix}\mathbf{X}_i^{\mathrm{T}} \\ \mathbf{D}_i^{\mathrm{T}}\end{pmatrix}\boldsymbol{\Delta}_i\mathbf{V}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{V}_i^{-1}\boldsymbol{\Delta}_i\,(\mathbf{X}_i \quad \mathbf{D}_i).$$

Again, the most important features are that there is no need to undersmooth to ensure the $n^{1/2}$-consistency of $\hat{\boldsymbol{\beta}}$ and that $\hat{\boldsymbol{\beta}}$ is automatically semiparametric efficient when the within-cluster correlation is correctly specified. Moreover, it is easy to see that there is a simple iterative algorithm for computing $\hat{\boldsymbol{\beta}}$, which is based on solving for $\boldsymbol{\alpha}$ with a fixed $\boldsymbol{\beta}$ using (13) and solving for $\boldsymbol{\beta}$ with a fixed $\boldsymbol{\alpha}$ using (12).

6. DISCUSSION

The major advantage of using a piecewise local polynomial approximation in the proposed estimation method is not only that the $n^{1/2}$-consistency and asymptotic normality of the slope estimate is ensured for all regular working correlation matrices, but also that the bandwidth selection is the least restrictive and allows use of the conventional criterion. Under the assumption of normal errors, semiparametric efficiency is achieved when the within-cluster correlation is correctly specified. Some existing estimation approaches ignore the within-cluster correlation and possibly lead to loss of efficiency (e.g., He et al. 2002). Some others estimate and use the within-cluster correlation but do not in general ensure $n^{1/2}$-consistency and efficiency (e.g., Lin and Carroll 2001a). We believe that the method developed in this article outperforms existing estimation methods in very important aspects.

APPENDIX: PROOFS OF THE PROPOSITIONS

Let $(\mathbf{X}, \mathbf{T}, \boldsymbol{\epsilon}), (\mathbf{X}_i, \mathbf{T}_i, \boldsymbol{\epsilon}_i), i \geq 1$, be iid random triplets and let $\tilde{\boldsymbol{\epsilon}}, \tilde{\boldsymbol{\epsilon}}_i, i \geq 1$, be iid random vectors of dimension J with conditional means 0 and conditional variances $\tilde{\mathbf{V}}, \tilde{\mathbf{V}}_i, i \geq 1$, given $(\mathbf{X}, \mathbf{T}), (\mathbf{X}_i, \mathbf{T}_i)$, $i \geq 1$. For simplicity of notation, we assume that $f_{k,l}^n$ in the definition of $\tilde{\mathbf{V}}_i$ is free of n. For convenience, $\|\cdot\|$ for any $(Np + p) \times K$ or $K \times (Np + p)$ matrix, where K is fixed and does not vary with n, is defined as the square root of the sum of squares of all entries of the matrix. To avoid trivialities, the inverse of a noninvertible matrix is defined as the 0 matrix.

Lemma A.1. Assume that condition (C2) holds for the matrix $\tilde{\mathbf{V}}$. Then there exist functions $\tilde{g}_1, \ldots, \tilde{g}_d$ defined on $[0, 1]$ such that for any function g with $E[g^{\mathrm{T}}(\mathbf{T})\tilde{\mathbf{V}}^{-1}g(\mathbf{T})] < \infty$,

$$E\big([\mathbf{X} - \tilde{\mathbf{G}}(\mathbf{T})]^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}g(\mathbf{T})\big) = \mathbf{0}, \tag{A.1}$$

where $\tilde{\mathbf{G}}(\mathbf{T}) = (\tilde{g}_1(\mathbf{T}), \ldots, \tilde{g}_d(\mathbf{T}))$ is a $J \times d$ matrix. The foregoing equation can be expressed in the form of the Fredholm equations for $\tilde{g}_k, k = 1, \ldots, d$, as

$$\sum_{j=1}^{J} E(v^{jj}|T_j = t)f_j(t)\tilde{g}_k(t) + \sum_{j=1}^{J}\sum_{\substack{l=1\\ l\neq j}}^{J} E\big(v^{jl}\tilde{g}_k(T_l)|T_j = t\big)f_j(t)$$

$$= \sum_{j=1}^{J}\sum_{l=1}^{J} E(v^{jl}x_{lk}|T_j = t)f_j(t), \tag{A.2}$$

where v^{jl} is the limit of $f_{j,l}^n$ as $n \to \infty$, x_{lk} is the (l, k)th element of $\mathbf{X}$, and f_j is the marginal density of T_j, the jth component of $\mathbf{T}$. In particular, if $\tilde{\mathbf{V}}$ is the identity matrix, then $\tilde{g}_k(t) = (\sum_{j=1}^{J} E(x_{jk}|T_j = t)f_j(t))/(\sum_{j=1}^{J} f_j(t))$, and if $J = 1$, then $\tilde{g}_k(t) = E(x_{1k}v^{11}|T_1 = t)/E(v^{11}|T_1 = t)$.

Proof. For simplicity and without loss of generality, we present proofs for $p = 0$ and $d = 1$. In this case, $\tilde{\mathbf{G}}(\mathbf{T}) = \tilde{g}_1(\mathbf{T})$. Set

$$\mathcal{Q} = \big\{g^{\mathrm{T}}(\mathbf{T})\tilde{\mathbf{V}}^{-1}\tilde{\boldsymbol{\epsilon}} : \text{all functions } g \text{ on } [0, 1] \text{ such that } E[g^{\mathrm{T}}(\mathbf{T})\tilde{\mathbf{V}}^{-1}g(\mathbf{T})] < \infty\big\}.$$

Note that $\mathcal{Q}$ is a linear subspace of the Hilbert space of random variables with mean 0 and finite variance, with the inner product being the covariance.

