Statistica Sinica **18**(2008), 987-1006

ESTIMATION OF TIME-VARYING PARAMETERS IN DETERMINISTIC DYNAMIC MODELS

Jianwei Chen and Hulin Wu

San Diego State University and University of Rochester

Abstract: In this paper, we develop local polynomial estimation procedures to fit deterministic dynamic models with a focus on the estimation of time-varying parameters. Three local estimation methods for estimating time-varying parameters are proposed: two-step local linear estimation, two-step local quadratic estimation, and a two-step local hybrid method. Although the proposed estimation methods are applicable for general dynamic models, we establish the asymptotic properties of the proposed estimators for a linear deterministic dynamic model and show that the proposed estimators for linear dynamic models achieve the optimal convergence rate. Simulation studies reveal that the proposed two-step estimation methods perform better than the naive one-step local estimator. An application from an AIDS clinical trial is presented to illustrate the methodologies.

Key words and phrases: Asymptotic conditional bias and variance, deterministic dynamic models, differential equation models, HIV/AIDS, one-step local estimators, two-step local estimators, time-varying parameters.

1. Introduction

In the study of infectious diseases such as HIV and hepatitis viruses, modeling immune response and viral dynamics is critical for understanding pathogenesis of viral infection and providing guidance toward development of treatment strategies. Typically, the immune response and viral dynamic models are deterministic differential equations that may not have closed form solutions. See Perelson and Nelson (1999), Nowak and May (2000) and Tan and Wu (2005) for a good survey on these models. For example, the HIV dynamic model with antiviral treatment can be expressed as (Huang, Rosenkranz, Wu (2003), Michele et al. (2004), Wu (2005) and Huang, Liu and Wu (2006))

$$
\begin{aligned}
\frac{d}{dt}X_1(t) &= \lambda - k[1-r(t)]X_1(t)X_3(t) - \rho X_1(t) \equiv \theta_1(t) - \rho X_1(t),\\
\frac{d}{dt}X_2(t) &= k[1-r(t)]X_1(t)X_3(t) - \delta X_2(t) \equiv \theta_2(t) - \delta X_2(t),\\
\frac{d}{dt}X_3(t) &= N\delta X_2(t) - cX_3(t) \equiv \theta_3(t) - cX_3(t),
\end{aligned}
\tag{1.1}
$$

where $X_1(t)$ is the concentration of uninfected target cells, $X_2(t)$ is the concentration of infected cells, and $X_3(t)$ is the concentration of free plasma virus (viral load) at time t; ρ is the death rate of uninfected T cells, δ is the death rate of infected cells, and c is the clearance rate of free virions; $\theta_1(t) = \lambda - k[1 - r(t)]X_1(t)X_3(t)$ denotes the T cell production rate, $\theta_2(t) = k[1 - r(t)]X_1(t)X_3(t)$ denotes the rate of target cells becoming infected cells, and $\theta_3(t) = N\delta X_2(t)$ denotes the virus production rate at time t. The time-varying functions $\theta_1(t)$, $\theta_2(t)$ and $\theta_3(t)$ are three important indices that characterize viral dynamics in HIV infected patients. In general, it is difficult to establish a closed form solution to the nonlinear differential equation model (1.1) without making more assumptions. Perelson et al. (1996) considered a simple HIV dynamic model under the assumptions of a constant uninfected cell concentration $X_1(t)$ and a perfect antiviral treatment effect. Given these assumptions, closed form solutions of the infected cell concentration $X_2(t)$ and the viral load $X_3(t)$ are given by

$$\begin{aligned} X_2(t) &= \frac{cX_0}{N\delta(c-\delta)}\left[c\exp\{-\delta t\} - \delta\exp\{-ct\}\right], \\ X_3(t) &= X_0 e^{-ct} + \frac{cX_0}{c-\delta}\left\{\frac{c}{c-\delta}\left[e^{-\delta t} - e^{-ct}\right] - \delta t e^{-ct}\right\}, \end{aligned} \tag{1.2}$$

where $X_0 = X_3(0)$ is the initial value of viral load (baseline value). In AIDS clinical studies, the viral load $X_3(t)$ can be observed with measurement error. Thus the parameters δ and c can be estimated by fitting the nonlinear parametric model

$$Y_3(t) = X_3(t) + e_3(t), \tag{1.3}$$

$$X_3(t) = X_0 e^{-ct} + \frac{cX_0}{c-\delta}\left\{\frac{c}{c-\delta}\left[e^{-\delta t} - e^{-ct}\right] - \delta t e^{-ct}\right\}, \tag{1.4}$$

where $Y_3(t)$ is the observed viral load at time t, and $e_3(t)$ is the unobserved measurement error with mean zero and variance $\sigma_3^2(t)$. Furthermore, the time-varying parameter $\boldsymbol{\theta}(t) = [\theta_1(t), \theta_2(t), \theta_3(t)]^T$ can be estimated by using (1.1) with fixed c and δ. Over the past several years, a variety of nonlinear exponential-type parametric models have been used to fit short-term clinical data from AIDS clinical trials, see Ho et al. (1995), Perelson et al. (1997, 1996), Wu, Ding and DeGruttola (1998), Wu and Ding (1999), Han and Chaloner (2004) and Wu (2005), among others.

During long-term AIDS treatment, the antiviral treatment effect may be imperfect and the concentration of uninfected target cells may change over time. Thus short-term models such as (1.2) are not applicable. Since the dynamic model (1.1) is composed of differential equations without a closed form solution, one requires numerical solutions, see papers by Putter et al. (2002), Xia (2003), Michele et al. (2004) and Huang, Liu and Wu (2006). Numerical methods can lead to intensive computations when complex statistical methods are employed for parameter estimation.

In this paper we develop nonparametric smoothing methods for the unknown time-varying parameter $\boldsymbol{\theta}(t)$ in the general deterministic dynamic model specified in Section 2. In order to achieve this aim, first we need to estimate the state variable $\mathbf{X}(t)$ and its derivative $\mathbf{X}'(t)$. In Section 2, we introduce local estimation procedures for this purpose. Three local estimation methods for the time-varying parameter $\boldsymbol{\theta}(t)$ are proposed: two-step local linear estimation, two-step local quadratic estimation and a two-step local hybrid estimation. The proposed methods can avoid numerical solutions to the differential equations. In Section 3, we investigate the proposed two-step estimators for a linear dynamic model and derive explicit formulas for them. The asymptotic biases and variances of the estimators are obtained, and we show that the proposed estimators achieve the optimal convergence rate. We show that a naive one-step local estimator of the time-varying parameter $\boldsymbol{\theta}(t)$, unlike the proposed two-step estimators, does not achieve the optimal convergence rate. We also discuss bandwidth selection problems and interval estimates in Section 3. In Section 4, the proposed methods are illustrated and compared, via simulation, with the data generated from an HIV dynamic model. The proposed methods perform well in estimating the time-varying parameter. An application to HIV dynamics data from an AIDS clinical trial is also presented in Section 4. Some concluding remarks are given in Section 5. Technical proofs of the theorems are provided in an online supplement.

2. Models and Methodology

2.1. Deterministic dynamic models

A deterministic dynamic model can be written as

$$\frac{d}{dt}\mathbf{X}(t) = \mathbf{F}(\mathbf{X}(t), \boldsymbol{\theta}(t)), \tag{2.1}$$

where $\mathbf{X}(t) = [X_1(t), \ldots, X_m(t)]^T$ is an unobserved state vector, $\boldsymbol{\theta}(t) = [\theta_1(t), \ldots, \theta_m(t)]^T$ is an unknown time-varying parameter vector, and $\mathbf{F}(\cdot) = [F_1(\cdot), \ldots, F_m(\cdot)]^T$ is a known linear or nonlinear function vector. The dynamic model may also contain constant parameters, but we only focus on time-varying parameters in this paper. Some components of the state vector $\mathbf{X}(t)$ are often observable. A linear observation or measurement function can be written as

$$\mathbf{Y}(t) = \Gamma\mathbf{X}(t) + \mathbf{e}(t), \tag{2.2}$$

where Γ is usually a known design matrix, and $\mathbf{Y}(t) = [Y_1(t), \ldots, Y_m(t)]^T$ is an observation vector with measurement noise being $\mathbf{e}(t) = [e_1(t), \ldots, e_m(t)]^T$.

Dynamic systems (2.1) and (2.2) are widely used in modeling viral dynamics for infectious diseases, as mentioned, in Section 1. The goal of this paper is to apply local polynomial techniques to estimate the time-varying parameter $\boldsymbol{\theta}(t)$.

2.2. Two-step local pth order polynomial estimation

First we suggest a two-step local estimation procedure for (2.1)−(2.2). Assume that the component of the state vector $X_k(\cdot)$, $k = 1, \ldots, m$, possesses $p+1$ ($p \geq 1$) derivatives and that the component of the time-varying parameter vector $\theta_j(.)$, $j = 1, \ldots, m$, has a second order derivative. Write $\{Y_k(t_i), i = 1, \ldots, n\}$ as the observed data of the kth component of the state vector at time points: $t_1, \ldots, t_n$. The proposed procedure consists of two steps.

