Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Efficient Local Estimation for Time-Varying Coefficients in Deterministic Dynamic Models With Applications to HIV-1 Dynamics

Jianwei Chen & Hulin Wu

Efficient Local Estimation for Time-Varying Coefficients in Deterministic Dynamic Models With Applications to HIV-1 Dynamics

Jianwei CHEN and Hulin WU

Recently deterministic dynamic models have become very popular in biomedical research and other scientific areas; examples include modeling human immunodeficiency virus (HIV) dynamics, pharmacokinetic/pharmacodynamic analysis, tumor cell kinetics, and genetic network modeling. In this article we propose estimation methods for the time-varying coefficients in deterministic dynamic systems that are usually described by a set of differential equations. Three two-stage local polynomial estimators are proposed, and their asymptotic normality is established. An alternative approach, a discretization method that is widely used in stochastic diffusion models, is also investigated. We show that the discretization method that uses the simple Euler discretization approach for the deterministic dynamic model does not achieve the optimal convergence rate compared with the proposed two-stage estimators. We use Monte Carlo simulations to study the finite-sample performance, and use a real data application to HIV dynamics to illustrate the proposed methods.

KEY WORDS: Asymptotic properties; Stochastic diffusion models; Time-varying coefficient dynamic model; Two-stage local polynomial estimation.

1. INTRODUCTION

Dynamic systems often are described by a set of linear/nonlinear ordinary differential equations. Deterministic dynamic systems are very popular in physics and engineering. Recently, dynamic models have been used to delineate biomedical systems, including human immunodeficiency virus type 1 (HIV-1) dynamics, pharmacokinetic/pharmacodynamics, tumor cell kinetics, and human gene expression. In HIV/AIDS research, viral dynamic studies have led to an improved understanding of the pathogenesis of HIV infection and have been used to evaluate the antiviral response of antiretroviral regimens (Ho et al. 1995; Perelson, Neumann, Markowitz, Leonard, and Ho 1996, Perelson et al. 1997; Wu and Ding 1999; Wu 2002; Huang, Rosenkranz, and Wu 2003; Michele, Ruy, M. R., Martin, Ho, and Perelson 2004). For example, Ho et al. (1995) used a simple dynamic model,

$$\frac{d}{dt}X(t) = P(t) - cX(t), \tag{1}$$

to describe HIV-1 dynamics, where $X(t)$ is HIV-1 RNA concentrations (viral load) in plasma, $P(t)$ is the virus production rate, and c is the clearance rate of free virions. If the dynamic system is in a steady state before antiviral treatment, then $dX(t)/dt = 0$ or $P(t) = cX(t) = constant$. Ho et al. (1995) treated the patients in their study with a potent protease inhibitor drug that was assumed to be perfect, completely blocking virus production [i.e., $P(t) = 0$]. But this assumption may not be valid, because treatment efficacy may vary due to many factors, such as the patient's immune status (CD4+ T-cell counts), drug resistance, adherence, and pharmacokinetics. Virus production may be affected by these factors. For example, we may assume that a patient's immune status or CD4+ T-cell count is linearly related to the virus production rate, $P(t) = a_1(t) + a_2(t)CD4(t)$, where $a_1(t)$ and $a_2(t)$ are time-varying coefficients. Thus we can rewrite model (1) as

$$\frac{d}{dt}X(t) = a_1(t) + a_2(t)CD4(t) - cX(t). \tag{2}$$

In clinical studies or clinical practice, the viral load $X(t)$ is monitored periodically; that is, we have the observation equation

$$Y(t) = X(t) + \varepsilon_1(t), \tag{3}$$

where $\varepsilon_1(t)$ are measurement errors. It is important to characterize the time-varying effects of CD4+ T-cell counts, also monitored at the same time with viral load, on virus production, that is, to estimate the time-varying coefficients $a_1(t)$ and $a_2(t)$ based on the measurements of viral load. Motivated by analysis of these biomedical dynamic systems, in this article we investigate a more general deterministic dynamic model with time-varying coefficients,

$$\frac{d}{dt}X(t) = \sum_{l=1}^{d} a_l(t)Z_l(t) - g[X(t)] \tag{4}$$

and

$$Y(t) = X(t) + \sigma(t)e(t), \tag{5}$$

where $X(t)$ is called a state variable or state function, $Y(t)$ is the measurement of $X(t)$ at time t, $e(t)$ are iid measurement errors over time t with mean 0 and variance 1, $\gamma(t) = (Z_1(t), \dots, Z_d(t))^T$ $(d \geq 1)$, are covariates with $Z_1(t) = 1$, and $\mathbf{a}(t) = (a_1(t), \dots, a_d(t))^T$ is a time-varying coefficient vector. The function $g[\cdot]$ is assumed to be known; for example, it can be simply $g[X(t)] = cX(t)$, where c is a known constant. Equation (4) is often called a state equation or dynamic equation, and (5) is called a measurement or observation equation. This model allows us to study how the covariates affect the dynamics of $X(t)$. Although this type of model is widely used in many biomedical dynamic systems and other scientific fields, statistical estimation and inference methods have not been well developed and

Jianwei Chen is Assistant Professor, Department of Mathematics and Statistics, San Diego State University, San Diego, CA 92182 (E-mail: *jchen@sciences.sdsu.edu*). Hulin Wu is Professor, Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642 (E-mail: *hwu@bst.rochester.edu*). This work was completed while the first author was a postdoctoral research fellow at the University of Rochester. This work was supported by National Institute of Allergy and Infectious Diseases/National Institutes of Health grants RO1 AI 52765, RO1 AI 055290, and N01-AI-50020. The authors thank Jin-Ting Zhang and Jack Hall for insightful discussions, Jeanne Holden-Wiltse for helpful assistance, and the joint editors, associate editor, and two referees for their constructive comments.

Journal of the American Statistical Association
March 2008, Vol. 103, No. 481, Theory and Methods
DOI 10.1198/016214507000001382

investigated, except for some simple least squares–based methods scattered throughout the literature. Note that in this article we focus on the models and methods for the dynamic data from a single subject; in future work, we will generalize these models and methods for longitudinal data from a population of subjects.

Developing estimation and inference methods to estimate the time-varying coefficients in dynamic models is challenging, because a differential equation model may not have a closed-form solution, and the numerical solutions may require intensive computations, particularly when complex statistical methods are used for parameter estimation (Li, Brown, Lee, and Gupta 2002; Wu et al. 2005; Huang, Liu, and Wu 2006). Li et al. (2002) studied a time-varying dynamic model with pharmacokinetics applications using a smoothing spline approach. Wu et al. (2005) and Huang et al. (2006) modeled a time-varying coefficient in a HIV-1 dynamic model through a parametric approach using Bayesian techniques for inference. In the econometrics literature, nonparametric smoothing methods for estimating time-varying functions, such as the drift (mean) and diffusion (variance) in stochastic diffusion models, have been developed (Arfi 1995, 1998; Stanton 1997; Chapman and Pearson 2000; Fan and Zhang 2003; Fan, Jiang, Zhang, and Zhou 2003; Fan 2005). Arfi (1995, 1998) developed the Nadaraya–Watson kernel estimator for drift and diffusion. Stanton (1997) introduced a drift and diffusion estimator through higher-order approximation and kernel estimation. Fan and Zhang (2003) and Fan et al. (2003) applied Euler discretization and the local polynomial estimation approach to estimate the time-varying functions in stochastic diffusion models. Little work has been done to investigate nonparametric methods for fitting deterministic dynamic models, however. In this article we propose two-stage local polynomial estimation procedures to estimate the time-varying coefficients in the deterministic dynamic model (4)–(5). We also study the discretization method, similar to that proposed by Stanton (1997) and Fan and Zhang (2003). Both methods share the same first-order asymptotic bias, whereas the discretization local estimator has a larger asymptotic variance compared with the proposed two-stage local estimators.

The rest of the article is organized as follows. In Section 2 we introduce the three proposed two-stage local polynomial estimators and derive. The explicit formulas for the proposed estimators. We study the asymptotic properties of the proposed estimators in Section 3. In Section 4 we investigate the discretization local estimation method for deterministic dynamic models and discuss the corresponding asymptotic results. In Section 5 we propose bandwidth selection methods for both stages in the estimation procedure for estimating the state variables and the time-varying coefficients. In Section 6 we present numerical simulation and application examples to illustrate the proposed methods. We provide some concluding remarks in Section 7, and present technical proofs of the theoretical results in the Appendix.

2. TWO-STAGE LOCAL ESTIMATORS

In this section we focus on estimating the unknown time-varying coefficient vector $\mathbf{a}(t)$ under the model (4)–(5). To achieve this aim, we first estimate the state function $X(t)$ and its first-order derivative $X^{(1)}(t) = dX(t)/dt$. We use local polynomial regression techniques for this purpose. Assume that the $(p+1)$th-order derivative of the state function $X(t)$ and the $(q+1)$th-order derivative of the time-varying coefficients $\mathbf{a}(t)$ exist. Let $\{(\mathbf{Z}(t_i), Y(t_i)), i = 1, \dots, n\}$ denote the observed data at time points $t_1, t_2, \dots, t_n$ from the time-varying coefficient dynamic model (4)–(5). We propose a two-stage local polynomial modeling approach to estimate the functions $X(t)$, $X^{(1)}(t)$, and $\mathbf{a}(t)$. The estimation procedure comprises two stages. The first stage involves estimating $X(t)$ and $X^{(1)}(t)$ based on the observation equation (5) with a bandwidth h_1. In the second stage, a nonparametric regression model is fitted for estimating the time-varying coefficients $\mathbf{a}(t) = (a_1(t), \dots, a_d(t))^T$ using the estimated functions $\widehat{X}(t)$ and $\widehat{X}^{(1)}(t)$, with a second bandwidth h_2.

We use a standard local polynomial regression method (Fan and Gijbels 1996) to estimate $X(t)$ and $X^{(1)}(t)$ at a given time point t_0 in the first stage. Based on a Taylor expansion, $X(t_i)$, $i = 1, \dots, n$, is approximated locally by

$$X(t_i) \approx \beta_0(t_0) + \beta_1(t_0)(t_i - t_0) + \cdots + \beta_p(t_0)(t_i - t_0)^p,$$

for t_i in a neighborhood of a given time point t_0. Let $\beta = (\beta_0(t_0), \dots, \beta_p(t_0))^T$; then the local polynomial estimator $\widehat{\beta} = (\widehat{\beta}_0(t_0), \dots, \widehat{\beta}_p(t_0))^T$ can be obtained by minimizing the locally weighted least squares criterion

$$\sum_{i=1}^n \big\{Y(t_i) - [\beta_0(t_0) + \beta_1(t_0)(t_i - t_0) + \cdots + \beta_p(t_0)(t_i - t_0)^p]\big\}^2 K_{h_1}(t_i - t_0), \quad (6)$$

where $K_{h_1}(\cdot) = K(\cdot/h_1)/h_1$, with K a kernel function and h_1 a bandwidth. Then we obtain

$$\begin{aligned}\widehat{\beta}(t_0) &= \big(\widehat{\beta}_0(t_0), \widehat{\beta}_1(t_0), \dots, \widehat{\beta}_p(t_0)\big)^T \\ &= (\mathbf{X}_p^T W_1 \mathbf{X}_p)^{-1} \mathbf{X}_p^T W_1 Y, \end{aligned} \quad (7)$$

where $Y = (Y(t_1), \dots, Y(t_n))^T$, $W_1 = \mathrm{diag}(K_{h_1}(t_1 - t_0), \dots, K_{h_1}(t_n - t_0))$, and

$$\mathbf{X}_p = \begin{pmatrix} 1 & t_1 - t_0 & \cdots & (t_1 - t_0)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_n - t_0 & \cdots & (t_n - t_0)^p \end{pmatrix}.$$

Thus we can estimate the state function $X(t_0)$ and its derivative $X^{(1)}(t_0)$ by

$$\widehat{X}_p(t_0) = \widehat{\beta}_0(t_0) \qquad \text{and} \qquad \widehat{X}_p^{(1)}(t_0) = \widehat{\beta}_1(t_0). \quad (8)$$

In the second stage, we substitute $\widehat{X}_p(t_i)$ and $\widehat{X}_p^{(1)}(t_i)$ in the time-varying coefficient dynamic model (4) to obtain a working time-varying coefficient regression model,

$$\widehat{X}_p^{(1)}(t_i) = \sum_{l=1}^d a_l(t_i) Z_l(t_i) - g[\widehat{X}_p(t_i)] + \varepsilon_2(t_i), \quad (9)$$

where $\varepsilon_2(t_i)$ denotes the approximation error term. From model (9), we again apply the local polynomial regression technique to estimate the time-varying coefficients $\mathbf{a}(t) = (a_1(t), \dots, a_d(t))^T$. Because the time-varying coefficient vector $\mathbf{a}(t)$ has the $(q+1)$th-order derivative, for each given time point t_0, we approximate the function $a_l(t_i)$, $i = 1, \dots, n, l = 1, \dots, d$, locally by

$$a_l(t_i) \approx \alpha_{l,0}(t_0) + \cdots + \alpha_{l,q}(t_0)(t_i - t_0)^q$$

for t_i in a neighborhood of the point t_0. Following the work of Fan and Zhang (1999), we can obtain the local polynomial estimator $\widehat{\alpha}_l = (\widehat{\alpha}_{l,0}, \ldots, \widehat{\alpha}_{l,q})^T$, $l = 1, \ldots, d$, minimizing the locally weighted function,

$$\sum_{i=1}^{n} \left\{ \widehat{X}_p^{(1)}(t_i) + g[\widehat{X}_p(t_i)] - \sum_{l=1}^{d} \sum_{r=0}^{q} \alpha_{l,r}(t_i - t_0)^r Z_l(t_i) \right\}^2 \times K_{h_2}(t_i - t_0), \quad (10)$$

where $K_{h_2}(\cdot)$ is a kernel function and h_2 is a bandwidth.