Because $\mathcal{Q}$ is a linear subspace, let $\tilde{g}_1^{\mathrm{T}}(\mathbf{T})\tilde{\mathbf{V}}^{-1}\tilde{\boldsymbol{\epsilon}} \in \mathcal{Q}$ be the projection of $\mathbf{X}^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\tilde{\boldsymbol{\epsilon}}$ onto $\mathcal{Q}$. Then the orthogonality of the projection implies that

$$\begin{aligned} &E\big[(\mathbf{X} - \tilde{g}_1(\mathbf{T}))^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}g(\mathbf{T})\big] \\ &= E\big[\big(\mathbf{X}^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\tilde{\boldsymbol{\epsilon}} - \tilde{g}_1(\mathbf{T})^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\tilde{\boldsymbol{\epsilon}}\big)\big(g(\mathbf{T})^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\tilde{\boldsymbol{\epsilon}}\big)^{\mathrm{T}}\big] \\ &= 0, \end{aligned}$$

for any g satisfying $E[g^{\mathrm{T}}(\mathbf{T})\tilde{\mathbf{V}}^{-1}g(\mathbf{T})] < \infty$. Then (A.1) holds, and (A.2) follows from (A.1) by direct calculation of expectation and considering the Dirac delta functions for g. The proof is complete.

Lemma A.2. Assume that conditions (C1) and (C2) hold and that $\mathbf{V}_i = \tilde{\mathbf{V}}_i$ for $i \geq 1$. Let $\mathbf{D}$ be defined based on $\mathbf{T}$ in the same way as $\mathbf{D}_i$ based on $\mathbf{T}_i$, and set $\bar{\mathbf{B}}_n = nE(\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\mathbf{D})$ and $\bar{\mathbf{C}}_n = nE(\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\mathbf{X})$. Let $\mathbf{a}_i$, $1 \leq i \leq n$, be any random vectors of J dimension. Then

$$\left\|\sum_{i=1}^{n}\mathbf{D}_i^{\mathrm{T}}\right\| = O_P\big(nh^{1/2}\big) = \|\mathbf{C}_n\|, \tag{A.3}$$

$$\left\|(\mathbf{C}_n^{\mathrm{T}}\mathbf{B}_n^{-1} - \bar{\mathbf{C}}_n^{\mathrm{T}}\bar{\mathbf{B}}_n^{-1})\sum_{i=1}^{n}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{a}_i\right\| = O_P\big(n^{1/2}h^{-1/2}\big)\sup_{1\leq i\leq n}\|\mathbf{a}_i\|, \tag{A.4}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\mathbf{B}_n^{-1}\mathbf{C}_n\|^2 = o_P(1). \tag{A.5}$$

Proof. For notational simplicity but without loss of generality, we prove the lemma for the case with $p = 0$ and $d = 1$; the proofs of general cases are only slightly different. Throughout the proof, $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum eigenvalues of a square matrix, and c is a certain large but fixed constant that does not depend on n or $\mathbf{b}$. Observe that for any vector $\mathbf{b} = (b_1, \ldots, b_N)^{\mathrm{T}}$ of N-dimension, $\mathbf{D}^{\mathrm{T}}\mathbf{D}\mathbf{b} = (b_1l_1, \ldots, b_Nl_N)^{\mathrm{T}}$, where $l_k = \sum_{j=1}^{J} I(T_j \in I_k)$. Then $\mathbf{b}^{\mathrm{T}}E(\mathbf{D}^{\mathrm{T}}\mathbf{D})\mathbf{b} = E(\|\mathbf{D}\mathbf{b}\|^2) = \sum_{k=1}^{N} b_k^2\sum_{j=1}^{J} P(T_j \in I_k)$. Note that $\inf\{\sum_{j=1}^{J} f_j(t) : t \in [0, 1]\} > 0$. Therefore, $(h/c)\|\mathbf{b}\|^2 \leq E(\|\mathbf{D}\mathbf{b}\|^2) \leq ch\|\mathbf{b}\|^2$ for some fixed $c > 0$ and all $\mathbf{b} \in \mathbb{R}^N$. Hence

$$\frac{h}{c} \leq \lambda_{\min}(E[\mathbf{D}^{\mathrm{T}}\mathbf{D}]) \leq \lambda_{\max}(E[\mathbf{D}^{\mathrm{T}}\mathbf{D}]) \leq ch. \tag{A.6}$$

Write

$$\begin{aligned} &\left\|\sum_{i=1}^{n}\mathbf{D}_i^{\mathrm{T}}\mathbf{D}_i - nE(\mathbf{D}^{\mathrm{T}}\mathbf{D})\right\|^2 \\ &= \sup_{\mathbf{b}\in\mathbb{R}^N}\left\{\left\|\left[\sum_{i=1}^{n}\mathbf{D}_i^{\mathrm{T}}\mathbf{D}_i - nE(\mathbf{D}^{\mathrm{T}}\mathbf{D})\right]\mathbf{b}\right\|^2 \Big/ \|\mathbf{b}\|^2\right\} \\ &= \sup_{\mathbf{b}\in\mathbb{R}^N}\left\{\sum_{k=1}^{N} b_k^2\left\{\sum_{i=1}^{n}\sum_{j=1}^{J}[I(T_{ij} \in I_k) - P(T_{ij} \in I_k)]\right\}^2 \Big/ \|\mathbf{b}\|^2\right\} \\ &\leq \sup_{k=1,\ldots,N}\left|\sum_{i=1}^{n}\sum_{j=1}^{J}[I(T_{ij} \in I_k) - P(T_{ij} \in I_k)]\right|^2. \end{aligned}$$

Applying Bernstein's inequality (e.g., de la Peña and Giné 1999, p. 166), we have, for all $n \geq 1$,

$$\begin{aligned} &P\left(\sup_{k=1,\ldots,N}\left|\sum_{i=1}^{n}\sum_{j=1}^{J}[I(T_{ij} \in I_k) - P(T_{ij} \in I_k)]\right| > c(nh\log(n))^{1/2}\right) \\ &\leq \sum_{k=1}^{N} P\left(\left|\sum_{i=1}^{n}\sum_{j=1}^{J}[I(T_{ij} \in I_k) - P(T_{ij} \in I_k)]\right| > c(nh\log(n))^{1/2}\right) \\ &\leq cn^{-2}, \end{aligned}$$

where c is some large but fixed positive constant. The Borel–Cantelli lemma ensures that

$$\left\|\sum_{i=1}^{n}\mathbf{D}_i^{\mathrm{T}}\mathbf{D}_i - nE(\mathbf{D}^{\mathrm{T}}\mathbf{D})\right\| = O\big(\{nh\log(n)\}^{1/2}\big), \tag{A.7}$$

