HHS Public Access

Author manuscript

Technometrics. Author manuscript; available in PMC 2016 January 01.

Published in final edited form as:

Technometrics. 2015 July 1; 57(3): 341–350. doi:10.1080/00401706.2015.1006338.

Regularized Semiparametric Estimation for Ordinary Differential Equations

Yun Li, **Ji Zhu**, and **Naisyin Wang**

Department of Statistics, University of Michigan

Abstract

Ordinary differential equations (ODEs) are widely used in modeling dynamic systems and have ample applications in the fields of physics, engineering, economics and biological sciences. The ODE parameters often possess physiological meanings and can help scientists gain better understanding of the system. One key interest is thus to well estimate these parameters. Ideally, constant parameters are preferred due to their easy interpretation. In reality, however, constant parameters can be too restrictive such that even after incorporating error terms, there could still be unknown sources of disturbance that lead to poor agreement between observed data and the estimated ODE system. In this paper, we address this issue and accommodate short-term interferences by allowing parameters to vary with time. We propose a new regularized estimation procedure on the time-varying parameters of an ODE system so that these parameters could change with time during transitions but remain constants within stable stages. We found, through simulation studies, that the proposed method performs well and tends to have less variation in comparison to the non-regularized approach. On the theoretical front, we derive finite-sample estimation error bounds for the proposed method. Applications of the proposed method to modeling the hare-lynx relationship and the measles incidence dynamic in Ontario, Canada lead to satisfactory and meaningful results.

Keywords

B-spline; Nonparametric; Penalized Estimation

1 Introduction

Dynamic systems are widely used in science and engineering, and they are often modeled through a set of ordinary differential equations (ODEs). Most ODE dynamic systems are fully determined by the parameters and initial values. They usually have non-linear structures and non-trivial analytic solutions. Given the parameters and initial values, there exist various numerical methods to solve non-linear ODEs, including the well known family of Runge-Kutta methods. In reality, the parameters of an ODE system are often unknown and need to be estimated using observed data.

SUPPLEMENTARY MATERIALS

The supplementary materials include approximation of the degrees of freedom, assumptions for theoretical properties, proofs of Lemma 1 and Theorem 1, and additional numerical results for the modified procedure of Cao et al. (2012).

Suppose that an ODE dynamic system has the following general structure:

$$\frac{dX}{dt}=F\{X(t),\theta,t\} \quad (1)$$

where $X(t) = \{X_1(t), \cdots, X_m(t)\}^T$ is the state vector (also referred to as ODE curves) to describe the dynamic system, $\theta = (\theta_1, \cdots, \theta_d)^T$ denotes the unknown parameters to be estimated, and $F(\cdot) = \{F_1(\cdot), \cdots, F_m(\cdot)\}^T$ is a known force functional structure, which is usually highly non-linear. Instead of directly observing the true state vector $X(t)$, we assume that we observe the surrogate $Y(t)$ at discrete time points

$$Y_{ij}=Y_j(t_{ij})=X_j(t_{ij})+\varepsilon_{ij}, i=1,\ldots,n_j; j=1,\ldots,m. \quad (2)$$

In most of the current statistics literature, the parameters θ are assumed as constants, and there are mainly two categories of methods for estimating the constant θ. The first category consists of various two-stage methods: one estimates the ODE curves $X(t)$ and their first derivatives in stage-one by a nonparametric smoothing fit to the data, and then, in the second stage, finds the parameter estimates through the classical least-square optimization with $X(t)$ and $dX(t)/dt$ replaced by the nonparametric estimates obtained from the first stage. For example, Varah (1982) estimated $X(t)$ and $dX(t)/dt$ using a spline smoothing technique in stage-one. Liang and Wu (2008) extended the work of Varah (1982) by using the local polynomial regression as the smoothing approach and they further provided statistical properties of the estimator. The use of non-parametric kernel estimation was proposed and studied in Brunel (2008). These approaches can be easily implemented and can perform very well with moderate to large data sets with densely observed data points. However, if the level of observation noise is relatively high and/or the sample size is relatively small, the two-stage method may not be able to obtain sufficiently precise estimates of $dX(t)/dt$ in the first stage and consequently the estimation of parameters in the second stage also suffers.

The second category of methods is built on profile estimation. The approach was introduced by Ramsay et al. (2007), and it has been referred to as the parameter cascade method. Instead of estimating the ODE curves directly from the data, one first constructs the ODE curves as functions of the parameters in an inner step. These estimated functions are then included into an outer step which minimizes a loss function between the observed data and the estimated ODE curves. In Ramsay et al. (2007) and several follow-up papers, a penalty term is included in the inner step with the intention of accommodating ODE model misspecifications while still maintaining the faithfulness of the estimated ODE curves towards the assumed system.

We note that the above methods all assume the parameters θ as constants. In reality, however, the parameters θ may not always remain constant as the system evolves with time. For example, Chen and Wu (2008) noticed that the ODE parameters in the HIV/AIDS dynamics could vary with time and they applied a two-stage method to estimate the time-varying ODE parameters. Cao et al. (2012) considered a specific ODE structure in which $dX(t)/dt$ is linked to a known function of $X(t)$ and a set of other observed covariates $Z(t)$ via time-varying coefficient functions. They proposed to include penalties, controlled by tuning

parameters, in both the inner and the outer steps of the parameter cascade method. In this paper, we consider a different modeling approach that allows the ODE parameters to vary with time, while at the same time retains the interpretation advantage of a parametric ODE system.

Taking the Lotka-Volterra dynamic model as an example, which is widely used to study the population evolution of predator and prey in ecological sciences: when the two components of the Lotka-Volterra model are dynamically balanced with each other, the parameters of the model are constants. However, when certain unpredictable interferences occur, such as environmentally unsound logging practice, the balance of the system may be broken and the ODE parameter values will change. If the interferences do not last long or they become part of the ecological system, after a certain time period, another balanced system will be re-established and the parameters would again become constants, usually at different values from before. Note in this situation, estimation methods that treat ODE parameters as constants are no longer suitable, while simply assuming time varying parameters through out the whole time domain will result in parameter estimates that are difficult to interpret and lose the understanding of the system provided by constant parameters. Our modeling strategy, as in Ramsay et al (2007), is motivated by the desire of allowing the assumed parametric model to differ from the true underlying one. With the assumed model, likely specified by scientists for its specific meaningful interpretation, our setup focuses on allowing a specific type of interpretable model violation through varying coefficient functions.

With this setup in mind, we wish to achieve a compromise between the two. Specifically, we propose a semi-parametric method that encourages the ODE parameters to stay as constants whenever possible (for interpretability) and at the same time also allows the ODE parameters to vary with time when needed (for flexibility). An additional new contribution comes from a penalty term that we propose to add in the outer step of the parameter cascade method, which will be described in detail in Section 2. We also show in Section 3 that, under certain regularity conditions, the difference between the parameter curves estimated by the proposed method and the truth is bounded at a certain rate.

The rest of the paper is organized as follows. In Section 2, we propose the estimation method and discuss various important issues in the algorithm including the corresponding degrees of freedom and the choice of the penalty parameter. Non-asymptotic bounds on the errors of the proposed estimator are developed in Section 3. In Section 4, we compare our method with other methods by simulation studies. The two models we have investigated are the FitzHugh-Nagumo model and the Lotka-Volterra model. Additional numerical results and the theoretical details are provided in the supplementary materials. In Section 5, we apply the proposed method to analyze a lynx-hare dynamic data set and a measles incidence dynamic data set collected in Ontario, Canada. We conclude the paper with a short discussion in Section 6.

2 Estimation Procedure

In this section, we propose a penalized method for estimating the ODE parameters and address issues arising in the selection of the tuning parameter. Our estimation procedure is, in part, motivated by the parameter cascade approach provide by Ramsay et al. (2007). We made two key modifications, one in the inner step and the other in the outer one, in order to build the estimation procedures that serve the goal of our investigation. For the purpose of variance reduction and estimation stability, we remove the additional penalty term introduced by Ramsay et al. (2007) in the inner step. The original motivation in Ramsay et al. (2007) of adding the penalty term was to accommodate scenarios where the ODE model is mis-specified. However, Ramsay and co-authors have noted the potential variation inflation due to this additional penalty (Poyton et al., 2006). In the numerical investigations in the first author's Ph.D. dissertation (Li, 2012), it is further noted that adding such an additional term in the inner step in estimating $X(t)$ via a set of B-spline basis functions increases the sensitivity of the results with respect to the tuning parameters, which include the number of B-spline bases and the penalty parameter(s); this is so even under constant-coefficient ODE models. In comparison to the original approach, the removal of the penalty term from the inner step leads to outcomes that are much less affected by tuning parameters.

The penalty term we add is in the outer step, and it does not suffer from the same instability and variation inflation as that with the penalty in the inner step. It further serves the dual purposes of regularizing roughness when the estimated parameter function is not a constant and encouraging a constant estimate when the parameter function does not vary much with time. Consequently, we only need to determine a minimum number of tuning parameters.

