

Numerical Discretization-Based Estimation Methods for Ordinary Differential Equation Models via Penalized Spline Smoothing with Applications in Biomedical Research

Hulin Wu,^{1,*} Hongqi Xue,¹ and Arun Kumar²

¹Department of Biostatistics and Computational Biology, University of Rochester, 601 Elmwood Avenue, Box 630, Rochester, New York 14642, U.S.A.

²Indian Institute of Management Indore, Indore 453331, India

*email: hwu@bst.rochester.edu

SUMMARY. Differential equations are extensively used for modeling dynamics of physical processes in many scientific fields such as engineering, physics, and biomedical sciences. Parameter estimation of differential equation models is a challenging problem because of high computational cost and high-dimensional parameter space. In this article, we propose a novel class of methods for estimating parameters in ordinary differential equation (ODE) models, which is motivated by HIV dynamics modeling. The new methods exploit the form of numerical discretization algorithms for an ODE solver to formulate estimating equations. First, a penalized-spline approach is employed to estimate the state variables and the estimated state variables are then plugged in a discretization formula of an ODE solver to obtain the ODE parameter estimates via a regression approach. We consider three different order of discretization methods, Euler's method, trapezoidal rule, and Runge–Kutta method. A higher-order numerical algorithm reduces numerical error in the approximation of the derivative, which produces a more accurate estimate, but its computational cost is higher. To balance the computational cost and estimation accuracy, we demonstrate, via simulation studies, that the trapezoidal discretization-based estimate is the best and is recommended for practical use. The asymptotic properties for the proposed numerical discretization-based estimators are established. Comparisons between the proposed methods and existing methods show a clear benefit of the proposed methods in regards to the trade-off between computational cost and estimation accuracy. We apply the proposed methods to an HIV study to further illustrate the usefulness of the proposed approaches.

KEY WORDS: Euler's method; HIV dynamics; Ordinary differential equation; Penalized splines; Runge–Kutta method; Trapezoidal rule; Two-stage smoothing method.

1. Introduction

Differential equations are very popular for modeling dynamics of a physical process in many fields such as engineering, physics, and biomedical sciences. Parameter estimation of differential equation models using the observed data is an important step in characterizing the physical process under investigation. In this article, we propose a new class of methods for estimating parameters in a nonlinear ordinary differential equation (ODE) model. A general nonlinear ODE model can be written as

$$\frac{d\mathbf{X}(t)}{dt} = \mathbf{f}\{t, \mathbf{X}(t), \mathbf{Y}(t), \boldsymbol{\beta}\}, \quad (1)$$

where $t \in [a, b]$ ($-\infty < a < b < +\infty$) is time (independent variable), $\mathbf{X}(t) = \{X_1(t), \dots, X_m(t)\}^T$ is an m -dimensional state variable vector, $\mathbf{Y}(t)$ is an input variable vector, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^T$ is a d -dimensional vector of unknown constant parameters with true value $\boldsymbol{\beta}_0$, and $\mathbf{f}(\cdot) = \{f_1(\cdot), \dots, f_m(\cdot)\}^T$ is a vector of known differentiable functions. We assume that $\mathbf{X}(t)$ is measured with noise at time points t_1, \dots, t_n and the measurement model can be written as

$$\mathbf{Y}_i = \mathbf{X}(t_i) + \mathbf{e}_i, \quad (2)$$

where the measurement errors $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ are independent with mean zero and a variance–covariance matrix $\boldsymbol{\Sigma}$. For presentation and notation simplicity, we mainly focus on the case of one equation (i.e., $m = 1$) without input variable \mathbf{Y} in the following methodology development and assume the variance-covariance matrix of the measurement errors to be $\sigma^2 I_n$, where I_n is an identity matrix. All of the methodology and computation algorithms developed in this article can be extended to the multidimensional case ($m \geq 2$) with input variables \mathbf{Y} . Our goal is to estimate $\boldsymbol{\beta}$ given the observed values $\{t_i, Y_i\}$.

The nonlinear least squares (NLS) method was the earliest and the most popular method developed for estimating the parameters of the ODE model (Hemker, 1972; Li, Osborne, and Pravan, 2005). Equation (1) usually does not have closed-form solutions for a general nonlinear function \mathbf{f} . In this case, numerical methods such as the Runge-Kutta algorithm (see Hairer, Norsett, and Wanner, 1993; Mattheij and Molenaar, 2002) have to be used to approximate the solution of the ODEs. Xue, Miao, and Wu (2010) studied the properties of the NLS estimator when both numerical error and measurement error are considered, and they showed that the NLS estimator is strongly consistent and asymptotically normal

with the same asymptotic covariance as that of the case with the true ODE solution exactly known. However, the NLS method suffers from some problems such as high computational cost, instability and convergence problems because numerical algorithms are often needed to solve the ODE repeatedly to obtain the NLS estimates. In addition, it also requires the initial or boundary conditions of the state variables to solve the ODE numerically. These drawbacks of the NLS method have motivated investigators to develop alternative estimation methods to avoid numerically solving the differential equations directly.

Varah (1982) proposed a two-stage smoothing method to estimate the ODE parameters. In the first stage, regression splines were used to obtain the estimates for both $X(t)$ and its first-order derivative $\hat{X}'(t) = dX(t)/dt$, say $\hat{X}(t)$ and $\hat{X}'(t)$, from the observed data, and in the second stage, parameters were estimated by minimizing the following objective function:

$$\hat{\beta}_n = \arg \min_{\beta} \sum_{i=1}^n [\hat{X}'(t_i) - f(t_i, \hat{X}(t_i), \beta)]^2. \quad (3)$$

Liang and Wu (2008) studied the asymptotic properties of this two-stage smoothing estimator by applying a local polynomial approach in the first stage. Brunel (2008) also studied the asymptotic properties for a similar two-stage smoothing estimator using regression splines in the first stage, while the second stage used an integral objective function, instead of the summation form (3). Chen and Wu (2008a, 2008b) established the theoretical results for such a method for a special ODE model with time-varying coefficients. This two-stage smoothing method avoids solving differential equations numerically when estimating ODE parameters. This method is computationally efficient and does not need the initial or boundary conditions of the state variables. However, there is a loss of some estimation accuracy and a large number of observations are required to attain a desired level of accuracy for the estimate of derivatives in the first stage. Ramsay (1996) proposed a principal differential analysis (PDA) method for the estimation of differential equation models. To obtain more accurate estimates, Poyton et al. (2006) and Varziri et al. (2008) further developed an iterative PDA algorithm. Ramsay et al. (2007) extended the PDA and developed a generalized profiling estimation procedure to estimate the parameters in ODE models, whose asymptotic properties were investigated by Qi and Zhao (2010).

