Estimating an Ordinary Differential Equation Model With Partially Observed Data

Kwang Woo Ahn

Division of Biostatistics

Medical College of Wisconsin, Milwaukee, WI 53226

email: kwooahn@mcw.edu

Michael Kosoy

Centers for Disease Control and Prevention

Fort Collins, CO 80521

email: mck3@cdc.gov

Kung–Sik Chan

Department of Statistics and Actuarial Science

The University of Iowa, Iowa City, Iowa 52242

email: kung-sik-chan@uiowa.edu

July 22, 2011

Abstract

Ordinary differential equation (ODE) models, e.g. the susceptible-infected-recovered model, are widely used in engineering, ecology, and epidemiology. Many ODE models specify the state process as driven by a system of nonlinear differential equations. In practice, ODE models may be observed partially in that only a proper subset of the state vector is observed with measurement errors over discrete time. While there are several methods for estimating ODE models with partially observed data, they are invariably subject to several problems including high computational cost, sensitivity to initial values or large sampling variability. We propose a new computationally efficient two-step method for estimating a differential equation model with partially observed data. We derive the consistency and large-sample distribution of the proposed method. The efficacy of the new method is illustrated by simulations and a real epidemiological time-series data on the prevalence of *Bartonella* in a wild population of cotton rats.

KEYWORDS: Asymptotic normality; Bartonella; Consistency; Local polynomial regression; Nonlinear state-space model; SIR model

1. INTRODUCTION

State-space models whose state process is driven by some system of nonlinear ordinary differential equation (ODE) are widely used to study dynamics in science (Farlow 1993; Diekmann and Heesterbeek 2000); such models are referred to as differential equation models below. For example, a variant of the susceptible-infected-recovered (SIR) model is driven by the following system of nonlinear differential equations:

$$\begin{aligned} \frac{dS}{dt} &= -\alpha\frac{SI}{N} + bN - \mu S, \quad & \frac{dR}{dt} &= \gamma I - \mu R, \\ \frac{dI}{dt} &= \alpha\frac{SI}{N} - \gamma I - \mu I, \quad & \frac{dN}{dt} &= -\mu N + bN, \end{aligned} \tag{1}$$

where S, I, R, and N are the sizes of susceptibles, infectives, recovered individuals, and total population, respectively, and α, μ, γ, b are the force of infection, death rate, recovery rate, and birth rate, respectively. If the proportions of the susceptible, infective and recovered are of main interest, equation (1) can be further reduced in dimension and simplified as follows:

$$\begin{aligned}
\frac{d}{dt}\Big(\frac{S}{N}\Big) &= -\alpha\frac{S}{N}\frac{I}{N} + \Big(1-\frac{S}{N}\Big)b,\\
\frac{d}{dt}\Big(\frac{I}{N}\Big) &= \alpha\frac{S}{N}\frac{I}{N} - (b+\gamma)\frac{I}{N},\\
\frac{d}{dt}\Big(\frac{R}{N}\Big) &= \gamma\frac{I}{N} - b\frac{R}{N},
\end{aligned} \tag{2}$$

where S/N, I/N, and R/N represent the proportions of susceptibles, infectives, and recovered individuals. Below, besides d/dt, we shall freely use several notations standing for the derivative operator including $'$, the superscript $^{(1)}$, and their higher order counterparts, for notational convenience and clarity. The third equation of (2) is redundant because $(S+I+R)/N = 1$ and μ is eliminated from equation (1). In practice, the state process $(S/N, I/N)^T$ is often partially observed in that only I/N is observed at discrete time with measurement errors. For example, Kosoy, Mandel, Green, Marston, Jones and Childs (2004a) monitored the prevalence of *Bartonella* infection in a wild population of cotton rats over a period of about 2 years. They reported monthly number of infected rats based on marked-capture-recapture study, but no information on the susceptibles is available; we shall return to this example below. Indeed, the problem of estimation with partially observed data from a process driven by a nonlinear ODE arises frequently in practice.

Consider the general case with the state equation driven by the following ODE:

$$\frac{d\boldsymbol{X}_0(t)}{dt} = Z(\boldsymbol{X}_0(t), \boldsymbol{\beta}), \tag{3}$$

where $\boldsymbol{X}_0(t) = (X_1(t), \dots, X_k(t))^T$ is the true state vector, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)^T$ is an unknown parameter vector, and $Z(\cdot) = (Z_1(\cdot), \dots, Z_k(\cdot))^T$ is a nonlinear vector function. Let $y(t) = a(\boldsymbol{X}_0(t))$ be a vector function of the state that is observable with additive measurement error:

$$Y(t) = y(t) + \epsilon(t), \tag{4}$$

where the errors $\{\epsilon(t)\}$ are independent and $\epsilon(t)$ is independent of $\{\boldsymbol{X}_0(s), s \le t\}$. The case of fully observed state corresponds to a being the identity function or some bijective function. For simplicity, we shall mainly consider the case of scalar-valued y.

Parameter estimation of an ODE model has been extensively studied in the literature. The state-space models using the extended Kalman filter (EKF) and the unscented Kalman filter (UKF) have been investigated from a Bayesian perspective; see Simon (2006). EKF approximates a nonlinear system by its first order Taylor expansion. However, the approximation error becomes non-negligible for strongly nonlinear systems. UKF partly overcomes this drawback by using unscented transformation, but it lacks theoretical justification.

The second approach to solving an ODE model is the collocation method which approximates the solution by some finite-dimensional basis-function expansion; see Ramsay, Hooker, Campbell and Cao (2007) who proposed the generalized profile estimation method. Some asymptotic properties of the latter method have been obtained recently by Brunel (2008) and Qi and Zhao (2010). The generalized profile estimation method works for both fully and partially observed data. However, the method is computationally expensive as it requires profiling out the generally high-dimensional coefficient in the functional expansion. Also, the associated optimization method appears to be quite sensitive to initial values; see simulation results in Section 5.

The third approach is the two-step approach that consists of (i) estimating the derivatives of the state process from data at the sampling epochs, via some nonparametric smoothing methods, and then (ii) estimating the model parameters by minimizing the sum of squared deviations between the left and right side of the defining ODE, with the derivatives replaced by their nonparametric estimates from (i); see Liang and Wu (2008) and Brunel (2008) where kernel smoothing and spline smoothing were used to implement (i), respectively. Brunel (2008) extended the two-step method for handling partially observed data for a class of models, the derivative of whose unobserved state component is partly linear in the unobserved state. Wu, Zhu, Miao and Perelson (2008) pointed out that the problem of partially observed data may be circumvented by deriving an equivalent ODE model in terms of the observed state components, assuming some Jacobian conditions. They proposed to estimate the ODE model by optimizing some criterion function based on the fit of the solution of the ODE to observed data, which results in a more complex objective function that is, also, sensitive to the estimation of the initial states of the ODE.

Here, our contributions are twofold. First, we provide in Section 2 a set of sufficient conditions, and necessary conditions, under which we can derive an ODE equivalent to (A.1) and whose state, known as derivate coordinates (Kantz and Schreiber 2003, p. 152), comprises an observed component of the state vector of (A.1) and its ith derivatives, $i = 1, \ldots, k-1$. Assuming the existence of such an equivalent ODE, we may then estimate the new state vector with discrete-time data on the observed component, via local polynomial regression, which can then be fed into the second step of estimation via least squares. In Section 3, we propose a new two-step method for estimating an ODE model with partially observed data, and derive the consistency and large-sample distribution of the proposed estimation method. In Section 4, we discuss conditions on the validity of the proposed method for estimating the SIR model with discrete-time data on the proportions of infectives. Empirical performance of the proposed method is studied by simulations in Section 5. We illustrate the efficacy of the proposed method by estimating an SIR model with seasonal birth rate, using the *Bartonella* data, in Section 6. We briefly conclude in Section 7.

2. VALIDITY OF THE METHOD OF DERIVATIVE COORDINATES

In this section, we study some sufficient conditions for the existence of an ODE that is equivalent to (A.1) and whose state vector comprises the derivative coordinates based on an observable function of the original state vector of (A.1). Since the derivative coordinates can be estimated by local polynomial regression, we can employ a two-step estimation scheme using the equivalent ODE.

Let $y^{(j)}$ be a function of $\boldsymbol{x}$ such that it equals the jth derivative of $y(t) = a(\boldsymbol{X}_0(t))$ with respect to t, evaluated at t with $\boldsymbol{X}_0(t) = \boldsymbol{x}$, for x in the state space $\mathcal{X} \subseteq \mathbb{R}^k$, an open subset. Since the ODE defined by (A.1) is autonomous, the preceding derivatives are invariant with respect to t and they can be evaluated at $t = 0$. Write y for $y^{(0)}$. Clearly, $y = a(\boldsymbol{x})$. Let $\boldsymbol{y}_{k-1} = (y^{(j)}, j = 0, \ldots, k-1)^T$, which can be interpreted as the observable y and its first $k-1$ derivatives along the solution of (A.1). The basic idea of our approach is based on the fact that the ODE defined by (A.1) is equivalent to an ODE with $\boldsymbol{y}_{k-1}$ as the state vector, i.e. there exists a bijective, differentiable map H such that $\boldsymbol{y}_{k-1} = H(\boldsymbol{x})$, under some mild regularity conditions. Assume for the moment that this is true and that (A.1) is equivalent to a k-order ODE in y:

$$y^{(k)}(t) = G(\boldsymbol{y}_{k-1}(t); \boldsymbol{\beta}). \tag{5}$$

Suppose, furthermore, that we observe $Y(t_i) = y(t_i) + \epsilon(t_i),\ i = 1, \ldots, n$. Denote the estimator by $\hat{y}^{(j)}(t_i),\ j = 0, \ldots, k,\ i = 1, \ldots, n$. In particular, $\hat{\boldsymbol{y}}_{k-1}(t_i) = (\hat{y}(t_i), \ldots \hat{y}^{(k-1)}(t_i))^T$. Let $|\cdot|$ be the Euclidean norm of the enclosed expression. Then we can estimate $\boldsymbol{\beta}$ by the following two-step estimation procedure. First, estimate y and its derivatives, i.e. $y^{(j)}(t_i),\ j = 0, \ldots, k$, via local polynomial fitting based on the observations $Y(t_i),\ i = 1, \ldots, n$. Next we estimate $\boldsymbol{\beta}$ by minimizing the objective function

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} |\hat{y}^{(k)}(t_i) - G(\hat{y}_{k-1}(t_i); \boldsymbol{\beta})|^2. \tag{6}$$

This approach circumvents the difficulties of being unable to observe the entire original state vector $\boldsymbol{X}_0$ up to some additive measurement error. The theoretical and empirical properties of the proposed estimation scheme will be developed in the following sections.