2.1 Set-up and notation

Let $\theta_\ell(t)$, $\ell = 1, \ldots, d$ denote the time-varying parameters in an ODE system. Using a p-dimensional basis $\psi(t) = \{\psi_1(t), \psi_2(t), \ldots, \psi_p(t)\}^T$, we expand $\theta_\ell(t)$ as follows:

$$\theta_\ell(t)=\xi_\ell(t)+e_{p,\ell}(t)=\psi(t)^T\eta_\ell+e_{p,\ell}(t), \ell=1,\ldots,d,$$

where η_ℓ is the coefficient vector of the basis expansion, and $e_{p,\ell}(t)$ represents the deviation of $\xi_\ell(t)$ from the true parameter curve $\theta_\ell(t)$. Note that our expansion allows imperfect modeling via $e_{p,\ell}(t)$. We also denote the space spanned by $\psi(t)$ as $\mathscr{L}_{\psi,p}$ and $\xi_\ell(\cdot) \in \mathscr{L}_{\psi,p}$. A common choice of $\psi(t)$ is the B-spline basis, which is what we will use. Throughout the paper, we use the notation of $\eta=(\eta_1^T,\ldots,\eta_d^T)^T$ and $\theta(\cdot) = \{\theta_1(\cdot), \ldots, \theta_d(\cdot)\}^T$.

For the j-th component of ODE curves, similarly we have

$$X_j(t)=\hat{X}_j(t)+e_{q,j}(t)=\phi(t)^Tc_j+e_{q,j}(t), j=1,\ldots,m,$$

where $\varphi(t) = \{\varphi_1(t), \ldots, \varphi_q(t)\}^T$ is a q-dimensional basis vector, c_j is the coefficient vector and $e_{q,j}(t)$ represents the deviation from the true ODE curve $X_j(t)$. Denote $c=(c_1^T,\ldots,c_m^T)^T$.

Further, we denote the initial-value vector as $X[0] = \{X_1(0), \ldots, X_m(0)\}^T$. Later, we will treat this vector as part of the unknown parameters. For notation convenience, we denote such a vector by x_0. We further let $\eta^* = (x_0^T, \eta^T)^T$, which is the entire parameter vector that we wish to estimate. Note that η and $\xi(\cdot)$ are equivalent, and we will use them interchangeably.

2.2 Algorithm

In what follows, we propose a two-step algorithm that regularizes the negative log-likelihood in the outer step. Throughout, unless otherwise specified, the integration in the algorithm was carried out by a Riemann integral approximation.

- Inner step. Given $\eta^* = \{x_0, \xi(\cdot)\}$, we build a profile estimator by solving for the ODE curves using the traditional least square criterion:

$$\hat{c}(\eta^*) = \underset{c}{\operatorname{argmin}} \int \sum_{j=1}^{m} w_j \left[\frac{d\hat{X}_j}{dt} - F_j\{\hat{X}, \xi(\cdot), t\} \right]^2 dt, \text{ subject to } \hat{X}[0] = x_0, \quad (3)$$

  where w_j's are normalizing weights with the purpose of making the numerical magnitudes of different components comparable. Given η^*, in this step, we construct the estimated ODE curves via $\hat{c}$, which will be adopted into the outer step as follows.

- Outer step. We estimate η^* by minimizing a penalized least square criterion:

$$\hat{\eta}^* = \underset{\eta^*}{\operatorname{argmin}} \sum_{j=1}^{m} \sum_{i=1}^{n_j} w_j \{Y_{ij} - \hat{X}_j(t_{ij}; \eta^*)\}^2 + \sum_{\ell=1}^{d} \lambda_\ell \int_0^T |\xi_\ell'(t)| dt, \quad (4)$$

  where $\hat{X_j}$ is estimated from the inner step, hence an implicit function of η^*.

The penalty on the first derivative of $\xi_\ell(t)$ serves dual purposes: encouraging interpretation and penalizing roughness. For the former, this penalty encourages the estimated parameter curve to be constant over time regions, hence providing highly interpretable results. To better understand the latter, we consider a general penalty term with the structure of $\lambda_\ell \int_0^T \mathscr{G}[\{\xi_\ell^{(s)}(t)\}^2] dt$. The most commonly used roughness penalty has $\mathscr{G}(\cdot)$ being the identity and $s = 2$, i.e. penalizing the square of the second derivative. In our case, we use $s = 1$, which corresponds to the first order roughness penalty (Green and Silverman, 1994), and $\mathscr{G}(\cdot)$ being the square-root function. This first order penalty assigns a high cost when neighboring values differ greatly. As a consequence, smoother estimates will be preferred, as discussed by Green and Silverman, provided that $\mathscr{G}(\cdot)$ is an increasing function in the range of the integration.

Note that λ_ℓ is the tuning parameter for the ℓ-th parameter curve, which balances the goodness of fit between the observations and the fitted ODE curve and the flexibility of the estimated parameter curve. When λ_ℓ is sufficiently large, the estimated parameter curve will be a constant over the entire time region, which reduces the problem to the constant parameter case. Furthermore, even though the inner-step is practically identical to the

traditional least-square ODE solver, with x_0 being part of η^* to be determined in the outer step, the final estimated x_0 may not be the same as the pre-determined initial values, which makes the proposed estimator differ from the traditional ODE solver. This property allows additional flexibility and robustness against mis-specified initial values in model fitting.

Denote the objective functions in the inner and outer steps above by $J(c, \eta^*)$ and $H(\hat{c}(\eta^*), \eta^*)$, respectively. For the latter, $\hat{c}(\eta^*)$ denotes the c-coefficients inside of $\hat{X}$ obtained from the inner step. We now describe the optimization procedure in each step with more details. Treating η^* as fixed, we can directly derive the gradient $\partial J(c, \eta^*) / \partial c$ given η^* within the inner step. As a result, we can utilize non-linear least squares to solve the minimization in the inner step. For the penalty term of the outer step, we use a grid of equally spaced points to approximate the integral in the penalty term of the outer step. Specifically, let $\tau_1 = T / K$, $\tau_2 = 2T / K, \cdots, \tau_K = T$, we approximate the penalty as

$$P_\lambda(\xi) = \sum_{\ell=1}^{d} \lambda_\ell T / K \sum_{k=1}^{K} |\xi'_\ell(\tau_k)|. \quad (5)$$

Then, we use the local quadratic approximation in Fan and Li (2001) to approximate the absolute value function in the above penalty. The exact procedure of how to do so is described by formulae and text between equations (3.6) and (3.7) in the original paper and we will not repeat it here. In practice, one can absorb T / K into the tuning parameter λ_ℓ. Now, using the implicit function theorem as in Ramsay et al. (2007), we can obtain the derivative $d\hat{c} / d\eta^*$, including components of $d\hat{c} / d\eta$ and $d\hat{c} / dx_0$, which are used to construct the non-linear least square gradient in the outer step. η^*, $\partial J(c, \eta^*) / \partial c = 0$ at $\hat{c}(\eta^*)$, we have, at $\{\hat{c}(\eta^*), \eta^*\}$

$$\frac{\partial^2 J}{\partial c^2} \frac{d\hat{c}}{d\eta^*} + \frac{\partial^2 J}{\partial c \partial \eta^*} = 0.$$

Therefore,

$$\frac{d\hat{c}}{d\eta^*} = -\left(\frac{\partial^2 J}{\partial c^2}\right)^{-1} \frac{\partial^2 J}{\partial c \partial \eta^*}.$$

Then, the outer step gradient at $\hat{c}(\eta^*)$ can be expressed as

$$\frac{dH}{d\eta^*} = \frac{\partial H(\cdot, \eta^*)}{\partial \eta^*} + \frac{\partial H(c, \cdot)}{\partial c} \frac{d\hat{c}}{d\eta^*} = \frac{\partial H}{\partial \eta^*} - \frac{\partial H}{\partial c} \left(\frac{\partial^2 J}{\partial c^2}\right)^{-1} \frac{\partial^2 J}{\partial c \partial \eta^*}.$$

Overall, using the derivatives $d\hat{c} / d\eta$ and $d\hat{c} / dx_0$, the discrete approximation to the integral and the local quadratic approximation to the absolute value function inside the penalty term, we can readily solve the minimization in the outer step, again via non-linear least squares.