In this article, we propose a novel class of estimation methods which are formed using the one-step discretization structure of the numerical algorithms for solving ODEs. We call such a new class of estimation methods as numerical discretization-based estimation (DBE) methods. We intend to improve the estimates from the two-stage smoothing method (Varah, 1982; Liang and Wu, 2008) while preserving its computational benefit.

The rest of the article is organized as follows. We introduce our new estimation methods in Section 2. The asymptotic properties of the proposed estimators are summarized in Section 3. In Section 4, we apply the proposed methods to the HIV dynamic data from HIV-1 infected patients treated by an antiretroviral therapy. Results from simulation studies are

presented to demonstrate the finite-sample behavior of the proposed methods in Section 5. In particular, we show that our new methods perform better than the general two-stage method (Varah, 1982; Liang and Wu, 2008) via our simulations. We conclude with some remarks in Section 6. The detailed technical proofs and assumptions as well as some additional simulation results are given in the Web Supplemental Materials.

2. Numerical Discretization-Based Estimation Methods

In this section, we introduce our new estimation methods. The proposed methods can be divided into two stages. In the first stage, we estimate the state variable as a smooth function of time t using P-splines. In the second stage, we use the discretization schemes of the ODE numerical algorithms to form a regression-like model to relate the smoothed-state variables to the unknown parameters β through the function f . Then a standard optimization technique is applied to the regression-like model to obtain the parameter estimates. This new framework of estimation methods allows us to efficiently use the well-designed ODE computational schemes, but not to implement the recursive scheme directly. Instead, a regression-like approach is applied to the “pseudo-data” generated from the estimated state variables from the first stage to obtain the final parameter estimates. We will demonstrate that the new methods can improve the estimation efficiency compared to similar methods. We present the two stages of our new estimation methods in the following two subsections. We describe the methodology for a one-dimensional ODE case for notation and presentation simplicity. Extension to the multi-dimensional case is straightforward but tedious.

2.1 Estimation of $X(t)$ and Its Derivatives Using P-Splines

In the first stage, we employ the penalized splines (P-splines) smoothing approach to estimate the curve $X(t)$. Although other smoothing approaches such as the local polynomial estimation, regression splines, and smoothing splines can also be used, we choose to use the P-splines here because of the following reasons: (1) its flexibility in fitting different types of curves and the compromise feature between regression and smoothing splines; (2) more computational efficiency compared to the local polynomial kernel smoothing; and (3) its convenience to implement using the R function, “smooth.spline.” Moreover, because the theoretical results for P-splines, in particular, for the estimate of derivatives are not established, it is very useful and has its own value to establish these theoretical results. The original idea of penalized splines smoothing was given in O’Sullivan (1986). The combination of B-splines and difference penalties, coined as P-splines, was introduced by Eilers and Marx (1996). Thereafter, the methodological investigation was pursued by several investigators (Yu and Ruppert, 2002; Ruppert, Wand, and Carroll, 2003; Wang and Ormerod, 2008), but only a few of them developed theoretical results for P-splines (Hall and Opsomer, 2005; Li and Ruppert, 2008; Claeskens, Krivobokova, and Opsomer, 2009; Kauermann, Krivobokova, and Fahrmeir, 2009).

We approximate $X(t)$ at time t by $X(t) \approx \sum_{j=-\nu}^K \delta_j N_{j,\nu+1}(t) = \mathbf{N}_{\nu+1}^T(t) \boldsymbol{\delta}$, where $\boldsymbol{\delta} = (\delta_{-\nu}, \dots, \delta_K)^T$ is the unknown coefficient vector to be estimated from the data,

and $\mathbf{N}_{\nu+1}(t) = \{N_{-\nu,\nu+1}(t), \dots, N_{K,\nu+1}(t)\}^T$ is the **B-spline basis function vector of degree ν** (order $\nu + 1$) and dimensional $1 \times (K + \nu + 1)$ at a sequence of knots $a = \tau_{-\nu} = \tau_{-\nu+1} = \dots = \tau_{-1} = \tau_0 < \tau_1 < \dots < \tau_K < \tau_{K+1} = \tau_{K+2} = \dots = \tau_{K+\nu+1} = b$ on the interval $[a, b]$ (see Schumaker, 1981). The estimation objective function contains a penalized sum of squared differences, i.e.,

$$L(\mathbf{X}) = \sum_{i=1}^n \{Y_i - \delta^T \mathbf{N}_{\nu+1}(t_i)\}^2 + \lambda \int_a^b [(\delta^T \mathbf{N}_{\nu+1}(t))^{(q)}]^2 dt, \quad (4)$$

with a penalized term by the integrated squared q th-order derivative of the spline function. It is common to assume $q \leq \nu$ and, in practice, q is often chosen to be 2, which is also used in our simulation studies and real data analysis. From de Boor (2001, Chapter X) and using similar notations as those in Claeskens, Krivobokova and Opsomer (2009), we know that the derivatives of spline functions can be expressed in terms of lower-order spline functions,

$$\left(\sum_{j=-\nu}^K \delta_j N_{j,\nu+1}(t) \right)^{(i)} = \sum_{j=-\nu+i}^K \delta_{j,i}^* N_{j,\nu+1-i}(t), \quad (5)$$

where the coefficients $\delta_{j,i}^*$ are defined, recursively, via

$$\delta_{j,i}^* = (\nu + 1 - i)(\delta_{j,i-1}^* - \delta_{j-1,i-1}^*) / (\tau_{j+\nu+1-i} - \tau_j), \quad (6)$$

