Analysis of Longitudinal Data With Semiparametric Estimation of Covariance Function

Jianqing FAN, Tao HUANG, and Runze LI

Improving efficiency for regression coefficients and predicting trajectories of individuals are two important aspects in the analysis of longitudinal data. Both involve estimation of the covariance function. Yet challenges arise in estimating the covariance function of longitudinal data collected at irregular time points. A class of semiparametric models for the covariance function by that imposes a parametric correlation structure while allowing a nonparametric variance function is proposed. A kernel estimator for estimating the nonparametric variance function is developed. Two methods for estimating parameters in the correlation structure—a quasi-likelihood approach and a minimum generalized variance method—are proposed. A semiparametric varying coefficient partially linear model for longitudinal data is introduced, and an estimation procedure for model coefficients using a profile weighted least squares approach is proposed. Sampling properties of the proposed estimation procedures are studied, and asymptotic normality of the resulting estimators is established. Finite-sample performance of the proposed procedures is assessed by Monte Carlo simulation studies. The proposed methodology is illustrated with an analysis of a real data example.

KEY WORDS: Kernel regression; Local linear regression; Profile weighted least squares; Semiparametric varying coefficient model.

1. INTRODUCTION

Estimating covariance functions is an important issue in the analysis of longitudinal data. It features prominently in forecasting the trajectory of an individual response over time and is closely related to improving the efficiency of estimated regression coefficients. Challenges arise in estimating the covariance function due to the fact that longitudinal data are frequently collected at irregular and possibly subject-specific time points. Interest in these kinds of challenges has surged in the recent literature. Wu and Pourahmadi (2003) proposed nonparametric estimation of large covariance matrices using a two-step estimation procedure (Fan and Zhang 2000), but their method can deal with only balanced or nearly balanced longitudinal data. Recently, Huang, Liu, Pourahmadi, and Liu (2006) introduced a penalized likelihood method for estimating a covariance matrix when the design is balanced and Yao, Müller, and Wang (2005a, b) approached the problem from the standpoint of functional data analysis.

In this article we consider a semiparametric varying-coefficient partially linear model,

$$y(t) = \mathbf{x}(t)^T \boldsymbol{\alpha}(t) + \mathbf{z}(t)^T \boldsymbol{\beta} + \varepsilon(t), \tag{1}$$

where $\boldsymbol{\alpha}(t)$ comprises p unknown smooth functions, $\boldsymbol{\beta}$ is a q-dimensional unknown parameter vector, and $E\{\varepsilon(t)|\mathbf{x}(t), \mathbf{z}(t)\} = 0$. Nonparametric models for longitudinal data (Lin and Carroll 2000; Wang 2003) can be viewed as special cases of model (1). Moreover, model (1) is a useful extension of the partially linear model, systematically studied by Härdle, Liang, and Gao (2000), and of the time-varying coefficient model (Hastie and Tibshirani 1993). It has been considered by Zhang, Lee, and Song (2002), Xia, Zhang, and Tong (2004), and Fan and Huang (2005) in the case of iid observations and by Martinussen and Scheike (1999) and Sun and Wu (2005) for longitudinal data. It is a natural extension of the models studied by Lin and Carroll (2001) (with identity link), He, Zhu, and Fung (2002), He, Fung, and Zhu (2005), Wang, Carroll, and Lin (2005), and Huang and Zhang (2004).

We focus on parsimonious modeling of the covariance function of the random error process $\varepsilon(t)$ for the analysis of longitudinal data, when observations are collected at irregular and possibly subject-specific time points. We approach this by assuming that $\text{var}\{\varepsilon(t)|\mathbf{x}(t), \mathbf{z}(t)\} = \sigma^2(t)$, which is a nonparametric smoothing function, but the correlation function between $\varepsilon(s)$ and $\varepsilon(t)$ has a parametric form, $\text{corr}\{\varepsilon(s), \varepsilon(t)\} = \rho(t, s, \boldsymbol{\theta})$, where $\rho(s, t, \boldsymbol{\theta})$ is a positive definite function of s and t, and $\boldsymbol{\theta}$ is an unknown parameter vector.

The covariance function is fitted by a semiparametric model, which allows the random error process $\varepsilon(t)$ to be nonstationary as its variance function $\sigma^2(t)$ may be time-dependent. Compared with a fully nonparametric fit, defined in (16) in Section 6, to the correlation function, our semiparametric model guarantees positive definiteness for the resulting estimate; it retains the flexibility of nonparametric modeling and parsimony and the ease of interpretation of parametric modeling. To improve the efficiency of the regression coefficient, one typically takes the weight matrix in the weighted least squares method to be the inverse of estimated covariance matrix. Thus the requirement on positive definiteness becomes necessary. Our semiparametric model allows a data analyst to easily incorporate prior information about the correlation structure. It can be used to improve the estimation efficiency of $\boldsymbol{\beta}$. For example, letting $\rho_0(s, t)$ be a working correlation function (e.g., working independence) and $\rho(s, t, \boldsymbol{\theta})$ be a family of correlation functions (e.g., an AR or ARMA correlation structure) that contains ρ_0, our method allows us to choose an appropriate $\boldsymbol{\theta}$ to improve the efficiency of the estimator of $\boldsymbol{\beta}$. Obviously, to improve the efficiency, the family of correlation functions $\{\rho(s, t, \boldsymbol{\theta})\}$ need not contain the true correlation structure.

We also introduce an estimation procedure for the variance function and propose two approaches to estimating the unknown vector $\boldsymbol{\theta}$, motivated from two different principles. We

Jianqing Fan is Frederick Moore Professor of Finance, Department of Operation Research and Financial Engineering, Princeton University, Princeton, NJ 08544 (E-mail: *jqfan@Princeton.edu*). Tao Huang is Assistant Professor, Department of Statistics, University of Virginia, Charlottesville, VA 22904 (E-mail: *th8e@Virginia.edu*). Runze Li is Associate Professor, Department of Statistics and the Methodology Center, Pennsylvania State University, University Park, PA 16802 (E-mail: *rli@stat.psu.edu*). Fan's research was supported in part by National Science Foundation (NSF) grant DMS-03-54223 and National Institutes of Health grant R01-GM072611. Li's research was supported by NSF grant DMS-03-48869 and National Institute on Drug Abuse grant P50 DA10075. The authors thank the associate editor and the referees for their constructive comments that substantially improved an earlier draft, and the MACS study for the data used in Section 5.3.

Journal of the American Statistical Association
June 2007, Vol. 102, No. 478, Theory and Methods
DOI 10.1198/016214507000000095

also propose an estimation procedure for the regression function $\boldsymbol{\alpha}(t)$ and coefficient $\boldsymbol{\beta}$ using the profile least squares. Asymptotic properties of the proposed estimators are investigated, and finite-sample performance is assessed through Monte Carlo simulation studies. A real data example is used to illustrate the proposed methodology.

The article is organized as follows. Estimation procedures for variance function and unknown parameters in the correlation matrix are proposed in Section 2. An efficient estimation procedure for $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}$ based on the profile least squares techniques is described in Section 3. Sampling properties of the proposed procedures are presented in Section 4, and simulation studies and real data analysis are given in Section 5. All technical proofs are relegated to the Appendix.

2. ESTIMATION OF COVARIANCE FUNCTION

Suppose that a random sample from model (1) consists of n subjects. For the ith subject, $i = 1, \dots, n$, the response variable $y_i(t)$ and the covariates $\{\mathbf{x}_i(t), \mathbf{z}_i(t)\}$ are collected at time points $t = t_{ij}, j = 1, \dots, J_i$, where J_i is the total number of observations for the ith subject. Denote

$$r_{ij} \equiv r_{ij}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = y_i(t_{ij}) - \mathbf{x}_i(t_{ij})^T \boldsymbol{\alpha}(t_{ij}) - \mathbf{z}_i(t_{ij})^T \boldsymbol{\beta}$$

and $\mathbf{r}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (r_{i1}, \dots, r_{iJ_i})^T$. Here we adopt the notation $r_{ij}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ to emphasize the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, although for true values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, $r_{ij}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \varepsilon_i(t_{ij})$.

To motivate the proposed estimation procedures that follow, assume for the moment that $\boldsymbol{\varepsilon}_i$ is normally distributed with mean 0 and covariance matrix Σ_i. Then the logarithm of the likelihood function for $\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2$, and $\boldsymbol{\theta}$ is

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) = -\frac{1}{2}\sum_{i=1}^{n} \log|\Sigma_i| - \frac{1}{2}\sum_{i=1}^{n} \mathbf{r}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})^T \Sigma_i^{-1} \mathbf{r}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) \quad (2)$$

after dropping a constant. Maximizing the log-likelihood function yields a maximum likelihood estimate (MLE) for the unknown parameters. The parameters can be estimated by iterating between estimation of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and estimation of $(\sigma^2, \boldsymbol{\theta})$. [We discuss the estimation procedure of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ for model (1) in detail in the next section.] Thus we may substitute their estimates into $r_{ij}(\boldsymbol{\alpha}, \boldsymbol{\beta})$, and $r_{ij}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ is computable and is denoted by $\hat{r}_{ij}$ for simplicity.