In the remainder of this section, we discuss when $\boldsymbol{y}_{k-1}$ provides an equivalent state vector for the ODE defined by (A.1). A similar problem in the context of a controlled ODE has been extensively studied, see Hunter, Su and Meyer (1983), Terrell (1999) and the references therein. However, these results do not apply to our setting without control. First, define some notations. Let $q = q(\boldsymbol{x})$ be a differentiable real-valued function of $\boldsymbol{x} \in \mathbb{R}^k$, with as many derivatives as needed in subsequent discussion. Define the Lie derivative $L_Z(q)$ whose value at $\boldsymbol{x} \in \mathbb{R}^k$ equals $< Z(\boldsymbol{x}), \nabla q(\boldsymbol{x}) >= Z(\boldsymbol{x})^T \nabla q(\boldsymbol{x})$ where Z is as defined in (A.1), $< \cdot >$ denotes taking the inner product, and $\nabla q(\boldsymbol{x})$ is the gradient of q evaluated at $\boldsymbol{x}$, i.e. if $\boldsymbol{x} = (x_1, \ldots, x_k)^T$, $\nabla q(\boldsymbol{x}) = (\partial q(\boldsymbol{x})/\partial x_1, \ldots, \partial q(\boldsymbol{x})/\partial x_k)^T$; we often suppress the parameter $\boldsymbol{\beta}$ for clarity. Let $Q(\boldsymbol{x}) = (a(\boldsymbol{x}), L_Z(a)(\boldsymbol{x}), \ldots, L_Z^{k-1}(a)(\boldsymbol{x}))^T$, where L_Z^j denotes composition of the function L_Z with itself j times. It can be shown by chain rule that $\boldsymbol{y}_{k-1} = Q(\boldsymbol{x})$. The dependence of Q on $\boldsymbol{\beta}$ is suppressed for clarity. Let $Q : \mathcal{X} \to \mathcal{Y}$, with both $\mathcal{X}$ and $\mathcal{Y}$ being open subsets of $\mathbb{R}^k$. The following main result of this section concerns when the function Q is globally invertible, in which case $\boldsymbol{y}$ provides an equivalent state vector for the ODE driven by (A.1). (This result is likely "known" among experts but we cannot find it stated in the literature.)

Theorem 1 *For the ODE defined by (A.1), $\boldsymbol{y}$ is an equivalent state vector with the parameter space $\Omega_{\boldsymbol{\beta}}$ if, for any fixed $\boldsymbol{\beta} \in \Omega_{\boldsymbol{\beta}}$, (i) Q is proper, i.e. for any compact set $K \subseteq \mathcal{Y}$, its pre-image $Q^{-1}(K)$ is a compact set, (ii) the Jacobian matrix $J(\boldsymbol{x}) = \partial Q/\partial \boldsymbol{x}^T$ is of full-rank, i.e. of non-zero determinant, for all $x \in \mathcal{X}$, (iii) $\mathcal{X}$ is a connected set and $\mathcal{Y}$ is a simply connected set, i.e. any two points in $\mathcal{Y}$ can be connected by a continuous curve lying entirely inside $\mathcal{Y}$. Conversely, if Q is globally invertible, then condition (ii) holds.*

The sufficiency part of the preceding theorem follows from the Hadamard global inverse function theorem, see Theorem 6.2.8 of Krantz and Parks (2002). The necessity part is trivial since an invertible Q must have an invertible Jacobian matrix.

Now, we illustrate this result with the SIR model defined by (2). We shall drop the last equation there as it is redundant since $S + I + R = N$. Write s for S/N and i for I/N. Let $\boldsymbol{x} = (s, i)^T$. The state space $\mathcal{X} = \{(s, i) : 0 < s < 1, 0 < i < 1, 0 < s + i < 1\}$. Here, the boundary $si = 0$ or $1 = s + i$ is excluded. Then, the state equation is given by

$$\boldsymbol{x}' = Z(\boldsymbol{x}) = \begin{pmatrix} -\alpha si + b(1-s) \\ \alpha si - (b+\gamma)i \end{pmatrix}.$$

Suppose we observe i up to some additive measurement errors. Hence, $y = a(\boldsymbol{x}) = i$. Then, $L_Z(a)(\boldsymbol{x}) = Z(\boldsymbol{x})^T(0,1)^T = \alpha si - (b+\gamma)i$ because $\nabla a(\boldsymbol{x}) = (0,1)^T$. So,

$$Q(\boldsymbol{x}) = \begin{pmatrix} i \\ \alpha si - (b+\gamma)i \end{pmatrix}.$$

Hence, $y_1 \in \mathcal{Y} = (0,1) \times \mathbb{R}$. The parameter space $\Omega_{\boldsymbol{\beta}} = \{(\alpha, \gamma, b)^T \in \mathbb{R}^3 : \alpha > 0, \gamma > 0, b > 0\}$. In this case, the global inversion of $y = Q(\boldsymbol{x})$ can be shown with ease as $i = y$ and $s = \{y^{(1)}/y + (b+\gamma)\}/\alpha$.

Alternatively, we now show the global invertibility of Q by checking the conditions of Theorem 1. We first verify condition (i). Let $K \subseteq \mathcal{Y}$ be a compact set. Clearly $Q^{-1}(K)$ is the intersection of a closed set and $\mathcal{X}$, because Q is a continuous function. It remains to show that $Q^{-1}(K)$ is bounded away from the boundary of $\mathcal{X}$ so that it is closed and bounded, and hence compact. The compactness of K implies that for any $\boldsymbol{y}_1 = (y, y^{(1)})^T \in K$, y is strictly bounded away from 0 and 1 and $y^{(1)}$ is bounded, hence $i = y$ is bounded away from 0 and 1 and so is $s = \{y^{(1)}/y + (b+\gamma)\}/\alpha$. That $s + i$ does not approach 1 follows from the third equation of the SIR model and the fact that i is bounded away from 0. Therefore condition (i) holds. The Jacobian matrix of Q equals

$$J(\boldsymbol{x}) = \begin{pmatrix} 0 & 1 \\ \alpha i & \alpha s - (b+\gamma) \end{pmatrix}.$$

J is of full-rank at x if and only if its determinant at $\boldsymbol{x}$ is non-zero. Now, $\det(J(\boldsymbol{x})) = -\alpha i$, which is non-zero for $\alpha \neq 0$ and $i = I/N \neq 0$, so condition (ii) holds. Finally, condition (iii) holds trivially.

3. A NEW TWO-STEP ESTIMATION METHOD

Without loss of generality, assume that the original ODE under study can be equivalently formulated as a pth order ODE in terms of a scalar process $\{X_1(t)\}$ which admits as many derivatives as required below and is observable with additive measurement errors over a set of epochs. The case of a vector process and how to determine p are briefly discussed at the end of this section. Specifically, let the pth order differential equation be

$$X_1^{(p)}(t) = F(\boldsymbol{X}(t), \boldsymbol{\beta}), \tag{7}$$

where $\boldsymbol{X}(t) = (X_1^{(0)}(t), \ldots, X_1^{(p-1)}(t))^T$ with $X_1^{(0)}(t) = X_1(t)$. For simplicity, we assume $a(\boldsymbol{X}_0) = (1, 0, \ldots, 0)\boldsymbol{X}_0$, i.e., the observation equation equals

$$Y(t_i) = X_1(t_i) + \epsilon(t_i),\ i = 1, \ldots, n, \tag{8}$$

where $\epsilon(t_i)$'s are independent with covariance matrix $\text{diag}(\sigma^2(t_1), \ldots, \sigma^2(t_n))$. The smoothness of the X-process and Taylor expansion imply that

$$X_1(t_j) \approx \sum_{i=0}^{\omega} \frac{X_1^{(i)}(t)}{i!}(t_j - t)^i \equiv \sum_{i=0}^{\omega} \alpha_i(t)(t_j - t)^i.$$

We can then estimate $\boldsymbol{\alpha}(t) = (\alpha_0(t), \ldots, \alpha_\omega(t))^T$ or $(X_1^{(0)}(t), \ldots, X_1^{(\omega)}(t))^T$ via local polynomial regression by minimizing

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

where $K(\cdot)$ is a symmetric kernel function, $K_{h_n}(\cdot) = K(\cdot/h_n)/h_n$, and $h_n > 0$ is the bandwidth. The derivatives at t can then be estimated as $\hat{X}^{(i)}(t) = i!\hat{\alpha}_i(t)$. For estimating the qth derivative, Fan and Gijbels (1996, Section 3.3) recommended setting the degree of the local polynomial to be larger than q by 1, i.e., $\omega = q + 1$, which will be adopted henceforth below .

For estimating the ODE model defined by (7) and (8), the parameter $\boldsymbol{\beta}$ can be estimated by minimizing the following objective function:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{\hat{X}_1^{(p)}(t_i) - F(\hat{\boldsymbol{X}}(t_i), \boldsymbol{\beta})\}^2.$$

This generalizes the approach by Liang and Wu (2008) who derived the large-sample properties of the estimator for the case of a vector differential equation with $p = 1$, which is appropriate when the state vector is fully observable. For the case of partially observable state vector, we now derive the large-sample properties of the proposed method. Define

$$D_n(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \sum_{i=1}^{n} \{F(\boldsymbol{X}(t_i), \boldsymbol{\beta}_1) - F(\boldsymbol{X}(t_i), \boldsymbol{\beta}_2)\}^2.$$

The function $D_n(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)/n$ converges to a function $D(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = E\{F(\boldsymbol{X}(t), \boldsymbol{\beta}_1) - F(\boldsymbol{X}(t), \boldsymbol{\beta}_2)\}^2$ as $n \to \infty$ by the uniform law of large numbers, under some regularity conditions. For simplicity, we suppress t in $F(\boldsymbol{X}(t), \boldsymbol{\beta})$ in the following regularity assumptions.

Assumption 2

(i) The function $X_1^{(j)}(t)$ *is continuous on* $\mathcal{T}$*, the compact support of* t*, for* $j = 0, 1, \ldots, p+2$.

(ii) The kernel function K *is symmetric about 0 and supported over* $[-1, 1]$*. In addition,* $K(x_1 + x_2/h_n) = O(h_n)$ *uniformly for* $x_1 \in [-1, 1]$ *and* $x_2 \neq 0$ *where* $h_n \to \infty$ *as* $n \to \infty$*. This condition is satisfied by most kernel functions practically used including the Gaussian kernel function, the triangular kernel function, and the following kernel functions for* $j \geq 1$,

$$K(x) = \frac{(1-x^2)^j}{2^{2j+1}B(j+1, j+1)} 1_{\{|x| \leq 1\}},$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$*. The kernel functions of* $j = 1, 2, 3$ *are Epanechnikov, biweight, and triweight kernel functions, respectively.*

(iii) The bandwidth h_n *is a sequence of positive numbers such that* $h_n \to 0$ *but* $nh_n^{2p+1} \to \infty$ *as* $n \to \infty$.

(iv) The sampling epochs t_i*'s are independent and identically distributed with a compact support* $\mathcal{T}$*; the common density function,* $f(t)$*, is bounded away from 0 and has continuous second derivatives; the variance* $\sigma^2(t)$ *is bounded away from 0 and* ∞ *for* $t \in \mathcal{T}$.

Assumption 3

(i) The function $F(\boldsymbol{X}, \boldsymbol{\beta})$ *is a continuous function of* $\boldsymbol{X}$ *and* $\boldsymbol{\beta}$ *for* $\boldsymbol{X} \in \mathcal{X}$ *and* $\boldsymbol{\beta} \in \Omega_{\boldsymbol{\beta}}$*, a compact subset of* $\mathbb{R}^m$ *in whose interior lies the true parameter value* $\boldsymbol{\beta}_0$.