Step 1. For a given point t_0, use a pth order polynomial ($p \geq 1$) to fit model (2.2) to obtain the estimates of $\mathbf{X}(t_0)$ and $\mathbf{X}'(t_0) = d\mathbf{X}(t)/dt|_{t=t_0}$ simultaneously, say $\widehat{\mathbf{X}}_p(t_0) = [\widehat{X}_{1,p}(t_0), \ldots, \widehat{X}_{m,p}(t_0)]^T$ and $\widehat{\mathbf{X}}'_p(t_0) = [\widehat{X}'_{1,p}(t_0), \ldots, \widehat{X}'_{m,p}(t_0)]^T$.

Step 2. Substitute the estimates $\widehat{\mathbf{X}}_p(t_i)$ and $\widehat{\mathbf{X}}'_p(t_i)$, $i = 1, \ldots, n$, in the dynamic equation (2.1) to obtain the regression model

$$\widehat{\mathbf{X}}'_p(t_i) = \mathbf{F}\left(\widehat{\mathbf{X}}_p(t_i), \boldsymbol{\theta}(t_i)\right) + \mathbf{e}_2(t_i), \tag{2.3}$$

where $\mathbf{e}_2(t_i)$ denotes the substitution error vector, then apply a local linear method to estimate the time-varying parameter $\boldsymbol{\theta}(t_0)$.

The first step is to use a standard local polynomial smoothing method. Since the components of the state vector $\mathbf{X}(t)$ possess $p+1$ derivatives, for each given time point t_0 we approximate the function $X_k(t_i)$, $k = 1, \ldots, m$, by

$$X_k(t_i) \approx X_k(t_0) + (t_i - t_0)X_k^{(1)}(t_0) + \cdots + \frac{(t_i - t_0)^p X_k^{(p)}(t_0)}{p!} = \mathbf{z}_{i,p}^T \boldsymbol{\beta}_{k,p}(t_0),$$

for t_i, $i = 1, \ldots, n$, in a neighborhood of the point t_0, where $\boldsymbol{\beta}_{k,p}(t_0) = [\beta_{k,0}(t_0), \ldots, \beta_{k,p}\ (t_0)]^T$, with $\beta_{k,v}(t_0) = X_k^{(v)}(t_0)/v!$, $v = 0, \ldots, p$, and $\mathbf{z}_{i,p} = [1, (t_i - t_0) \ldots, (t_i - t_0)^p]^T$. Following the local polynomial fitting (Fan and Gijbels (1996)), the estimators $\widehat{\mathbf{X}}_p(t_0) = [\widehat{\beta}_{1,0}(t_0), \ldots, \widehat{\beta}_{m,0}(t_0)]^T$ and $\widehat{\mathbf{X}}'_p(t_0) = [\widehat{\beta}_{1,1}(t_0), \ldots, \widehat{\beta}_{m,1}\ (t_0)]^T$ can be obtained by minimizing the locally weighted least-squares criterion,

$$\sum_{k=1}^{m} \left(\mathbf{Y}_k - \Gamma \mathbf{Z}_p \boldsymbol{\beta}_{k,p}(t_0)\right)^T W_{h_k} \left(\mathbf{Y}_k - \Gamma \mathbf{Z}_p \boldsymbol{\beta}_{k,p}(t_0)\right), \tag{2.4}$$

with respect to $\boldsymbol{\beta}_{k,p}(t_0)$, $k = 0, \ldots, m$. Here $\mathbf{Y}_k = [Y_k(t_1), \ldots, Y_k(t_n)]^T$, $\mathbf{Z}_p = [\mathbf{z}_{1,p}, \ldots, \mathbf{z}_{n,p}]^T$ and $W_{h_k} = \mathrm{diag}(K_{h_k}(t_1 - t_0), \ldots, K_{h_k}(t_n - t_0))$, with $K_{h_k}(\cdot) =$

$K(\cdot/h_k)/h_k$ obtained by re-scaling a kernel function $K(\cdot)$ with bandwidth $h_k > 0$ to control the size of the associated neighborhood for the kth component.

In the second step, we apply a local linear method to estimate the time-varying parameter vector $\boldsymbol{\theta}(t)$ based on (2.3). We assume that the component functions $F_k(\cdot)$, $k = 1, \ldots, m$, have a continuous second derivative. Then $\theta_k(t_i)$, $k = 1, \ldots, m$, $i = 1, \ldots, n$, can be approximated around the point t_0 by

$$\theta_k(t_i) \approx \alpha_{k,0}(t_0) + \alpha_{k,1}(t_0)(t_i - t_0).$$

Thus, $\boldsymbol{\theta}(t_0) = [\theta_1(t_0), \ldots, \theta_m(t_0)]^T$ can be estimated via minimizing the locally weighted function

$$\sum_{k=1}^{m}\sum_{i=1}^{n}\left\{\widehat{X}'_{k,p}(t_i) - F_k\left(\widehat{\mathbf{X}}_p(t_i), (I_m \otimes \mathbf{z}_{i,1}^T)\boldsymbol{\alpha}(t_0)\right)\right\}^2 K_b(t_i - t_0), \tag{2.5}$$

where $\otimes$ denotes the Kronecker product and I_m denotes an m-dimensional identity matrix, $\boldsymbol{\alpha}(t_0) = [\boldsymbol{\alpha}_1^T(t_0), \ldots, \boldsymbol{\alpha}_m^T(t_0)]^T$ with $\boldsymbol{\alpha}_j(t_0) = [\alpha_{j,0}(t_0), \alpha_{j,1}(t_0)]^T$, while $K_b(\cdot)$ is a kernel function with b being a properly selected bandwidth (see Section 3.3). Let $\widehat{\boldsymbol{\alpha}}(t_0)$ minimize the locally weighted function (2.5). Then the local estimator of $\boldsymbol{\theta}(t_0)$ is $\widehat{\boldsymbol{\theta}}_p(t_0) = [\widehat{\alpha}_{1,0}(t_0), \ldots, \widehat{\alpha}_{m,0}(t_0)]^T$. We use the notation $\widehat{\boldsymbol{\theta}}_p(t_0)$ to denote a two-step pth order local polynomial estimator of $\boldsymbol{\theta}(t_0)$. In particular, $\widehat{\boldsymbol{\theta}}_1(t_0)$ and $\widehat{\boldsymbol{\theta}}_2(t_0)$ denote the two-step local linear estimator and the local quadratic estimator, respectively, and are popular special cases in practical implementations.

2.3. Two-step local hybrid estimation

The two-step estimation procedure estimates $\mathbf{X}(t_0)$ and $\mathbf{X}'(t_0)$ simultaneously in Step 1 using the same local polynomial approximation. This may not be efficient. In this subsection, we extend the procedure to a two-step local hybrid estimation method in which $\mathbf{X}(t_0)$ and its derivative $\mathbf{X}'(t_0)$ are estimated separately using different orders of local polynomial approximation. If necessary, different smoothing parameters may also be used. For example, we may estimate $\mathbf{X}(t_0)$ using a local linear estimator and $\mathbf{X}'(t_0)$ using a local quadratic estimator, since $\mathbf{X}'(t_0)$ needs a higher order polynomial approximation to achieve a similar approximation accuracy compared to that of $\mathbf{X}(t_0)$. The two-step local hybrid estimation procedure can be described as follows.

Step 1. Use a local linear estimator $\widehat{\mathbf{X}}_1(t_i) = [\widehat{X}_{1,1}(t_0), \ldots, \widehat{X}_{m,1}(t_0)]^T$ to estimate the function $\mathbf{X}(t_0)$, and use a local quadratic estimator $\widehat{\mathbf{X}}'_2(t_0) = [\widehat{X}'_{1,2}(t_0), \ldots, \widehat{X}'_{m,2}(t_0)]^T$ to estimate the derivative function $\mathbf{X}'(t_0)$.

Step 2. Substitute $\widehat{\mathbf{X}}_1(t_i)$ and $\widehat{\mathbf{X}}_2'(t_i)$ into (2.1) to obtain the regression model

$$\widehat{\mathbf{X}}_2'(t_i) = \mathbf{F}\left(\widehat{\mathbf{X}}_1(t_i), \boldsymbol{\theta}(t_i)\right) + \mathbf{e}_3(t_i), \tag{2.6}$$

where $\mathbf{e}_3(t_i)$ denotes the approximation error vector; use a local linear approach to estimate the time-varying parameter $\boldsymbol{\theta}(t_0)$ based on (2.6).

We denote the two-step local hybrid estimator of the time-varying parameter by $\widehat{\boldsymbol{\theta}}(t_0)$. We expect the two-step local hybrid estimator to be more efficient than the two-step local linear estimator or the two-step local quadratic estimator due to its increased flexibility. In order to evaluate the three estimation procedures, we study their asymptotic properties for a linear dynamic model in the next section.

