Published in final edited form as:
J Am Stat Assoc. 2008 December 1; 103(484): 1570–1583. doi:10.1198/016214508000000797.

Parameter Estimation for Differential Equation Models Using a Framework of Measurement Error in Regression Models

Hua Liang and **Hulin Wu**

Hua Liang (E-mail: hliang@bst.rochester.edu) is Associate Professor and Hulin Wu (hwu@bst.rochester.edu) is Professor, Department of Biostatistics and Computational Biology, University of Rochester School of Medicine and Dentistry, 601 Elmwood Avenue, Box 630, Rochester, New York 14642

Abstract

Differential equation (DE) models are widely used in many scientific fields that include engineering, physics and biomedical sciences. The so-called “forward problem”, the problem of simulations and predictions of state variables for given parameter values in the DE models, has been extensively studied by mathematicians, physicists, engineers and other scientists. However, the “inverse problem”, the problem of parameter estimation based on the measurements of output variables, has not been well explored using modern statistical methods, although some least squares-based approaches have been proposed and studied. In this paper, we propose parameter estimation methods for ordinary differential equation models (ODE) based on the local smoothing approach and a pseudo-least squares (PsLS) principle under a framework of measurement error in regression models. The asymptotic properties of the proposed PsLS estimator are established. We also compare the PsLS method to the corresponding SIMEX method and evaluate their finite sample performances via simulation studies. We illustrate the proposed approach using an application example from an HIV dynamic study.

Keywords

AIDS; HIV viral dynamics; ordinary differential equations (ODE); local polynomial smoothing; measurement errors models; nonparametric regression; principal differential analysis; regression calibration; SIMEX

1. Introduction

Differential equations are widely used to describe dynamic systems in many scientific fields including physics, engineering, economics, and biomedical sciences. The studies of differential equations have mainly focused on the so-called forward problem, i.e., simulation and analysis of the behavior of state variables for a given system. However, the inverse problem, using the measurements of state variables to estimate the parameters that characterize the system, has not been well studied particularly from statistical perspectives. Statistical methods for estimating parameters in differential equation models are very sparse in the statistical literature. In this paper, we intend to propose new statistical estimation methods for a general ordinary differential equation (ODE) model that can be written as:

Correspondence to: Hulin Wu.

$$\frac{d\mathbf{X}(t)}{dt}=F\{\mathbf{X}(t),\beta\}, \tag{1.1}$$

where $\mathbf{X}(t) = \{ X_1(t), \ldots, X_k(t)\}^{\mathrm{T}}$ is an unobserved state vector, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^{\mathrm{T}}$ is a vector of unknown parameters, and $\mathbf{F}(\cdot) = \{F_1(\cdot), \ldots, F_k(\cdot)\}^{\mathrm{T}}$ is a known linear or nonlinear function vector. In practice, we may not observe $\mathbf{X}(t)$ directly, but we can observe its surrogate $\mathbf{Y}(t)$. For simplicity, here we assume an additive measurement error model to relate $\mathbf{X}(t)$ to the surrogate $\mathbf{Y}(t)$, i.e.,

$$\boldsymbol{Y}(t)=\mathbf{X}(t)+e(t), \tag{1.2}$$

where the measurement error $e(t)$ is independent of $X(t)$ with a covariance matrix Σ_e.

Parameter estimation for ODE models has been investigated using the least squares principle by mathematicians (Hemker, 1972; Bard, 1974; Li, Osborne and Prvan, 2005), computer scientists (Varah, 1982), and chemical engineers (Ogunnaike and Ray, 1994; Poyton et al., 2006). Mathematicians have focused on the development of efficient and stable algorithms to solve the least squares problem. Recently statisticians have started to develop various statistical methods to estimate dynamic parameters in ODE models. For example, Putter et al. (2002), Huang and Wu (2006), and Huang, Liu and Wu (2006) have developed hierarchical Bayesian approaches to estimate dynamic parameters in HIV dynamic models for longitudinal data. Li et al. (2002) proposed a spline-based approach to estimate time-varying parameters in ODE models. Ramsay (1996) proposed a technique named principal differential analysis (PDA) for estimation of differential equation models (see a comprehensive survey in Ramsay and Silverman, 2005). The basic idea of PDA is to fit the discrete measurements of the output variables $\mathbf{Y}(t)$ using a spline approach, and to obtain the estimated derivative curves. These estimated values are then substituted into the ODEs, and the estimated differential equation parameters can be obtained by a simple least squares procedure. Ramsay et al. (2007) applied a penalized spline method to estimate the constant dynamic parameters in ODE models. Chen and Wu (2008a, 2008b) proposed a two-step approach to estimate time-varying parameters in ODE models. Miao et al (2008a) explored the identifiability, global optimization techniques, model selection, and multi-model inference under the framework of the nonlinear least squares approach for ODE models. Overall the statistical literature for ODE models is generally sparse. Many statistical inference issues for ODE models have not been well addressed. In addition, there are some drawbacks with the existing estimation methods. First, the standard nonlinear least squares (NLS) method needs to minimize the error sum of squares which requires numerically solving the ODEs repeatedly. The initial values of the state variables of the ODEs need to be known and given. The conventional gradient-based optimization methods such as the Gauss-Newton method, the Levenberg-Marquardt method and the quasi-Newton method may fail to converge or may converge to a local minima if the initial values of the state variables and unknown parameters are not close enough to the true values. Thus, the computationally-intensive global optimization method may need to be used to solve the problem. The parameter estimates from the NLS method are also sensitive to the initial values of state variables which are not available in many biomedical applications. Second, the spline smoothing-based approaches (Varah, 1982; Ramsay and Silverman, 2005; Poyton et al, 2006; Ramsay et al., 2007) may not be flexible enough to deal with the complicated local features of the data. The rigorous asymptotic properties of these estimators have not been established. The PDA method and penalized spline approaches (Ramsay and Silverman, 2005; Ramsay et al., 2007) also need more efficient optimization techniques and complicated iterative computation algorithms to obtain an estimator. The convergence of the computational algorithms needs to be justified

(Ramsay et al., 2007). Third, the computational cost is high for most existing methods due to repeatedly solving the ODEs numerically or complicated optimization algorithms.

In this paper we attempt to develop a local kernel smoothing-based method as an alternative approach to estimate the unknown parameters $\boldsymbol{\beta}$ for the general ODE model (1.1). At the same time, we expect that our new method can ease the aforementioned problems of the existing methods. In Section 2, we formulate the estimation problem of the ODE model into a framework of measurement error in linear or nonlinear regression models. We also introduce a local polynomial smoothing procedure for estimation of the state function $\mathbf{X}(t)$ and its derivative that will be used to derive the main results in Section 3. In Section 2, we also briefly introduce the SIMEX approach to deal with measurement error in nonlinear regression models. We present our proposed method and main theoretical results in Section 3. We consider two examples for numerical illustration and compare our proposed method to the SIMEX method via Monte Carlo simulation studies in Section 4. In Section 5, an application to HIV dynamics data from an AIDS clinical trial is presented to illustrate the usefulness of the proposed method. We conclude the paper with some discussions in Section 6.

2. Estimation Procedure under a Framework of Measurement Error in Regression Models

Since the model (1.2) assumes that the state variables $\mathbf{X}(t)$ are observable with noise, we are able to estimate both $\mathbf{X}(t)$ and its derivative $\mathbf{X}'(t) = d\mathbf{X}(t)/dt$. Suppose $\hat{\mathbf{X}}'(t)$ is an estimator of $\mathbf{X}'(t)$. Substituting the estimates $\hat{\mathbf{X}}'(t_i)$, $i = 1, \ldots, n$, in the dynamic equation (1.1), we obtain a regression model:

$$\widehat{\mathbf{X}}'(t_i)=F\{\mathbf{X}(t_i),\beta\}+\Delta(t_i), \qquad (2.1)$$

where $\Delta(t_i)$ denotes the substitution error vector, that is $\Delta(t_i) = \hat{\mathbf{X}}'(t_i) - \mathbf{X}'(t_i)$. If $\hat{\mathbf{X}}'(t_i)$ is an unbiased estimator of $\mathbf{X}'(t_i)$, $\Delta(t_i)$ are errors with mean zero but are not independent. However, if the estimator $\hat{\mathbf{X}}'(t_i)$ is a biased estimator (e.g., the local polynomial estimator in our case), $\Delta(t_i)$ are not mean zero errors. Thus, $\Delta(t_i)$ are different from the conventional measurement error. This complexity makes it challenging to study the properties of the proposed estimator for $\boldsymbol{\beta}$.

In the regression model (2.1), the predictor $\mathbf{X}(t)$ is not directly observed, and instead one observes $\mathbf{Y}(t) = \mathbf{X}(t) + e(t)$, which adds another complexity to model (2.1). We need to deal with the problem of linear/nonlinear regression with measurement error in covariates. Otherwise, if we naively replace $\mathbf{X}(t)$ by $\mathbf{Y}(t)$ in the model (2.1), the parameter estimates are biased (Carroll, Ruppert, Stefanski, and Crainiceanu, 2006). An alternative choice is to replace $\mathbf{X}(t)$ by its estimate. This idea is essentially similar to the regression calibration technique in measurement error models, i.e., the error-prone covariate is replaced by an estimator from the regression on its surrogate. For details on regression calibration methods, see Carroll et al. (2006).

In this paper, the covariate or predictor $\mathbf{X}(t)$ is a solution to the ODE models and is assumed to be a smooth function of time t. Thus, we propose replacing the error-prone variable $\mathbf{X}(t)$ by its estimator obtained from a nonparametric smoothing method. Another alternative method for nonlinear regression models with measurement error in covariates is the simulation extrapolation (SIMEX) algorithm (Cook and Stefanski, 1994; Carroll et al., 2006). In the following two subsections, we briefly introduce the local polynomial smoothing method for the estimation of $\mathbf{X}(t)$ and its derivative and the SIMEX algorithm.

2.1. Local polynomial estimation of X(t) and X′(t)

To estimate the parameters of interest in the ODE model (1.1) under the framework of measurement errors in a nonlinear regression model, we first need to estimate the state variable $\mathbf{X}(t)$ and its derivative $\mathbf{X}'(t)$. For notational simplicity, we consider the univariate state variable case ($k = 1$) in the following methodology development and denote $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ by $X(t)$ and $Y(t)$, respectively. Extension to the multivariate case ($k > 1$) is straightforward although cumbersome.

Let $(Y_1, \ldots, Y_n)$ be the observations at the time points $t_1, \ldots, t_n$. Rewriting (1.2) as

$$Y_i = X(t_i) + e_i, \quad i = 1, \ldots, n,$$

where $(e_1, e_2, \ldots, e_n)$ are independent with mean zero and finite variance $\sigma^2(t_i)$. This is a traditional nonparametric regression model, and conventional regression techniques such as local polynomial regression, smoothing spline, and regression spline, among others, can be used to estimate $X(t)$ and $X'(t)$. Here we employ the local polynomial approach.

In this paper we use local linear regression to estimate $X(t)$ and local quadratic regression to estimate $X'(t)$. It is noteworthy that the higher degree polynomial kernel methods can also be employed to estimate $X(t)$ and $X'(t)$. We chose the local linear and local quadratic smoothers due to their simplicity and efficiency as suggested by Fan and Gijbels (1996). Also the bandwidth (smoothing parameter) selection is more critical than the degrees of polynomial smoother.