2.1 Estimation of Variance Function

We first propose an estimation procedure for $\sigma^2(t)$. Note that

$$\sigma^2(t_{ij}) = E\{\varepsilon^2(t) | t = t_{ij}\}.$$

A natural estimator for $\sigma^2(t)$ is the kernel estimator

$$\hat{\sigma}^2(t) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{J_i} \hat{r}_{ij}^2 K_{h_1}(t - t_{ij})}{\sum_{i=1}^{n}\sum_{j=1}^{J_i} K_{h_1}(t - t_{ij})},$$

where $K_{h_1}(x) = h_1{}^{-1}K(x/h_1)$, $K(x)$ is a kernel density function, and h_1 is a smoothing parameter. Note that locally around a time point, few subjects contribute more than one data point to the estimation of $\sigma^2(t)$. Thus the estimator should behave locally as if the data were independent. Ruppert, Wand, Holst, and Hössjer (1997) studied local polynomial estimation of the variance function when observations are independently taken from the canonical nonparametric regression model, $Y = m(X) + \varepsilon$ with $E(\varepsilon|X) = 0$ and $\text{var}(\varepsilon|X) = \sigma^2(X)$. Fan and Yao (1998) further showed that the local linear fit of variance function performs as well as the ideal estimator, which is a local linear fit to the true squared residuals $\{(Y_i - m(X_i))^2\}$, allowing data to be taken from a stationary mixing process. A similar result was obtained by Müller and Stadtmüller (1993). The consistency and asymptotic behavior of $\hat{\sigma}^2(t)$ are studied in Theorem 2(b) in Section 4, from which we may choose an optimal bandwidth for $\hat{\sigma}^2(t)$ using various existing bandwidth selectors for independent data (see, e.g., Ruppert, Sheather, and Wand 1995).

2.2 Estimation of $\boldsymbol{\theta}$

Decompose the covariance matrix Σ_i into variance–correlation form, that is,

$$\Sigma_i = V_i C_i(\boldsymbol{\theta}) V_i,$$

where $V_i = \text{diag}\{\sigma(t_{i1}), \dots, \sigma(t_{iJ_i})\}$ and $C_i(\boldsymbol{\theta})$ is the correlation matrix of $\boldsymbol{\varepsilon}_i$ with (k, l) element equaling $\rho(t_{ik}, t_{il}, \boldsymbol{\theta})$. To construct an estimator for $\boldsymbol{\theta}$, we maximize $\ell(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. In other words,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left(-\frac{1}{2}\sum_{i=1}^{n} \left\{ \log|C_i(\boldsymbol{\theta})| + \hat{\mathbf{r}}_i^T \hat{V}_i^{-1} C_i^{-1}(\boldsymbol{\theta}) \hat{V}_i^{-1} \hat{\mathbf{r}}_i \right\} \right), \quad (3)$$

where $\hat{V}_i = \text{diag}\{\hat{\sigma}(t_{i1}), \dots, \hat{\sigma}(t_{iJ_i})\}$ and $\hat{\mathbf{r}}_i = (\hat{r}_{i1}, \dots, \hat{r}_{iJ_i})^T$. The estimator in (3) is referred to as a quasi-likelihood (QL) estimator.

Optimizing QL may provide a good estimate for $\boldsymbol{\theta}$ when the correlation structure is correctly specified, but when it is misspecified, the QL might not be the best criterion to optimize. For example, we may be interested in improving the efficiency for $\boldsymbol{\beta}$, treating $\boldsymbol{\alpha}$, σ^2, and $\boldsymbol{\theta}$ as nuisance parameters. In such a case, we are interested in choosing $\boldsymbol{\theta}$ to minimize the estimated variance of $\hat{\boldsymbol{\beta}}$. For example, for a given working correlation function $\rho_0(s, t)$ (e.g., working independence), we can embed this matrix into a family of parametric models $\rho(s, t, \boldsymbol{\theta})$ [e.g., autocovariance function of the ARMA(1, 1) model]. Even though $\rho(s, t, \boldsymbol{\theta})$ might not be the true correlation function, we can always find a $\boldsymbol{\theta}$ to improve the efficiency of $\boldsymbol{\beta}$. More generally, suppose that the current working correlation function is $\rho_0(s, t; \boldsymbol{\theta}_0)$. Let $\rho_1(s, t), \dots, \rho_m(s, t)$ be given family of correlation functions. We can always embed the current working correlation function $\rho_0(s, t; \boldsymbol{\theta}_0)$ into the family of the correlation functions,

$$\rho(s, t; \boldsymbol{\theta}) = \tau_0 \rho_0(s, t; \boldsymbol{\theta}_0) + \tau_1 \rho_1(s, t) + \cdots + \tau_m \rho_m(s, t),$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_0, \tau_0, \dots, \tau_m)$ and $\tau_0 + \cdots + \tau_m = 1$ with all $\tau_i \geq 0$. Thus, by optimizing the parameters $\boldsymbol{\theta}_0, \tau_0, \dots, \tau_m$, the efficiency of the resulting estimator $\hat{\boldsymbol{\beta}}$ can be improved.

To fix the idea, let $\Gamma(\hat{\sigma}^2, \boldsymbol{\theta})$ be the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$ derived in (11) in Section 3 for a given working correlation function $\rho(s, t, \boldsymbol{\theta})$. Define the generalized variance of $\hat{\boldsymbol{\beta}}$ as the determinant of $\Gamma(\hat{\sigma}^2, \boldsymbol{\theta})$. Minimizing the volume of the confidence ellipsoid of $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \Gamma^{-1}(\hat{\sigma}^2, \boldsymbol{\theta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) < c$ for any positive constant c is equivalent to minimizing the generalized variance. Thus we may choose $\boldsymbol{\theta}$ to minimize the volume of the confidence ellipsoid,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} |\Gamma(\hat{\sigma}^2, \boldsymbol{\theta})|. \tag{4}$$

We refer to this approach as the minimum generalized variance (MGV) method.

3. ESTIMATION OF REGRESSION COEFFICIENTS

As mentioned in Section 2, the estimation of σ^2 and $\boldsymbol{\theta}$ depends on the estimation of $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}$. On the other hand, improving the efficiency of the estimate for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ relies on the estimation of σ^2 and $\boldsymbol{\theta}$. In practice, therefore, estimation must be done in steps. The initial estimates of $(\boldsymbol{\alpha}(t), \boldsymbol{\beta})$ are constructed by ignoring within subject correlation. With this initial estimate, we can further estimate $\sigma^2(t)$ and $\boldsymbol{\theta}$. Finally, we can now estimate $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}$ more efficiently by using the estimate of $\sigma^2(t)$ and $\boldsymbol{\theta}$. In this section we propose efficient estimates for $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}$ using the profile least squares techniques.

For a given $\boldsymbol{\beta}$, let $y^*(t) = y(t) - \mathbf{z}(t)^T\boldsymbol{\beta}$. Then model (1) can be written as

$$y^*(t) = \mathbf{x}(t)^T\boldsymbol{\alpha}(t) + \varepsilon. \tag{5}$$

This is a varying coefficient model, studied by Fan and Zhang (2000) in the context of longitudinal data and by Hastie and Tibshirani (1993) for the case of iid observations. Thus $\boldsymbol{\alpha}(t)$ can be easily estimated using any linear smoother. Here we use local linear regression (Fan and Gijbels 1996). For any t in a neighborhood of t_0, it follows from Taylor's expansion that

$$\begin{aligned}\alpha_l(t) &\approx \alpha_l(t_0) + \alpha_l'(t_0)(t - t_0)\\ &\equiv a_l + b_l(t - t_0) \quad \text{for } l = 1, \ldots, q.\end{aligned}$$

Let $K(\cdot)$ be a kernel function and h be a bandwidth. Thus we can find local parameters $(a_1, \ldots, a_q, b_1, \ldots, b_q)$ that minimize

$$\sum_{i=1}^{n}\sum_{j=1}^{J_i}\left[y_i^*(t_{ij}) - \sum_{l=1}^{q}\{a_l + b_l(t_{ij} - t_0)\}x_{il}(t_{ij})\right]^2 \times K_h(t_{ij} - t_0), \tag{6}$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$. The local linear estimate for $\boldsymbol{\alpha}(t_0)$ is then simply $\hat{\boldsymbol{\alpha}}(t_0, \hat{\boldsymbol{\beta}}) = (a_1, \ldots, a_q)^T$. Note that because the data are localized in time, the covariance structure does not greatly affect the local linear estimator.

The profile least squares estimator of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ has a closed form using the following matrix notation. Let $\mathbf{y}_i = (y_i(t_{i1}), \ldots, y_i(t_{iJ_i}))^T$, $\mathbf{X}_i = (\mathbf{x}_i(t_{i1}), \ldots, \mathbf{x}_i(t_{iJ_i}))^T$, $\mathbf{Z}_i = (\mathbf{z}_i(t_{i1}), \ldots, \mathbf{z}_i(t_{iJ_i}))^T$, and $\mathbf{m}_i = (\mathbf{x}_i(t_{i1})^T\boldsymbol{\alpha}(t_{i1}), \ldots, \mathbf{x}_i(t_{iJ_i})^T\boldsymbol{\alpha}(t_{iJ_i}))^T$. Write $\mathbf{y} = (\mathbf{y}_1^T, \ldots, \mathbf{y}_n^T)^T$, $\mathbf{X} = (\mathbf{X}_1^T, \ldots, \mathbf{X}_n^T)^T$, $\mathbf{Z} = (\mathbf{Z}_1^T, \ldots, \mathbf{Z}_n^T)^T$, and $\mathbf{m} = (\mathbf{m}_1^T, \ldots, \mathbf{m}_n^T)^T$. Then, model (5) can be written as

$$\mathbf{y} - \mathbf{Z}\boldsymbol{\beta} = \mathbf{m} + \boldsymbol{\varepsilon}, \tag{7}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1(t_{11}), \ldots, \varepsilon_n(t_{nJ_n}))^T$. It is known that the local linear regression results in a linear estimate in $y^*(t_{ij})$ for $\boldsymbol{\alpha}(\cdot)$ (Fan and Gijbels 1996). Thus the estimate of $\boldsymbol{\alpha}(\cdot)$ is linear in $\mathbf{y} - \mathbf{Z}\boldsymbol{\beta}$, and the estimate of $\mathbf{m}$ is of the form $\hat{\mathbf{m}} = \mathbf{S}(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})$. The matrix $\mathbf{S}$, usually called a smoothing matrix of the local linear smoother, depends only on the observations $\{t_{ij}, \mathbf{x}_i(t_{ij}), j = 1, \ldots, J_i, i = 1, \ldots, n\}$. Substituting $\hat{\mathbf{m}}$ into (7) results in the synthetic linear model

$$(I - \mathbf{S})\mathbf{y} = (I - \mathbf{S})\mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{8}$$

where I is the identity matrix of order $n^* = \sum_{i=1}^n J_i$.