Let $\widehat{\alpha} = (\widehat{\alpha}_{1,0}, \ldots, \widehat{\alpha}_{1,q}, \ldots, \widehat{\alpha}_{d,0}, \ldots, \widehat{\alpha}_{d,q})^T$ minimize the locally weighted function (10). Then the local estimator of the time-varying coefficient vector $\mathbf{a}(t_0)$ is $\widehat{\mathbf{a}}(t_0) = (\widehat{\alpha}_{1,0}(t_0), \ldots, \widehat{\alpha}_{d,0}(t_0))^T$. We call this estimator a two-stage pth-order local polynomial estimator. In particular, we designate $\widehat{\mathbf{a}}_L(t_0)$ and $\widehat{\mathbf{a}}_Q(t_0)$ the two-stage local linear and local quadratic estimators, two popular special cases used for practical implementations.

For convenience, we can write the solution in (10) as a matrix form. Let $W_{n2} = \mathrm{diag}(K_{h_2}(t_1 - t_0), \ldots, K_{h_2}(t_n - t_0))$ and

$$\mathbf{Z}_n = \begin{pmatrix} Z_1(t_1) & \cdots & (t_1 - t_0)^q Z_1(t_1) & \cdots & Z_d(t_1) & \cdots & (t_1 - t_0)^q Z_d(t_1) \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_1(t_n) & \cdots & (t_n - t_0)^q Z_1(t_n) & \cdots & Z_d(t_n) & \cdots & (t_n - t_0)^q Z_d(t_n) \end{pmatrix}.$$

In addition, let $\mathbf{X}_{p(i)}$ be the matrix $\mathbf{X}_p$ with $t_0 = t_i$ and let $W_{1(i)}$ be the matrix W_1 with $t = t_i$. If we estimate $X(t_0)$ and $X^{(1)}(t_0)$ simultaneously using a local linear or local quadratic regression, then the two-stage local linear estimator $\widehat{\mathbf{a}}_L(t_0)$ or local quadratic estimator $\widehat{\mathbf{a}}_Q(t_0)$ can be expressed as

$$\widehat{\mathbf{a}}_L(t_0) = (I_d \otimes e_{1,q+1}^T)(\mathbf{Z}_n^T W_{n2} \mathbf{Z}_n)^{-1} \times \mathbf{Z}_n^T W_{n2}(\mathbf{U}_{1,1} + \mathbf{g}[\mathbf{U}_{0,1}]) \quad (11)$$

and

$$\widehat{\mathbf{a}}_Q(t_0) = (I_d \otimes e_{1,q+1}^T)(\mathbf{Z}_n^T W_{n2} \mathbf{Z}_n)^{-1} \times \mathbf{Z}_n^T W_{n2}(\mathbf{U}_{1,2} + \mathbf{g}[\mathbf{U}_{0,2}]), \quad (12)$$

where $\otimes$ denotes the Kronecker product, I_d denotes a d-dimensional identity matrix, $\mathbf{U}_{j,p} = (\widehat{X}_p^{(j)}(t_1), \ldots, \widehat{X}_p^{(j)}(t_n))^T$; $\mathbf{g}[\mathbf{U}_{j,p}] = (g[\widehat{X}_p^{(j)}(t_1)], \ldots, g[\widehat{X}_p^{(j)}(t_n)])^T$, $j = 0, 1$, $p = 1, 2$, with $\widehat{X}_p^{(j)}(t_i) = e_{j+1,p+1}^T(\mathbf{X}_{p(i)}^T W_{1(i)} \mathbf{X}_{p(i)})^{-1} \mathbf{X}_{p(i)}^T W_{1(i)} Y$, $i = 1, \ldots, n$, and $e_{r,s}$, $r = 1, \ldots, s$, being a s-dimensional vector with 1 at the rth position and 0 elsewhere.

Similarly, if we estimate $X(t_0)$ and $X^{(1)}(t_0)$ separately by different orders of the local polynomial approximation [say, local linear and local quadratic smoothing, denoted by $\widehat{X}_1(t_0)$ and $\widehat{X}_2^{(1)}(t_0)$], then we can obtain an alternative two-stage local mixture estimator,

$$\widehat{\mathbf{a}}_M(t_0) = (I_d \otimes e_{1,q+1}^T)(\mathbf{Z}_n^T W_{n2} \mathbf{Z}_n)^{-1} \times \mathbf{Z}_n^T W_{n2}(\mathbf{U}_{1,2} + \mathbf{g}[\mathbf{U}_{0,1}]). \quad (13)$$

Intuitively the two-stage local mixture estimator $\widehat{\mathbf{a}}_M(t_0)$ may be expected to be more efficient than either the two-stage local linear or the local quadratic estimator, because the former may use different orders of the local polynomial and different bandwidths (if necessary) to more efficiently estimate $X(t_0)$ and its derivative $X'(t_0)$ separately. Our asymptotic results in the next section demonstrate that the two-stage local linear or quadratic estimator and the two-stage local mixture estimator have the same asymptotic variance but different asymptotic biases, with the two-stage local quadratic and local mixture estimators having a rather simpler asymptotic bias expression compared with the two-stage local linear estimator.

3. ASYMPTOTIC RESULTS

In this section we investigate the asymptotic properties of the three proposed two-stage local estimators $\widehat{\mathbf{a}}_L(t_0)$, $\widehat{\mathbf{a}}_Q(t_0)$, and $\widehat{\mathbf{a}}_M(t_0)$ for the time-varying coefficients in the dynamic model (4)–(5). First, we define the following notations. Let $\mu_j = \int u^j K(u)\,du$,
$\nu_j = \int u^j K(u)^2\,du$, $j = 0, 1, 2, \ldots$; $c_q = (\mu_{q+1}, \ldots, \mu_{2q+1})^T$ be a $(q+1)$ vector; and

$$S = (\mu_{i+j-2})_{1 \le i,j \le q+1}, \qquad S^* = (\nu_{i+j-2})_{1 \le i,j \le q+1},$$

and

$$\Gamma(t) = \gamma(t)\gamma^T(t),$$

where $\gamma(t) = (Z_1(t), \ldots, Z_d(t))^T$ is defined in Section 1. Furthermore, let $f(t)$ be a density function of t, let $\mathcal{D} = (t_1, t_2, \ldots, t_n)^T$ be the observed vector, and let $\mathbf{a}^{(q+1)}(t_0) = (a_1^{(q+1)}(t_0), \ldots, a_d^{(q+1)}(t_0))^T$.

Under the regularity conditions given in the Appendix, we summarize the asymptotic normality of the three two-stage local estimators in the following theorems, the proofs of which are given in the Appendix.

Theorem 1. Under conditions given in the Appendix, as $h_1 \to 0$, $nh_1/\log(1/h_1) \to \infty$, $nh_1^3 \to \infty$, $h_2 \to 0$, and $nh_2 \to \infty$, the two-stage local linear estimator $\widehat{\mathbf{a}}_L(t_0)$ in (11) has

$$\sqrt{nh_2}\{\widehat{\mathbf{a}}_L(t_0) - \mathbf{a}(t_0) - \lambda_L(t_0) + o_P(h_1^2 + h_2^{q+1})\} \xrightarrow{\mathcal{D}} \mathrm{N}(0, \Sigma_L(t_0)),$$

where the asymptotic bias is

$$\begin{aligned} \lambda_L(t_0) = {} & \frac{1}{(q+1)!} h_2^{q+1} e_{1,q+1}^T S^{-1} c_q \mathbf{a}^{(q+1)}(t_0) \\ & + \frac{1}{2} h_1^2 \mu_2 X^{(2)}(t_0) g^{(1)}[X(t_0)] \Gamma(t_0)^{-1} \gamma(t_0) \\ & + \frac{1}{3!} h_1^2 \frac{\mu_4}{\mu_2} \left(3 \frac{f'(t_0)}{f(t_0)} X^{(2)}(t_0) + X^{(3)}(t_0) \right) \\ & \times \Gamma(t_0)^{-1} \gamma(t_0) \end{aligned}$$

and the asymptotic variance is

$$\Sigma_L(t_0) = \frac{\sigma^2(t_0)}{f(t_0)} \{g^{(1)}[X(t_0)]\}^2 \times e_{1,q+1}^T S^{-1} S^* S^{-1} e_{1,q+1} \Gamma(t_0)^{-1}.$$

In particular, for the case of linear fitting in the second stage ($q = 1$), we have the asymptotic expansion

$$\sqrt{nh_2}\{\widehat{\mathbf{a}}_L(t_0) - \mathbf{a}(t_0) - \lambda_L(t_0) + o_P(h_1^2 + h_2^2)\} \xrightarrow{\mathcal{D}} \mathrm{N}(0, \Sigma_L(t_0)),$$

where

$$\lambda_L(t_0) = \frac{1}{2}h_2^2\mu_2\mathbf{a}^{(2)}(t_0) + \frac{1}{2}h_1^2\mu_2 X^{(2)}(t_0)g^{(1)}[X(t_0)]\Gamma(t_0)^{-1}\gamma(t_0) + \frac{h_1^2}{3!}\frac{\mu_4}{\mu_2}\left(3\frac{f'(t_0)}{f(t_0)}X^{(2)}(t_0) + X^{(3)}(t_0)\right)\Gamma(t_0)^{-1}\gamma(t_0)$$

and

$$\Sigma_L(t_0) = \frac{\nu_0\sigma^2(t_0)}{f(t_0)}\{g^{(1)}[X(t_0)]\}^2\Gamma(t_0)^{-1}.$$

Theorem 2. Under conditions given in the Appendix, as $h_1 \to 0, nh_1/\log(1/h_1) \to \infty, nh_1^3 \to \infty, h_2 \to 0$, and $nh_2 \to \infty$, the two-stage local quadratic estimator $\widehat{\mathbf{a}}_Q(t_0)$ in (12) has

$$\sqrt{nh_2}\{\widehat{\mathbf{a}}_Q(t_0) - \mathbf{a}(t_0) - \lambda_Q(t_0) + o_P(h_1^2 + h_2^{q+1})\} \xrightarrow{\mathcal{D}} \mathrm{N}(0, \Sigma_Q(t_0)),$$

where the asymptotic bias is

$$\lambda_Q(t_0) = \frac{1}{(q+1)!}h_2^{q+1}e_{1,q+1}^T S^{-1}c_q\mathbf{a}^{(q+1)}(t_0) + \frac{1}{3!}h_1^2\frac{\mu_4}{\mu_2}X^{(3)}(t_0)\Gamma(t_0)^{-1}\gamma(t_0),$$

and the asymptotic variance is

$$\Sigma_Q(t_0) = \frac{\sigma^2(t_0)}{f(t_0)}\{g^{(1)}[X(t_0)]\}^2 \times e_{1,q+1}^T S^{-1}S^*S^{-1}e_{1,q+1}\Gamma(t_0)^{-1}.$$

In particular, for the case of linear fitting in the second stage ($q = 1$), we have the asymptotic expansion

$$\sqrt{nh_2}\{\widehat{\mathbf{a}}_Q(t_0) - \mathbf{a}(t_0) - \lambda_Q(t_0) + o_P(h_1^2 + h_2^2)\} \xrightarrow{\mathcal{D}} \mathrm{N}(0, \Sigma_Q(t_0)),$$

where

$$\lambda_Q(t_0) = \frac{1}{2}h_2^2\mu_2\mathbf{a}^{(2)}(t_0) + \frac{1}{3!}h_1^2\frac{\mu_4}{\mu_2}X^{(3)}(t_0)\Gamma(t_0)^{-1}\gamma(t_0)$$

and

$$\Sigma_Q(t_0) = \frac{\nu_0\sigma^2(t_0)}{f(t_0)}\{g^{(1)}[X(t_0)]\}^2\Gamma(t_0)^{-1}.$$

Theorem 3. Under conditions given in the Appendix, as $h_1 \to 0, nh_1/\log(1/h_1) \to \infty, nh_1^3 \to \infty, h_2 \to 0$, and $nh_2 \to \infty$, the two-stage local mixture estimator $\widehat{\mathbf{a}}_M(t_0)$ in (13) has

$$\sqrt{nh_2}\{\widehat{\mathbf{a}}_M(t_0) - \mathbf{a}(t_0) - \lambda_M(t_0) + o_P(h_1^2 + h_2^{q+1})\} \xrightarrow{\mathcal{D}} \mathrm{N}(0, \Sigma_M(t_0)),$$

where the asymptotic bias is

$$\lambda_M(t_0) = \frac{1}{(q+1)!}h_2^{q+1}e_{1,q+1}^T S^{-1}c_q\mathbf{a}^{(q+1)}(t_0) + \frac{1}{2}h_1^2 X^{(2)}(t_0)g^{(1)}[X(t_0)]\Gamma(t_0)^{-1}\gamma(t_0) + \frac{1}{3!}h_1^2\frac{\mu_4}{\mu_2}X^{(3)}(t_0)\Gamma^{-1}(t_0)\gamma(t_0),$$

and the asymptotic variance is

$$\Sigma_M(t_0) = \frac{\sigma^2(t_0)}{f(t_0)}\{g^{(1)}[X(t_0)]\}^2 \times e_{1,q+1}^T S^{-1}S^*S^{-1}e_{1,q+1}\Gamma(t_0)^{-1}.$$

In particular, for the case of linear fitting in the second stage ($q = 1$), we have the asymptotic expansion

$$\sqrt{nh_2}\{\widehat{\mathbf{a}}_M(t_0) - \mathbf{a}(t_0) - \lambda_M(t_0) + o_P(h_1^2 + h_2^2)\} \xrightarrow{\mathcal{D}} \mathrm{N}(0, \Sigma_M(t_0)),$$

where

$$\lambda_M(t_0) = \frac{1}{2}h_2^2\mu_2\mathbf{a}^{(2)}(t_0) + \frac{1}{2!}h_1^2 X^{(2)}(t_0)g^{(1)}[X(t_0)]\Gamma(t_0)^{-1}\gamma(t_0) + \frac{1}{3!}h_1^2\frac{\mu_4}{\mu_2}X^{(3)}(t_0)\Gamma(t_0)^{-1}\gamma(t_0)$$

and

$$\Sigma_M(t_0) = \frac{\nu_0\sigma^2(t_0)}{f(t_0)}\{g^{(1)}[X(t_0)]\}^2\Gamma(t_0)^{-1}.$$

The results in Theorems 1, 2, and 3 show that the asymptotic biases of the two-stage local quadratic estimator $\widehat{\mathbf{a}}_Q(t_0)$ and the two-stage local mixture estimator $\widehat{\mathbf{a}}_M(t_0)$ have simpler structures and do not involve $f^{(1)}(t_0)$, which appears in the asymptotic bias of the two-stage local linear estimator $\widehat{\mathbf{a}}_L(t_0)$, whereas the local quadratic estimator $\widehat{\mathbf{a}}_Q(t_0)$ has the simplest expression of the asymptotic bias. It also is interesting to notice that the asymptotic conditional variances of the three two-stage estimators are the same and are independent of the first bandwidth h_1. Thus we can choose the first bandwidth as small as possible, subject to the constraints $nh_1^3 \to \infty$ and $nh_1/\log(1/h_1) \to \infty$ in the conditions of Theorems 1, 2, and 3. In particular, when the first bandwidth $h_1 = o_P(h_2^{(q+1)/2})$, the square root of the mean squared errors (MSEs) for each of the three two-stage local estimators achieves the optimal convergence rate of order $O_P(n^{-(q+1)/(2q+3)})$.