For presentation completeness, we briefly summarize the local polynomial regression procedure. We assume that the third derivative of $X(t)$ exists. For each given time point t_0, we approximate the function $X(t_i)$ locally by a pth-order polynomial; that is,

$$X(t_i) \approx X(t_0) + (t_i - t_0)X^{(1)}(t_0) + \cdots + X^{(p)}(t_0)(t_i - t_0)^p / p! \triangleq \sum_{j=0}^{p} \alpha_j(t_0)(t_i - t_0)^j$$

, for t_i, $i = 1, \ldots, n$, in a neighborhood of the point t_0, where $\alpha_j(t_0) = X^{(j)}(t_0)$ for $j = 0, 1, \ldots, p$. Following the local polynomial fitting (Fan and Gijbels, 1996), the estimators $\hat{X}^{(v)}(t)$ of $X^{(v)}(t)$ ($v = 0, 1$ in our case) can be obtained by minimizing the locally weighted least-squares criterion,

$$\sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} \alpha_j (t_i - t)^j \right\}^2 K_h(t_i - t),$$

where $K(\cdot)$ is a symmetric kernel function, $K_h(\cdot) = K(\cdot/h)/h$, and h is a proper bandwidth. Assuming that the matrix $\mathbf{T}_{p,t}^{\mathrm{T}} \mathbf{W}_t \mathbf{T}_{p,t}$ is not singular, the standard weighted least squares theory leads to the solution

$$\widehat{\alpha} = (\mathbf{T}_{p,t}^{\mathrm{T}} \mathbf{W}_t \mathbf{T}_{p,t})^{-1} \mathbf{T}_{p,t}^{\mathrm{T}} \mathbf{W}_t Y,$$

where $Y = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ is the vector of responses, here $p = 1$ or 2, and

$$\mathbf{T}_{p,t}=\begin{bmatrix} 1 & t_1-t & \dots & (t_1-t)^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_n-t & \dots & (t_n-t)^p \end{bmatrix}$$

is an $n \times (p + 1)$ design matrix and

$$\mathbf{W}_t=\text{diag}\{K_h(t_1-t),\dots,K_h(t_n-t)\}$$

is an $n \times n$ diagonal matrix of kernel weights. As a consequence, the estimators $\hat{X}(t)$ and $\hat{X}'(t)$ can be expressed as

$$\widehat{X}(t)=\xi_1^{\mathrm{T}}(\mathbf{T}_{1,t}^{\mathrm{T}}\mathbf{W}_t\mathbf{T}_{1,t})^{-1}\mathbf{T}_{1,t}^{\mathrm{T}}\mathbf{W}_tY,$$
$$\widehat{X}'(t)=\xi_2^{\mathrm{T}}(\mathbf{T}_{2,t}^{\mathrm{T}}\mathbf{W}_t\mathbf{T}_{2,t})^{-1}\mathbf{T}_{2,t}\mathbf{W}_tY,$$

where $\boldsymbol{\xi}_1$ is the 2×1 vector having 1 in the first entry and zero in the 2nd entry, while ξ_2 is the 3×1 vector having 1 in the 2nd entry and zeros in the other entries. Note that $\hat{X}'(t)$ is actually the slope of the local quadratic fit.

The asymptotic biases and the variances of the local linear estimator of $\hat{X}(t)$ and the local quadratic estimator of $\hat{X}'(t)$, under Assumption A in Section 3, are given below (see Fan and Gijbels (1996) for detailed derivation of these results).

$$\text{bias}\{\widehat{X}(t)|t_1,\cdots,t_n\}=\frac{1}{2}h^2X''(t)\mu_2(K)+o(h^2)+O(n^{-1}), \tag{2.2}$$

$$\text{var}\{\widehat{X}(t)|t_1,\cdots,t_n\}=(nh)^{-1}\mu_0(K^2)/f(t)+o\{(nh)^{-1}\}, \tag{2.3}$$

$$\text{bias}\{\widehat{X}'(t)|t_1,\cdots,t_n\}\}=\frac{1}{3!}\mu_4(K)X^{(3)}(t)h^2, \tag{2.4}$$

$$\text{var}\{\widehat{X}'(t)|t_1,\cdots,t_n\}=n^{-1}h^{-3}\mu_2(K^2)/f(t)+o(n^{-1}h^{-3}), \tag{2.5}$$

where and below $f(t)$ is the density of t, and $\mu_\ell(K)=\int_{-1}^{1}z^\ell K(z)\,dz$ for $\ell = 0, 1 \dots 4$. These results will be used to derive the asymptotic properties of our proposed estimator in the next section. Note that the estimator of $X'(t)$ achieves the second-order kernel bias rate of order h^2 which is the same as that of the estimator of $\hat{X}(t)$. However, the asymptotic variance rate of the estimator $\hat{X}'(t)$ is higher than that of $\hat{X}(t)$ (i.e., the order h^{-2}). We also noticed (as pointed out by one referee) that the local quadratic estimator of $X'(t)$ improves its local linear estimator. The orders of the bias and the variance of the local quadratic estimator of $X'(t)$ are the same as those of the local linear estimator of $X'(t)$, but an extra constant in the bias expression of the local quadratic estimator of $X'(t)$ creates an opportunity for significant bias reduction especially in the boundary and highly clustered design regions (although the order of the convergence rate of the two estimators are the same). This argument is similar to that the local linear estimator is preferable compared to the local constant estimator for estimating the original function (Fan and Gijbels, 1996, Section 3.3).

2.2. The SIMEX algorithm

The SIMEX algorithm is a useful tool for dealing with measurement error in covariates for nonlinear regression models. It is a functional method which can be applied without making any assumption about the distribution of unobservable covariates. We have formulated the parameter estimation problem for the ODE model into a framework of measurement error in a nonlinear regression model (2.1). Thus, we can directly apply the SIMEX approach to our model (2.1), which will be used to serve as a comparison basis for the PsLS method that will be proposed in the next section. The SIMEX method was initially proposed by Cook and Stefanski (1994). A detailed description of this method can be found in Carroll et al. (2006). Here we briefly outline the algorithm based on the SIMEX principle.

Assume that there is a function $\mathcal{J}$ for estimating $\boldsymbol{\beta}$ when $X(t)$ is measured without error, and we call this estimator a naive estimator of $\boldsymbol{\beta}$ and it is denoted by $\hat{\boldsymbol{\beta}}_{\text{naive}} = \mathcal{J}(X)$. Also the measurement error variance Σ_e is assumed to be known exactly. The first step of the algorithm is to create additional data sets via simulations by adding increasingly large measurement error $(1 + \boldsymbol{\psi})\Sigma_e$ for $\boldsymbol{\psi} \geq 0$. For B simulated data sets with a theoretical measurement error $(1 + \boldsymbol{\psi})\Sigma_e$ for each data set, we compute the average estimates of $\hat{\boldsymbol{\beta}}$. For each of the data sets $b = 1, \ldots, B$, we define $\widehat{\beta}_{\psi,b}=\mathcal{J}\{W^*_{i,b}(\psi)\}$, where $W^*_{i,b}(\psi)=X(t_i)+\psi^{1/2}V_{i,b}$, and the $V_{i,b}$ are independently generated from a normal distribution with mean 0 and variance Σ_e. Define $\widehat{\beta}_{\psi}=B^{-1}\sum_{b=1}^{B}\widehat{\beta}_{\psi,b}$ and the extrapolant function $\mathcal{G}(\boldsymbol{\psi}) = \hat{\boldsymbol{\beta}}_{\boldsymbol{\psi}}$ as a function of $\boldsymbol{\psi}$. Note that $\mathcal{G}(0) = \hat{\boldsymbol{\beta}}$ (naive). The extrapolation step extrapolates the function $\mathcal{G}(\boldsymbol{\psi})$ back to $\boldsymbol{\psi} = -1$, i.e., $\mathcal{G}(-1)$ is the SIMEX estimator of $\boldsymbol{\beta}$. More details on how to select the extrapolant function and the implementation of the SIMEX method can be found in Carroll et al. (2006).

3. Pseudo-LS Estimator and Main Results

In this section, we propose a straightforward idea to estimate the unknown parameters in model (2.1). First we substitute a smoothing estimate of $X(t)$ in model (2.1), and then use the least squares principle to obtain the estimates of unknown parameters. Denote $\Delta(t) = \hat{X}'(t) - X'(t)$, then we have $\hat{X}'(t) = F\{X(t), \boldsymbol{\beta}\} + \Delta(t)$ and $\Delta(t)$ can be regarded as the “error.” The estimator of $\boldsymbol{\beta}$ is defined as the value of $\boldsymbol{\beta}$ which minimizes

$$S_n(\beta)=\sum_{i=1}^{n}\left[\widehat{X}'(t_i) - F\{\widehat{X}(t_i),\beta\}\right]^2 \tag{3.6}$$

subject to $\boldsymbol{\beta} \in \Omega_\beta$ (parameter space). Note that in this objective function, $\{\hat{X}(t_i), \hat{X}'(t_i); i = 1, \ldots, n\}$ are not the observed data and measured covariates, instead they are the smoothing estimates of the state variable $X(t)$ and its derivative which are not independently distributed. Thus, the estimator obtained by minimizing this objective function is not the true least squares (LS) estimator, instead we call this estimator the pseudo-least squares (PsLS) estimator denoted by $\hat{\boldsymbol{\beta}}_n$. In addition, the “error” term $\Delta(t_i)$ is neither independent nor mean zero as in a conventional nonlinear least squares (NLS) regression model. As a consequence, the study of the asymptotic properties for the proposed estimator is not trivial.

Implementation for obtaining the PsLS estimator is simple. If $F(\cdot, \cdot)$ is a linear function, an ordinary least squares procedure for linear regression models can be used to get the estimate of $\boldsymbol{\beta}$. Similarly, for a nonlinear function of $F(\cdot, \cdot)$, the nonlinear regression procedure from standard statistical packages such as SAS, Splus or R can be used to obtain the PsLS estimates. However, we need to set the smoothing estimate of the derivative function $\hat{X}'(t)$ as the response variable and the smoothing estimate of the state variable $\hat{X}(t)$ as the covariate at the observation time points $t = t_1, t_2, \ldots, t_n$.

Although the idea of the PsLS estimate is simple, it is critical to show that the PsLS estimator has good asymptotic properties such as consistency and asymptotic normality. For the standard nonlinear least-squares (NLS) estimator, the asymptotic properties have been established (Jennrich, 1969; Malinvaud, 1970; Wu, 1981). Similar ideas can be used to study the asymptotic properties of the proposed PsLS estimator. However, since the PsLS estimator is based on the nonparametric kernel estimator of the state variable and its derivative, the asymptotic results from the nonparametric kernel estimation need to be used. Here we present the asymptotic results of the proposed PsLS estimator, while a sketch of the main ideas of the proofs for these results is given in the Appendix. Let

$$B_n(\beta_1,\beta_2)=\sum_{i=1}^{n} F\{X(t_i),\beta_1\}F\{X(t_i),\beta_2\},$$
$$D_n(\beta_1,\beta_2)=\sum_{i=1}^{n}[F\{X(t_i),\beta_1\}-F\{X(t_i),\beta_2\}]^2.$$

The strong law of large numbers for iid random variables implies that $B_n(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)/n$ converges to a function, say $B(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, for all $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ uniformly, and then $D_n(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)/n$ converges to $D(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = B(\boldsymbol{\beta}_1, \boldsymbol{\beta}_1) + B(\boldsymbol{\beta}_2, \boldsymbol{\beta}_2) - 2B(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. Now we give the following assumptions that are standard in NLS regression and local linear kernel estimation.

Assumption A

- **i.** The function $X^{(3)}(t)$ is continuous on [0, 1].
- **ii.** The kernel function K is symmetric about zero and is supported on [−1, 1].
- **iii.** The bandwidth $h = h_n = n^{-2/7} a_n$ is a sequence satisfying $h \to 0$ as $n \to \infty$, where a_n is a sequence tending to 0 slower than $\log^{-1} n$.
- **iv.** t_i are iid and have a common compact support and their density function, $f(t)$, is bounded away from zero and has bounded and continuous second derivatives.

Assumption B

- **i.** $F(x, \boldsymbol{\beta})$ is a continuous function of $\boldsymbol{\beta}$ for $\boldsymbol{\beta} \in \Omega_\beta$.
- **ii.** Ω_β is a closed, bounded compact subset of $\mathbb{R}^m$.
- **iii.** $D(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = 0$ if and only if $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

Assumption C

- **i.** The first and second partial derivatives, $\frac{\partial F(x,\beta)}{\partial\beta}, \frac{\partial^2 F(x,\beta)}{\partial x\partial\beta}, \frac{\partial^2 F(x,\beta)}{\partial\beta\partial\beta^{\mathrm{T}}}$, exist and are continuous for all $\boldsymbol{\beta} \in \Omega_\beta$, $x \in \chi$, and

$$|\frac{\partial F(x_1,\beta)}{\partial \beta} - \frac{\partial F(x_2,\beta)}{\partial \beta}| \le C_1|x_1 - x_2|^{\zeta}$$

for some $0 < \zeta \le 1$.

ii. The first partial derivative $\frac{\partial F(x,\beta)}{\partial x}$ is continuous for $x \in \chi$ and satisfy:

$$\sup_{x\in\chi}|\frac{\partial F(x,\beta)}{\partial x}| \le M_{\beta}.$$

We present our main results on the consistency and the asymptotic distribution of the proposed PsLS estimator as follows (the proofs are relegated in the Appendix).

