Statistica Sinica **21** (2011), 1145-1170

A TWO-STAGE ESTIMATION METHOD FOR RANDOM COEFFICIENT DIFFERENTIAL EQUATION MODELS WITH APPLICATION TO LONGITUDINAL HIV DYNAMIC DATA

Yun Fang, Hulin Wu and Li-Xing Zhu

Shanghai Normal University, University of Rochester and Hong Kong Baptist University

Abstract: We propose a two-stage estimation method for random coefficient ordinary differential equation (ODE) models. A maximum pseudo-likelihood estimator (MPLE) is derived based on a mixed-effects modeling approach and its asymptotic properties for population parameters are established. The proposed method does not require repeatedly solving ODEs, and is computationally efficient although it does pay a price with the loss of some estimation efficiency. However, the method does offer an alternative approach when the exact likelihood approach fails due to model complexity and high-dimensional parameter space, and it can also serve as a method to obtain the starting estimates for more accurate estimation methods. In addition, the proposed method does not need to specify the initial values of state variables and preserves all the advantages of the mixed-effects modeling approach. The finite sample properties of the proposed estimator are studied via Monte Carlo simulations and the methodology is also illustrated with application to an AIDS clinical data set.

Key words and phrases: AIDS/HIV data, local polynomial kernel smoothing, longitudinal data, mixed-effects models, ordinary differential equation, pseudo likelihood, viral dynamics.

1. Introduction

Ordinary differential equation (ODE) models are widely used in such scientific fields as engineering, physics, econometrics, and, recently, biomedical sciences. ODE models have been applied to quantify HIV viral dynamics and have led to many important findings for AIDS pathogenesis in the past two decades (Ho et al. (1995), Perelson et al. (1996), Perelson et al. (1997), Perelson and Nelson (1999), Wu et al. (1999), Wu and Ding (1999), Nowak and May (2000), Tan and Wu (2005), and Wu (2005)).

In this paper, we use a two-stage method to estimate the random coefficient parameters in ODE models for longitudinal data. In the first stage, we estimate the state variables and time-varying covariates, as well as their derivatives, by using local polynomial smoothing for the nonparametric mixed effects models (Wu and Zhang (2002)). In the second stage, a maximum pseudo-likelihood (MPL) estimation is developed to estimate the unknown parameters, including both the population and individual coefficients in a random coefficient ODE model. The asymptotic properties of the proposed estimator for population parameters are established. The spirit of two-stage method was initiated by Varah (1982) who used a cubic spline to smooth the data in the first stage, and employed the least squares method for parameter estimation in the second stage. Ellner, Seifu, and Smith (2002) fitted the dynamic models to time series data using the local polynomial regression. Liang and Wu (2008) established the theoretical properties of the method for ODE models by applying the local polynomial smoothing in the first stage. They proved that the two-step estimator has strong consistency and asymptotic normality. Brunel (2008) obtained similar asymptotic results by using regression splines. Chen and Wu (2008a,b) adopted a similar two-stage method for ODE models with time-varying coefficients and also obtained the asymptotic results for the proposed estimators. However, this literature deals only with cross-section iid data.

Longitudinal dynamic systems (random coefficient ODE models) have been suggested by Putter et al. (2002), Huang and Wu (2006), and Huang, Liu, and Wu (2006), in which the hierarchical Bayesian approach was used to estimate dynamic parameters in HIV dynamic models from longitudinal clinical data. Lahiri (2003) proposed a spline-enhanced population model to study pharmacokinetics using a random time-varying coefficient ODE model. Guedj, Thiébaut, and Commenges (2007) used the maximum likelihood approach directly to estimate unknown parameters in random coefficient ODE models.

In this paper, we extend the two-step estimation method (Varah (1982), Liang and Wu (2008)) to longitudinal dynamic systems by adopting the mixed-effects modeling approach. The extension is not trivial. When the nonparametric mixed-effects (NPME) modeling approach is used in the first step, we need to resort to asymptotic independence and a pseudo-likelihood idea to derive the parameter estimates in the second step. The asymptotic theories for the proposed pseudo-likelihood estimator are established. We also demonstrate, via Monte Carlo simulations, that the NPME modeling approach in the first step provides a good estimate for the state variables and their derivatives. Thus, the proposed method preserves the advantages of both the mixed-effects modeling approach and the two-step estimation method for ODE models: it avoids repeatedly solving the ODE numerically so that it is computationally efficient; it does not require initial values of state variables; and implementation is easy compared to the exact maximum likelihood method for ODE models.

The remainder of this paper is organized as follows. The two-stage estimation procedure is described in Section 2. The theoretical properties for the population estimates are established in Section 3. In Section 4, we fit a viral dynamic model to a longitudinal HIV dynamic data from an AIDS clinical study to further illustrate the usefulness of the proposed method. In Section 5, we conduct simulation studies to evaluate the performance of the proposed estimates. We conclude our paper with some discussion in Section 6. The proofs of the theoretical results are provided in the Appendix.

2. Estimation Procedure

The proposed method is motivated by a HIV dynamic study with a popular ODE model:

$$\begin{aligned}\frac{d}{dt}T_U(t) &= \lambda - \rho T_U(t) - \eta(t)T_U(t)V(t),\\ \frac{d}{dt}T_I(t) &= \eta(t)T_U(t)V(t) - \delta T_I(t),\\ \frac{d}{dt}V(t) &= N\delta T_I(t) - cV(t),\end{aligned} \tag{2.1}$$

where $T_U(t)$ is the concentration of uninfected target T cells, $T_I(t)$ is the concentration of infected cells, and $V(t)$ is the concentration of plasma virus (viral load) at time t. The functions $V(t)$, $T_U(t)$, and $T_I(t)$ are called state variables. Parameter λ represents the rate at which new T cells are continuously generated, ρ is the death rate of uninfected T cells, $\eta(t)$ is the time-varying infection rate of T cells, δ is the death rate of infected cells, c is the clearance rate of free virus, and N is the average number of virus produced from each infected cell.

In the dynamic system (2.1), viral load $V(t)$ and total CD4+ T cell counts $T(t) = T_U(t)+T_I(t)$ can be measured in AIDS clinical studies. Using some simple algebra, Liang and Wu (2008) transformed the system (2.1) into a single ODE model:

$$V'(t) = \alpha_0 + \alpha_1 T(t) + \alpha_2 T'(t) - cV(t), \tag{2.2}$$

where $V'(t) = dV(t)/dt$ and $T'(t) = dT(t)/dt$, and parameters $\alpha_0 = -N\delta\lambda/(\rho - \delta)$, $\alpha_1 = -N\delta\rho/(\rho - \delta)$ and $\alpha_2 = N\delta/(\rho - \delta)$. Model (2.2) can be extended into a random coefficient ODE model as

$$V_i'(t_{ij}) = \alpha_{0i} + \alpha_{1i}T_i(t_{ij}) + \alpha_{2i}T_i'(t_{ij}) - c_iV_i(t_{ij}), \tag{2.3}$$

with subject index $i = 1, \ldots, n$ and the measurement index for the ith subject $j = 1, \ldots, n_i$. In general, model (2.3) can be written as

$$X_i'(t_{ij}) = F(X_i(t_{ij}), Z_i(t_{ij}), Z_i'(t_{ij}), \theta_i)^\tau, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i, \tag{2.4}$$

where X_i is a vector of state variables for the ith subject. The vector $Z_i(t_{ij}) = (Z_{i1}(t_{ij}), \ldots, Z_{ik}(t_{ij}))^\tau$ is a $k \times 1$ vector of input variables (covariates) and $Z_i'(t_{ij})$ denotes its derivative (the notation "τ" stands for transposition). $F(\cdot) = (F_1(\cdot), \ldots, F_q(\cdot))^\tau$ is a known linear or nonlinear function vector. The random coefficient (unknown parameter) vector can be written as $\theta_i = \theta + b_{\theta,i}$, where θ $(q \times 1)$ is the population parameter vector and $b_{\theta,i}$'s are random components of the parameters that are independent and identically distributed (i.i.d.) with mean 0 and covariance matrix D_θ. For simplicity, we only consider the case that X_i is a univariate state variable and the ODE model (2.4) is linear for unknown parameters θ_i, i.e., the linear mixed-effects ODE model

$$X_i'(t_{ij}) = F(X_i(t_{ij}), Z_i(t_{ij}), Z_i'(t_{ij}))^\tau \theta_i, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i. \tag{2.5}$$

Note that (2.5) has no closed-form solution in general. Although the theoretical results are difficult to establish and the computation is more costly, our methodologies are applicable to general nonlinear mixed-effects ODE models (models with nonlinear for unknown parameters). In this paper, for simplicity, our methodology and theoretical development focus on model (2.5).

In model (2.5), the state variable $X_i(t)$ and input variables $Z_{il}(t)$, for $1 \leq l \leq k$, are observed longitudinally, i.e., for $i = 1, \ldots, n$, $j = 1, \ldots, n_i$,

$$Y_i(t_{ij}) = X_i(t_{ij}) + \epsilon_i(t_{ij}), \;\; X_i(t_{ij}) = u(t_{ij}) + v_i(t_{ij}), \tag{2.6}$$

$$S_{il}(t_{ij}) = Z_{il}(t_{ij}) + \epsilon_{il}(t_{ij}), \;\; Z_{il}(t_{ij}) = u_l(t_{ij}) + v_{il}(t_{ij}), \tag{2.7}$$

where $u(t)$ and $u_l(t)$ are the population mean (fixed-effect) functions of the longitudinal data; $v_i(t)$ and $v_{il}(t)$, which are the subject-specific effects or random-effect functions, model the departure of the i-th individual effect from the population mean functions $u(t)$ and $u_l(t)$, respectively, and $\epsilon_i(t)$ and $\epsilon_{il}(t)$ are the measurement error functions, respectively. Models (2.6) and (2.7) can be considered as nonparametric mixed-effects (NPME) models (Shi, Weiss, and Taylor (1996), Rice and Wu (2001), and Wu and Zhang (2002, 2006)).