with probability 1 for all large n. Combining (A.6) with (A.7), we have, with probability 1 for all large n,

$$\frac{nh}{c} \le \lambda_{\min}\left(\sum_{i=1}^n \mathbf{D}_i^{\mathrm{T}}\mathbf{D}_i\right) \le \lambda_{\max}\left(\sum_{i=1}^n \mathbf{D}_i^{\mathrm{T}}\mathbf{D}_i\right) \le nhc$$

and

$$\begin{aligned}\frac{(nh)^{-1}}{c} &\le \lambda_{\min}\left(\left\{\sum_{i=1}^n \mathbf{D}_i^{\mathrm{T}}\mathbf{D}_i\right\}^{-1}\right)\\ &\le \lambda_{\max}\left(\left\{\sum_{i=1}^n \mathbf{D}_i^{\mathrm{T}}\mathbf{D}_i\right\}^{-1}\right)\\ &\le (nh)^{-1}c\end{aligned}$$

for some large but fixed $c > 0$. Therefore,

$$\begin{aligned}\|\mathbf{B}_n^{-1}\| + \|\bar{\mathbf{B}}_n^{-1}\| &= O_P((nh)^{-1}) \qquad \text{and}\\ \|\mathbf{B}_n\| + \|\bar{\mathbf{B}}_n\| &= O_P(nh),\end{aligned} \tag{A.8}$$

because $\tilde{\mathbf{V}}_i$ and $\tilde{\mathbf{V}}_i^{-1}$ are assumed to be uniformly bounded. Write

$$\begin{aligned}E\left(\left\|\sum_{i=1}^n \mathbf{D}_i^{\mathrm{T}}\right\|^2\right) &= \sum_{j=1}^J E\left(\sum_{k=1}^N\left[\sum_{i=1}^n I(T_{ij}\in I_k)\right]^2\right)\\ &= O(1)\sum_{j=1}^J\sum_{k=1}^N\left((nh)^2 + \sum_{i=1}^n \operatorname{var}\{I(T_{ij}\in I_k)\}\right)\\ &= O(1)\sum_{j=1}^J\sum_{k=1}^N\big((nh)^2 + nh\big)\\ &= O(n^2h),\end{aligned}$$

because $N = O(1/h)$. Therefore, $\|\sum_{i=1}^n \mathbf{D}_i^{\mathrm{T}}\| = O_P(nh^{1/2})$ and $\|\sum_{i=1}^n \mathbf{D}_i^{\mathrm{T}}\mathbf{a}_i\| = O_P(nh^{1/2})\sup_{1\le i\le n}\|\mathbf{a}_i\|$ for any random vectors $\mathbf{a}_i \in \mathbb{R}^J$. Hence $\|\mathbf{C}_n\| = O_P(nh^{1/2})$ and (A.3) holds because $\mathbf{X}_i$ and $\tilde{\mathbf{V}}_i^{-1}$ are assumed to be uniformly bounded. With similar calculations, we can show that $E(\|\mathbf{C}_n - \bar{\mathbf{C}}_n\|^2) = O(n)$ and $\|\bar{\mathbf{C}}_n\| = O_P(nh^{1/2})$. Moreover,

$$E\big(\|(\mathbf{B}_n - \bar{\mathbf{B}}_n)\mathbf{b}\|^2\big) \le nhc\|\mathbf{b}\|^2 \tag{A.9}$$

for some fixed $c > 0$ and for all $\mathbf{b} \in \mathbb{R}^N$. Apply (A.8) and (A.9) and write

$$\begin{aligned}&\|\mathbf{B}_n^{-1}\mathbf{C}_n - \bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n\|\\ &\quad= \|\mathbf{B}_n^{-1}[\mathbf{C}_n - \bar{\mathbf{C}}_n - (\mathbf{B}_n - \bar{\mathbf{B}}_n)\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n]\|\\ &\quad\le \|\mathbf{B}_n^{-1}\|\big(\|\mathbf{C}_n - \bar{\mathbf{C}}_n\| + \|(\mathbf{B}_n - \bar{\mathbf{B}}_n)\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n\|\big)\\ &\quad= O_P((nh)^{-1})\big[O\big(n^{1/2}\big) + O_P\big(n^{1/2}h^{1/2}\big)\|\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n\|\big]\\ &\quad= O_P((nh)^{-1})\big[O\big(n^{1/2}\big) + O_P\big(n^{1/2}h^{1/2}\big)(nh)^{-1}nh^{1/2}\big]\\ &\quad= O_P\big(n^{-1/2}h^{-1}\big).\end{aligned}$$

Therefore,

$$\begin{aligned}&\left\|(\mathbf{C}_n^{\mathrm{T}}\mathbf{B}_n^{-1} - \bar{\mathbf{C}}_n^{\mathrm{T}}\bar{\mathbf{B}}_n^{-1})\sum_{i=1}^n \mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i\mathbf{a}_i\right\|\\ &\qquad= O_P\big(n^{-1/2}h^{-1}nh^{1/2}\big)\sup_{1\le i\le n}\|\mathbf{a}_i\|\\ &\qquad= O_P\big((nh^{-1})^{1/2}\big)\sup_{1\le i\le n}\|\mathbf{a}_i\|,\end{aligned}$$

and (A.4) holds. Moreover,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n \|\mathbf{D}_i^{\mathrm{T}}(\mathbf{B}_n^{-1}\mathbf{C}_n - \bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n)\|^2\\ &\quad= O\left(\frac{1}{n}\right)\|\mathbf{B}_n^{-1}\mathbf{C}_n - \bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n\|^2\left\|\sum_{i=1}^n \mathbf{D}_i\mathbf{D}_i^{\mathrm{T}}\right\|\\ &\quad= O_P\big(n^{-1}(n^{-1}h^{-2})nh\big) = O_P((nh)^{-1})\\ &\quad= o_P(1).\end{aligned} \tag{A.10}$$