Theorem 1—Under Assumptions A–C, the PsLS estimator $\hat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ is strongly consistent.

Theorem 2—Under Assumptions A–C, $nh^{3/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ asymptotically follows a normal distribution with mean zero and covariance matrix given in (A.10).

Remark 1—Note that in the proof of Theorem 2, we need to deal with the local polynomial estimators of X(t) and $X'(t)$ which make it different and more complex compared to the proof of the asymptotic normality of the standard NLS estimator. It is also noteworthy that if F(X, $\boldsymbol{\beta}$) is a linear function, say F(X, $\boldsymbol{\beta}$) = $X^T\boldsymbol{\beta}$, the assumptions (B) and (C) are satisfied. As a consequence, the corresponding linear LS estimator $\hat{\boldsymbol{\beta}}_n$ is $n^{-1}h^{-3/2}$-consistent and asymptotically normal with the asymptotic covariance $\sigma_e^2\mu_2^{-2}(K)\mu(K^2)\{E(XX^{\mathrm{T}})\}^{-1}$.

Remark 2—Theorem 2 shows that the proposed PsLS estimator of $\boldsymbol{\beta}$ is still asymptotically normal. However, the convergence rate of the PsLS estimator is not root–n as that of the standard NLS estimator (Jennrich, 1969; Malinvaud, 1970; Wu, 1981; Seber and Wild, 1989), instead the convergence rate is $n^{-1}h^{-3/2}$, which is faster than the conventional root–n. The reason for this interesting result is because the variance of the error term $\Delta(t_i)$ in the regression model (2.1) is not a constant, instead it goes to zero with the rate of $(nh)^{-1}$, which is a consequence of data smoothing from the first step. This smaller variance results in a faster convergence rate of the estimator of β compared to the standard root–n convergence rate of the nonlinear least squares estimate.

Remark 3—Bandwidth selection is critical in local polynomial regression. The bandwidths for smoothing X(t) and $X'(t)$ in the first step of our estimation procedure need to satisfy some conditions in order to guarantee the consistency and asymptotic normality of the PsLS estimator. Note that, for the standard local linear estimator, the optimal bandwidth for estimating X(t) can be obtained using the data-driven cross-validation method or the substitution method based on the asymptotic mean integrated squared error (Ruppert, Sheather and Wand, 1995). This optimal bandwidth, $\hat{h}_{opt}$, is of order $n^{-1/5}$. However, the order of this optimal bandwidth does not satisfy the Assumption A (iii) for Theorems 1 and 2 which requires the bandwidth $h = h_n = n^{-2/7}a_n$, where a_n is a sequence tending to 0 slower than $\log^{-1} n$. This assumption is needed to stabilize the asymptotic variance in Theorem 2. In addition, we need to select a bandwidth which lets the asymptotic bias of the PsLS estimator approach zero as fast as possible. Thus, we need to undersmooth the data in the first step. For example, we may select the bandwidth $h = \hat{h}_{opt} \times n^{-3/35}a_n$ which will satisfy the Assumption A (iii), where a_n can be selected as $\log^{-r} n$ with r being a positive fractional number. This result only provides an ad hoc guidance for bandwidth selection since the constant in the asymptotic results cannot be determined. The data-driven approach for bandwidth selection is complicated under our model setting and is a worthy topic for future research.

4. Simulation Studies

FitzHugh (1961) and Nagumo et al. (1962) simplified the Hodgkin-Huxley model (1952) for the behavior of spike potentials in the giant axon of squid neurons. They reduced the original Hodgkin-Huxley model from four variables to two variables so that phase plane techniques could be used for the analysis of the model. The FitzHugh-Nagumo model can be described by the following two equations:

$$\begin{cases} \frac{dx_1(t)}{dt} = \{x_1(t) + x_2(t) - x_1^3(t)\}\gamma, \\ \frac{dx_2(t)}{dt} = -\{x_1(t) - \alpha + \beta x_2(t)\}/\gamma, \end{cases} \tag{4.1}$$

where α, β, and γ are the parameters of interest, while $x_1(t)$ and $x_2(t)$ are the state variables indicating the voltage across an axon membrane and outward currents respectively. This model has been widely used due to its simplicity and flexibility. This model is flexible in its ability to reproduce many qualitative characteristics of electrical impulses along nerve and cardiac fibers, such as the existence of an excitation threshold, relative and absolute refractory periods, and the generation of pulse trains under the action of external currents. It is also very useful in genetics, biology, and heat and mass transfer systems.

The study of HIV viral dynamics over the past decade has led to a good understanding of the pathogenesis of HIV infection (Ho et al., 1995; Perelson et al., 1996, 1997; Notermans et al., 1998, Wu et al., 1999). Ordinary differential equation (ODE) models were originally proposed to describe the interactions between HIV virus and immune cellular response. See Perelson and Nelson (1999), Nowak and May (2000) and Tan and Wu (2005) for recent reviews of these models.

One popular HIV dynamic model can be written as

$$\frac{d}{dt}T_U(t) = \lambda - \rho T_U(t) - \eta(t)T_U(t)V(t), \tag{4.2}$$

$$\frac{d}{dt}T_I(t) = \eta(t)T_U(t)V(t) - \delta T_I(t), \tag{4.3}$$

$$\frac{d}{dt}V(t) = N\delta T_I(t) - cV(t), \tag{4.4}$$

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

$(c, \delta, \lambda, \rho, N, \eta(t))^T$ are unknown dynamic parameters. Similar HIV dynamic models have been proposed and studied by many investigators since the early 1990's (Ho et al., 1995; Perelson and Nelson, 1999; Nowak and May, 2000, Tan and Wu, 2005).

In this section, we present the results from simulation experiments generated from models (4.2)–(4.4) and (4.1) for studying the finite sample properties of the proposed methods, the PsLS estimates and the SIMEX estimates. In local polynomial smoothing, we used the kernel function $K(u) = 3/4(1 - u^2)I_{(|u|\leq 1)}$. We selected the bandwidth using the strategy given in Remark 3. We first obtained the standard optimal bandwidth, h_{opt}, using the substitution method based on the asymptotic mean integrated squared error (Ruppert, Sheather and Wand, JASA, 1995). Then we used the result, $h = \hat{h}_{opt} \times n^{-3/35}\, a_n$, where a_n was selected as $a_n = \log^{-1/16} n$ based on our experience. In implementing the SIMEX algorithm, we use the quadratic extrapolating function and take $\psi = 0, 0.2, \ldots, 2$ and $B = 100$. For each configuration below, we ran 500 replications. To evaluate the performance of different methods, we define the average relative estimation error (ARE) of a parameter θ as

$$\mathrm{ARE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\widehat{\theta} - \theta|}{|\theta|} \times 100\%,$$

where $\hat{\theta}$ is the estimate of θ and N is the number of simulation runs (here $N = 500$).

Example 1

First we perform simulations for the FitzHugh-Nagumo equations. We generated the data from the FitzHugh-Nagumo equation (4.1). Our true parameter values are taken as $\alpha_0 = 0.34$, $\beta_0 = 0.2$, and $\gamma_0 = 3$, and initial conditions $\{x_1, x_2\}$ are (0, 0.1). We selected σ_1^2, and σ_2^2 as 0.05, 0.06, …, 0.10 respectively. Our data were obtained by solving the equations (4.1) at every 0.1 time units on the interval [0, 20], and then measurement errors were added as follows.

$$y_{1i} = x_1(t_i) + \varepsilon_{1i},$$
$$y_{2i} = x_2(t_i) + \varepsilon_{2i},$$

where ε_{1i} and ε_{2i} are independently normally distributed with mean 0 and standard deviations σ_1 and σ_2 respectively. We therefore have a total of 36 scenarios of different variance parameter combinations and each simulation data set has 201 observations.

We applied the proposed PsLS and SIMEX methods to the simulated data sets to estimate the unknown parameters (α, β, γ) in the FitzHugh-Nagumo equations. We report the averages of the estimated values, associated errors and coverage probabilities of the PsLS estimates and SIMEX estimates for all 36 scenarios in Table 1, and the associated AREs in Table 2. Table 1 shows that the point estimates of the parameters are reasonably close to the true values and the coverage probabilities are close to the nominal level for both methods. From Table 2, we can see that the AREs of the estimates for α and β are quite similar between the PsLS method and SIMEX method. However, the AREs of the estimate for γ from the PsLS method are consistently smaller than those from the SIMEX method for all cases.

To evaluate the goodness-of-fit, we obtained the predicted (fitted) values of $X_1(t)$ and $X_2(t)$ and their derivatives by solving the ODEs (4.1) with the estimated parameter values. We present the predicted curves of $X_1(t)$ and $X_2(t)$ and their derivatives for the case of $\sigma_1^2 = \sigma_2^2 = 0.1$ from the two methods, PsLS and SIMEX, as well as the corresponding true curves (by solving the ODEs using the true parameter values) in Figure 1, in which the associated 95% pointwise confidence intervals, i.e., the 2.5% and 97.5% quantiles of the estimates from the 500 replications, of these state variables and their derivatives are also delineated. Although the presented results in Figure 1 are from the case with the largest measurement errors, we can see that the predicted curves of $X_1(t)$ and $X_2(t)$ and their derivatives have good agreement with the corresponding true curves for both estimation methods.

Example 2

In the HIV dynamic example, we generated data from models (4.2)–(4.4) with the initial values $(T_{U0}, T_{I0}, V_0) = (600, 30, 10^5)$ and the true values of parameters $(\lambda_0, \rho_0, N_0, \delta_0, c_0) = (36, 0.108, 10^3, 0.5, 3)$ and the time-varying parameter $\eta(t) = 9 * 10^{-5}\{1 - 0.9\cos(\pi t/1000)\}$.

In AIDS clinical studies or clinical practice, only plasma viral load $V(t)$ and the total CD4+ T cell counts $T(t) = T_I(t) + T_U(t)$ can be measured. We therefore combine equations (4.2) and (4.3), and obtain

$$\frac{d}{dt}\{T_U(t)+T_I(t)\}=\lambda - \rho T_U(t) - \delta T_I(t).$$

Notice that $T(t) = T_I(t) + T_U(t)$ and substitute $T_U(t) = T(t) - T_I(t)$ in the above equation, we obtain

$$\frac{d}{dt}T(t)=\lambda - \rho\{T(t) - T_I(t)\} - \delta T_I(t).$$

Denote $T' = dT(t)/dt$ and from the above equation, we can get

$$T_I=\frac{-\lambda}{\rho-\delta}+\frac{\rho}{\rho-\delta}T+\frac{1}{\rho-\delta}T'.$$

Substitute this into equation (4.4) and let $\alpha_0=-\frac{N\delta\lambda}{\rho-\delta}$, $\alpha_1=\frac{N\delta\rho}{\rho-\delta}$ and $\alpha_2=\frac{N\delta}{\rho-\delta}$, we have

$$V'(t)=\alpha_0+\alpha_1 T(t)+\alpha_2 T'(t) - cV(t), \tag{4.5}$$

where $V(t)$ and $T(t)$ for $t = t_1, t_2, \ldots, t_n$ are measurements from AIDS clinical studies. If we obtain the estimates of $(\alpha_0, \alpha_1, \alpha_2)$, we can derive the estimates of important viral dynamic parameters using the relationships:

$$\lambda=-\alpha_0/\alpha_2,\ \ \rho=\alpha_1/\alpha_2,\ \ N=\alpha_1/\delta-\alpha_2.$$

Here we assume that the parameters δ and c are known and can be obtained from the literature (Perelson et al., 1996; Perelson et al., 1997; Wu, Ding, and DeGruttola, 1998; Wu and Ding, 1999; Fitgerald et al., 2002; Wu, 2005; Han and Chaloner, 2004). Our primary interest is to estimate parameters (α_0, α_1, α_2) or (λ, ρ, N) which have never been estimated from clinical data.