Remark 1. When the local linear fitting ($q = 1$) in the second stage is used for estimating time-varying coefficients and the first bandwidth $h_1 = o_P(n^{-1/5})$ is used in Theorems 1, 2, and 3, the asymptotic biases of the three two-stage local estimators are the same, given by

$$\frac{1}{2}h_2^2\mu_2\mathbf{a}^{(2)}(t_0) + o_P(h_2^2).$$

Furthermore, the optimal bandwidth for the second stage is $h_2 = O_P(n^{-1/5})$, and the optimal convergence rate is $O_P(n^{-2/5})$.

From Remark 1, we can see that the proposed two-stage estimators have the same asymptotic bias and variance as long as the first bandwidth h_1 is between the rates $O_P(n^{-1/3})$ and $O_P(n^{-1/(2q+3)})$, and the optimal convergence rate $O_P(n^{-(q+1)/(2q+3)})$ for estimating the time-varying coefficient vector $\mathbf{a}(t_0) = (a_1(t_0), \ldots, a_d(t_0))^T$ can be achieved.

4. A DISCRETIZATION METHOD

An alternative approach for estimating the time-varying coefficients in the dynamic model (4) is a discretization method, which we briefly investigate in this section. Stanton (1997) introduced nonparametric estimators for the drift and diffusion in the stochastic diffusion model using higher-order approximation and kernel estimation techniques. Fan and Zhang (2003) and Fan et al. (2003) applied the Euler discretization and the local polynomial estimation approach for the stochastic diffusion model, and showed that the higher-order method can reduce the numerical approximation error in the asymptotic bias, but the variance escalates nearly exponentially with the order of approximation. We can apply the discretization approach and nonparametric estimation techniques of those author to our deterministic dynamic model. Let

$$\Delta X(t_i) = X(t_{i+1}) - X(t_i) \qquad \text{and} \qquad \Delta_i = t_{i+1} - t_i. \tag{14}$$

We use the Euler method to discretize the time-varying coefficient differential equation (4) as

$$\Delta X(t_i) = \left\{ \sum_{l=1}^{d} a_l(t_i) Z_l(t_i) - g[X(t_i)] \right\} \Delta_i + O(\Delta_i^2), \qquad i = 1, \ldots, n, \tag{15}$$

where $O(\Delta_i)$ is the discretization error. Note that when the time difference Δ_i is small, (15) is a good approximation of model (4). By combining (5) and (15), we obtain

$$E\big\{[Y(t_{i+1}) - Y(t_i)]/\Delta_i | \mathcal{D}\big\} = \sum_{l=1}^{d} a_l(t_i) Z_l(t_i) - g[X(t_i)] + O(\Delta_i). \tag{16}$$

Based on model (16), using the local polynomial techniques developed in Section 2, we can obtain the local estimator of the time-varying coefficients, $\mathbf{a}(t_0) = (a_1(t_0), \ldots, a_d(t_0))^T$, by minimizing the estimated locally weighted function

$$\sum_{i=1}^{n-1} \left\{ [Y(t_{i+1}) - Y(t_i)]/\Delta_i - \sum_{l=1}^{d} \sum_{r=0}^{q} \alpha_{l,r}(t_i - t_0)^r Z_l(t_i) + g[\widehat{X}_1(t_i)] \right\}^2 K_{h_2}(t_i - t_0), \tag{17}$$

where $K_h(\cdot)$ is a kernel function with h_2 a bandwidth and $\widehat{X}_1(t_i)$, $i = 1, \ldots, n$, is a local linear estimator of the state variable $X_1(t_i)$ developed in (8).

Let $\widehat{\alpha}_D = (\widehat{\alpha}_{1,0}, \ldots, \widehat{\alpha}_{1,q}, \ldots, \widehat{\alpha}_{d,0}, \ldots, \widehat{\alpha}_{d,q})^T$ minimize the locally weighted function (17). Then the local estimator of the time-varying coefficient vector $\mathbf{a}(t_0)$ is $\widehat{\mathbf{a}}_D(t_0) = (\widehat{\alpha}_{1,0}(t_0), \ldots, \widehat{\alpha}_{d,0}(t_0))^T$. We call this estimator a direct local polynomial estimator. Similar to theorem 4 of Fan and Zhang (2003), the asymptotic bias and variance of the direct local polynomial estimator $\widehat{\mathbf{a}}_D(t_0)$ are described in the following theorem.

Theorem 4. Under conditions given in the Appendix, as $h_1 \to 0$, $nh_1/\log(1/h_1) \to \infty$, and $nh_1^3 \to \infty$, $h_2 \to 0$, and $nh_2^3 \to \infty$, the asymptotic bias of the direct local polynomial estimator $\widehat{\mathbf{a}}_D(t_0)$ is

$$\begin{aligned} \operatorname{bias}\{\widehat{\mathbf{a}}_D(t_0)|\mathcal{D}\} &= \frac{1}{2} X^{(2)}(t_0)\Delta\gamma(t_0) + O(\Delta^2) \\ &\quad + \frac{h_2^{q+1}}{(q+1)!} e_{1,q+1}^T S^{-1} c_q \mathbf{a}^{(q+1)}(t_0) \\ &\quad + \frac{1}{2} h_1^2 \mu_2 X^{(2)}(t_0) g^{(1)}[X(t_0)]\Gamma(t_0)^{-1}\gamma(t_0) \\ &\quad + o_P(h_1^2 + h_2^{q+1}), \end{aligned}$$

and the asymptotic variance is

$$\operatorname{var}\{\widehat{\mathbf{a}}_D(t_0)|\mathcal{D}\} = \frac{\{\sigma^2(t_0)\}^{(1)}}{f(t_0)nh_2\Delta} e_{1,q+1}^T S^{-1} S^* S^{-1} e_{1,q+1} \Gamma(t_0)^{-1}\{1 + o(1)\},$$

provided that $\{\sigma^2(\cdot)\}^{(1)} = d\sigma^2(\cdot)/dt$ is continuous in a neighborhood of t_0, $\Delta_i = \Delta \to 0$, and $h_2 = O(\Delta^{1/2})$.

Comparing these results with Theorems 1, 2, and 3 in Section 3 shows that the direct local estimator has the same asymptotic bias as the proposed two-stage local estimators, but the asymptotic variance of the direct local estimator has an additional term Δ compared with that of the two-stage estimators. If we use the optimal bandwidth $h_2 = O_P(n^{-1/5})$ for the local linear fitting $q = 1$, then the squared root of the MSE of the direct local estimator $\widehat{\mathbf{a}}_D(t_0)$ can achieve only the convergence rate of order $O_P(n^{-2/7})$, which is slower than the $O_P(n^{-2/5})$ rate of the two-stage estimators. This result demonstrates that the asymptotic properties of the three proposed two-stage estimators are more appealing than those of the direct local estimator.

5. BANDWIDTH SELECTION AND IMPLEMENTATION

5.1 Optimal Bandwidth

In this section we discuss the selection of smoothing parameters for the two-stage local estimators. Bandwidth selection is very important in nonparametric smoothing. The asymptotic optimal bandwidth $h_{1,opt}$ for estimating $X(t)$ and $X'(t)$ in the first stage is standard (Fan and Gijbels 1996) and is not discussed further here. From Theorems 1, 2, and 3 developed in Section 3, the optimal bandwidth $h_{2,opt}$ for estimating $\mathbf{a}(t)$ can be chosen by minimizing their asymptotic weighted mean integrated squared error (MISE),

$$\operatorname{MISE}\{\widehat{\mathbf{a}}(h_2)\} = \int \Big\{ \operatorname{bias}(\widehat{\mathbf{a}}(t)|D)^T \Gamma(t) \operatorname{bias}(\widehat{\mathbf{a}}(t)|D) + \operatorname{tr}\big[\Gamma(t)\operatorname{var}(\widehat{\mathbf{a}}(t)|D)\big] \Big\} w(t)\, dt,$$

where $w(\cdot)$ is a weight function. By using the results in Theorems 1, 2, and 3, the asymptotic weighted MISE for the two-stage local estimators $\widehat{\mathbf{a}}_L(t)$, $\widehat{\mathbf{a}}_Q(t)$, and $\widehat{\mathbf{a}}_M(t)$ can be obtained.

As an example, the asymptotic weighted MISE of $\widehat{\mathbf{a}}_Q(t)$ can be expressed as

$$\begin{aligned}
&\mathrm{MISE}\{\widehat{\mathbf{a}}_Q(h_2)\}\\
&= \int \Bigg\{ \bigg[\frac{1}{(q+1)!} h_2^{q+1} e_{1,q+1}^T S^{-1} c_q \mathbf{a}^{(q+1)}(t) \\
&\quad + \frac{1}{3!} h_1^2 \frac{\mu_4}{\mu_2} X^{(3)}(t) \Gamma^{-1}(t) \gamma(t) \bigg]^T \\
&\quad \times \Gamma(t) \bigg[\frac{1}{(q+1)!} h_2^{q+1} e_{1,q+1}^T S^{-1} c_q \mathbf{a}^{(q+1)}(t) \\
&\quad + \frac{1}{3!} h_1^2 \frac{\mu_4}{\mu_2} X^{(3)}(t) \Gamma^{-1}(t) \gamma(t) \bigg] \\
&\quad + \frac{d\sigma^2(t)}{f(t) n h_2} \big\{ g^{(1)}[X(t)] \big\}^2 e_{1,q+1}^T S^{-1} S^* S^{-1} e_{1,q+1} \Bigg\} \\
&\quad \times w(t)\, dt. \qquad (18)
\end{aligned}$$

By taking the derivative of MISE with respect to h_2, the asymptotic optimal bandwidth $h_{2,opt}$ for estimator $\widehat{\mathbf{a}}_Q(t)$ can be obtained. In particular, if we take $h_1 = o_P(h_2^{(q+1)/2})$, then the asymptotic weighted MISE of $\widehat{\mathbf{a}}_Q(t)$ is given by

$$\begin{aligned}
&\mathrm{MISE}\{\widehat{\mathbf{a}}_Q(h_2)\}\\
&= \int \Bigg\{ \frac{1}{[(q+1)!]^2} h_2^{2q+2} [e_{1,q+1}^T S^{-1} c_q]^2 \mathbf{a}^{(q+1)^T}(t) \\
&\quad \times \Gamma(t) \mathbf{a}^{(q+1)}(t) \\
&\quad + \frac{d\sigma^2(t)}{f(t) n h_2} \big\{ g^{(1)}[X(t)] \big\}^2 e_{1,q+1}^T S^{-1} S^* S^{-1} e_{1,q+1} \Bigg\} \\
&\quad \times w(t)\, dt. \qquad (19)
\end{aligned}$$

Thus the asymptotic optimal bandwidth $h_{2,opt}$ is given by

$$h_{2,opt} = C_q(K) \left[\frac{d \int \sigma^2(t) w(t)/f(t)\, dt}{\int \{\mathbf{a}^{(p+1)^T}(t) \Gamma(t) \mathbf{a}^{(p+1)}(t)\} w(x)\, dx} \right]^{1/(2q+3)} \times n^{-1/(2q+3)}, \qquad (20)$$

where

$$C_q(K) = \left(\frac{[(q+1)!]^2 \{g^{(1)}[X(t)]\}^2 e_{1,q+1}^T S^{-1} S^* S^{-1} e_{1,q+1}}{(2q+2)[e_{1,q+1}^T S^{-1} c_q]^2} \right)^{1/(2q+3)}.$$

This theoretical optimal bandwidth depends on some unknown quantities and is not directly accessible. In the next section we provide a data-driven bandwidth selection procedure based on the generalized pre-asymptotic method (Fan and Gijbels 1996) and the generalized cross-validation (GCV) method. Note that similar asymptotic optimal bandwidths can be obtained for the two-stage local linear estimator $\widehat{\mathbf{a}}_L(t)$ and the mixture estimator $\widehat{\mathbf{a}}_M(t)$.

5.2 Bandwidth Selection

In this article we need to determine the first bandwidth h_1 for the estimation of the state function and its first derivative, and select the second bandwidth h_2 for estimating the time-varying coefficients. For simplicity, we consider only the local linear case ($q = 1$) for estimating the time-varying coefficients, which was recommended by Fan and Gijbels (1996) in practice. There exist several popular methods for selecting the bandwidth, including the plug-in method (Ruppert, Sheather, and Wand 1995), the pre-asymptotic substitution method (Fan and Gijbels 1996), the empirical bias bandwidth selection method (Ruppert 1997), the cross-validation method, and the GCV method. In Theorems 1, 2, and 3, we have shown that if the bandwidth h_1 in the first stage is selected between the rates $O_P(n^{-1/3})$ and $O_P(n^{-1/5})$, then the two-stage estimators can achieve an optimal convergence rate. First, we use the pre-asymptotic substitution method of Fan and Gijbels (1996) to obtain the initial optimal bandwidth $h_0 = \widehat{h}((Y(t_i), t_i), i = 1, \ldots, n)$, which is a consistent estimator of the asymptotic optimal bandwidth and is of the order $O_P(n^{-1/5})$. Then we suggest selecting the first bandwidth $h_1 = b\widehat{h}_0$, where $.5 < b < .9$ and $\widehat{h}_0$ is the optimal bandwidth for the local linear estimation of $X(t)$. Based on our experience with simulation studies, the estimated results do not differ much with different choices of the first bandwidth h_1. Thus the two-stage estimators are not very sensitive to the choice of the first bandwidth as long as it is small enough such that the bias in the first-stage smoothing is negligible.