To improve the efficiency for estimating $\boldsymbol{\beta}$, we minimize the weighted least squares,

$$(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})^T(\mathbf{I} - \mathbf{S})^T\mathbf{W}(\mathbf{I} - \mathbf{S})(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta}), \tag{9}$$

where $\mathbf{W}$ is a weight matrix, called a working covariance matrix. As usual, misspecification of the working covariance matrix does not affect the consistency of the resulting estimate, but does affect the efficiency. The weighted least squares estimator for $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \{\mathbf{Z}^T(I - \mathbf{S})^T\mathbf{W}(I - \mathbf{S})\mathbf{Z}\}^{-1}\mathbf{Z}^T(I - \mathbf{S})^T\mathbf{W}(I - \mathbf{S})\mathbf{y}. \tag{10}$$

This estimator is called the *profile weighted least squares estimator*. The profile least squares estimator for the nonparametric component is simply $\hat{\boldsymbol{\alpha}}(\cdot; \hat{\boldsymbol{\beta}})$. Using (8), it follows that when the weight matrix does not depend on $\mathbf{y}$,

$$\operatorname{cov}\{\hat{\boldsymbol{\beta}} | t_{ij}, \mathbf{x}_i(t_{ij}), \mathbf{z}_i(t_{ij})\} = \mathbf{D}^{-1}\mathbf{V}\mathbf{D}^{-1} \hat{=} \Gamma(\sigma^2, \boldsymbol{\theta}), \tag{11}$$

where $\mathbf{D} = \mathbf{Z}^T(I - \mathbf{S})^T\mathbf{W}(I - \mathbf{S})\mathbf{Z}$ and $\mathbf{V} = \operatorname{cov}\{\mathbf{Z}^T(I - \mathbf{S})^T\mathbf{W}\boldsymbol{\varepsilon}\}$. In practice, $\hat{\Gamma}(\sigma^2, \boldsymbol{\theta})$ is estimated using a sandwich formula by taking $\hat{\mathbf{V}} = \mathbf{Z}^T(I - \mathbf{S})^T\mathbf{W}\mathbf{R}\mathbf{W}^T(I - \mathbf{S})\mathbf{Z}$, where $\mathbf{R} = \operatorname{diag}\{\mathbf{r}_1\mathbf{r}_1^T, \ldots, \mathbf{r}_n\mathbf{r}_n^T\}$ with $\mathbf{r}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i$. Speckman (1988) derived a partial residual estimator of $\boldsymbol{\beta}$ for partially linear models with independent and identically distributed data; the form of this estimator is the same as that in (9), with $\mathbf{W}$ set to be an identity matrix. However, the partial residual approach is difficult to implement for model (1).

4. SAMPLING PROPERTIES

In this section we investigate sampling properties of the profile weighted least squares estimator. The proposed estimation procedures are applicable for various formulations for collecting longitudinal data. Here we consider the collected data as a random sample from the population process $\{y(t), \mathbf{x}(t), \mathbf{z}(t)\}$, $t \in [0, T]$. To facilitate the presentation, we assume that J_i, $i = 1, \ldots, n$, are independent and identically distributed with $0 < E(J_i) < \infty$, and for a given J_i, $t_{ij}, j = 1, \ldots, J_i$, are independent and identically distributed according to a density $f(t)$. Furthermore, suppose that the weight matrix $\mathbf{W}$ in (9) is block diagonal, that is, $\mathbf{W} = \operatorname{diag}\{\mathbf{W}_1, \ldots, \mathbf{W}_n\}$, where $\mathbf{W}_i$ is a $J_i \times J_i$ matrix. Moreover, assume that the (u, v)-element of $\mathbf{W}_i$ is set to be $w(t_{iu}, t_{iv})$ for a bivariate positive function $w(\cdot, \cdot)$. When the weight function $w(\cdot, \cdot)$ is data-dependent, assume that it tends to a positive definite function in probability. Thus for simplicity, assume that $w(\cdot, \cdot)$ is deterministic.

Let $\mathbf{G}(t) = E\mathbf{x}(t)\mathbf{x}^T(t)$, $\Psi(t) = E\mathbf{x}(t)\mathbf{z}^T(t)$, and write

$$\tilde{\mathbf{X}}_i = \left(\Psi(t_{i1})\mathbf{G}^{-1}(t_{i1})\mathbf{x}_i(t_{i1}), \ldots, \Psi(t_{iJ_i})\mathbf{G}^{-1}(t_{iJ_i})\mathbf{x}_i(t_{iJ_i})\right)^T.$$

Set

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^{n}\{\mathbf{Z}_i - \tilde{\mathbf{X}}_i\}^T\mathbf{W}_i\{\mathbf{Z}_i - \tilde{\mathbf{X}}_i\}$$

and

$$\xi_n = \frac{1}{n}\sum_{i=1}^{n}\{\mathbf{Z}_i - \tilde{\mathbf{X}}_i\}^T \mathbf{W}_i \boldsymbol{\varepsilon}_i,$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_i(t_{i1}), \ldots, \varepsilon_i(t_{iJ_i}))^T$. Let

$$\mathbf{A} = E\{(\mathbf{Z}_1 - \tilde{\mathbf{X}}_1)^T \mathbf{W}_1 (\mathbf{Z}_1 - \tilde{\mathbf{X}}_1)\}$$

and

$$\mathbf{B} = E\{(\mathbf{Z}_1 - \tilde{\mathbf{X}}_1)^T \mathbf{W}_1 \boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_1^T \mathbf{W}_1 (\mathbf{Z}_1 - \tilde{\mathbf{X}}_1)\}.$$

Let $\boldsymbol{\alpha}_0(t)$ and $\boldsymbol{\beta}_0$ denote the true values of $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}$.

Theorem 1. Under the regularity conditions (1)–(5) in the Appendix, if the matrices $\mathbf{A}$ and $\mathbf{B}$ exist, and if $\mathbf{A}$ is positive definite, then, as $n \to \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \sqrt{n}\Sigma_n^{-1}\xi_n + o_P(1) \xrightarrow{\mathcal{L}} \mathrm{N}(0, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}),$$

where n is the number of subjects.

When $\mathbf{W}_i$ is taken to be the inverse of the conditional variance–covariance matrix of $\boldsymbol{\varepsilon}_i$ given $\mathbf{x}_i(t_{ij})$ and $\mathbf{z}_i(t_{ij})$ for $j = 1, \ldots, J_i$, then $\mathbf{A} = \mathbf{B}$. In this case

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{L}} \mathrm{N}(0, \mathbf{B}_0^{-1}),$$

where $\mathbf{B}_0 = E\{(\mathbf{Z}_1 - \tilde{\mathbf{X}}_1)^T \operatorname{cov}^{-1}(\boldsymbol{\varepsilon}_1 | \mathbf{X}_1, \mathbf{Z}_1)(\mathbf{Z}_1 - \tilde{\mathbf{X}}_1)\}$. We show in the Appendix that for any weight matrix $\mathbf{W}_i$,

$$\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1} - \mathbf{B}_0^{-1} \geq 0, \tag{12}$$

where $D \geq 0$ means that the matrix D is nonnegative definite. Thus the most efficient estimator for $\boldsymbol{\beta}$ among the profile weighted least squares estimates given in (10) is the one that uses the inverse of the true variance–covariance matrix of $\boldsymbol{\varepsilon}_i$ as the weight matrix $\mathbf{W}_i$.

One could also use a working independence correlation structure, that is, let $\mathbf{W}$ be a diagonal matrix. Under the conditions of Theorem 1, the resulting estimate of $\boldsymbol{\beta}$ is still root–n–consistent.

Let $\mu_i = \int u^i K(u)\,du$ and $\nu_i = \int u^i K^2(u)\,du$. For a vector of functions $\boldsymbol{\alpha}(u)$ of u, denote $\dot{\boldsymbol{\alpha}}(u) = d\boldsymbol{\alpha}(u)/du$ and $\ddot{\boldsymbol{\alpha}}(u) = d^2\boldsymbol{\alpha}(u)/du^2$, which are the componentwise derivatives. The following theorem presents the asymptotic normality for $\hat{\boldsymbol{\alpha}}(t)$ and $\hat{\sigma}^2(t)$; its proof was given the earlier version of this article (Fan, Huang, and Li 2005).