Set $\mathcal{Q}_h = \{g(\mathbf{T})^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\tilde{\boldsymbol{\epsilon}} : g(t) = \sum_{k=1}^N c_kI_k(t)$, where $c_k, k = 1,\ldots,N$, are real constants$\}$. Then $[\mathbf{D}\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n]^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\tilde{\boldsymbol{\epsilon}} \in \mathcal{Q}_h$, and it is the projection of $\mathbf{X}^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\tilde{\boldsymbol{\epsilon}}$ onto $\mathcal{Q}_h$. Recall the definition of $\mathcal{Q}$ in the proof of Lemma A.1. Because $\lim_{h\to 0}\mathcal{Q}_h = \mathcal{Q}$, the projection of $\mathbf{X}^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\tilde{\boldsymbol{\epsilon}}$ onto $\mathcal{Q}_h$ also converges to that onto $\mathcal{Q}$, that is,

$$E\big\{\big\|[\mathbf{D}\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n - \tilde{\mathbf{G}}(\mathbf{T})]^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\tilde{\boldsymbol{\epsilon}}\big\|^2\big\} \to 0.$$

It follows that $E\|\mathbf{D}\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n - \tilde{\mathbf{G}}(\mathbf{T})\|^2 \to 0$. Therefore,

$$\frac{1}{n}\sum_{i=1}^n\big\|\mathbf{D}_i\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n - \tilde{\mathbf{G}}(\mathbf{T}_i)\big\|^2 = o_P(1). \tag{A.11}$$

Then (A.5) follows from (A.10) and (A.11). The proof is complete.

Lemma A.3. Assume that conditions (C1) and (C2) hold and that $\mathbf{V}_i = \tilde{\mathbf{V}}_i$. Then $\mathbf{A}_n/n \to \boldsymbol{\Sigma}_2^*$ in probability and $n^{1/2}\mathbf{A}_n^{-1}\sum_{i=1}^n \mathbf{Z}_i^T\boldsymbol{\epsilon}_i \to \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}^*)$, where $\boldsymbol{\Sigma}^* = (\boldsymbol{\Sigma}_2^*)^{-1}\boldsymbol{\Sigma}_1^*(\boldsymbol{\Sigma}_2^*)^{-1}$ and $\boldsymbol{\Sigma}_1^*$ and $\boldsymbol{\Sigma}_2^*$ are defined in (A.13) and (A.12). Moreover, $n\mathbf{A}_n^{-1}\sum_{i=1}^n(\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{Z}_i)\mathbf{A}_n^{-1}$ converges to $\boldsymbol{\Sigma}^*$ in probability.

Proof. Recall the definition of $\tilde{\mathbf{G}}$ in Lemma A.1. Write

$$\begin{aligned}\frac{1}{n}\mathbf{A}_n &= \frac{1}{n}\sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\mathbf{X}_i\\ &= \frac{1}{n}\sum_{i=1}^n(\mathbf{X}_i - \tilde{\mathbf{G}}(\mathbf{T}_i))^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{X}_i\\ &\quad+ \frac{1}{n}\sum_{i=1}^n\big(\tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\mathbf{B}_n^{-1}\mathbf{C}_n\big)^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{X}_i.\end{aligned}$$

The second term is of order $O_P(1/n)\sum_{i=1}^n\|\tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\mathbf{B}_n^{-1}\mathbf{C}_n)\|$, which converges to 0 in probability by (A.5) of Lemma A.2. The first term is an average of iid random variables. It then follows from the law of large numbers that $(1/n)\mathbf{A}_n$ converges in probability to

$$\begin{aligned}\boldsymbol{\Sigma}_2^* &\equiv E\big[(\mathbf{X} - \tilde{\mathbf{G}}(\mathbf{T}))^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}(\mathbf{X} - \tilde{\mathbf{G}}(\mathbf{T}))\big]\\ &= E\big[(\mathbf{X} - \tilde{\mathbf{G}}(\mathbf{T}))^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\mathbf{X}\big].\end{aligned} \tag{A.12}$$

Write

$$\begin{aligned}n^{-1/2}\sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\epsilon}_i &= n^{-1/2}\sum_{i=1}^n(\mathbf{X}_i - \tilde{\mathbf{G}}(\mathbf{T}_i))^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\epsilon}_i\\ &\quad+ n^{-1/2}\sum_{i=1}^n\big(\tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\mathbf{B}_n^{-1}\mathbf{C}_n\big)^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\epsilon}_i.\end{aligned}$$

The conditional variance of the second term is of order $(1/n) \times \sum_{i=1}^n\|\tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\mathbf{B}_n^{-1}\mathbf{C}_n\|^2$, which converges to 0 in probability by (A.5) of Lemma A.2. Therefore, the second term is $o_P(1)$. The summands of the first term are iid random vectors with mean 0 and variance

$$\boldsymbol{\Sigma}_1^* = E\big[(\mathbf{X} - \tilde{\mathbf{G}}(\mathbf{T}))^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\boldsymbol{\Sigma}\tilde{\mathbf{V}}^{-1}(\mathbf{X} - \tilde{\mathbf{G}}(\mathbf{T}))\big], \tag{A.13}$$

where $\boldsymbol{\Sigma} = \mathrm{var}(\boldsymbol{\epsilon}|\mathbf{X}, \mathbf{T})$. Hence it follows from the classical central limit theorem that it converges in distribution to $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_1^*)$. As a result,

$$n^{1/2}\mathbf{A}_n^{-1}\sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\epsilon}_i \to \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}^*),$$

where $\boldsymbol{\Sigma}^* = (\boldsymbol{\Sigma}_2^*)^{-1}\boldsymbol{\Sigma}_1^*(\boldsymbol{\Sigma}_2^*)^{-1}$ can be consistently approximated by $n\mathbf{A}_n^{-1}\sum_{i=1}^n(\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{Z}_i)\mathbf{A}_n^{-1}$. The proof is complete.