For the second bandwidth h_2, we choose to use the GCV method that minimizes the GCV score

$$\mathrm{GCV}(h_2) = \frac{\sum_{i=1}^n \{\widehat{X}^{(1)}(t_i) + g[\widehat{X}(t_i)] - \sum_{l=1}^d \widehat{a}_l(t_i) Z_l(t_i)\}^2}{\{1 - \mathrm{tr}(\mathbf{S}_{h_2})/n\}^2}, \qquad (21)$$

where $\widehat{X}(t_i)$ and $\widehat{X}^{(1)}(t_i)$ are the estimates of $X(t)$ and $X^{(1)}(t)$ in the first stage with the bandwidth h_1; and $\widehat{a}_l(t_i), l = 1, \ldots, d$, is the two-stage estimator; and $\mathbf{S}_{h_2}$ is the corresponding smoother matrix. Now we summarize the bandwidth selection procedure for the two-stage estimators as follows:

1. Use the pre-asymptotic substitution method to obtain an initial optimal bandwidth for the local linear estimation of $X(t)$, $\widehat{h}_0 = \widehat{h}((y(t_i), t_i), i = 1, \ldots, n)$.
2. Select the first-stage bandwidth $h_1 = b\widehat{h}_0$, where $.5 < b < .9$ for the first-stage smoothing, and obtain the estimates $\widehat{X}_p(t_i)$ and $\widehat{X}_p^{(1)}(t_i)$, $p = 1, 2$, for $i = 1, \ldots, n$.
3. For the two-stage linear and quadratic estimators, use the GCV method to find a second bandwidth h_2 by using the estimates $(\widehat{X}_p(t_i), \widehat{X}_p^{(1)}(t_i))$, $p = 1, 2$, in (21). Then the optimal bandwidth h_2 can be used in (11)–(12) to obtain the two-stage local linear estimator $\widehat{\mathbf{a}}_L(t_0)$ and the two-stage local quadratic estimator $\widehat{\mathbf{a}}_Q(t_0)$.
4. For the two-stage mixture estimator, the optimal bandwidth h_2 can be obtained by minimizing the GCV function (21). This bandwidth can be used to obtain the two-stage mixture estimator $\widehat{\mathbf{a}}_M(t_0)$ in (13).

5.3 Local Ridge Regression

In the practical implementation of the proposed estimation procedure, we often encounter the ill-conditioned problem due to the collinearity of the local covariates in the second stage. Here we propose using the local ridge regression method to tackle this problem. For simplicity of illustration, we consider only a deterministic dynamic model with two time-varying covariates ($d = 2$). Based on the formula (12), the two-stage local quadratic estimator is given by

$$\begin{pmatrix} \widehat{a}_{1,Q} \\ \widehat{a}_{2,Q} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \times \begin{pmatrix} u_{0,11} & u_{1,11} & u_{0,12} & u_{1,12} \\ u_{1,11} & u_{2,11} & u_{1,12} & u_{2,12} \\ u_{0,21} & u_{1,21} & u_{0,22} & u_{1,22} \\ u_{1,21} & u_{2,21} & u_{1,22} & u_{2,22} \end{pmatrix}^{-1} \begin{pmatrix} v_{0,1} \\ v_{1,1} \\ v_{0,2} \\ v_{1,2} \end{pmatrix}, \quad (22)$$

where

$$u_{k,lr} = \sum_{i=1}^{n} (t_i - t_0)^k K_{h_2}(t_i - t_0) Z_l(t_i) Z_r(t_i)$$

and

$$v_{j,l} = \sum_{i=1}^{n} (t_i - t_0)^j K_{h_2}(t_i - t_0) Z_l(t_i) \{ U_{1,2}(i) - g[U_{0,2}(i)] \},$$

$$k, = 0, 1, 2;\ j = 0, 1;\ l, r = 1, 2;\ i = 1, \ldots, n,$$

with $U_{j,2}(i) = e_{j+1,3}^T (\mathbf{X}_{3(i)}^T W_{1(i)} \mathbf{X}_{3(i)})^{-1} \mathbf{X}_{3(i)}^T W_{1(i)} Y$, $j = 0, 1$; $i = 1, \ldots, n$.

During practical implementation, the matrix involving the inverse in (22) can be ill-conditioned because of the sparse local data points. A popular method for dealing with this problem is the ridge regression technique (Seifert and Gasser 1996; Fan and Chen 1999). The ridge parameter selection in the ridge regression is an important issue to address. Based on the conditions in Appendix, if $h_2 \to 0$ and $nh_2 \to \infty$, then we have the asymptotic approximation

$$u_{k,lr} = nh_2 f(t_0) h_2^{k-1} \int u^k K(u)\, du \times Z_l(t_0) Z_r(t_0)(1 + o_P(1)), \quad (23)$$

where $nh_2 f(t_0)$ can be intuitively understood as the effective number of local data points around the local time point t_0. Therefore, we can select the ridge parameter as

$$h_2^{k-1} \int u^k K(u)\, du\, Z_l(t_0) Z_r(t_0) \quad \text{for } k = 0, 2,$$

and replace $u_{k,lr}$ with $u_{k,lr} + h_2^{k-1} \int u^k K(u)\, du\, Z_l(t_0) Z_r(t_0)$ for $k = 0, 2$ in the matrix (22). It follows from (23) that the new matrix will not alter the asymptotic behavior and will avoid near singularity of the matrix when the effective number of local data points is small. In the same way, the ridge regression also can be used in the implementation of the two-stage local linear estimator $\widehat{\mathbf{a}}_L(t_0)$ and the mixture estimator $\widehat{\mathbf{a}}_M(t_0)$.

6. NUMERICAL EXAMPLES

In this section we design numerical studies to evaluate the finite-sample properties of the proposed two-stage estimators by Monte Carlo simulations and illustrate the use of the proposed methods through an application example in HIV-1 dynamics. We examine whether the time-varying coefficients in the deterministic dynamic models can be efficiently estimated by the proposed estimation procedures. A long-term HIV-1 dynamic model is used in the simulation studies and in the real data application.

6.1 Simulation Models

We designed our simulation studies based on a HIV-1 dynamic model with antiviral treatment that was investigated by several authors (Huang et al. 2003; Michele et al. 2004; Wu 2005). The model can be specified as a set of three differential equations,

$$T'(t) = \lambda - \rho T(t) - k[1 - r(t)] T(t) X(t), \quad (24)$$

$$T^{*\prime}(t) = k[1 - r(t)] T(t) X(t) - \delta T^*(t), \quad (25)$$

and

$$X'(t) = N\delta T^*(t) - cX(t), \quad (26)$$

where $T(t)$ is the concentration of uninfected CD4+ T cells, which are the targets of HIV-1 infection; $T^*(t)$ is the concentration of infected T cells; and $X(t)$ is the virus concentration at time t. Parameter λ represents the rate at which new T cells are generated from sources within the body, such as the thymus; ρ is the death rate of T cells; k is the infection rate of T cells infected by virus; δ is the death rate of infected cells; N is the number of new virions produced from each of the infected cells during their life time; and c is the clearance rate of free virions. In our simulation studies, the true parameter values and initial values of the state variables in the HIV-1 dynamic model (24)–(26) were taken as $\lambda = 36.0$, $\rho = .108$, $k = 5 \times 10^{-4}$, $\delta = .1$, $N = 1{,}000.0$, $c = 3.5$, $X(0) = 10{,}000.0$, $T(0) = 350$, and $T^*(0) = 20$. We selected these values based on the work of Huang et al. (2003). The time-varying antiviral drug efficacy parameter was assumed to be $r(t) = \cos(3.14t/500.0)$.

In clinical trials or clinical practice, we are able to measure the plasma virus concentration (viral load), $X(t)$, at each clinical visit; that is, we have a measurement or observation model for viral load of

$$Y(t) = X(t) + \sigma(t) e(t), \quad (27)$$

where $e(t)$ are measurement errors.

First, we generated data, $\{T(t_i), T^*(t_i), X(t_i)\}$, at time points $t_i = \text{day} \times t_i^*$ with $t_i^* = i/(n+1)$, $i = 1, \ldots, n$, by solving the differential equations (24)–(26). We generated the observed viral load data $\{Y(t_i), i = 1, \ldots, n\}$ based on the model (27) with the error term $e(t_i)$ following an iid standard normal distribution. The variance $\sigma^2(t_i) = (1 + t_i^{1/2})\sigma^2$ where $\sigma = 100$. We generated the data $\{Y(t_i), T(t_i), T^*(t_i), i = 1, \ldots, n\}$ for $n = 200$, 100, 50, and 20.

The HIV-1 dynamic model (24)–(26) can be transformed into a simpler dynamic model with time-varying coefficients as given in Section 1. Combining (25) and (26), leads to

$$X'(t) = -NT^{*\prime}(t) + Nk[1 - r(t)] X(t) T(t) - cX(t). \quad (28)$$

Let $a_1(t) = -NT^{*\prime}(t)$, $a_2(t) = Nk[1 - r(t)] X(t)$, $Z_1(t) = 1$, and $Z_2(t) = T(t)$. We also assume that the measured CD4+ T-cell counts are proportional to $T(t)$; that is, we can use CD4+ T-cell counts to replace $T(t)$ in the foregoing model. Then we obtain the time-varying coefficient dynamic model

$$X'(t) = a_1(t) Z_1(t) + a_2(t) Z_2(t) - cX(t) \quad (29)$$

and

$$Y(t) = X(t) + \sigma(t)e(t), \tag{30}$$

where $Z_1(t) = 1$ and $Z_2(t)$ are CD4+ T-cell counts that also are available from clinical studies. Thus this model allows us to assess how CD4+ T-cell counts affect virus production. Also note that the parameter c can be estimated separately by using a short-term HIV dynamic model, which can be expressed as (Perelson et al. 1996; Han and Chaloner 2004)

$$X(t) = X(0)e^{-ct} + \frac{cX(0)}{c-\delta}\left\{\frac{c}{c-\delta}[e^{-\delta t} - e^{-ct}] - \delta t e^{-ct}\right\} \tag{31}$$

and

$$Y(t) = X(t) + \sigma(t)e(t), \tag{32}$$

where $X(0)$ is the initial value of viral load (baseline value). Based on the short-term (0–7 days) viral load data, the parameter c can be estimated by fitting a nonlinear parametric model. Thus our focus here is on estimating the time-varying coefficients $a_1(t)$ and $a_2(t)$. Note that the constant parameter c is not the focus of this article. For the simulation studies, we fixed c to be the true value ($c = 3.5$). For the real data example, we have frequent viral load data in the first week, which we used to estimate the constant parameter c for each patient based on the methods of Perelson et al. (1996). We then used these values of c in model (29) to estimate the time-varying parameters using our proposed methods.

6.2 Performance of the Proposed Estimators

In the simulation studies, we applied the two-stage estimation methods proposed in the previous sections to model (29)–(30). To evaluate these estimators, we used the Epanechnikov kernel $K(t) = .75(1 - t^2)_+$ for all of the estimators. We used bandwidth selection method discussed in Section 5 to determine the first-stage bandwidth h_1 and the optimal bandwidth $h_{2,opt}$. We used the ridge regression method (Sec. 5.3) to overcome the potential singularity problem in the estimation computation. We assessed the performance of these estimators using the square root of average squared errors (RASE), defined as

$$RASE(\widehat{\mathbf{a}}) = \left\{ n_{grid}^{-1} \sum_{j=1}^{n_{grid}} (\widehat{\mathbf{a}}(t_j^0) - \mathbf{a}(t_j^0))^T (\widehat{\mathbf{a}}(t_j^0) - \mathbf{a}(t_j^0)) \right\}^{1/2}, \tag{33}$$

where $\{t_j^0, j = 1, \ldots, n_{grid}\}$ are the grid points at which the time-varying coefficients $\mathbf{a}(\cdot)$ were estimated. We summarize the performance of the three two-stage local estimators based on 500 simulation runs with a sample sizes of $n = 200$ and a smaller sample sizes of $n = 20$. The results for sample sizes $n = 100$ and $n = 50$ are basically the same as those for sample size $n = 200$ and thus are omitted.

We first compared the three two-stage estimators with a sample size of $n = 200$. Table 1 presents the results of the average of RASEs over different values of the first-stage bandwidth $h_{1,opt} = bh_0$, with $b = .5, .6, \ldots, 1.1$ and using the pre-asymptotic substitution method to obtain h_0. From the table, we can see that the RASEs are not sensitive to the choice of the first-stage bandwidth h_1. This result is consistent with the asymptotic properties of the proposed two-stage estimators that we investigated in Section 3. The boxplots of the RASEs for the three estimators with the first-stage bandwidth $h_1 = .7h_0$ are given in Figure 1, from which we can see that the two-stage quadratic estimator is slightly better than the two-stage local linear and local mixture estimators. From Table 1, we note that this trend is generally true for most cases. Figure 2 depicts a typical estimated function that attains the median performance among 500 simulations. In particular, the local linear estimate of $X(t)$ and the local quadratic estimate of $X'(t)$ are presented in Figures 2(a) and (b). Figure 2(c) and (d) plot the estimated time-varying coefficients $a_1(t)$ and $a_2(t)$ from the three two-stage methods. These plots, do not show a clear difference between the three estimators; all of them are very close to the true functions, although we can see differences between the three estimators from the numerical comparisons (Table 1).

Table 1. The RASEs for the three two-stage local estimators with sample size $n = 200$ and different bandwidths h_1

Initial bandwidth h_1	RASE($\widehat{\mathbf{a}}_L$)		RASE($\widehat{\mathbf{a}}_Q$)		RASE($\widehat{\mathbf{a}}_M$)	
	Mean	SD	Mean	SD	Mean	SD
.5 h_0	575	67	503	56	561	62
.6 h_0	518	72	477	56	514	65
.7 h_0	488	60	448	56	484	58
.8 h_0	451	59	421	56	452	57
.9 h_0	434	59	402	54	437	56
1.0 h_0	409	55	387	52	416	54
1.1 h_0	397	54	379	49	405	52

NOTE: SD, standard deviation.

To investigate the behavior of the methods for smaller samples, we conducted a simulation study with a sample size of $n = 20$. The performance of the three two-stage estimators is summarized in Figure 3. Similar to Figure 2, the proposed methods still give reasonable estimation results. We also conducted simulations for larger measurement errors and different sample sizes (data not shown); which produced similar conclusions.