Theorem 2. Suppose that conditions of Theorem 1 hold. Then the following results hold:

(a) If $nh^5 = O(1)$ as $n \to \infty$, then

$$\sqrt{nh}\left(\hat{\boldsymbol{\alpha}}(t) - \boldsymbol{\alpha}(t) - \frac{1}{2}\mu_2 h^2 \ddot{\boldsymbol{\alpha}}(t)\right) \xrightarrow{\mathcal{L}} \mathrm{N}\left(0, \frac{\nu_0}{f(t)E(J_1)}\sigma^2(t)\Gamma^{-1}(t)\right).$$

(b) Under conditions 5 and 6 in the Appendix, if $c < nh_1^5 < C$, and $c < h/h_1 < C$ for some positive constants c and C, then, as $n \to \infty$,

$$\sqrt{nh_1}(\hat{\sigma}^2(t) - \sigma^2(t) - b(t)) \xrightarrow{\mathcal{L}} \mathrm{N}(0, v(t)),$$

where the bias is

$$b(t) = \frac{h_1^2}{2}\left\{\ddot{\sigma}^2(t) + \frac{2\dot{\sigma}^2(t)f'(t)}{f(t)}\right\}\mu_2$$

and the variance is

$$v(t) = \frac{\operatorname{var}\{\varepsilon^2(t)\}\nu_0}{f(t)E(J_1)}.$$

Because the parametric convergence rate of $\hat{\boldsymbol{\beta}}$ is faster than the nonparametric convergence rate of $\hat{\boldsymbol{\alpha}}(t)$, the asymptotic bias and variance have forms similar to those of the varying-coefficient model (Cai, Fan, and Li 2000). The choice of the weight matrix $\mathbf{W}$ determines the efficiency of $\hat{\boldsymbol{\beta}}$, but it does not affect the asymptotic bias and variance of $\boldsymbol{\alpha}(t)$.

From Theorem 2(b), the asymptotic bias and variance do not depend on the choice of the weight matrix $\mathbf{W}$. Therefore, one may use the residuals obtained using the working independence correlation matrix to estimate $\sigma^2(t)$. This is consistent with our empirical findings from the simulation studies. Therefore, in next section, $\sigma^2(t)$ is estimated using residuals obtained under working independence. Theorem 2(b) implies that we may choose a bandwidth by modifying one of existing bandwidth selectors used for independent data.

5. NUMERICAL COMPARISON AND APPLICATION

In this section we investigate finite-sample properties of the estimators proposed in Sections 2 and 3 through Monte Carlo simulation. All simulation studies were conducted using Matlab code. We examined the finite-sample performance and numerical comparisons for the proposed estimate $\hat{\sigma}^2(t)$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\alpha}}(t)$ in the earlier version of this article. (See Fan et al. 2005 for details.) To save space, in this section we focus on the inference on $\boldsymbol{\beta}$.

5.1 Simulation Study

We generate 1,000 datasets, each consisting of $n = 50$ subjects, from the following model:

$$y(t) = \mathbf{x}(t)^T\boldsymbol{\alpha}(t) + \mathbf{z}(t)^T\boldsymbol{\beta} + \varepsilon(t). \tag{13}$$

In practice, observation times are usually scheduled but may be randomly missed. Thus we generate the observation times in the following way. Each individual has a set of "scheduled" time points, $\{0, 1, 2, \ldots, 12\}$, and each scheduled time, except time 0, has a 20% probability of being skipped. The actual observation time is a random perturbation of a scheduled time: a uniform $[0, 1]$ random variable is added to a nonskipped scheduled time. This results in different observed time points t_{ij} per subject.

In our simulation, the random error process $\varepsilon(t)$ in (13) is taken to be a Gaussian process with mean 0, variance function

$$\sigma^2(t) = .5\exp(t/12),$$

and ARMA(1, 1) correlation structure

$$\operatorname{corr}(\varepsilon(s), \varepsilon(t)) = \gamma\rho^{|t-s|}$$

for $s \neq t$. We consider three pairs of (γ, ρ)—(.85, .9), (.85, .6), and (.85, .3)—which correspond to strongly, moderately, and weakly correlated errors.

We let the coefficients of both $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}$ be two-dimensional in our simulation, and further set $x_1(t) \equiv 1$ to include an intercept term. We generate the covariates in the following way: For a given t, $(x_2(t), z_1(t))^T$ follows a bivariate normal distribution with mean 0, variance 1, and correlation .5, and $z_2(t)$ is a Bernoulli-distributed random variable with success probability .5 and independent of $x_2(t)$ and $z_1(t)$. In this simulation we set $\boldsymbol{\beta} = (1, 2)^T$,

$$\alpha_1(t) = \sqrt{t/12}, \qquad \text{and} \qquad \alpha_2(t) = \sin(2\pi t/12).$$

Presumably we can gain some efficiency by incorporating the correlation structure, and it is of interest to study the size of gain. We consider the case in which the working correlation structure is taken to be the true one, which is an ARMA(1, 1) correlation structure. For comparison, we also estimate $\boldsymbol{\beta}$ using a working independence correlation structure and the true correlation structure in which the parameter $(\gamma, \rho)^T$ is set to be the true value. The profile weighted least squares estimate using the true correlation is shown to be the most efficient estimate among the profile weighted least squares estimates and serves as a benchmark, whereas the working independence correlation structure is supposed to be commonly used in practice.

Table 1 summarizes of the results over 1,000 simulations. In the table, "bias" represents the sample average over 1,000 estimates subtracting the true value of $\boldsymbol{\beta}$, "SD" represents the sample standard deviation over 1,000 estimates. "Median" represents the median of the 1,000 estimates subtracting the true value, and "MAD" represents the median absolute deviation of the 1,000 estimates divided by a factor of .6745. From Table 1, both quasi-likelihood (QL) and minimum generalized variance (MGV) approaches yield estimates for $\boldsymbol{\beta}$ as good as those obtained using the true correlation function, and is much better than the estimate using working independence correlation structure. The relative efficiency [MAD(Independence)/MAD(QL)] is about 3 for highly correlated random error, 2 for moderately correlated error, and 1.3 for weakly correlated error.

The simulation results also indicate that the MGV method is more stable and robust than the QL method. This is demonstrated in the case of weakly correlated random error, in which the estimates apparently were quite bad (i.e., the SD is much higher than the MAD) for a few realizations. Note that the object function to optimize in (3) may not be a concave function of $\boldsymbol{\theta}$. Thus the numerical algorithm may not converge when it stops. This may yield a bad estimate for $\boldsymbol{\beta}$ and contribute to the issues of the robustness of the algorithm. In addition, the QL criterion is similar to the least squares criterion and hence is not very robust. On the other hand, the MGV method, aimed directly to minimize the precision of estimated standard errors, does not allow estimates to have large standard errors.

We next study the impact of misspecification of correlation structure, by comparing the performance of $\hat{\boldsymbol{\beta}}$ using independent and AR(1) working correlation structures when the true correlation structure is ARMA(1, 1). The top part of Table 2 summarizes the simulation results. From Table 2, we can see that the AR(1) working correlation structure produces much more efficient estimates than the working independence correlation structure. For example, the relative efficiency for highly correlated random error is about $(30.066/19.975)^2 \approx 2.3$. Thus, even when the true correlation structure is unavailable, choosing a structure close to the truth is still quite desirable.

In practice, we could try several values for ρ and choose the best one using the QL or MGV method rather than an optimization algorithm. We call such a search a *rough grid point* search. We next examine how such search works in practice, using the points $\{.05, .1, .25, .5, .75, .9, .95\}$ for ρ. The bottom part of Table 2 presents the simulation results. Comparing the bottom and top parts of Table 2 shows that the performance of the resulting estimates using the rough grid point search is very close to that using an optimization algorithm.

Now we test the accuracy of the proposed standard error formula (11). Table 3 depicts the simulation results for the case where $(\gamma, \rho) = (.85, .9)$. Results for other cases are similar. In Table 3, "SD" represents for the sample standard deviation of 1,000 estimates of $\boldsymbol{\beta}$ and can be viewed as the true standard deviation of the resulting estimate. "SE" represents for the sample average of 1,000 estimated standard errors using formula (11), and "Std" represents the standard deviation of these 1,000 standard errors. Table 3 demonstrates that the standard error for-

*Table 1. Performance of $\hat{\boldsymbol{\beta}}$**

	$\hat{\beta}_1$				$\hat{\beta}_2$			
Method	*SD*	*Bias*	*MAD*	*Median*	*SD*	*Bias*	*MAD*	*Median*
$(\gamma, \rho) = (.85, .9)$								
Independence	47.780	−1.9730	44.575	−1.2802	82.488	−1.7276	79.580	−2.7890
True	25.061	−1.2565	25.905	−.7676	45.003	.1211	45.543	−.1568
QL	25.156	−1.2545	25.536	−.7709	44.932	.1749	44.654	−.6489
MGV	25.205	−1.2040	25.575	−.9126	45.585	.2663	45.033	−.5308
$(\gamma, \rho) = (.85, .6)$								
Independence	47.499	−2.6415	49.465	−.8980	82.094	−1.1161	82.553	−3.0444
True	34.308	−1.6807	34.569	−1.5081	62.596	−.2047	61.871	−.3016
QL	46.365	−.2651	34.807	−1.2672	62.650	−.0023	62.485	−.3322
MGV	34.634	−1.3411	35.450	−.5676	64.393	−.2691	61.090	−1.8051
$(\gamma, \rho) = (.85, .3)$								
Independence	46.991	−2.8990	47.457	−1.6817	81.798	−1.0896	83.991	−1.2721
True	40.123	−1.9687	40.184	−2.1143	73.031	−.5122	73.278	.1861
QL	95.506	−6.7632	41.841	−1.9187	288.389	−5.7357	77.514	.1459
MGV	40.389	−1.6740	40.685	−1.4153	74.798	−.5055	73.465	.1435

*Values in the columns of SD, bias, MAD, and median are multiplied by a factor of 1,000.