Lemma A.4. Assume that conditions (C1) and (C2) hold and that $\mathbf{V}_i = \tilde{\mathbf{V}}_i$. Then

$$\sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{T}_i) = o_P(n^{1/2}). \tag{A.14}$$

Proof. Recall the definition of $\mathbf{Z}_i$ in (3) and the fact that

$$\sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\mathbf{X}_i = \mathbf{A}_n \qquad \text{and} \qquad \sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\mathbf{D}_i = \mathbf{0}. \tag{A.15}$$

It follows that $E(\mathbf{Z}_1^{\mathrm{T}}\mathbf{D}_1) = \cdots = E(\mathbf{Z}_n^{\mathrm{T}}\mathbf{D}_n) = 0$. Define $\mathcal{G}_i = \{\mathbf{D}_i\boldsymbol{\alpha} : \text{for all } \boldsymbol{\alpha}\}$ as the collection of all J-vectors of the form $g(\mathbf{T}_i)$, where g is piecewise polynomial of order p as given in (2). Let Π denote the projection of a random vector onto a linear space of random vectors of the same dimension. Write $\Pi(\boldsymbol{\theta}(\mathbf{T}_i)|\mathcal{G}_i) = \mathbf{D}_i\boldsymbol{\alpha}$ for some $(Np+p)$-vector $\boldsymbol{\alpha}$ and let $\boldsymbol{\Delta}_i = \boldsymbol{\theta}(\mathbf{T}_i) - \mathbf{D}_i\boldsymbol{\alpha}$. Likewise, let $\Pi(\tilde{\mathbf{G}}(\mathbf{T}_i)|\mathcal{G}_i) = \mathbf{D}_i\tilde{\boldsymbol{\alpha}}$ for some $(Np+p)\times d$ matrix $\tilde{\boldsymbol{\alpha}}$, and let $\tilde{\boldsymbol{\Delta}}_i = \tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\tilde{\boldsymbol{\alpha}}$. Note that $\boldsymbol{\Delta}_i$ and $\tilde{\boldsymbol{\Delta}}_i$ are the d-vector and $J\times d$ matrix. The Taylor expansion implies that $\|\boldsymbol{\Delta}_i\| = O(h^{p+1}) = \|\tilde{\boldsymbol{\Delta}}_i\|$. By the definitions of $\bar{\mathbf{B}}_n$ and $\bar{\mathbf{C}}_n$, we know that $E([\mathbf{V}_i^{-1}(\mathbf{X}_i - \mathbf{D}_i\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n)]^{\mathrm{T}} \times \mathbf{D}_i) = \mathbf{0}$. Lemma A.1 implies that $E([\mathbf{V}_i^{-1}(\mathbf{X}_i - \tilde{\mathbf{G}}(\mathbf{T}_i))]^{\mathrm{T}}\mathbf{D}_i) = \mathbf{0}$. Hence $E([\tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n]^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{D}_i) = \mathbf{0}$. Therefore,

$$\begin{aligned}
&E(\tilde{\boldsymbol{\Delta}}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\tilde{\boldsymbol{\Delta}}_i)\\
&\quad = E([\tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n]^{\mathrm{T}}\mathbf{V}_i^{-1}[\tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n])\\
&\qquad + E([\mathbf{D}_i(\tilde{\boldsymbol{\alpha}} - \bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n)]^{\mathrm{T}}\mathbf{V}_i^{-1}[\mathbf{D}_i(\tilde{\boldsymbol{\alpha}} - \bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n)])\\
&\quad \ge E([\tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n]^{\mathrm{T}}\mathbf{V}_i^{-1}[\tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n]).
\end{aligned}$$

As a result, $E(\|\tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n\|^2) = O(1)E(\tilde{\boldsymbol{\Delta}}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\tilde{\boldsymbol{\Delta}}_i) = O(h^{2p+2})$ and $\|\tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n\| = O_P(h^{p+1})$. Observe (A.15) and write

$$\begin{aligned}
&\sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{T}_i)\\
&\quad = \sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\{\boldsymbol{\theta}(\mathbf{T}_i) - \mathbf{D}_i\boldsymbol{\alpha}\}\\
&\quad = \sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\Delta}_i\\
&\quad = \sum_{i=1}^n (\mathbf{X}_i - \mathbf{D}_i\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n)^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\Delta}_i\\
&\qquad + \sum_{i=1}^n (\mathbf{D}_i\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n - \mathbf{D}_i\mathbf{B}_n^{-1}\mathbf{C}_n)^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\Delta}_i\\
&\quad = \sum_{i=1}^n [\mathbf{X}_i - \tilde{\mathbf{G}}(\mathbf{T}_i)]^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\Delta}_i + \sum_{i=1}^n [\tilde{\mathbf{G}}(\mathbf{T}_i) - \mathbf{D}_i\bar{\mathbf{B}}_n^{-1}\bar{\mathbf{C}}_n]^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\Delta}_i\\
&\qquad + (\bar{\mathbf{C}}_n^{\mathrm{T}}\bar{\mathbf{B}}_n^{-1} - \mathbf{C}_n^{\mathrm{T}}\mathbf{B}_n^{-1})\sum_{i=1}^n \mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\Delta}_i\\
&\quad = (I) + (II) + (III), \quad \text{say.}
\end{aligned}$$

The summands of the first term (I) are n independent random vectors of d dimension, and the mean is 0 by Lemma A.1. The variance of (I) is on the order of $nh^{2(p+1)}$ as $\boldsymbol{\Delta}_i = O(h^{p+1})$. Therefore, (I) is of order $o_P(n^{1/2})$. The second term (II) is of order $O_P(nh^{2p+2}) = o_P(n^{1/2})$ by the preceding argument. It follows from (A.4) of Lemma A.2 that the third term (III) is of order $O_P(n^{1/2}h^{-1/2})\sup_{1\le i\le n}\|\boldsymbol{\Delta}_i\| = O_P(n^{1/2}h^{-1/2}h^{p+1}) = o_P(n^{1/2})$. The proof is complete.