(ii) Equation $D(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = 0$ *if and only if* $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

Assumption 4

(i) The derivatives, $\partial F(\boldsymbol{X}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}$*,* $\partial F(\boldsymbol{X}, \boldsymbol{\beta})/\partial \boldsymbol{X}$*,* $\partial^2 F(\boldsymbol{X}, \boldsymbol{\beta})/(\partial \boldsymbol{X} \partial \boldsymbol{\beta}^T)$*, and* $\partial^2 F(\boldsymbol{X}, \boldsymbol{\beta})/(\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T)$ *exist and are continuous functions of* $\boldsymbol{\beta} \in \Omega_{\boldsymbol{\beta}}$*,* $\boldsymbol{X} \in \mathcal{X}$*. Also, there exist two positive constants* d *and* $0 < \zeta \leq 1$ *such that for all* $\boldsymbol{X}_{\nu_1}, \boldsymbol{X}_{\nu_2} \in \mathcal{X}$

$$\Big| \frac{\partial F(\boldsymbol{X}_{\nu_1}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \frac{\partial F(\boldsymbol{X}_{\nu_2}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big| \leq d |\boldsymbol{X}_{\nu_1} - \boldsymbol{X}_{\nu_2}|^{\zeta}.$$

(ii) The partial derivative $\partial F(\boldsymbol{X},\boldsymbol{\beta})/\partial \boldsymbol{X}$ is a continuous function of $\boldsymbol{X} \in \mathcal{X}$ and satisfies

$$\sup_{\boldsymbol{X}\in\mathcal{X}} \left| \frac{\partial F(\boldsymbol{X},\boldsymbol{\beta})}{\partial \boldsymbol{X}} \right| \leq M,$$

for some constant $M < \infty$.

Assumption 1 imposes standard regularity conditions on the kernel and the density of the sampling epochs in order to ensure standard asymptotic properties of the kernel estimators of the derivatives of the state process. Assumption 2 is related to the identifiability of the ODE model, see Xia and Moog (2003). Assumption 3 imposes some smoothness conditions on the ODE. For the case $p = 1$, these assumptions are similar to those used in Liang and Wu (2008).

Let

$$\mu_j = \int u^j K(u)du, \quad \nu_j = \int u^j K^2(u)du.$$

Define two $(p+2)\times(p+2)$ matrices $\boldsymbol{S}_{p+1}$ and $\boldsymbol{S}^*_{p+1}$ with their (i,j)th entries equal to μ_{i+j-2} and ν_{i+j-2}, respectively. Let $\boldsymbol{c}_{p+1} = (\mu_{p+2},\ldots,\mu_{2p+3})^T$. Let $\boldsymbol{\xi}_{p+1}$ be the $(p+2)\times 1$ unit vector having 1 in the $(p+1)$th entry and $\boldsymbol{\beta}_0$ be the true parameter vector; the subscript $p+1$ in $\boldsymbol{c}_{p+1}$ and $\boldsymbol{\xi}_{p+1}$ refers to the degree of the local polynomial employed in estimating $X_1^{(p)}$. We now state the main results on the consistency and the large-sample distribution of the proposed estimation method, where p, the order of the differential equation, is an arbitrary but fixed positive integer. The proof can be found in Appendix.

Theorem 5 *Under Assumptions 1-3, $\hat{\boldsymbol{\beta}}_n$ is consistent, i.e. $\hat{\boldsymbol{\beta}}_n$ converges to $\boldsymbol{\beta}_0$ in probability as $n \to \infty$. In addition, $nh_n^{(2p+1)/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 + \boldsymbol{C}h_n^2)$ converges weakly to a normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$, where*

$$\boldsymbol{C} = \frac{\boldsymbol{\xi}_{p+1}^T \boldsymbol{S}_{p+1}^{-1}\boldsymbol{c}_{p+1}}{p+1} E\Big(X^{(p+2)}(t)\frac{\partial F(\boldsymbol{X}(t),\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\Big),$$

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = {} & p!^2 \boldsymbol{\xi}_{p+1}^T \boldsymbol{S}_{p+1}^{-1}\boldsymbol{S}^*_{p+1}\boldsymbol{S}_{p+1}^{-1}\boldsymbol{\xi}_{p+1}\Big[E\Big\{\frac{\partial F(\boldsymbol{X}(t),\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\Big\}\Big\{\frac{\partial F(\boldsymbol{X}(t),\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\Big\}^T\Big]^{-1} \\ & \times E\Big[\Big\{\frac{\sigma(t)\partial F\{\boldsymbol{X}(t),\boldsymbol{\beta}_0\}}{\sqrt{f(t)}\partial\boldsymbol{\beta}}\Big\}\Big\{\frac{\sigma(t)\partial F\{\boldsymbol{X}(t),\boldsymbol{\beta}_0\}}{\sqrt{f(t)}\partial\boldsymbol{\beta}}\Big\}^T\Big] \\ & \times \Big[E\Big\{\frac{\partial F(\boldsymbol{X}(t),\beta_0)}{\partial \boldsymbol{\beta}}\Big\}\Big\{\frac{\partial F(\boldsymbol{X}(t),\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\Big\}^T\Big]^{-1}.\end{aligned}$$

Remark 6 *Although Theorem 5 concerns the case of random design in the sampling scheme, Fan and Gijbels (1996, Section 3.2.4) pointed out that local polynomial estimators adapt to both random and fixed designs. Especially, the bias and variance expressions in Appendix remain valid for both designs, so the asymptotic properties stated in Theorem 5 hold for fixed design as well, assuming the uniform convergence of* $D_n(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)/n \to D(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ *as* $n \to \infty$.

Remark 7 *For bandwidth selection, one may use the optimal bandwidth* h_{opt} *given in Section 3.2.3 of Fan and Gijbels (1996), with the unknown quantity there replaced by estimates from data using some initial bandwidth value. Indeed, we adopted this approach in the simulations and real application below. The optimal bandwidth* h_{opt} *is of order* $n^{-1/(2p+5)}$*, which satisfies Assumption 1(iii). We observed this simple optimal bandwidth worked well in our simulation study.*

Remark 8 *Theorem 5 implies that the convergence rate may be faster than the root-n convergence rate, which was also observed and justified by Liang and Wu (2008) for* $p = 1$*. The convergence rate generally does not exceed the root-n rate for regular models. However, the faster convergence rate in our case obtains because the variance of* $\hat{X}_1^{(p)}(t_i) - F(\hat{\boldsymbol{X}}(t_i), \boldsymbol{\beta})$ *in* $S(\boldsymbol{\beta})$ *is proportional to* $1/(nh_n^{2p+1})$*, which converges to zero under the regularity conditions.*

Although so far we have considered the case that only a univariate component of the original state vector is observable, the method extends readily to the case when more than one component of the state vector are observable, see Appendix.

4. SIR MODEL REVISITED

We now illustrate the new method using the SIR model. Consider the SIR model defined by the first two equations of (2) because the third equation is redundant. Letting $S(t)/N(t) = s(t)$ and $I(t)/N(t) = i(t)$ and suppressing t for simplicity, these two equations become

$$\begin{aligned} s' &= -\alpha si + (1-s)b, \\ i' &= \alpha si - (b+\gamma)i. \end{aligned} \tag{9}$$

Furthermore, the birth rate b is relaxed to be some smooth function of t, up to some parameters that are included in the vector parameter $\boldsymbol{\beta}$. For illustration, we shall consider the following simple seasonal birth rate model:

$$b = b(t) = q\sin(\frac{\pi}{6}t) + r\cos(\frac{\pi}{6}t) + v, \tag{10}$$

where the period is 12. In practice, we may only have information on the proportion of infectives, i.e. only i is observed at epochs $t_1, \ldots, t_n \in \mathcal{T} \subset \mathbb{R}$. It follows from the discussions in Section 2 that $(i, i^{'})^T$ is an equivalent state vector. Indeed, s can be solved from the second equation of (A.4) as follows:

$$s = \frac{1}{\alpha}\Big(\frac{i^{'}}{i} + b + \gamma\Big). \tag{11}$$

and hence we can also compute $s^{'}$ from the first equation of (A.4) as follows:

$$s^{'} = -\alpha s i + (1-s)b = -(i^{'} + bi + \gamma i) + \Big\{1 - \frac{1}{\alpha}\Big(\frac{i^{'}}{i} + b + \gamma\Big)\Big\}b. \tag{12}$$

In order to construct the second-order differential equation in terms of i and $i^{'}$, we differentiate the second equation of (A.4) with respect to t. Using (11) and (12), we obtain

$$\begin{aligned} i^{''} &= \alpha s^{'} i + \alpha s i^{'} - b^{'} i - (b+\gamma) i^{'} \\ &= (i^{'} + bi + \gamma i)(-\alpha i - b) + \frac{i^{'2}}{i} + (\alpha b - b^{'}) i \equiv F(\boldsymbol{X}, \boldsymbol{\beta}), \end{aligned} \tag{13}$$

where $\boldsymbol{X} = (i, i^{'})^T$.

Based on local polynomial fitting, we can estimate the higher derivatives of i. Denoting the estimators of the first and second derivatives by $\hat{i}^{'}$ and $\hat{i}^{''}$ respectively, the proposed method then estimates the unknown parameter vector $\boldsymbol{\beta}$ by minimizing the following objective function:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{\hat{i}^{''}(t_i) - F(\hat{\boldsymbol{X}}(t_i), \boldsymbol{\beta})\}^2. \tag{14}$$

We show in Appendix that Assumptions 1-3 for the SIR model with seasonal birth rate are valid under very general conditions.