Table 2. Impacts of Misspecification of Correlation on $\hat{\boldsymbol{\beta}}$[*]

	$\hat{\beta}_1$				$\hat{\beta}_2$			
Method	*SD*	*Bias*	*MAD*	*Median*	*SD*	*Bias*	*MAD*	*Median*
Optimization algorithm search								
$(\gamma, \rho) = (.85, .9)$								
Independence	47.7800	−1.9730	44.5759	−1.2802	82.4880	−1.7276	79.5815	−2.7890
QL	31.8570	−.4859	29.6149	−.0837	60.8860	−.0684	54.6946	.2275
MGV	33.1210	−.5275	31.8003	.0535	63.4840	.3800	58.0557	.7224
$(\gamma, \rho) = (.85, .6)$								
Independence	47.4990	−2.6415	49.4655	−.8980	82.0940	−1.1161	82.5541	−3.0444
QL	37.0470	−1.0667	36.2184	−.8100	68.9890	−.0925	61.8111	−1.6883
MGV	37.9660	−1.1648	36.9805	−1.0777	71.3970	.1137	65.4138	−2.2338
$(\gamma, \rho) = (.85, .3)$								
Independence	46.9910	−2.8990	47.4580	−1.6817	81.7980	−1.0896	83.9923	−1.2721
QL	41.0240	−1.6139	40.7671	−1.0700	74.8700	−.3320	73.7801	.0931
MGV	42.4130	−1.6264	42.2526	−.7556	79.1190	−.1797	73.2301	−2.0012
Rough grid point search								
$(\gamma, \rho) = (.85, .9)$								
Independence	47.7800	−1.9730	44.5759	−1.2802	82.4880	−1.7276	79.5815	−2.7890
QL	31.9390	−.4489	29.3436	−.0714	60.7410	.0272	54.9896	.7295
MGV	33.2930	−.5232	31.4578	−.2297	63.8040	.4463	58.2365	.6303
$(\gamma, \rho) = (.85, .6)$								
Independence	47.4990	−2.6415	49.4655	−.8980	82.0940	−1.1161	82.5541	−3.0444
QL	37.2570	−1.1533	36.4912	−1.1077	6.9254	.2263	63.1202	−1.2543
MGV	40.6740	−1.1390	39.7678	−1.5885	77.0800	.5339	71.0284	.7412
$(\gamma, \rho) = (.85, .3)$								
Independence	46.9910	−2.8990	47.4580	−1.6817	81.7980	−1.0896	83.9923	−1.2721
QL	41.3200	−1.6483	40.9850	−1.8832	75.1910	.0079	73.4095	.3885
MGV	48.4380	−1.5369	47.8895	−1.9910	91.9430	.2399	84.2413	1.8811

[*] Values in the columns of SD, bias, MAD, and median are multiplied by a factor of 1,000.

mula works very well for both correctly specified and misspecified correlation structures.

5.2 Comparison With the Traditional Approach

In this section we demonstrate the flexibility and efficiency of model (1) by comparing its performance with linear models for longitudinal data,

$$y(t) = \mathbf{x}(t)^T \boldsymbol{\alpha} + \mathbf{z}(t)^T \boldsymbol{\beta} + \varepsilon(t), \tag{14}$$

which can be viewed as a special case of model (1) with constant function $\boldsymbol{\alpha}(\cdot)$. We used the weighted least squares method to estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in model (14). To make the comparison fair, we generated 1,000 datasets, each consisting of $n = 50$ samples, from model (13) with the following:

- Case I: $\alpha_1(t) = \sqrt{t/12}$ and $\alpha_2(t) = \sin(2\pi t/12)$, exactly the same as given in Section 5.1.
- Case II: $\alpha_1(t) = 2$ and $\alpha_2(t) = 1$; that is, both $\alpha_1(t)$ and $\alpha_2(t)$ are constant functions.

Table 3. Standard Errors

	$\hat{\beta}_1$		$\hat{\beta}_2$	
	SD	*SE (Std)*	*SD*	*SE (Std)*
ARMA(1, 1) working correlation matrix				
Independence	.0478	$.0464_{(.0065)}$	.0825	$.0800_{(.0108)}$
QL	.0252	$.0254_{(.0030)}$	.0449	$.0440_{(.0047)}$
MGV	.0252	$.0257_{(.0031)}$	.0456	$.0446_{(.0049)}$
AR(1) working correlation matrix				
QL	.0319	$.0307_{(.0078)}$	.0609	$.0541_{(.0131)}$
MGV	.0331	$.0316_{(.0084)}$	.0635	$.0557_{(.0141)}$

All other parameters and generation schemes of observation times are the same as those specified in Section 5.1.

To illustrate the flexibility of model (1), we fit data generated under the setting of Case I using the linear model (14). The error correlation structure is no longer ARMA if model (14) is fitted under the setting of Case I. Thus we did not include the "true" correlation structure in our simulation. Simulation results are summarized in the top part of Table 4, in which the notation is the same as that given in Tables 1 and 2. To save space, we present only the simulation results with $(\gamma, \rho) = (.85, .6)$; results for other (γ, ρ) pairs are similar. Compared with the results in Tables 1 and 2, misspecification $\boldsymbol{\alpha}(t)$ may yield a less efficient estimate with larger bias.

Simulation results of models (14) and (1) for Case II are summarized in the middle and bottom parts of Table 4. The bias of the resulting estimates for all estimation procedures are in the same magnitude. Comparing the simulation of models (14) and (1) with independent working correlation matrix and with the true/QL ARMA(1, 1) correlation matrix shows that the proposed models do not lose much efficiency. In summary, the proposed estimation procedure with model (1) offers a good balance between model flexibility and estimation efficiency.

5.3 An Application

Here we demonstrate the newly proposed procedures through an analysis of a subset of data from the Multi-Center AIDS Cohort study. The dataset comprises the human immunodeficiency virus (HIV) status of 283 homosexual men who were infected with HIV during the following-up period of 1984–1991. This dataset has been analyzed by Fan and Zhang (2000) and Huang, Wu, and Zhou (2002) using functional linear models. Details of

*Table 4. Comparison to the Linear Model**

			$\hat{\beta}_1$		$\hat{\beta}_2$	
Model	*Correlation*	*Method*	*MAD*	*Median(bias)*	*MAD*	*Median(bias)*
Case I: $\alpha_1(t) = \sqrt{t/12}$, $\alpha_2(t) = \sin(2\pi t/12)$, and $(\gamma, \rho) = (.85, .6)$						
(14)	Independence		62.5252	−3.7274	102.7234	2.3116
	AMRA(1, 1)	QL	43.3809	−5.2141	75.1942	1.0293
	ARMA(1, 1)	MGV	60.7346	−4.1006	98.3291	1.8330
	AR(1)	QL	52.8866	−2.2510	93.2711	−3.9622
	AR(1)	MGV	59.9004	−3.2324	96.4753	1.1836
Case II: $\alpha_1(t) = 2$, $\alpha_2(t) = 1$, and $(\gamma, \rho) = (.85, .6)$						
(14)	Independence		47.7871	−3.1878	82.1597	−2.2803
	AMRA(1, 1)	True	32.9404	−1.9984	61.6268	.5491
	ARMA(1, 1)	QL	33.1803	−2.8015	61.8600	.1782
	ARMA(1, 1)	MGV	47.0792	−1.2353	76.6334	−.6911
	AR(1)	QL	35.1901	−.8354	64.2013	−.3883
	AR(1)	MGV	47.0576	−1.4820	76.8226	−.8559
(1)	Independence		49.4474	−1.0333	82.7413	−3.0255
	AMRA(1, 1)	True	34.3453	−1.6239	63.0509	.2820
	ARMA(1, 1)	QL	35.3995	−1.7503	62.9548	−.5040
	ARMA(1, 1)	MGV	35.6286	−.3130	62.1033	−2.5856
	AR(1)	QL	36.2746	−.8732	63.2304	−1.2967
	AR(1)	MGV	39.8883	−.8650	72.1075	1.2003

*Values in the columns of MAD and median are multiplied by a factor of 1,000.

the study design, methods, and medical implications have been given by Kaslow et al. (1987).

All participants were scheduled to undergo measurement during semiannual visits, but because many participants missed some of their scheduled visits and the HIV infections occurred randomly during the study, there are unequal numbers of repeated measurements and different measurement times per individual. Our interest is in describing the trend in mean CD4 percentage depletion over time and to evaluate the effects of cigarette smoking, pre-HIV infection CD4 percentage, and age at infection on the mean CD4 percentage after the infection. Huang et al. (2002) took the response $y(t)$ to be CD4 cell percentage and considered the functional linear model

$$y(t) = \beta_0(t) + \beta_1(t)\text{Smoking} + \beta_2(t)\text{Age} + \beta_3(t)\text{PreCD4} + \varepsilon(t). \quad (15)$$

The results of the hypothesis testing of Huang et al. (2002) indicate that the baseline function varies over time; neither Smoking nor Age has a significant impact on the mean CD4 percentage, and whether or not PreCD4 has a constant effect over time is nuclear. The p value for testing whether or not $\beta_3(t)$ varies over time is .059. Thus we fit the data using a simpler semiparametric varying-coefficient partially linear model,

$$y(t) = \alpha_1(t) + \alpha_2(t)X_1 + \beta_1 Z_1 + \beta_2 Z_2 + \varepsilon(t),$$

where, for numerical stability, X_1 is the standardized variable for PreCD4, Z_1 is the smoking status (1 for a smoker and 0 for a nonsmoker), Z_2 is the standardized variable for age, and the unit for observation time t is 1 month.