Proof of Proposition 1

Proof of (a). Write

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \mathbf{A}_n^{-1}\sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\mathbf{Y}_i\\
&= \mathbf{A}_n^{-1}\sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}[\mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\theta}(\mathbf{T}_i) + \boldsymbol{\epsilon}_i]\\
&= \boldsymbol{\beta} + \mathbf{A}_n^{-1}\sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{T}_i) + \mathbf{A}_n^{-1}\sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\epsilon}_i.
\end{aligned} \tag{A.16}$$

If $\mathbf{V}_i = \tilde{\mathbf{V}}_i$, then (6) follows from Lemmas A.3 and A.4. In general, $\mathbf{V}_i \ne \tilde{\mathbf{V}}_i$. Replacing $\mathbf{V}_i$ by $\tilde{\mathbf{V}}_i$, we can analogously define $\tilde{\mathbf{A}}_n, \tilde{\mathbf{B}}_n$, and $\tilde{\mathbf{C}}_n$. It is straightforward to verify that $\sum_{i=1}^n \tilde{\mathbf{Z}}_i^{\mathrm{T}}\mathbf{X}_i = \tilde{\mathbf{A}}_n$ and $\sum_{i=1}^n \tilde{\mathbf{Z}}_i^{\mathrm{T}}\mathbf{D}_i = 0$. Note that condition (C2) is $\sup_{1\le i\le n}\|\tilde{\mathbf{V}}_i^{-1} - \mathbf{V}_i^{-1}\| = O_P(n^{-1/4})$. Observe (A.3)–(A.5) and (A.8) of Lemma A.2. We can use the Delta method to show that $\|\tilde{\mathbf{B}}_n^{-1} - \mathbf{B}_n^{-1}\| = O_P(n^{-1/4})\|\tilde{\mathbf{B}}_n^{-1}\| = O_P(n^{-5/4}h^{-1})$, $\|\tilde{\mathbf{C}}_n - \mathbf{C}_n\| = O_P(n^{3/4}h^{1/2}) = O_P(n^{-1/4})\|\tilde{\mathbf{C}}_n\|$, and $\|\tilde{\mathbf{B}}_n^{-1}\tilde{\mathbf{C}}_n - \mathbf{B}_n^{-1}\mathbf{C}_n\| = O_P(n^{-1/4}h^{-1/2})$. Observe that $\|\boldsymbol{\theta}(\mathbf{T}_i) - \mathbf{D}_i\boldsymbol{\alpha}\| = O(h^{p+1}) = o(n^{-1/4})$ by condition (C1), and that $\|\sum_{i=1}^n \mathbf{D}_i^{\mathrm{T}}\tilde{\mathbf{V}}_i^{-1}\boldsymbol{\Delta}_i\| = O_P(nh^{1/2}h^{p+1}) = o_P(n^{3/4}h^{1/2})$. Write

$$\begin{aligned}
&\sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{T}_i)\\
&\quad = \sum_{i=1}^n \mathbf{Z}_i^{\mathrm{T}}\{\boldsymbol{\theta}(\mathbf{T}_i) - \mathbf{D}_i\boldsymbol{\alpha}\}\\
&\quad = \sum_{i=1}^n (\mathbf{X}_i - \mathbf{D}_i\mathbf{B}_n^{-1}\mathbf{C}_n)^{\mathrm{T}}\{\mathbf{V}_i^{-1} - \tilde{\mathbf{V}}_i^{-1}\}\boldsymbol{\Delta}_i\\
&\qquad + (\tilde{\mathbf{B}}_n^{-1}\tilde{\mathbf{C}}_n - \mathbf{B}_n^{-1}\mathbf{C}_n)^{\mathrm{T}}\sum_{i=1}^n \mathbf{D}_i^{\mathrm{T}}\tilde{\mathbf{V}}_i^{-1}\boldsymbol{\Delta}_i + \sum_{i=1}^n \tilde{\mathbf{Z}}_i^{\mathrm{T}}\boldsymbol{\Delta}_i\\
&\quad = O_P(nn^{-1/4}h^{p+1}) + o_P(n^{-1/4}h^{-1/2}n^{3/4}h^{1/2}) + \sum_{i=1}^n \tilde{\mathbf{Z}}_i^{\mathrm{T}}\boldsymbol{\Delta}_i\\
&\quad = \sum_{i=1}^n \tilde{\mathbf{Z}}_i^{\mathrm{T}}\boldsymbol{\Delta}_i + o_P(n^{1/2})\\
&\quad = \sum_{i=1}^n \tilde{\mathbf{Z}}_i^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{T}_i) + o_P(n^{1/2})\\
&\quad = o_P(n^{1/2}),
\end{aligned}$$

where the last equality holds by Lemma A.4. Analogously, it can be shown that $\|\mathbf{A}_n/n - \tilde{\mathbf{A}}_n/n\| = O_P(n^{-1/4}) = o_P(1)$ and $\sum_{i=1}^n(\tilde{\mathbf{Z}}_i - \mathbf{Z}_i)^{\mathrm{T}}\boldsymbol{\epsilon}_i = o_P(n^{1/2})$. Hence (6) holds in general, and the proof of part (a) is complete.

Proof of (b). Under the normality assumption, the log-likelihood function based on the ith cluster is

$$loglik(\boldsymbol{\beta},\theta) \propto -\frac{1}{2}\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \boldsymbol{\theta}(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \boldsymbol{\theta}(\mathbf{T}_i)\}.$$

Differentiating with respect to $\boldsymbol{\beta}$ and (the smoothing function) $\theta(\cdot)$, we obtain

$$\dot{l}_\beta = \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i \qquad \text{and} \qquad \dot{l}_\theta = \{\mathbf{g}(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i,$$

where $\mathbf{g}$ is an arbitrary smooth function. Let $\mathbf{x}_{ik}$ be the kth column of $\mathbf{X}_i$. Denote the projection of $\mathbf{x}_{ik}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i$ onto the linear space spanned by all $\dot{l}_\theta$'s as $\{\mathbf{g}_k^*(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i$. Then the $\tilde{\mathbf{G}}(\mathbf{T})$ in the proof of (a) becomes $\mathbf{G}^*(\mathbf{T}) = (\mathbf{g}_1^*(\mathbf{T}),\ldots,\mathbf{g}_d^*(\mathbf{T}))$. Note that the efficient score for $\boldsymbol{\beta}$ is $(\mathbf{X}_i - \mathbf{G}^*(\mathbf{T}_i))^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i$, and the efficient Fisher information for $\boldsymbol{\beta}$ is

$$\begin{aligned}\mathcal{I}(\boldsymbol{\beta}) &\equiv E\{(\mathbf{X}_i - \mathbf{G}^*(\mathbf{T}_i))^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{X}_i - \mathbf{G}^*(\mathbf{T}_i))\}\\ &= E\{(\mathbf{X}_i - \mathbf{G}^*(\mathbf{T}_i))^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\}.\end{aligned}$$

In contrast, if $\mathbf{V}_i$ is replaced by $\boldsymbol{\Sigma}_i$, then it is seen that $\boldsymbol{\Sigma}_1^*$ and $\boldsymbol{\Sigma}_2^*$ in the proof of (a) satisfies $\boldsymbol{\Sigma}_1^* = \boldsymbol{\Sigma}_2^* = \mathcal{I}(\boldsymbol{\beta})$. Therefore, $\hat{\boldsymbol{\beta}}$ is semiparametric efficient with asymptotic variance $\{\mathcal{I}(\boldsymbol{\beta})\}^{-1}$. The proof is complete.