for $i = 1, 2, \dots$ with $\delta_{j,0}^* = \delta_j$. Thus, we can rewrite the penalty term in equation (4) as $\lambda \delta^T \Delta_q^T \mathbf{R} \Delta_q \delta$, where matrix \mathbf{R} has elements $R_{ij} = \int_a^b N_{j,\nu+1-q}(t) N_{i,\nu+1-q}(t) dt$ for $i, j = -\nu, \dots, K$ and Δ_q denotes the matrix corresponding to the weighted difference operator defined in equation (6), which is just $\mathbf{D}^{(i)}$ defined by the expression (6) in Zhou and Wolfe (2000). Then $\delta_i^* \equiv (\delta_{-\nu+i,i}^*, \dots, \delta_{K,i}^*)^T = \Delta_i \delta$. Further, define the $n \times (K + \nu + 1)$ spline design matrix

$$\mathbf{Z} = \{\mathbf{N}_{\nu+1}(t_1), \dots, \mathbf{N}_{\nu+1}(t_n)\}^T,$$

and let $\mathbf{D}_q = \Delta_q^T \mathbf{R} \Delta_q$, then equation (4) can be written as

$$L(\mathbf{X}) = (\mathbf{Y} - \mathbf{Z}\delta)^T (\mathbf{Y} - \mathbf{Z}\delta) + \lambda \delta^T \mathbf{D}_q \delta. \quad (7)$$

The minimizer of equation (7) takes the form

$$\hat{\delta} = (\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{D}_q)^{-1} \mathbf{Z}^T \mathbf{Y}. \quad (8)$$

Then $\hat{\mathbf{X}}(t) = \mathbf{N}_{\nu+1}^T(t) \hat{\delta}$. Moreover, we can easily obtain the following expression for the estimate of i th-order derivative, $\hat{\mathbf{X}}^{(i)}(t)$, from equation (5) as

$$\hat{\mathbf{X}}^{(i)}(t) = \mathbf{N}_{\nu+1-i}^T(t) \Delta_i \hat{\delta}. \quad (9)$$

Denote $X^{(0)}(t) = X(t)$, $\hat{X}^{(0)}(t) = \hat{X}(t)$, and $\delta_0 = \mathbf{I}$. To determine the smoothing parameter λ , the generalized cross validation method can be used (Craven and Wahba, 1979).

The asymptotic bias and variance as well as the average mean squared error of the P-spline estimator $\hat{X}(t)$ have been obtained by Claeskens, Krivobokova, and Opsomer (2009). However, more general asymptotic results (including the asymptotic bias and variance, the mean squared error (MSE), and asymptotic normality) of the P-spline estimator and its derivatives are required to establish the asymptotic properties of the proposed numerical discretization-based

estimators. The details of the theoretical development are provided in Lemmas 1–4 in the Supplementary Materials.

2.2 ODE Parameter Estimation

In the second stage, by one-step discretization methods for ODEs (Hairer, Norsett, and Wanner, 1993; Mattheij and Molenaar, 2002), we have

$$\frac{X(t_{i+1}) - X(t_i)}{t_{i+1} - t_i} = F(t_i, X(t_i), X(t_{i+1}), \beta_0) + O(h^p), \quad (10)$$

$$i = 1, \dots, n-1,$$

where $h = \max_i h_i$ with $h_i = t_{i+1} - t_i$. Obviously, $h = O(n^{-1})$.

The form of $F(t_i, X(t_i), X(t_{i+1}), \beta)$ and the order p are determined by the discretization methods. In this article, we consider three different discretization methods, Euler's method, trapezoidal rule and Runge-Kutta method. For each of these methods, the form of $F(t_i, X(t_i), X(t_{i+1}), \beta)$ and order p are given as follows. For Euler's method, we have

$$F(t_i, X(t_i), X(t_{i+1}), \beta) = f(t_i, X(t_i), \beta) \quad (11)$$

with $p = 1$; the trapezoidal rule gives the form

$$F(t_i, X(t_i), X(t_{i+1}), \beta) = \frac{1}{2} [f(t_i, X(t_i), \beta) + f(t_{i+1}, X(t_{i+1}), \beta)] \quad (12)$$

with $p = 2$; for the fourth-order Runge-Kutta method, we have

$$F(t_i, X(t_i), X(t_{i+1}), \beta) = \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6}, \quad (13)$$

with $p = 4$, where $k_1 = f(t_i, X(t_i), \beta)$, $k_2 = f(t_i + h_i/2, X(t_i) + h_i k_1/2, \beta)$, $k_3 = f(t_i + h_i/2, X(t_i) + h_i k_2/2, \beta)$, $k_4 = f(t_i + h_i, X(t_i) + h_i k_3, \beta)$.

Replacing the state variable, $X(t_i)$, with its estimate, $\hat{X}(t_i)$ from the first stage, the expression (10) can be rewritten as

$$\frac{\hat{X}(t_{i+1}) - \hat{X}(t_i)}{t_{i+1} - t_i} = F(t_i, \hat{X}(t_i), \hat{X}(t_{i+1}), \beta_0) + \Gamma_i, \quad (14)$$

where Γ_i is the sum of discretization error and estimation error from the first stage. We propose a numerical discretization-based estimator $\hat{\beta}_n$ of β by minimizing the following weighted least squares criterion:

$$\mathbf{S}_n(\beta) = \sum_{i=1}^{n-1} w(t_i) \left[\frac{\hat{X}(t_{i+1}) - \hat{X}(t_i)}{t_{i+1} - t_i} - F(t_i, \hat{X}(t_i), \hat{X}(t_{i+1}), \beta) \right]^2, \quad (15)$$

where $w(t)$ is a prescribed nonnegative weight function on $[a, b]$. A more precise assumption for $w(\cdot)$ required for the asymptotic results is discussed in Section 3. For a multidimensional ODE model (1) with $m \geq 2$, we use the following objective function:

$$\mathbf{S}_n(\beta) = \sum_{i=1}^{n-1} \sum_{j=1}^m w_j(t_i) \left[\frac{\hat{X}_j(t_{i+1}) - \hat{X}_j(t_i)}{t_{i+1} - t_i} - F_j(t_i, \hat{X}_j(t_i), \hat{X}_j(t_{i+1}), \beta) \right]^2,$$

where $\hat{X}_j(t)$ is the estimator of $X_j(t)$ via the first stage using P-splines, $F_j(\cdot)$ is the function given by the associated discretization method, and $w_j(\cdot)$ is the weight function.