6.3 An Application Example

We illustrate the proposed methods by analyzing real data from an AIDS clinical study. In this study, HIV-1–infected patients were treated with highly actively antiretroviral therapy (HAART) in combination with immune-based therapy. Thus it is interesting to study the effect of CD4+ T-cell counts (target of immune-based therapy) on the virus production (target of HAART therapy). In this study, the plasma virus concentration (viral load) and CD4+ T-cell counts were monitored frequently. We have both viral load and CD4+ T-cell count data at baseline and at weeks 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 56, 64, 72, 80, 88, and 96 after initiating the treatment. (Note that not all patients followed this measurement schedule exactly, and that some variations always exist in practice.) More frequent viral load data within the first 3 days also are available from these patients and can be used to estimate the constant viral dynamic parameters in model (31)–(32). Using the proposed

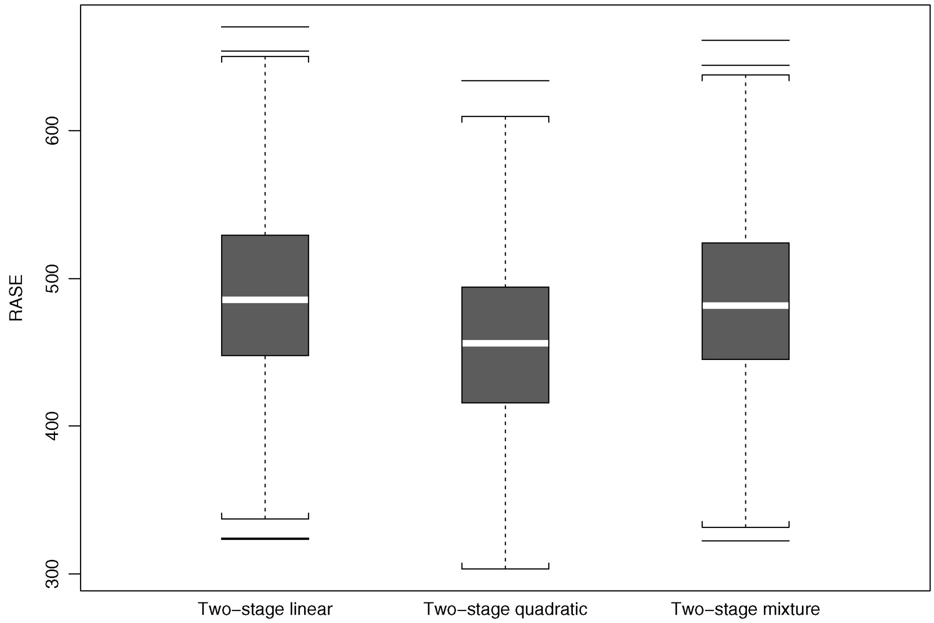

Figure 1. Simulation comparisons of different estimation methods with a sample size of $n = 200$. Boxplots of the RASEs for the two-stage local linear estimator $\widehat{\mathbf{a}}_L(t)$, the two-stage local quadratic estimator $\widehat{\mathbf{a}}_Q(t)$, and the two-stage local mixture estimator $\widehat{\mathbf{a}}_M(t)$.

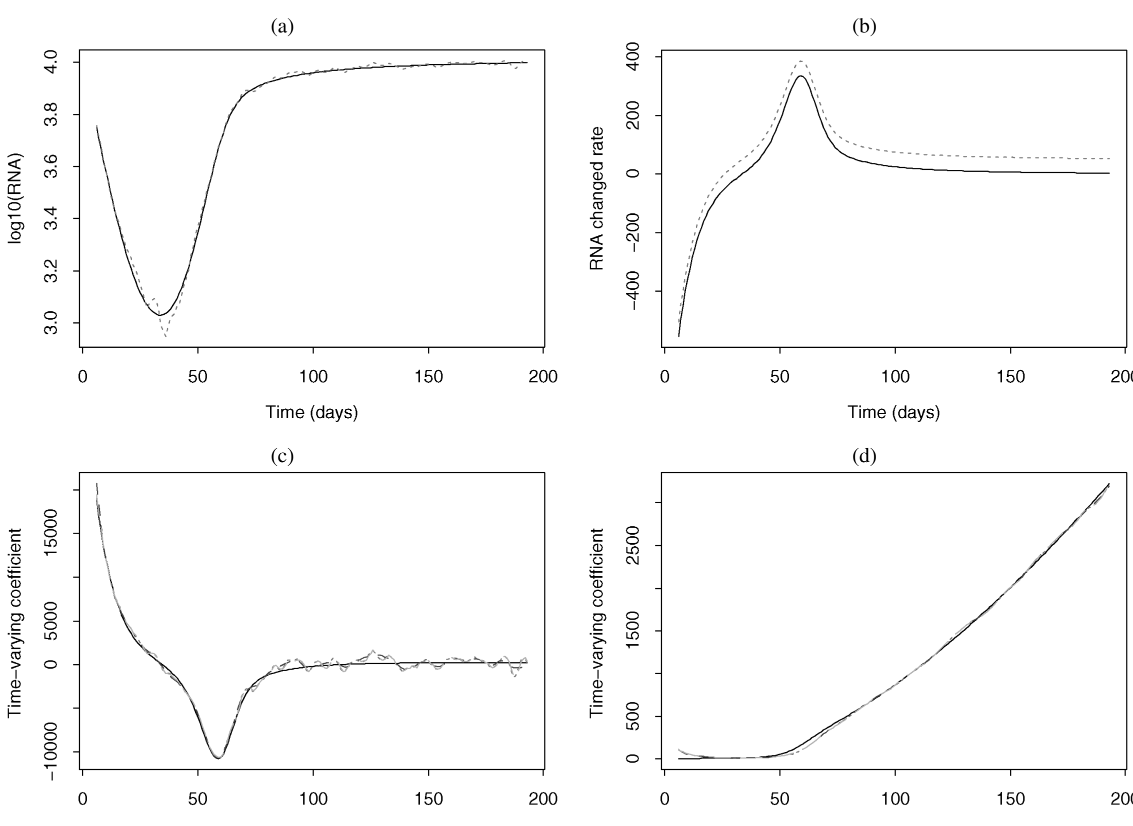

Figure 2. Estimation results from the two-stage estimation methods with a sample size of $n = 200$ (—, true functions; - - -, estimated functions). (a) Typical local linear estimates $\widehat{X}_1(t)$. (b) Typical local quadratic estimates $\widehat{X}_2'(t)$. (c) The three two-stage local estimates $\widehat{a}_1(t)$. (d) The three two-stage local estimates $\widehat{a}_2(t)$. In (c) and (d), dash curves (from shortest to longest dash) are the two-stage local estimators $\widehat{a}_L(x)$, $\widehat{a}_Q(x)$, and $\widehat{a}_M(x)$.

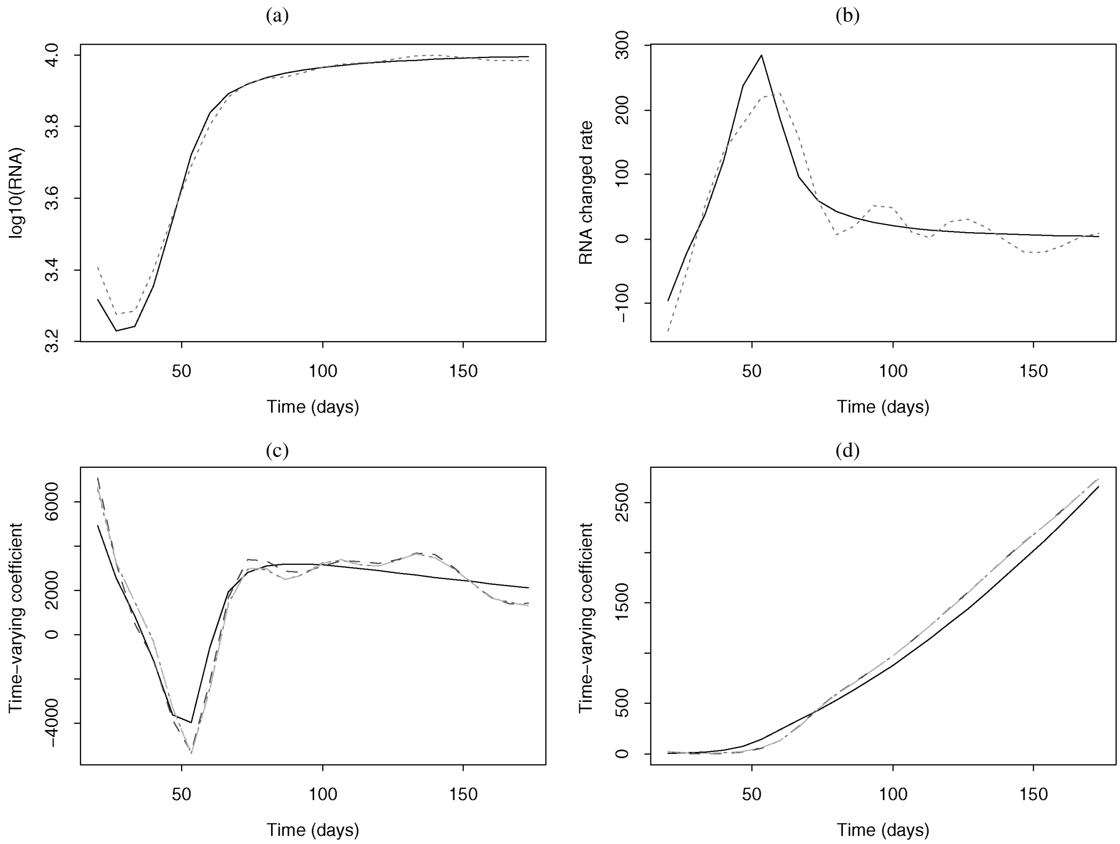

Figure 3. Estimation results from the two-stage estimation methods with a sample size of $n = 20$ (—, true functions; - - -, estimated functions). (a) Typical local linear estimates $\widehat{X}_1(t)$. (b) Typical local quadratic estimates $\widehat{X}_2'(t)$. (c) The three two-stage local estimates $\widehat{a}_1(t)$. (d) The three two-stage local estimates $\widehat{a}_2(t)$. In (c) and (d), dash curves (from shortest to longest dash) are the two-stage local estimators $\widehat{a}_L(x)$, $\widehat{a}_Q(x)$, and $\widehat{a}_M(x)$.

methods, we fitted the data to the time-varying coefficient dynamic model

$$X'(t_i) = a_1(t_i) + a_2(t_i)CD4(t_i) - cX(t_i)$$

and

$$Y(t_i) = X(t_i) + \sigma(t_i)e(t_i),$$

where $Y(t_i), i = 1, \ldots, n$, are viral load measurements. For illustration purposes, we present the model-fitting results for the data only from three HIV-1–infected subjects; we will report the complete results for all study subjects and the biomedical findings in later work.

Figure 4 presents plots of the estimated state variable function $\widehat{X}(t)$ (viral load trajectory during the treatment period) by local linear fitting and the estimates of the time-varying coefficients $\widehat{a}_1(t)$ and $\widehat{a}_2(t)$ by the two-stage local quadratic method for the three patients. From Figure 4(a), we can see that the viral load fitting is pretty good. The estimate of $\widehat{a}_2(t)$ shows that the effect of CD4+ T-cell counts on virus production is nonlinear, which can be reasonably explained by prey–predator dynamics. At the beginning of treatment, the viral load was high, and the abundance of CD4+ T cells, the targets of HIV-1 infection, was important and necessary for virus replication, which resulted in a greater CD4+ T-cell effect. After initiating the antiviral therapy, the viral load reduced to a lower level, and the effect of CD4+ T-cell counts also was attenuated, because the large number of targets or CD4+ T cells was not necessary. In a later stage, the viral load rebounded to a higher level, demonstrating that higher CD4+ T-cell counts will have a greater effect on the viral rebound. Thus the estimate of time-varying coefficients in the dynamic model is important for characterizing the time-varying effects of covariates on the viral dynamics.

7. SUMMARY AND DISCUSSION

Deterministic dynamic systems are widely used in modeling viral dynamics and other biological processes. Typically, the dynamic models are specified as differential equations that may or may not have closed-form solutions. Numerical solutions often are costly when computation-based statistical methods are used for parameter estimation. In this article we have introduced a time-varying coefficient dynamic model to study the dynamic relationship between the state variable and potential covariates in the dynamic system. Nonparametric regression techniques have been developed to fit the proposed time-varying coefficient dynamic models. Three two-stage estimators—the local linear estimator, the local quadratic estimator, and the local mixture estimator—have been proposed, and their asymptotic properties have been carefully studied for the time-varying dynamic models. An alternative approach—a discretization method that is popular in the stochastic diffusion (differential) models, also has been briefly investigated. We have conducted Monte Carlo simulations based on HIV dynamic models. Simulation results