In the following subsections, we introduce the two-stage estimation procedure for the random coefficient ODE model (2.5). The basic idea is to apply a nonparametric smoothing approach to (2.6) and (2.7) to estimate the time-varying state variables and input variables (covariates) as well as their derivatives in the first stage, and then substitute the estimates from the first stage to model (2.5) to form a parametric regression model to estimate unknown parameters in the second stage.

2.1. Stage I: nonparametric estimation of state variables and their derivatives

In the first stage, we fit the NPME model (2.6) and (2.7) to obtain the estimates of the time-varying state variables, $X_i(t)$, and the input variables, $Z_i(t)$, as well as their derivatives for each individual subject using the local polynomial approach (Wu and Zhang (2002)). Here we only present the results for the state variable $X_i(t)$, it is the same for the input variable $Z_i(t)$. For convenience, we assume that the unobserved random-effect functions $v_i(t)$ are i.i.d. sampling trajectories of the underlying Gaussian process (GP) with mean function 0 and a covariance function $\gamma(s,t)$. We also assume that the measurement errors $\epsilon_i(t)$ are i.i.d. copies of a GP $\epsilon(t)$ with mean 0 and a covariance function $\gamma_\epsilon(s,t) = \sigma^2(t)I(s=t)$, where $I(\cdot)$ is an index function. For $i = 1, \ldots, n$, we have

$$v_i \sim GP(0, \gamma), \qquad \epsilon_i \sim GP(0, \gamma_\epsilon). \tag{2.8}$$

Let $t_{ij}, j = 1, \ldots, n_i$ be the design time points for the ith individual subject. The model (2.6) can be rewritten as

$$Y_{ij} = u(t_{ij}) + v_i(t_{ij}) + \epsilon_i(t_{ij}), \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i. \tag{2.9}$$

Assume that $u(t)$ and $v_i(t)$ are smooth functions and have up to $(p+1)$-order continuous derivatives at each time point within some interval of interest, $v_i^{(r)}(t)$ with covariance function $\gamma_r(t,s)$, for each $r = 1, \ldots, p$. Then, for each t_{ij}, $u(t_{ij})$ and $v_i(t_{ij})$ can be approximated by p-th degree polynomials within a neighborhood of t_0, i.e.,

$$u(t_{ij}) \approx u(t_0) + u'(t_0)(t_{ij} - t_0) + \ldots + \frac{u^{(p)}(t_0)}{p!}(t_{ij} - t_0)^p = H_{ij,p}^\tau(t_0)\beta, \tag{2.10}$$

$$v_i(t_{ij}) \approx v_i(t_0) + v_i'(t_0)(t_{ij} - t_0) + \ldots + \frac{v_i^{(p)}(t_0)}{p!}(t_{ij} - t_0)^p = H_{ij,p}^\tau(t_0)b_i, \tag{2.11}$$

where $H_{ij,p}(t_0) = (1, (t_{ij} - t_0), \ldots, (t_{ij} - t_0)^p)^\tau$, and

$$\beta = \Big[u(t_0), u'(t_0), \ldots, \frac{u^{(p)}(t_0)}{p!}\Big]^\tau, \quad b_i = \Big[v_i(t_0), v_i'(t_0), \ldots, \frac{v_i^{(p)}(t_0)}{p!}\Big]^\tau. \tag{2.12}$$

Thus, within a neighborhood of t_0, the NPME model (2.9) can be approximated by a LME model

$$Y_{ij} = H_{ij,p}^\tau(t_0)\beta + H_{ij,p}^\tau(t_0)b_i + \epsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i, \tag{2.13}$$

where ϵ_{ij} include the measurement and approximation errors, and b_i are the random effects. Let $K_h(\cdot) = K(\cdot/h)/h$, K a kernel function and h denoting bandwidth. For model (2.13), Wu and Zhang (2002) used a local likelihood method to give the estimator of β and predictor of b_i. Also they pointed out that their method is equivalent to fitting the standard LME model

$$\tilde{Y}_{ij} = \tilde{H}_{ij}^{\tau}(t_0)\beta + \tilde{H}_{ij}^{\tau}(t_0)b_i + \epsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i, \tag{2.14}$$

where $\tilde{Y}_{ij} = K_h^{1/2}(t_{ij} - t_0)Y_{ij}$ is the response variable, $\tilde{H}_{ij,p}(t_0) = K_h^{1/2}(t_{ij} - t_0)H_{ij,p}(t_0)$ are fixed-effects and random-effects covariates. Standard statistical software packages such as the R *lme* function or SAS procedure PROC MIXED can be used to fit (2.14).

Denote the final estimation of the state variable for individual subjects as $\hat{X}_i(t) = \hat{u}(t) + \hat{v}_i(t)$ and its derivative as $\hat{X}_i'(t) = \hat{u}'(t) + \hat{v}_i'(t)$. By Proposition 1 in Wu and Zhang (2002), we have

$$\begin{cases} \hat{u}(t) = \mathbf{e}_1^{\tau}\{\sum_{i=1}^{n}(I + G_iD)^{-1}G_i^{-1}\}^{-1} \times \sum_{i=1}^{n}(I + G_iD)^{-1}\psi_i, \\ \hat{u}'(t) = \mathbf{e}_2^{\tau}\{\sum_{i=1}^{n}(I + G_iD)^{-1}G_i^{-1}\}^{-1} \times \sum_{i=1}^{n}(I + G_iD)^{-1}\psi_i, \\ \hat{v}_i(t) = \mathbf{e}_1^{\tau}D(I + G_iD)^{-1}\mathbf{g}_i, \\ \hat{v}_i'(t) = \mathbf{e}_2^{\tau}D(I + G_iD)^{-1}\mathbf{g}_i, \end{cases} \tag{2.15}$$

where $\mathbf{e}_k^{\tau}$ is a $(p+1)$ vector with 1 at the kth element and 0 otherwise;

$$G_i = \begin{pmatrix} s_{i,0} & s_{i,1} & \cdots & s_{i,p} \\ s_{i,1} & s_{i,2} & \cdots & s_{i,p+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_{i,p+1} & s_{i,p+2} & \cdots & s_{i,2p} \end{pmatrix}, \psi_i = \begin{pmatrix} \psi_{i,0} \\ \psi_{i,1} \\ \vdots \\ \psi_{i,p} \end{pmatrix}, \mathbf{g}_i = \begin{pmatrix} g_{i,0} \\ g_{i,1} \\ \vdots \\ g_{i,p} \end{pmatrix}; \tag{2.16}$$

$$s_{i,r} = \sum_{j=1}^{n_i} \frac{K_h(t_{ij} - t)(t_{ij} - t)^r}{\sigma^2(t_{ij})}, \quad r = 0, \ldots, 2p; \tag{2.17}$$

$$\psi_{i,r} = \sum_{j=1}^{n_i} \frac{K_h(t_{ij} - t)(t_{ij} - t)^r Y_{ij}}{\sigma^2(t_{ij})}, \quad r = 0, \ldots, p; \tag{2.18}$$

$$\mathbf{g}_{i,r} = \sum_{j=1}^{n_i} K_h(t_{ij} - t)(t_{ij} - t)^r \frac{[Y_{ij} - H_{ij,r}^{\tau}(t)\hat{\beta}]}{\sigma^2(t_{ij})}, \; r = 0, \ldots, p. \tag{2.19}$$

Here we use local linear smoothing ($p = 1$) to estimate the curve functions $u(t)$ and $v_i(t)$ and local quadratic kernel ($p = 2$) to estimate the derivative functions $u'(t)$ and $v_i'(t)$. Similarly we can obtain the estimates of the input variable and its derivative, $\hat{Z}_i(t)$ and $\hat{Z}_i'(t)$.

For bandwidth selection, there are the criteria of leave-one-subject-out cross validation (SCV, Rice and Silverman (1991)) and leave-one-point-out cross validation (PCV). Wu and Zhang (2002) compared four bandwidth selection strategies. They concluded that when estimating the individual curve, the best method is the bias-corrected hybrid bandwidth (BCHB) and the second best is the hybrid bandwidth (HB) method, i.e., using SCV for population curve estimate and PCV for random-effects curve estimate. Readers are referred to Wu and Zhang (2002) for details. Due to the heavy computational burden of BCHB, we prefer the HB method here. That is, we select $\hat{h}_{01}$ for $u(t)$ by SCV and choose $\hat{h}_{02}$, the bandwidth for the random effects curves $v_i(t)$, by PCV. The consistency of SCV has been proved by Hart and Wehrly (1993). Thus the $\hat{h}_{01}$ is of order $\tilde{N}^{-1/5}$ where $\tilde{N} = \sum_1^n n_i$, and $\hat{h}_{02}$ is of order $n_i^{-1/5}$. However, our asymptotic theories in Section 3 (Condition C8) requires the bandwidth to be of order $O_p(n^{-1/4}a_n)$, where a_n is a sequence going to 0 with a slower rate than $log^{-1}(n)$. Thus, to meet this requirement, we need to use modified bandwidths, $h_{01} = \hat{h}_{01} \times n^{-1/20} \times \tilde{m}^{1/5}a_n$ and $h_{02} = \hat{h}_{02} \times n^{-1/4} \times \tilde{m}^{1/5}a_n$, where $\tilde{m} = n/\sum_{i=1}^n(1/n_i)$, which is at the same order as n_i, and $a_n = log^{-\varsigma}(n)$ with ς being a positive number less than 1. Since the constants in the limiting bandwidths are unknown, the asymptotic order of the bandwidths just provides us with a rough guideline for determining the practical bandwidths. Besides, for estimation of derivatives of a curve under the framework of a nonparametric mixed-effects model, the issue of bandwidth selection is not well resolved. It is an interesting research topic worthy of more attention in the future. We adopt an ad hoc approach: apply the local quadratic polynomial approach to fit the NPME model, and use the SCV and PCV method to choose the bandwidths, say $\hat{h}_{11}$ and $\hat{h}_{12}$, to optimize the estimates of $u(t)$ and $v_i(t)$. Then, to satisfy condition (C8), we use the modified bandwidths, $h_{11} = \hat{h}_{11} \times n^{-1/20} \times \tilde{m}^{1/5}a_n$ and $h_{12} = \hat{h}_{12} \times n^{-1/4} \times \tilde{m}^{1/5}a_n$ to estimate $u'(t)$ and $v_i'(t)$. Similar ideas of partially data-driven bandwidth selection were also used in Carroll et al. (1997) and Stute and Zhu (2005) for cross-section iid data.