3. Linear Deterministic Dynamic Model

3.1. Estimation

We derive the explicit formulas of the proposed estimators for a linear dynamic model. The asymptotic biases and variances of these estimators are then developed. We expect asymptotic results for the linear dynamic model to shed some light on the performance and behavior of the proposed general estimators. The linear dynamic model can be written as

$$\frac{d}{dt}\mathbf{X}(t) = \boldsymbol{\theta}(t) - \mathbf{a}\mathbf{X}(t), \tag{3.1}$$

$$\mathbf{Y}(t) = \mathbf{X}(t) + \mathbf{e}(t), \tag{3.2}$$

where $\mathbf{X}(t) = [X_1(t), \ldots, X_m(t)]^T$ is an unobserved state vector and $\mathbf{Y}(t) = [Y_1(t), \ldots, Y_m(t)]^T$ is a measurement vector of $\mathbf{X}(t)$ with $\mathbf{e}(t) = [e_1(t), \ldots, e_m(t)]^T$ being a the measurement error vector. We assume that $e_k(t)$, $k = 1, \ldots, m$, has mean zero and covariance $\mathrm{Cov}\left(e_k(t), e_k(s)\right) = \sigma_k^2(t) 1_{\{t=s\}}$, with $\mathrm{Cov}\left(\mathbf{e}(t), \mathbf{e}(t)\right) = \mathrm{diag}\left(\sigma_1^2(t), \ldots, \sigma_m^2(t)\right)$. The time-varying parameter $\boldsymbol{\theta}(t) = [\theta_1(t), \ldots, \theta_m(t)]^T$ is an unknown vector, while $\mathbf{a} = (a_1, \ldots, a_m)^T$, with $a_k = (a_{k,1}, \ldots, a_{k,m})^T$, is a known constant. This model is useful in modeling infectious diseases; for example, the HIV viral dynamic model (1.1) can be cast in this form with $\mathbf{a} = \mathrm{diag}\,(\rho, \delta, c)$.

It is worth noting that for the linear dynamic model (3.1)−(3.2), a naive estimator of $\boldsymbol{\theta}(t_0)$ is

$$\tilde{\boldsymbol{\theta}}_p(t_0) = \widehat{\mathbf{X}}_p'(t_0) + \mathbf{a}\widehat{\mathbf{X}}_p(t_0). \tag{3.3}$$

We call this direct substitution estimator a one-step local pth order polynomial estimator. In particular, we take $\tilde{\boldsymbol{\theta}}_1(t)$ and $\tilde{\boldsymbol{\theta}}_2(t)$ as the one-step local linear and local quadratic estimators, respectively. Similarly, $\mathbf{X}(t_0)$ and $\mathbf{X}'(t_0)$ can be estimated separately by different orders of the local polynomial approximation, say, local linear or local quadratic smoothing, denoted by $\widehat{\mathbf{X}}_1(t_0)$ and $\widehat{\mathbf{X}}'_2(t_0)$, respectively. Then we obtain an alternative one-step local hybrid estimator as

$$\tilde{\boldsymbol{\theta}}(t_0) = \widehat{\mathbf{X}}'_2(t_0) + \mathbf{a}\widehat{\mathbf{X}}_1(t_0). \tag{3.4}$$

In the following, we show that the square-root of the conditional mean squared errors for the naive one-step estimators can only achieve a convergence rate of $O_P(n^{-2/7})$; this is slower than those of the three two-step local estimators that are discussed below.

Based on (3.1)–(3.2), the solution to (2.4) can be expressed as

$$\widehat{\mathbf{X}}_p(t_0) = [\widehat{X}_{1,p}(t_0), \ldots, \widehat{X}_{m,p}(t_0)]^T, \tag{3.5}$$
$$\widehat{\mathbf{X}}'_p(t_0) = [\widehat{X}'_{1,p}(t_0), \ldots, \widehat{X}'_{m,p}(t_0)]^T, \tag{3.6}$$

with

$$\widehat{X}_{k,p}(t_0) = e^T_{1,p+1}\left(\mathbf{Z}_p^T W_{h_{k;0,p}}\mathbf{Z}_p\right)^{-1}\mathbf{Z}_p^T W_{h_{k;0,p}}\mathbf{Y}_k,$$
$$\widehat{X}'_{k,p}(t_0) = e^T_{2,p+1}\left(\mathbf{Z}_p^T W_{h_{k;1,p}}\mathbf{Z}_p\right)^{-1}\mathbf{Z}_p^T W_{h_{k;1,p}}\mathbf{Y}_k,$$

where $e_{v+1,p+1}$, $v = 1, 2$, is a $(p+1)$ dimensional vector with 1 on the $(v+1)^{th}$ position and 0 elsewhere, and $h_{k;0,p}$ and $h_{k;1,p}$ are the bandwidths for estimating $X_k(t_0)$ and $X'_k(t_0)$, respectively. The matrices $\mathbf{Y}_k$, $\mathbf{Z}_p$ and W_h were defined in Section 2.1.

For (3.1)–(3.2), the two-step local estimation is reduced to a linear locally weighed least squares problem. Based on (2.5), the two-step local pth order polynomial estimator $\widehat{\boldsymbol{\theta}}_p(t_0) = [\widehat{\alpha}_{1,0}(t_0), \ldots, \widehat{\alpha}_{m,0}(t_0)]^T$ of $\boldsymbol{\theta}(t)$ can be obtained by minimizing the linear locally weighted function

$$\sum_{k=1}^{m}\sum_{i=1}^{n}\left\{\widehat{X}'_{k,p}(t_i) + a_k\widehat{\mathbf{X}}_p(t_i) - \mathbf{z}_{i,1}^T\boldsymbol{\alpha}_k(t_0)\right\}^2 K_{b_k}(t_i - t_0), \tag{3.7}$$

where $\mathbf{z}_{i,1} = [1, t_i - t_0]^T$, $\boldsymbol{\alpha}_k(t_0) = [\alpha_{k,0}(t_0), \alpha_{k,1}(t_0)]^T$, and K_{b_k} is a kernel function with b_k as bandwidth for the kth component $\theta_k(t)$.

In order to obtain explicit expressions for the two-step local linear estimator, the two-step local quadratic estimator, and the hybrid estimator proposed in the previous section, we further denote by $\mathbf{Z}_{p(i)}$ the matrix $\mathbf{Z}_p$ at $t_0 = t_i$, and by $W_{h(i)}$ the matrix W_h at $t_0 = t_i$. For $v = 0, 1$; $p = 1, 2$ and $k = 1, \ldots, m$, define

$$U_{k;v,p} = \begin{pmatrix} e^T_{v+1,p+1}\left(\mathbf{Z}^T_{p(1)}W_{h_{k;v,p}(1)}\mathbf{Z}_{p(1)}\right)^{-1}\mathbf{Z}^T_{p(1)}W_{h_{k;v,p}(1)} \\ \vdots \\ e^T_{v+1,p+1}\left(\mathbf{Z}^T_{p(n)}W_{h_{k;v,p}(n)}\mathbf{Z}_{p(n)}\right)^{-1}\mathbf{Z}^T_{p(n)}W_{h_{k;v,p}(n)} \end{pmatrix}.$$

Thus the two-step local linear and quadratic estimators can be expressed as $\widehat{\boldsymbol{\theta}}_1(t_0) = [\widehat{\theta}_{1,1}(t_0), \ldots, \widehat{\theta}_{m,1}(t_0)]^T$ and $\widehat{\boldsymbol{\theta}}_2(t_0) = [\widehat{\theta}_{1,2}(t_0), \ldots, \widehat{\theta}_{m,2}(t_0)]^T$ with

$$\widehat{\theta}_{k,1}(t_0) = e_{1,2}^T \left(\mathbf{Z}_1^T W_{b_k} \mathbf{Z}_1\right)^{-1} \mathbf{Z}_1^T W_{b_k} \Big(U_{k;1,1}\mathbf{Y}_k + \sum_{j=1}^m a_{k,j} U_{j;0,1}\mathbf{Y}_j\Big), \tag{3.8}$$

$$\widehat{\theta}_{k,2}(t_0) = e_{1,2}^T \left(\mathbf{Z}_1^T W_{b_k} \mathbf{Z}_1\right)^{-1} \mathbf{Z}_1^T W_{b_k} \Big(U_{k;1,2}\mathbf{Y}_k + \sum_{j=1}^m a_{k,j} U_{j;0,2}\mathbf{Y}_j\Big). \tag{3.9}$$

Similarly, the two-step local hybrid estimator of $\boldsymbol{\theta}(t_0)$ for (3.1)−(3.2) is given by $\widehat{\boldsymbol{\theta}}(t_0) = [\widehat{\theta}_1(t_0), \ldots, \widehat{\theta}_m(t_0)]^T$, with

$$\widehat{\theta}_{k}(t_0) = e_{1,2}^T \left(\mathbf{Z}_1^T W_{b_k} \mathbf{Z}_1\right)^{-1} \mathbf{Z}_1^T W_{b_k} \Big(U_{k;1,2}\mathbf{Y}_k + \sum_{j=1}^m a_{k,j} U_{j;0,1}\mathbf{Y}_j\Big). \tag{3.10}$$

Based on (3.8), (3.9) and (3.10), one sees that different bandwidths can be used for different components in the estimation procedure. We bandwidth selection in Section 3.3.

3.2. Asymptotic results

In this subsection, we present the asymptotic biases and variances of the proposed estimators for (3.1)–(3.2). We summarize the notation of these estimators and the corresponding bandwidths in Table 1.

For convenience, take $\mu_j = \int u^j K(u)du$, $\nu_j = \int u^j K(u)^2 du$, $j = 0, 1, 2, \cdots$. Let $D = (t_1, \ldots, t_n)^T$ denote the observed time point vector. We make the following regularity assumptions.

(1) The density function f has a continuous second derivative in some neighborhood of t_0, and $f(t_0) \neq 0$.

(2) The component functions $X_k^{(4)}(t)$, $\theta_k^{(2)}(t)$ and $\sigma_k^2(t)$, $k = 1, \ldots, m$, are continuous in some neighborhood of t_0.

(3) The kernel function $K(\cdot)$ is a symmetric density function with compact support.