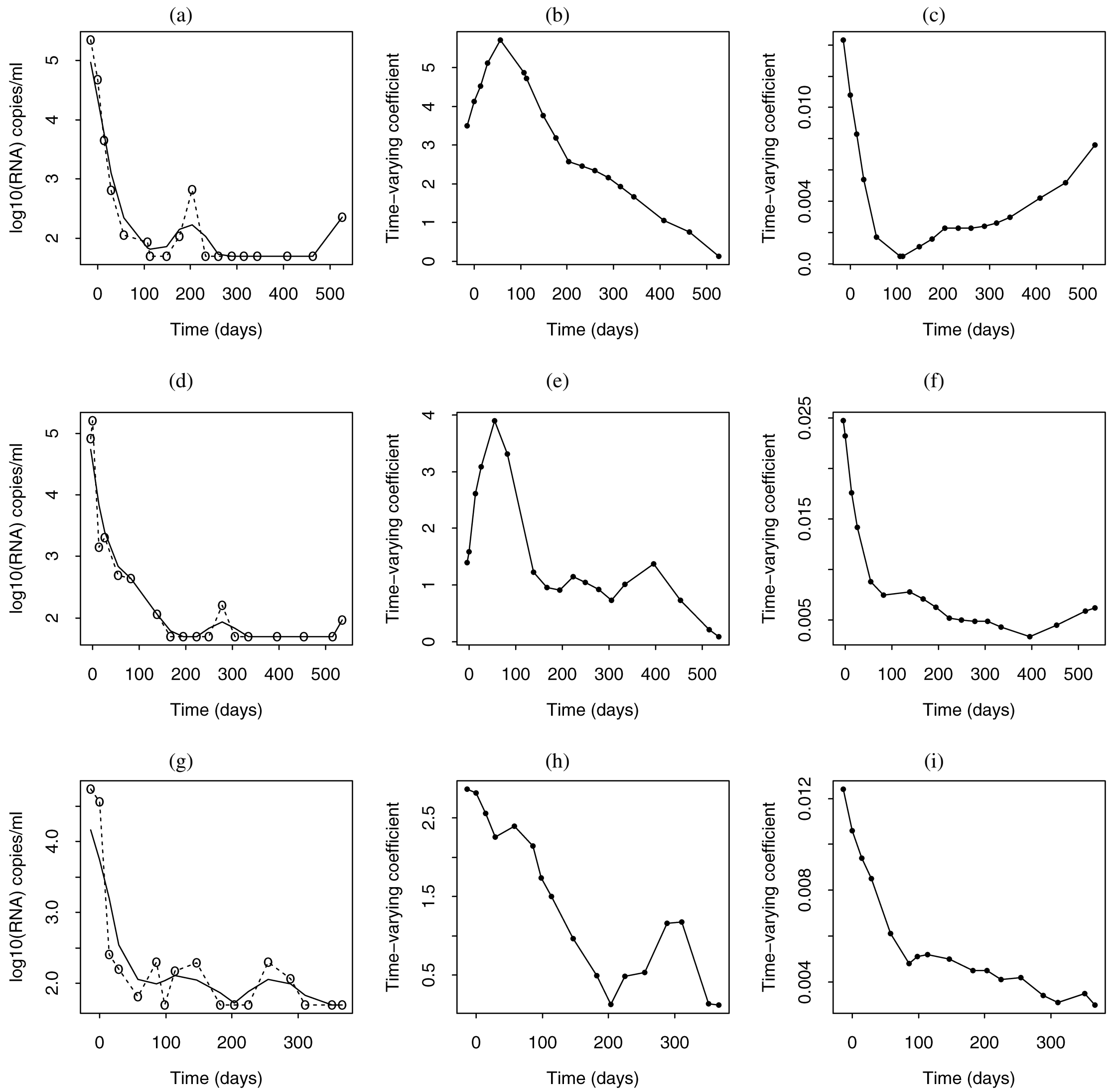

Figure 4. Data analysis results from three patients from an AIDS clinical trial. (a), (d), and (g) The estimated viral load over the treatment days [—, the local linear estimate of $X(t)$; the –o–o, the observed viral load]. (b), (e), and (h) The estimated time-varying coefficient $a_1(t)$ over the treatment days [the two-stage local quadratic estimate $\widehat{a}_1(t)$]. (c), (f), and (i) The estimated time-varying coefficient $a_2(t)$ over the treatment days [the two-stage local quadratic estimate $\widehat{a}_2(t)$]. (a), (b), and (c) Patient 3. (d), (e), and (f) Patient 2. (g), (h), and (i) Patient 3.

have shown that all of the three two-stage estimators perform well, whereas the two-stage local quadratic estimator is more stable and performs slightly better than the other two estimators. An application example from an AIDS clinical study also has been presented to illustrate the usefulness of the proposed methodologies.

Statistical literature dealing with deterministic dynamic models, particularly with time-varying parameters, is very sparse (Li et al. 2002; Wu et al. 2005; Huang et al. 2006), although dynamic models have become very popular in modeling biomedical systems and in other scientific areas. We are making an effort to fill this gap. The proposed new models and methods have several innovative and unique features compared with the existing methods:

1. To the best of our knowledge, the proposed two-stage methods for estimating time-varying parameters in differential equation models are quite new and innovative.
2. Compared with the methods of parameter estimation for differential equation models proposed by Li et al. (2002) and Huang et al. (2006), our new methods do not require numerically solving the differential equations. Moreover our methods are regression-based approaches and thus can avoid the high computational cost and possible numerical errors caused by numerical evaluation of nonlinear differential equations.
3. Li et al. (2002) proposed a spline-based method to deal with the time-varying parameters in differential equations, but provided no formal theoretical justifications or asymptotic results. In contrast, we applied the local polynomial approach and provided formal theoretical justifications for the proposed estimators.
4. Wu et al. (2005) and Huang et al. (2006) applied a hierarchical Bayesian method to estimate dynamic parameters in differential equation models; however, they focused on constant parameters instead of time-varying parameters as

we do in this article. Their methods also require numerous numerical evaluations of the nonlinear differential equations, and sometimes the convergence is problematic.

5. Wu and Ding (1999), Wu (2005), and other related work by Wu and his group have mainly applied mixed-effects models to estimate the dynamic parameters in differential equation models. But they usually simplified the nonlinear differential equation models first by making some assumptions, so that a closed-form solution to the simplified version of the models could be obtained. Then the problem can be reduced to a standard nonlinear mixed-effects regression model. These simplification methods may suffer from the approximation errors and are not robust when additional assumptions are made.

Our new methods avoid these problems.

In summary, we believe that the proposed models and methods are novel for estimating the time-varying parameters in differential equation models, which are very important and useful for modeling HIV dynamics and other dynamic biomedical processes. We also acknowledge some limitations of our methods. The bandwidth selection for both stages may require some attention to achieve the optimal estimates. We expect that more formal and efficient inference procedures for the dynamic models and their parameters will be developed in the future. It will be also interesting and useful to extend the proposed models and the estimation methods to longitudinal data.

APPENDIX: PROOFS OF THEOREMS

Here we impose some technical regularity conditions for Theorems 1, 2, 3, and 4. For each given point t_0, the following conditions are needed:

1. The design time points t_i, $i=1,\dots,n$, are iid with density function $f(t)$, which has continuous first derivative in some neighborhood of t_0 and $f(t_0)\neq 0$.
2. The function $g^{(3)}[\cdot]$ is continuous, and $g^{(1)}[X(t_0)]\neq 0$.
3. The functions $\sigma^2(t)$, $X^{(4)}(t)$, and $a_j^{(q+1)}(t)$, $j=1,\dots,d$, are continuous in some neighborhood of t_0.
4. The functions $Z_j(t)$, $j=1,\dots,d$, are continuous in some neighborhood of t_0, and $\Gamma(t_0)$ is a positive matrix.
5. The kernel function $K(\cdot)$ is a symmetric density function with a compact support.

The following lemma will be needed in the proofs of Theorems 1, 2, and 3. For convenience of expression, let $b_{n1}=h_1^2+(\log(1/h_1)/nh_1)^{1/2}$ and $b_{n2}=h_1^2+(\log(1/h_1)/nh_1^3)^{1/2}$. We outline the key idea in the proofs of the Theorems 1, 2, 3, and 4.

Lemma 1. Let $(t_1,Y_n),\dots,(t_n,Y_n)$ be iid random vectors, where the Y_i's are scalar random variables. Assume further that $E|Y|^s<\infty$ and $\sup_t\int|y|^s f(t,y)\,dy<\infty$, where f denotes the joint density of (t,Y). Let K be a bounded positive function with a bounded support, satisfying the Lipschitz condition. Then

$$\sup_{t\in D}\left|n^{-1}\sum_{i=1}^n\{K_h(t_i-t_0)Y_i-E[K_h(t_i-t_0)Y_i]\}\right| = O_P\{[nh/\log(1/h)]^{-1/2}\}, \quad (\text{A.1})$$

provided that $n^{2\varepsilon-1}h\to\infty$ for $\varepsilon<1-s^{-1}$.

Proof. The proof follows immediately from the result obtained by Mack and Silverman (1982).

Proof of Theorem 1

Let $\widehat{\mathbf{a}}_L^*(t_0)=\sqrt{nh_2}\{\widehat{\mathbf{a}}_L(t_0)-\mathbf{a}(t_0)\}$. By Taylor's expansion, we obtain

$$\begin{pmatrix}\gamma^T(t_1)\mathbf{a}(t_1)\\ \vdots\\ \gamma^T(t_n)\mathbf{a}(t_n)\end{pmatrix}=\mathbf{Z}\alpha(t_0)+\frac{1}{(q+1)!}\times\begin{pmatrix}\gamma^T(t_1)\mathbf{a}^{(q+1)}(\eta_1)(t_1-t_0)^{q+1}\\ \vdots\\ \gamma^T(t_n)\mathbf{a}^{(q+1)}(\eta_n)(t_n-t_0)^{q+1}\end{pmatrix}, \quad (\text{A.2})$$

where $\alpha(t_0)=(a_1(t_0),\dots,a_1^{(q)}(t_0)/q!,\dots,a_d(t_0),\dots,a_d^{(q)}(t_0)/q!)^T$ and η_i is between t_i and t_0 for $i=1,\dots,n$. Based on (4), (11), and (A.2), $\widehat{\mathbf{a}}_L^*(t_0)$ can be expressed as

$$\begin{aligned}\widehat{\mathbf{a}}_L^*(t_0)&=\sqrt{nh_2}\{(I_d\otimes e_{1,q+1}^T)(\mathbf{Z}_n^TW_{n2}\mathbf{Z}_n)^{-1}\\ &\quad\times\mathbf{Z}_n^TW_{n2}(\mathbf{U}_{1,1}+\mathbf{g}[\mathbf{U}_{0,1}])-\mathbf{a}(t_0)\}\\ &=\sqrt{nh_2}(I_d\otimes e_{1,q+1}^T)\left(\frac{1}{n}\mathbf{Z}_n^TW_{n2}\mathbf{Z}_n\right)^{-1}\\ &\quad\times(\mathbf{A}_{n1}+\mathbf{A}_{n2}),\end{aligned}$$

where

$$\mathbf{A}_{n1}=\frac{1}{(q+1)!}\frac{1}{n}\sum_{i=1}^n(\Gamma(t_i)\otimes T_i)K_{h_2}(t_i-t_0)\times\mathbf{a}^{(q+1)}(\eta_i)(t_i-t_0)^{q+1}$$

and

$$\mathbf{A}_{n2}=\frac{1}{n}\sum_{i=1}^n(\gamma(t_i)\otimes T_i)K_{h_2}(t_i-t_0)\times\big[\big(\widehat{X}_1^{(1)}(t_i)-X^{(1)}(t_i)\big)+\big(g[\widehat{X}_1(t_i)]-g[X(t_i)]\big)\big],$$

with $T_i=(1,(t_i-t_0),\dots,(t_i-t_0)^q)^T$, $i=1,\dots,n$. Using the fact that $(\mathbf{A}_n)_{ij}=(E\mathbf{A}_n)_{ij}+O_P[\{\text{var}(\mathbf{A}_n)_{ij}\}^{1/2}]$, it can be shown that

$$\begin{aligned}\frac{1}{n}\mathbf{Z}_n^TW_{n2}\mathbf{Z}_n&=f(t_0)(\Gamma(t_0)\otimes H_2SH_2)(1+o_P(1))\\ &\equiv\mathbf{B}(1+o_P(1))\end{aligned}\quad(\text{A.3})$$

and

$$\begin{aligned}\mathbf{A}_{n1}&=\frac{1}{(q+1)!}E\{(\Gamma(t_1)\otimes T_1)K_{h_2}(t_1-t_0)\\ &\quad\times\mathbf{a}^{(q+1)}(\eta_1)(t_1-t_0)^{q+1}\}(1+o_P(1))\\ &=\frac{1}{(q+1)!}f(t_0)h_2^{q+1}(\Gamma(t_0)\otimes H_2c_q)\mathbf{a}^{(q+1)}(t_0)+o_P(h_2^{q+1})\\ &\equiv\mathbf{A}_1+o_P(h_2^{q+1}),\end{aligned}\quad(\text{A.4})$$

where $c_q=(\mu_{q+1},\dots,\mu_{2q+1})^T$. Using the results of Fan and Gijbels (1996) and Huang and Fan (1998), we obtain that

$$\begin{aligned}\widehat{X}_1(t_i)&=X(t_i)+\frac{1}{2!}\mu_2h_1^2X^{(2)}(t_i),\\ &\quad+\frac{1}{nf(t_i)}\sum_{k=1}^nK_{h_1}(t_k-t_i)\sigma(t_k)e(t_k)+o_P(b_{n1}),\end{aligned}$$

$$\widehat{X}_1^{(1)}(t_i)=X_1^{(1)}(t_i)+\frac{1}{3!}\frac{\mu_4}{\mu_2}h_1^2\left(X^{(3)}(t_i)+3X^{(2)}(t_i)\frac{f^{(1)}(t_i)}{f(t_i)}\right)$$

$$+\frac{1}{\mu_2 n h_1^2 f(t_i)}\sum_{k=1}^{n}(t_k-t_i)K_{h_1}(t_k-t_i)\sigma(t_k)e(t_k)$$
$$+o_P(b_{n2}), \tag{A.5}$$

uniformly in a neighborhood of t_i for each $i=1,\ldots,n$. Based on Lemma A.1 and Taylor's expansion, for each $i=1,\ldots,n$, we find that

$$g[\widehat{X}_1(t_i)] = g[X(t_i)] + g^{(1)}[X(t_i)](\widehat{X}_1(t_i)-X(t_i)) + O_P(b_{n1}^2). \tag{A.6}$$

Note that the term $o_P(1)$ holds uniformly in i such that t_i falls in the neighborhood of t_0. It follows from (A.3), (A.5), and (A.6) that

$$\begin{aligned}
\mathbf{A}_{n2} &= \frac{1}{n}\sum_{i=1}^{n}(\gamma(t_i)\otimes T_i)K_{h_2}(t_i-t_0)h_1^2\\
&\quad\times\left[\frac{1}{2!}\mu_2 X^{(2)}(t_i)g^{(1)}[X(t_i)]\right.\\
&\quad\left.+\frac{1}{3!}\frac{\mu_4}{\mu_2}\left(X^{(3)}(t_i)+3X^{(2)}(t_i)\frac{f^{(1)}(t_i)}{f(t_i)}\right)\right]\\
&\quad+\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n}\frac{1}{\mu_2 n h_1^2 f(t_i)}(\gamma(t_i)\otimes T_i)(t_k-t_i)\\
&\quad\times K_{h_2}(t_i-t_0)K_{h_1}(t_k-t_i)\sigma(t_k)e(t_k)\\
&\quad+\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n}\frac{1}{nf(t_i)}(\gamma(t_i)\otimes T_i)K_{h_2}(t_i-t_0)\\
&\quad\times K_{h_1}(t_k-t_i)g^{(1)}[X(t_i)]\sigma(t_k)e(t_k)\\
&\quad+o_P(b_{n1}+b_{n2})\\
&\equiv \mathbf{D}_{n1}+\mathbf{D}_{n2}+\mathbf{D}_{n3}+o_P(b_{n1}+b_{n2}).
\end{aligned} \tag{A.7}$$