2.2. Stage II: ODE parameter estimation

The idea for ODE parameter estimation in the second stage is simple. We substitute the estimates of $X_i(t)$, $X_i'(t)$, $Z_i(t)$, and $Z_i'(t)$ from the first stage into the ODE model, say model (2.5), to formulate a regression-like model for unknown parameters, and then apply regression approaches to obtain the parameter estimates. However, the justification for the estimator may not be trivial since the formulated regression models are not standard, instead both sides of the models are functions of the nonparametric estimates from the first stage rather than the measurement data. The theoretical properties for the final estimator need to be rigorously established. In this subsection, we derive the maximum pseudo-likelihood estimation (MPLE) method.

First we employ the linear mixed-effects regression approach (Davidian and Giltinan (1995), Vonesh and Chinchilli (1996)) to derive the parameter estimates for model (2.5). We assume that the $\hat{X}_i'(t_{ij})$ are obtained by the local polynomial mixed-effects (LP-MIX) approach from the first stage, as introduced in the last subsection. We substitute $\hat{X}_i'(t_{ij})$ into (2.5) and it follows that

$$\hat{X}_i'(t_{ij}) = F(X_i(t_{ij}), Z_i(t_{ij}), Z_i'(t_{ij}))^\tau \theta_i + \Delta_i(t_{ij}), \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i, \tag{2.20}$$

where $\theta_i = \theta + b_{\theta,i}$; we assume that $b_{\theta,i} \sim N(0, D_\theta)$. By (2.15), the error term is

$$\begin{aligned} \Delta_i(t_{ij}) &= \hat{X}_i'(t_{ij}) - X_i'(t_{ij}) \\ &= \mathbf{e}_2^\tau \{\sum_{i=1}^n (I + G_i D)^{-1} G_i^{-1}\}^{-1} \\ &\quad \times \sum_{i=1}^n (I + G_i D)^{-1} \psi_i + \mathbf{e}_2^\tau D (I + G_i D)^{-1} \mathbf{g}_i - u'(t_{ij}) - v_i'(t_{ij}). \end{aligned} \tag{2.21}$$

In order to derive the estimates of the unknown parameters in model (2.20), we need to study the properties of the error term $\Delta_i(t)$. Let $\mathcal{D} = \{t_{ij}, i = 1, \ldots, n, j = 1, \ldots, n_i\}$ denote the collection of design time points. Let $B(K) = \int K(t) t^2 dt$, $V(K) = \int K^2(t) dt$, $E(K) = \int K(t) t^4 dt$, $C(K) = \int K^2(t) t^2 dt$. We prove the following lemma in the Appendix.

Lemma 2.1. *Under conditions* (C1) *to* (C9) *in Section* 3*, we have*

$$E[\Delta_i(t)|\mathcal{D}] = \frac{h_{11}^2 u^{(3)}(t) E(K)}{3! B(K)} + o_p(h_{11}^2), \tag{2.22}$$

$$\text{Var}\,[\Delta_i(t)|\mathcal{D}] = \frac{\tau^2(t) C(K)}{n_i h_{12}^3 B^2(K) f(t)} + o_p[(n_i h_{12}^3)^{-1}], \tag{2.23}$$

where h_{11} and h_{12} denote the bandwidths for estimating the derivatives $u'(t)$ and $v_i'(t)$, respectively.

Let $\tau^2(t) = \gamma(t,t) + \sigma^2(t)$. Note that the $v_i(t)$'s and $\epsilon_i(t)$'s are mutually independent Gaussian processes. The random vector $(Y_{11}, \ldots, Y_{1n_1}, \ldots, Y_{n1}, \ldots, Y_{nn_i}, v_i'(t_{ij}))^\tau$ is normally distributed. Let $\mathbf{\Delta_i} = (\Delta_i(t_{i1}), \ldots, \Delta_i(t_{in_i}))^\tau$, by (2.21) a linear transformation of $(Y_{11}, \ldots, Y_{1n_1}, \ldots, Y_{n1}, \ldots, Y_{nn_i}, v_i'(t_{ij})^\tau$, so $\mathbf{\Delta}_i$ is normally distributed. For $i \neq j$, $\mathbf{\Delta}_i$ and $\mathbf{\Delta}_j$ are not independent, but are asymptotically independent, defined as follows (Lahiri (2003), Hürlimann (2004), Draisma et al. (2004)).

Definition 2.1. Two sequences of random vectors $\{U_n\}$ in R^p and $\{W_n\}$ in R^q, defined on a common probability space, are asymptotically independent if there exist constants $l_n > 0$, $s_n > 0$, and vectors $\mu_n \in R^p$ and $\omega_n \in R^q$ such that the random vector $(l_n[U_n - \mu_n]^\tau, s_n[W_n - \omega_n]^\tau)$ converges in distribution to some random vector (U, W), and U and W are independent.

Lemma 2.2. *Under the observation design $\mathcal{D}$ and conditions* (C1) *to* (C9) *in Section* 3, $\mathbf{\Delta}_i$ *and* $\mathbf{\Delta}_j$ *are asymptotically independent for* $i \neq j$.

By Lemmas 2.1 and 2.2, $\{\mathbf{\Delta}_i, i = 1, \ldots, n\}$ are normal vectors which are mutually and asymptotically independent with asymptotic mean 0. Let $\mathbf{X}_i = (X_i(t_{i1}), \ldots, X_i(t_{in_i}))^\tau$, $\mathbf{F}_i = \{F(X_i(t_{i1}), Z_i(t_{i1}), Z_i'(t_{i1})), \ldots, F(X_i(t_{in_i}), Z_i(t_{in_i}), Z_i'(t_{in_i}))\}^\tau$, $\mathbf{Z}_i = (Z_i(t_{i1}), \ldots, Z_i(t_{in_i}))^\tau$, and $\mathbf{Z}_i' = (Z_i'(t_{i1}), \ldots, Z_i'(t_{in_i}))^\tau$. Since $\{\mathbf{\Delta}_i, i = 1, \ldots, n\}$ are asymptotically independent, we define the pseudo-likelihood for model (2.20) as the product of the likelihoods for $\mathbf{\Delta}_i$'s:

$$
\begin{aligned}
&PL(\mathbf{\Delta}_i, b_{\theta,i}|\hat{\mathbf{X}}_i') \\
&= (2\pi)^{-n}\prod_{i=1}^n |R_i|^{-1/2}\prod_{i=1}^n \exp\Big\{-\frac{1}{2}[\hat{\mathbf{X}}_i' - \mathbf{F}_i\theta - \mathbf{F}_i b_{\theta,i}]^\tau \mathbf{R}_i^{-1}[\hat{\mathbf{X}}_i' - \mathbf{F}_i\theta - \mathbf{F}_i b_{\theta,i}]\Big\} \\
&\quad |D_\theta|^{-n/2}\prod_{i=1}^n \exp\Big(-\frac{1}{2}b_{\theta,i}^\tau D_\theta^{-1} b_{\theta,i}\Big), \qquad (2.24)
\end{aligned}
$$

where $D_\theta = \mathrm{Var}\,(b_{\theta,i})$ and $\mathbf{R}_i = \mathrm{Var}\,(\mathbf{\Delta}_i)$. The pseudo-likelihood idea has been used under different frameworks by others (Besag (1974, 1977), Troxel, Lipsitz, and Harrington (1998)), in which the correlations among dependent variables are ignored.

If $\mathbf{X}_i$, $\mathbf{Z}_i$, and $\mathbf{Z}_i'$ are exactly known, the estimate of θ and the predictor of $b_{\theta,i}$ are the maximizers of (2.24), that can be obtained as

$$
\hat{\theta} = \Big(\sum_{i=1}^n \mathbf{F}_i^\tau \mathbf{V}_i^{-1}\mathbf{F}_i\Big)^{-1}\sum_{i=1}^n \mathbf{F}_i^\tau \mathbf{V}_i^{-1}\hat{\mathbf{X}}_i', \qquad (2.25)
$$

$$
\hat{b}_{\theta,i} = D_\theta \mathbf{F}_i^\tau \mathbf{V}_i^{-1}(\hat{\mathbf{X}}_i' - \mathbf{F}_i\hat{\theta}). \qquad (2.26)
$$

However, in the above expressions, the $\mathbf{X}_i$, $\mathbf{Z}_i$, and $\mathbf{Z}_i'$ are not exactly known but measured with error. Then we can substitute their estimates from the first stage, $\hat{\mathbf{X}}_i$, $\hat{\mathbf{Z}}_i$ and $\hat{\mathbf{Z}}_i'$ into (2.25) and (2.26). The final maximum pseudo-likelihood estimates (MPLE) are

$$
\hat{\theta} = \Big(\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1}\hat{\mathbf{F}}_i\Big)^{-1}\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1}\hat{\mathbf{X}}_i', \qquad (2.27)
$$

$$
\hat{b}_{\theta,i} = D_\theta \hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1}(\hat{\mathbf{X}}_i' - \hat{\mathbf{F}}_i\hat{\theta}), \qquad (2.28)
$$

where $\hat{\mathbf{F}}_i = \{F(\hat{X}_i(t_{i1}), \hat{Z}_i(t_{i1}), \hat{Z}_i'(t_{i1})), \ldots, F(\hat{X}_i(t_{in_i}), \hat{Z}_i(t_{in_i}), \hat{Z}_i'(t_{in_i}))\}^\tau$ and $\hat{\mathbf{V}}_i = \hat{\mathbf{F}}_i D_\theta \hat{\mathbf{F}}_i^\tau + \mathbf{R}_i$. This plug-in approach for unknown state variables and their derivatives is similar to the estimation procedure in Gong and Samaniego (1981), Liang, Wu, and Carroll (2003), Wu and Liang (2004), and Liang and Wu (2008).