Note that our observation (measurement) models for this example are

$$y_{1i}=T(t_i)+\varepsilon_{1i},$$
$$y_{2i}=V(t_i)+\varepsilon_{2i}.$$

In our simulations, we assumed that (ε_{1i}, ε_{2i}) are independent and follow normal distributions with mean zero and variances $\sigma_1^2=20, 30, 40$ and $\sigma_2^2=100, 150, 200$ respectively. The simulated data were generated by numerically solving equations (4.2)–(4.4) and two output schedules were used: (i) at every 0.1 time units on the interval [0, 20], and (ii) at every 0.2 time units on the interval [0, 20] which correspond to two sample size cases, 200 and 100 respectively. Measurement noise was then added to the numerically generated data based on the above observation equations.

First we employed a local smoothing method to obtain the estimates of $V'(t)$, $V(t)$, $T'(t)$, and $T(t)$, say, $\hat{V}'(t)$, $\hat{V}(t)$, $\hat{T}'(t)$, and $\hat{T}(t)$, respectively, then we have

$$\widehat{V}'(t)=\alpha_0+\alpha_1\widehat{T}(t)+\alpha_2\widehat{T}'(t)-c\widehat{V}(t)+\Delta(t). \tag{4.6}$$

We applied the proposed PsLS and SIMEX methods in Sections 2 and 3 to estimate the parameters (α_0, α_1, α_2) or (λ, ρ, N). We report the averages of the estimated values, associated errors and coverage probabilities of the PsLS estimates and SIMEX estimates for all 18 scenarios in Table 3, and the associated AREs in Table 4.

Table 3 shows that the point estimates of the parameters are close to the true values and the coverage probabilities are close to the nominal level. Table 4 shows that the average relative errors of the estimates of parameter N from both PsLS and SIMEX methods is reasonably small, while the estimates of λ and ρ are less accurate, in particular for the small sample size case. However, from Table 4 we can clearly see that the AREs of the proposed PsLS method are smaller for the estimates of parameters λ and ρ. For comparisons, we present the fitted curves and the associated 95% pointwise confidence intervals (dashed lines from the PsLS method and dotted lines from the SIMEX method) for the case of

$\sigma_1^2=40, \sigma_2^2=200$ and sample size n=200 superimposed on the corresponding true curves of the state variables and their derivatives (solid lines) in Figure 2. We can see that the PsLS method also fitted the true curve better.

5. Applications

The experimental data for the FitzHugh-Nagumo equations are rarely available (Ramsay et al., 2007). But in the study of HIV dynamics, extensive clinical data have been collected from many clinical trials. A clinical trial was designed to monitor HIV dynamics frequently by one of the authors of this paper and his clinical collaborators. In this study, HIV-1 infected patients were recruited to be treated by antiviral therapies and immune-based treatment. This study measured HIV viral load at hours 0, 1, 2, 3, 4, 6, 8, 10, 12, 14, 16, 18, 20, 24, 28, 32, 40, 46, 52, 58, 64, 70, 144, 240, and 336 during the first two weeks of treatment, and then at weeks 3, 4, 6, 8, 10, 12, 14, 16, 20, 24, 28, 32, 36, 40, 44 and 48 during treatment. At most weekly clinical visits, total CD4 T cell counts were also measured.

Similar to the simulation study example, we fitted model (4.5) to the viral load data using the proposed PsLS and SIMEX methods. Similar bandwidth selection method was used, i.e., the formula $h = \hat{h}_{opt} \times n^{-3/35} \log^{-1/16} n$ was employed. To save space, we present the parameter estimation results from two patients as follows (the delta method was used to obtain the standard error of the kinetic parameters):

1. Patient #1
    a. PsLS: $\lambda = 47.4$ (s.e. 14.3), $\rho = 0.085$ (s.e. 0.057), $N = 623$ (s.e. 17.4), $c = 0.074$ (s.e. 0.003)
    b. SIMEX: $\lambda = 43.2$ (s.e. 25.1), $\rho = 0.075$ (s.e. 0.082), $N = 598$ (s.e. 24.51), $c = 0.136$ (s.e. 0.104)
2. Patient #2
    a. PsLS: $\lambda = 45.6$ (s.e. 12.4), $\rho = 0.071$ (s.e. 0.004), $N = 469$ (s.e. 47.6), $c = 0.083$ (s.e. 0.004)
    b. SIMEX: $\lambda = 39.3$ (s.e. 20.3), $\rho = 0.094$ (s.e. 0.005), $N = 512$ (s.e. 36.5), $c = 0.103$ (s.e. 0.004)

The fitted (predicted) curves of viral load and total CD4 T cell counts and their derivatives are shown in Figure 3. From this figure, we can see that the fitted curves compare well to the observed data. The estimates of the derivatives of the viral load and CD4 T cell counts are reasonably estimated. These estimation results may provide important information for clinicians to make treatment decisions for individual AIDS patients.

6. Discussion

Formal statistical estimation methods for parameters in ordinary differential equation (ODE) models are relatively new in statistical literature (Li et al., 2002; Huang and Wu, 2006; Huang, Liu and Wu, 2006; Ramsay et al., 2007; Chen and Wu 2008a, 2008b; Miao et al. 2008a). In this paper, we have proposed a PsLS method to deal with this problem under the framework of measurement error models. We also compared our PsLS method to a popular method for dealing with measurement errors in nonlinear regression models, the SIMEX method. We found out that the performance of the proposed PsLS method is as good as the SIMEX method for most cases, and is better than the SIMEX method for some other cases based on our simulation studies although we did not expect this. What we expected was that the proposed PsLS method should be comparable to the SIMEX method in the sense of estimation error, but should achieve significant benefits in computational cost, which is true based on our simulation studies and real data applications (the PsLS method is more than 20 times faster than the SIMEX method). The proposed methods do not require numerically solving the ODEs, but instead use local smoothing methods to estimate the state functions and their derivatives. We also established the consistency and asymptotic normality of the proposed PsLS estimator.

Note that the intention of the proposed PsLS method is not to try to improve the existing methods such as the standard nonlinear LS (NLS) method (Seber and Wild 1989 and Bates and Watts 1988) and penalized spline method (Ramsay 1996, Ramsay et al. 2007) in the sense of estimation efficiency or accuracy, but instead our method provides an alternative estimation approach for ODE models in the framework of measurement error models to avoid some critical problems of these existing methods that include: i) the requirement and sensitivity of initial values of the state variables for ordinary differential equation (ODE) models on the parameter estimation, in particular for the NLS method; ii) the convergence problem of the NLS method and other existing methods; iii) high computational cost due to iteratively solving the ODEs numerically in the estimation procedure; and iv) high computational cost due to complicated optimization techniques. However, there is a cost associated with the proposed method. Our PsLS method does alleviate these problems, but pays a price in terms of efficiency (the estimation error will be a little bit larger as we expected). Another limitation of the proposed method is that it requires frequent measurement data of state variables since the first step of the proposed method is to apply local smoothing methods to estimate the state variables and their derivatives. In particular, reliable estimates of derivative functions require a relatively large sample size.

In summary, the proposed PsLS estimation method has several advantages compared to the existing methods although it may not improve the performance of the existing methods in the sense of estimation accuracy. These advantages include: 1) computational efficiency; 2) easing of the convergence problem; 3) the initial values of the state variables of the differential equations not required; and 4) providing good initial estimates of the unknown parameters for other computationally-intensive methods to further refine the estimates rapidly. We are currently investigating how to combine the proposed PsLS method with other existing methods (e.g. the NLS method) to overcome the computational problems of the existing methods, while at the same time also improve the estimation efficiency (accuracy). We hope to report some promising results along this line in the near future.

In this paper, we also assumed that the parameters in the ODE models are uniquely identifiable. The ODE model identifiability is another interesting topic, but beyond scope of this paper. Some references for nonlinear ODE model identifiability, including HIV dynamic models, can be found in Conte, Moog and Perdon (1999), Tunali and Tarn (1987), Diop and Fliess (1991), Ljung and Glad (1994), Xia and Moog (2003), Jeffrey and Xia (2005), Miao et al. (2008a), Miao et al. (2008b), and Wu et al. (2008). Another interesting extension of the proposed methods is to incorporate mixed-effects modeling idea to deal with longitudinal data (Huang and Wu, 2006; Huang, Liu and Wu, 2006). This will be the next focus of our research.

Acknowledgments

The authors would like to thank Yun Fang for her careful reading of an earlier version of this manuscript and her help in correcting an error in the theoretical results. The authors are also grateful to an associate editor and three referees for their constructive comments and suggestions. This research was partially supported by the NIAID/NIH grants AI62247, AI059773, AI50020, AI052765, AI055290, and AI27658.

References

Bard, Y. Nonlinear Parameter Estimation. London: Academic; 1974.

Bates, DM.; Watts, DB. Nonlinear Regression Analysis and Its Applications. New York: Wiley; 1988.

Bock, HG. Numerical Treatment of Inverse Problems in Chemical Reaction Kinetics. In: Ebert, K.; Deuflhard, P.; Jager, W., editors. Modelling of Chemical Reaction Systems. New York: Springer; 1981. p. 102-125.

Carroll RJ, Küchenhoff H, Lombard F, Stefanski LA. Asymptotics for The SIMEX Estimator in Structural Measurement Error Models. Journal of the American Statistical Association 1996;91:242–50.

Carroll, RJ.; Ruppert, D.; Stefanski, LA.; Crainiceanu, CM. Measurement Error in Nonlinear Models. Vol. 2. New York: Chapman and Hall; 2006.

Chen J, Wu H. Efficient Local Estimation for Time-varying Coefficients in Deterministic Dynamic Models with Applications to HIV-1 Dynamics. Journal of the American Statistical Association 2008a; 103:369–384.

Chen J, Wu H. Estimation of Time-varying Parameters in Deterministic Dynamic Models With Application to HIV Infections. Statistica Sinica. 2008bin press

Conte, G.; Moog, CH.; Perdon, AM. Nonlinear Control Systems: An Algebraic Setting. London: Springer; 1999.

Cook JR, Stefanski LA. Simulation-Extrapolation Estimation in Parametric Measurement Error Models. Journal of the American Statistical Association 1994;89:1314–28.

Diop, S.; Fliess, M. On Nonlinear Observability. Proc of the first Europ Control Conf; Paris, Hermes. 1991. p. 152-157.

Fan, J.; Gijbels, I. Local Polynomial Modeling and its Applications. London: Chapman and Hall; 1996.

Fitzgerald AP, DeGruttola V, Vaida F. Modeling HIV Viral Rebound Using Non-Linear Mixed Effects Models. Statistics in Medicine 2002;21:2093–2108. [PubMed: 12111889]

FitzHugh R. Impulses And Physiological States in Models of Nerve Membrane. Biophysical Journal 1961;1:445–466. [PubMed: 19431309]

Han C, Chaloner K. Bayesian Experimental Design for Nonlinear Mixed-Effects Models With Application to HIV Dynamics. Biometrics 2004;60:25–33. [PubMed: 15032770]

Hemker, PW. Numerical Methods for Differential Equations in System Simulation And in Parameter Estimation. In: Hemker, HC.; Hess, B., editors. Analysis and Simulation of Biochemical Systems. 1972. p. 59-80.

Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid Turnover of Plasma Virions And CD4 Lymphocytes in HIV-1 Infection. Nature 1995;373:123–126. [PubMed: 7816094]

Hodgkin AL, Huxley AF. A Quantitative Description of Membrane Current And Its Application to Conduction And Excitation in Nerve. Journal of Physiology 1952;133:444–479.

Huang Y, Wu H. A Bayesian Approach for Estimating Antiviral Efficacy in HIV Dynamic Models. Journal of Applied Statistics 2006;33:155–174.

Huang Y, Liu D, Wu H. Hierarchical Bayesian Methods for Estimation of Parameters in A Longitudinal HIV Dynamic System. Biometrics 2006;62:413–423. [PubMed: 16918905]

Jeffrey, AM.; Xia, X. Identifiability of HIV/AIDS Model. In: Tan, WY.; Wu, H., editors. Deterministic And Stochastic Models Of AIDS Epidemics and HIV Infections With Intervention. Singapore: World Scientific; 2005.