It is easily seen by calculating the second moment that

$$\begin{aligned}
\mathbf{D}_{n1} &= \frac{1}{3!}f(t_0)h_1^2(\gamma(t_0)\otimes H_2U_0)\\
&\quad\times\frac{\mu_4}{\mu_2}\left(X^{(3)}(t_0)+3X^{(2)}(t_0)\frac{f^{(1)}(t_0)}{f(t_0)}\right)\\
&\quad+\frac{1}{2!}f(t_0)h_1^2(\gamma(t_0)\otimes H_2U_0)\mu_2X^{(2)}(t_0)g^{(1)}[X(t_0)]\\
&\quad+o_P(b_{n1})\\
&\equiv \mathbf{D}_1'+o_P(b_{n1}),
\end{aligned} \tag{A.8}$$

$$\begin{aligned}
\mathbf{D}_{n3} &= \frac{1}{n}\sum_{k=1}^{n}(\gamma(t_k)\otimes T_k)K_{h_2}(t_k-t_0)g^{(1)}[X(t_k)]\sigma(t_k)e(t_k)\\
&\quad+o_P(b_{n1})\\
&\equiv \mathbf{D}_{n3}'+o_P(b_{n1}),
\end{aligned}$$

with $H_2=\operatorname{diag}(1,\ldots,h_2^q)$ and $U_0=(1,\mu_1,\ldots,\mu_q)^T$. In addition, we have $\mathbf{D}_{n2}=o_P(b_{n2})$, because if we let

$$d_{ki}=\frac{1}{f(t_i)}(\gamma(t_i)\otimes T_i)(t_k-t_i)K_{h_2}(t_i-t_0)K_{h_1}(t_k-t_i)\sigma(t_k)e(t_k),$$

then $E(\mathbf{D}_{n2})=o(b_{n2})$ and

$$\begin{aligned}
&\operatorname{var}(\mathbf{D}_{n2})\\
&= E\left\{\frac{1}{\mu_2^2n^4h_1^4}\left[\sum_{k=k',i=i'}+\sum_{k=k',i\neq i'}\right]d_{ki}d_{k'i'}^T\right\}+o(b_{n2}^2)\\
&= \frac{\nu_2}{\mu_2^2n^3h_1^3}E\left\{\sum_{k=1}^{n}\frac{1}{f(t_k)}(\Gamma(t_k)\otimes T_kT_k^T)K_{h_2}^2(t_k-t_0)\sigma^2(t_k)\right\}\\
&\quad+o(b_{n2}^2)\\
&= O\left(\frac{1}{nh_2nh_1^3}\right)+o(b_{n2}^2).
\end{aligned}$$

Using the results (A.3), (A.3), (A.4), (A.7), and (A.8), the estimator $\widehat{\mathbf{a}}_L^*(t_0)$ can be rewritten as

$$\begin{aligned}
\widehat{\mathbf{a}}_L^*(t_0) &= \sqrt{nh_2}(I_d\otimes e_{1,q+1}^T)\\
&\quad\times\mathbf{B}^{-1}\{\mathbf{A}_1+\mathbf{D}_1'+\mathbf{D}_{n3}'+o_P(h_2^{q+1}+h_1^2)\}\\
&= \sqrt{nh_2}\{\lambda_L(t_0)+o_P(h_2^{q+1}+h_1^2)\}\\
&\quad+\frac{\sqrt{nh_2}}{nf(t_0)}\left(\Gamma^{-1}(t_0)\otimes e_{1,q+1}^TS^{-1}H_2^{-1}\right)\\
&\quad\times\sum_{k=1}^{n}(\gamma(t_k)\otimes T_k)K_{h_2}(t_k-t_0)g^{(1)}[X(t_k)]\sigma(t_k)e(t_k)\\
&\equiv \sqrt{nh_2}\{\lambda_L(t_0)+o_P(h_2^{q+1}+h_1^2)\}+\mathbf{V}_n,
\end{aligned} \tag{A.9}$$

where

$$\begin{aligned}
\lambda_L(t_0) &= \frac{1}{(q+1)!}h_2^{q+1}e_{1,q+1}^TS^{-1}c_q\mathbf{a}^{(q+1)}(t_0)\\
&\quad+\frac{1}{2}h_1^2\mu_2X^{(2)}(t_0)g^{(1)}[X(t_0)]\Gamma(t_0)^{-1}\gamma(t_0)\\
&\quad+\frac{1}{3!}h_1^2\frac{\mu_4}{\mu_2}\left(3\frac{f'(t_0)}{f(t_0)}X^{(2)}(t_0)+X^{(3)}(t_0)\right)\Gamma(t_0)^{-1}\gamma(t_0).
\end{aligned}$$

Because $\mathbf{V}_n$ is a sum of iid random vectors, we can establish the asymptotic normality of $\sqrt{nh_2}\{\widehat{\mathbf{a}}_L(t_0)-\mathbf{a}(t_0)-\lambda_L(t_0)\}$ by calculating the first two moments. Obviously, we obtain that the expectation of $\mathbf{V}_n$ is 0 and the variance is

$$\begin{aligned}
\operatorname{var}(\mathbf{V}_n) &= \frac{nh_2}{nf^2(t_0)}\left(\Gamma^{-1}(t_0)\otimes e_{1,q+1}^TS^{-1}H_2^{-1}\right)\\
&\quad\times\operatorname{var}\left\{(\gamma(t_1)\otimes T_1)K_{h_2}(t_1-t_0)g^{(1)}[X(t_1)]\sigma(t_i)e(t_1)\right\}\\
&\quad\times\left(\Gamma^{-1}(t_0)\otimes H_2^{-1}S^{-1}e_{1,q+1}\right)\\
&= \frac{\sigma^2(t_0)}{f(t_0)}\left\{g^{(1)}[X(t_0)]\right\}^2\Gamma^{-1}(t_0)\\
&\quad\times e_{1,q+1}^TS^{-1}S^*S^{-1}e_{1,q+1}(1+o_P(1)).
\end{aligned} \tag{A.10}$$

By the central limit theorem, we have

$$\sqrt{nh_2}\{\widehat{\mathbf{a}}_L(t_0)-\mathbf{a}(t_0)-\lambda_L(t_0)+o_P(h_2^{q+1}+h_1^2)\}\xrightarrow{D}\mathrm{N}(0,\operatorname{var}(\mathbf{V}_n)).$$

This, together with (A.9) and (A.10), proves the result in Theorem 1.

Proof of Theorem 2

Let $\widehat{\mathbf{a}}_Q^*(t_0)=\sqrt{nh_2}\{\widehat{\mathbf{a}}_Q(t_0)-\mathbf{a}(t_0)\}$ and $b_{1n}'=h_1^4+[\frac{\log(1/h_1)}{nh_1}]^{1/2}$. Similar to the (A.2)–(A.3) in the proof of Theorem 1, we find that

$$\begin{aligned}
\widehat{\mathbf{a}}_Q^*(t_0) &= \sqrt{nh_2}\{(I_d\otimes e_{1,q+1}^T)(\mathbf{Z}_n^TW_{n2}\mathbf{Z}_n)^{-1}\\
&\quad\times\mathbf{Z}^TW_2(\mathbf{U}_{1,2}+\mathbf{g}[\mathbf{U}_{0,2}])-\mathbf{a}(t_0)\}\\
&= \sqrt{nh_2}(I_d\otimes e_{1,q+1}^T)\left(\frac{1}{n}\mathbf{Z}_n^TW_{n2}\mathbf{Z}_n\right)^{-1}\\
&\quad\times(\mathbf{A}_{n1}+\mathbf{A}_{n2}),
\end{aligned} \tag{A.11}$$

where

$$\mathbf{A}_{n1} = \frac{1}{(q+1)!}\frac{1}{n}\sum_{i=1}^{n}(\Gamma(t_i)\otimes T_i)K_{h_2}(t_i - t_0) \times \mathbf{a}^{(q+1)}(\eta_i)(\eta_i - t_0)^{q+1}$$

and

$$\mathbf{A}_{n2} = \frac{1}{n}\sum_{i=1}^{n}(\gamma(t_i)\otimes T_i)K_{h_2}(t_i - t_0) \times \big[\big(\widehat{X}_2^{(1)}(t_i) - X^{(1)}(t_i)\big) + \big(g[\widehat{X}_2(t_i)] - g[X(t_i)]\big)\big].$$

Again, using the results of Fan and Gijbels (1996) and Huang and Fan (1998), we obtain that

$$\begin{aligned}\widehat{X}_2(t_i) &= X(t_i) + \frac{1}{4!}\frac{\mu_4^2 - \mu_2\mu_6}{\mu_4 - \mu_2^2}h_1^4\left(X^{(4)}(t_i) + 4X^{(2)}(t_i)\frac{f^{(1)}(t_i)}{f(t_i)}\right)\\ &\quad + \frac{1}{nf(t_i)(\mu_4 - \mu_2^2)}\sum_{k=1}^{n}\left[\mu_4 - \mu_2\frac{(t_k - t_i)^2}{h_1^2}\right]\\ &\quad \times K_{h_1}(t_k - t_i)\sigma(t_k)e(t_k) + o_P(b'_{1n}),\\ \widehat{X}_2^{(1)}(t_i) &= X_1^{(1)}(t_i) + \frac{1}{3!}\frac{\mu_4}{\mu_2}h_1^2X^{(3)}(t_i)\\ &\quad + \frac{1}{\mu_2 nh_1^2 f(t_i)}\sum_{k=1}^{n}(t_k - t_i)K_{h_1}(t_k - t_i)\sigma(t_k)e(t_k)\\ &\quad + o_P(b_{n2}),\end{aligned} \tag{A.12}$$

uniformly in a neighborhood of t_i for each $i = 1, \ldots, n$. Based on Lemma A.1 and Taylor's expansion, for each $i = 1, \ldots, n$, we find that

$$g[\widehat{X}_2(t_i)] = g[X(t_i)] + g^{(1)}[X(t_i)]\big(\widehat{X}_2(t_i) - X(t_i)\big) + O_P(b_{n1}'^2). \tag{A.13}$$

Note that the term $o_P(1)$ holds uniformly in i such that t_i falls in the neighborhood of t_0. It follows from (A.11), (A.12), and (A.13) that

$$\begin{aligned}\mathbf{A}_{n2} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{3!}\frac{\mu_4}{\mu_2}(\gamma(t_i)\otimes T_i)K_{h_2}(t_i - t_0)h_1^2X^{(3)}(t_i)\\ &\quad + \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n}\frac{1}{\mu_2 nh_1^2 f(t_i)}(\gamma(t_i)\otimes T_i)\\ &\quad \times K_{h_2}(t_i - t_0)K_{h_1}(t_k - t_i)(t_k - t_i)\sigma(t_k)e(t_k)\\ &\quad + \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n}\frac{1}{nf(t_i)(\mu_4 - \mu_2^2)}(\gamma(t_i)\otimes T_i)K_{h_2}(t_i - t_0)\\ &\quad \times \left[\mu_4 - \mu_2\frac{(t_k - t_i)^2}{h_1^2}\right]K_{h_1}(t_k - t_i)g^{(1)}[X(t_i)]\sigma(t_k)e(t_k)\\ &\quad + o_P(b_{n1} + b_{n2})\\ &\equiv \mathbf{D}_{n1} + \mathbf{D}_{n2} + \mathbf{D}_{n3} + o_P(b_{n1} + b_{n2}).\end{aligned} \tag{A.14}$$

By calculating the second moment, we obtain that

$$\begin{aligned}&\mathbf{D}_{n1} - \mathbf{D}'_1 \xrightarrow{P} 0, \qquad \mathbf{D}_{n2} \xrightarrow{P} 0, \qquad \text{and}\\ &\mathbf{D}_{n3} - \mathbf{D}'_{n3} \xrightarrow{P} 0,\end{aligned} \tag{A.15}$$

where

$$\mathbf{D}'_1 = \frac{1}{3!}f(t_0)\frac{\mu_4}{\mu_2}h_1^2(\gamma(t_0)\otimes H_2U_0)X^{(3)}(t_0)$$

and

$$\mathbf{D}'_{n3} = \frac{1}{n}\sum_{k=1}^{n}(\gamma(t_k)\otimes T_k)K_{h_2}(t_k - t_0)g^{(1)}[X(t_k)]\sigma(t_k)e(t_k).$$

Using the results (A.3) and (A.4) in the proof of Theorem 1, along with (A.11), (A.14), and (A.15), we find that the estimator $\widehat{\mathbf{a}}_Q^*(t_0) = \sqrt{nh_2}\{\widehat{\mathbf{a}}_Q(t_0) - \mathbf{a}(t_0)\}$ can be expressed as

$$\begin{aligned}\widehat{\mathbf{a}}_Q^*(t_0) &= \sqrt{nh_2}(I_d\otimes e_{1,q+1}^T)\mathbf{B}^{-1}\\ &\quad \times \{\mathbf{A}_1 + \mathbf{D}'_1 + \mathbf{D}'_{n3} + o_P(h_2^{q+1} + h_1^2)\}\\ &= \sqrt{nh_2}\{\lambda_Q(t_0) + o_P(h_2^{q+1} + h_1^2)\}\\ &\quad + \frac{\sqrt{nh_2}}{nf(t_0)}\big(\Gamma^{-1}(t_0)\otimes e_{1,q+1}^T S^{-1}H_2^{-1}\big)\\ &\quad \times \sum_{k=1}^{n}(\gamma(t_k)\otimes T_k)K_{h_2}(t_k - t_0)g^{(1)}[X(t_k)]\sigma(t_k)e(t_k)\\ &\equiv \sqrt{nh_2}\{\lambda_Q(t_0) + o_P(h_2^{q+1} + h_1^2)\} + \mathbf{V}_n,\end{aligned} \tag{A.16}$$

where

$$\lambda_Q(t_0) = \frac{1}{(q+1)!}h_2^{q+1}e_{1,q+1}^T S^{-1}c_q\mathbf{a}^{(q+1)}(t_0) + \frac{1}{3!}h_1^2\frac{\mu_4}{\mu_2}X^{(3)}(t_0)\Gamma(t_0)^{-1}\gamma(t_0).$$

By a similar argument as in the proof of Theorem 1, we establish the asymptotic normality of $\sqrt{nh_2}\{\widehat{\mathbf{a}}_Q(t_0) - \mathbf{a}(t_0) - \lambda_Q(t_0)\}$ with zero expectation and variance

$$\operatorname{var}(\mathbf{V}_n) = \frac{\sigma^2(t_0)}{f(t_0)}\big\{g^{(1)}[X(t_0)]\big\}^2\Gamma^{-1}(t_0) \times e_{1,q+1}^T S^{-1}S^*S^{-1}e_{1,q+1}(1 + o_P(1)).$$

Therefore, the proof of Theorem 2 has been completed directly.