2.3 Tuning parameter selection

Tuning parameters are often selected using criteria such as BIC, AIC or GCV. One key component in the above criteria is the degrees of freedom of the fitted model. For an estimator $\hat{\mu}$, the conventional definition of the degrees of freedom is

$$df = \sum_{i=1}^{n} \mathrm{cov}(\hat{\mu}_i, y_i)/\sigma^2.$$

Efron (2004) proposed a bootstrap procedure and Shen and Ye (2002), a data perturbation method, to estimate the degrees of freedom associated with $\hat{\mu}$. Even though these procedures can be implemented in a straightforward manner, they are computationally demanding. To reduce the computational cost, we design an approximation to the degrees of freedom by mimicking how it is estimated in ridge regression. The essential idea is to approximate the objective function in (4) with a ridge regression type criterion, and then to estimate the degrees of freedom mimicking what has been done for the ridge regression. The details are given in the supplementary materials. Our numerical studies indicate that the proposed approximation obtains similar results in estimating the degrees of freedom as the computationally more intensive bootstrap method.

There are also other practical issues with tuning parameter selection. For example, to further reduce the computational cost, we use the same value for λ_ℓ across all ℓ. Regarding K (the number of grid points for the discrete integral in the penalty term) and q (the number of basis functions for approximating the ODE curves), we found in our numerical studies that the results are not sensitive to them as long as they are large enough. As for p, the number of basis functions that expand the parameter curves, one may use the theoretical results in Section 3 as a guideline, and we also found in our numerical studies that it can be much smaller than both K and q.

3 Theoretical Results

In this section we develop non-asymptotic bounds for our proposed estimation method. We outline the main results. The key assumptions and the proofs are provided in the supplementary materials.

3.1 Set-up and notation

To clearly present the essential concepts, we focus on the scenario that there is only one time-varying parameter, i.e., $d = 1$, and the ODE system (1) has one X with $m = 1$. The entire time range is also scaled to [0, 1], i.e., $T = 1$.

Recall that we use an element in $\mathscr{L}_{\psi,p}$, the space spanned by the basis functions $\psi(t)$, to approximate the ODE parameter curve $\theta(t)$, i.e.,

$$\theta(t) = \xi(t) + e_p(t) = \psi(t)^T \eta + e_p(t). \quad (6)$$

Throughout, we consider bounded basis functions and conduct all the numerical studies using the B-spline bases. We denote $\omega_p = \|\theta - \xi\|_\infty = \|e_p\|_\infty$, where $\|\cdot\|_\infty$ is the supremum norm.

With $0 < \tau_1 < \tau_2 < \cdots < \tau_K = 1$ forming a grid of K evenly spaced points between 0 and 1, we let $A = \{\nabla\psi(\tau_1), \nabla\psi(\tau_2), \cdots, \nabla\psi(\tau_K)\}^T$, where $\nabla\psi(\tau_k) = K\{\psi(\tau_k) - \psi(\tau_{k-1})\}$ and $\tau_0 = 0$. Letting $\gamma_k = K\{\xi(\tau_k) - \xi(\tau_{k-1})\}$, we have

$$\gamma = A\eta, \quad (7)$$

where $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_K)^T$. For exposition simplicity in our theoretical development, we let $K = p$; consequently, A is a $p \times p$ square matrix. For certain choices of ψ, for example B-splines, A is also invertible. Thus, there is a one-to-one correspondence between γ and η. This strategy is also adopted in the theoretical proofs of James et al. (2009). Denote $\gamma^* = (x_0, \gamma^T)^T$. With the equivalence between γ^* and η^*, we may write the estimated ODE curve $\hat{X}$ in the inner step as $\hat{X}(t; \gamma^*)$ or $\hat{X}(t; x_0, \xi)$.

Using the discrete approximation to the integral in the penalty with $K = p$ evenly spaced points, due to (5), the outer criterion (4) becomes

$$\hat{\gamma}^* = \underset{\gamma^*}{\operatorname{argmin}} \frac{1}{n}\|Y - \hat{X}(t;\gamma^*)\|^2 + \frac{\lambda}{p}\|\gamma\|_1, \quad (8)$$

where $\|\gamma\|_1 = \sum_{k=1}^{p}|\gamma_k|$, $Y = (Y_1, Y_2, \ldots, Y_n)^T$ and $t = (t_1, t_2, \ldots, t_n)^T$.

3.2 Main results

We need four major assumptions for our theoretical analysis and we describe them in the supplementary materials. We briey comment on them here before we describe the theoretical properties. Most of the assumptions we have are regularity ones, which ensure that the ODE system can be solved and that the solution follows reasonable structures. Helpful remarks about the regularity assumptions regarding the ODE system can be found in Qi and Zhao (2010). The assumption (A1) indicates that the force function F in the ODE system (1) is sufficiently smooth. Since we approximate the ODE solution and its derivative function by an element within a q-dimensional functional space, the assumption (A2) warrants the estimation feasibility and the precision. For example, if cubic splines are used to approximate the ODE curve, the assumption (A2) can be easily satisfied based on Theorem 1 in Hall (1968). Part of our theoretical derivations extend the properties reported in Qi and Zhao (2010), and similar to them, we also focus on γ^* that belongs to a compact set.

We now introduce Lemma 1, which provides precision in estimation of the ODE curve.

Lemma 1 *Under assumptions* (A1) and (A2) given in supplementary materials, *for* $\hat{X}(\cdot; x_0, \xi) = \hat{X}(\cdot; \gamma^*)$ *which is estimated using the inner criterion (3), we have, given* γ^*,

$$\|\hat{X}(\cdot;\gamma^*) - X(\cdot;\gamma^*)\|_\infty \le c_2 r_q, \quad (9)$$

where c_2 is a constant that only depends on c_1 in the assumption (A1), and r_q is the upper bound in the assumption (A2).

This lemma indicates that, given parameter γ*, the ODE curve can be estimated with a high precision. The precision rate is determined only by the estimation rate r_q in the assumption (A2), where the rate r_q is determined by the number of basis functions, q, or the choice of knots when B-spline bases are used to approximate the ODE curve. It is worth pointing out that Lemma 1 holds for any parameter curve θ, not only for ξ in $\mathscr{L}_{\psi,p}$ if the assumption (A2) is made based on θ.

Assumption (A3) is given to ensure that the approximation of the ODE curve in $\mathscr{L}_{\psi,p}$ is uniquely determined by the vector η* (and equivalently γ*), and vice versa.

Our last assumption is similar to the restricted eigenvalue (RE) assumption in Bickel et al. (2009). This assumption for regularized estimators prevents certain problematic setups analogous to unregularized singular design matrices in linear models. It is known to hold for design matrices with orthogonal or weakly dependent columns (Raskutti et al., 2010). It also supports our use of B-spline basis functions, where two basis functions with disjoint supports are orthogonal to each other.

Before we state the major theorem, we need more notations. We use $\theta_0(\cdot)$, $\xi_0(\cdot)$ and $\gamma_0^*=(X_0[0],\gamma_0^T)^T$ to respectively denote the true ODE parameter curve, the approximation function in $\mathscr{L}_{\psi,p}$, and the corresponding coefficient vector. We let $J_{\mathscr{F}}=\{2,\ldots,p+1\}$ and $J(\gamma)=\{k\in J_{\mathscr{F}}:\gamma_k\neq 0\}$, and also let $|J|$ denote the cardinality of J. We denote $\hat{\xi}_0(\cdot)$ as the estimator for $\xi_0(\cdot)$ using our method, and let $S_p=|J(\gamma_0)|$, the number of non-zero parameters. Further, we define $K_1=\vee_{k=2}^{p+1}\|\mathscr{M}_{\cdot k}^U\|\vee(4^{-1}\|\mathscr{M}_{\cdot 1}^U\|)/\sqrt{n}$ and $K_2=\vee_{k=1}^{p+1}(\|\mathscr{M}_{\cdot k}^U\|_1/n)$, where $a\vee b=\max(a,b)$, $\mathscr{M}^U$ is defined in Assumption (A3) given in the supplementary materials, and $\mathscr{M}_{\cdot k}^U$ denotes the k-th column of $\mathscr{M}^U$. Denote $\alpha_p(t)=\|\psi(t)^TA^{-1}\|$ and $\omega_{p,q}=c_2r_q+c_3\omega_p$, where c_2 is from Lemma 1 and c_3 is from Assumption (A3) in the supplementary materials. Then we have the following theorem.