Table 1: Empirical performance of the proposed estimation method. Data were simulated from a state-space model whose state is driven by the ODE defined by equations (1) and (12). The parameter n denotes the sample size and Δ the time interval. Average parameter estimates are reported with their standard deviation in the row with title 'SD' just below it, whereas the corresponding coverage rate of the nominal 95% confidence interval reported in the row with heading 'CR' two rows below it.

		b known		q, r known			r known				all parameters unknown				
Simulation Study I															
n	Δ	α	γ	α	γ	v	α	γ	q	v	α	γ	q	r	v
		(1)	(0.15)	(1)	(0.15)	(0.15)	(1)	(0.15)	(0.06)	(0.15)	(1)	(0.15)	(0.06)	(-0.10)	(0.15)
		[0.5]	[0.5]	[0.5]	[0.5]	[0.5]	[0.5]	[0.5]	[0]	[0.5]	[0.5]	[0.5]	[-2]	[0]	[0.5]
17	1	0.969	0.147	0.984	0.150	0.145	0.987	0.151	0.065	0.149	0.857	0.143	0.057	-0.047	0.161
	SD	0.335	0.038	0.533	0.079	0.127	0.649	0.096	0.082	0.176	0.501	0.065	0.061	0.062	0.089
	CR	88.2	96.0	85.5	94.3	94.2	90.7	95.6	90.4	95.7	99.5	98.5	93.2	95.2	99.5
100	1	0.969	0.151	0.961	0.151	0.151	1.027	0.152	0.044	0.147	0.884	0.151	0.057	-0.061	0.164
	SD	0.079	0.009	0.104	0.018	0.032	0.194	0.038	0.017	0.045	0.267	0.078	0.027	0.030	0.050
	CR	91.3	98.8	90.8	98.3	98.6	98.7	92.1	99.0	96.8	99.1	97.1	96.3	95.1	99.8
200	1	0.970	0.150	0.959	0.151	0.150	1.027	0.155	0.046	0.148	0.889	0.151	0.057	-0.063	0.163
	SD	0.054	0.006	0.070	0.012	0.021	0.077	0.015	0.010	0.025	0.262	0.032	0.025	0.028	0.047
	CR	91.3	99.5	91.1	99.1	99.6	97.1	90.7	99.7	98.2	96.9	91.2	95.4	92.5	99.9
400	1	0.971	0.151	0.964	0.152	0.152	1.029	0.156	0.045	0.148	0.873	0.149	0.061	-0.064	0.161
	SD	0.037	0.004	0.049	0.009	0.015	0.061	0.011	0.008	0.018	0.177	0.023	0.019	0.020	0.033
	CR	89.9	99.9	90.2	99.5	99.7	99.1	96.3	99.8	99.6	96.3	90.7	96.7	90.6	99.5
Simulation Study II															
100	1	0.969	0.151	0.961	0.151	0.151	1.027	0.152	0.044	0.147	0.884	0.151	0.057	-0.061	0.164
	SD	0.079	0.009	0.104	0.018	0.032	0.194	0.038	0.017	0.045	0.267	0.078	0.027	0.030	0.050
	CR	91.3	98.8	90.8	98.3	98.6	98.7	92.1	99.0	96.8	99.1	97.1	96.3	95.1	99.8
200	0.5	0.982	0.152	0.963	0.152	0.148	1.025	0.154	0.047	0.145	0.922	0.147	0.057	-0.066	0.167
	SD	0.057	0.007	0.080	0.015	0.028	0.187	0.032	0.022	0.041	0.210	0.024	0.027	0.022	0.032
	CR	94.8	98.4	93.5	98.2	98.8	98.4	94.4	98.2	96.8	95.2	95.3	90.3	88.5	99.9
400	0.25	0.989	0.152	0.972	0.154	0.150	1.027	0.154	0.049	0.147	0.922	0.147	0.062	-0.071	0.170
	SD	0.042	0.005	0.063	0.013	0.025	0.190	0.034	0.023	0.041	0.213	0.025	0.028	0.024	0.033
	CR	93.2	98.8	94.2	98.1	98.7	97.5	92.0	97.1	96.1	88.6	92.8	84.4	90.1	99.5

Table 2: Comparison of three methods of estimation, with data simulated as in part of simulation study II. Averages and standard deviations of the parameters estimates of the proposed method, the methods of Wu et al. (2008) and Ramsay et al. (2007) were reported in rows with headings 'Two-Step', 'Nonlinear Least Squares' and 'Generalized Profile Estimation', respectively.

			Initial Values Set at True Values						Initial Values Different from True Values					
			b known		r known				b known		r known			
n	Δ		α	γ	α	γ	q	v	α	γ	α	γ	q	v
			(1)	(0.15)	(1)	(0.15)	(0.06)	(0.15)	(1)	(0.15)	(1)	(0.15)	(0.06)	(0.15)
			[1]	[0.15]	[1]	[0.15]	[0.06]	[0.15]	[0.5]	[0.5]	[0.5]	[0.5]	[0]	[0.5]
200	0.5	Two-	0.981	0.152	1.025	0.156	0.047	0.145	0.981	0.152	1.016	0.153	0.046	0.143
	SD	Step	0.059	0.007	0.084	0.015	0.015	0.025	0.059	0.007	0.229	0.036	0.023	0.041
400	0.25		0.985	0.152	1.026	0.157	0.049	0.148	0.985	0.152	1.028	0.155	0.048	0.147
	SD		0.043	0.005	0.078	0.013	0.014	0.024	0.043	0.005	0.169	0.029	0.021	0.037
200	0.5	Nonlinear	1.126	0.162	1.476	0.205	0.028	0.251	1.127	0.162	1.445	0.185	0.082	0.458
	SD	Least	0.059	0.005	0.362	0.048	0.059	0.172	0.060	0.005	0.312	0.050	0.052	0.342
400	0.25	Squares	1.059	0.156	1.283	0.176	0.049	0.239	1.060	0.156	1.241	0.181	0.107	0.408
	SD		0.040	0.004	0.250	0.030	0.051	0.177	0.042	0.004	0.115	0.014	0.029	0.136
200	0.5	Generalized	1.018	0.144	1.016	0.146	0.048	0.146	0.535	0.145	2.313	0.020	-0.010	1.509
	SD	Profile	0.036	0.007	0.054	0.005	0.014	0.009	0.010	0.005	0.168	0.005	0.022	0.085
400	0.25	Estimation	1.003	0.145	1.002	0.147	0.051	0.148	0.534	0.162	0.783	0.205	0.010	0.652
	SD		0.018	0.006	0.006	0.006	0.012	0.008	0.012	0.006	0.006	0.007	0.006	0.017

5. SIMULATION STUDIES

In this section, we study the empirical performance of the proposed method for estimating an ODE model via simulation. We simulated data from the SIR model defined by (1) with seasonal birth rate defined by (10). Given the population size $N(t)$, the sample size $m(t)$ was randomly drawn from the binomial distribution Bin($N(t)$,a) where a is the capture probability. Since the asymptotic results remain valid for the case of fixed sampling design that often occurs in practice, we simulated observations at equally-spaced epochs $t = t_1, \ldots, t_n$ where $t_j - t_{j-1} \equiv \Delta, j = 2, \ldots, n$; at time t, $w(t)$ is drawn from the binomial distribution $\text{Bin}(m(t), i(t))$ where $i(t) = I(t)/N(t)$; $y(t) = w(t)/m(t) = i(t) + \epsilon(t)$ is the observed sample proportion of infectives at time t. The w's are conditionally independent given $m(t), I_t/N_t, t = t_1, t_2, \ldots, t_n$. Thus, $\epsilon(t_j), j = 1, \ldots, n$ are independent, of zero mean and variance $\sigma^2(t_j) = i(t_j)\{1 - i(t_j)\}/m(t)$ which is bounded away from 0 and infinity if the i's are bounded away from 0. The true parameter vector $(\alpha, \mu, \gamma, q, r, v)^T = (1, 0.15, 0.15, 0.06, -0.1, 0.15)^T$, with the initial proportion of susceptibles being 0.25 and that of infectives equal to 0.55. The capture probability a is set to be 0.2. The fourth-order Runge-Kutta method was employed to generate the underlying continuous-time process. The discretization step size is set to 1/30 corresponding to 1 day, whereas the proportion of infectives was measured as the sample proportion of infected subjects, once per month, twice a month, or four times a month. Given the observations $\{y(t_j)\}$, we estimated $\boldsymbol{\beta} = (\alpha, \gamma, q, r, v)^T$ by minimizing (14) with F there given by (13). We used the triweight kernel function for local polynomial smoothing. Two kinds of simulation studies were performed. All experiments were replicated 1000 times, unless stated otherwise:

1. Simulation Study I: We fix the time interval between two consecutive observations as 1, i.e. a month, and try various sample sizes including 17, 100, 200, and 400. Since our real data example in Section 6 has 17 observations, $n = 17$ is considered in the simulations.

2. Simulation Study II: The study period is fixed to be $[0, 100]$ with the equal time interval

between two consecutive observations being either 1, 0.5 or 0.25. The sample size equals 100 with time interval 1, 200 with time interval 0.5, and 400 with time interval 0.25.

We summarize in Table 1 the simulation results for the proposed estimation method based on four scenarios: (i) known birth rate b; (ii) known r; and (iii) all parameters unknown. Initial values for the optimization are enclosed by square brackets. Table 1 reports the averages of the estimates, their standard deviations, and the empirical coverage rates of the nominal 95% confidence interval constructed based on the limiting distribution given in Theorem 5. Simulation study I concerns the case of increasing sampling domain asymptotic framework whereas simulation study II concerns the case of infill asymptotic framework for which sampling is increasingly dense over a fixed domain. Strictly speaking, the theoretical results in Section 3 hold only under the infill asymptotic framework in which the consistency of the derivative estimates from the local polynomial regression obtains. Nevertheless, it is of practical interest to examine the performance of the proposed method under the increasing domain asymptotic framework as monitoring studies on prevalence of infectious disease generally take place over equally-spaced epochs.

In preliminary simulation studies, we observed that the high variability in the derivative estimates close to the boundaries significantly drive up the variability and bias of the proposed estimation method. Consequently, we restricted the summation in the objective function (14) to summands with $3.5 \leq t \leq n - 2.5$ based on our experience. This is consistent with the finding in Brunel (2008) showing that excluding the derivative estimates on the boundaries improves the convergence rate in the two-step method with spline-based derivative estimates. Table 1 shows that the bias and sample standard deviation of the estimators generally decrease with increasing sample size. It is harder to estimate the birth rate function because the observations consist of percents of infectives which yield indirect information about the birth rate. Nevertheless, if the parameter r is fixed, the other shape parameter (q) and the overall birth rate (v) appear to be well estimated with reasonable bias and variability. The empirical coverage rates generally get closer to the nominal level with increasing sample size.

We also compare the proposed method with the methods of Wu et al. (2008) and Ramsay et al. (2007) by repeating part of simulation study II, which is reported in Table 2. Since the generalized profile method of Ramsay et al. (2007) is computationally expensive, all results in Table 2 are based on 200 replications. In order to explore the effects of the initial values on optimization, we implemented two sets of initial values for the parameters of the ODE, namely, the true parameter values and values different from the true values. For the method of Wu et al. (2008), the differential equation (13) is numerically solved via the Euler scheme with the initial values of the state vector estimated by local polynomial at $t = 3.5$ because we observed that (i) a more accurate numerical solver, Runge-Kutta scheme, induces a more complex objective function, which causes severe initial value problems on optimization and high variability; (ii) using the initial values of the state vector estimated at boundary points leads to more variability and bias as in the case of the proposed method. Parameter estimates were then obtained by the least squares method based on estimates from $t = 3.5$ to $t = 100$. We implemented the generalized profile estimation method of Ramsay et al. (2007) via the R-package CollocInfer, following the computer code for fitting an example given in Ramsay et al. (2007). The performance of the generalized profile estimation method is very sensitive to initial values. For true initial values, it generally enjoys the smallest bias and variability among the three methods although the proposed method compares well with it. On the other hand, with non-true initial values, it incurs much larger bias and slightly higher variability. Both the proposed method and the method of Wu et al. (2008) are more robust to initial values, although the proposed method consistently outperforms the latter method in terms of smaller bias and variability. Moreover, the computation time for the methods of Wu et al. (2008) and Ramsay et al. (2007) were, respectively, 15–40 times and ≥ 2000 times higher than that of the proposed method. In sum, these simulation results suggest that the proposed method is a computationally quick and yet relatively efficient, robust method for estimating an ODE model with partially observed data.