Bandwidth Selection. We use a multifold cross-validation method to select a bandwidth for $\hat{\boldsymbol{\alpha}}(t)$. We partition the data into Q groups, each of which has approximately the same number of subjects. For each k, $k = 1, \ldots, Q$, we fit model (15) for the data excluding the k-group of data. The cross-validation score is defined as the sum of residual squares,

$$CV(h) = \sum_{k=1}^{Q} \sum_{i \in d_k} \sum_{j=1}^{J_i} \left\{ y_i(t_{ij}) - \hat{y}_{-d_k}(t_{ij}) \right\}^2,$$

where $\hat{y}_{-d_k}(t_{ij})$ is the fitted value for the ith subject at observed time t_{ij} with the data in d_k deleted, using a working independence correlation matrix. In the implementation, we choose $Q = 15$. Figure 1(a) depicts the cross-validation score function, $CV(h)$, that gives the optimal bandwidth $h = 21.8052$. Note that $\hat{\sigma}^2(t)$ is a one-dimensional kernel regression of the squared residuals over time. Thus various bandwidth selectors for one-dimensional smoothing can be used to choose a bandwidth for $\hat{\sigma}^2(t)$. In this application we directly use the plug-in bandwidth selector (Ruppert et al. 1995) and choose the bandwidth $h_1 = 12.7700$.

Estimation. The resulting estimate of $\boldsymbol{\alpha}(t)$ is depicted in Figures 1(b) and 1(c). The intercept function decreases with

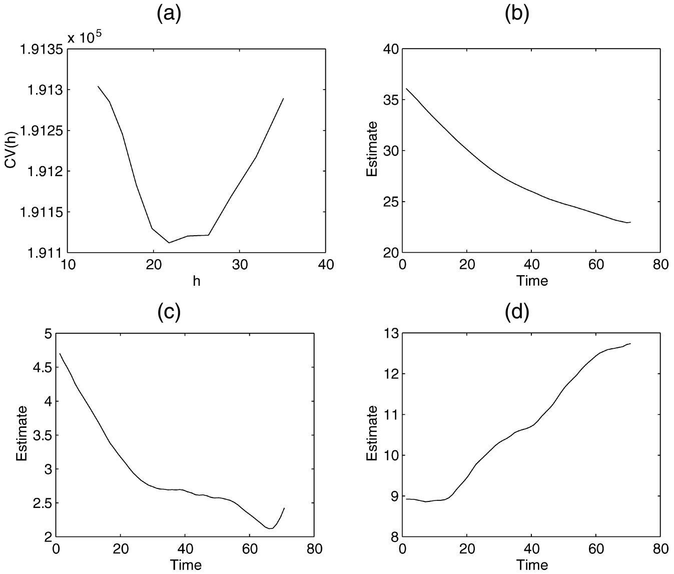

Figure 1. Plots of (a) the Cross-Validation Score Against the Bandwidth, (b) and (c) Estimate of $\alpha_1(t)$ and $\alpha_2(t)$ With Bandwidth 21.8052, Chosen by the Cross-Validation Method, and (d) Estimated $\sigma(t)$ With Bandwidth 12.7700, Chosen by the Plug-in Method.

time, implying an overall trend of decreasing CD4 cell percentage over time. The trend for $\alpha_2(t)$ implies that the impact of PreCD4 on CD4 cell percentage decreases gradually as time evolves. The results are consistent with our expectation; they quantify the extent to which the mean CD4 percentage decreases over time and how the association between CD4 percentage and PreCD4 varies as time evolves. The resulting estimate of $\hat{\sigma}(t)$, depicted in Figure 1(d), indicates that $\sigma(t)$ seems to be constant during the first and half year and then increases as time increases. This shows that predicting the CD4 percentage becomes harder over time.

We next estimate $\boldsymbol{\beta}$. Here we consider an ARMA(1, 1) correlation structure. The proposed estimation procedures in Section 2 were applied for estimating (γ, ρ). The resulting estimates are displayed in the top part of Table 5, and the corresponding estimates for $\boldsymbol{\beta}$ are depicted in the bottom part of Table 5. The QL approach yields a correlation structure with moderate correlation, and the standard error for the resulting estimate of $\boldsymbol{\beta}$ is smaller than that obtained using the independence correlation structure. The MVG method results in a correlation structure with low correlation but with corresponding standard error still smaller than that of the independence correlation structure. Table 5 shows that the effects of smoking status and age are not significant under the three estimation schemes.

Prediction of Individual Trajectory. We now illustrate how to incorporate correlation information into prediction. Let us assume that given the covariates $\mathbf{x}(t)$ and $\mathbf{z}(t)$, the error process $\varepsilon(t)$ is a Gaussian process with mean 0 and covariance function $c(t, s)$. Denote $\mu(t) = \mathbf{x}(t)^T\boldsymbol{\alpha}(t) + \mathbf{z}(t)^T\boldsymbol{\beta}$. Suppose that data for an individual are collected at $t = t_1, \dots, t_J$ and we want to predict his or her $y(t)$ at $t = t^*$ with covariates $\mathbf{x}(t^*)$ and $\mathbf{z}(t^*)$. Let $\mathbf{y}_o = (y(t_1), \dots, y(t_J))^T$ be the observed response and let $\boldsymbol{\mu} = (\mu(t_1), \dots, \mu(t_J))^T$ be its associated mean. Let Σ be the covariance matrix of $(\varepsilon(t_1), \dots, \varepsilon(t_J))^T$, and let $\mathbf{c}^* = (c(t_1, t^*), \dots, c(t_J, t^*))^T$. Then, by the properties of the multivariate normal distribution, we have

$$E\{y(t^*)|\mathbf{y}_o\} = \mu(t^*) + \mathbf{c}^{*T}\Sigma^{-1}(\mathbf{y}_o - \boldsymbol{\mu})$$

and

$$\operatorname{var}\{y(t^*)|\mathbf{y}_o\} = \sigma^2(t^*) - \mathbf{c}^{*T}\Sigma^{-1}\mathbf{c}^*.$$

Thus the prediction of $y(t^*)$ is

$$\hat{y}(t^*) = \hat{\mu}(t^*) + \hat{\mathbf{c}}^{*T}\hat{\Sigma}^{-1}(\mathbf{y}_o - \hat{\boldsymbol{\mu}}).$$

Because the errors in estimating the unknown regression coefficients and parameters of the covariance matrix are negligible relative to random error, the $(1 - \alpha)100\%$ predictive interval is

$$\hat{y}(t^*) \pm z_{1-\alpha/2}\sqrt{\hat{\sigma}^2(t^*) - \hat{\mathbf{c}}^{*T}\hat{\Sigma}^{-1}\hat{\mathbf{c}}^*},$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$th quantile of the standard normal distribution. In particular, it is easy to verify that when t^* is one of the observed time points, the prediction error is zero, a desired property.

Table 5. Estimates of (γ, ρ) and β

	Independence	QL	MGV
$\hat{\gamma}$		.8575	.5334
$\hat{\rho}$		.9852	.0804
$\hat{\beta}_1$	$.8726_{(1.1545)}$	$.6848_{(.9972)}$	$.6328_{(1.0864)}$
$\hat{\beta}_2$	$-.5143_{(.6110)}$	$.0556_{(.4718)}$	$-.3658_{(.5488)}$

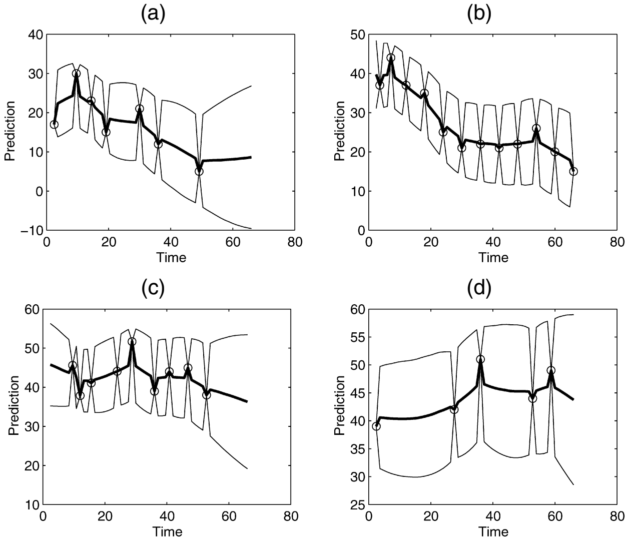

Figure 2. Plot of Pointwise Predictions and Their 95% Predictive Intervals for Four Typical Subjects. The solid line is the prediction, the dashed and dotted lines represent the limits for the 95% pointwise predictive confidence interval, and "○" is the observed value of y(t).

We now apply the prediction procedure for this application. Assume that $\varepsilon(t)$ has AMRA(1, 1) correlation structure. As an illustration, here we consider only the prediction with (γ, ρ) estimated by the QL approach, that is, $(\hat{\gamma}, \hat{\rho}) = (.8575, .9852)$. Predictions and their 95% predictive intervals for four typical subjects are displayed in Figure 2.