Theorem 1 *Suppose $Y_i=X(t_i;X_0[0],\theta_0)+\varepsilon_i=\hat{X}(t_i;\gamma_0^*)+\zeta_i$, where $\zeta_i=\varepsilon_i+e_i$. Assume $\varepsilon_i\sim N(0,\sigma_\varepsilon^2)$ and let*

$$\frac{\lambda}{2p}=a\sigma_\varepsilon K_1\sqrt{\frac{\log(p+1)}{n}}+2K_2\omega_{p,q},$$

where $a>2\sqrt{2}$. Then, under assumptions (A1)–(A4) given in the supplementary materials, *with probability at least $1-(p+1)^{1-a^2/8}$, we have*

$$|\hat{X}_0[0] - X_0[0]| \leq \frac{4a\sigma_\varepsilon K_1}{\kappa^2}\sqrt{\frac{(S_p+1)\log(p+1)}{n}} + \frac{8K_2\sqrt{S_p+1}}{\kappa^2}\omega_{p,q},$$

$$|\hat{\xi}_0(t) - \theta_0(t)| \leq \frac{16\alpha_p(t)a\sigma_\varepsilon K_1(S_p+1)}{\kappa^2}\sqrt{\frac{\log(p+1)}{n}} + \frac{32\alpha_p(t)K_2(S_p+1)}{\kappa^2}\omega_{p,q} + \omega_p.$$

In addition, if $\omega_p = o[\{\log(p+1)/n\}^{1/2}]$ *so that approximately* $\zeta_i \sim N(0, \sigma_\varepsilon^2)$, *then with probability at least* $1 - (p+1)^{1-a^2/8}$, *we have the following error bounds*:

$$|\hat{X}_0[0] - X_0[0]| \leq \frac{4a\sigma_\varepsilon K_1}{\kappa^2}\sqrt{\frac{(S_p+1)\log(p+1)}{n}},\ |\hat{\xi}_0(t) - \theta_0(t)| \leq \frac{16\alpha_p(t)a\sigma_\varepsilon K_1(S_p+1)}{\kappa^2}\sqrt{\frac{\log(p+1)}{n}}.$$

As in James et al. (2009), one can also establish assumptions to ensure the rates of $\alpha_p(t)S_p\{\log(p+1)/n\}^{1/2}$ and/or $\alpha_p(t)S_p\omega_{p,q}$ going to zero as n, p and q grow to ∞; see for example the assumption A3 therein. Hence, the non-asymptotic bounds could hold under the high-dimensional scenario when $p \gg n$. One can also consider S_p to be of order $n^{1/5}$ while p can remain to be a very large number by updating the choices of basis functions. The finite-sample bounds we establish here enable researchers to build additional asymptotic results according to the scenarios fitting their specific interests.

Further, as in the case when $\theta(t) \equiv \theta$ which is time invariant, we can safely use a very large q in the approximation of $X(t)$. This is because, by removing the penalty in the inner step, there is no bias-variance tradeoff in the inner step estimation. A larger number of q simply leads to higher numerical precision in the spline approximation.

4 Simulation Studies

In this section, we apply the proposed method to two simulated ODE dynamic systems; both have wide scientific applications.

4.1 The FitzHugh-Nagumo model

This model was invented by FitzHugh (1961) and Nagumo et al. (1962) to simplify the Hodgkin-Huxley model (1952), which was used to study the behavior of spike potential in the giant axon of squid neurons. Specifically, the ODEs are

$$\frac{dV}{dt} = c\left(V - \frac{V^3}{3} + R\right),\ \frac{dR}{dt} = -\frac{1}{c}(V - a + bR), \quad (10)$$

where V describes the voltage across an axon membrane, R is the recovery variable summarizing outward currents, and a, b, and c are ODE parameters in the dynamic system.

In this simulation study, we let the parameters a, b and c in (10) vary over time according to the following structure:

$$a(t)=0.2(\mathbb{I}(0 \le t \le \pi)+[1+\sin\{(t-\pi)/2\}]\mathbb{I}(\pi<t \le 3\pi)+\mathbb{I}(3\pi<t \le 20)),$$
$$b(t)=0.2(\mathbb{I}(0 \le t \le 3.5\pi)+[1+\sin\{(t-3.5\pi)/2\}]\mathbb{I}(3.5\pi<t \le 5.5\pi)+\mathbb{I}(5.5\pi<t \le 20)),$$
$$c(t)=3.0(\mathbb{I}(0 \le t \le 2.3\pi)+[1+\sin\{(t-2.3\pi)/2\}/2]\mathbb{I}(2.3\pi<t \le 4.3\pi)+\mathbb{I}(4.3\pi<t \le 20)),$$

where $\mathbb{I}(\cdot)$ is the indicator function.

The true V and R are calculated based on (10) with initial values $V(0)=-1.0$ and $R(0)=1.0$. We then generate 201 pairs of observed V and R on equally-distanced grids in [0, 20] with the observation errors drawn from $N(0, 0.5^2)$. To approximate the ODE curves $V(t)$ and $R(t)$, we use cubic B-splines with knots placed at each observed time point. To estimate the ODE parameter curves, we use cubic B-splines with 21 equally spaced knots. For the penalty term in the outer-step criterion, we use 201 equally spaced τ_k's to approximate the integral.

We consider four estimation methods, Methods I–IV. Method I is the proposed method that has a regularization term in the outer-step criterion. Method II is the time-varying estimation method without the regularization term in the outer step. In Method III, the parametric model with constant parameters is used. In addition to these three methods, we also consider Method IV, in which the true constant regions of parameters are assumed to be known. Specifically, within true constant regions, the parameters are estimated as constant values, and for non-constant regions, similar to Method II, cubic B-splines are used to approximate the parameter curves. Note that Method IV can only be carried out in simulation studies and can be considered as a benchmark. For Method I, we also compare the results using BIC, AIC and GCV for tuning parameter selection.

We use $\theta_{0\ell}(t)$, $\ell = 1, \ldots, d$, to denote the true ODE parameter curves, and $\hat{\theta_\ell}(t)$ its estimate. In the current example, θ_0's are $a(t)$, $b(t)$ and $c(t)$. Let $\mathscr{F}_\ell=\{t:\theta'_{0\ell}(t)=0\}$ and $\mathscr{F}_\ell^c=\{t:\theta'_{0\ell}(t) \neq 0\}$. Further, let $|\mathscr{F}_\ell|$ and $|\mathscr{F}_\ell^c|$ be the lengths of the regions of $\mathscr{F}_\ell$ and $\mathscr{F}_\ell^c$ respectively.

To compare different methods, we use the mean integrated square error (MISE), which is defined as follows:

$$\mathrm{MISE}_{\mathscr{F}_\ell}=\mathrm{E}\int_{\mathscr{F}_\ell}\{\hat{\theta}_\ell(t)-\theta_{0\ell}(t)\}^2dt, \mathrm{MISE}_{\mathscr{F}_\ell^c}=\mathrm{E}\int_{\mathscr{F}_\ell^c}\{\hat{\theta}_\ell(t)-\theta_{0\ell}(t)\}^2dt.$$

Further, denote the standard error of the estimator $\hat{\theta_\ell}$ at t as $\mathrm{SE}_{\hat{\theta_\ell}}(t)$ and define the average estimation standard error (AVSE) as

$$\mathrm{AVSE}_{\mathscr{F}_\ell}=|\mathscr{F}_\ell|^{-1}\int_{\mathscr{F}_\ell}\mathrm{SE}_{\hat{\theta}_\ell}(t)dt,\ \mathrm{AVSE}_{\mathscr{F}_\ell^c}=|\mathscr{F}_\ell^c|^{-1}\int_{\mathscr{F}_\ell^c}\mathrm{SE}_{\hat{\theta}_\ell}(t)dt.$$

For each t, the $\mathrm{SE}_{\hat{\theta_\ell}}(t)$ is assessed using the Monte Carlo standard deviation of the corresponding estimates. For the estimation of the ODE curves $X_j(t)$, which are $V(t)$ and $R(t)$ in this simulation, we also compare the MISE and AVSE, which are defined as follows:

$$\mathrm{MISE}_{X_j} = \mathrm{E}\int_0^T \{\hat{X}_j(t) - X_j(t)\}^2 dt,\ \mathrm{AVSE}_{X_j} = T^{-1}\int_0^T \mathrm{SE}_{\hat{X}_j}(t)dt,$$

where $\hat{X_j}$ is the estimated curve for the jth ODE component X_j and $\mathrm{SE}_{\hat{X_j}}(t)$ denotes the standard error of $\hat{X_j}$ at time t.

Table 1 (upper half) contains the MISE results over 100 repetitions for both the parameter curves and the ODE curves. The AVSE results are also compared in Table 1 (lower half). From the MISE results, we can see that our method (Method I) performs the best among the four methods considered, and for our method, the BIC results are slightly better than those of AIC and GCV. The performance of Method II is the worst among all methods, especially for the MISE results of $a(t)$ and $b(t)$. For $c(t)$, Method II performs slightly better, particularly in the non-constant region $\mathscr{F}_c^c$. This observation implies that the FitzHugh-Nagumo dynamic system may be more sensitive to the change in parameter c, which makes c easier to be estimated than a and b. The MISE results of Method III further suggest that $c(t)$ can not be approximated well by a constant. It is interesting to see that Method IV performs worse than our method in terms of MISE. Comparison of the AVSE results explains the reason: recall that Method IV does not regularize the estimation in the non-constant region, which results in relatively high estimation standard errors, and consequently jeopardizes the MISE. In the comparison of the AVSE, we also note that our method has comparable standard errors with Method III, in which only five parameters, i.e., constants a, b, c and two initial values, need to be estimated.