The computation procedure is the same as that to obtain $\hat{\theta}$ and $\hat{b}_{\theta,i}$ as the maximum likelihood estimation for the standard LME model. For implementation, we can directly use R function *lme* or SAS procedure PROC MIXED for the model

$$\hat{\mathbf{X}}_i' = \hat{\mathbf{F}}_i\theta + \hat{\mathbf{F}}_i b_{\theta,i} + \mathbf{\Delta}_i,\ i = 1, \ldots, n. \tag{2.29}$$

Note that we call our estimate the maximum pseudo-likelihood estimate (MPLE). The word "pseudo" has a two-layer meaning here: the likelihood function is formulated using the concept of asymptotic independence instead of exact independence; the unknown time-varying functions are replaced by their nonparametric estimates.

3. Asymptotic Properties

In this section, we study the asymptotic properties of the MPLE of the population parameters. For notational simplicity, we ignore the input or covariate variables $Z_i(t)$ in the theoretical development, i.e., we consider the model

$$X_i'(t_{ij}) = F(X_i(t_{ij}))^\tau(\theta + b_{\theta,i}). \tag{3.1}$$

Results still hold and the proofs are similar when the model contains the input or covariate variables $Z_i(t_{ij})$ and their derivatives $Z_i'(t_{ij})$. First we introduce the following conditions.

C1 The design points t_{ij}, $j = 1, \ldots, n_i$, $i = 1, \ldots, n$ are *i.i.d.* variables with a density function $f(t)$.

C2 The time point t is in the interior of f where $f(t) \neq 0$ and $f'(t)$ exists.

C3 The curves of the fixed and random effects $u(t)$ and $v_i(t), i = 1, \ldots, n$, have third order continuous derivatives at t.

C4 The covariance functions $\gamma(s,t)$ of $v_i(t)$ and $\gamma_1(s,t)$ of $v_i'(t)$ have twice-continuous derivatives in s and t.

C5 The variance function of the measurement error $\sigma^2(t)$ is continuous at t.

C6 The kernel function K is a bounded symmetric density function with support $[-1,1]$ satisfying $\int_{-1}^1 K(t)dt = 1$, $\int_{-1}^1 tK(t)dt = 0$. If $B(K) = \int K(t)t^2dt$, $V(K) = \int K^2(t)dt$, $E(K) = \int K(t)t^4dt$, $C(K) = \int K^2(t)t^2dt$, then $B(K)$, $V(K)$, $C(K)$ and $E(K) < +\infty$.

C7 As $n \to +\infty$, $n_i \to +\infty$, $n_i h_{12}^3 \to +\infty$, $n_i h_{02} \to +\infty$.

C8 The bandwidths h_{01}, h_{02}, h_{11}, and h_{12} are of order $O_p(n^{-1/4}a_n)$ for the LP-MIX estimate in Stage I, where a_n is a sequence tending to 0 at a rate slower than $log^{-1}(n)$.

C9 D is a diagonal matrix, say $D = diag(d_1^2(t), d_2^2(t))$ when $p = 1$ and $D = diag(d_1^2(t), d_2^2(t), d_3^2(t))$ when $p = 2$, with $d_r^2(t) > 0$ for $r = 1, 2, 3$.

C10 The matrices $\mathbf{F}_i$ and $\hat{\mathbf{F}}_i$ are of full rank in column. The first and second derivatives of the function F, $\partial F(x)/\partial x$ and $\partial^2 F(x)/\partial x^2$ exist and are continuous for x in its domain χ. There exists a positive constant M, for $x \in \chi$, with

$$\sup_{x\in\chi}\left|\frac{\partial F(x)}{\partial x}\right| \le M.$$

The conditions (C1) to (C6) and (C10) are commonly used. Conditions (C7) and (C8) guarantee the asymptotic normality of $\hat{\theta}$ in Theorem 3.2. Condition (C9) is set for technical convenience. By (C7) and (C8), we can see that if we have $n_i = O(n^{\omega})$, then $\omega > 3/4$.

Lemma 3.1. *For the LP-MIX estimator in the first stage, under* (C1) *to* (C9) *we have*

$$E[\Lambda_i(t)|\mathfrak{D}] = \frac{h_{01}^2 u^{(2)}(t) B(K)}{2} + o_p(h_{01}^2),\ \mathrm{Var}\,[\Lambda_i(t)|\mathfrak{D}] = \frac{\tau^2(t) V(K)}{n_i h_{02} f(t)} + o_p[(n_i h_{02})^{-1}],$$

where $\tau^2(t) = \gamma(t,t) + \sigma^2(t)$ *and* $\Lambda_i(t) = \hat{X}_i(t) - X_i(t)$.

The proofs of next results are in the Appendix.

Theorem 3.1. *Under the conditions* (C1) *to* (C10)*, the population parameter estimator* $\hat{\theta}$ *is a consistent estimator of the parameter* θ.

Theorem 3.2. *Under the conditions* (C1) *to* (C10)*,*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, D_\theta), \tag{3.2}$$

where $D_\theta = \mathrm{Var}\,(b_{\theta,i})$.

Remark 3.1. The convergence rate of the proposed MPL estimator in Theorem 3.2 is the standard root-n, where n is the number of subjects or clusters in the longitudinal study. This rate is the same as that of the standard LME model (Vonesh and Chinchilli (1996)), but here we assume that the number of measurements for each individual subject or cluster $\to \infty$. This, combined with our model feature of the same covariate for both population parameters and random components of the parameters, results in a simpler asymptotic variance-covariance matrix compared to that of the standard LME model (see details of the proof in Appendix). Theorem 3.2 indicates that the asymptotic variance of the proposed MPL estimator of the population parameters only depends on the between-subject variation of the longitudinal data.

Remark 3.2. Note that if additional input variables or covariates, $Z_i(t_{ij})$ and $Z_i'(t_{ij})$ exist as in model (2.5), the above results still hold. We only need to make similar assumptions to (C3) and (C4) for $Z_i(t_{ij})$, and obtain the estimates of $\hat{Z}_i(t)$ and $\hat{Z}_i'(t)$ by using the LP-MIX with bandwidths of order $n^{-1/4}a_n$ as specified in the condition (C8). The proofs of the consistency and asymptotic normality of the parameter estimates are similar.

4. Data Analysis

We fit the random coefficient ODE model (2.3) to an AIDS clinical data set to further illustrate the usefulness of the proposed methods. The viral load and numbers of CD4+ T cells from four patients were frequently measured in this study after initiating an antiretroviral regimen. Viral load was measured at 13 time points during the first day, 14 measurements from day 2 to week 2, and then one measurement at each of every four weeks, for all subjects. The measurements of total CD4+ T cell counts were also scheduled at week 2 and every four weeks. Note that the observed viral load and concentration of CD4+ T cells are modeled by $V_i(t_{ij})$ and $T_i(t_{ij})$ with

$$V_i(t_{ij}) = V(t_{ij}) + V_i^*(t_{ij}) + \epsilon_{1i}(t_{ij}), \quad T_i(t_{ij}) = T(t_{ij}) + T_i^*(t_{ij}) + \epsilon_{2i}(t_{ij}), \quad (4.1)$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$, where $V(t)$ and $T(t)$ model population-specific mean functions, $V_i^*(t)$ and $T_i^*(t)$ model subject-specific variations from the mean curve functions, and $\epsilon_{1i}(t)$ and $\epsilon_{2i}(t)$ are measurement errors.

In order to estimate the dynamic parameters in the original model (2.1), we first apply the nonlinear regression approach and the model suggested in Perelson et al. (1996) to estimate the parameters δ_i and c_i for all subjects. Then we assume that these two parameters are known and we focus on the estimation of parameters λ, ρ, and N in model (2.3). Assume that the underlying decomposition of random coefficients α_{0i}, α_{1i}, and α_{2i}, are, respectively,

$$\alpha_{0i} = \alpha_0 + b_{0i}, \qquad \alpha_{1i} = \alpha_1 + b_{1i}, \qquad \alpha_{2i} = \alpha_2 + b_{2i}, \qquad (4.2)$$

where α_0, α_2, and α_2 are fixed effects, and b_{0i}, b_{1i}, and b_{2i} are random effects. From (2.3), we first estimate the population parameters α_0, α_1, and α_2 and individual parameters α_{0i}, α_{1i}, and α_{2i} using the proposed MPL method. Then

Table 1. The estimates (95% confidence intervals) of λ_i, ρ_i and N_i for individual patients and the fixed effects λ, ρ, and N using the MPL method.

Patient	λ_i cells/day			ρ_i day^{-1}			N_i virions/cell		
1	37.8	(29.43,	272.37)	0.043	(0.034,	0.308)	497.2	(63.53,	776.38)
2	21.9	(13.17,	157.56)	0.043	(0.026,	0.323)	447.9	(24.16,	981.60)
3	43.5	(21.78,	296.14)	0.047	(0.022,	0.324)	553.6	(45.96,	1189.65)
4	28.8	(21.69,	224.34)	0.047	(0.035,	0.323)	474.4	(33.44,	713.35)
fixed effects	λ			ρ			N		
	33.5	[22.83,	215.27]	0.045	[0.030,	0.291]	497.2	[33.44,	740.45]

we recover the estimates of $(\lambda,\ \rho,\ N)$ using the relationships (Liang and Wu (2008)): $\hat{\lambda}_i = -\hat{\alpha}_{0i}/\hat{\alpha}_{2i}$, $\hat{\rho}_i = \hat{\alpha}_{1i}/\hat{\alpha}_{2i}$, $\hat{N}_i = \hat{\alpha}_{1i}/\delta_i - \hat{\alpha}_{2i}$, and for population parameters $\hat{\lambda} = -\hat{\alpha}_0/\hat{\alpha}_2$, $\hat{\rho} = \hat{\alpha}_1/\hat{\alpha}_2$, $\hat{N} = \hat{\alpha}_1/\delta - \hat{\alpha}_2$.

Some outliers and the early data due to the shoulder effect (Wu and Ding (1999)) were excluded from the model fitting. We report the estimation results for both population (fixed-effects) and individual parameters in Table 1. The 95% bootstrap confidence interval estimates are also given in the table.