Table 3: Estimates and 95% confidence intervals of the SIR model fitted to the *Bartonella* data.

	α	γ	q	r	u
$\hat{\theta}$	1.139	0.057	-0.157	-0.201	0.451
95% CI	$(0.337, 1.940)$	$(-0.191, 0.304)$	$(-0.520, 0.206)$	$(-0.290, -0.113)$	$(0.032, 0.871)$

6. PREVALENCE OF BARTONELLA IN A WILD POPULATION OF COTTON RATS

Here, we apply the proposed method for analyzing an infectious time-series data. The data were collected by Kosoy, Mandel, Green, Marston, Jones and Childs (2004*b*) from a marked-capture-recapture study on bartonella prevalence in a cotton rat population in Walton Co., Georgia, USA, over a period of 17 months, from March, 1996 to July, 1997 after conducting a two-month pilot study on February and October, 1995. But trapping was not done in December, 1996. Trapped animals were inspected for whether they were infected by *Bartonella* based on blood tests. As is commonly the case, the *Bartonella* monitoring data contains no information on the recovered and susceptibles. In this application, we fit the SIR model with seasonal birth rate defined by (A.4) assuming seasonal birth rate (10) of period 12, with the time series of monthly bartonella infection rates from October, 1995 to July 1997, with missing data in November 1995 to February 1996 and December 1996. To avoid issues related to vertical transmission of bartonella infection from parent subjects to their children, we excluded 19 newborns in June and July in 1996 from the analysis.

In the left upper diagram of Figure 1, the solid black line plots sample infection rates, the red dashed line plots the fitted values from a local linear fit, and the blue dotted line connects the fitted values from the SIR model defined by (2) with seasonal birth rate (10) estimated by the proposed method (see Table 3); the fitted values from the fitted SIR model was computed via the fourth-order Runge-Kutta method with the initial proportions of susceptibles and

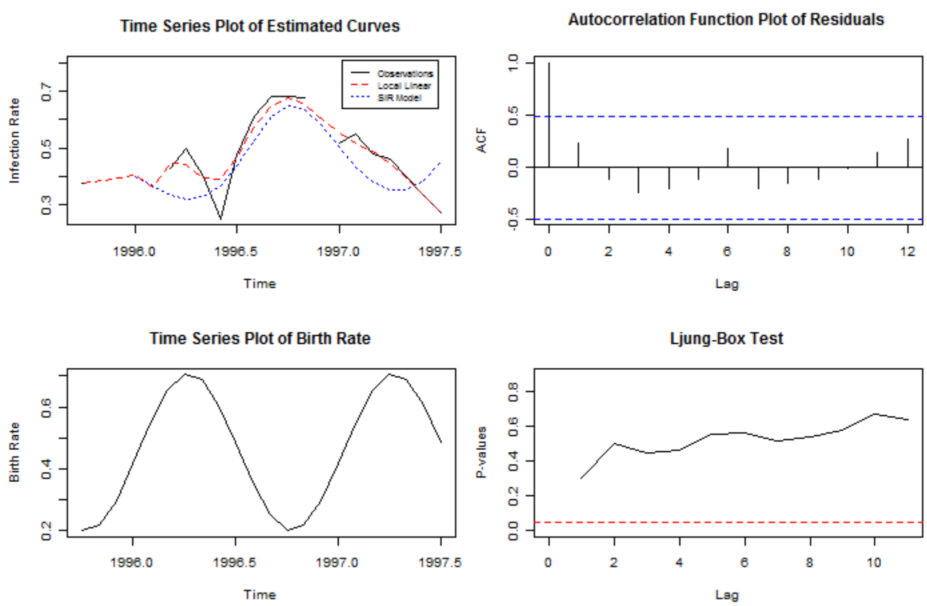

Figure 1: Upper left diagram plot the observed (solid line) fitted values by the local polynomial regression (dashed line) and fitted values from the SIR model (dotted line). Lower left diagram displays the estimated birth rate function. Upper right diagram shows the sample autocorrelation function of the standardized residuals from the fitted SIR model, whereas the lower right digram shows the p-values of Box-Ljung test for the standardized residuals.

infectives on January, 1996 calculated by formulas given in Section 4. The estimated birth rate curve in the left lower diagram of Figure 1 shows that as the birth rate increases, the infection rate decreases and vice versa, which is consistent with intuition.

Table 3 reports the parameter estimates with their 95% confidence intervals obtained by the asymptotic results derived in Theorem 5. For interpretation, 5.7% ($\hat{\gamma}$) of infectives recovered a month after infection on average. The estimate $\hat{\alpha} = 1.139$ is the product of the transmission probability and the number of contacts per month. The birth rate b attained maximum in April and minimum in October. We now assess the goodness of fit of the fitted SIR model, by checking whether or not the standardized residuals are approximately white noise, i.e. uncorrelated and of constant variance. The residuals are defined as the observed values minus the fitted values. The residuals are standardized by normalizing them by their standard deviations. The right upper diagram of Figure 1 plots the autocorrelation function of the standardized residuals; none of them are significantly different from 0 at 5% level. The white-noise assumption can be further tested by checking whether the first k residual autocorrelations all equal to zero by the Ljung-Box test. The right lower diagram of Figure 1 reports the the p-values of the Ljung-Box tests for $k = 1, \ldots, 11$, showing that the white-noise assumption cannot be rejected at 5% significance level. These plots suggest that there are no residual autocorrelations, and hence the model fits the data well.

7. CONCLUSION

We show that, under some conditions, an ODE model may admit an equivalent state vector in terms of the observed state component and its higher derivatives; this observation forms the basis of our proposed method for estimating an ODE model by conditional least squares with the higher derivatives first estimated via local polynomial regression. The proposed method enjoys desirable large-sample properties. Both simulation studies and the real application demonstrate the usefulness of the proposed method.

Here, we mention several interesting future research directions. While we consider the case of ODE, the proposed method may be readily extended to the case when the state process is driven by some partial differential equation. It is of interest to study the theoretical properties of such extension. Second, in practice, the choice of the bandwidth is of importance. The limiting distribution obtained in Section 3 may be useful for studying a more rigorous optimal choice of the bandwidth. For the *Bartonella* data, the observed monthly counts of infectives are binomially distributed, but this information is not explicitly used in the estimation of the derivatives via local polynomial regression. It is of interest to study possible gain by estimating the derivatives via local generalized linear models using Fan, Heckman and Wand (1995). Furthermore, simulations suggest that omitting several derivative estimates around the sampling boundaries improved the empirical convergence rate of the proposed estimation scheme. Although this observation is consistent with the theoretical result of Brunel (2008) that omitting the two extreme boundary derivative estimates, from spline smoothing, improves the convergence rate of the two-step method, it is of interest to investigate the impact of omitting multiple boundary derivative estimates on the convergence rate of the proposed method.

ACKNOWLEDGEMENTS

This work was partially supported by the U.S. National Science Foundation.

REFERENCES

Brunel, N. J.-B. (2008), "Parameter estimation of ODE's via nonparametric estimators," *Electronic Journal of Statistics*, 2, 1242–1267.

Diekmann, O., and Heesterbeek, J. A. P. (2000), *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, West Sussex: John Wiley & Sons Ltd.

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

Fan, J., Heckman, N. E., and Wand, M. P. (1995), "Local polynomial kernel regression for generalized linear models and quasi-likelihood functions," *Journal of American Statistical Association*, 90, 141–150.

Farlow, S. J. (1993), *Partial Differential Equations for Scientists and Engineers*, New York: Dover Publications.

He, D., and Earn, D. J. D. (2007), "Epidemiological effects of seasonal oscillations in birth rates," *Theoretical Population Biology*, 72(2), 274 – 291.

Hunter, L. R., Su, R., and Meyer, G. (1983), "Global transformations of nonlinear systems," *IEEE Transactions on Automatic Control*, 28, 24–31.

Ireland, J. M., Mestel, B. D., and Norman, R. A. (2007), "The effect of seasonal host birth rates on disease persistence," *Mathematical Biosciences*, 206(1), 31 – 45. Alcala Special Issue.

Kantz, H., and Schreiber, T. (2003), *Nonlinear Time Series Analysis (2nd ed.)*, Cambridge: Cambridge University Press.

Kosoy, M., Mandel, E., Green, D., Marston, E., Jones, D., and Childs, J. (2004*a*), "Prospective studies of bartonella of rodents. Part II. Diverse infections in a single rodent community," *Vector-Borne and Zoonotic Diseases*, 4, 296–305.

Kosoy, M. Y., Mandel, E. L., Green, D. C., Marston, E. L., Jones, D. C., and Childs, J. E. (2004*b*), "Prospective studies of *Bartonella* of rodents. Part I. Demographic and temporal patterns in population dynamics," *Vector-Borne and Zoonotic Disease*, 4, 285–295.

Krantz, S. G., and Parks, H. R. (2002), *The Implicit Function Theorem: History, Theory, and Applications*, Boston: Birkhäuser.

Liang, H., and Wu, H. (2008), "Parameter estimation for differential equation models using a framework of measurement error in regression models," *Journal of American Statistical Association*, 103, 1570–1583.

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., 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*, 69, 741–796.

Simon, D. (2006), *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches*, New Jersey: John Wiley and Sons, Inc.

Terrell, W. J. (1999), "Some Fundamental Control Theory II: Feedback Linearization of Single Input Nonlinear Systems," *The American Mathematical Monthly*, 106, 812–828.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge: Cambridge University Press.

Wu, H., Zhu, H., Miao, H., and Perelson, A. S. (2008), "Parameter Identifiability and Estimation of HIV/AIDS dynamic Models," *Bulletin of Mathematical Biology*, 70, 785–799.

Xia, X., and Moog, C. H. (2003), "Identifiability of Nonlinear Systems With Application to HIV/AIDS Models," *IEEE Transactions on Automatic Control*, 48, 330–336.

Zhang, J., Jin, Z., Xue, Y., and Li, Y. (2009), "The Dynamics of the Pulse Birth in an SIR Epidemic Model with Standard Incidence," *Discrete Dynamics in Nature and Society*, p. doi:10.1155/2009/490437.