Table 1 also contains the MISE and AVSE results for ODE curves V and R. Method I with BIC achieves the smallest MISE results and the smallest AVSE results, and Method III performs the worst.

Figure 1 shows the average estimated parameter curves, together with the true parameter curves. Even though it seems that Method IV outperforms Method III in Figure 1, the MISE results for $a(t)$ and $b(t)$ suggest otherwise. The reason is again due to AVSE, i.e., Method III gains by the smaller estimation standard errors. From Figure 1, we can also see that in the non-constant time regions $\mathscr{F}_a^c$, $\mathscr{F}_b^c$ and $\mathscr{F}_c^c$, the proposed method over-shrinks the peak and the valley, especially at time points where the true first derivatives are zero.

4.2 The Lotka-Volterra Model

As a second example, we study the well known Lotka-Volterra dynamic model (Lotka, 1910; Volterra, 1926). This model, also known as the predator-prey model, has wide applications in modeling the dynamics of ecological systems with predator-prey interactions, competition and disease dispersion. The model has two components H and L, described by the following ODEs:

$$\frac{dH}{dt} = aH - bHL, \frac{dL}{dt} = -cL + dHL, \quad (11)$$

where *a*, *b*, *c* and *d* are the ODE parameters, and we specify them as follows in the simulation:

$$\begin{aligned} a(t) &= 0.5\mathbb{I}(0 \le t \le 6) + \sin(\pi t/36)\mathbb{I}(6 < t \le 12) + \sqrt{3}/2\mathbb{I}(12 < t \le 20), \\ b(t) &= 0.25\mathbb{I}(0 \le t < 3) + [0.25 - 0.1\{(t-3)/6\}^4]\mathbb{I}((3 < t \le 9) + 0.15\mathbb{I}(9 < t \le 20), \\ c(t) &= 1.0\mathbb{I}(0 \le t < 11) + [0.8 + 0.2\{(t-14)/3\}^2]\mathbb{I}(11 < t \le 17) + 1.0\mathbb{I}(17 < t \le 20), \\ d(t) &= 0.3\mathbb{I}(0 \le t \le 6) + [0.5 - 0.2\{(t-9)/3\}^2]\mathbb{I}(6 < t \le 12) + 0.3\mathbb{I}(12 < t \le 20). \end{aligned}$$

The initial values are set as $H(0) = 3.5$ and $L(0) = 0.5$, and we generate 101 pairs of observations on equally-distanced grid points between 0 and 20. The observation errors are drawn from $N(0, 1)$. Cubic B-splines are used to approximate the two ODE curves with knots placed at each observed time point. The setting of cubic B-splines for estimating the parameter curves and the grid for approximating the penalty integral are the same as those in the previous example.

Again, we compare the performances of Methods I–IV. Table 2 contains the MISE and AVSE results over 100 repetitions, and the average estimated parameter curves are plotted in Figure 2. From Figure 2, we can see that the estimated parameter curves of Method II are not as much erratic as the estimation results of $a(t)$ and $b(t)$ in the previous example. It implies that the parameters in this example are easier to estimate, comparing to those in the FitzHugh-Nagumo model. We can also see that for both Method I and Method IV, the average estimated parameter curves are very close to the true curves. Table 2 suggests that, in terms of both MISE and AVSE, our method performs the best among the four methods considered. Different from the results of the previous example, Method III performs unanimously worse than other methods. Notice the large MISE and AVSE for *H* and *L* of Method III; it implies that Method III barely fits the data generated under our parameter setting.

Upon request from a reviewer, we have created code to carry out the procedure proposed in Cao et al. (2012). We note that in Cao et al. (2012), the observed covariates $Z(t)$ are assumed to have time varying coefficient functions while $dX(t)/dt$ are assumed to relate to these $Z(t)$ and a known function of the unobserved $X(t)$ via constant coefficients, thus their setup is different from ours and does not apply directly to our setting. Nevertheless, their approach is shown to have superior numerical performances than those of Chen and Wu (2008) under the models considered in Cao et al. (2012). When the relationships between $dX(t)/dt$ and $X(t)$ are also postulated via time-varying coefficients, we had to modify the estimation procedure in Cao et al. (2012) so that it can be applied, and our experience suggested that choosing tuning parameters in both the inner and outer steps become a demanding task and that the results are sensitive to the choice of tuning parameters, including the numbers of basis functions used to estimate $X(t)$ and time-varying coefficients, respectively. The inclusion of a roughness penalty on the second derivative of the coefficient function in Cao et al. (2012) is no doubt beneficial. The setup in the inner step serves the purpose of balancing the overfit to the ODE structure by the goodness of fit to the responses, *Y*, evidenced by the best choice of the tuning parameters in the inner step being moderate. However, this is at the cost of variation inflation and sometimes estimation instability.

Nevertheless, we report the best results we obtained, corresponding to Tables 1 and 2, as supplementary materials. This modified procedure of Cao et al. (2012) has performance between those of methods III and IV, and our method still performs the best.

5 Ecology and Epidemiology Examples

In this section, we investigate two data examples, one is in ecology and the other is from epidemiology.

5.1 The hare and lynx data example

The numbers of trapped lynx and snowshoe hares of North Canada were collected from 1900 to 1920 (Odum 1953), and the observed data are believed to reflect the relative populations of lynx and hare in the study region. The Lotka-Volterra model (11), with $H(t)$ and $L(t)$ representing the evolution function of the number of snowshoe hares and lynx respectively, is used to model this predator-prey dynamics. Cubic B-splines with 201 equally-spaced knots are used to estimate the ODE curves $H(t)$ and $L(t)$. The parameter curves $a(t)$, $b(t)$, $c(t)$ and $d(t)$ are estimated with cubic B-splines with 8 equally-spaced knots, due to the limited number of observations. BIC is used to select the tuning parameter.

Figure 3 presents the estimated parameter curves based on three methods: the non-regularized method (Method II), the constant fitting (Method III) and our method (Method I). The estimated parameter curves by Method II are erratic and fail to provide any structural information for interpreting the ODE dynamics. Method I, on the other hand, offers much improved structural information than Method II: the estimated parameter curve $\hat{a}(t)$ stays almost as a constant over the entire time period; the estimated parameter curve $\hat{b(t)}$ is basically a constant from 1900 to 1905, then starts to increase and later stabilizes into a slightly larger constant around 1912; similar to $\hat{a}(t)$, the estimated parameter curve $\hat{c(t)}$ stays as a constant over the whole time period; the estimated parameter curve $\hat{d(t)}$ varies quite a bit and has not reached a constant-stage at the end of the time range, but comparing with the estimated $\hat{d(t)}$ of Method II, it has a much smaller variation.

Figure 4 shows the estimated ODE curves of $H(t)$ and $L(t)$, together with the original observed data. Both Methods I and II fit better at the first peaks of the $H(t)$ and $L(t)$ dynamics than Method III, similarly for the valleys in the middle of $H(t)$ and $L(t)$. On the other hand, towards the tails of $H(t)$ and $L(t)$, Methods I and III perform better than Method II. Overall, Method I performs the best among the three methods. We also conducted a bootstrap analysis for the purpose of variation comparison. The result, which is not reported here, shows that standard errors of our regularized method are comparable to those of Method III, and are much smaller than the standard errors of Method II.

5.2 The Canadian measles incidence data example

This data set consists of weekly measles incidence reports for the province of Ontario, Canada, from 1939 through 1965. Following the analysis of Hooker et al. (2011), we model the measles incidence using the so called SEI dynamic equations:

$$\dot{S}=\rho(t)-\beta(t)\left(\frac{I}{p(t)}+\upsilon\right)S,\ \dot{E}=\beta(t)\left(\frac{I}{p(t)}+\upsilon\right)S-\sigma E,\ \dot{I}=p(t)\sigma E-\gamma I,$$

where S is the susceptible class, E is the exposed (infected with the disease but not infectious) class and I is the infectious class. We note that S increases with a recruitment rate $\rho(t)$ and moves into E with a rate of $\beta(t)$ (I / $p(t)$ + υ). E transforms into I with the rate $\sigma p(t)$ as I recovers with the rate γ. In this data set only I, the measles infectious class, is observed. The other two state variables S and E are unobserved. The parameters are $\rho(t)$, $\beta(t)$, $p(t)$ (= p_0 + $p_1 t$), υ, σ and γ. Of these parameters, $\rho(t)$ is interpolated from the monthly birth rate data at a five-year lag, σ is known to be around 365/8, and γ is roughly estimated by the five-day mean infectious period and equals to 365/5. Only the parameters $\beta(t)$, p_0, p_1 and υ need to be estimated from the data.