(4) Let $h_{k;0,p}$ and $h_{k;1,p}$, $k = 1, \ldots, m$, be the bandwidths of the pth order local polynomial fit for $X_k(t_0)$ and its first derivate $X_k'(t_0)$, respectively. When $p - v$ is odd, $h_{k;v,p} \to 0$, and $nh_{k;v,p} \to \infty$; when $p - v$ is even, $h_{k;v,p} \to 0$ and $nh_{k;v,p}^3 \to \infty$.

(5) Let b_k, $k = 1, \ldots, m$, be the bandwidth of the local linear fit for the kth component $\theta_k(t)$. There is an $s > 2$ and some $\varepsilon < 2 - s^{-1}$ such that $n^{2\varepsilon-1}h_{k;v,p} \to \infty$, for $k = 1, \ldots, m, v = 0, 1, p = 1, 2$, with $n^{2\varepsilon-1}b_k \to \infty$ and $nb_k^2/(\log n)^2 \to \infty$.

Table 1. Notation of the local estimators and the bandwidths

Notation	Definition
$\mathbf{X}(t_0)$	$\mathbf{X}(t_0) = [X_1(t_0), \ldots, X_m(t_0)]^T$ is a state vector
$\mathbf{X}'(t_0)$	$\mathbf{X}'(t_0) = [X_1'(t_0), \ldots, X_m'(t_0)]^T$ is a derivative vector of $\mathbf{X}(t_0)$
$\widehat{\mathbf{X}}_1(t_0)$	$\widehat{\mathbf{X}}_1(t_0) = [\widehat{X}_{1,1}(t_0), \ldots, \widehat{X}_{m,1}(t_0)]^T$ is a local linear estimator of $\mathbf{X}(t_0)$
$\widehat{\mathbf{X}}_2(t_0)$	$\widehat{\mathbf{X}}_2(t_0) = [\widehat{X}_{1,2}(t_0), \ldots, \widehat{X}_{m,2}(t_0)]^T$ is a local quadratic estimator of $\mathbf{X}(t_0)$
$\widehat{\mathbf{X}}_1'(t_0)$	$\widehat{\mathbf{X}}_1'(t_0) = [\widehat{X}_{1,1}'(t_0), \ldots, \widehat{X}_{m,1}'(t_0)]^T$ is a local linear estimator of $\mathbf{X}'(t_0)$
$\widehat{\mathbf{X}}_2'(t_0)$	$\widehat{\mathbf{X}}_2'(t_0) = [\widehat{X}_{1,2}'(t_0), \ldots, \widehat{X}_{m,2}'(t_0)]^T$ is a local quadratic estimator of $\mathbf{X}'(t_0)$
$\boldsymbol{\theta}(t_0)$	$\boldsymbol{\theta}(t_0) = [\theta_1(t_0), \ldots, \theta_m(t_0)]^T$ is a time-varying parameter vector
$\tilde{\boldsymbol{\theta}}_1(t_0)$	$\tilde{\boldsymbol{\theta}}_1(t_0) = [\tilde{\theta}_{1,1}(t_0), \ldots, \tilde{\theta}_{m,1}(t_0)]^T$ is one-step local linear estimator of $\boldsymbol{\theta}(t_0)$
$\tilde{\boldsymbol{\theta}}_2(t_0)$	$\tilde{\boldsymbol{\theta}}_2(t_0) = [\tilde{\theta}_{1,2}(t_0), \ldots, \tilde{\theta}_{m,2}(t_0)]^T$ is one-step local quadratic estimator of $\boldsymbol{\theta}(t_0)$
$\tilde{\boldsymbol{\theta}}(t_0)$	$\tilde{\boldsymbol{\theta}}(t_0) = [\tilde{\theta}_1(t_0), \ldots, \tilde{\theta}_m(t_0)]^T$ is one-step local hybrid estimator of $\boldsymbol{\theta}(t_0)$
$\widehat{\boldsymbol{\theta}}_1(t_0)$	$\widehat{\boldsymbol{\theta}}_1(t_0) = [\widehat{\theta}_{1,1}(t_0), \ldots, \widehat{\theta}_{m,1}(t_0)]^T$ is two-step local linear estimator of $\boldsymbol{\theta}(t_0)$
$\widehat{\boldsymbol{\theta}}_2(t_0)$	$\widehat{\boldsymbol{\theta}}_2(t_0) = [\widehat{\theta}_{1,2}(t_0), \ldots, \widehat{\theta}_{m,2}(t_0)]^T$ is two-step local quadratic estimator of $\boldsymbol{\theta}(t_0)$
$\widehat{\boldsymbol{\theta}}(t_0)$	$\widehat{\boldsymbol{\theta}}(t_0) = [\widehat{\theta}_1(t_0), \ldots, \widehat{\theta}_m(t_0)]^T$ is two-step local hybrid estimator of $\boldsymbol{\theta}(t_0)$
$h_{k;0,1}$	Bandwidth of the local linear fit for the kth state variable $X_k(t_0)$, $k = 1, \ldots, m$
$h_{k;0,2}$	Bandwidth of the local quadratic fit for the kth state variable $X_k(t_0)$, $k=1, \ldots, m$
$h_{k;1,1}$	Bandwidth of the local linear fit for the kth derivative $X_k'(t_0)$, $k = 1, \ldots, m$
$h_{k;1,2}$	Bandwidth of the local quadratic fit for the kth derivative $X_k'(t_0)$, $k = 1, \ldots, m$
b_k	Bandwidth of the local linear fit for the kth component $\theta_k(t_0)$, $k = 1, \ldots, m$

Theorem 1. *Suppose Conditions* (1)–(4) *hold. Let* $h_{0,1} = \max_{j=1,\ldots,m} h_{j;0,1}$ *and* $h_{0,2} = \max_{j=1,\ldots,m} h_{j;0,2}$.

(a) *The asymptotic conditional bias and variance of the one-step local linear estimator* $\tilde{\theta}_{k,1}(t_0)$ *for the* k*th component in* $\boldsymbol{\theta}(t_0) = [\theta_1(t_0), \ldots, \theta_m(t_0)]^T$ *are*

$$\begin{aligned}\text{bias}(\tilde{\theta}_{k,1}(t_0)|D) &= \frac{1}{3!} h_{k;1,1}^2 \frac{\mu_4}{\mu_2} \left(3\frac{f'(t_0)}{f(t_0)} X_k^{(2)}(t_0) + X_k^{(3)}(t_0) \right) \\ &\quad + \frac{1}{2}\mu_2 \sum_{j=1}^m a_{k,j} h_{j;0,1}^2 X_j^{(2)}(t_0) + o_P(h_{0,1}^2 + h_{k;1,1}^2), \\ \text{Var}\,(\tilde{\theta}_{k,1}(t_0)|D) &= \frac{\sigma_k^2(t_0)}{f(t_0)} \left(\sum_{j=1}^m \frac{a_{k,j}^2 \nu_0}{n h_{j;0,1}} + \frac{\nu_2}{\mu_2^2 n h_{k;1,1}^3} \right) (1 + o_P(1)).\end{aligned}$$

(b) *The asymptotic conditional bias and variance of the one-step local quadratic estimator* $\tilde{\theta}_{k,2}(t_0)$ *for the* k*th component in* $\boldsymbol{\theta}(t_0) = [\theta_1(t_0), \ldots, \theta_m(t_0)]^T$ *are*

$$\begin{aligned}\text{bias}(\tilde{\theta}_{k,2}(t_0)|D) &= \frac{1}{4!} \frac{\mu_4^2 - \mu_2\mu_6}{\mu_4 - \mu_2^2} \sum_{j=1}^m a_{k,j} h_{j;0,2}^4 \left(4\frac{f'(t_0)}{f(t_0)} X_j^{(3)}(t_0) + X_j^{(4)}(t_0) \right) \\ &\quad + \frac{1}{3!} h_{k;1,2}^2 \frac{\mu_4}{\mu_2} X_k^{(3)}(t_0) + o_P(h_{0,2}^4 + h_{k;1,2}^2),\end{aligned}$$

$$\mathrm{Var}\,(\tilde{\theta}_{k,2}(t_0)|D) = \frac{\sigma_k^2(t_0)}{f(t_0)}\left(\sum_{j=1}^{m}\frac{a_{k,j}^2(\nu_0\mu_4^2 - 2\nu_2\mu_2\mu_4 + \mu_2^2\nu_4)}{(\mu_4-\mu_2^2)^2 nh_{j;0,2}} + \frac{\nu_2}{\mu_2^2 nh_{k;1,2}^3}\right)$$
$$\times\,(1+o_P(1))\,.$$

(c) *The asymptotic conditional bias and variance of the one-step local hybrid estimator $\tilde{\theta}_k(t_0)$ for the kth component in $\boldsymbol{\theta}(t_0) = [\theta_1(t_0), \ldots, \theta_m(t_0)]$ are*

$$\mathrm{bias}(\tilde{\theta}_k(t_0)|D) = \frac{1}{2}\sum_{j=1}^{m} a_{k,j}h_{j;0,1}^2\mu_2 X_j^{(2)}(t_0) + \frac{1}{3!}h_{k;1,2}^2\frac{\mu_4}{\mu_2}X_k^{(3)}(t_0)$$
$$+o_P(h_{0,1}^2 + h_{k;1,2}^2),$$
$$\mathrm{Var}\,(\tilde{\theta}_k(t_0)|D) = \frac{\sigma_k^2(t_0)}{f(t_0)}\left(\sum_{j=1}^{m}\frac{a_{k,j}^2\nu_0}{nh_{j;0,1}} + \frac{\nu_2}{\mu_2^2 nh_{k;1,2}^3}\right)(1+o_P(1))\,.$$

The proof of Theorem 1 follows from Theorem 3.1 in Fan and Gijbels (1996). It is then clear that the asymptotic biases and variances are different for the three one-step estimators, and that the asymptotic bias and variance of the one-step local hybrid estimator $\tilde{\theta}_k(t_0)$ are the simplest. Based on the arguments in Fan and Gijbels (1996), we can see that the one-step hybrid estimator takes advantage of the smaller variance in estimating $\mathbf{X}(t)$ using a local linear smoother compared to the one-step local quadratic estimator, and gains the benefit of a smaller bias in estimating $\mathbf{X}'(t)$ using a local quadratic smoother compared to the one-step local linear estimator.