In general, a higher-order discretization method is expected to give a better estimation accuracy since its discretization error is smaller, but its computational cost is higher. The trade-off between the estimation accuracy and computational cost needs to be considered in practical applications. For convenience, we use the abbreviations EDB, TDB, and RDB for the Euler's discretization-based estimator, the trapezoidal discretization-based estimator, and the Runge-Kutta discretization-based estimator, respectively. Compared to the existing two-stage smoothing methods (Varah, 1982; Brunel, 2008; Liang and Wu, 2008), the proposed method avoids the estimation of the derivatives of the state variables, which may not be easy to do for sparse data. We study the asymptotic properties of the proposed estimators in Section 3 and evaluate the finite-sample performance of these estimators in Section 5.

3. Asymptotic Properties

In this section, we summarize the theoretical properties for the proposed numerical discretization-based estimators. Due to space limitations, we provide detailed regularity conditions and the technical proofs in the Supplementary Materials. First, we derived the asymptotic bias and variance of the P-spline estimators for the derivatives of the state variables and also establish their asymptotic normality. Then we apply such results to obtain the following consistency and asymptotic normality for the proposed estimators.

PROPOSITION 1: *Under Assumptions B1–B11 and B13 in the Supplementary Materials, if $K \sim cn^\varsigma$ with $\varsigma \geq 1/(2\nu + 3)$ and $\lambda = O(n^\gamma)$ with $\gamma \leq (\nu + 2 - q)/(2\nu + 3)$, then $\hat{\beta}_n$ is a strongly consistent estimator of β_0 .*

PROPOSITION 2: *Under Assumptions B1–B11 and B13 in the Supplementary Materials, (i) For $w(a) = w(b) = 0$, if $K \sim cn^\varsigma$ with $\varsigma \geq 1/(2\nu + 3)$ and $\lambda = O(n^\gamma)$ with $\gamma \leq (\nu + 2 - q)/(2\nu + 3)$, then $\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, \Sigma_1^*)$ with Σ_1^* given by (A.7) in the Supplementary Materials.*

(ii) For $w(a) \neq 0$ or $w(b) \neq 0$, if $K \sim cn^\varsigma$ with $\varsigma > 1/(2\nu + 3)$ and $\lambda = O(n^\gamma)$ with $\gamma \leq (\nu + 2 - q)/(2\nu + 3)$, then $\sqrt{n\kappa}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, \Sigma_2^)$ with Σ_2^* given by (A.8) in the Supplementary Materials and $\kappa = \max_{0 \leq j \leq K} (\tau_{j+1} - \tau_j)$.*

Remark 1. All of these theoretical results (Lemmas 1–5 and Propositions 1–2) can be extended to random designs if the random design points, t_1, \dots, t_n are independent and identically distributed (i.i.d.) with a cumulative distribution function $Q(t)$, where $Q(t)$ is a continuous distribution function with a positive and continuous density $\rho(t)$ on $[a, b]$. We only need to replace $E(\hat{X}^{(i)}(t))$, $\text{Var}(\hat{X}^{(i)}(t))$ and $o(\cdot)$ by $E(\hat{X}^{(i)}(t)|t_1, \dots, t_n)$, $\text{Var}(\hat{X}^{(i)}(t)|t_1, \dots, t_n)$ and $o_P(\cdot)$, respectively, and in this case $\hat{\beta}_n$ is a weakly consistent estimator of β_0 in Proposition 1.

Remark 2. From Lemma 5 in the Supplementary Materials, when h approaches to zero as n goes to infinity, we have

$F(t_i, X(t_i), X(t_{i+1}), \beta) \rightarrow f(t_i, X(t_i), \beta)$. Therefore, for large n , different numerical methods are expected to behave similarly, which is supported by our asymptotic results that the rates for consistency and the asymptotic variance for different numerical discretization-based methods are the same. Our numerical studies also demonstrate this trend. But for small sample sizes, our simulation results in the following section show that there is some difference among different discretization methods.

Remark 3. In Proposition 2, the estimator $\hat{\beta}_n$ has different convergence rates for different boundary conditions of the weight function. The standard root n rate can be achieved if the weight function is zero at both boundaries. Otherwise, the proposed estimators only achieve the nonparametric rate. Similar results were obtained by employing a similar strategy for handling the boundary problem in literature (Gasser and Müller, 1983; Bickel and Ritov, 1988; Huang and Fan, 1999; Brunel, 2008). Liang and Wu (2008) did not notice this phenomenon, which resulted in an error in their derivation of the asymptotic variance.

Remark 4. In Proposition 2(ii), to ensure that the asymptotic bias of $\hat{\beta}_n$ is zero, a faster rate of K compared to the optimal rate $n^{1/(2\nu+3)}$ is required. In this case, the choice of K gives an undersmoothed $\hat{X}(t)$. Then we cannot directly use the knots selected according to standard generalized cross validation (GCV) or other similar methods. Zhou, Shen, and Wolfe (1998) proposed an alternative two-step procedure.

4. Application: HIV Dynamics

The work presented in this article is motivated by a clinical data set obtained from HIV-1 infected patients who were treated with an antiretroviral therapy. This clinical study was conducted to monitor HIV dynamics for individual patients. Very frequent viral load measurements were collected from the patients after initiating the antiretroviral regimen 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 follow-up treatment. In addition, the measurements of total CD4+ T cell counts were also taken at Day 1, weeks 2 and 4, and monthly thereafter. Liang and Wu (2008) fitted a three-dimensional ODE model to this HIV data set.

We apply the proposed methods to estimate the unknown parameters in a HIV dynamic model based on the clinical data set. One popular HIV dynamic model is given by the following set of differential equations (Huang, Liu, and Wu, 2006; Liang and Wu, 2008):

$$\begin{cases} \frac{d}{dt}T_U(t) = \lambda - \rho T_U(t) - \eta(t)T_U(t)V(t), \\ \frac{d}{dt}T_I(t) = \eta(t)T_U(t)V(t) - \delta T_I(t), \\ \frac{d}{dt}V(t) = N\delta T_I(t) - cV(t), \end{cases} \quad (16)$$

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

Table 1

Application of HIV dynamics: Point estimates with 95% bootstrap confidence intervals obtained from different methods when N and c were fixed at 264.74 and 2.46 for the first patient, and 1114.37 and 3.78 for the second patient based on our preliminary data analysis