The structure of $\beta(t)$ within each year has been studied in Bauch and Earn (2003), which consists of a high-level component during the summer season and a low-level one for the rest of the year. Adopting this yearly structure, we further use the proposed method to find the long-term pattern of $\beta(t)$. Following Hooker et al. (2011), we let

$$\beta(t)=\alpha(t)+s(t),$$

where $s(t)$ is a cyclic function that describes the same within-year pattern across all years, subject to the constraint $\int_0^1 s(t)dt=0$, and $\alpha(t)$ is the general parameter curve that describes the long- term trend. We use the cyclic cubic B-splines with knots at each month to expand $s(t)$ while using the regular cubic B-splines with knots at each year to approximate $\alpha(t)$. To find the long-term yearly pattern, only $\alpha(t)$ is regularized in the outer step of the proposed method.

We compare our method with two other methods: the time-varying approach without regularization in the outer step (Method II) and a set of short-term constant-fitting conducted every two years. In the latter approach, we assume $\alpha(t) = c$ for two neighboring years and only fit the data within those two years; we repeat this process for about 25 times from 1939 to 1963. Figure 5 shows the estimated $\alpha(t)$, in log scale, based on these three methods. We can see that the estimated $\alpha(t)$ based on our method is relatively large in the early years and deceases gradually to a constant after 1958. This implies that the rate at which the susceptible class moving into the exposed class decreases in the long-term pattern and gradually becomes stabilized. This pattern of $\alpha(t)$ might be related to an introduction of measles vaccine around 1954. After the measles vaccine took effect, $\alpha(t)$ could be modeled as a constant and $\beta(t)$ only contains the seasonal pattern, as shown in the bottom panel of Figure 5. On the other hand, the estimated $\alpha(t)$ of the other two methods are much more unstable and fail to provide any useful understanding for the long-term pattern of $\alpha(t)$. To further illustrate the differences among the three methods, we also plot the estimated I curve in Figure 5 from 1952 to 1954. We can see that in terms of goodness-of-fit to the data, our method performs similarly to the non-regularized approach; the result of the two-year constant fitting follows the data more closely, but there is a potential of overfitting.

6 Summary

We have proposed a regularized parameter estimation method for the ODE dynamic system. The proposed method aims at keeping a balance between interpretability and flexibility of the parameter curves. The proposed regularization not only helps in recovering the parametric structure, but also plays a role in smoothing and reducing the estimation variance. When the parameter curves are constants over some time regions, the proposed method in general performs much better than the ordinary non-regularized method. Our theoretical analysis indicates that under certain regularity conditions, the estimated nonparametric curves can obtain estimation error bounds that are functions of $\sqrt{\log p}$. This implies that a large p may not cause much harm in prediction accuracy for the proposed method. Numerical studies also show that the proposed method consistently out-performs other competitors and provides results that are helpful for understanding the dynamic system.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

The authors thank the editor, associate editor, and two anonymous referees for their comments that have resulted in significant improvements in the article. We also gratefully acknowledge Professor David J.D. Earn of McMaster University in Canada for making the Canadian measles incidence data available for research purposes. This research was partially supported by a grant from the National Cancer Institute (CA74552), two grants from the National Science Foundation (NSF DMS 0748389 and NSF DMS 1407698), and a grant from the National Institute of Health (NIH R01 GM096194).

References

1. Bauch CT, Earn DJD. Transients and attractors in epidemics. Proceedings of Royal Society B. 2003; 270:1573–1578.
2. Bickel PJ, Ritov Y, Tsybakov AB. Simultaneous analysis of lasso and Dantzig selector. Annals of Statistics. 2009; 37:1705–1732.
3. Brunel NJ-B. Parameter estimation of ODEs via nonparametric estimators. Electronic Journal of Statistics. 2008; 2:1242–1267.
4. Cao J, Huang JZ, Wu H. Penalized nonlinear least squares estimation of time-varying parameters in ordinary differential equations. Journal of Computational and Graphical Statistics. 2012; 21:42–56. [PubMed: 23155351]
5. Chen J, Wu H. Efficient local estimation for time-varying coefficients in deterministic dynamic models with applications to HIV-1 dynamics. Journal of the American Statistical Association. 2008; 103:369–384.
6. Efron B. The estimation of prediction error: Covariance penalties and cross-validation. Journal of the American Statistical Association. 2004; 99:619–642.
7. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association. 2001; 96:1348–1360.
8. FitzHugh R. Impulses and physiological states in models of nerve membrane. Biophysical Journal. 1961; 1:445–466. [PubMed: 19431309]
9. Green, PJ.; Silverman, BW. Nonparametric Regression and Generalized Linear Models. A Roughness Penalty Approach. New York: Chapman and Hall; 1994.
10. Hall CA. On error bounds for spline interpolation. Journal of Approximation Theory. 1968; 1:209–218.

11. Hodgkin AL, Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve. Journal of Physiology. 1952; 133:444–479.
12. Hooker G, Ellner SP, Roditi L, Earn DJD. Parameterizing state - space models for infectious disease dynamics by generalized profiling: measles in Ontario. Journal of the Royal Society Interface. 2011; 8:961–974.
13. Hutzinger, O. The Handbook of Environmental Chemistry, Vol. 1, Part D: The Natural Environment and the Biogeochemical Cycles. Berlin: Springer-Verlag; 1985.
14. James G, Wang J, Zhu J. Functional linear regression that's interpretable. Annals of Statistics. 2009; 37:2083–2108.
15. Li, Y. PhD dissertation. Department of Statistics, University of Michigan; 2012. Regularized Statistical Methods for Data of Grouped or Dynamic Nature.
16. Liang H, Wu H. Parameter estimation for differential equation models using a framework of measurement error in regression models. Journal of the American Statistical Association. 2008; 103:1570–1583. [PubMed: 19956350]
17. Lotka AJ. Contribution to the theory of periodic reaction. Journal of Physical Chemistry. 1910; 14:271–274.
18. Nagumo JS, Arimoto S, Yoshizama S. An active pulse transmission line simulating a nerve axon. Proceeding of the IRE. 1962; 50:2061–2070.
19. Odum, EP. Fundamentals of Ecology. Philadelphia: Springer; 1953.
20. Qi X, Zhao H. Asymptotic efficiency and finite-sample properties of the generalized profiling estimation of parameters in ordinary differential equations. Annals of Statistics. 2010; 38:435–481.
21. Poyton AA, Varziri MS, McAuley KB, McLellan PJ, Ramsay JO. Parameter estimation in continuous-time dynamic models using principal differential analysis. Computers & Chemical Engineering. 2006; 30:698–708.
22. Ramsay JO, Hooker G, Campbell D, Cao J. Parameter estimation for differential equations: a generalized smoothing approach (with discussion). Journal of the Royal Statistical Society, Series B. 2007; 69:741–796.
23. Raskutti G, Wainwright MJ, Yu B. Restricted eigenvalue properties for correlated Gaussian designs. Journal of Machine Learning Research. 2010; 11:2241–2259.
24. Shen X, Ye J. Adaptive model selection. Journal of the American Statistical Association. 2002; 97:210–221.
25. Varah JM. A spline least squares method for numerical parameter estimation in differential equations. SIAM Journal on Scientific Computing. 1982; 3:131–141.
26. Volterra V. Variazioni e uttuazioni del numero d'individui in specie animali conviventi. Mem. Acad. Lincei Roma. 1926; 2:31–113.