Jennrich RI. Asymptotic Properties of Nonlinear Least Squares Estimators. Annals of Mathematical Statistics 1969;40:633–643.

Li L, Brown MB, Lee KH, Gupta S. Estimation And Inference for A Spline-Enhanced Population Pharmacokinetic Model. Biometrics 2002;58:601–611. [PubMed: 12229995]

Li Z, Osborne M, Prvan T. Parameter Estimation in Ordinary Differential Equations. IMA Journal of Numerical Analysis 2005;25:264–285.

Ljung L, Glad T. On Global Identifiability For Arbitrary Model Parameterizations. Automatica 1994;30:265–276.

Mack Y, Silverman B. Weak And Strong Uniform Consistency of Kernel Regression Estimates. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 1982;60:405–415.

Malinvaud E. The Consistency of Nonlinear Regressions. Annals of Mathematical Statistics 1970;41:956–969.

Miao H, Dykes C, Demeter L, Wu H. Differential Equation Modeling of HIV Viral Fitness Experiments: Model Identification, Model Selection, and Multi-Model Inference. Biometrics. 2008ain press

Miao H, Dykes C, Demeter LM, Cavenaugh J, Park SY, Perelson AS, Wu H. Modeling and Estimation of Kinetic Parameters and Replicative Fitness of HIV-1 from Flow-Cytometry-Based Growth Competition Experiments. Bulletin of Mathematical Biology. 2008bin press

Nagumo JS, Arimoto S, Yoshizawa S. An Active Pulse Transmission Line Simulating a Nerve Axon. Proceedings of the IRE 1962;50:2061–2070.

Notermans DW, Goudsmit J, Danner SA, de Wolf F, Perelson AS, Mitter J. Rate of HIV-1 Decline Following Antiretroviral Therapy is Related to Viral Load at Baseline And Drug Regimen. AIDS 1998;12:1483–1490. [PubMed: 9727569]

Nowak, MA.; May, RM. Virus Dynamics: Mathematical Principles of Immunology and Virology. Oxford: Oxford University Press; 2000.

Ogunnaike, BA.; Ray, WH. Process Dynamics, Modeling, And Control. New York: Oxford University Press; 1994.

Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, And Viral Generation Time. Science 1996;271:1582–1586. [PubMed: 8599114]

Perelson AS, Essunger P, Cao YZ, Vesanen M, Hurley A, Saksela K, Markowitz M, Ho DD. Decay Characteristics of HIV-1-Infected Compartments During Combination Therapy. Nature 1997;387:188–191. [PubMed: 9144290]

Perelson AS, Nelson PW. Mathematical Analysis of HIV-1 Dynamics in Vivo. SIAM Review 1999;41:3–44.

Poyton AA, Varziri MS, McAuley KB, McLellan PJ, Ramsay JO. Parameter Estimation in Continuous-Time Dynamic Models Using Principal Differential Analysis. Computer and Chemical Engineering 2006;30:698–708.

Putter H, Heisterkamp SH, Lange JMA, De Wolf F. A Bayesian Approach to Parameter Estimation in HIV Dynamical Models. Statistics in Medicine 2002;21:2199–2214. [PubMed: 12210633]

Ramsay JO. Principal Differential Analysis: Data Reduction by Differential Operators. Journal of the Royal Statistical Society, Series B 1996;58:495–508.

Ramsay JO, Hooker G, Campbell D, Cao J. Parameter Estimation for Differential Equations: A Generalized Smoothing Approach (with Discussions). Journal of the Royal Statistical Society, Series B 2007;69:741–796.

Ramsay, JO.; Silverman, BW. Functional Data Analysis. Vol. 2. New York: Springer; 2005.

Ruppert D, Sheather SJ, Wand MP. An Effective Bandwidth Selector for Local Least Squares Regression. Journal of the American Statistical Association 1995;90:1257–1270.

Seber, GAF.; Wild, CJ. Nonlinear Regression. New York: John Wiley & Sons; 1989.

Stone CJ. Optimal Global rates of Convergence for Nonparametric Regression. Annals of Statistics 1982;10:1348–1360.

Tan, WY.; Wu, H. Deterministic and Stochastic Models of AIDS Epidemics and HIV Infections with Intervention. Singapore: World Scientific; 2005.

Tunali T, Tarn TJ. New Results for Identifiability of Nonlinear Systems. IEEE Transactions on Automatic Control 1987;32:146–154.

Varah JM. A Spline Least Squares Method for Numerical Parameter Estimation in Differential Equations. SIAM Journal on Scientific Computing 1982;3:131–141.

Wu CF. Asymptotic Theory of Nonlinear Least Squares Estimation. Annals of Statistics 1981;9:501–513.

Wu H. Statistical Methods for HIV Dynamic Studies in AIDS Clinical Trials. Statistical Methods in Medical Research 2005;14:171–192. [PubMed: 15807150]

Wu H, Ding AA. Population HIV-1 Dynamics in Vivo: Applicable Models And Inferential Tools for Virological Data from AIDS Clinical Trials. Biometrics 1999;55:410–418. [PubMed: 11318194]

Wu H, Ding AA, DeGruttola V. Estimation of HIV Dynamic Parameters. Statistics in Medicine 1998;17:2463–2485. [PubMed: 9819839]

Wu H, Kuritzkes DR, McClernon DR, Kessler H, Connick E, Landay A, Spear G, Heath-Chiozzi M, Rousseau F, Fox L, Spritzler J, Leonard JM, Lederman MM. Characterization of Viral Dynamics in Human Immunodeficiency Virus Type 1-Infected Patients Treated With Combination Antiretroviral Therapy: Relationships to Host Factors, Cellular Restoration And Virological Endpoints. Journal of Infectious Diseases 1999;179:799–807. [PubMed: 10068574]

Wu H, Zhu H, Miao H, Perelson AS. Parameter Identifiability and Estimation of HIV/AIDS Dynamic Models. Bulletin of Mathematical Biology 2008;70:785–799. [PubMed: 18247096]

Xia X, Moog CH. Identifiability of Nonlinear Systems With Applications to HIV/AIDS Models. IEEE Transactions on Automatic Control 2003;48:330–336.

Appendix: Proofs

Before we prove Theorems 1 and 2, we state the following lemma for the proofs of the main results.

Lemma 1

Under Assumptions A and C,

$$\sup_t |\widehat{X}'(t) - X'(t)| = O_p(b_n) \ \ and \ \ \sup_t |\widehat{X}(t) - X(t)| = O_p(c_n),$$

where $b_n = h^2 + n^{-1/2}h^{-3/2} \log n$ and $c_n = h^2 + n^{-1/2}h^{-1/2} \log n$.

Proof

The proof of this lemma is similar to that in Mack and Silverman (1982). See Stone (1982) for a detailed discussion on uniform convergence rates for nonparametric estimation.

From Lemma 1, we have

$$\max_{1 \le i \le n} |\Delta(t_i)| = O_p(b_n). \tag{A.1}$$

Proof of Theorem 1—The key step of the proof of the consistency is to show that $\boldsymbol{\beta}_0$, the true value of the m–dimensional parameter vector $\boldsymbol{\beta}$, uniquely minimizes $\lim_{n\to\infty} S_n(\boldsymbol{\beta})$. Note that

$$\begin{aligned}
&\sum_{i=1}^{n}\left[\widehat{X}'(t_i) - F\{\widehat{X}(t_i),\beta\}\right]^2 = \sum_{i=1}^{n}\left[F\{X(t_i),\beta_0\} + \Delta(t_i) - F\{\widehat{X}(t_i),\beta\}\right]^2 \\
&= \sum_{i=1}^{n}\left[F\{X(t_i),\beta_0\} - F\{\widehat{X}(t_i),\beta\}\right]^2 + \sum_{i=1}^{n}\Delta^2(t_i) \\
&+ 2\sum_{i=1}^{n}\left[F\{X(t_i),\beta_0\} - F\{\widehat{X}(t_i),\beta\}\right]\Delta(t_i).
\end{aligned} \tag{A.2}$$

The second term is order of (nb_n^2) from (A.1), while the order of the third term is lower than that of the first term based on the Cauchy-Schwarz inequality if $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. Now we consider the first term which can be decomposed as

$$\begin{aligned}
&\sum_{i=1}^{n}\left[F\{X(t_i),\beta_0\} - F\{\widehat{X}(t_i),\beta\}\right]^2 = \\
&\sum_{i=1}^{n}[F\{X(t_i),\beta_0\} - F\{X(t_i),\beta\}]^2 \\
&+ \sum_{i=1}^{n}\left[F\{X(t_i),\beta\} - F\{\widehat{X}(t_i),\beta\}\right]^2 \\
&+ 2\sum_{i=1}^{n}[F\{X(t_i),\beta_0\} - F\{X(t_i),\beta\}]\left[F\{X(t_i),\beta\} - F\{\widehat{X}(t_i),\beta\}\right].
\end{aligned}$$

By Assumption C(ii) and Lemma 1, we know that the second term from above is bounded as follows:

$$\sum_{i=1}^{n}\left[F\{X(t_i),\beta\} - F\{\widehat{X}(t_i),\beta\}\right]^2 \le n\sup_t |X(t) - \widehat{X}(t)|^2 \sup_{x \in \chi}\left|\frac{\partial F(x,\beta)}{\partial x}\right|^2 \le nM_\beta^2 c_n^2. \tag{A.3}$$

In a similar argument, we know that, if, $\boldsymbol{\beta}_0 \neq \boldsymbol{\beta}$,

$$\sum_{i=1}^{n}[F\{X(t_i),\beta_0\}-F\{X(t_i),\beta\}]\left[F\{X(t_i),\beta\}-F\{\widehat{X}(t_i),\beta\}\right]=O(nc_n)=0(n). \tag{A.4}$$

The strong law of large number yields, if $\boldsymbol{\beta}_0 \neq \boldsymbol{\beta}$,

$$\sum_{i=1}^{n}[F\{X(t_i),\beta_0\}-F\{X(t_i),\beta\}]^2=nD(\beta_0,\beta)+o(n). \tag{A.5}$$

Combining (A.3)–(A.5), we can see that the first term of (A.2) is dominated by the term

$$\sum_{i=1}^{n}[F\{X(t_i),\beta_0\}-F\{X(t_i),\beta\}]^2,$$

which has a unique minimum at $\boldsymbol{\beta}_0$ by Assumption B (iii) when n is large enough. Therefore the PsLS estimator defined in (3.6) is strongly consistent.

Note that the results (A.3)–(A.5) in the above proof utilized the asymptotic properties of the local linear estimators which are critical for establishing the consistency of the proposed PsLS estimator. More discussions on the assumptions of NLS estimators and the proofs can be found in Seber and Wild (1989) or Bates and Watts (1988).