Patient	Parameter	LW	EDB	TDB	RDB
1	λ	198.94 (128.02,518.85)	269.31 (139.33,472.08)	223.45 (15.60,270.90)	259.60 (119.06,292.42)
	ρ	0.31 (0.16,0.68)	0.44 (0.17,0.63)	0.35 (0.06,0.36)	0.42 (0.14,0.45)
	δ	0.34 (0.17,2.12)	0.50 (0.20,1.91)	0.39 (0.01,0.52)	0.48 (0.16,0.57)
2	λ	47.01 (28.74,89.24)	46.36 (33.08,89.45)	78.52 (35.62,92.89)	82.95 (27.85,95.03)
	ρ	0.10 (0.06,0.19)	0.10 (0.07,0.19)	0.16 (0.07,0.20)	0.17 (0.06,0.20)
	δ	0.11 (0.06,0.24)	0.11 (0.07,0.24)	0.21 (0.08,0.25)	0.22 (0.06,0.26)

represents the rate at which new T cells are continuously generated; ρ is the death rate of uninfected T cells; $\eta(t)$ is the time varying infection rate of T cells which depends on antiviral drug efficacy; δ is the death rate of infected cells; c is the clearance rate of free virions; and 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. Estimation of these parameters is a challenging problem and has been studied by many authors (Wu and Ding, 1999; Putter et al., 2002; Huang, Liu, and Wu, 2006; Huang and Wu, 2006; Chen and Wu, 2008a, 2008b; Liang, Miao, and Wu, 2009). Note that we only have the observation data for total CD4+ T cell count $T(t) = T_U(t) + T_I(t)$ and viral load $V(t)$. Viral load $V(t)$ was observed in all the clinical visits, but total CD4+ T cell count, considered as an input variable, was measured less frequently. The viral load data were taken very frequently in the first two days (20 observations) of antiviral treatment because a significant drop in the viral load is expected during the early time period. We can first simplify the ODE model (16) as

$$V'(t) = \alpha_0 + \alpha_1 T(t) + \alpha_2 T'(t) - cV(t), \quad (17)$$

where $\alpha_0 = \frac{-N\delta\lambda}{\rho-\delta}$, $\alpha_1 = \frac{N\delta\rho}{\rho-\delta}$, and $\alpha_2 = \frac{N\delta}{\rho-\delta}$, so that measurement data can be directly used. Derivation of model (17) is given in Liang and Wu (2008).

In the first stage, we use the P-spline smoothing method to obtain the estimates of $V(t)$, $V'(t)$, $T(t)$, and $T'(t)$, respectively. R function “smooth.spline” was used to fit a P-spline to the data. This function uses the generalized cross-validation criterion to estimate the smoothing parameter. Note that we imputed the CD4+ T cell count data and its derivative if missing at the time points with viral load data available using the smoothing estimate as suggested by Liang and Wu (2008). Then from equation (17), we have

$$\hat{V}'(t) \approx \alpha_0 + \alpha_1 \hat{T}(t) + \alpha_2 \hat{T}'(t) - c\hat{V}(t).$$

We apply the proposed EDB, TDB, and RDB methods to estimate the parameters $(\alpha_0, \alpha_1, \alpha_2)$ or (λ, ρ, δ) . Here we used $w(t_i) = 1$ in the objective function (15). For comparison, we also obtain the estimates of the two-stage smoothing method

in Liang and Wu (2008) (named LW). But our implementation of the LW method is slightly different from the original article (Liang and Wu, 2008). We used P-splines instead of the local polynomial method for the first-stage smoothing to make the comparisons fair. Note that we fixed the parameter values for N and c as the values from the literature and our preliminary data analysis (Liang and Wu, 2008) to avoid identifiability problems. Parameters λ , ρ , and δ were estimated from the data for the two patients. The estimation results are reported in Table 1.

From Table 1, we can see that the estimation results for the first patient from all the methods are close to each other although there are some differences for different parameters. For the second patient, the estimates from the TDB and RDB methods are very close, while the estimates from the LW and EDB methods are quite similar. The difference in the estimates from these two groups of methods is large. We suspect that this is probably due to the difference in the performance of these different orders of discretization-based estimation methods as we observed in our simulation studies in the following section.

5. Simulation Studies

We designed two simulation studies to evaluate the finite-sample performance of the proposed numerical discretization-based estimation methods. The average relative error (ARE), calculated by the following formula, is used to compare the performance of different methods:

$$\text{ARE} = \frac{1}{500} \sum_{i=1}^{500} \left| \frac{\hat{\beta}_i - \beta_0}{\beta_0} \right| \times 100\%,$$

where $\hat{\beta}_i$ is the estimate and β_0 is the true value of a parameter β . To summarize the overall performance for all simulation scenarios, we use the total average of the AREs for all cases. In addition, to evaluate the trade-off between the estimation accuracy and the computational cost, we calculate the overall average computational cost for all simulation cases.

Simulation Example I: A linear ODE model. To assess the difference in finite-sample performance for different orders of the discretization algorithms, EDB, TDB, and RDB

Table 2

Simulation results: Average relative errors in percentage (sample mean \pm sample standard deviation) for the estimates of the parameters obtained from 500 replications (note that true $\alpha_0 = 3.0$ and $\alpha_1 = 0.33$ for Example I; $a = 1.5$, $b = 1$, and $c = 2$ for Example II)