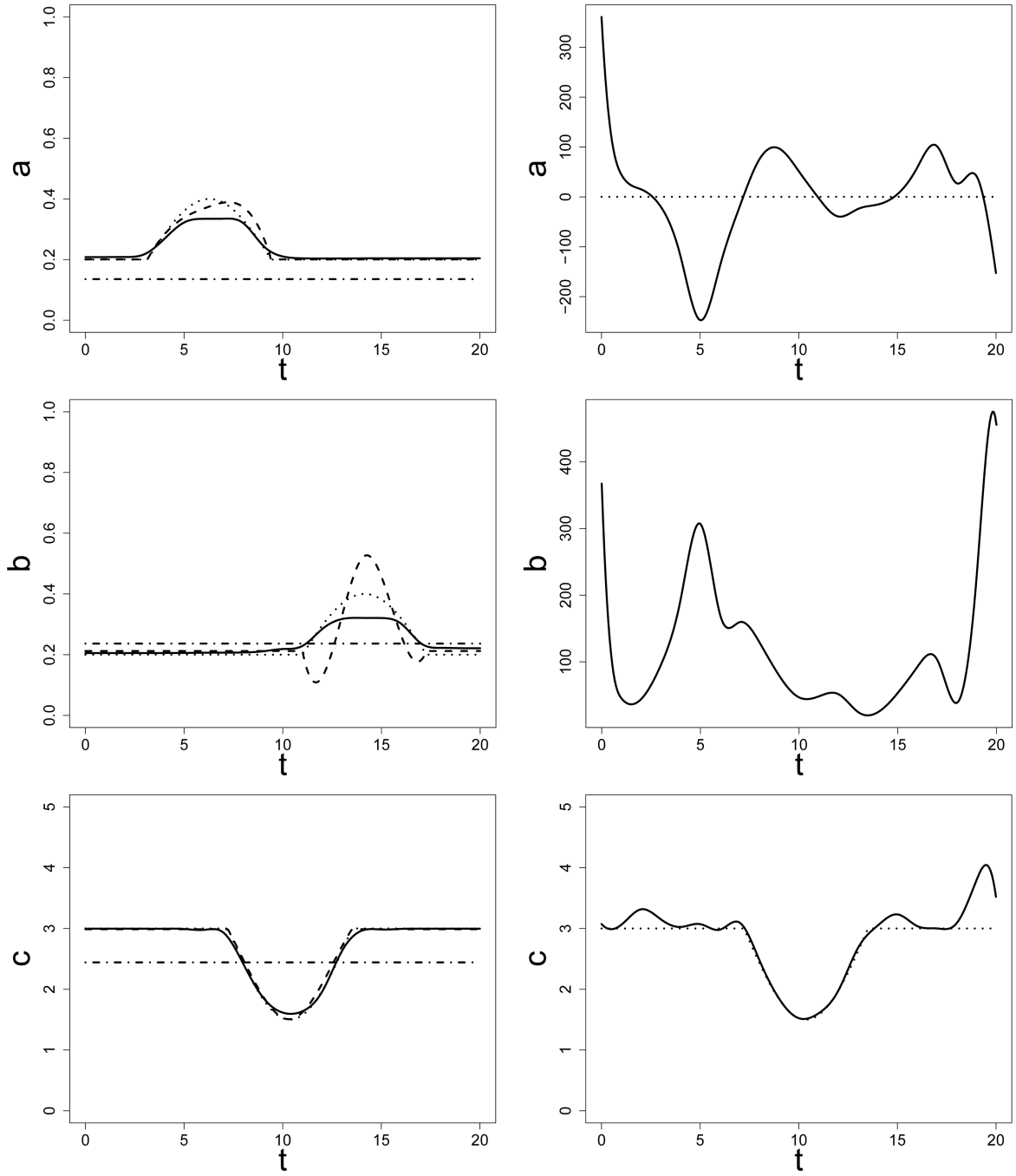

Figure 1.

The average of the estimated parameter curves plotted with true curves in the FitzHugh-Nagumo model. Left: Method I (solid); Method III (dash-dot); Method IV (dashed); Truth (dotted). Right: Method II (solid); Truth (dotted).

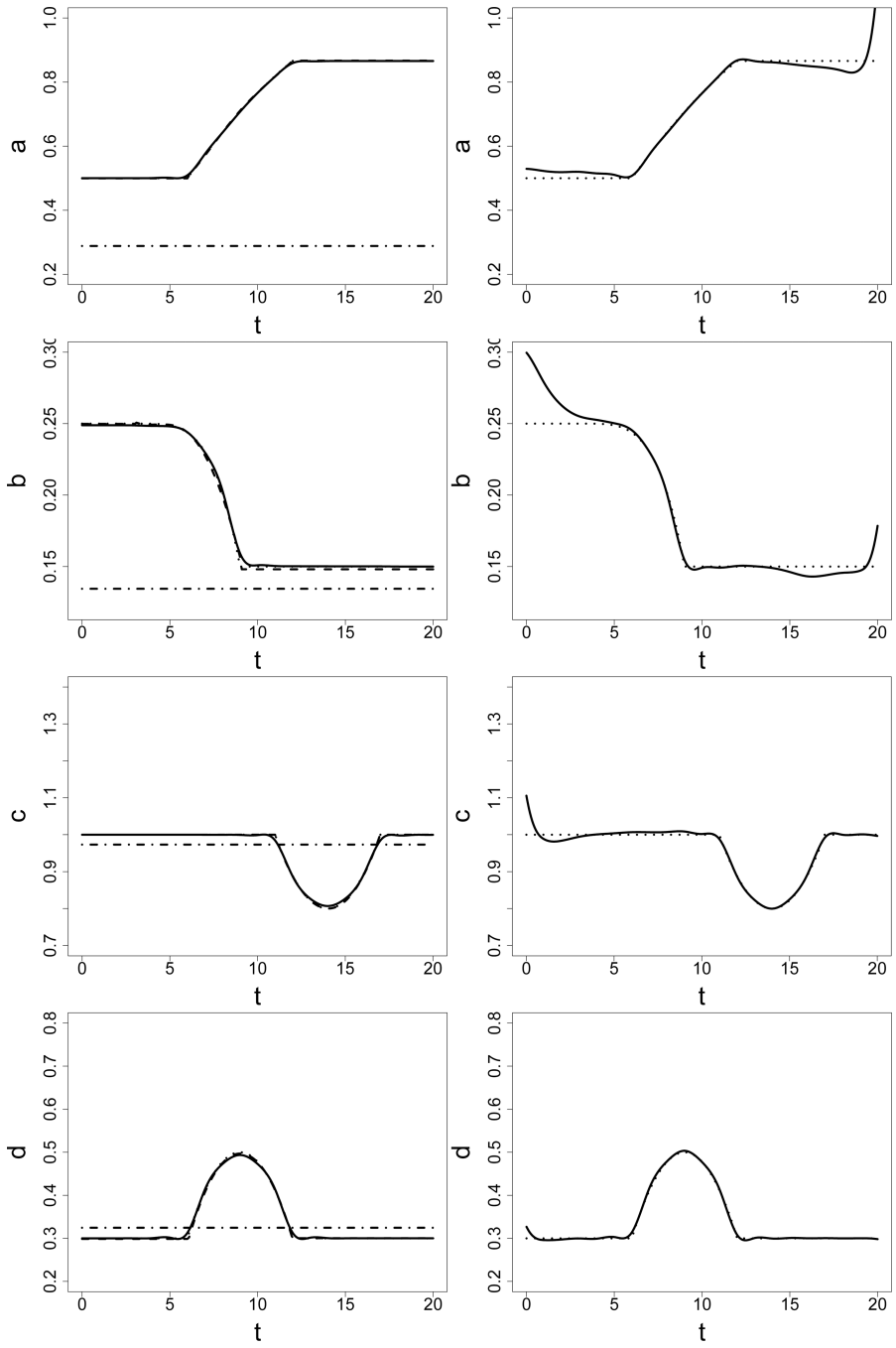

Figure 2.

The average of the estimated parameter curves plotted with true curves in the Lotka-Volterra model. Left: Method I (solid); Method III (dash-dot); Method IV (dashed); Truth (dotted). Right: Method II (solid); Truth (dotted).