Remark 1. The optimal bandwidths for the one-step estimators are

$$h_{k;v,p}^{opt} = \begin{cases} O_P(n^{-\frac{1}{2p+3}}) & \text{if } \ p-v \ \text{ is odd,} \\ O_P(n^{-\frac{1}{2p+5}}) & \text{if } \ p-v \ \text{ is even.} \end{cases} \tag{3.11}$$

When the optimal bandwidths ($k = 1, \ldots, m; v = 0, 1; p = 1, 2$) are used in Theorem 1, the asymptotic conditional biases of the three one-step estimators for $\theta_k(t_0)$, $k = 1, \ldots, m$, are, respectively,

$$bias(\tilde{\theta}_{k,1}(t_0)|D) = \frac{1}{3!}h_{k;1,1}^2\frac{\mu_4}{\mu_2}\left(3\frac{f'(t_0)}{f(t_0)}X_k^{(2)}(t_0) + X_k^{(3)}(t_0)\right) + o_P(h_{k;1,1}^2),$$
$$bias(\tilde{\theta}_{k,2}(t_0)|D) = \frac{1}{3!}h_{k;1,2}^2\frac{\mu_4}{\mu_2}X_k^{(3)}(t_0) + o_P(h_{k;1,2}^2),$$
$$bias(\tilde{\theta}_k(t_0)|D) = \frac{1}{3!}h_{k;1,2}^2\frac{\mu_4}{\mu_2}X_k^{(3)}(t_0) + o_P(h_{k;1,2}^2),$$

while the corresponding asymptotic conditional variances are each

$$\frac{\nu_2\sigma_k^2(t_0)}{\mu_2^2 f(t_0)nh_{k;1,2}^3}(1+o_P(1)),$$

where the optimal bandwidths are $h_{k;1,1} = h_{k;1,2} = O_P(n^{-1/7})$. Thus, the square-root of the conditional mean squared errors (MSEs) of the three one-step local estimators can only achieve the convergence rate $O_P(n^{-2/7})$, which is slower than the standard optimal convergence rate of $O_P(n^{-2/5})$.

Notice that, when the optimal bandwidth is used, the asymptotic biases and variances of the one-step local quadratic estimator and hybrid estimator are the same. This is because the asymptotic bias and variance are dominated by that of the estimate of $X_k'(t)$, $k = 1, \dots, m$, which is in the same order for both estimators. The asymptotic biases of both the one-step local quadratic estimator and the hybrid estimator are simpler under the optimal bandwidth case compared to that of the one-step local linear estimator.

The asymptotic conditional biases and variances for the three two-step local estimators are given next.

Theorem 2. *Suppose Conditions* (1)–(5) *hold. Let* $h_{0,1} = \max_{j=1,\dots,m} h_{j;0,1}$ *and* $h_{0,2} = \max_{j=1,\dots,m} h_{j;0,2}$.

(a) *The asymptotic conditional bias and variance of the two-step local linear estimator* $\widehat{\theta}_{k,1}(t_0)$ *for the kth component in* $\boldsymbol{\theta}(t_0) = [\theta_1(t_0), \dots, \theta_m(t_0)]^T$ *are*

$$\begin{aligned}
\mathrm{bias}(\widehat{\theta}_{k,1}(t_0)|D) &= \frac{1}{2} b_k^2 \mu_2 \theta_k^{(2)}(t_0) + \frac{1}{3!} h_{k;1,1}^2 \frac{\mu_4}{\mu_2} \left(3 \frac{f'(t_0)}{f(t_0)} X_k^{(2)}(t_0) + X_k^{(3)}(t_0) \right) \\
&\quad + \frac{1}{2} \sum_{j=1}^{m} a_{k,j} h_{j;0,1}^2 \mu_2 X_j^{(2)}(t_0) + o_P(b_k^2 + h_{0,1}^2 + h_{k;1,1}^2) \\
\mathrm{Var}\,(\widehat{\theta}_{k,1}(t_0)|D) &= \frac{\nu_0 \sigma_k^2(t_0)}{f(t_0) n b_k} \left[\sum_{j=1}^{m} \left(\frac{a_{k,j}^2 (n-1)}{n} + \frac{a_{k,j}^2 \nu_0}{f(t_0) n h_{j;0,1}} \right) + \frac{\nu_2}{\mu_2^2 f(t_0) n h_{k;1,1}^3} \right] \\
&\quad \times (1 + o_p(1)).
\end{aligned}$$

(b) *The asymptotic conditional bias and variance of the two-step local quadratic estimator* $\widehat{\theta}_{k,2}(t_0)$ *for the kth component in* $\boldsymbol{\theta}(t_0) = [\theta_1(t_0), \dots, \theta_m]^T$ *are*

$$\begin{aligned}
\mathrm{bias}(\widehat{\theta}_{k,2}(t_0)|D) &= \frac{1}{2} b_k^2 \mu_2 \theta_k^{(2)}(t_0) + \frac{1}{3!} h_{k;1,2}^2 \frac{\mu_4}{\mu_2} X_k^{(3)}(t_0) + o_P(b_k^2 + h_{0,2}^4 + h_{k;1,2}^2) \\
&\quad + \frac{1}{4!} \frac{\mu_4^2 - \mu_2 \mu_6}{\mu_4 - \mu_2^2} \sum_{j=1}^{m} a_{k,j} h_{j;0,2}^4 \left(4 \frac{f'(t_0)}{f(t_0)} X_j^{(3)}(t_0) + X_j^{(4)}(t_0) \right), \\
\mathrm{Var}\,(\widehat{\theta}_{k,2}(t_0)|D) &= \frac{\nu_0 \sigma_k^2(t_0)}{f(t_0) n b_k} \left[\sum_{j=1}^{m} \left(\frac{a_{k,j}^2 (n-1)}{n} + \frac{a_{k,j}^2 (\nu_0 \mu_4^2 - 2\nu_2 \mu_2 \mu_4 + \mu_2^2 \nu_4)}{(\mu_4 - \mu_2^2)^2 f(t_0) n h_{j;0,2}} \right) \right. \\
&\quad \left. + \frac{\nu_2}{\mu_2^2 f(t_0) n h_{k;1,2}^3} \right] (1 + o_p(1)).
\end{aligned}$$

(c) *The asymptotic conditional bias and variance of the two-step local hybrid estimator $\widehat{\theta}_k(t_0)$ for the kth component in $\boldsymbol{\theta}(t_0) = [\theta_1(t_0), \ldots, \theta_m(t_0)]^T$ are*

$$\begin{aligned}
\mathrm{bias}(\widehat{\theta}_k(t_0)|D) &= \frac{1}{2} b_k^2 \mu_2 \theta_k^{(2)}(t) + \frac{1}{2} \sum_{j=1}^m a_{k,j} h_{j;0,1}^2 \mu_2 X_j^{(2)}(t_0) + \frac{1}{3!} h_{k;1,2}^2 \frac{\mu_4}{\mu_2} X_k^{(3)}(t_0) \\
&\quad + o_P(b_k^2 + h_{0,1}^2 + h_{k;1,2}^2), \\
\mathrm{Var}\,(\widehat{\theta}_k(t_0)|D) &= \frac{\nu_0 \sigma_k^2(t_0)}{f(t_0) n b_k} \Big[\sum_{j=1}^m \Big(\frac{a_{k,j}^2 (n-1)}{n} + \frac{a_{k,j}^2 \nu_0}{f(t_0) n h_{j;0,1}} \Big) + \frac{\nu_2}{\mu_2^2 f(t_0) n h_{k;1,2}^3} \Big] \\
&\quad \times (1 + o_p(1)) \, .
\end{aligned}$$

Proof. see Appendix (online supplement).

The results in Theorems 2 show that the asymptotic bias of the two-step local hybrid estimator $\widehat{\theta}_k(t_0)$ has the simplest structure. It is interesting to notice that the asymptotic conditional variances of the three two-step estimators for the kth component $\theta_k(t_0)$, $k = 1, \ldots, m$, are asymptotically same, i.e. $[(\nu_0 \sigma_k^2(t_0))/(f(t_0) n b_k)] \sum_{j=1}^m a_{k,j}^2 \,(1 + o_p(1))$, and the asymptotic conditional variances are independent of the initial bandwidths $h_{k;v,p}$. Thus we can choose the initial bandwidths as small as possible as long as they satisfies the constraints in Conditions (4) and (5). In particular, when the initial bandwidths $h_{k;v,p} = O_P(b_k) \equiv h_k$ for $k = 1, \ldots, m, v = 0, 1, p = 1, 2$, the square-root of the conditional MSEs of the two-step estimators achieve the optimal convergence rate of $O_P(n^{-2/5})$.