	Parameter	n	(σ_1, σ_2)	LW	EDB	TDB	RDB
Ex. I	α_0	51	0.2	7.10 (2.80 \pm 0.12)	10.19 (2.70 \pm 0.11)	5.34 (2.85 \pm 0.11)	5.11 (2.87 \pm 0.11)
			0.8	17.09 (2.55 \pm 0.36)	19.38 (2.52 \pm 0.44)	14.67 (2.61 \pm 0.33)	16.29 (2.70 \pm 0.51)
		201	0.2	4.79 (2.86 \pm 0.07)	5.79 (2.83 \pm 0.06)	4.26 (2.87 \pm 0.07)	4.31 (2.87 \pm 0.07)
			0.8	11.51 (2.67 \pm 0.20)	11.78 (2.65 \pm 0.18)	11.35 (2.67 \pm 0.19)	10.97 (2.69 \pm 0.20)
	α_1	51	0.2	7.90 (0.31 \pm 0.02)	10.65 (0.30 \pm 0.01)	5.94 (0.32 \pm 0.01)	5.72 (0.32 \pm 0.01)
			0.8	18.72 (0.28 \pm 0.04)	21.10 (0.28 \pm 0.06)	16.22 (0.29 \pm 0.04)	18.40 (0.30 \pm 0.07)
		201	0.2	5.32 (0.32 \pm 0.01)	6.25 (0.31 \pm 0.01)	4.72 (0.32 \pm 0.01)	4.79 (0.32 \pm 0.01)
			0.8	12.68 (0.29 \pm 0.02)	12.91 (0.29 \pm 0.02)	12.53 (0.29 \pm 0.02)	12.08 (0.30 \pm 0.02)
Ex. II	a	26	(0.02,0.01)	11.24 (1.34 \pm 0.10)	16.93 (1.25 \pm 0.07)	7.18 (1.41 \pm 0.09)	6.84 (1.42 \pm 0.09)
			(0.05,0.03)	19.49 (1.24 \pm 0.20)	21.73 (1.19 \pm 0.17)	14.50 (1.32 \pm 0.18)	13.53 (1.35 \pm 0.19)
		51	(0.02,0.01)	9.19 (1.38 \pm 0.09)	12.30 (1.32 \pm 0.07)	6.97 (1.40 \pm 0.07)	6.68 (1.41 \pm 0.08)
			(0.05,0.03)	16.27 (1.28 \pm 0.16)	18.09 (1.25 \pm 0.16)	13.88 (1.32 \pm 0.15)	13.54 (1.34 \pm 0.17)
	b	26	(0.02,0.01)	12.93 (0.88 \pm 0.08)	18.08 (0.82 \pm 0.06)	8.23 (0.93 \pm 0.07)	7.91 (0.94 \pm 0.07)
			(0.05,0.03)	22.42 (0.80 \pm 0.16)	23.71 (0.77 \pm 0.13)	16.74 (0.86 \pm 0.14)	15.50 (0.89 \pm 0.15)
		51	(0.02,0.01)	10.58 (0.90 \pm 0.07)	13.38 (0.87 \pm 0.05)	7.98 (0.93 \pm 0.05)	7.66 (0.93 \pm 0.06)
			(0.05,0.03)	18.68 (0.83 \pm 0.12)	20.24 (0.82 \pm 0.13)	15.98 (0.86 \pm 0.12)	15.58 (0.88 \pm 0.14)
	c	26	(0.02,0.01)	2.68 (1.95 \pm 0.03)	7.82 (1.84 \pm 0.02)	1.01 (1.98 \pm 0.02)	0.73 (2.00 \pm 0.02)
			(0.05,0.03)	4.94 (1.91 \pm 0.06)	8.56 (1.83 \pm 0.05)	2.70 (1.96 \pm 0.05)	2.29 (1.97 \pm 0.05)
		51	(0.02,0.01)	1.77 (1.97 \pm 0.02)	4.31 (1.91 \pm 0.01)	1.01 (1.98 \pm 0.02)	0.88 (1.99 \pm 0.02)
			(0.05,0.03)	3.49 (1.93 \pm 0.05)	5.37 (1.89 \pm 0.04)	2.33 (1.96 \pm 0.04)	2.22 (1.96 \pm 0.04)

in the linear HIV dynamic model (17), first we selected to use a simple two-parameter ODE model for our simulation study:

$$V'(t) = \alpha_0 - \alpha_1 V(t). \quad (18)$$

We used parameter values $(\alpha_0, \alpha_1) = (3, 1/3)$ and $V(0) = -1$ to generate the data. Two different data generation schemes were used. The first is to generate the observations at every time interval of 0.4 in the range of $t \in [0, 20]$ which gives $n = 51$ observations; and the second is to generate the observations at every time interval of 0.1 in the range of $t \in [0, 20]$, which produces $n = 201$ observations. The observations were generated by numerically solving the ODE model (18) using the fourth-order Runge–Kutta method, and then measurement errors, generated from a normal distribution with

mean 0 and standard deviation σ , were added to the observations. The measurement error standard deviation is taken as $\sigma \in (0.2, 0.4, 0.6, 0.8)$.

We compare the proposed EDB, TDB, and RDB methods with the LW method. We run simulations for the NLS method as well. But the results from the NLS method varied from very good to very bad depending on the assumption about the prior knowledge of the initial conditions. The unstable results from the NLS method due to the variation in the search range for the initial conditions made it difficult to compare the results from the NLS method with the results from other methods. Therefore, we have excluded the NLS method from the simulation comparisons. In this simulation study, we used $w(t_i) = 1$ in the objective function (15). For each simulation case, 500 runs were replicated.

Table 3

Summary of simulation results: Average simulation time (cost) in seconds and overall average AREs for Examples I and II

	<i>n</i>	Parameter	LW		EDB		TDB		RDB	
			cost	ARE	cost	ARE	cost	ARE	cost	ARE
Ex. I	51	α_0	0.05	12.25%	0.05	14.74%	0.04	10.24%	0.09	10.85%
		α_1		13.57%		15.84%		11.35%		12.13%
	201	α_0	0.08	8.16%	0.09	8.76%	0.09	7.73%	0.14	7.65%
		α_1		9.00%		9.54%		8.55%		8.44%
Ex. II	26	<i>a</i>	1.23	15.18%	1.19	19.15%	1.17	10.96%	1.52	10.39%
		<i>b</i>	1.23	17.48%	1.19	20.74%	1.17	12.57%	1.52	11.91%
		<i>c</i>	1.23	3.82%	1.19	8.22%	1.17	1.94%	1.52	1.60%
	51	<i>a</i>	1.55	12.65%	1.23	15.09%	1.21	10.41%	2.66	10.26%
		<i>b</i>	1.55	14.50%	1.23	16.72%	1.21	11.94%	2.66	11.77%
		<i>c</i>	1.55	2.69%	1.23	4.82%	1.21	1.72%	2.66	1.59%

We report the AREs for the estimates of different methods in Table 2. From Table 2, we see a clear difference in the performance of different methods. When the sample size is larger, this difference is smaller. We observe a general trend that the EDB method produces the largest AREs for almost all the simulation cases. The TDB and RDB methods are quite similar, and about 50% of the cases the TDB is better than the RDB method in terms of AREs. The RDB tends to be slightly better than the TDB when the sample size is larger. The LW method seems to be somewhere in between the EDB and TDB.