Proof of Proposition 2

By (A.16),

$$(8) = I + II,$$

where

$$I = \left[\sum_{i=1}^n \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i\mathbf{T}_i(t_0)\right]^{-1} \times \left[\sum_{i=1}^n \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i\boldsymbol{\theta}(\mathbf{T}_i) - \sum_{i=1}^n \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i\mathbf{X}_i\mathbf{A}_n^{-1}\sum_{j=1}^n \mathbf{Z}_j^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{T}_j)\right]$$

and

$$II = \left[\sum_{i=1}^n \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i\mathbf{T}_i(t_0)\right]^{-1} \times \sum_{i=1}^n\left[\mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i - \sum_{j=1}^n \mathbf{T}_j^{\mathrm{T}}(t_0)\mathbf{w}_j\mathbf{X}_j\mathbf{A}_n^{-1}\mathbf{Z}_i^{\mathrm{T}}\right]\boldsymbol{\epsilon}_i.$$

It is easy to see that the asymptotic bias of $\hat{\theta}(t_0)$ can be calculated from the term I and its variance can be calculated from the term II.

Proof of (a). By (A.14) and the Taylor expansion, it follows that

$$\begin{aligned}&Bias\{\hat{\theta}(t_0)|\mathcal{H}_n\}\\ &= \mathbf{e}_1^{\mathrm{T}}\mathbf{I} - \theta(t_0)\\ &= \mathbf{e}_1^{\mathrm{T}}\left[\sum_{i=1}^n \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i\mathbf{T}_i(t_0)\right]^{-1}\left[\sum_{i=1}^n \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i\boldsymbol{\theta}(\mathbf{T}_i)\right]\\ &\quad + o_P(n^{-1/2}) - \theta(t_0)\\ &= \mathbf{e}_1^{\mathrm{T}}\left[\sum_{i=1}^n \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i\mathbf{T}_i(t_0)\right]^{-1}\\ &\quad \times \left[\sum_{i=1}^n \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i(\mathbf{T}_i - t_0)^{p+1}\left(\frac{\theta^{(p+1)}(t_0)}{(p+1)!} + o(1)\right)\right]\\ &\quad + o_P(n^{-1/2}).\end{aligned} \tag{A.17}$$

The $(k+1,l+1)$th $(k,l=0,\ldots,p)$ element of $\sum_{i=1}^n \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i \times \mathbf{T}_i(t_0)$ is

$$\sum_{i=1}^n\sum_{k_1=1}^J\sum_{k_2=1}^J (T_{ik_1} - t_0)^k w_{i,k_1,k_2}(T_{ik_2} - t_0)^l, \tag{A.18}$$

where w_{i,k_1,k_2} is the (k_1,k_2)th element of $(\mathbf{K}_i^{-1/2}\mathbf{V}_i\mathbf{K}_i^{-1/2})^{-1}$. The condition that the joint density of $(T_{i1},\ldots,T_{iJ})$ exists implies that

$$P(\text{at least two of } \{K_h(T_{ij} - t_0): j = 1,\ldots,J\} \text{ are positive}) = O(h^2). \tag{A.19}$$

This implies that as $h \to 0$,

$$w_{i,k_1,k_2} = \begin{cases} K_h(T_{ik_1} - t_0)\dfrac{(1+o(1))}{\tilde{v}_{i,k_1,k_1}} & \text{if } k_1 = k_2\\ K_h^{1/2}(T_{ik_1} - t_0)K_h^{1/2}(T_{ik_2} - t_0)o(1) & \text{otherwise,}\end{cases} \tag{A.20}$$

where $\tilde{v}_{i,k_1,k_1}$ is the (k_1,k_1)th element of $\mathbf{V}_i^{-1}$.

It follows from the law of large numbers, condition (C2), and (A.20) that

$$\frac{1}{n}\sum_{i=1}^n\sum_{k_1=1}^J\sum_{k_2=1}^J (T_{ik_1} - t_0)^k w_{i,k_1,k_2}(T_{ik_2} - t_0)^l = h^{k+l}\left\{\mu_{k+l}\sum_{j=1}^J \frac{f_j(t_0)}{v_j(t_0)} + o(1)\right\},$$

where $v_j(t_0)$ is the limit of (j,j)th element of $\tilde{\mathbf{V}}_1$. Therefore,

$$\frac{1}{n}\sum_{i=1}^n \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i\mathbf{T}_i(t_0) = \left\{\sum_{j=1}^J \frac{f_j(t_0)}{v_j(t_0)}\right\}\mathbf{HSH}(1+o(1)), \tag{A.21}$$

where $\mathbf{H} = \mathrm{diag}(1,h,\ldots,h^p)$. Similarly, we can show that

$$\frac{1}{n}\sum_{i=1}^n \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i(\mathbf{T}_i - t_0)^{p+1} = h^{p+1}\mathbf{Hc}_p\left(\sum_{j=1}^J \frac{f_j(t_0)}{v_j(t_0)}\right)(1+o(1)). \tag{A.22}$$

Thus (9) follows from condition (C1), (A.17), (A.21), and (A.22). It is well known that $\mathbf{e}_1^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{c}_p = 0$ if p is even.