From Table 1, we can see that the MPL estimates for individual patients are close to each other, which is not surprising since the mixed-effects model shrinks the estimates toward the mean. The estimation results show that the proliferation rate of uninfected CD4+ T cells is about 34 cells per day per *ml* blood, the death rate of uninfected CD4+ T cells is some 0.045 with a corresponding half-life of about 15 days, and the virus burst size (the number of virions produced per infected cell) is about 497. These results are consistent with some of the earlier estimates in the literature (Hellerstein et al. (1999), Haase et al. (1996)). To our best knowledge, investigators have not been able to estimate the HIV viral dynamic parameters from clinical data directly. Liang and Wu (2008) have made an effort to estimate the parameters from two patients of this study, but they were not able to estimate these parameters for all patients due to insufficient data and convergence problem for some patients. In contrast, the proposed method allows us to borrow information across-subjects using mixed-effects modeling approach so that we can estimate these parameters for all patients in the study. This is important for individualizing treatment and supporting clinical decisions.

5. Simulation Study

In this section, we report on the evaluation of the performance of the proposed MPL estimate using Monte Carlo simulations based on the HIV dynamic model (2.3). To evaluate the estimation performance, we take the average relative

Table 2. The AREs (%) of the MPL estimates for individual parameters α_{0i}, α_{1i}, α_{2i} and fixed effects α_0, α_1, α_2 for the first simulation scheme in Section 5. $(n = 4, n_i = 36)$.

γ_1^*	γ_2^*	α_{0i}	α_{1i}	α_{2i}	α_0	α_1	α_2
100	10	5.68	34.17	22.93	4.81	32.17	22.84
400	10	5.61	34.12	22.99	4.89	32.73	22.89
100	20	6.37	35.22	22.36	5.74	34.43	22.62
400	20	6.30	35.18	22.42	5.67	34.80	22.32

estimation error (ARE) of θ as

$$ARE = \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{|\hat{\theta} - \theta|}{|\theta|} \times 100\%,$$

where $\hat{\theta}$ is the estimate of θ and N_s is the number of simulation runs. Two simulation schemes were designed as suggested by the referee.

Simulation Scheme 1. To get the performance of the proposed MPL estimate in the application in Section 4, the first simulation experiment mimics the study. We used the estimated parameters and the estimated initial values $\hat{V}_i(0)$ and $\hat{T}_i(0)$ for individual subjects in Section 4 to generate data. Since we can only estimate $T_i(0)$, we assumed that the ratio of $T_{Ui}(0)$ and $T_{Ii}(0)$ is $1:10$ to obtain the initial values for $T_{Ui}(0)$ and $T_{Ii}(0)$. The time varying infection rate $\eta(t)$ of T cells was set at $9 \times 10^{-5}(1 - 0.9\cos(\pi t/100))$. The measurement errors $\epsilon_{1i}(t)$ and $\epsilon_{2i}(t)$ were independently generated from normal distributions with means zero and variances $\gamma_{\epsilon 1}(t,t) = 400$ and 100 and $\gamma_{\epsilon 2}(t,t) = 20$ and 10, respectively. We generated our data by numerically solving the ODE system (2.1) via the fourth-order Runge-Kutta algorithm, and then measurement errors were added to the simulated data $V_i(t)$ and $T_i(t)$. For simplicity, we took 36 measurements for each of the four patients as designed by this application study. We simulated 500 data sets and applied the proposed MPL estimation method to these simulated data sets. Note that the Epanechnikov kernel $K(u) = (3/4)(1 - u^2)I(|u| \leq 1)$ and the bandwidth selection method proposed in Section 2.1 were used in the estimation. The fixed-effect parameters α_0, α_1, α_2 and the individual parameters α_{0i}, α_{1i}, α_{2i} in (4.2) were estimated and their ARE's are reported in Table 2.

From Table 2, we can see that the AREs of the MPL estimates for fixed effects and individual parameters, ranging from 5% to 35%, are quite reasonable. This suggests that our estimates of these kinetic parameters for these patients are reliable.

Simulation Scheme 2. To further illustrate the performance for different sample sizes, we designed the second simulation scheme as follows. First we chose the fixed-effect parameters as $(\alpha_0, \alpha_1, \alpha_2)$ = (45,918, -138, -1,276), then the random effects of parameters (b_{0i}, b_{1i}, b_{2i}) were generated independently from a multivariate normal distribution with mean 0 and variance $diag(10^4, 4, 36)$, based on the data analysis results in Section 4. Thus the parameters α_{0i}, α_{1i}, α_{2i}, and c_i for individual subjects in model (2.3) were obtained by adding the fixed effects and random effects, together as in (4.2) . Secondly, we calculated the parameters λ_i, ρ_i, and N_i for the i-th individual in original ODE model (2.1) based on their relationships with α_{0i}, α_{1i}, α_{2i} and c_i (see Section 4). The fixed effects of initial values for the state variables $(T_U(0), T_I(0), V(0))$ were set as $(30, 600, 10^5)$. The random effects of initial values $(b_{Ui}(0), b_{Ii}(0), b_{Vi}(0))$ were independently generated from a multivariate normal distribution with mean 0 and variance $diag(4, 40, 10^6)$. The time-varying infection rate of T cells for each individual was $\eta_i(t) = 9 \times 10^{-5}(1 - 0.9\cos(\pi t/1,000))$. The death rate of infected CD4+ T cells is fixed as $\delta_i = 0.5$/day and the clearance rate of virus is fixed as $c_i = 3$ based on the estimates from the literature (Perelson et al. (1996)). The measurement errors $\epsilon_{1i}(t)$ and $\epsilon_{2i}(t)$ are independently generated from normal distributions with means zero and variances $\gamma_{\epsilon 1}(t,t) = \gamma_1^*[1 + 0.75\cos(t/40)]$ and $\gamma_{\epsilon 2}(t,t) = \gamma_1^*[1+0.1\sin(t/80)]$ with $\gamma_1^* = (50, 200)$ and $\gamma_2^* = (10, 40)$, respectively. Finally, we simulated our data by numerically solving the ODE system (2.1) via the fourth-order Runge-Kutta algorithm, and then the measurement errors were added to the simulated data $V_i(t)$ and $T_i(t)$.

We chose three different sample sizes for the simulated data: a) $n = 8$, $n_i = 30$+Poisson(8); b) $n = 8$, $n_i = 60$+Poisson(16); and c) $n = 16$, $n_i = 30$+Poisson(8), where a Poisson distribution was used to mimic the unbalanced data for individual subjects. The observation time points were equally-spaced for simplicity: $t_{ij} = 0.1 \times j$ with $j = 1, 2, \ldots, n_i$. We carried out $N_s = 500$ simulation runs. Similarly, we used the proposed MPL to estimate the fixed-effect parameters $(\alpha_0, \alpha_1, \alpha_2)$ and individual parameters $(\alpha_{0i}, \alpha_{1i}, \alpha_{2i})$. We report the ARE's of these estimates in Table 3. From Table 3, we can see that the overall performance for both population parameter estimates and individual parameter estimates were reasonably good for the proposed MPL method. For larger sample sizes, the performance of the estimator of fixed effects was better. Also, with n_i larger, the ARE's of the MPL estimator for both fixed and individual parameters are smaller, which suggests the importance of the sample size for individual subjects for the proposed method, since it is necessary to obtain good nonparametric estimates for the state variables in the first stage. The ARE's of individual parameter estimates are generally larger than those of fixed effects estimates, as expected. Additional simulations were performed for the data with

Table 3. The AREs (%) of the MPL estimates for individual parameters α_{0i}, α_{1i}, α_{2i} and fixed effects α_0, α_1, α_2 for the second simulation scheme in Section 5.

n	n_i	γ_1^*	γ_2^*	α_{0i}	α_{1i}	α_{2i}	α_0	α_1	α_2
8	30+Poisson(8)	50	10	11.23	14.03	10.51	7.56	11.85	8.85
		200	10	11.33	14.43	10.80	7.53	12.15	9.12
		50	40	18.63	25.01	19.73	13.39	20.57	17.10
		200	40	18.67	25.26	20.29	13.53	21.26	17.97
8	60+Poisson(16)	50	10	4.98	11.43	6.44	3.85	9.52	5.71
		200	10	4.99	11.44	6.43	3.83	9.58	5.74
		50	40	8.55	18.43	9.04	7.06	15.48	7.34
		200	40	8.65	18.66	9.18	7.16	15.59	7.42
16	30+Poisson(8)	50	10	7.51	10.00	8.27	5.99	9.59	7.46
		200	10	7.73	9.95	8.27	6.34	9.56	7.52
		50	40	9.14	17.38	11.69	6.59	15.76	9.75
		200	40	9.25	17.71	11.93	5.68	15.85	9.64

missing completely at random (MCAR) and with outliers, the results and conclusions are similar to the data with random early drop-out. These simulation results are not shown here.

6. Discussion

Differential equation models have been widely used to describe dynamic processes. However, the statistical literature is scant on the problems of parameter estimation and statistical inference for ODE models. Recently this field has started to attract more attention from the statistical research community (Putter et al. (2002), Lahiri (2003), Huang, Liu, and Wu (2006), Guedj, Thiébaut, and Commenges (2007) Ramsay et al. (2007), Chen and Wu (2008a,b) Liang and Wu (2008); Miao et al. (2009)). However, the statistical research for ODE models is still in its infancy and there are many unresolved methodological problems. In particular, the theory is not well established for many proposed statistical methods. We have extended the two-stage approach (Varah (1982), Chen and Wu (2008a,b), Liang and Wu (2008)) to fit longitudinal data to random coefficient ODE models. We derived the maximum pseudo-likelihood estimator (MPLE) and established its asymptotic properties.

The proposed method has some advantages that have been listed in the Introduction, but our approach also inherits some limitations of the two-stage methods. These include a reduced estimation efficiency since the estimated state variables from the first stage, instead of the data, are used in the ODE parameter estimation in the second stage. This is the price that the proposed method has to pay in order to achieve computational efficiency. In fact, the proposed method can be combined with the exact maximum likelihood approach (Guedj, Thiébaut, and Commenges (2007)) to improve the estimation efficiency, which is a worthy topic for future research.