6. DISCUSSION

In this article we proposed a class of semiparametric models for the covariance function of longitudinal data. We further developed an estimation procedure for $\sigma^2(t)$ using kernel regression, estimation procedures for $\boldsymbol{\theta}$ in correlation matrix using QL and MGV approaches, and estimation procedure for regression coefficients $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}$ using profile weighted least squares. Robust method estimation procedures have been proposed for semiparametric regression modeling with longitudinal data (He et al. 2002, 2005). In the presence of outliers, one should consider a robust method to estimate $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}$.

Although misspecification of the correlation structure $\rho(s, t, \boldsymbol{\theta})$ does not affect the consistency of the resulting estimate of $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}$, it may lead to nonexistence or inconsistency of the estimates of $\boldsymbol{\theta}$. Thus it is of interest to check whether the imposed correlation structure is approximately correct. To address this issue, we may consider a full nonparametric estimate for the correlation function $\rho(s, t)$,

$$\rho(s, t) = \frac{\sum_{i=1}^{n}\sum_{j\neq j'}^{J_i} \hat{e}_i(t_{ij})\hat{e}_i(t_{ij'})K_{h_2}(s - t_{ij})K_{h_2}(t - t_{ij'})}{\sum_{i=1}^{n}\sum_{j\neq j'}^{J_i} K_{h_2}(s - t_{ij})K_{h_2}(t - t_{ij'})} \tag{16}$$

for $s \neq t$, where $\hat{e}(t_{ij}) = \hat{r}_{ij}/\hat{\sigma}(t_{ij})$, the standardized residual.

The nonparametric covariance estimator cannot be guaranteed to be positive definite, but it may be useful in specifying an approximate correlation structure or in checking whether the imposed correlation structure $\rho(s,t,\boldsymbol\theta)$ its approximately correct. This is a two-dimensional smoothing problem, but the effective data points in (16) can be small unless the time points for each subject are nearly balanced.

Some alternative estimation procedures for $\boldsymbol\alpha(t)$ and $\boldsymbol\beta$ also may be considered. For example, an alternative strategy for estimating $\boldsymbol\beta$ is to first decorrelate data within subjects, and then apply the profile least squares techniques to the decorrelated data. Further research and comparison may be of interest.

In this article we have not discussed the sampling property of $\hat{\boldsymbol\theta}$ derived by the QL and MGV approaches. If the correlation function is correctly specified, then the asymptotic property of $\hat{\boldsymbol\theta}$ may be derived by following conventional techniques related to linear mixed-effects models. It would be interesting to investigate the asymptotic behaviors of $\hat{\boldsymbol\theta}$ when the correlation function is misspecified. Some new formulation may be needed to establish the asymptotic property of $\hat{\boldsymbol\theta}$. This research topic is beyond the scope of this article; further research is needed.

APPENDIX: CONDITIONS AND PROOFS

The following technical conditions are imposed. They are not the weakest possible conditions, but they are imposed to facilitate the proofs.

1. The density function $f(\cdot)$ is Lipschitz-continuous and bounded away from 0. The function $K(\cdot)$ is a symmetric density function with a compact support.
2. $nh^8 \to 0$ and $nh^2/(\log n)^3 \to \infty$.
3. $E\mathbf{x}(t)\mathbf{x}(t)^T$ and $E\mathbf{x}(t)\mathbf{z}(t)^T$ are Lipschitz-continuous.
4. J_i has a finite moment-generating function. In addition, $E\|\mathbf{x}(t)\|^4 + E\|\mathbf{z}(t)\|^2 < \infty$.
5. $\boldsymbol\alpha(t)$ has a continuous second derivative.
6. $\sigma^2(\cdot)$ has a continuous second derivative.

Proof of Theorem 1

First, by condition 4, we can easily show almost surely that $\max_{1\le i\le n} J_i = O(\log n)$. For each given $\boldsymbol\beta$, the estimator $\hat{\boldsymbol\alpha}(t;\boldsymbol\beta)$ is a local linear estimator by minimizing (6) based on data

$$\{t_{ij}, \mathbf{x}_i(t_{ij}), y_i^*(t_{ij})\}, \qquad j = 1,\dots,J_i, i = 1,\dots,n.$$

Observe that $\{y_i^*(t_{ij}), j = 1,\dots,J_i\}$ is a realization from the process

$$y^*(t) = \mathbf{x}(t)^T\boldsymbol\alpha_0(t) + \mathbf{z}(t)^T(\boldsymbol\beta_0 - \boldsymbol\beta) + \varepsilon(t).$$

Note that the consistency of $\hat{\boldsymbol\alpha}(t;\boldsymbol\beta)$ is not affected by ignoring the correlation within subjects. Following the proof of Fan and Huang (2005), $\hat{\boldsymbol\alpha}(t;\boldsymbol\beta)$ is a consistent estimator of the function

$$\boldsymbol\alpha(t;\boldsymbol\beta) = \boldsymbol\alpha_0(t) - \mathbf{G}^{-1}(t)\Psi(t)(\boldsymbol\beta - \boldsymbol\beta_0). \tag{A.1}$$

Indeed, uniformly in t,

$$\hat{\boldsymbol\alpha}(t;\boldsymbol\beta) - \boldsymbol\alpha(t;\boldsymbol\beta) = O_P(c_n), \tag{A.2}$$

where $c_n = h^2 + \{-\log h/(nh)\}^{1/2}$. Let $\hat m_{ij}(\boldsymbol\beta) = \mathbf{x}_i(t_{ij})^T\hat{\boldsymbol\alpha}(t_{ij};\boldsymbol\beta)$ and $\hat{\mathbf{m}}_i(\boldsymbol\beta) = (\hat m_{i1},\dots,\hat m_{iJ_i})^T$. Note that the profile weighted least squares estimate $\hat{\boldsymbol\beta}$ is the minimizer of the weighted quadratic function

$$\ell_n(\boldsymbol\beta) = \frac{1}{n}\sum_{i=1}^n (\mathbf{y}_i - \hat{\mathbf{m}}_i(\boldsymbol\beta) - \mathbf{Z}_i\boldsymbol\beta)^T\mathbf{W}_i(\mathbf{y}_i - \hat{\mathbf{m}}_i(\boldsymbol\beta) - \mathbf{Z}_i\boldsymbol\beta), \tag{A.3}$$

which is a convex and quadratic function of $\boldsymbol\beta$. This allows us to apply the convexity lemma and the quadratic approximation lemma (see, e.g., Fan and Gibjels 1996, pp. 209–210) to establish the asymptotic normality of $\hat{\boldsymbol\beta}$.

We next decompose $\ell_n(\boldsymbol\beta)$. Write

$$\mathbf{m}_i(\boldsymbol\beta) = \big(\mathbf{x}_i(t_{i1})^T\boldsymbol\alpha(t_{i1};\boldsymbol\beta),\dots,\mathbf{x}_i\big(t_{iJ_1}\big)^T\boldsymbol\alpha\big(t_{iJ_1};\boldsymbol\beta\big)\big)^T,$$

$$I_{n,1}(\boldsymbol\beta) = \frac{1}{n}\sum_{i=1}^n\{\mathbf{y}_i - \mathbf{m}_i(\boldsymbol\beta) - \mathbf{Z}_i\boldsymbol\beta\}^T\mathbf{W}_i\{\mathbf{y}_i - \mathbf{m}_i(\boldsymbol\beta) - \mathbf{Z}_i\boldsymbol\beta\},$$

$$I_{n,2}(\boldsymbol\beta) = \frac{2}{n}\sum_{i=1}^n\{\mathbf{y}_i - \mathbf{m}_i(\boldsymbol\beta) - \mathbf{Z}_i\boldsymbol\beta\}^T\mathbf{W}_i\{\mathbf{m}_i(\boldsymbol\beta) - \hat{\mathbf{m}}_i(\boldsymbol\beta)\},$$

and

$$I_{n,3}(\boldsymbol\beta) = \frac{1}{n}\sum_{i=1}^n\{\mathbf{m}_i(\boldsymbol\beta) - \hat{\mathbf{m}}_i(\boldsymbol\beta)\}^T\mathbf{W}_i\{\mathbf{m}_i(\boldsymbol\beta) - \hat{\mathbf{m}}_i(\boldsymbol\beta)\},$$

where $\mathbf{m}_i(\boldsymbol\beta) = (m_{i1},\dots,m_{iJ_i})$ with $m_{ij}(\boldsymbol\beta) = x_i(t_{ij})^T\boldsymbol\alpha(t_{ij},\boldsymbol\beta)$. Then

$$\ell_n(\boldsymbol\beta) = I_{n,1}(\boldsymbol\beta) + I_{n,2}(\boldsymbol\beta) + I_{n,3}(\boldsymbol\beta). \tag{A.4}$$

Note that $I_{n,2}(\boldsymbol\beta)$ and $I_{n,3}(\boldsymbol\beta)$ are quadratic in $\boldsymbol\beta$. Using techniques related to the approaches of Müller and Stadtmüller (1993) and Fan and Huang (2005), and after some tedious calculations, it follows that for each given $\boldsymbol\beta$,

$$I_{n,2}(\boldsymbol\beta) = I_{n,3}(\boldsymbol\beta) = O(c_n^2) = o_P\big(n^{-1/2}\big). \tag{A.5}$$