Simulation Example II: A nonlinear ODE model. This simulation example was designed to further evaluate the performance of the proposed estimators under a more general nonlinear ODE model setting. We adopted a simplified model for gene–protein interactions (Chen, He, and Church, 1999) which is given by

$$\begin{cases} \frac{dR}{dt} = \frac{a}{1 + \exp(-P)} - bR, \\ \frac{dP}{dt} = 2R - cP, \end{cases} \quad (19)$$

where R denotes mRNA concentration and P denotes protein concentration. Parameter a is the transcription effect, parameter b is the degradation rate for mRNA, and parameter c is the degradation rate for protein. We used parameter values $(a, b, c) = (1.5, 1, 2)$ and initial values $(R(0), P(0)) = (0, 1)$ to generate the observations at every time interval of 0.8 and 0.4 in the range of $t \in [0, 20]$ giving $n = 26$ and $n = 51$ observations, respectively. The high computational cost prohibited us from using too large sample sizes for the non-linear model. The observations were generated by numerically solving the ODE model (19) using the fourth-order Runge–Kutta method, and then adding measurement errors to $R(t)$ generated from a normal distribution with mean 0 and standard deviation σ_1 and adding measurement errors to $P(t)$ generated from a normal distribution with mean 0 and standard deviation σ_2 . The measurement error standard deviations were taken as $(\sigma_1, \sigma_2) \in \{(0.02, 0.01), (0.05, 0.03)\}$. For each simulation case, 500 runs were replicated.

We report the AREs for the estimates of different methods in Table 2. From Table 2, we can see a similar trend as that in simulation results of the linear ODE model above. The results

confirm a general trend that the EDB method produces the largest AREs for all the simulation cases. The RDB method is slightly better than the TDB method. The LW method seems to be somewhere in between the EDB and the TDB.

For all the above-mentioned simulations, a gradient-based method (the method “L-BFGS-B” in the R function “optim”) was used for optimizing the objective function. We report the overall average computational cost and overall average AREs for the proposed methods as well as the LW estimates in Table 3.

From Table 3, again we can see that the EDB produces the largest AREs among all the methods for all cases and the computational cost of the EDB method is similar to that of the TDB method. The RDB method is slightly better than the TDB method in terms of AREs. However, the computational cost of the TDB is much lower for all cases (ranging from 23% to 55% lower in computational cost) compared to the RDB method. Again, the LW method is in between the EDB and the TDB methods in terms of AREs, but the computational cost of the LW method is higher than that of the EDB and the TDB methods. To summarize these simulation results, we recommend the TDB method for practical applications. A theoretical explanation for these simulation results are provided in Remark 5 below.

In all above simulations, the weight function $w(t_i)$ in the objective function (15) is assumed to be a constant. To evaluate the effect of the weight function as suggested by one referee, we performed additional simulations based on model (19). To save space, we report the AREs for the estimates of different methods in the online Supplementary Materials (Appendix B). The results show considerable improvement in the ARE for all the methods if the appropriate weight function is used. Comparing different methods, the EDB method again has the largest ARE. The LW method performed better than the TDB method for the cases of small noise. For the larger noise cases, the TDB method has the smallest ARE. The RDB method was slightly worse than the LW and the TDB method for most of the cases. In addition, we also performed more simulations for the cases of unequally spaced observations as suggested by the referee, similar trends, and conclusions are obtained. These additional simulation results are also included in the online Supplementary Materials (Appendix B).

Remark 5. The finite-sample behavior of the proposed numerical discretization-based methods can be explained by the expression (A.3) in the proof of Proposition 1 in the Supplemental Materials. The bias in the estimate of β is due to replacement of state vectors by their estimates in the objective function (15) which introduces an estimation error given by $\hat{x}'(t_i) - x'(t_i) + O(h)$. In addition, the omission of remainder terms in the Taylor series expansion of $x(t_{i+1})$ introduces a numerical error given by an order of $O(h^p)$. Thus, the numerical error of the EDB method ($p = 1$) is in the same order as the estimation error ($O(h)$). But for the TDB method ($p = 2$) and the RDB method ($p = 4$), the numerical error is in a higher order compared to the estimation error. That is why we can see a significant improvement in terms of AREs when we move from the lower-order EDB method to the higher-order TDB method in our simulations. However, the improvement of AREs from the TDB method to the RDB method diminishes. The performance of the LW method is between that of the EDB and TDB. This is probably because that the LW method uses the estimates of derivative curves of the state variables to form a regression model by a first-order approximation (similar to the EDB method), and the estimate of derivatives of the LW method utilizes multiple data points, but the EDB method only uses the difference of two data points to approximate the derivative. That is why the LW method improves the first-order EDB method, but not to the level of the second-order TDB method.

6. Conclusion and Discussion

In this article, we have proposed a new class of numerical discretization-based estimation (DBE) methods to estimate unknown parameters in ODE models. This new class of methods are shown to have some benefits in computational efficiency and estimation accuracy over the existing approaches. In particular, the proposed methods intend to provide a flexible balance between the computational cost and estimation accuracy. The higher order of discretization algorithms can produce better accuracy in the estimates, but require a higher computational cost. Our simulation studies demonstrate that the trapezoidal discretization-based (TDB) method provides a good balance between the computational cost and estimation efficiency and is recommended for practical use. Our theoretical analysis of the proposed methods suggests that the benefit of using a higher-order discretization algorithm will diminish when the sample size is large. Thus, a lower-order discretization method can be used when the sample size is large. Our simulation studies also support this conclusion. Our simulation results and theoretical analysis also show that using an appropriate weight at the second stage to discount the boundary effect of the first stage smoothing can also improve the estimates considerably.

Because the proposed estimates are formulated on the basis of the numerical algorithms of ODE solvers, the reasonably frequent observations may be required to obtain reasonable estimates of ODE parameters. Although more frequent smoothing estimates of the state variables from the first stage can be used in the regression estimation of the ODE parameters in the second stage, this data augmentation strategy in the second stage may need to be carefully evaluated from both

theoretical and practical perspectives, which is an interesting future research topic.