Proof of Theorem 2—Note that, under the assumptions that continuous derivatives exist and using the mean-value theorem, we have

$$\mathbf{0}=\frac{\partial S_n(\widehat{\beta}_n)}{\partial\beta}=\frac{\partial S_n(\beta_0)}{\partial\beta}+\frac{\partial^2 S_n(\beta_n^*)}{\partial\beta\partial\beta^{\mathrm{T}}}(\widehat{\beta}_n-\beta_0),$$

where $\frac{\partial S_n(\tilde{\beta})}{\partial\beta}$ represents $\frac{\partial S_n(\beta)}{\partial\beta}$ evaluated at $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$, and β_n^* lies between $\hat{\boldsymbol{\beta}}_{\mathrm{n}}$ and $\boldsymbol{\beta}_0$. Then we have

$$\widehat{\beta}_n-\beta_0=-\left\{\frac{\partial^2 S_n(\beta_n^*)}{\partial\beta\partial\beta^{\mathrm{T}}}\right\}^{-1}\frac{\partial S_n(\beta_0)}{\partial\beta}. \tag{A.6}$$

We first study the derivative $\frac{\partial S_n(\beta)}{\partial\beta}$, which can be expressed as:

NIH-PA Author Manuscript
NIH-PA Author Manuscript

$$\frac{\partial S_n(\beta)}{\partial \beta} = -2\sum_{i=1}^{n}\left[F\{X(t_i),\beta\} - F\{\widehat{X}(t_i),\beta\}\right]\frac{\partial F\{\widehat{X}(t_i),\beta\}}{\partial \beta} - 2\sum_{i=1}^{n}\Delta(t_i)\frac{\partial F\{\widehat{X}(t_i),\beta\}}{\partial \beta}$$
$$\triangleq I_{1n}+I_{2n}.$$

By Assumption C(ii) and Lemma 1, we obtain that

$$|I_{1n}| \le M_{\beta_0} c_n n.$$

Note that I_{2n} can be expressed as

$$-2\sum_{i=1}^{n}\left[\frac{\partial F\{\widehat{X}(t_i),\beta\}}{\partial \beta} - \frac{\partial F\{X(t_i),\beta\}}{\partial \beta}\right]\Delta(t_i) - 2\sum_{i=1}^{n}\frac{\partial F\{X(t_i),\beta\}}{\partial \beta}\Delta(t_i). \tag{A.7}$$

The first term is bounded by

$$nC_1\sup_t|\widehat{X}(t)=X(t)|^{\zeta}\sup_t|\Delta(t)|=O(nb_nc_n^{\zeta}).$$

Write $F'_{j,\beta}=\frac{\partial F\{X(t_j),\beta\}}{\partial \beta}$ for $j = 1 \sim n$, and $\mathbf{F}_\beta=(F'_{1,\beta},\ldots,F'_{n,\beta})^{\mathbf{T}}$. The second summand of (A.7) can be expressed as

$$\mathbf{F}_\beta^{\mathbf{T}}\{\Delta(t_1),\ldots,\Delta(t_n)\}^{\mathbf{T}}.$$

Using the notation in Section 2.1, let $\xi_{2,t}=\xi_2^{\mathbf{T}}(\mathbf{T}_{2,t}^{\mathbf{T}}\mathbf{W}_t\mathbf{T}_{2,t})^{-1}\mathbf{T}_{2,t}^{\mathbf{T}}\mathbf{W}_t$ and $\boldsymbol{\Xi}_2=(\xi_{2,t_1}^{\mathbf{T}},\ldots,\xi_{2,t_n}^{\mathbf{T}})^{\mathbf{T}}$, then we have

$$\{\Delta(t_1),\ldots,\Delta(t_n)\}^{\mathbf{T}}=\boldsymbol{\Xi}_2 Y - X'(\mathbf{t})=\boldsymbol{\Xi}_2X(\mathbf{t}) - X'(\mathbf{t})+\boldsymbol{\Xi}_2\mathbf{e}.$$

Recall the expression of bias given in (2.4) for $X'(t)$. A direct calculation yields that

$$\mathbf{F}_\beta^{\mathbf{T}}\{\boldsymbol{\Xi}_2X(\mathbf{t}) - X'(\mathbf{t})\}=Q_1nh^2+o(nh^2), \tag{A.8}$$

where Q_1 is a constant independent of n.

Furthermore, $2\mathbf{F}_\beta^{\mathbf{T}}\boldsymbol{\Xi}_2\mathbf{e}$ is a sum of weighted independent variables $\{e_i, i = 1,\ldots, n\}$ with mean zero and covariance matrix of the form:

$$4\mathbf{F}_\beta^{\mathrm{T}}\mathbf{\Xi}_2\mathrm{cov}(\mathbf{e})\mathbf{\Xi}_2^{\mathrm{T}}\mathbf{F}_\beta.$$

Note that the (i, j)th entry of $\mathbf{\Xi}_2\mathbf{\Xi}_2^{\mathrm{T}}$, denoted by ζ_{ij}, can be expressed as

$$\xi_2^{\mathrm{T}}(\mathbf{T}_{2,t_i}^{\mathrm{T}}\mathbf{W}_{t_i}\mathbf{T}_{2,t_i})^{-1}\mathbf{T}_{2,t_i}^{\mathrm{T}}\mathbf{W}_{t_i}\mathbf{W}_{t_j}\mathbf{T}_{2,t_j}(\mathbf{T}_{2,t_j}^{\mathrm{T}}\mathbf{W}_{t_j}\mathbf{T}_{2,t_j})^{-1}\xi_2.$$

By direct calculations similar to deriving the bias and variance of $\hat{X}(t)$, we have that

$$n^{-1}\mathbf{T}_{2,t}^{\mathrm{T}}\mathbf{W}_t\mathbf{T}_{2,t}=A_3\{f(t)N_3+hf'(t)Q_3\}A_3+o_p(hA_3\mathbf{1}A_3),$$
$$n^{-1}\mathbf{T}_{2,t}^{\mathrm{T}}\mathbf{W}_t^2\mathbf{T}_{2,t}=h^{-1}f(t)A_3S_3A_3+o_p(h^{-1}A_3\mathbf{1}A_3),$$

where $A_3 = \mathrm{diag}(1, h, h^2)$, N_3, Q_3, S_3 and $\mathbf{1}$ are all 3×3 matrices whose (i, j) entry are $\mu_{i+j-2}(K)$, $\mu_{i+j-1}(K)$, $\mu_{i+j-2}(K^2)$, and 1, respectively. Then

$$\zeta_2^{\mathrm{T}}(n^{-1}\mathbf{T}_{2,t}^{\mathrm{T}}\mathbf{W}_t\mathbf{T}_{2,t})^{-1}=f^{-1}(t)h^{-1}\zeta_2^{\mathrm{T}}\{N_2^{-1}-hf'(t)/f(t)N_2^{-1}Q_2N_2^{-1}\}A_2^{-1}+o_p(hA_2^{-1}).$$

A simplification yields that

$$\zeta_{ij}=\begin{cases} n^{-1}h^{-3}\mu_2^{-2}(K)\mu_2(K^2)+o(n^{-1}h^{-3}) & \text{if } i=j\\ o(n^{-1}h^{-3}) & \text{if } i\neq j.\end{cases}$$

As a result,

$$\mathbf{F}_\beta^{\mathrm{T}}\mathbf{\Xi}_2\mathrm{cov}(\mathbf{e})\mathbf{\Xi}_2^{\mathrm{T}}\mathbf{F}_\beta=\sigma_e^2\mu_2^{-2}(K)\mu_2(K^2)h^{-3}E\{\frac{\partial F(X,\beta)}{f(t)\,\partial\beta}\}^{\otimes 2}+o(h^{-3}), \tag{A.9}$$

where $A^{\otimes 2} = AA^{\mathrm{T}}$. On the other hand, we have

$$\frac{1}{n}\frac{\partial^2 S_n(\beta_n^*)}{\partial\beta\partial\beta^{\mathrm{T}}}=-\frac{2}{n}\sum_{i=1}^n[\widehat{X}'(t_i)-F\{\widehat{X}(t_i),\beta_n^*\}]\frac{\partial^2F\{\widehat{X}(t_i),\beta_n^*\}}{\partial\beta\partial\beta^{\mathrm{T}}}$$
$$+\frac{2}{n}\sum_{i=1}^n\frac{\partial F\{\widehat{X}(t_i),\beta_n^*\}}{\partial\beta}\left[\frac{\partial F\{\widehat{X}(t_i),\beta_n^*\}}{\partial\beta}\right]^{\mathrm{T}}.$$

Using an argument similar to (A.2) and Assumption C, we know that the first term of $\frac{1}{n}\frac{\partial^2 S_n(\beta_n^*)}{\partial\beta\partial\beta^{\mathrm{T}}}$ is $o(1)$, while the second term converges to $2E\left[\left\{\frac{\partial F(X,\beta)}{\partial\beta}\right\}^{\otimes 2}\right]$. Combining (A.6)–(A.9) and recalling Assumption A(iii) on the bandwidth h, we may apply the Lindeberg central limit theorem and obtain

$$nh^{3/2}(\widehat{\beta}_n - \beta_0) \rightarrow \text{Normal}(\mu_\beta, \sum\nolimits_\beta)$$

in distribution, where $\mu_\beta = \lim_{n\to\infty} Q_1 nh^2 nh^{3/2} n^{-1} = 0$ from (A.8), and

$$\sum\nolimits_\beta = \sigma_e^2 \mu_2^{-2}(K)\mu_2(K^2)\left[E\left\{\frac{\partial F(X,\beta)}{\partial\beta}\right\}^{\otimes 2}\right]^{-1} E\left\{\frac{\partial F(X,\beta)}{f(t)\ \partial\beta}\right\}^{\otimes 2}\left[E\left\{\frac{\partial F(X,\beta)}{\partial\beta}\right\}^{\otimes 2}\right]^{-1} \quad \text{(A.10)}$$

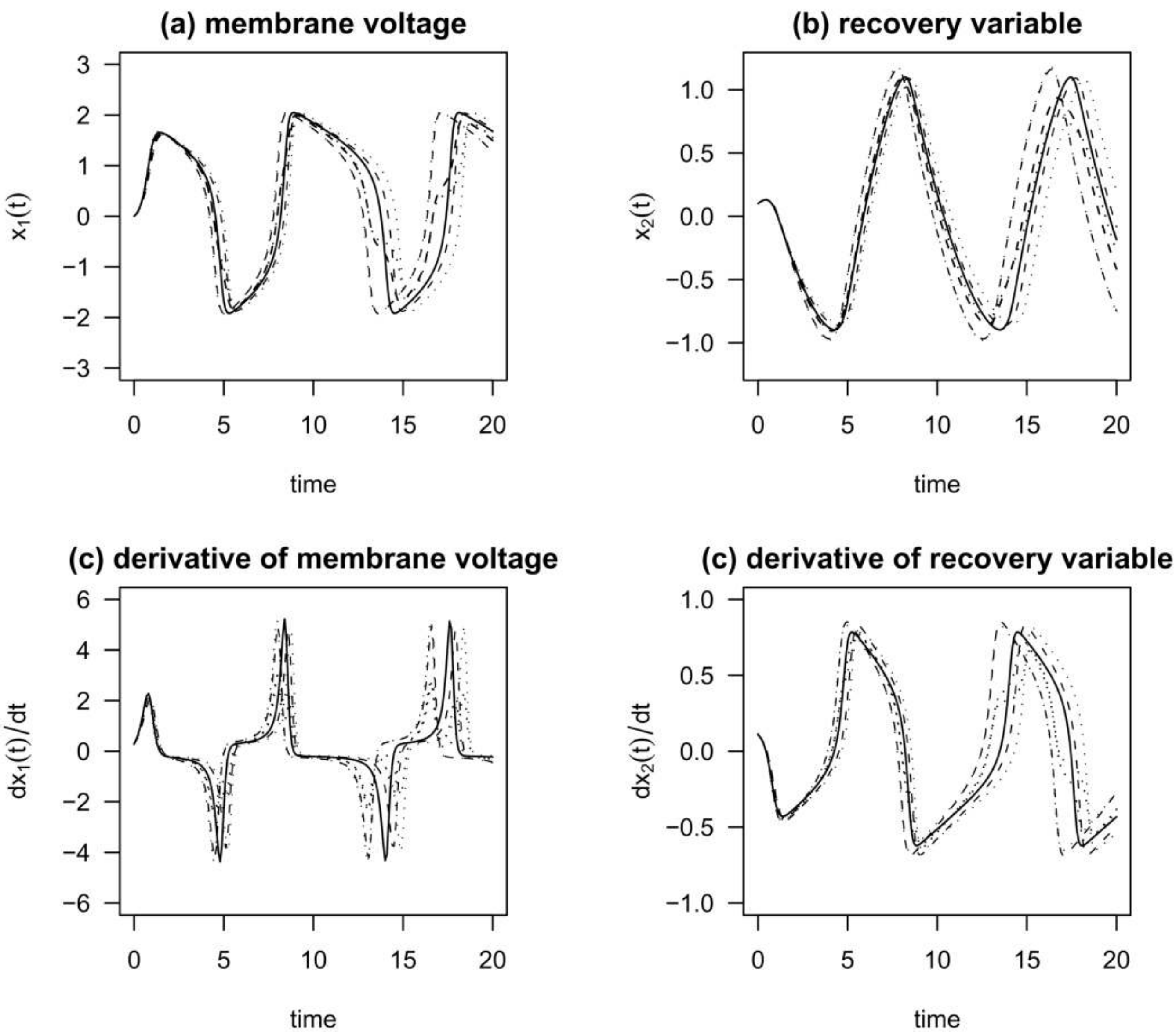

Figure 1.