We have worked under the assumption that the second stage parametric model is linear in unknown parameters so that the linear mixed-effects (LME) model can be used. However, the proposed estimation procedure is applicable to a general nonlinear model in the second stage. In this case, however, the nonlinear mixed-effects model (Davidian and Giltinan (1993), Davidian and Giltinan (1995), Vonesh and Chinchilli (1996)) instead of LME model, should be fitted in the second stage. The theoretical development is also more tedious in the nonlinear model case. We considered a single ODE model for notational simplicity and computational convenience, but our method can be generalized to multivariate ODE models. It is still an open question how to use the two-stage approach to deal with latent (unmeasurable) state variables. In this paper, we employed the local polynomial nonparametric approach in the first stage. In fact, many other nonparametric smoothing methods such as regression splines, smoothing splines, and penalized splines (Wu and Zhang (2006)) can be used to fit the nonparametric mixed-effects model in the first stage. Also note that the standard regression estimation for each individual subject, instead of a mixed-effects regression model, can be used to estimate the unknown parameters in the second stage, though this may fail if the data from some individual subjects are too sparse. Our methodological development is motivated by HIV dynamic studies, but we also expect that our method can be applied to other ODE models with longitudinal data.

Acknowledgement

Fang was supported by a NSF grant from the National Natural Science Foundation of China (No. 10701035), ChenGuang project of Shanghai Education Development Foundation (No. 2007CG33); Wu was partially supported by USA NIH grants AI055290, AI50020, AI078498, AI078842, and AI087135; and Zhu was supported by a grant (HKU 2030/07P) from the Research Grants Council of Hong Kong and a FRG grant from Hong Kong Baptist University, Hong Kong. This paper was completed while the first author was visiting the Department of Biostatistics & Computational Biology, University of Rochester School of Medicine and Dentistry as a visiting PhD student under the supervision of the second author and she appreciates the hospitality and support from the Department. The authors are also grateful to Drs. Hua Liang and Hongqi Xue for their helpful discussions on the proofs of the theoretical results. The authors thank an AE and a referee for their constructive comments and suggestions.

Appendix

Proof of Lemma 2.1. For estimating the derivatives $u'(t)$ and $v_i'(t)$, a local quadratic ($p = 2$) linear mixed-effects (LME) model is used, i.e.,

$$G_i = \begin{pmatrix} s_{i,0} & s_{i,1} & s_{i,2} \\ s_{i,1} & s_{i,2} & s_{i,3} \\ s_{i,2} & s_{i,3} & s_{i,4} \end{pmatrix} = \frac{f(t)}{\sigma^2(t)} \begin{pmatrix} n_i & 0 & n_i h^2 B(K) \\ 0 & n_i h^2 B(K) & 0 \\ n_i h^2 B(K) & 0 & n_i h^4 E(K) \end{pmatrix} \times [1 + O_p((n_i h)^{-1/2})]. \tag{A.1}$$

Since (C9), $D = diag(d_1^2(t), d_2^2(t), d_3^2(t))$, it follows that

$$G_i^{-1} D^{-1} = \frac{\sigma^2(t)}{f(t)} \begin{pmatrix} \frac{E(K) d_1^2(t)}{n_i(E(K) - B^2(K))} & 0 & n_i h^2 B(K) \\ 0 & \frac{d_2^2(t)}{n_i h^2 B(K)} & 0 \\ n_i h^2 B(K) & 0 & \frac{d_3^2(t)}{n_i h^4 (E(K) - B^2(K))} \end{pmatrix} \times [1 + O_p((n_i h)^{-1/2})]. \tag{A.2}$$

When different bandwidths, h_{11} and h_{12}, are used to estimate $u'(t)$ and $v_i'(t)$, the h in (A.1) and (A.2) can be replaced by h_{11} and h_{12}. We prove the lemma in the following three steps.

Step 1. By the proof of Theorem 1 in Wu and Zhang (2002), we have

$$\mathbf{e}_2^\tau \{\sum_{k=1}^{n} (I + G_k D)^{-1} G_k^{-1}\}^{-1} \times (I + G_i D)^{-1} G_i = \frac{1}{n}[1 + O_p(n_i^{-1})]\mathbf{e}_2^\tau. \tag{A.3}$$

Recall the expression for $\hat{u}(t)$ in (2.15). We have

$$\begin{aligned} &\hat{u}'(t) - u'(t) \\ &=: \Bigg\{ n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \mathbf{e}_2^\tau G_i^{-1} H_{ij,2}(t) \frac{K_{h,11}(t_{ij} - t)}{\sigma^2(t)} \\ &\quad \times \Big[u(t_{ij}) - u(t) - (t_{ij} - t) u'(t) - \frac{(t_{ij} - t)^2}{2} u^{(2)}(t) \Big] \\ &\quad + n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \mathbf{e}_2^\tau G_i^{-1} H_{ij,2}(t) \frac{K_{h,11}(t_{ij} - t)}{\sigma^2(t)} [v_i(t_{ij}) + e_i(t_{ij})] \Bigg\} [1 + o_p(1)] \\ &=: \left[\frac{h_{11}^2 u^{(3)}(t) E(K)}{3! B(K)} + n^{-1} \sum_{i=1}^{n} \xi_{1i}(t) \right] [1 + o_p(1)]. \end{aligned} \tag{A.4}$$

The $\{\xi_{1i}(t), i = 1, \ldots, n\}$ are independent. Here $H_{ij,2}(t) = (1, (t_{ij} - t), (t_{ij} - t)^2)^\tau$ and $K_{h,11}(\cdot) = h_{11}^{-1} K(\cdot / h_{11})$.

Step 2. By the proof of Theorem 1 in Wu and Zhang (2002), we have $\mathbf{e}_2^\tau D(I+G_iD)^{-1}G_i = \mathbf{e}_2^\tau + o_p(1)$. Then, based on $\hat{v}'(t)$ given in (2.15) in Section 2, we obtain that

$$\begin{aligned}
\hat{v}_i'(t) &= \sum_{j=1}^{n_i} \mathbf{e}_2^\tau G_i^{-1} H_{ij,2}(t) \frac{K_{h,12}(t_{ij}-t)}{\sigma^2(t_{ij})} \\
&\quad \times \left\{ Y_{ij} - H_{ij,2}^\tau(t) \frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{n_i} \mathbf{e}_2^\tau G_i^{-1} H_{ij,2}(t) \frac{K_{h,12}(t_{il}-t)}{\sigma^2(t_{il})} Y_{ij} \right\} \\
&= \left\{ \sum_{j=1}^{n_i} \mathbf{e}_2^\tau G_i^{-1} H_{ij,2}(t) \frac{K_{h,12}(t_{ij}-t)}{\sigma^2(t_{ij})} [v_i(t_{ij}) + e_i(t_{ij})] \right. \\
&\quad \left. - n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \mathbf{e}_2^\tau G_i^{-1} H_{ij,2}(t) \frac{K_{h,12}(t_{ij}-t)}{\sigma^2(t)} [v_i(t_{ij}) + e_i(t_{ij})] \right\} [1 + O_p(\tilde{m}^{-1})] \\
&=: \left[\xi_{2i}(t) - \frac{1}{n} \sum_{i=1}^{n} \xi_{2i}(t) \right] [1 + o_p(1)].
\end{aligned}$$

Step 3. In summary,

$$\Delta_i(t) = \left[\frac{h_{11}^2 u^{(3)}(t) E(K)}{3! B(K)} + \xi_{2i}(t) - v_i'(t) + n^{-1} \sum_{i=1}^{n} \xi_{1i}(t) - \frac{1}{n} \sum_{i=1}^{n} \xi_{2i}(t) \right] [1 + o_p(1)],$$

where $\{\xi_{1i}(t), i=1,\ldots,n\}$ and $\{\xi_{2i}(t), i=1,\ldots,n\}$ are series of i.i.d. random variables with mean 0. We have

$$\mathrm{Var}\,(\xi_{2i}(t)|\mathfrak{D}) = d_2^2(t) + \frac{\tau^2(t) C(K)}{n_i h_{12}^3 B^2(K) f(t)} + o_p[(n_i h_{12}^3)^{-1}]. \tag{A.5}$$

Consequently, $\mathrm{Var}\,(n^{-1}\sum_{i=1}^n \xi_{2i}(t) - n^{-1}\sum_{i=1}^n \xi_{2i}(t)|\mathfrak{D}) = O_p[(nn_ih^3)^{-1}]$. Under (C7), similarly, $n^{-1}\sum_{i=1}^n \xi_{2i}(t) - n^{-1}\sum_{i=1}^n \xi_{2i}(t) = O_p[(nn_ih^3)^{-1/2}]$. On the other hand,

$$\mathrm{Var}\,(\xi_{2i}(t) - v_i'(t|\mathfrak{D}) = \frac{\tau^2(t) C(K)}{n_i h_{12}^3 B^2(K) f(t)} + o_p[(n_i h_{12}^3)^{-1/2}]. \tag{A.6}$$

Therefore, by (C7), $n^{-1}\sum_{i=1}^n \xi_{1i}(t) - n^{-1}\sum_{i=1}^n \xi_{2i}(t)$ is of higher order than $\xi_{2i}(t) - v_i'(t)$. Finally,

$$\Delta_i(t) = \frac{h_{11}^2 u^{(3)}(t) E(K)}{3! B(K)} + \xi_{2i}(t) - v_i'(t) + o_p[h_{11}^2 + (n_i h_{12}^3)^{-1/2}], \tag{A.7}$$

By (A.7) and (A.6), we conclude that

$$\begin{aligned} E[\Delta_i(t)|\mathfrak{D}] &= \frac{h_{11}^2 u^{(3)}(t)E(K)}{3!B(K)} + o_p(h_{11}^2), \\ \mathrm{Var}\,[\Delta_i(t)|\mathfrak{D}] &= \frac{\tau^2(t)C(K)}{n_i h_{12}^3 B^2(K) f(t)} + o_p[(n_i h_{12}^3)^{-1/2}]. \end{aligned} \tag{A.8}$$

This completes the proof of Lemma 2.1.