Besides the P-spline smoothing method, we are currently exploring other smoothing methods such as regression splines, smoothing splines and local polynomial techniques for the first stage of the proposed methods. We plan to compare how these different smoothing methods affect the proposed numerical discretization-based estimates from theoretical perspectives and finite-sample behaviors. We may also need to extend the proposed methods to deal with the high-dimensional case and with the latent (unobservable) state variables. In summary, this article reflects our initial work on this new promising class of estimation methods for ODE models, and more work in this direction is necessary in the future.

7. Supplementary Materials

Web Appendix A and B, referenced in Sections 3 and 5, are available under the Paper Information link at the Biometrics website <http://www.biometrics.tibs.org>.

ACKNOWLEDGEMENT

This research was partially supported by the NIAID/NIH grants AI50020, AI087135, AI078842, and AI078498. The authors appreciate the AE and two referees for their insightful comments and suggestions.

REFERENCES

- Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā* **50**, 381–393.
- Brunel, N. (2008). Parameter estimation of ODE's via nonparametric estimators. *Electronic Journal of Statistics*, **2**, 1242–1267.
- Chen, J. and Wu, H. (2008a). Efficient local estimation for time-varying coefficients in deterministic dynamic models with applications to HIV-1 dynamics. *Journal of the American Statistical Association* **103**, 369–384.
- Chen, J. and Wu, H. (2008b). Estimation of time-varying parameters in deterministic dynamic models with application to HIV infections. *Statistical Sinica* **18**, 987–1006.
- Chen, T., He, H.L., and Church, G.M. (1999). Modeling gene expression with differential equations. *Pacific Symposium of Biocomputing* **4**, 29–40.
- Claeskens, G., Krivobokova, T., and Opsomer, J. (2009). Asymptotic properties of penalized spline estimators. *Biometrika* **96**, 529–544.
- Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the generalized cross-validation. *Numerische Mathematik* **31**, 337–403.
- de Boor, C. (2001). *A Practical Guide to Splines*. Revised Edition. New York: Springer.
- Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science* **11**, 89–121.
- Gasser, T. and Müller, H. G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics* **11**, 171–185.
- Hairer, E., Norsett, S., and Wanner, G. (1993). *Solving Ordinary Differential Equation I: Nonstiff Problems*. Berlin: Springer-Verlag.
- Hall, P. and Opsomer, J. D. (2005). Theory of penalised spline regression. *Biometrika* **92**, 105–118.
- Hemker, P. (1972). Numerical methods for differential equations in system simulation and in parameter estimation. In *Analysis and*

- Simulation of Biochemical Systems*, H. C. Hemker and B. Hess (eds), 59–80. Amsterdam and London: North-Holland/New York: American Elsevier.
- Hemker, P. (1972). Numerical methods for differential equations in system simulations and in parameter estimation. *Analysis and Simulation of Biochemical Systems*, 59–80.
- Huang, L. and Fan, J. (1999). Nonparametric estimation of quadratic regression functionals. *Bernoulli* **5**, 927–949.
- Huang, Y., Liu, D., and Wu, H. (2006). Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics* **62**, 413–423.
- Huang, Y. and Wu, H. (2006). A Bayesian approach for estimating antiviral efficacy in HIV dynamic models. *Journal of Applied Statistics* **33**, 155–174.
- Kauermann, G., Krivobokova, T., and Fahrmeir, L. (2009). Some asymptotic results on generalized penalized spline smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**, 487–503.
- Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika* **95**, 415–436.
- Li, Z., Osborne, M. R., and Pravan, T. (2005). Parameter estimation of ordinary differential equations. *IMA Journal of Numerical Analysis* **25**, 264–285.
- Liang, H., Miao, H., and Wu, H. (2009). Estimation of constant and time-varying dynamic parameters of HIV infection in a nonlinear differential equation model. *Search Results Annals of Applied Statistics* **4**, 460–483.
- Liang, H. and Wu, H. (2008). Parameter estimation for differential equation models using a framework of measurement error in regression. *Journal of the American Statistical Association* **103**, 1570–1583.
- Matteij, R. and Molenaar, J. (2002). *Ordinary Differential Equations in Theory and Practice*. Philadelphia, Pennsylvania: SIAM.
- O’Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion). *Statistical Science* **1**, 505–527.
- Poyton, A. A., Varziri, M. S., McAuley, K. B., McLellan, P. J., and Ramsay, J. O. (2006). Parameter estimation in continuous-time dynamic models using principal differential analysis. *Computers and Chemical Engineering* **30**, 698–708.
- Putter, H., Heisterkamp, S., Lange, J., and Wolf, F. (2002). A Bayesian approach to parameter estimation in HIV dynamic models. *Statistics in Medicine* **21**, 2199–2214.
- Qi, X. and Zhao, H. (2010). Asymptotic efficiency and finite-sample properties of the generalized profiling estimation of parameters in ordinary differential equations. *Annals of Statistics* **38**, 435–481.
- Ramsay, J. O. (1996). Principal differential analysis: Data reduction by differential operators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **58**, 495–508.
- Ramsay, J. O., Hooker, G., Campbell, D., and Cao, J. (2007). Parameter estimation for differential equations: A generalized smoothing approach (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **5**, 741–796.
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge, UK: Cambridge University Press.
- Schumaker, L. L. (1981). *Spline Functions: Basis Theory*. New York: Wiley.
- Varah, J. (1982). A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal on Scientific and Statistical Computing* **3**, 28–46.
- Varziri, M. S., Poyton, A. A., McAuley, K. B., McLellan, P. J., and Ramsay, J. O. (2008). Selecting optimal weighting factors in iPDA for parameter estimation in continuous-time dynamic models. *Computers and Chemical Engineering* **32**, 3011–3022.
- Wand, M. and Ormerod, J. (2008). On semiparametric regression with O’Sullivan penalised splines. *Australian and New Zealand Journal of Statistics* **50**, 179–198.
- 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.
- Xue, H., Miao, H. and Wu, H. (2010). Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error. *Annals of Statistics* **38**, 2351–2387.
- Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single index model. *Journal of the American Statistical Association* **97**, 1042–1054.
- Zhou, S., Shen, X. and Wolfe, D. (1998). Local asymptotics for regression splines and confidence regions. *Annals of Statistics* **26**, 1760–1782.
- Zhou, S. and Wolfe, D. (2000). On derivative regression in spline estimation. *Statistica Sinica* **10**, 93–108.

Received February 2011. Revised September 2011.

Accepted September 2011.