The trajectories of state variables $X_1(t)$ and $X_2(t)$ with their derivatives for the simulated data from example 1. The solid lines indicate the true curves, and the dashed and dotted lines indicate the average fitted curves and the associated 95% pointwise confidence intervals obtained by the PsLS and SIMEX estimation procedures respectively.

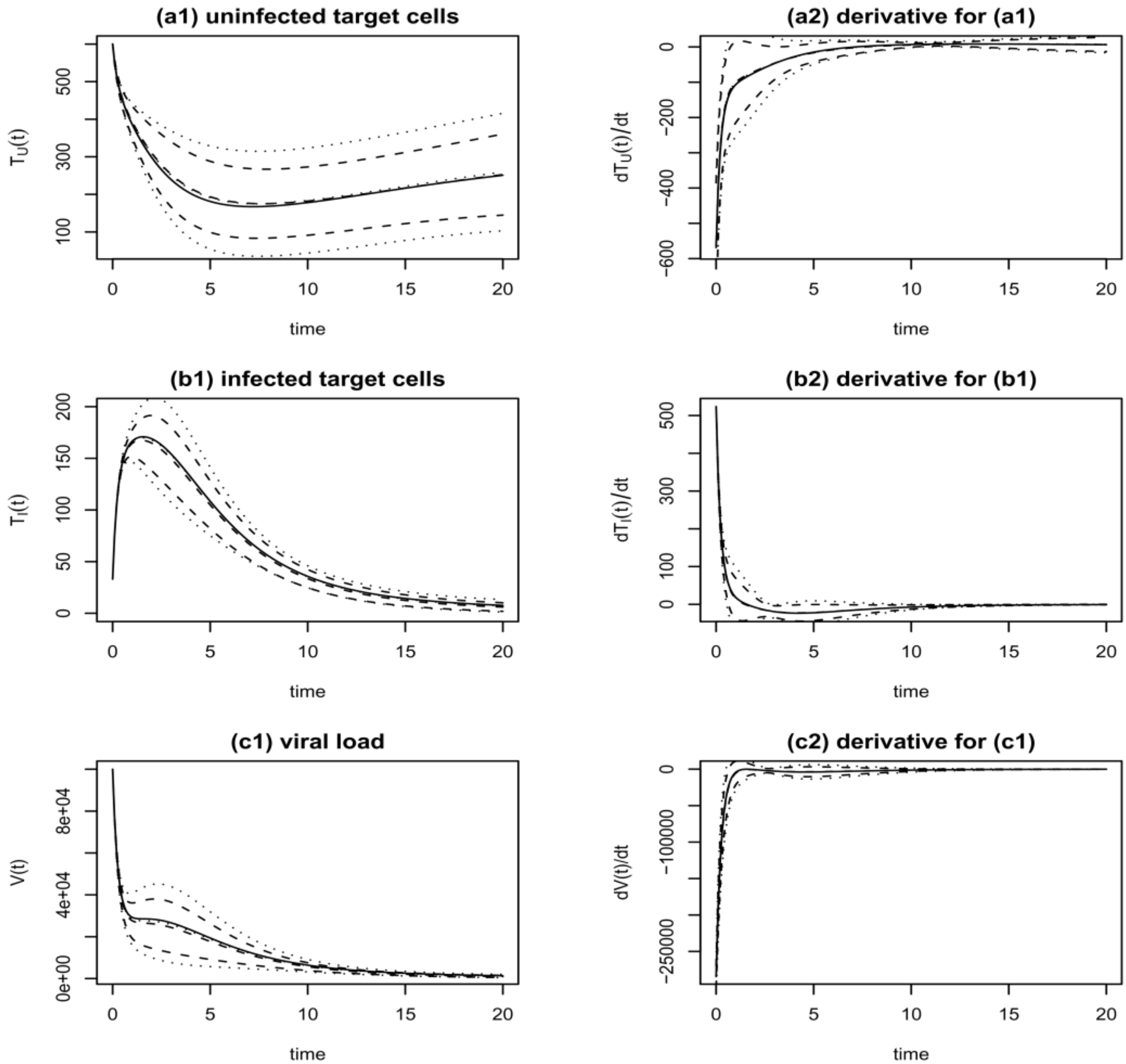

Figure 2.

The trajectories of state variables $T_U(t)$, $T_I(t)$ and $V(t)$ with their derivatives for the simulated data from example 2. The solid lines indicate the true curves, and the dashed and dotted lines indicate the average fitted curves and the associated 95% pointwise confidence intervals obtained by the PsLS and SIMEX estimation procedures respectively.

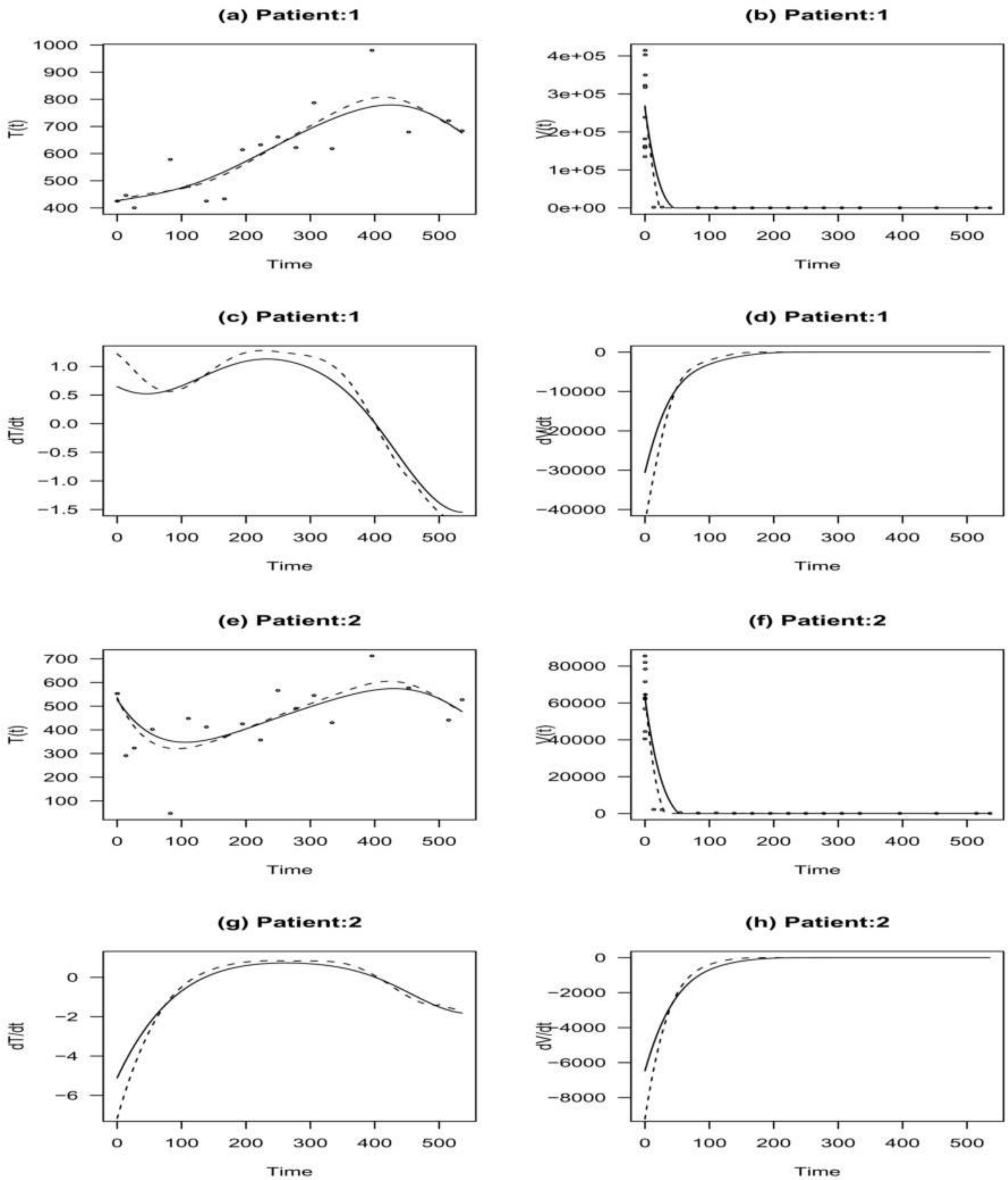

Figure 3.

The fitted curves of $T(t)$ and $V(t)$ with their derivatives for two patients from an HIV dynamics study. Dots indicate the observations. The solid and dotted lines are the fitted curves by the PsLS and SIMEX methods.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

The averages of the estimated values, standard errors (s.e.), and 95% coverage probability (CP) of the estimators of the parameters (α = .34, β = 0.2, and γ = 3) in Example 1.