Proof of Lemma 2.2. By (A.7) in the proof of Lemma 2.1, it is easy to derive that $Cov(\Delta_i(t_1), \Delta_j(t_2)|\mathfrak{D}) = o_p[(n_i h_{12}^3)^{-1}]$. Let $\mu(t) = (3!B(K)^{-1})h_{11}^2 E(K)$ $u^{(3)}(t)$. Note that $\Delta_i(t_1)$ and $\Delta_j(t_2)$ are normal vectors. Thus

$$\begin{aligned} &\left(\sqrt{n_i h_{12}^3}[\Delta_i(t_1) - \mu(t_1)], \sqrt{n_j h_{12}^3}[\Delta_j(t_2) - \mu(t_2)]|\mathfrak{D}\right) \\ &\xrightarrow{d} N\left(0, \frac{C(K)}{B^2(K)} diag\left(\frac{\tau^2(t_1)}{f(t_1)}, \frac{\tau^2(t_2)}{f(t_2)}\right)\right). \end{aligned}$$

By Definition 2.1, under the design $\mathfrak{D}$, $\Delta_i(t_1)$ and $\Delta_j(t_2)$ are asymptotically independent for $i \neq j$. Then it is obvious that the vectors $\boldsymbol{\Delta}_i$ and $\boldsymbol{\Delta}_j$ are asymptotically conditionally independent.

Proof of Lemma 3.1. Let $H_{ij,1}(t) = (1, t_{ij} - t)^\tau$. Similar to proof of Lemma 2.1, we have

$$\begin{aligned} \Lambda_i(t) &= \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \mathbf{e}_2^\tau G_i^{-1} H_{ij,1}(t) \frac{K_{h,01}(t_{ij} - t)}{\sigma^2(t)} \Big[u(t_{ij}) - u(t) - (t_{ij} - t)u'(t) \Big] \right. \\ &\quad \left. + \xi_{3i}(t) - v_i(t) \right\} [1 + o_p(1)] \\ &= \left[\frac{h_{01}^2 u^{(2)}(t) B(K)}{2} + \xi_{3i}(t) - v_i(t) \right] [1 + o_p(1)], \end{aligned}$$

where

$$\xi_{3i}(t) = \sum_{j=1}^{n_i} \mathbf{e}_1^\tau G_i^{-1} H_{ij,1}(t) \frac{K_{h,02}(t_{ij} - t)}{\sigma^2(t_{ij})} [v_i(t_{ij}) + e_i(t_{ij})].$$

Obviously $E(\xi_{3i}(t)|\mathfrak{D}) = 0$, and the conditional variance of $\xi_{3i}(t)$ is

$$\mathrm{Var}\,(\xi_{3i}(t) - v_i(t)|\mathfrak{D}) = \frac{\tau^2(t)V(K)}{n_i h_{02} f(t)} + o_p[(n_i h_{02})^{-1}].$$

So in summary, we find

$$E(\Lambda_i(t)|\mathfrak{D}) = \frac{h_{01}^2 u^{(2)}(t) B(K)}{2} + o_p(h_{01}^2),$$

$$Var(\Lambda_i(t)|\mathfrak{D}) = \frac{\tau^2(t)V(K)}{n_i h_{02} f(t)} + o_p[(n_i h_{02})^{-1}].$$

The proof of Lemma 3.1 is completed.

Proof of Theorem 3.1. Note that $\hat{\theta} = [\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{F}}_i]^{-1} [\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{X}}_i']$ in (2.27), where $\hat{\mathbf{F}}_i = (F(\hat{X}_i(t_{i1})), \ldots, F(\hat{X}_i(t_{in_i})))^\tau$, $\hat{\mathbf{V}}_i = \hat{\mathbf{F}}_i D_\theta \hat{\mathbf{F}}_i^\tau + \mathbf{R}_i$. First, for the matrix, $\hat{\mathbf{V}}_i^{-1}$, we have

$$\hat{\mathbf{V}}_i^{-1} = \mathbf{R}_i^{-1} - \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i (D_\theta^{-1} + \hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i)^{-1} \hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1}. \tag{A.9}$$

Then

$$\begin{aligned}
&\hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{F}}_i \\
&= \hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i - \hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i \left\{ D_\theta^{-1} + \hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i \right\}^{-1} \hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i \\
&= \hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i - (\hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i)^{1/2} \\
&\quad \times \left\{ (\hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i)^{-1/2} D_\theta^{-1/2} D_\theta^{-1/2} (\hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i)^{-1/2} + I \right\}^{-1} (\hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i)^{1/2},
\end{aligned} \tag{A.10}$$

and furthermore,

$$\begin{aligned}
&\left\{ (\hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i)^{-1/2} D_\theta^{-1/2} D_\theta^{-1/2} (\hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i)^{-1/2} + I \right\}^{-1} \\
&\quad = I - (\hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i)^{-1/2} D_\theta^{-1/2} \left[I + D_\theta^{-1/2} (\hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i)^{-1} D_\theta^{-1/2} \right]^{-1} \\
&\quad\quad \times D_\theta^{-1/2} (\hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i)^{-1/2}.
\end{aligned} \tag{A.11}$$

Substitute (A.11) into (A.10). Note that R_i is the variance-covariance matrix of $\mathbf{\Delta}_i$ and each element of R_i is $o_p(1)$. $\hat{\mathbf{F}}$ is a $n_i \times q$ matrix. So each element of $(\hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i)^{-1}$ goes to $+\infty$ as $n_i \to +\infty$. With $(\hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i)^{-1} = o_p(1)$, we have

$$\begin{aligned}
\hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{F}}_i &= D_\theta^{-1/2} \left[I + D_\theta^{-1/2} (\hat{\mathbf{F}}_i^\tau \mathbf{R}_i^{-1} \hat{\mathbf{F}}_i)^{-1} D_\theta^{-1/2} \right]^{-1} D_\theta^{-1/2} \\
&= D_\theta^{-1/2} [I + o_p(1)]^{-1} D_\theta^{-1/2} = D_\theta^{-1} + o_p(1).
\end{aligned} \tag{A.12}$$

Similarly, $\hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1} \mathbf{F}_i = D_\theta^{-1} + o_p(1)$. So we have $\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{F}}_i = nD_\theta^{-1} + o_p(n)$. For notation simplicity, we use b_i to represent the random component of θ_i. Moreover, we have

$$\begin{aligned}
\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{X}}_i' &= \sum_{i=1} \hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1} [\mathbf{F}_i \theta + \mathbf{F}_i b_i + \mathbf{\Delta}_i] \\
&= [nD_\theta^{-1} \theta + D_\theta^{-1} \sum_{i=1}^n b_i + \sum_{i=1}^n \hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1} \mathbf{\Delta}_i][1 + o_p(1)].
\end{aligned}$$

Consequently, it is obvious that

$$\hat{\theta} = \left\{D_\theta^{-1} + o_p(1)\right\}^{-1} \times \left\{D_\theta^{-1}\theta + D_\theta^{-1}\frac{1}{n}\sum_{i=1}^n b_i + \frac{1}{n}\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1}\mathbf{\Delta}_i\right\}[1 + o_p(1)]. \quad (A.13)$$

By a similar procedure from (A.10) to (A.12),

$$\frac{1}{n}\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1}\mathbf{\Delta}_i = \frac{1}{n}\sum_{i=1}^n D_\theta^{-1}(\hat{\mathbf{F}}_i\mathbf{R}_i^{-1}\hat{\mathbf{F}}_i)^{-1}\hat{\mathbf{F}}_i\mathbf{R}_i^{-1}\mathbf{\Delta}_i[1 + o_p(1)].$$

By Lemma 2.1, $E(\mathbf{\Delta}_i|\mathfrak{D}) = O_p(h_{12}^2)$, one has

$$E(\frac{1}{n}\sum_{i=1}^n D_\theta^{-1}(\hat{\mathbf{F}}_i\mathbf{R}_i^{-1}\hat{\mathbf{F}}_i)^{-1}\hat{\mathbf{F}}_i\mathbf{R}_i^{-1}\mathbf{\Delta}_i|\mathfrak{D}) = O_p(h_{12}^2). \quad (A.14)$$

By Lemma 2.2, $\mathbf{\Delta}_i$ and $\mathbf{\Delta}_j$ are asymptotically independent. Since $(\hat{\mathbf{F}}_i^\tau\mathbf{R}_i^{-1}\hat{\mathbf{F}}_i)^{-1} = o_p(1)$,

$$Var\Big(\frac{1}{n}\sum_{i=1}^n D_\theta^{-1}(\hat{\mathbf{F}}_i^\tau\mathbf{R}_i^{-1}\hat{\mathbf{F}}_i)^{-1}\hat{\mathbf{F}}_i^\tau\mathbf{R}_i^{-1}\mathbf{\Delta}_i|\mathfrak{D}\Big) = \frac{1}{n^2}\sum_{i=1}^n D_\theta^{-1}(\hat{\mathbf{F}}_i^\tau\mathbf{R}_i^{-1}\hat{\mathbf{F}}_i)^{-1}D_\theta^{-1} = o_p(n^{-1}). \quad (A.15)$$

Then based on (A.14) and (A.15),

$$\frac{1}{n}\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau \hat{\mathbf{V}}_i^{-1}\mathbf{\Delta}_i = O_p(h_{11}^2) + o_p(n^{-1/2}). \quad (A.16)$$

On the other hand, $n^{-1}\sum_{i=1}^n b_i = O_p(n^{-1/2})$. Then by (A.13), $\hat{\theta} = \theta + o_p(1)$, and $\hat{\theta}$ is a consistent estimator of θ.