Proof of Theorem 3

The proof of Theorem 3 is completed along the same lines as that of Theorems 1 and 2.

Proof of Theorem 4

We denote $\Delta_i = t_{i+1} - t_i$, $\Delta e(t_i) = \sigma(t_{i+1})e(t_{i+1}) - \sigma(t_i)e(t_i)$, and $\Delta Y(t_i) = Y(t_{i+1}) - Y(t_i)$, $i = 1, \ldots, n-1$. Let $\mathbf{Y}^* = (\frac{\Delta Y(t_1)}{\Delta_1}, \ldots, \frac{\Delta Y(t_{n-i})}{\Delta_{n-1}})^T$ and $\mathbf{g} = (g[\widehat{X}(t_1)], \ldots, g[\widehat{X}(t_{n-1})])^T$. Write $S_n = \mathbf{Z}_{n-1}^T W_{(n-1)2}\mathbf{Z}_{n-1}$ and $T_n = \mathbf{Z}_{n-1}^T W_{(n-1)2}(\mathbf{Y}^* + \mathbf{g})$. Then, by (17) and (A.2), we can write $\widehat{\mathbf{a}}_D(t_0) = (I_d\otimes e_{1,q+1}^T)S_n^{-1}T_n$, and thus

$$\widehat{\mathbf{a}}_D(t_0) - \mathbf{a}(t_0) = (I_d\otimes e_{1,q+1}^T)nS_n^{-1}(\mathbf{A}_{n1} + \mathbf{A}_{n2}), \tag{A.17}$$

where

$$\mathbf{A}_{n1} = \frac{1}{(q+1)!}\frac{1}{n}\sum_{i=1}^{n-1}(\Gamma(t_i)\otimes T_i)K_{h_2}(t_i - t_0) \times \mathbf{a}^{(q+1)}(\eta_i)(t_i - t_0)^{q+1}$$

and

$$\begin{aligned}\mathbf{A}_{n2} &= \frac{1}{n}\sum_{i=1}^{n-1}(\gamma(t_i)\otimes T_i)K_{h_2}(t_i - t_0)\\ &\quad \times \left\{\left[\frac{\Delta Y(t_i)}{\Delta_i} - \gamma^T(t_i)\mathbf{a}(t_i) + g[X(t_i)]\right]\right.\\ &\quad \left. + \big(g[\widehat{X}_1(t_i)] - g[X(t_i)]\big)\right\}.\end{aligned}$$

We first establish the conditional asymptotic bias of the direct local estimator $\widehat{\mathbf{a}}_D(t_0)$. By Taylor's expansion, for $i = 1, \ldots, n-1$, we find that

$$\frac{X(t_{i+1}) - X(t_i)}{\Delta_i} = \gamma^T(t_i)\mathbf{a}(t_i) - g[X(t_i)] + \frac{1}{2}X^{(2)}(t_i)\Delta_i + O(\Delta_i^2) \quad \text{(A.18)}$$

and

$$g[\widehat{X}_1(t_i)] - g[X(t_i)] = g^{(1)}[X(t_i)](\widehat{X}_1(t_i) - X(t_i)) + g^{(2)}[s_i](\widehat{X}_1(t_i) - X(t_i))^2, \quad \text{(A.19)}$$

where s_i is between $\widehat{X}_1(t_i)$ and $X(t_i)$ for $i = 1, \ldots, n-1$. Note that $g^{(2)}[\cdot]$ is a bounded function. Then the expected value of the term $g^{(2)}[s_i](\widehat{X}_1(t_i) - X(t_i))^2$ in (A.19) is bounded by $O_p(1/nh_1)$. Note that $S_n = (n-1)f(t_0)(\Gamma(t_0) \otimes H_2SH_2)(1 + o_P(1))$. Using (A.3), (A.17), (A.18), and (A.19), the conditional asymptotic bias of $\widehat{\mathbf{a}}_D(t_0)$ is given by

$$\begin{aligned}
&\operatorname{bias}(\widehat{\mathbf{a}}_D(t_0)|\mathcal{D}) \\
&= (I_d \otimes e_{1,q+1}^T)nS_n^{-1} \\
&\quad \times \Bigg\{ \frac{1}{(q+1)!} f(t_0)h_2^{q+1}(\Gamma(t_0) \otimes H_2c_q)\mathbf{a}^{(q+1)}(t_0) \\
&\quad + o_P(h_2^{q+1}) + \frac{1}{n}\sum_{i=1}^{n-1}(\gamma(t_i) \otimes T_i)K_{h_2}(t_i - t_0) \\
&\quad \times \Bigg[\bigg(\frac{\Delta X(t_i)}{\Delta_i} - \gamma^T(t_i)\mathbf{a}(t_i) + g[X(t_i)]\bigg) \\
&\quad + E\big(g[\widehat{X}_1(t_i)] - g[X(t_i)]|\mathcal{D}\big)\Bigg]\Bigg\} \\
&= \frac{1}{2}X^{(2)}(t_0)\Delta\gamma(t_0) + O(\Delta^2) \\
&\quad + \frac{h_2^{q+1}}{(q+1)!}e_{1,q+1}^T S^{-1}c_q\mathbf{a}^{(q+1)}(t_0) \\
&\quad + \frac{1}{2}h_1^2\mu_2X^{(2)}(t_0)g^{(1)}[X(t_0)]\Gamma(t_0)^{-1}\gamma(t_0) \\
&\quad + o_P(h_1^2 + h_2^{q+1}). \qquad \text{(A.20)}
\end{aligned}$$

Next we consider the conditional asymptotic variance of $\widehat{\mathbf{a}}_D(t_0)$. By Taylor's expansion, we have $g[\widehat{X}_1(t_i)] = g[X(t_i)] + g^{(1)}[s_i'](\widehat{X}_1(t_i) - X(t_i))$, where s_i' is between $\widehat{X}_1(t_i)$ and $X(t_i)$ for $i = 1, \ldots, n-1$. By (A.17), (A.18), and (A.19), we have

$$\operatorname{var}(\widehat{\mathbf{a}}_D(t_0)|\mathcal{D}) = (I_d \otimes e_{1,q+1}^T)n^2S_n^{-1}\big[\operatorname{var}(I_{n1}|\mathcal{D}) + \operatorname{var}(I_{n2}|\mathcal{D}) + 2\operatorname{cov}(I_{n1}, I_{n2}|\mathcal{D})\big]S_n^{-1}(I_d \otimes e_{1,q+1}), \quad \text{(A.21)}$$

where $I_{n1} = \frac{1}{n}\sum_{i=1}^{n-1}(\gamma(t_i) \otimes T_i)K_{h_2}(t_i - t_0)\frac{\Delta e(t_i)}{\Delta_i}$ and $I_{n2} = \frac{1}{n}\sum_{i=1}^{n-1}(\gamma(t_i) \otimes T_i)K_{h_2}(t_i - t_0)g^{(1)}[s_i'](\widehat{X}_1(t_i) - X(t_i))$. Again, using Taylor's expansion, we obtain that

$$\begin{aligned}
\operatorname{var}(I_{n1}|\mathcal{D}) &= \frac{1}{n^2\Delta^2}\sum_{i=1}^{n-1}(\Gamma(t_i) \otimes T_iT_i^T)K_{h_2}^2(t_i - t_0) \\
&\quad \times (\sigma^2(t_{i+1}) + \sigma^2(t_i)) \\
&\quad - \frac{2}{n^2\Delta^2}\sum_{i=1}^{n-2}(\gamma(t_i) \otimes T_i)(\gamma(t_{i+1})^T \otimes T_{i+1}^T) \\
&\quad \times K_{h_2}(t_i - t_0)K_{h_2}(t_{i+1} - t_0)\sigma^2(t_{i+1}) \\
&= \frac{1}{n^2\Delta^2}\Bigg\{(\Gamma(t_1) \otimes T_1T_1^T)K_{h_2}^2(t_1 - t_0)\sigma^2(t_1) \\
&\quad + \sum_{i=1}^{n-1}(\Gamma(t_i) \otimes T_iT_i^T)K_{h_2}^2(t_i - t_0) \\
&\quad \times \big[\{\sigma^2(t_i)\}^{(1)}\Delta + O(\Delta^2)\big]\Bigg\} \\
&= \frac{f(t_0)\{\sigma^2(t_i)\}^{(1)}}{nh_2\Delta}(\Gamma(t_0) \otimes H_2S^*H_2) \\
&\quad \times (1 + o_P(1)). \qquad \text{(A.22)}
\end{aligned}$$

Because $g^{(1)}[\cdot]$ is a bounded function, the second term $\operatorname{var}(I_{n2}|\mathcal{D})$ in (A.21) is bounded by $O_p(\frac{1}{n^2h_1h_2}) = o_P(\frac{1}{nh_2\Delta})$. By Schwarz's inequality, the last term in (A.21) is bounded by $O_P(\frac{1}{nh_2\sqrt{nh_1\Delta}}) = o_P(\frac{1}{nh_2\Delta})$. Combining (A.21) and (A.22), leads to the expression of the conditional asymptotic variance in Theorem 4. Thus the proof is completed.

[Received November 2005. Revised August 2007.]

REFERENCES

Arfi, M. (1995), "Non-Parametric Drift Estimation From Estimation From Ergodic Samples," *Journal of Nonparametric Statistics*, 5, 381–389.

——— (1998), "Non-Parametric Variance Estimation From Estimation From Ergodic Samples," *Scandinavia Journal of Statistics*, 25, 225–334.

Chapman, D. A., and Pearson, N. D. (2000), "Is the Short Rate Drift Actually Nonlinear?" *Journal of Finance*, LV, 355–388.

Fan, J. (2005), "A Selective Overview of Nonparametric Methods in Financial Econometrics," *Statistical Science*, 20, 317–357.

Fan, J., and Chen, J. (1999), "One-Step Local Quasi-Likelihood Estimation," *Journal of the Royal Statistical Society*, Ser. B, 61, 927–943.

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

Fan, J., and Zhang, C. (2003), "A Re-Examination of Stanton's Diffusion Estimations With Applications to Financial Model Validation," *Journal of the American Statistical Association*, 98, 118–134.

Fan, J., and Zhang, W. (1999), "Statistical Estimation in Varying-Coefficient Models," *The Annals of Statistics*, 27, 1491–1518.

Fan, J., Jiang, J., Zhang, C., and Zhou, Z. (2003), "Time-Dependent Diffusion Model for Term Structure Dynamics and the Stock Price Volatility," *Statistica Sinica*, 13, 965–992.

Han, C., and Chaloner, K. (2004), "Bayesian Experimental Design for Nonlinear Mixed-Effects Models With Application to HIV Dynamics," *Biometrics*, 60, 25–33.

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*, 374, 123–126.

Huang, L. S., and Fan, J. (1998), "Nonparametric Estimation of Quadratic Regression Functionals," *Bernoulli*, 5, 927–949.

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., Rosenkranz, S. L., and Wu, H. (2003), "Modeling HIV Dynamics and Antiviral Responses With Consideration of Time-Varying Drug Exposures, Sensitivities and Adherence," *Mathematical Biosciences*, 184, 165–186.

Li, L., Brown, M. B., Lee, K., and Gupta, S. (2002), "Estimation and Inference for a Spline-Enhanced Population Pharmacokinetic Model," *Biometrics*, 58, 601–611.

Mack, Y. P., and Silverman, B. W. (1982), "Weak and Strong Uniform Consistency of Kernel Regression and Density Estimation," *Zeitschript für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 61, 405–415.

Michele, D. M., Ruy, M. R., Martin, M., Ho, D. D., and Perelson, A. S. (2004), "Modeling the Long-Term Control of Viremia in HIV-1–Infected Patients Treated With Antiretroviral Therapy," *Mathematical Biosciences*, 188, 47–62.

Perelson, A. S., Essunger, P., Cao, Y., 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, 188–191.

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.

Ruppert, D. (1997), "Empirical-Bias, Bandwidths for Local Polynomial Nonparametric Regression and Density Estimation," *Journal of the American Statistical Association*, 92, 1049–1062.

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.

Seifert, B., and Gasser, T. (1996), "Finite-Sample Variance of Local Polynomials: Analysis and Solutions," *Journal of the American Statistical Association*, 91, 267–275.

Stanton, H. (1997), "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *Journal of Finance*, 5, 1973–2002.

Wu, H. (2005), "Statistical Methods for HIV Dynamic Studies in AIDS Clinical Trials," *Statistical Methods in Medical Research*, 14, 171–192.

Wu, H., and Ding, A. A. (1999), "Population HIV-1 Dynamics in vivo: Applicable Models and Inferential Tool for Virological Data From AIDS Clinical Trials," *Biometrics*, 55, 410–418.

Wu, H., Huang, H., Acosta, E. P., Rosenkranz, S. L., Kuritzkes, D. R., Eron, J. J., Perelson, A. S., and Gerber, J. G. (2005), "Modeling Long-Term HIV Dynamics and Antiretroviral Response: Effects of Drug Potency, Pharmacokinetics, Adherence, and Drug Resistance," *JAIDS*, 39, 272–283.

Wu, L. (2002), "A Joint Model for Nonlinear Mixed-Effects Models With Censoring and Covariates Measured With Error, With Application to AIDS Studies," *Journal of the American Statistical Association*, 97, 955–963.