APPENDIX A. EXTENSION TO THE CASE OF OBSERVING MULTIPLE COMPONENTS OF THE STATE PROCESS

The basic idea is to augment the observable components with enough number of their higher derivatives to make an equivalent state vector. Consider the general case with the state equation driven by the following ODE:

$$\frac{d\boldsymbol{X}_0(t)}{dt} = Z(\boldsymbol{X}_0(t), \boldsymbol{\beta}), \tag{A.1}$$

Let the original state vector be $\boldsymbol{X}_0(t) = (X_1(t), \ldots, X_k(t))^T$, some of whose components are unobservable. Without loss of generality, assume that $\boldsymbol{X}_{obs}(t) = (X_1(t), \ldots, X_u(t))^T$ is observed up to some additive error, but $\boldsymbol{X}_{mis}(t) = (X_{u+1}(t), \ldots, X_k(t))^T$ is unobservable, where $1 \leq u < k$. Partition (A.1) according to $\boldsymbol{X}_{obs}$ and $\boldsymbol{X}_{mis}$, and let $Z_{obs}(\boldsymbol{X}_0, \boldsymbol{\beta})$ be the right hand side of (A.1) corresponding to $\boldsymbol{X}_{obs}$ so that $\boldsymbol{X}'_{obs}(t) = Z_{obs}(\boldsymbol{X}_0(t), \boldsymbol{\beta})$. To solve $\boldsymbol{X}_{mis}(t)$ in terms of $\boldsymbol{X}_{obs}(t)$, its derivatives, and $\boldsymbol{\beta}$, we set up the following equations:

$$\boldsymbol{X}_{obs}^{(1)}(t) = Z_{obs}(\boldsymbol{X}_0(t), \boldsymbol{\beta}), \ldots, \boldsymbol{X}_{obs}^{(\ell)}(t) = Z_{obs}^{(\ell-1)}(\boldsymbol{X}_0^{(i)}(t), i = 0, \ldots, \ell-1, \boldsymbol{\beta}), \tag{A.2}$$

where $\boldsymbol{X}^{(i)}$ denotes the ith derivative vector of $\boldsymbol{X}$ at time t. To determine ℓ, notice that all derivatives of $\boldsymbol{X}_0(t)$ is a function of $\boldsymbol{X}_0(t)$ given $\boldsymbol{\beta}$. For example, it follows from the chain rule that $\boldsymbol{X}_0^{(2)}(t)$ is a function of $\boldsymbol{X}_0(t)$ and $\boldsymbol{X}_0^{(1)}(t)$, and since $\boldsymbol{X}_0^{(1)}(t)$ is a function of $\boldsymbol{X}_0(t)$, $\boldsymbol{X}_0^{(2)}(t)$ is a function of $\boldsymbol{X}_0(t)$. Consequently, the right side of each equation in (A.2) depends only on $\boldsymbol{X}_0(t)$ and $\boldsymbol{\beta}$. Because u components of $\boldsymbol{X}_0$ are observable while the remaining $k-u$ components are unobservable, hence $k-u$ equations are generally needed to solve for the $k-u$ unknowns, resulting in

$$\ell = \left\lceil \frac{k-u}{u} \right\rceil,$$

where $\lceil x \rceil$ represents the ceiling function of x. We assume that the first $k-u$ equations in (A.2) uniquely determine $\boldsymbol{X}_{mis}$, given $\boldsymbol{\beta}$, and hence also all its higher derivatives. Then write the solutions $\boldsymbol{X}_{mis}(t) = R(\boldsymbol{X}_{obs}^{(i)}(t), i = 0, \ldots, \ell, \boldsymbol{\beta})$ for some function R. Differentiate the last equation of (A.2) followed by substituting in it the preceding expression for $\boldsymbol{X}_{mis}(t)$

to yield

$$\boldsymbol{X}_{obs}^{(\ell+1)}(t) = Z_{obs}^{(\ell)}(\boldsymbol{X}_{obs}^{(i)}(t), i = 0, \ldots, \ell, \boldsymbol{\beta}) \equiv F(\boldsymbol{X}_{obs}^{(i)}(t), i = 0, \ldots, \ell, \boldsymbol{\beta}). \tag{A.3}$$

Finally, we propose to estimate $\boldsymbol{\beta}$ by minimizing the total conditional sum of squared errors:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} |\hat{\boldsymbol{X}}_{obs}^{(\ell+1)}(t_i) - F(\hat{\boldsymbol{X}}_{obs}^{(j)}(t_i), j = 0, \ldots, \ell, \boldsymbol{\beta})|^2,$$

where $\hat{\boldsymbol{X}}_{obs}^{(j)}(t_i), j = 0, \ldots, \ell + 1$ are obtained from local polynomial fitting based on data at epochs $t_i, i = 1, \ldots, n$.

APPENDIX B. VALIDITY OF ASSUMPTIONS 1-3 FOR THE SIR MODEL

Consider

$$\begin{aligned} s' &= -\alpha s i + (1 - s) b, \\ i' &= \alpha s i - (b + \gamma) i. \end{aligned} \tag{A.4}$$

For Assumption 1, (i) is clearly satisfied whereas (ii) generally holds with a suitable kernel function. The bandwidth condition in (iii) can be readily implemented and the density condition in (iv) refers to the sampling protocol. Part (i) of Assumption 2 is clearly satisfied if we restrict the parameter space is a sufficiently large, relatively compact set. Part (ii) there presupposes the existence of an ergodic stationary solution; this is a difficult problem, but see Ireland, Mestel and Norman (2007) and He and Earn (2007) for conditions under which the SIR model with seasonal birth rate admits a chaotic solution and hence an ergodic stationary solution. We now show that Assumption 3 holds, under some regularity conditions. We assume that i admits a positive lower bound, see Zhang, Jin, Xue and Li (2009) for relevant discussion. We verify Assumption 3(i) first. Let $\boldsymbol{\beta} = (\alpha, \gamma, q, r, v)^T$ and recall $\boldsymbol{X} = (i, i')^T$, and that

$$F(\boldsymbol{X}, \boldsymbol{\beta}) = (i' + bi + \gamma i)(-\alpha i - b) + \frac{i'^2}{i} + (\alpha b - b')i.$$

Then,

$$\frac{\partial F(\boldsymbol{X},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \begin{pmatrix} -(i^{'}+bi+\gamma i)i+bi \\ -\alpha i^2 - bi \\ -\alpha \sin\left(\frac{\pi}{6}t\right)i^2 + \left\{(-2b-\gamma+\alpha)\sin\left(\frac{\pi}{6}t\right) - \frac{\pi}{6}\cos\left(\frac{\pi}{6}t\right)\right\}i - \sin\left(\frac{\pi}{6}t\right)i^{'} \\ -\alpha \cos\left(\frac{\pi}{6}t\right)i^2 + \left\{(-2b-\gamma+\alpha)\cos\left(\frac{\pi}{6}t\right) + \frac{\pi}{6}\sin\left(\frac{\pi}{6}t\right)\right\}i - \cos\left(\frac{\pi}{6}t\right)i^{'} \\ -\alpha i^2 + (-2b-\gamma+\alpha)i - i^{'} \end{pmatrix}.$$

Thus,

$$\Big|\frac{\partial F(\boldsymbol{X}_{\nu_1},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \frac{\partial F(\boldsymbol{X}_{\nu_2},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|$$
$$\leq d_1|i_{\nu_1}-i_{\nu_2}| + d_2|i^{'}_{\nu_1}-i^{'}_{\nu_2}| + d_3|i^2_{\nu_1}-i^2_{\nu_2}| = d_1|i_{\nu_1}-i_{\nu_2}| + d_2|i^{'}_{\nu_1}-i^{'}_{\nu_2}| + d_3(i_{\nu_1}+i_{\nu_2})|i_{\nu_1}-i_{\nu_2}|,$$

where

$$\begin{aligned} d_1 &= |i^{'}_{\nu_1}| + 2b + \left|(-2b-\gamma+\alpha)\sin\left(\frac{\pi}{6}t\right) + \frac{\pi}{6}\cos\left(\frac{\pi}{6}t\right)\right| \\ &\quad + \left|(-2b-\gamma+\alpha)\cos\left(\frac{\pi}{6}t\right) - \frac{\pi}{6}\sin\left(\frac{\pi}{6}t\right)\right| + |-2b-\gamma+\alpha|, \\ d_2 &= |i_{\nu_2}| + \left|\sin\left(\frac{\pi}{6}t\right) + \cos\left(\frac{\pi}{6}t\right) + 1\right|, \\ d_3 &= (b+\gamma+2\alpha) + \left|\alpha\sin\left(\frac{\pi}{6}t\right)\right| + \left|\alpha\cos\left(\frac{\pi}{6}t\right)\right|. \end{aligned}$$

From (A.4), $i^{'}$ is a function of s and i, and s and i are bounded, so $i^{'}$ is bounded. Thus, d_1, d_2, and $d_3(i_{\nu_1}+i_{\nu_2})$ are bounded, so Assumption 3(i) is satisfied. Now,

$$\frac{\partial F(\boldsymbol{X},\boldsymbol{\beta})}{\partial \boldsymbol{X}} = \begin{pmatrix} (b+\gamma)(-\alpha i - b) - \alpha(i^{'}+bi+\gamma i) - \frac{i^{'2}}{i^2} + (\alpha b - b^{'}) \\ -\alpha i - b + 2\frac{i^{'}}{i} \end{pmatrix}.$$

Since i has a positive lower bound, Assumption 3(ii) is also satisfied.

APPENDIX C. PROOF OF THEOREM 2

For simplicity, we assume that t_i's are independent and identically distributed and $\sigma_1^2 = \cdots = \sigma_n^2 \equiv \sigma_\epsilon^2$, in the following proofs.

C.1 Proof of Consistency

Let $\delta(t_i) = \hat{X}^{(p)}(t_i) - X^{(p)}(t_i) = \hat{X}^{(p)}(t_i) - F(\boldsymbol{X}(t_i), \boldsymbol{\beta}_0)$. First, we have

$$
\begin{aligned}
&\sum_{i=1}^{n} \{\hat{X}^{(p)}(t_i) - F(\hat{\boldsymbol{X}}(t_i), \boldsymbol{\beta})\}^2 \\
&= \sum_{i=1}^{n} \{F(\boldsymbol{X}(t_i), \boldsymbol{\beta}_0) + \delta(t_i) - F(\hat{\boldsymbol{X}}(t_i), \boldsymbol{\beta})\}^2 \\
&= \sum_{i=1}^{n} \{F(\boldsymbol{X}(t_i), \boldsymbol{\beta}_0) - F(\hat{\boldsymbol{X}}(t_i), \boldsymbol{\beta})\}^2 + \sum_{i=1}^{n} \delta^2(t_i) \\
&\quad + 2\sum_{i=1}^{n} \{F(\boldsymbol{X}(t_i), \boldsymbol{\beta}_0) - F(\hat{\boldsymbol{X}}(t_i), \boldsymbol{\beta})\}\delta(t_i).
\end{aligned}
\tag{A.5}
$$