Proof of Theorem 3.2. First, we can rewrite $\hat{\theta} = [n^{-1}\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau\hat{\mathbf{V}}_i^{-1}\hat{\mathbf{F}}_i]^{-1}[n^{-1}\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau\hat{\mathbf{V}}_i^{-1}\hat{\mathbf{X}}_i']$. By (A.12), we have

$$\begin{aligned}
&\hat{\theta} - \theta \quad (A.17)\\
&= [n^{-1}\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau\hat{\mathbf{V}}_i^{-1}\hat{\mathbf{F}}_i]^{-1}[n^{-1}\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau\hat{\mathbf{V}}_i^{-1}\hat{\mathbf{X}}_i' - n^{-1}\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau\hat{\mathbf{V}}_i^{-1}\hat{\mathbf{F}}_i\theta]\\
&= [D_\theta^{-1} + o_p(1)]^{-1}[n^{-1}\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau\hat{\mathbf{V}}_i^{-1}(\mathbf{F}_i - \hat{\mathbf{F}}_i)\theta + n^{-1}\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau\hat{\mathbf{V}}_i^{-1}\mathbf{\Delta}_i\\
&\quad + n^{-1}\sum_{i=1}^n \hat{\mathbf{F}}_i^\tau\hat{\mathbf{V}}_i^{-1}\mathbf{F}_i b_i]
\end{aligned}$$

$$= n^{-1}\sum_{i=1}^{n} b_i + o_p(n^{-1/2}) + D_\theta \left[\frac{1}{n}\sum_{i=1}^{n}\hat{\mathbf{F}}_i^\tau \mathbf{V}_i^{-1}\boldsymbol{\Delta}_i - \frac{1}{n}\sum_{i=1}^{n}\hat{\mathbf{F}}_i \mathbf{V}_i^{-1}\boldsymbol{\Lambda}_i \mathbf{F}_i^{*\prime}\theta\right] \times[1+o_p(1)], \qquad \text{(A.18)}$$

where $\mathbf{F}_i^* = (F(X_i(t_{i1}) + \phi_{i1} * \Lambda_i(t_{i1})), \ldots, F(X_i(t_{in_i}) + \phi_{in_i} * \Lambda_i(t_{in_i})))^\tau$, $\mathbf{F}_i^{*\prime}$ denotes the derivative of $\mathbf{F}_i^*$, and $\{\phi_{i1}, \ldots, \phi_{in_i}\}$ is a series of numbers between 0 and 1. $\boldsymbol{\Lambda}_i = diag(\Lambda_i(t_{i1})), \ldots, \Lambda_i(t_{in_i})))$. Then by (A.16), we can obtain that

$$\frac{1}{n}\sum_{i=1}^{n} D_\theta \hat{\mathbf{F}}_i^\tau \mathbf{V}_i^{-1}\boldsymbol{\Delta}_i = O_p(h_{11}^2) + o_p(n^{-1/2}).$$

Similarly, we find

$$\frac{1}{n}\sum_{i=1}^{n} D_\theta \hat{\mathbf{F}}_i \mathbf{V}_i^{-1}\boldsymbol{\Lambda}_i \mathbf{F}_i^{*\prime}\theta = O_p(h_{01}^2) + o_p(n^{-1/2}).$$

Under condition (C8), $h_{01} = o_p(n^{-1/4})$, $h_{11} = o_p(n^{-1/4})$, so

$$\frac{1}{n}\sum_{i=1}^{n} D_\theta \hat{\mathbf{F}}_i^\tau \mathbf{V}_i^{-1}\boldsymbol{\Delta}_i = o_p(n^{-1/2}), \quad \frac{1}{n}\sum_{i=1}^{n} D_\theta \hat{\mathbf{F}}_i \mathbf{V}_i^{-1}\boldsymbol{\Lambda}_i \mathbf{F}_i^{*\prime}\theta = o_p(n^{-1/2}).$$

Thus, under (C8) and based on (A.18), we represent $\hat{\theta} - \theta$ as

$$\hat{\theta} - \theta = \frac{1}{n}\sum_{i=1}^{n} b_i + o_p(n^{-1/2}).$$

It is assumed that $\{b_i, i = 1, \ldots, n\}$ are i.i.d. random vectors with mean 0 and variance matrix D_θ. By Central Limit Theorem, we get $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, D_\theta)$, the conclusion of Theorem 3.2.

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36**, 192-236.

Besag, J. (1977). Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika* **64**, 616-618.

Brunel, N. (2008). Parameter estimation of ODE's via nonparametric estimators. *Electronic J. Statist.* **2**, 1242-1267.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92**, 477-489.

Chen, J. and Wu, H. (2008a). Efficient local estimation for time-varying coefficients in deterministic dynamic models with applications to HIV-1 dynamics. *J. Amer. Statist. Assoc.* **103**, 369-384.

Chen, J. and Wu, H. (2008b). Estimation of time-varying parameters in deterministic dynamic models. *Statist. Sinica* **18**, 987-1006.

Davidian, M. and Giltinan, D. M. (1993). Some general estimation methods for nonlinear mixed effects models. *J. Biopharmaceutical Statist.* **3**, 23-55.

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

Draisma, G., Dree, H., Ferreira, A. and Hann, L. D. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* **10**, 251-280.

Ellner, S., Seifu, Y. and Smith, R. H. (2002). Fitting population dynamic models to time-series data by gradient matching. *Ecology* **83**, 2256-2270.

Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: theory and application. *Ann. Statist.* **9**, 861-869.

Guedj, J., Thiébaut, R. and Commenges, D.(2007). Maximum likelihood estimation in dynamical models of HIV. *Biometrics* **63**, 1198-1206.

Haase, A. T., Henry, K., Zupancic, M., Sedgewick, G., Faust, R. A., Melroe, H., Cavert, W., Gebhard, K., Staskus, K., Zhang, Z.-Q., Dailey, P. J., Balfour Jr., H. H., Erice, A. and Perelson, A. S. (1996). Quantitative image analysis of HIV-1 infection in lymphoid tissue. *Science* **274**, 985-989.

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

Hellerstein, M., Hanley, M. B., Cesar, D., Siler, S., Papageorgopoulos, C., Wieder, E., Schmidt, D., Hoh, R., Neese, R., Macallan, D., Deeks, S. and McCune, J. M. (1999). Directly measured kinetics of circulating T lymphocytes in normal and HIV-1-infected humans. *Nature Medicine* **5**, 83-89.

Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M. and Markowitz, M. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123-126.

Huang, Y., Liu, D. and Wu, H. (2006). Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics* **62**, 413-423.

Huang, Y. and Wu, H. (2006). A Bayesian approach for estimating antiviral efficacy in HIV dynamic models. *J. Appl. Statist.* **33**, 155-174.

Hürlimann, W. (2004). On the rate of convergence to asymptotic independence between order statistics. *Statist. Probab. Lett.* **66**, 355-362.

Lahiri, S. N. (2003). A necessary and sufficient condition for asymptotic independence of discrete fourier transforms under short-and long-range dependence. *Ann. Statist.* **31**, 613-641.

Liang, H., Wu, H. and Carroll, R.J. (2003). The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient semiparametric models with Measurement Error. *Biostatistics* **4**, 297-312.

Liang, H. and Wu, H. (2008). Parameter estimation for differential equation models using a framework of measurement error in regression. *J. Amer. Statist. Assoc.* **103**, 1570-1583.

Miao, H., Dykes, C., Demeter, L. M. and Wu, H. (2009). Differential equation modeling of HIV viral fitness experiments: model identification, model Selection, and multi-model inference. *Biometrics* **65**, 292-300.

Nowak, M. A. and May, W. H. (2000). *Virus Dynamics: Mathematical Principles of Immunology and Virology.* Oxford University Press, Oxford.

Perelson, A. S., Essunger, P., Cao, Y. Z., Vesanen, M., Hurley, A., Saksela, K., Markowitz, M. and Ho, D. D. (1997). Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387**, 181-191.

Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. and Ho, D. D. (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582-1586.

Perelson, A. S. and Nelson, P. W. (1999). Mathematical analysis of HIV-1 dynamics in vivo. *SIAM Rev.* **41**, 3-44.

Putter, H., Heisterkamp, S. H., Lange, J. M. and De Wolf, F. (2002). A Bayesian approach to parameter estimation in HIV dynamical models. *Stat. Med.* **21**, 2199-2214.

Ramsay, J. O., Hooker, G., Campbell, D. and Cao, J. (2007). Parameter estimation for differential equations: a generalized smoothing approach (with discussion). *J. Roy. Statist. Soc. Ser. B* **69**, 741-796.

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

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

Shi, M., Weiss, R. E. and Taylor, J. M. (1996). An analysis of pediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Appl. Statist.* **45**, 151-163.

Stute, W. and Zhu, L. X. (2005). Nonparametric checks for single-index models. *Ann. Statist.* **33**, 1048-1083.

Tan, W. Y. and Wu, H. (2005). *Deterministic and Stochastic Models of AIDS Epidemics and HIV Infections with Intervention*. World Scientific, Singapore.

Troxel, A. B., Lipsitz, T. R. and Harrington, D. P. (1998). Marginal models for the analysis of longitudinal measurements with nonignorable non-monotone missing data. *Biometrika* **85**, 661-672.

Varah, J. M. (1982). A spline least squares method for numerical parameter estimation in differential equations. *SIAM J. Sci. Comput.* **3**, 28-46.

Vonesh, E. F. and Chinchilli, V. M. (1996). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker, New York.

Wu, H. (2005). Statistical methods for HIV dynamic studies in AIDS clinical trials. *Statist. Meth. Medical Res.* **14**, 171-192.

Wu, H. and Ding, A. (1999). Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics* **55**, 410-418.

Wu, H. and Liang, H. (2004). Backfitting random varying-coefficient models with time-dependent smoothing covariates. *Scan. J. Statist.* **31**, 3-19.

Wu, H., Kuritzkes, D. R., McClernon, D. R., Kessler, H., Connick, E., Landay, A., Spear, G., Heath-Chiozzi, M., Rousseau, F., Fox, L., Spritzler, J., Leonard, J. M. and Lederman, M. M. (1999). Characterization of viral dynamics in human immunodeficiency virus type 1-infected patients treated with combination antiretroviral therapy: Relationships to host factors, cellular restoration, and virologic end points. *J. Infectious Diseases* **179**, 799-807.

Wu, H. and Zhang, J. T. (2002). Local polynomial mixed-effects models for longitudinal data. *J. Amer. Statist. Assoc.* **97**, 883-897.

Wu, H. and Zhang, J. T. (2006). Nonparametric regression methods for longitudinal data analysis. Wiley, New Jersey.

Mathematics and Science College, Shanghai Normal University, Shanghai, China.

E-mail: kellyfang0221@gmail.com

Department of Biostatistics and Computational Biology, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642 , U.S.A.

E-mail: hwu@bst.rochester.edu

Department of Biostatistics and Department of Mathematics, Hong Kong Baptist University, HK, China.

E-mail: lzhu@hkbu.edu.hk

(Received June 2009; accepted March 2010)