Remark 2. When the initial bandwidths $h_{k;v,p} = o_P(n^{-1/5}) \equiv h_k$ are used in Theorem 2, then the asymptotic conditional biases of the three two-step estimators of $\theta_k(t_0)$, $k = 1, \ldots, m$, are each $(1/2) b_k^2 \mu_2 \theta_k^{(2)}(t_0) + o_P(b_k^2)$. Furthermore, the optimal bandwidth for the second step is $b_k = O_P(n^{-1/5})$ for $k = 1, \ldots, m$.

In summary, the asymptotic properties of the proposed two-step local estimators are more appealing than those of the simple one-step estimators, with the latter only achieving the convergence rate of the order $O_P(n^{-2/7})$. We compare the performance of these methods via finite sample simulations in Section 4.

3.3. Bandwidth selection

The selection of the smoothing parameters is important in all nonparametric model fitting. Here we need to determine the bandwidths $h_{k;v,p}$, $k = 1, \ldots, m$, for the one-step estimators in (3.3)−(3.4), and the bandwidths h_k and b_k, $k = 1, \ldots, m$, for the two-step estimators in (3.8)−(3.10).

For the one-step local estimators, we use the pre-asymptotic substitution method in Fan and Gijbels (1996) to obtain the estimated optimal bandwidths

$\widehat{h}^{opt}_{k;v,p} = \widehat{h}((Y_k(t_i), t_i), i = 1, \ldots, n)$, $k = 1, \ldots, m$, to estimate $X_k(t)$ and $X'_k(t)$, respectively. Fan and Huang (1997) have shown that the estimated optimal bandwidth by the pre-asymptotic substitution method is a consistent estimator of the asymptotically optimal bandwidth, and is of order $O_p(n^{-1/5})$ when the local linear fitting is used.

For the two-step local estimators, we have shown that, if the initial bandwidth $h_{k;0}$ in Step 1 is selected between the rates $O_P(n^{-1/3})$ and $O_P(n^{-1/5})$, we can achieve the optimal convergence rate. The choice of the initial bandwidth is not very sensitive as long as it is small enough to satisfy the optimal rate conditions. Based on our experience, we suggest selecting the initial bandwidth $h_k = d \times \widehat{h}^{opt}_{k;0,1}$ where $0.5 < d < 0.9$ and $\widehat{h}^{opt}_{k,0,1}$ is the optimal bandwidth for local linear estimation of $X_k(t)$, $k = 1, \ldots, m$. We summarize the bandwidth selection procedure for the two-step estimators as follows.

1. Obtain the optimal bandwidth for the local linear estimation of $X_k(t)$, $\widehat{h}^{opt}_{k;0,1}$ for $k = 1, \ldots, m$.
2. Select the initial bandwidths $h_k = d \times \widehat{h}^{opt}_{k;0,1}$, $k = 1, \ldots, m$, where $0.5 < d < 0.9$ for the first step smoothing, and obtain the estimates $\widehat{\mathbf{X}}_p(t_i) = [\widehat{X}_{1,p}(t_i), \ldots, \widehat{X}_{m,p}(t_i)]^T$ and $\widehat{\mathbf{X}}'_p(t_i) = [\widehat{X}'_{1,p}(t_i), \ldots, \widehat{X}'_{m,p}(t_i)]^T$, $p = 1, 2$, for $i = 1, \ldots, n$.
3. For the two-step local linear and local quadratic estimators, compute a rough estimate of $\boldsymbol{\theta}(t)$ by $\boldsymbol{\theta}^*(t_i) = \widehat{\mathbf{X}}'_p(t_i) + a\widehat{\mathbf{X}}_p(t_i)$, $i = 1, \ldots, n$, $(p = 1, 2)$. Then use the pre-asymptotic substitution method to find the second bandwidths $\widehat{b}^{opt}_k = \widehat{h}((\theta^*_k(t_i), t_i), i = 1, \ldots, n)$ for $k = 1, \ldots, m$. The estimated optimal bandwidths $\widehat{b}^{opt}_k$, $k = 1, \ldots, k$, can be used in (3.8)$-$(3.9) to obtain $\widehat{\boldsymbol{\theta}}_1(t_0)$ and $\widehat{\boldsymbol{\theta}}_2(t_0)$, respectively.
4. For the two-step local hybrid estimator, compute the rough estimate of $\boldsymbol{\theta}(t)$ by $\boldsymbol{\theta}^*(t_i) = \widehat{\mathbf{X}}'_2(t_i) + a\widehat{\mathbf{X}}_1(t_i)$, $i = 1, \ldots, n$. Use the pre-asymptotic substitution method to determine the estimated optimal bandwidth for the second step, $\widehat{b}^{opt}_k = \widehat{h}((\theta^*_k(t_i), t_i), i = 1, \ldots, n)$. These bandwidths can be applied to (3.10) to obtain $\widehat{\boldsymbol{\theta}}(t_0)$.

3.4. An approximate confidence interval

Without loss of generality, we consider an approximate confidence interval estimate based on $\widehat{\theta}_{k,2}(t_0)$, $k = 1, \ldots, m$. Since $f(t_0)$, $\sigma^2(t_0)$ and $\text{bias}(\widehat{\theta}_{k,2}(t_0)|D)$ are unknown, Theorem 2 cannot be directly used to construct the confidence interval for $\theta_k(t_0)$, so we bring in estimates of the conditional bias and variance. Following the generalized pre-asymptotic method of Fan and Gijbels (1996), the conditional bias of $\widehat{\theta}_{k,2}(t_0)$ can be approximated by

$$\widehat{\mathbf{B}}_{k,2}(t_0) = e^T_{1,2} \left(\mathbf{Z}^T_1 W_{b_k} \mathbf{Z}_1\right)^{-1} \mathbf{Z}^T_1 W_{b_k} [\widehat{r}_{k,1}(t_0), \ldots, \widehat{r}_{k,n}(t_0)]^T, \qquad (3.12)$$

where $\widehat{r}_{k,i}(t_0) = \widehat{\alpha}^*_{k,2}(t_0)(t_i - t_0)^2 + \widehat{\alpha}^*_{k,3}(t_0)(t_i - t_0)^3$. The estimators $\widehat{\alpha}^*_{k,2}(t_0)$ and $\widehat{\alpha}^*_{k,3}(t_0)$ can be obtained by minimizing the locally weighted function

$$\sum_{i=1}^{n} \left\{ \widehat{X}'_{k,p}(t_i) + a_k \widehat{\mathbf{X}}_p(t_i) - \sum_{l=0}^{3} \alpha_{k,l}(t_0)(t_i - t_0)^l \right\}^2 K_{b_k}(t_i - t_0).$$

The conditional variance of $\widehat{\theta}_{k,2}(t_0)$ is approximated by

$$\widehat{\mathbf{V}}_{k,2}(t_0) = e_{1,2}^T \big(\mathbf{Z}_1^T W_{b_k} \mathbf{Z}_1\big)^{-1} \mathbf{Z}_1^T W_{b_k} \Big[(U_{k,1,2} + a_{k,k} U_{k,0,2})(U_{k,1,2} + a_{k,k} U_{k,0,2})^T \sigma_k^2(t_0) + \sum_{j=1; j\neq k}^{m} a_{k,j}^2 U_{j,0,2} U_{j,0,2}^T \sigma_j^2(t_0) \Big] W_{b_k} \mathbf{Z}_1 \left(\mathbf{Z}_1^T W_{b_k} \mathbf{Z}_1\right)^{-1} e_{1,2}, \qquad (3.13)$$

where $\sigma_j^2(t_0)$, $j = 1, \ldots, m$, can be estimated by

$$\widehat{\sigma}_j^2(t_0) = \frac{\sum_{i=1}^n (Y_j(t_i) - \widehat{Y}_j(t_i))^2 K_h(t_i - t_0)}{\sum_{i=1}^n K_h(t_i - t_0)}.$$

With the estimated bias and variance in (3.12) and (3.13), an approximate $(1 - \alpha)100\%$ confidence interval of the two-step local quadratic estimator for $\theta_k(t_0)$, $k = 1, \ldots, m$, is

$$\widehat{\theta}_{k,2}(t_0) - \widehat{\mathbf{B}}_{k,2}(t_0) \pm z_{1-\frac{\alpha}{2}} \{\widehat{\mathbf{V}}_{k,2}(t_0)\}^{\frac{1}{2}},$$

where $z_{1-\alpha/2}$ denotes $(1 - \alpha/2)$ quantile of the standard Gaussian distribution. The approximate confidence intervals of the two-step local linear estimator $\widehat{\theta}_{k,1}(t_0)$ and the two-step local hybrid estimator $\widehat{\theta}_k(t_0)$ can be constructed in a similar fashion.