From Theorem 3.1 of Fan and Gijbels (1996), we have, for any $t_0 \in \mathcal{T}$

$$
\begin{aligned}
\mathrm{E}(\hat{X}^{(j)}(t_0)) - X^{(j)}(t_0) &= \xi_{j+1}^T S_{j+1}^{-1} c_{j+1} \frac{X^{(j+1)}(t_0)}{j+1} h_n^2 + o_p(h_n^2), \\
\mathrm{Var}(\hat{X}^{(j)}(t_0)) &= \xi_{j+1}^T S_{j+1}^{-1} S_{j+1}^* S_{j+1}^{-1} \xi_{j+1} \frac{j!^2 \sigma^2(t_0)}{f(t_0) n h_n^{2j+1}} + o_p\Big(\frac{1}{n h_n^{2j+1}}\Big),
\end{aligned}
\tag{A.6}
$$

where the $o_p(\cdot)$ terms hold uniformly for $t_0 \in \mathcal{T}$. Thus, $\hat{X}^{(j)}(t_0) - X^{(j)}(t_0) = O_p(h_n^2 + 1/\sqrt{n h_n^{2j+1}}) = O_p(h_n^2 + 1/\sqrt{n h_n^{2p+1}}) \equiv O_p(b_n)$ uniformly over $\mathcal{T}$ for $j = 0, \ldots, p$, which is $o_p(1)$ provided that $h_n \to 0$ and $n h_n^{2j+1} \to \infty$, $j = 0, \ldots, p$. The second term of the last equation of (A.5) is hence $O_p(n b_n^2)$. Consider the first term there, which can be decomposed as

$$
\begin{aligned}
&\sum_{i=1}^{n} \{F(\boldsymbol{X}(t_i), \boldsymbol{\beta}_0) - F(\hat{\boldsymbol{X}}(t_i), \boldsymbol{\beta})\}^2 \\
&= \sum_{i=1}^{n} \{F(\boldsymbol{X}(t_i), \boldsymbol{\beta}_0) - F(\boldsymbol{X}(t_i), \boldsymbol{\beta})\}^2 + \sum_{i=1}^{n} \{F(\boldsymbol{X}(t_i), \boldsymbol{\beta}) - F(\hat{\boldsymbol{X}}(t_i), \boldsymbol{\beta})\}^2 \\
&\quad + 2\sum_{i=1}^{n} \{F(\boldsymbol{X}(t_i), \boldsymbol{\beta}_0) - F(\boldsymbol{X}(t_i), \boldsymbol{\beta})\}\{F(\boldsymbol{X}(t_i), \boldsymbol{\beta}) - F(\hat{\boldsymbol{X}}(t_i), \boldsymbol{\beta})\}.
\end{aligned}
$$

By Assumption 3(i) and the mean value theorem, the second term of the preceding equation can be bounded as follows:

$$
\begin{aligned}
&\sum_{i=1}^{n} \{F(\boldsymbol{X}(t_i), \boldsymbol{\beta}) - F(\hat{\boldsymbol{X}}(t_i), \boldsymbol{\beta})\}^2 = \sum_{i=1}^{n} |\boldsymbol{X}(t_i) - \hat{\boldsymbol{X}}(t_i)|^2 \Big|\frac{\partial F(\tilde{\boldsymbol{x}}_i, \boldsymbol{\beta})}{\partial \boldsymbol{x}}\Big|^2 \\
&\leq M^2 \sum_{i=1}^{n} |\boldsymbol{X}(t_i) - \hat{\boldsymbol{X}}(t_i)|^2 = O_p(n b_n^2),
\end{aligned}
\tag{A.7}
$$

where $\tilde{\boldsymbol{x}}_i$ lies between $\boldsymbol{X}(t_i)$ and $\hat{\boldsymbol{X}}(t_i)$. Due to the continuity of F and the compactness of $\mathcal{T}$ and $\Omega_{\boldsymbol{\beta}}$, $F(\boldsymbol{X}(t_i), \boldsymbol{\beta}_0) - F(\boldsymbol{X}(t_i), \boldsymbol{\beta})$ is uniformly bounded. Thus, similar to the derivation of (A.7), we have

$$\begin{aligned} &\sum_{i=1}^{n}[\{F(\boldsymbol{X}(t_i), \boldsymbol{\beta}_0) - F(\boldsymbol{X}(t_i), \boldsymbol{\beta})\}\{F(\boldsymbol{X}(t_i), \boldsymbol{\beta}) - F(\hat{\boldsymbol{X}}(t_i), \boldsymbol{\beta})\}] \\ &= O_p(nb_n) = o_p(n). \end{aligned} \tag{A.8}$$

Finally, the third term of (A.5) can be similarly shown to be $O_p(nb_n)$.

Note that Assumptions 1(i), 2(i) and 3(i) imply that $\forall t \in \mathcal{T}, \forall \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \Omega_{\boldsymbol{\beta}}$, there exists a constant K such that

$$|F(\boldsymbol{X}(t), \boldsymbol{\beta}_1) - F(\boldsymbol{X}(t), \boldsymbol{\beta}_2)| \le K|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2|. \tag{A.9}$$

Hence the class of functions $\{F(\boldsymbol{X}(t), \boldsymbol{\beta}), \boldsymbol{\beta} \in \Omega_{\boldsymbol{\beta}}\}$ is P-Glivenko-Cantelli so that that the uniform law of large numbers hold; c.f. van der Vaart (1998, Example 19.7). The uniform law of large numbers then entails the almost sure convergence of

$$\frac{1}{n}\sum_{i=1}^{n}\{F(\boldsymbol{X}(t_i), \boldsymbol{\beta}_0) - F(\boldsymbol{X}(t_i), \boldsymbol{\beta})\}^2 \to D(\boldsymbol{\beta}_0, \boldsymbol{\beta}), \tag{A.10}$$

uniformly in $\boldsymbol{\beta}$, where $D(\boldsymbol{\beta}_0, \boldsymbol{\beta}) = E\{F(\boldsymbol{X}(t), \boldsymbol{\beta}_0) - F(\boldsymbol{X}(t), \boldsymbol{\beta})\}^2$. Consistency then follows from (A.7)-(A.10) and Assumption 2(ii).

C.2 Proof of the Asymptotic Distribution

Note that, under the assumptions of continuous derivatives and using the mean-value theorem, we have

$$\mathbf{0} = \frac{\partial S(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} = \frac{\partial S(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} + \frac{\partial^2 S(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0),$$

where $\partial S(\boldsymbol{\beta}^*)/\partial \boldsymbol{\beta}$ equals $\partial S(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$ evaluated at $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ and $\tilde{\boldsymbol{\beta}}$ lies between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. Then, we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = -\Big\{\frac{\partial^2 S(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\Big\}^{-1}\frac{\partial S(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}. \tag{A.11}$$

Consider $\partial S(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$, which can be expressed as

$$\frac{\partial S(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = -2\sum_{i=1}^{n}\{F(\boldsymbol{X}(t_i),\boldsymbol{\beta}) - F(\hat{\boldsymbol{X}}(t_i),\boldsymbol{\beta})\}\frac{\partial F(\hat{\boldsymbol{X}}(t_i),\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} - 2\sum_{i=1}^{n}\delta(t_i)\frac{\partial F(\hat{\boldsymbol{X}}(t_i),\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}$$
$$\equiv I_{1n} + I_{2n}.$$

Due to the continuity of $\partial F(\boldsymbol{X}(t_i),\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ and the compactness of $\mathcal{T}$ and $\Omega_{\boldsymbol{\beta}}$, $|\partial F(\boldsymbol{X}(t_i),\boldsymbol{\beta})/\partial\boldsymbol{\beta}|$ is bounded by some constant, say, d_0. Using the mean value theorem and Assumption 3(i), we obtain

$$\begin{aligned}
I_{1n} &= -2\sum_{i=1}^{n}\{F(\boldsymbol{X}(t_i),\boldsymbol{\beta}) - F(\hat{\boldsymbol{X}}(t_i),\boldsymbol{\beta})\}\frac{\partial F(\hat{\boldsymbol{X}}(t_i),\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\\
&= -2\sum_{i=1}^{n}\{F(\boldsymbol{X}(t_i),\boldsymbol{\beta}) - F(\hat{\boldsymbol{X}}(t_i),\boldsymbol{\beta})\}\Big\{\frac{\partial F(\hat{\boldsymbol{X}}(t_i),\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} - \frac{\partial F(\boldsymbol{X}(t_i),\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} + \frac{\partial F(\boldsymbol{X}(t_i),\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big\}\\
&\le 2dM\sum_{i=1}^{n}|\boldsymbol{X}(t_i) - \hat{\boldsymbol{X}}(t_i)|^{1+\zeta} + 2d_0\sum_{i=1}^{n}|\boldsymbol{X}(t_i) - \hat{\boldsymbol{X}}(t_i)|\\
&= O_p(nb_n^{1+\zeta}) + O_p(nb_n) = O_p(nb_n),
\end{aligned}$$

where recall $0 < \zeta \le 1$. Define

$$\boldsymbol{T}_{j,t} = \begin{pmatrix} 1 & (t_1 - t) & \cdots & (t_1 - t)^j \\ \vdots & \vdots & & \vdots \\ 1 & (t_n - t) & \cdots & (t_n - t)^j \end{pmatrix}.$$

Write K_i for $K_{h_n}(t_i - t)$ and let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$. Then, the $n \times n$ matrix of weights $\boldsymbol{W}$ becomes

$$\boldsymbol{W} = \mathrm{diag}\{K_1, \ldots, K_n\}.$$

It can be readily seen that I_{2n} can be expressed as

$$-2\sum_{i=1}^{n}\Big\{\frac{\partial F(\hat{\boldsymbol{X}}(t_i),\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} - \frac{\partial F(\boldsymbol{X}(t_i),\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big\}\delta(t_i) - 2\sum_{i=1}^{n}\frac{\partial F(\boldsymbol{X}(t_i),\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\delta(t_i). \tag{A.12}$$

By Assumption 3(i), the first term is $O_p(nb_n{}^{1+\zeta})$. Write $\boldsymbol{F}'_{p-1,i} = \partial F\{\boldsymbol{X}(t_i),\boldsymbol{\beta}\}/\partial\boldsymbol{\beta}$ for $i = 1, \ldots, n$, and $\boldsymbol{F}_{p-1} = (F'_{p-1,1}, \ldots, F'_{p-1,n})^T$. The second term of (A.12) can be expressed as $\boldsymbol{F}_{p-1}^T\{\delta(t_1), \ldots, \delta(t_n)\}^T$. Recall $\boldsymbol{\xi}_{p+1}$ stands for the $(p+2) \times 1$ unit vector having 1 in the $(p+1)$th entry. Let $\boldsymbol{\xi}_{p+1,t} = p!\xi_{p+1}^T(\boldsymbol{T}_{p+1,t}^T\boldsymbol{W}_t\boldsymbol{T}_{p+1,t})^{-1}\boldsymbol{T}_{p+1,t}^T\boldsymbol{W}_t$ and $\boldsymbol{\Xi}_{p+1} =$

$(\boldsymbol{\xi}_{p+1,t_1}^T, \ldots, \boldsymbol{\xi}_{p+1,t_n}^T)^T$. Letting $\boldsymbol{X}_t^{(p)} = (X^{(p)}(t_1), \ldots, X^{(p)}(t_n))^T$, $\boldsymbol{X}_t = (X(t_1), \ldots, X(t_n))^T$ and $\boldsymbol{\epsilon} = (\epsilon(t_1), \ldots, \epsilon(t_n))^T$, we have

$$\{\delta(t_1), \ldots, \delta(t_n)\}^T = \boldsymbol{\Xi}_{p+1}\boldsymbol{Y} - \boldsymbol{X}_t^{(p)} = \boldsymbol{\Xi}_{p+1}\boldsymbol{X}_t - \boldsymbol{X}_t^{(p)} + \boldsymbol{\Xi}_{p+1}\boldsymbol{\epsilon}.$$