We now deal with the main term $I_{n,1}(\boldsymbol\beta)$. Using the model

$$y(t) = \mathbf{x}(t)^T\boldsymbol\alpha_0(t) + \mathbf{z}(t)^T\boldsymbol\beta_0 + \varepsilon(t)$$

and (A.1), we have

$$I_{n,1}(\boldsymbol\beta) = \frac{1}{n}\sum_{i=1}^n \varepsilon_i^T W_i\varepsilon_i - 2(\boldsymbol\beta - \boldsymbol\beta_0)^T\xi_n + (\boldsymbol\beta - \boldsymbol\beta_0)^T\Sigma_n(\boldsymbol\beta - \boldsymbol\beta_0). \tag{A.6}$$

The minimization of $I_{n,1}$ is given by

$$\hat{\boldsymbol\beta}_0 = \boldsymbol\beta_0 + \Sigma_n\xi_n,$$

where Σ_n and ξ_n are defined before Theorem 1. By the weak law of large numbers and central limit theorem,

$$\sqrt{n}(\hat{\boldsymbol\beta}_0 - \boldsymbol\beta_0) \xrightarrow{\mathcal{L}} \mathrm{N}(0, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}), \tag{A.7}$$

where $\mathbf{A}$ and $\mathbf{B}$ are as defined in Section 3.2. Finally, we apply the convexity lemma to show that

$$\sqrt{n}(\hat{\boldsymbol\beta} - \boldsymbol\beta_0) = \sqrt{n}\Sigma_n^{-1}\xi_n + o_P(1). \tag{A.8}$$

This together with (A.7) proves the results. To show (A.8), first, by the convexity lemma, $\hat{\boldsymbol\beta}$ is a consistent estimator of $\boldsymbol\beta_0$. From (A.4), we have

$$\begin{aligned}0 &= \dot I_{n,1}(\hat{\boldsymbol\beta}) + \dot I_{n,2}(\hat{\boldsymbol\beta}) + \dot I_{n,3}(\hat{\boldsymbol\beta})\\ &= 2\Sigma_n(\hat{\boldsymbol\beta} - \boldsymbol\beta_0) - 2\xi_n + \dot I_{n,2}(\hat{\boldsymbol\beta}) + \dot I_{n,3}(\hat{\boldsymbol\beta}).\end{aligned}$$

Because $I_2(\boldsymbol\beta)$ and $I_3(\boldsymbol\beta)$ are quadratic in $\boldsymbol\beta$, it follows from (A.5) that

$$\dot I_{n,2}(\hat{\boldsymbol\beta}) = o_P\big(n^{-1/2}\big) \qquad \text{and} \qquad \dot I_{n,3}(\hat{\boldsymbol\beta}) = o_P\big(n^{-1/2}\big).$$

This completes the proof of Theorem 1.

Proof of (12)

Write $\mathbf{U} = (\mathbf{Z}_1 - \tilde{\mathbf{X}}_1)$, and $\mathbf{W}_0 = \operatorname{cov}(\boldsymbol{\varepsilon}|\mathbf{X}_1, \mathbf{Z}_1)$. Define

$$\mathbf{D} = \{E(\mathbf{U}^T\mathbf{W}_1\mathbf{U})\}^{-1}\mathbf{U}^T\mathbf{W}_1\mathbf{W}_0^{1/2} - \{E(\mathbf{U}^T\mathbf{W}_0^{-1}\mathbf{U})\}^{-1}\mathbf{U}^T\mathbf{W}_0^{-1/2}.$$

Then

$$\begin{aligned}\mathbf{D}\mathbf{D}^T &= \{E(\mathbf{U}^T\mathbf{W}_1\mathbf{U})\}^{-1}(\mathbf{U}^T\mathbf{W}_1\mathbf{W}_0\mathbf{W}_1\mathbf{U})\{E(\mathbf{U}^T\mathbf{W}_1\mathbf{U})\}^{-1}\\ &\quad - \{E(\mathbf{U}^T\mathbf{W}_1\mathbf{U})\}^{-1}(\mathbf{U}^T\mathbf{W}_1\mathbf{U})\{E(\mathbf{U}^T\mathbf{W}_0^{-1}\mathbf{U})\}^{-1}\\ &\quad - \{E(\mathbf{U}^T\mathbf{W}_0^{-1}\mathbf{U})\}^{-1}(\mathbf{U}^T\mathbf{W}_1\mathbf{U})\{E(\mathbf{U}^T\mathbf{W}_1\mathbf{U})\}^{-1}\\ &\quad + \{E(\mathbf{U}^T\mathbf{W}_0^{-1}\mathbf{U})\}^{-1}(\mathbf{U}^T\mathbf{W}_0^{-1}\mathbf{U})\{E(\mathbf{U}^T\mathbf{W}_0^{-1}\mathbf{U})\}^{-1}.\end{aligned}$$

Because $\mathbf{D}\mathbf{D}^T$ is nonnegative definite, we have that

$$\begin{aligned}E(\mathbf{D}\mathbf{D}^T) = \{E(\mathbf{U}^T\mathbf{W}_1\mathbf{U})\}^{-1}E(\mathbf{U}^T\mathbf{W}_1\mathbf{W}_0\mathbf{W}_1\mathbf{U})\{E(\mathbf{U}^T\mathbf{W}_1\mathbf{U})\}^{-1}\\ - \{E(\mathbf{U}^T\mathbf{W}_0^{-1}\mathbf{U})\}^{-1}\end{aligned}$$

is nonnegative definite. Hence

$$\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1} - \mathbf{B}_0^{-1} \geq 0.$$

The equality holds if and only if $\mathbf{D} = 0$, which occurs when $\mathbf{W} = \mathbf{W}_0^{-1}$.

[Received March 2006. Revised October 2006.]

REFERENCES

Cai, Z., Fan, J., and Li, R. (2000), "Efficient Estimation and Inferences for Varying-Coefficient Models," *Journal of the American Statistical Association*, 95, 888–902.

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

Fan, J., and Huang, T. (2005), "Profile Likelihood Inferences on Semiparametric Varying Coefficient Partially Linear Models," *Bernoulli*, 11, 1031–1059.

Fan, J., Huang, T., and Li, R. (2005), "Analysis of Longitudinal Data With Semiparametric Estimation of Covariance Function," Technical Report 05-074, Methodology Center, Pennsylvania State University.

Fan, J., and Yao, Q. (1998), "Efficient Estimation of Conditional Variance Functions in Stochastic Regression," *Biometrika*, 85, 645–660.

Fan, J., and Zhang, J. (2000), "Two-Step Estimation of Functional Linear Models With Applications to Longitudinal Data," *Journal of the Royal Statistical Society*, Ser. B, 62, 303–322.

Härdle, W., Liang, H., and Gao, J. (2000), *Partially Linear Models*, New York: Springer-Verlag.

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

He, X., Fung, W. K., and Zhu, Z. Y. (2005), "Robust Estimation in Generalized Partial Linear Models for Clustered Data," *Journal of the American Statistical Association*, 100, 1176–1184.

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

Huang, J. Z., Liu, N., Pourahmadi, M., and Liu, L. (2006), "Covariance Selection and Estimation via Penalized Normal Likelihood," *Biometrika*, 93, 85–98.

Huang, J. Z., Wu, C. O., and Zhou, L. (2002), "Varying-Coefficient Models and Basis Function Approximations for the Analysis of Repeated Measurements," *Biometrika*, 89, 111–128.

Huang, J. Z., and Zhang, L. (2004), "Efficient Estimation in Marginal Partially Linear Models for Longitudinal/Clustered Data Using Splines," manuscript.

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

Lin, X., and Carroll, R. (2000), "Nonparametric Function Estimation for Clustered Data When the Predictor Is Measured Without/With Error," *Journal of the American Statistical Association*, 95, 520–534.

——— (2001), "Semiparametric Regression for Clustered Data Using Generalized Estimating Equations," *Journal of the American Statistical Association*, 96, 1045–1056.

Martinussen, T., and Scheike, T. H. (1999), "A Semiparametric Additive Regression Model for Longitudinal Data," *Biometrika*, 86, 691–702.

Müller, H. G., and Stadtmüller, U. (1993), "On Variance Function Estimation With Quadratic Forms," *Journal of Statistical Planning and Inference*, 35, 213–231.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995), "An Effective Bandwidth Selector for Local Least Squares Regression," *Journal of the American Statistical Association*, 90, 1257–1270.

Ruppert, D., Wand, M. P., Holst, U., and Hössjer, O. (1997), "Local Polynomial Variance Function Estimation," *Technometrics*, 39, 262–273.

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

Sun, Y., and Wu, H. (2005), "Semiparametric Time-Varying Coefficients Regression Model for Longitudinal Data," *Scandinavian Journal of Statistics*, 32, 21–47.

Wang, N. (2003), "Marginal Nonparametric Kernel Regression Accounting Within-Subject Correlation," *Biometrika*, 90, 29–42.

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.

Wu, W. B., and Pourahmadi, M. (2003), "Nonparametric Estimation of Large Covariance Matrices of Longitudinal Data," *Biometrika*, 90, 831–844.

Xia, Y., Zhang, W., and Tong, H. (2004), "Efficient Estimation for Semivarying-Coefficient Models," *Biometrika*, 91, 661–681.

Yao, F., Müller, H. G., and Wang, J.-L. (2005a), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100, 577–590.

——— (2005b), "Functional Regression Analysis for Longitudinal Data," *The Annals of Statistics*, 33, 2873–2903.

Zhang, W., Lee, S. Y., and Song, X. (2002), "Local Polynomial Fitting in Semivarying Coefficient Models," *Journal of Multivariate Analysis*, 82, 166–188.