4. Numerical Examples

4.1. Simulation studies

Monte Carlo simulation studies were designed to evaluate the finite properties of the proposed estimators. We generated the simulation data from the HIV dynamic model at (1.1),

$$\mathbf{X}'(t_i) = \boldsymbol{\theta}(t_i) - \mathbf{a}\mathbf{X}(t_i), \qquad i = 1, \ldots, n, \qquad (4.1)$$

where $\mathbf{X}(t_i) = [X_1(t_i), X_2(t_i), X_3(t_i)]^T$, $\mathbf{a} = \mathrm{diag}(\rho, \delta, c)$, and $\boldsymbol{\theta}(t_i) = [\theta_1(t_i), \theta_2(t_i), \theta_3(t_i)]^T$ with $\theta_1(t_i) = \lambda - k[1 - r(t_i)]X_1(t_i)X_3(t_i)$, $\theta_2(t_i) = k[1 - r(t_i)]X_1(t_i) X_3(t_i)$ and $\theta_3(t_i) = N\delta X_2(t_i)$. The measurement models for $X_k(t_i)$, $k = 1, 2, 3$, are

$$Y_k(t_i) = X_k(t_i) + e_k(t_i). \qquad (4.2)$$

First we generated data, $\{(X_1(t_i), X_2(t_i), X_3(t_i))\}$, at time points $t_i = \text{day} \times t_i^*$ with $t_i^* = i/(n+1)$, $i = 1, \ldots, n$, by solving (4.1) with the following parameter values and initial values: $\lambda = 36.0$, $\rho = 0.108$, $k = 9 \times 10^{-5}$, $\delta = 0.5$, $N = 1,000.0$, $c = 3.0$, $X_1(0) = 600$, $X_2(0) = 33$, $X_3(0) = 100,000$, and $r(t) = 0.9\cos(\pi t/10000)$. The observed data $\{(Y_1(t_i), Y_2(t_i), Y_3(t_i)), i = 1, \ldots, n\}$ were generated based on the model (4.2) with the error term $e_k(t_i)$, $k = 1, 2, 3$, following an iid normal distribution with mean 0 and variance $\sigma_k^2(t_i) = (1+t_i)^{1/2}\sigma_k^2$, where $[\sigma_1, \sigma_2, \sigma_3] = [20, 5, 100]$ or $[40, 10, 200]$. Our objective was to estimate the time-varying parameter $\boldsymbol{\theta}(t) = [\theta_1(t), \theta_2(t), \theta_3(t)]^T$ from the observed data $\{(Y_1(t_i), Y_2(t_i), Y_3(t_i)), i = 1, \ldots, n\}$, where $n = 100$ and 200.

We applied the one-step and the two-step estimation methods proposed in the previous sections to the simulated data. To evaluate these estimators, we employed the Epanechnikov kernel $K(t) = 0.75(1-t^2)_+$ for all the estimators. The performance of these estimators was assessed using the Square-Root of Average Squared Errors (RASE) defined as

$$RASE(\widehat{\theta}_k(t_j^*)) = \left(n_{grid}^{-1} \sum_{j=1}^{n_{grid}} \{\widehat{\theta}_k(t_j^*) - \theta_k(t_j^*)\}^2 \right)^{\frac{1}{2}}, \tag{4.3}$$

where $\{t_j^*, j = 1, \ldots, n_{grid}\}$ were the grid points at which the time-varying parameters $\theta_k(\cdot)$, $k = 1, 2, 3$, were estimated. The bandwidth selection strategy discussed in Section 3.3 was used to determine the optimal bandwidths $\widehat{h}_{k,v,p}^{opt}$ and $\widehat{b}_k^{opt}$, $k = 1, 2, 3, v = 0, 1, p = 1, 2$.

For the one-step estimators, only the hybrid estimator was used since it is the best among the three one-step estimators. We compared the three two-step estimators to the one-step hybrid estimator based on 500 simulations. Boxplots of the RASEs for the four estimators for $\theta_1(t)$, $\theta_2(t)$ and $\theta_3(t)$ are given in Figure 1 (a)$-$(c), respectively. There we can see that the three two-step estimators outperform the one-step hybrid estimator, and the two-step hybrid estimator is slightly better than the two-step local linear and local quadratic estimators. The initial bandwidths in the two-step local estimators were $h_k = 0.7 \times \widehat{h}_{k;0,1}^{opt}$, $k = 1, 2, 3$, with the estimated optimal bandwidths $\widehat{h}_{k;0,1}^{opt}$ obtained using the pre-asymptotic substitution method.

The estimates of $[X_k(t), X_k'(t), \theta_k(t)]$, $k = 1, 2, 3$, using the two-step hybrid method are presented in Figures 2. We can see that these estimates are very close to the true functions. Other estimators also gave reasonable estimation results (data not shown). We conducted simulations for larger measurement errors and different sample sizes, with similar conclusions reached.

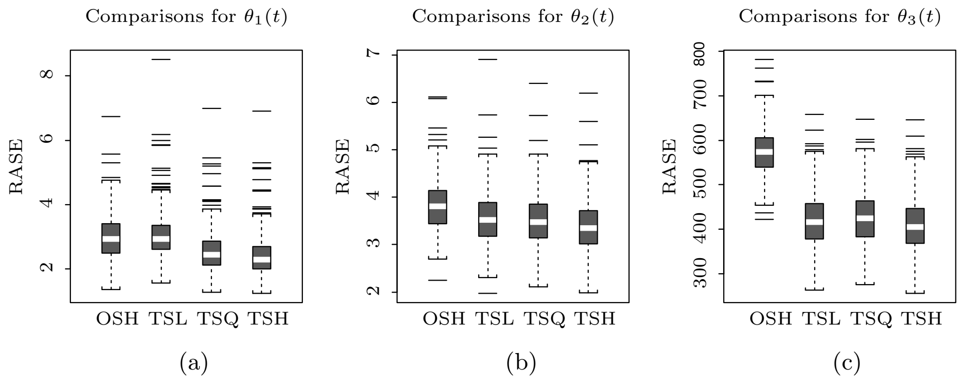

Figure 1. Simulation comparisons of RASE for different estimation methods with the sample size $n = 100$ and the variances $(\sigma_1, \sigma_2, \sigma_3) = (20, 5, 100)$: (a)$-$(c) Boxplots of the RASEs for the one-step local hybrid (OSH) estimates, the two-step local linear (TSL) estimates, the two-step local quadratic (TSQ) estimates and the two-step local hybrid (TSH) estimates for $\theta_1(t)$, $\theta_2(t)$ and $\theta_3(t)$, respectively.

4.2. Application to AIDS clinical trial data

We applied the proposed methods to a viral load data set from an AIDS clinical trial to further illustrate the usefulness of the proposed estimation methods. The AIDS clinical trial was developed by the Adult AIDS Clinical Trials Group (AACTG), and the viral load (HIV RNA copies in plasma) was monitored at weeks 1, 3, 4, 5, 6, 8, 9, 12, 14, 16, 21, 25, 29 and 33 in HIV-1 infected patients after receiving highly active antiretroviral therapy (HAART). Viral load data from two patients are plotted in Figure 3. More frequent viral load data within the first three days are also available from these patients and could be used to estimate the constant viral dynamic parameters in the model (1.3)$-$(1.4). For methodological details, see papers by Perelson et al. (1996), Han and Chaloner (2004) and Wu (2005).

We use the proposed nonparametric local estimation methods to fit the HIV dynamic model (4.1) to the viral load data from individual AIDS patients in this clinical study. In particular, we smoothed the viral load $X_3(t)$, estimated the viral load change profile $X'(t)$, and the production rate of free HIV virions $\theta_3(t)$ for individual patients. The two-step local hybrid estimator was used to estimate the production rate $\theta_3(t)$. The corresponding estimated bandwidths (h_3, b_3) of $\theta_3(t)$ for patients 1 and 2 were $(58.19, 70.53)$ and $(95.42, 40.55)$, respectively. The model fitting results for the two patients are plotted in Figure 3. From this figure, we can see that the fitted data compare well to the observed viral load data. The