From (A.6), we have

$$\boldsymbol{F}_{p-1}^T\{\boldsymbol{\Xi}_{p+1}\boldsymbol{X}_t - \boldsymbol{X}_t^{(p)}\} = \boldsymbol{C}nh_n^2 + o_p(nb_n), \tag{A.13}$$

where

$$\boldsymbol{C} = \frac{\boldsymbol{\xi}_{p+1}^T\boldsymbol{S}_{p+1}^{-1}\boldsymbol{c}_{p+1}}{p+1}E\Big(X^{(p+1)}(t)\frac{\partial F\{\boldsymbol{X}(t), \boldsymbol{\beta}\}}{\partial \boldsymbol{\beta}}\Big).$$

Furthermore, $2\boldsymbol{F}_{p-1}^T\boldsymbol{\Xi}_{p+1}\boldsymbol{\epsilon}$ is a sum of weighted independent variables $\{\epsilon_i, i = 1, \ldots, n\}$ with mean $\boldsymbol{0}$ and covariance matrix of the form

$$4\boldsymbol{F}_{p-1}^T\boldsymbol{\Xi}_{p+1}Cov(\boldsymbol{\epsilon})\boldsymbol{\Xi}_{p+1}^T\boldsymbol{F}_{p-1}.$$

Let $\boldsymbol{H} = \text{diag}(1, h_n, \ldots, h_n^{p+1})$. Noting $Cov(\boldsymbol{\epsilon}) = \sigma_\epsilon^2 I$, consider the (i, j)th entry of $\boldsymbol{\Xi}_{p+1}\boldsymbol{\Xi}_{p+1}^T$ denoted by η_{ij}:

$$\eta_{ij} = p!^2\boldsymbol{\xi}_{p+1}^T(\boldsymbol{T}_{p+1,t_i}^T\boldsymbol{W}_t\boldsymbol{T}_{p+1,t_i})^{-1}\boldsymbol{T}_{p+1,t_i}^T\boldsymbol{W}_{t_i}\boldsymbol{W}_{t_j}\boldsymbol{T}_{p+1,t_j}(\boldsymbol{T}_{p+1,t_j}^T\boldsymbol{W}_t\boldsymbol{T}_{p+1,t_j})^{-1}\boldsymbol{\xi}_{p+1}.$$

When $i = j$, from Section 3.7 of Fan and Gijbels (1996), we have

$$\begin{aligned}&(\boldsymbol{T}_{p+1,t_i}^T\boldsymbol{W}_{t_i}\boldsymbol{T}_{p+1,t_i})^{-1}\boldsymbol{T}_{p+1,t_i}^T\boldsymbol{W}_{t_i}\boldsymbol{W}_{t_i}\boldsymbol{T}_{p+1,t_i}(\boldsymbol{T}_{p+1,t_i}^T\boldsymbol{W}_{t_i}\boldsymbol{T}_{p+1,t_i})^{-1}\\ &= \frac{1}{f(t_i)nh_n}\boldsymbol{H}^{-1}\boldsymbol{S}_{p+1}^{-1}\boldsymbol{S}_{p+1}^*\boldsymbol{S}_{p+1}^{-1}\boldsymbol{H}^{-1}(1 + o_p(1)).\end{aligned}$$

Thus, we have

$$\eta_{ii} = \frac{p!^2}{f(t_i)nh_n^{2p+1}}\boldsymbol{\xi}_{p+1}^T\boldsymbol{S}_{p+1}^{-1}\boldsymbol{S}_{p+1}^*\boldsymbol{S}_{p+1}^{-1}\boldsymbol{\xi}_{p+1}(1 + o_p(1)),$$

which holds uniformly over $\mathcal{T}$. When $i \neq j$, the (r, s)th entry of $\boldsymbol{T}_{p+1,t_i}^T\boldsymbol{W}_{t_i}\boldsymbol{W}_{t_j}\boldsymbol{T}_{p+1,t_j}$ equals

$$G_{rs} = \frac{1}{h_n^2}\sum_{k=1}^n K\Big(\frac{t_k - t_i}{h_n}\Big)(t_k - t_i)^{r-1}K\Big(\frac{t_k - t_j}{h_n}\Big)(t_k - t_j)^{s-1}.$$

Without loss of generality, assume $t_i \neq t_j$ if $i \neq j$. Upon substituting $(t_k - t_i)/h_n$ by u, we get

$$E(G_{rs}) = nh_n^{r+s-3}\int K(u)u^{r-1}K\Big(u + \frac{t_i - t_j}{h_n}\Big)\Big(u + \frac{t_i - t_j}{h_n}\Big)^{s-1}f(t_i + h_nu)du.$$

Using Taylor expansion and the fact that $h_n \to 0$ as $n \to \infty$, we have

$$f(t_i + h_n u) = f(t_i) + o(1). \tag{A.14}$$

Consider

$$\int K(u)u^{r-1}K\Big(u + \frac{t_i - t_j}{h_n}\Big)\Big(u + \frac{t_i - t_j}{h_n}\Big)^{s-1} du.$$

Since the support of K is $[-1, 1]$,

$$\Big|\Big(u + \frac{t_i - t_j}{h_n}\Big)^{s-1}\Big| \le 1. \tag{A.15}$$

In addition, by Assumption 1(i), $K(u+(t_i-t_j)/h_n) = O(h_n)$. Combining (A.14) and (A.15), we have

$$E(G_{rs}) = O(nh_n^{r+s-2}).$$

Applying similar arguments yields

$$\sqrt{Var(G_{rs})} = O(nh_n^{r+s-2}).$$

Then,

$$G_{rs} = E(G_{rs}) + O_p(\sqrt{Var(G_{rs})}) = O_p(nh_n^{r+s-2}).$$

Therefore, for some matrix $\boldsymbol{S}^{**}$ independent of h_n and n,

$$\boldsymbol{T}_{p+1,t_i}^T \boldsymbol{W}_{t_i} \boldsymbol{W}_{t_j} \boldsymbol{T}_{p+1,t_j} = O_p(n\boldsymbol{H}\boldsymbol{S}^{**}\boldsymbol{H}) = o_p(\frac{n}{h_n}\boldsymbol{H}\boldsymbol{S}^{**}\boldsymbol{H}). \tag{A.16}$$

It follows from Section 3.7 of Fan and Gijbels (1996) that

$$\boldsymbol{T}_{p+1,t_i}^T \boldsymbol{W}_{t_i} \boldsymbol{T}_{p+1,t_i} = nf(t_i)\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}(1 + o_p(1)). \tag{A.17}$$

Combining (A.16) and (A.17),

$$\eta_{ij} = o_p(\frac{1}{nh_n^{2p+1}}),$$

uniformly over $\mathcal{T}$. Thus, we have

$$\boldsymbol{\Xi}_{p+1} Cov(\boldsymbol{\epsilon}) \boldsymbol{\Xi}_{p+1}^T = \frac{p!^2\sigma_\epsilon^2}{nh_n^{2p+1}} \boldsymbol{\xi}_{p+1}^T \boldsymbol{S}_{p+1}^{-1} \boldsymbol{S}_{p+1}^* \boldsymbol{S}_{p+1}^{-1} \boldsymbol{\xi}_{p+1} \boldsymbol{A}_f (1 + o_p(1)),$$

where $\boldsymbol{A}_f = \text{diag}(1/f(t_1), \ldots, 1/f(t_n))$. Consequently,

$$\begin{aligned}
&\boldsymbol{F}_{p-1}^T \boldsymbol{\Xi}_{p+1} Cov(\boldsymbol{\epsilon}) \boldsymbol{\Xi}_{p+1}^T F_{p-1} \\
&= \frac{p!^2 \sigma_\epsilon^2}{h_n^{2p+1}} \boldsymbol{\xi}_{p+1}^T \boldsymbol{S}_{p+1}^{-1} \boldsymbol{S}_{p+1}^* \boldsymbol{S}_{p+1}^{-1} \boldsymbol{\xi}_{p+1} E\Big[\Big(\frac{\partial F(\boldsymbol{X}(t), \boldsymbol{\beta})}{\sqrt{f(t)}\partial \boldsymbol{\beta}}\Big)\Big(\frac{\partial F(\boldsymbol{X}(t), \boldsymbol{\beta})}{\sqrt{f(t)}\partial \boldsymbol{\beta}}\Big)^T\Big] \\
&\quad + o_p(\frac{1}{h_n^{2p+1}}).
\end{aligned} \tag{A.18}$$

On the other hand, we have

$$\begin{aligned}
\frac{1}{n}\frac{\partial^2 S(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= -\frac{2}{n}\sum_{i=1}^n [\hat{X}^{(p)}(t_i) - F(\hat{\boldsymbol{X}}(t_i), \tilde{\boldsymbol{\beta}})] \frac{\partial^2 F(\hat{\boldsymbol{X}}(t_i), \tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \\
&\quad + \frac{2}{n}\sum_{i=1}^n \frac{\partial F(\hat{\boldsymbol{X}}(t_i), \tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \Big[\frac{\partial F(\hat{\boldsymbol{X}}(t_i), \tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}}\Big]^T.
\end{aligned}$$

Using an argument similar to that for deriving (A.5) and Assumption 1 and 3, it can be checked that the first term on the right side of the preceding equation is $o_p(1)$, while the second term converges to $2E[\{\partial F(\boldsymbol{X}, \boldsymbol{\beta}_0)/\partial \boldsymbol{\beta}\}\{\partial F(\boldsymbol{X}, \boldsymbol{\beta}_0)/\partial \boldsymbol{\beta}\}^T]$. Combining (A.11) - (A.18) and recalling Assumption 1(iii) on the bandwidth h_n, we apply the Lindeberg-Feller central limit theorem to obtain the desirable result:

$$n h_n^{(2p+1)/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 + \boldsymbol{C} h_n^2) \to N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$$

in distribution, where

$$\begin{aligned}
\boldsymbol{C} &= \frac{\boldsymbol{\xi}_{p+1}^T \boldsymbol{S}_{p+1}^{-1} \boldsymbol{c}_{p+1}}{p+1} E\Big(X^{(p+2)}(t) \frac{\partial F(\boldsymbol{X}(t), \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\Big), \\
\boldsymbol{\Sigma}_{\boldsymbol{\beta}} &= p!^2 \sigma_\epsilon^2 \boldsymbol{\xi}_{p+1}^T \boldsymbol{S}_{p+1}^{-1} \boldsymbol{S}_{p+1}^* \boldsymbol{S}_{p+1}^{-1} \boldsymbol{\xi}_{p+1} \Big[E\Big\{\frac{\partial F(\boldsymbol{X}(t), \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\Big\}\Big\{\frac{\partial F(\boldsymbol{X}(t), \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\Big\}^T\Big]^{-1} \\
&\quad \times E\Big[\Big\{\frac{\partial F(\boldsymbol{X}(t), \boldsymbol{\beta}_0)}{\sqrt{f(t)}\partial \boldsymbol{\beta}}\Big\}\Big\{\frac{\partial F(\boldsymbol{X}(t), \boldsymbol{\beta}_0)}{\sqrt{f(t)}\partial \boldsymbol{\beta}}\Big\}^T\Big] \\
&\quad \times \Big[E\Big\{\frac{\partial F(\boldsymbol{X}(t), \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\Big\}\Big\{\frac{\partial F(\boldsymbol{X}(t), \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\Big\}^T\Big]^{-1}.
\end{aligned}$$