Proof of (b). From the expression of II and $\mathrm{var}(\boldsymbol{\epsilon}_i|\mathcal{H}_n) = \boldsymbol{\Sigma}_i$, it is easy to see that

$$\begin{aligned}\mathrm{var}\{\hat{\theta}(t_0)|\mathcal{H}_n\} &= \mathbf{e}_1^{\mathrm{T}}\,\mathrm{var}\{II|\mathcal{H}_n\}\mathbf{e}_1\\ &= \mathbf{e}_1^{\mathrm{T}}\left[\sum_{i=1}^n \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i\mathbf{T}_i(t_0)\right]^{-1}\mathbf{L}_n\\ &\quad \times \left[\sum_{i=1}^n \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i\mathbf{T}_i(t_0)\right]^{-1}\mathbf{e}_1,\end{aligned} \tag{A.23}$$

where

$$\begin{aligned}\mathbf{L}_n &= \sum_{i=1}^n\left[\left(\mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i - \sum_{j=1}^n \mathbf{T}_j^{\mathrm{T}}(t_0)\mathbf{w}_j\mathbf{X}_j\mathbf{A}_n^{-1}\mathbf{Z}_i^{\mathrm{T}}\right)\right.\\ &\quad \left.\times \boldsymbol{\Sigma}_i\left(\mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i - \sum_{j=1}^n \mathbf{T}_j^{\mathrm{T}}(t_0)\mathbf{w}_j\mathbf{X}_j\mathbf{A}_n^{-1}\mathbf{Z}_i^{\mathrm{T}}\right)^{\mathrm{T}}\right]\\ &= II_1 - II_2 - II_3 + II_4,\end{aligned}$$

with

$$II_1 = \sum_{i=1}^{n} [\mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i\boldsymbol{\Sigma}_i\mathbf{w}_i\mathbf{T}_i(t_0)],$$

$$II_2 = \left[\sum_{i=1}^{n} \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i\mathbf{X}_i\right]\mathbf{A}_n^{-1}\left[\sum_{i=1}^{n} \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{w}_i\mathbf{T}_i(t_0)\right],$$

$$II_3 = II_2^{\mathrm{T}},$$

and

$$II_4 = \left[\sum_{i=1}^{n} \mathbf{T}_i^{\mathrm{T}}(t_0)\mathbf{w}_i\mathbf{X}_i\right]\mathbf{A}_n^{-1}\left[\sum_{i=1}^{n} \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{Z}_i\right]\mathbf{A}_n^{-1}\left[\sum_{i=1}^{n} \mathbf{X}_i^{\mathrm{T}}\mathbf{w}_i\mathbf{T}_i(t_0)\right].$$

By the law of large numbers and condition (C2), it follows that

$$\frac{1}{n}II_2 = O(1) \qquad \text{and} \qquad \frac{1}{n}II_4 = O(1).$$

Similar to (A.21),

$$\frac{1}{n}II_1 = h^{-1}\left\{\sum_{j=1}^{J} \sigma_j^2(t_0)\frac{f_j(t_0)}{v_j^2(t_0)}\right\}\mathbf{H}\mathbf{S}^*\mathbf{H}(1+o(1));$$

thus

$$\mathbf{L}_n = nh^{-1}\left\{\sum_{j=1}^{J} \sigma_j^2 \frac{f_j(t_0)}{v_j^2(t_0)}\right\}\mathbf{H}\mathbf{S}^*\mathbf{H}(1+o(1)). \qquad \text{(A.24)}$$

By (A.21), (A.23), and (A.24), the result (10) follows.

Application of the Cauchy–Schwarz inequality leads to

$$\left[\sum_{j=1}^{J} \sigma_j^2(t_0)\frac{f_j(t_0)}{v_j^2(t_0)}\right]\left[\sum_{j=1}^{J} \frac{f_j(t_0)}{\sigma_j^2(t_0)}\right] \ge \left[\sum_{j=1}^{J} \frac{f_j(t_0)}{v_j(t_0)}\right]^2.$$

The equality holds if and only if $v_j(t_0) = \sigma_j^2(t_0)$ for $j = 1, \dots, J$. Therefore,

$$\frac{\sum_{j=1}^{J} \sigma_j^2(t_0) f_j(t_0)/v_j^2(t_0)}{[\sum_{j=1}^{J} f_j(t_0)/v_j(t_0)]^2}$$

is minimized if and only if $v_j(t_0) = \sigma_j^2(t_0)$ for $j = 1, \dots, J$ with the minimum

$$\frac{1}{\sum_{j=1}^{J} f_j(t_0)/\sigma_j^2(t_0)}.$$

This completes the proof of (b).

[Received September 2003. Revised April 2005.]

REFERENCES

Atkinson, A. C. (1985), *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Oxford, U.K.: Clarendon Press.

Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997), "Generalized Partially Linear Single-Index Models," *Journal of the American Statistical Association*, 92, 477–489.

Chen, H. (1988), "Convergence Rates for Parametric Components in a Partly Linear Model," *The Annals of Statistics*, 1, 136–146.

Chen, K., and Jin, Z. (2005), "Local Polynomial Regression Analysis of Clustered Data," *Biometrika*, 92, 54–79.

de la Peña, V., and Giné, E. (1999), *Decoupling: From Dependence to Independence*, New York: Springer-Verlag.

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

Hastie, T., and Tibshirani, R. (1990), *Generalized Additive Models*, London: Chapman & Hall.

He, X., Zhu, Z., and Fung, W. K. (2002), "Estimation in a Semiparametric Model for Longitudinal Data With Unspecified Dependence Structure," *Biometrika*, 89, 579–590.

Lin, X., and Carroll, R. J. (2001a), "Semiparametric Regression for Clustered Data Using Generalized Estimating Equations," *Journal of the American Statistical Association*, 96, 1045–1056.

——— (2001b), "Semiparametric Regression for Clustered Data," *Biometrika*, 88, 1179–1185.

Severini, T. A., and Staniswalis, J. G. (1994), "Quasi-Likelihood Estimation in Semiparametric Models," *Journal of the American Statistical Association*, 89, 501–511.

Speckman, P. (1988), "Kernel Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society*, Ser. B, 50, 413–436.

Wang, N. (2003), "Marginal Nonparametric Kernel Regression Accounting for Within-Subject Correlation," *Biometrika*, 90, 43–52.

Wang, N., Carroll, R. J., and Lin, X. (2005), "Efficient Semiparametric Marginal Estimation for Longitudinal/Clustered Data," *Journal of the American Statistical Association*, 100, 147–157.

Zhang, D., Lin, X., Raz, J., and Sower, M. F. (1998), "Semiparametric Stochastic Mixed Models for Longitudinal Data," *Journal of the American Statistical Association*, 93, 710–719.