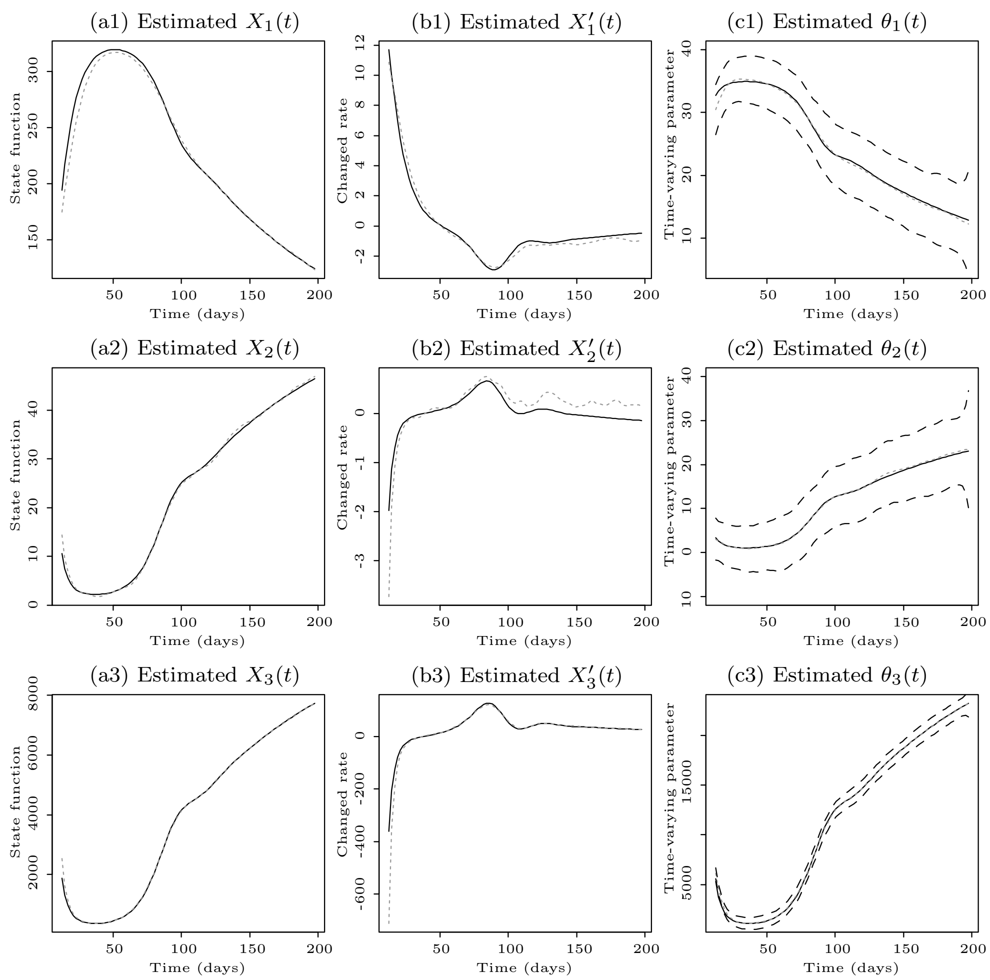

Figure 2. Estimation results from the two-step hybrid estimation method with the sample size $n = 100$ and the variances $(\sigma_1, \sigma_2, \sigma_3) = (20, 5, 100)$ in the simulation study: solid curves – true functions and dashed curves – estimated functions. (a1-a3) pointwise averages of local linear estimates for the state functions $X_1(t)$, $X_2(t)$ and $X_3(t)$ from 500 simulations,respectively; (b1-b3) pointwise averages of local quadratic estimates for the derivatives $X_1'(t)$, $X_2'(t)$ and $X_3'(t)$ from 500 simulations,respectively; (c1-c3) pointwise averages of the two-step hybrid estimates and their 95% confidential intervals for $\theta_1(t)$, $\theta_2(t)$ and $\theta_3(t)$ from 500 simulations, respectively.

estimates of the derivative of the viral load (viral load change) and the time-varying parameters (viral production rate) are reasonably estimated (Figure 3). These estimation results may provide important information and help clinicians make treatment decisions for individual AIDS patients.

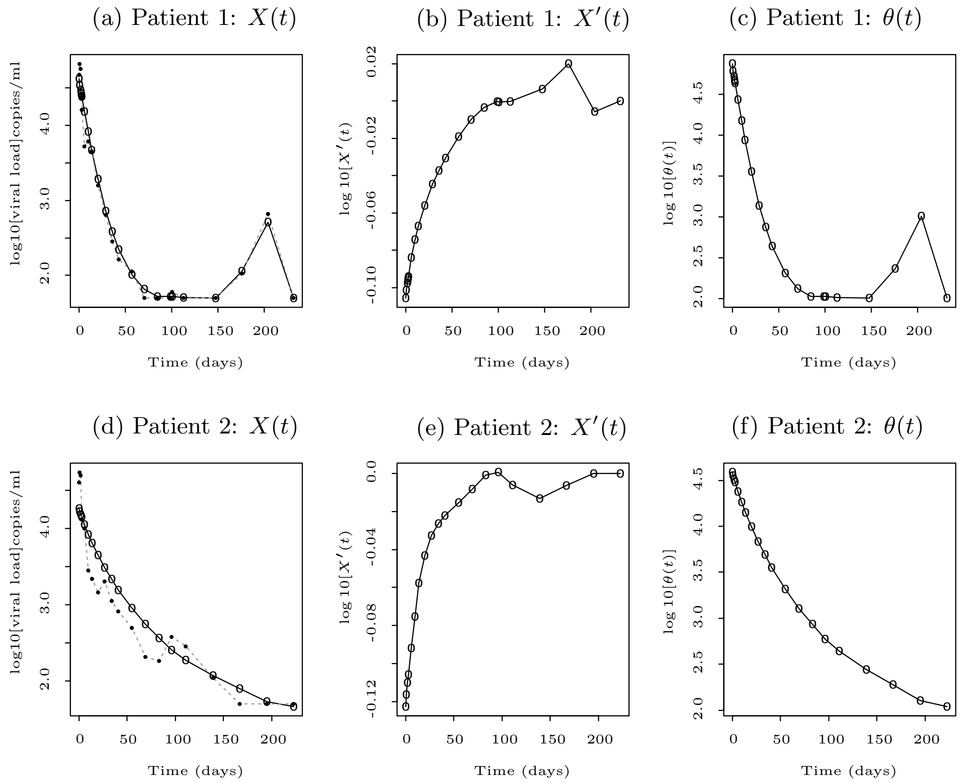

Figure 3. Estimation results for two patients from an AIDS clinical study. The bandwidths for patient 1 are $(h_3, b_3) = (58.19, 70.53)$ and for patient 2 are $(h_3, b_3) = (95.42, 40.55)$. (a) and (d): The estimated viral load over the treatment days. The solid line indicates the local linear estimator of the viral load $X_3(t)$ and the dashed line is the observed viral load profile for the two patients, respectively. (b) and (e): The estimated viral load change profile over the treatment days. The solid line indicates the local quadratic estimator of the derivative $X_3'(t)$ for the two patients, respectively. (c) and (f): The estimated virus production rate of free HIV virions over the treatment days. The solid line indicates the two-step local hybrid estimator of the time-varying virus production rate $\theta_3(t)$ for the two patients, respectively.

5. Concluding Remarks

Although we have mainly focused on a linear deterministic dynamic model with a single time-varying parameter in this paper, these techniques can be extended to multivariate models with multiple unknown time-varying parameters. This is not a trivial generalization; in the case of multiple time-varying parameters, the second step of our proposed methods may need to resort to time-varying coefficient models or generalized time-varying coefficient models that have been actively studied by many authors in the past years (e.g., Hoover et al. (1998), Fan and Zhang (1999), Cai, Fan and Li (2002), Fan and Zhang (2003) and Wu and Zhang (2006)). In the development of theoretical results, we have focused on linear dynamic models although our methodologies can apply to more general dynamic models. In particular it is more difficult to derive asymptotic theories for nonlinear dynamic models, and this is a good future research topic. Also note that we only considered the estimation of the time-varying parameters in the dynamic models. We have made an assumption that all other parameters in the model are known or can be estimated from other sources. Thus, we do not have the identifiability problem in our numerical examples. However, in practice, it is very important to study the identifiability of the dynamic models before parameter estimation. This is beyond the scope of this paper; we refer readers to Xia (2003), Xia and Moog (2003) and Jeffrey and Xia (2005).

Acknowledgement

The authors would like to thank the Editors, an Associate Editor and two referees for their insightful and constructive comments and suggestions. This work was partially supported by the National Institutes of Health (NIH) research grants RO1 AI052765, RO1 AI055290 and NO1 AI050020.

References

Cai, Z., Fan, J. and Li, R. (2002). Efficient estimation and inference for varying-coefficient models. *J. Amer. Statist. Assoc.* **95**, 888-902.

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

Fan, J. and Huang, L. S. (1997). Rates of convergence for the pre-asymptotic substitution bandwidth selector. *Statist. Probab. Lett.* **43**, 309-316.

Fan, J. and Zhang, W. (1999). Statistical estimation in varying-coefficient models. *Ann. Statist.* **27**, 1491-1518.

Fan, J. and Zhang, W. (2003). A re-examination of Stanton's diffusion estimations with applications to finncial model validation. *J. Amer. Statist. Assoc.* **98**, 118-134.

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.

Hoover, D. R., Rice, J. A., Wu, C.O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809-822.

Huang, Y., Rosenkranz, S., Wu, H. (2003). Modeling HIV dynamics and antiviral responses with consideration of time-varying drug exposures, sensitivities and adherence. *Math. Biosci.* **184**, 165-186.

Huang, H., Liu, D. and Wu, H. (2006). Hierarchical Bayesian Methods for Estimation of Parameters in a Longitudinal HIV Dynamic System. *Biometrics* **62**, 413-423.

Jeffrey, A. M. and Xia, X. (2005). Identifiability of HIV/AIDS model. In Tan, W.Y., Wu, H.(Ed.), *Deterministic And Stochastic Models Of AIDS Epidemics and HIV Infections With Intervention*. World Scientific.

Michele, D. M., Ruy, M. R., Martin, M., Ho, D. D. and Perelson, A. S. (2004). Modeling the long-term control of vivrmia in HIV-1 infected patients treated with antiretroviral therapy. *Math. Biosci.* **188**, 47-62.

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

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

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.

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

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

Wu, H., Ding, A. A. and De Gruttola, V. D. (1998). Estimation of HIV dynamic parameters. *Statist. Medicine* **17**, 2463-2485.

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. (2005). Statistical methods for HIV dynamic studies in AIDS Clinical Trials. *Statist. Meth. Medical Res.* **14**, 171-192.

Wu, H. and Zhang, J. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches*. John Wiley, New York.

Xia, X. (2003). Estimation of HIV/AIDS parameters. *Automatica* **39** 1983-1988.

Xia and Moog(2003). Identifiability of nonlinear systems with applications to HIV/AIDS models. *IEEE Trans. Autom. Control* **48**, 330-336.

Department of Mathematics and Statistics, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, U.S.A.

E-mail: jchen@sciences.sdsu.edu

Department of Biostatistics and Computational Biology, University of Rochester, 601 Elmwood Avenue, Box 630, Rochester, New York 14642, U.S.A.

E-mail: hwu@bst.rochester.edu

(Received August 2006; accepted February 2007)