		PsLS(s.e., CP)			SIMEX (s.e., CP)		
σ_1^2	σ_2^2	α	β	γ	α	β	γ
0.05	0.05	0.330(0.08,93.4)	0.223(0.12,94.0)	3.076(0.17,93.0)	0.341(0.10,94.9)	0.198(0.13,93.8)	2.925(0.25,92.8)
	0.06	0.336(0.08,97.3)	0.216(0.12,92.0)	3.069(0.17,96.2)	0.343(0.10,96.2)	0.205(0.13,94.6)	2.942(0.26,96.7)
	0.07	0.335(0.09,94.3)	0.213(0.12,97.0)	3.079(0.18,94.5)	0.342(0.12,95.7)	0.191(0.15,94.6)	2.909(0.23,94.3)
	0.08	0.333(0.10,94.0)	0.214(0.13,92.0)	3.086(0.18,94.1)	0.337(0.11,97.3)	0.209(0.15,94.5)	2.886(0.22,93.8)
	0.09	0.344(0.10,98.0)	0.216(0.14,92.0)	3.089(0.18,93.8)	0.342(0.11,96.1)	0.206(0.18,97.3)	2.860(0.20,94.1)
	0.1	0.341(0.10,98.0)	0.220(0.12,94.0)	3.081(0.18,97.3)	0.340(0.12,97.9)	0.210(0.14,96.4)	2.825(0.19,96.8)
0.06	0.05	0.336(0.08,98.1)	0.221(0.10,95.0)	3.038(0.18,93.3)	0.338(0.10,95.8)	0.199(0.12,93.2)	2.887(0.27,95.6)
	0.06	0.338(0.08,96.0)	0.216(0.13,91.0)	3.038(0.17,91.4)	0.339(0.11,95.8)	0.202(0.13,96.2)	2.850(0.24,93.9)
	0.07	0.338(0.09,97.1)	0.217(0.11,97.0)	3.040(0.18,92.1)	0.339(0.10,94.8)	0.212(0.13,93.9)	2.817(0.24,92.3)
	0.08	0.341(0.08,94.9)	0.218(0.14,92.0)	3.037(0.18,94.6)	0.345(0.11,97.2)	0.211(0.16,96.5)	2.898(0.22,94.2)
	0.09	0.337(0.10,96.7)	0.211(0.12,94.2)	3.037(0.19,97.4)	0.341(0.10,95.2)	0.211(0.14,93.8)	2.789(0.22,97.3)
	0.1	0.338(0.11,98.0)	0.209(0.16,92.0)	3.051(0.20,95.3)	0.352(0.13,98.0)	0.192(0.17,96.0)	2.866(0.19,94.9)
0.07	0.05	0.331(0.07,97.2)	0.219(0.10,94.0)	3.017(0.19,93.7)	0.345(0.09,95.9)	0.203(0.12,94.4)	2.780(0.25,92.9)
	0.06	0.334(0.08,97.0)	0.223(0.11,92.0)	3.022(0.17,94.3)	0.342(0.10,96.5)	0.207(0.13,93.7)	2.796(0.24,93.2)
	0.07	0.336(0.10,95.0)	0.219(0.11,93.2)	3.018(0.18,94.3)	0.345(0.10,93.5)	0.215(0.13,93.6)	2.865(0.20,93.9)
	0.08	0.337(0.11,95.0)	0.212(0.11,92.0)	3.018(0.20,95.8)	0.341(0.10,97.0)	0.201(0.13,94.3)	2.829(0.20,94.3)
	0.09	0.336(0.11,95.0)	0.220(0.15,94.0)	3.023(0.19,96.1)	0.338(0.10,96.1)	0.213(0.17,94.6)	2.720(0.22,97.2)
	0.1	0.340(0.11,92.0)	0.218(0.14,93.0)	3.008(0.19,94.8)	0.343(0.12,95.0)	0.201(0.18,97.0)	2.893(0.21,94.3)
0.08	0.05	0.335(0.08,96.0)	0.221(0.10,92.0)	2.984(0.20,94.2)	0.345(0.11,96.3)	0.200(0.13,95.3)	2.714(0.25,93.8)
	0.06	0.336(0.09,94.0)	0.217(0.12,95.0)	2.990(0.18,93.6)	0.343(0.11,95.4)	0.200(0.15,93.2)	2.726(0.24,93.0)
	0.07	0.337(0.09,98.0)	0.213(0.10,94.5)	2.989(0.20,93.6)	0.346(0.10,92.9)	0.195(0.14,94.2)	2.805(0.20,93.5)
	0.08	0.337(0.10,92.0)	0.215(0.12,96.0)	2.975(0.19,93.4)	0.345(0.11,94.9)	0.205(0.15,96.3)	2.854(0.20,92.9)
	0.09	0.339(0.10,94.0)	0.211(0.14,93.0)	2.991(0.19,93.8)	0.347(0.10,96.2)	0.201(0.16,95.8)	2.865(0.19,92.9)
	0.1	0.347(0.10,93.0)	0.212(0.15,92.0)	2.994(0.22,92.3)	0.343(0.11,97.0)	0.210(0.16,93.0)	2.916(0.21,92.1)
0.09	0.05	0.331(0.08,95.8)	0.227(0.12,93.0)	2.976(0.17,92.9)	0.342(0.10,97.1)	0.211(0.13,95.7)	2.870(0.26,93.0)
	0.06	0.336(0.08,96.0)	0.226(0.10,92.0)	2.961(0.21,93.2)	0.342(0.09,96.4)	0.223(0.14,92.9)	2.637(0.22,96.5)
	0.07	0.338(0.10,97.0)	0.219(0.11,98.0)	2.963(0.20,92.9)	0.344(0.10,96.7)	0.209(0.12,93.6)	2.834(0.21,93.0)
	0.08	0.342(0.10,97.0)	0.219(0.13,94.0)	2.983(0.21,95.8)	0.349(0.11,96.2)	0.212(0.14,94.8)	2.840(0.22,96.7)
	0.09	0.338(0.08,96.0)	0.217(0.13,93.2)	2.966(0.21,96.7)	0.343(0.10,98.0)	0.203(0.18,96.1)	2.908(0.22,93.4)
	0.1	0.342(0.10,90.0)	0.210(0.15,94.0)	2.966(0.20,93.2)	0.343(0.11,97.0)	0.209(0.17,95.0)	2.883(0.21,93.6)
0.1	0.05	0.334(0.07,98.2)	0.220(0.11,93.1)	2.951(0.20,93.7)	0.342(0.09,96.8)	0.199(0.13,94.3)	2.812(0.23,93.2)
	0.06	0.338(0.08,96.9)	0.214(0.10,96.0)	2.946(0.20,95.7)	0.348(0.08,96.3)	0.204(0.14,94.3)	2.597(0.22,93.8)
	0.07	0.338(0.08,94.0)	0.221(0.11,92.1)	2.948(0.22,96.3)	0.346(0.10,93.6)	0.215(0.15,96.2)	2.885(0.26,94.7)
	0.08	0.341(0.09,97.0)	0.212(0.13,93.0)	2.936(0.20,93.7)	0.349(0.10,95.9)	0.205(0.16,94.3)	2.859(0.22,93.9)
	0.09	0.344(0.11,93.0)	0.209(0.14,92.1)	2.954(0.22,97.5)	0.349(0.11,96.0)	0.202(0.15,97.5)	2.867(0.21,96.9)
	0.1	0.345(0.10,91.0)	0.211(0.12,92.3)	2.945(0.20,92.8)	0.347(0.12,97.0)	0.206(0.15,93.8)	2.842(0.23,92.9)

Table 2

Relative error for the simulated data from example 1

		PsLS			SIMEX	
σ_1^2	σ_2^2	α	β	γ	αβ	γ
0.05	0.05	6.21	17.765	16.329	9.8221.531	18.551
	0.06	7.27	17.355	15.829	9.8222.141	17.981
	0.07	7.21	20.625	15.659	12.1224.681	19.081
	0.08	7.17	26.955	14.529	11.3927.591	19.871
	0.09	7.27	30.595	14.159	11.1230.671	20.731
	0.1	7.72	24.415	14.079	11.1025.131	21.881
0.06	0.05	6.70	16.655	18.379	10.2220.481	19.831
	0.06	7.33	17.995	17.759	10.2722.711	21.061
	0.07	6.06	20.845	17.269	9.6622.141	22.161
	0.08	5.75	26.665	16.969	10.2928.901	22.791
	0.09	7.32	22.785	16.549	9.6024.711	23.111
	0.1	7.90	29.705	16.069	13.6729.251	23.851
0.07	0.05	6.44	14.615	19.219	10.0820.671	23.411
	0.06	7.70	18.715	18.649	10.5122.501	22.861
	0.07	7.95	17.295	18.589	10.2923.751	23.901
	0.08	6.66	19.365	18.079	10.1322.891	25.111
	0.09	8.18	27.565	17.629	9.3629.321	25.391
	0.1	8.09	29.935	18.139	11.9230.481	26.281
0.08	0.05	6.28	16.405	20.939	11.4821.931	25.611
	0.06	6.90	21.505	20.139	10.2925.911	25.201
	0.07	7.33	18.545	20.069	10.8524.501	25.891
	0.08	7.95	21.385	20.229	10.9225.691	27.581
	0.09	7.78	25.045	18.609	10.3726.671	27.231
	0.1	7.75	30.925	18.859	11.0227.101	28.851
0.09	0.05	7.31	17.755	21.769	9.9322.681	27.081
	0.06	7.22	21.755	21.479	9.0625.951	28.181
	0.07	7.38	15.435	21.179	10.1820.161	28.271
	0.08	7.38	22.845	20.299	11.7823.271	28.061
	0.09	7.04	28.695	20.329	10.1532.801	29.111
	0.1	8.45	29.775	20.389	11.1530.581	29.971
0.1	0.05	6.42	18.885	22.679	9.2923.711	28.981
	0.06	6.78	19.325	21.869	9.0622.991	29.511
	0.07	6.62	22.085	21.789	10.3026.821	29.911
	0.08	7.80	23.195	22.119	10.5227.421	30.781
	0.09	8.30	24.395	20.849	12.1927.591	30.511
	0.1	8.57	26.495	20.989	12.6927.041	31.341

Table 3

The averages of the estimated values, standard errors (s.e.), and 95% coverage probablity (CP) of the estimators of the parameters ($\lambda = 36$, $\rho = 0.108$, and $N = 1000$) in example 2 under the different configurations

			PsLS (s.e., CP)		SIMEX (s.e., CP)		
case	σ_1^2	σ_2^2	**λ**	**ρ**	**Nλ**	**ρ**	**N**
(i)	20	100	35.5(2.53, 96.5)	0.112(0.29, 94.0)	976.6(29.9, 96.5)35.5(7.49, 94.7)	0.106(0.47, 93.8)	995.9(37.8, 95.9)
		150	35.8(2.61, 96.0)	0.106(0.19, 94.0)	977.7(30.5, 94.0)36.0(4.45, 96.5)	0.108(0.57, 93.7)	927.2(46.7, 97.1)
		200	36.9(2.54, 93.5)	0.111(0.19, 94.0)	949.1(30.0, 98.5)36.9(2.61, 96.9)	0.113(0.49, 93.3)	997.3(33.7, 92.7)
	30	100	35.9(2.84, 93.5)	0.104(0.22, 92.5)	957.0(34.3, 94.0)35.1(7.25, 92.7)	0.102(0.62, 93.1)	972.4(44.4, 94.2)
		150	35.8(2.79, 93.4)	0.104(0.21, 93.5)	961.6(34.0, 92.5)35.4(5.97, 93.2)	0.105(0.59, 93.0)	936.0(45.3, 93.8)
		200	36.1(2.76, 95.1)	0.105(0.20, 96.0)	957.7(34.4, 98.0)35.7(6.96, 96.1)	0.105(0.57, 94.0)	974.4(46.8, 93.5)
	40	100	36.4(2.89, 95.1)	0.105(0.23, 93.8)	939.0(36.4, 93.8)34.5(4.54, 94.3)	0.098(0.48, 94.0)	935.6(48.9, 91.9)
		150	35.1(2.85, 94.0)	0.097(0.21, 97.5)	953.7(38.1, 94.5)33.3(4.26, 94.8)	0.090(0.38, 96.7)	938.3(50.4, 93.6)
		200	35.8(2.81, 97.1)	0.102(0.21, 92.9)	948.2(38.8, 97.2)34.8(6.41, 97.5)	0.106(0.60, 93.9)	949.6(50.7, 96.8)
(ii)	20	100	33.3(2.67, 98.0)	0.097(0.19, 95.5)	941.4(30.2, 96.5)33.5(5.62, 96.8)	0.114(0.44, 94.5)	1029.3(47.4, 93.8)
		150	33.8(2.66, 94.2)	0.100(0.19, 96.5)	951.5(33.1, 93.0)33.9(6.29, 93.8)	0.109(0.37, 96.8)	937.4(50.2, 93.5)
		200	33.1(2.63, 95.5)	0.101(0.19, 93.9)	945.1(30.7, 93.5)32.6(9.78, 95.9)	0.109(0.55, 93.6)	921.2(55.8, 92.9)
	30	100	32.1(2.70, 95.9)	0.091(0.19, 96.0)	1045.1(31.7, 92.5)34.1(4.30, 97.2)	0.107(0.28, 94.3)	948.5(58.2, 96.7)
		150	33.4(2.75, 93.5)	0.093(0.20, 95.8)	939.8(35.6, 93.9)33.3(8.86, 96.8)	0.107(0.64, 96.4)	1048.4(58.5, 95.8)
		200	31.5(2.73, 95.5)	0.092(0.20, 92.5)	946.1(37.6, 96.0)31.5(6.97, 96.1)	0.106(0.57, 92.4)	946.4(56.7, 96.8)
	40	100	33.6(3.11, 95.5)	0.083(0.22, 96.0)	962.1(39.9, 93.5)33.3(9.71, 96.2)	0.110(0.80, 96.3)	943.6(68.5, 96.2)
		150	32.4(2.94, 97.5)	0.083(0.20, 94.0)	944.6(39.1, 93.0)35.1(9.60, 97.8)	0.109(0.58, 94.7)	939.4(62.6, 93.5)
		200	31.1(2.26, 94.5)	0.090(0.22, 94.5)	935.6(41.9, 96.5)30.8(9.44, 96.7)	0.114(0.72, 94.8)	938.1(66.1, 97.3)

Table 4

Relative error for the simulated data from example 2

case	σ_1^2	σ_2^2	PsLS λρ	PsLS N	SIMEX λ	SIMEX ρ	SIMEX N
(i)	20	100	9.5716.03	2.45	10.38	17.35	2.76
		150	10.2218.14	2.36	10.50	18.65	2.87
		200	10.1917.49	2.15	10.87	18.77	2.71
	30	100	12.9721.62	3.86	12.54	20.35	4.26
		150	12.1721.41	3.39	12.05	20.51	3.74
		200	13.2422.34	3.54	12.08	20.05	4.09
	40	100	13.8224.37	4.52	12.00	19.40	5.22
		150	15.7727.11	5.04	14.31	24.10	6.30
		200	15.6326.26	4.98	13.06	21.10	6.06
(ii)	20	100	12.6920.48	3.74	15.16	22.95	4.21
		150	11.7620.91	3.44	13.28	23.66	3.84
		200	12.7322.34	3.57	15.02	26.38	3.92
	30	100	15.7523.83	5.78	14.66	23.24	5.73
		150	15.7528.70	5.68	15.76	27.94	5.43
		200	16.1628.68	5.97	16.36	28.48	5.61
	40	100	17.1536.69	8.92	19.47	37.97	8.08
		150	20.8038.98	8.61	20.20	41.86	8.74
		200	23.6744.74	8.27	25.85	52.17	8.76