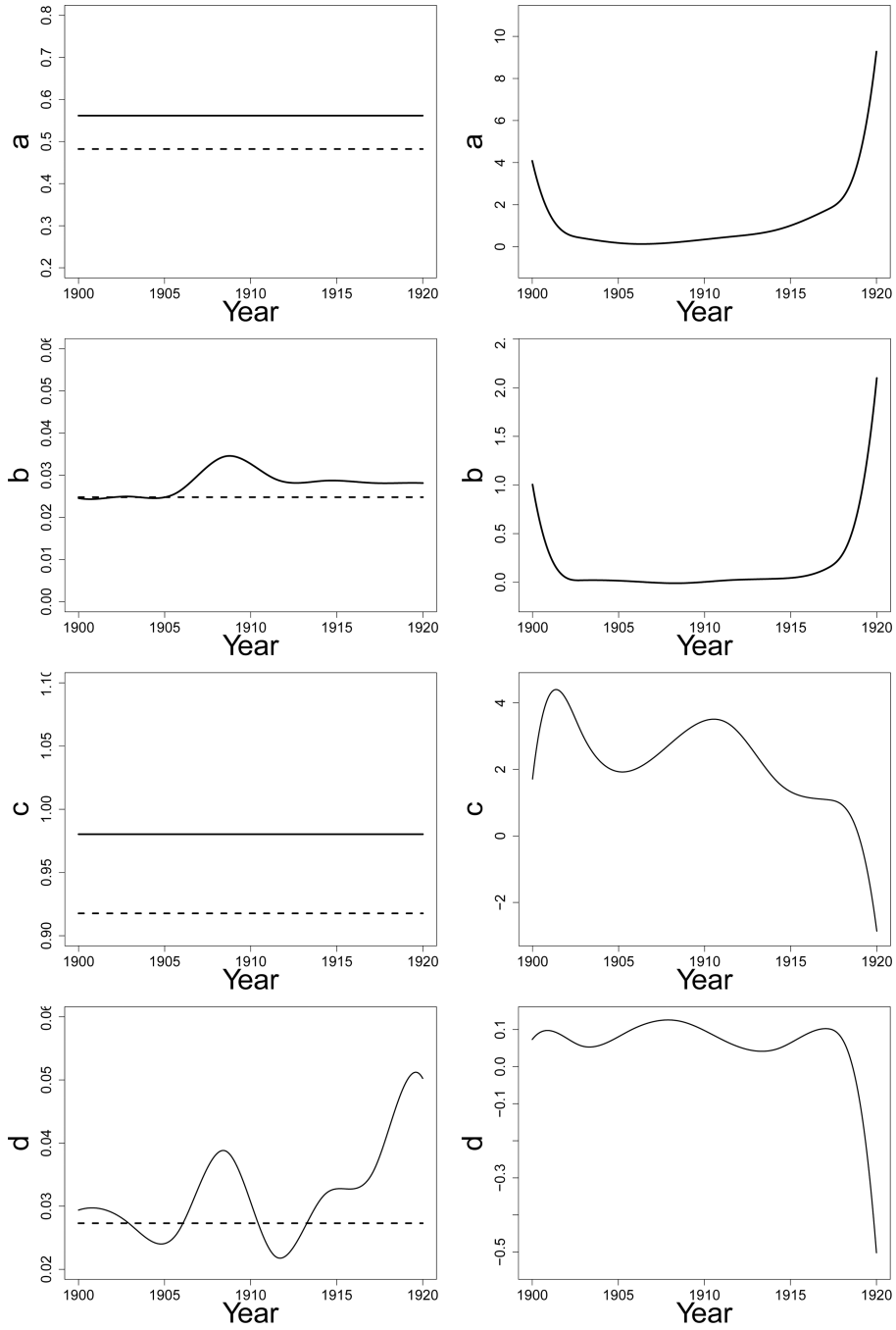

Figure 3.

Estimated parameter curves for the lynx-hare data set through the Lotka-Volterra model. Left: Method I (solid); Method III (dashed). Right: Method II (solid).

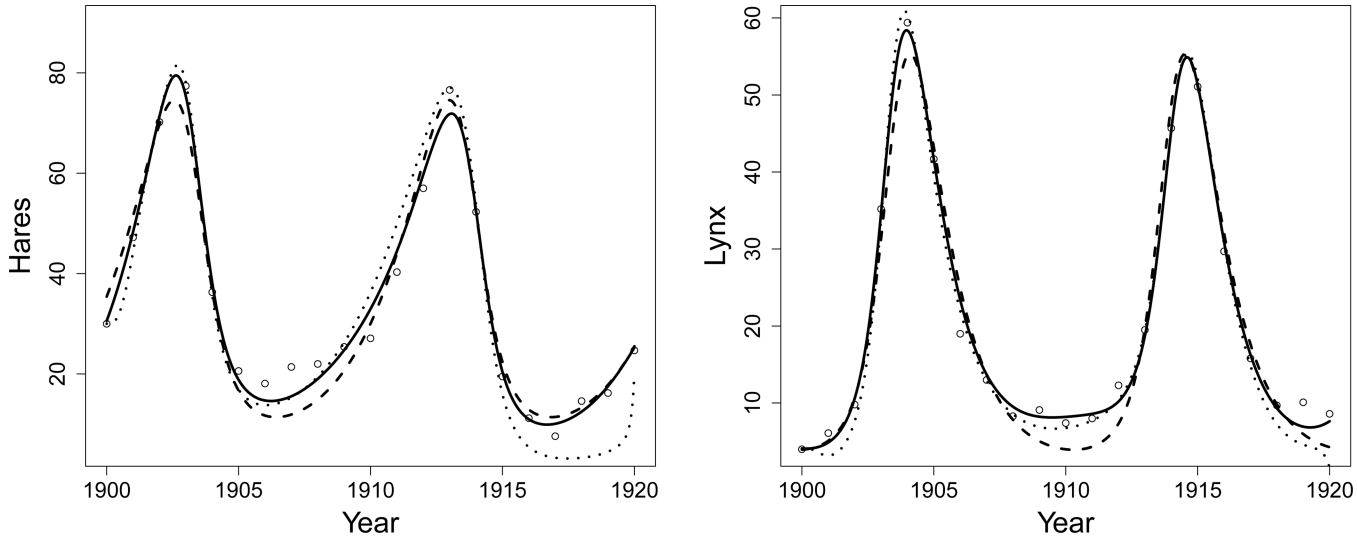

Figure 4.

Estimated $H(\cdot)$ and $L(\cdot)$ for the lynx-hare data set. Method I (solid); Method II (dotted); Method III (dashed).

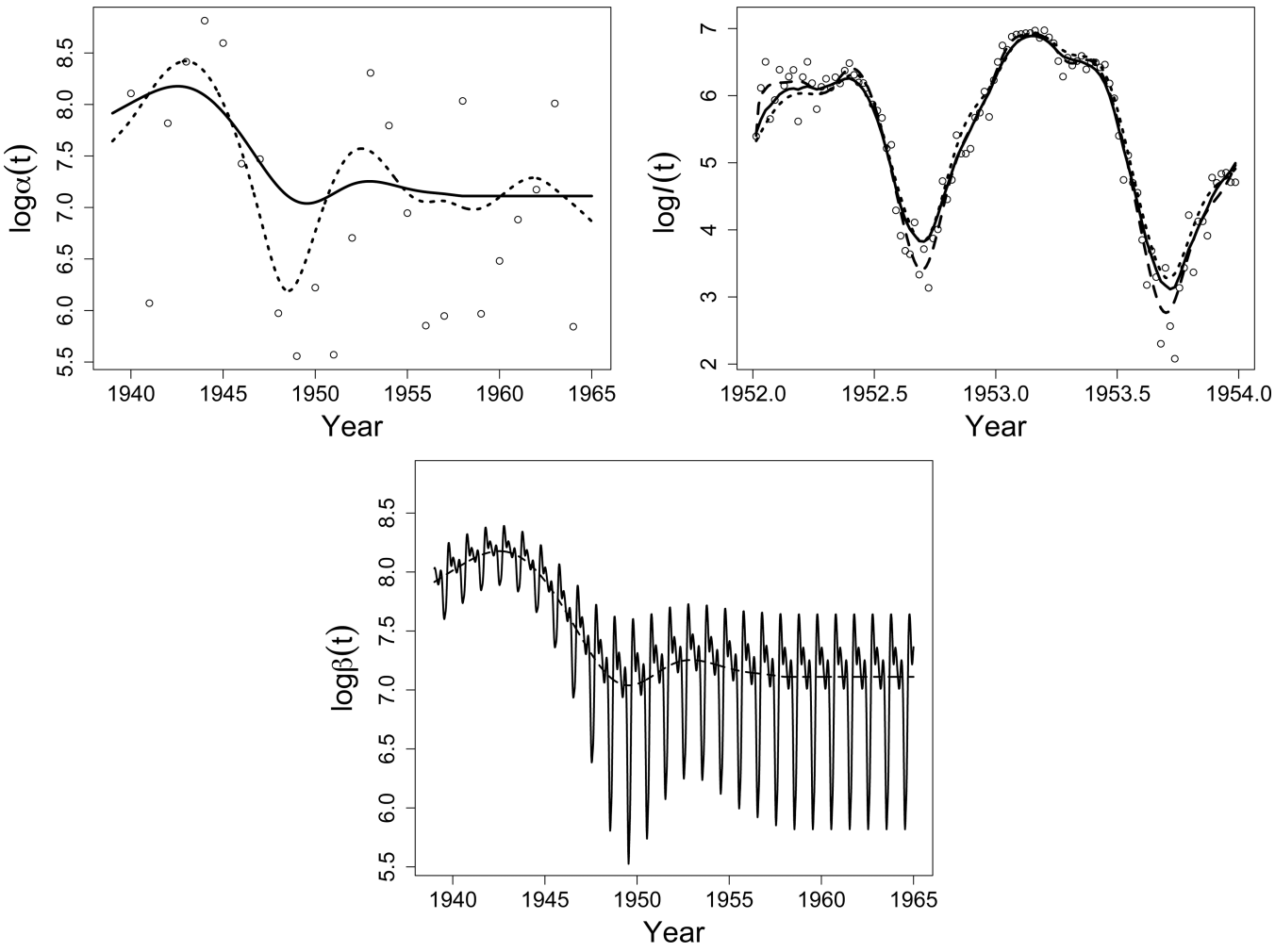

Figure 5.

Top Left: estimated $\alpha(t)$ based on the regularized method (solid) and the non-regularized method (dotted); circles are based on the two-year constant fitting. Top Right: estimated $I(t)$ for 1952–1954; regularized method (solid), non-regularized method (dotted) and two-year constant fitting (dashed). Bottom: regularized estimate of $\beta(t)$ (solid) with the regularized global time trend $\hat{\alpha}(t)$ (dashed) imposed.

Table 1

The average MISE and AVSE results of the four different estimation methods over 100 repetitions for the FitzHugh-Nagumo model.

	MISE Results							
	$\mathscr{F}_a$	$\mathscr{F}_a^c$	$\mathscr{F}_b$	$\mathscr{F}_b^c$	$\mathscr{F}_c$	$\mathscr{F}_c^c$	V	R
BIC	0.0016	0.0102	0.0039	0.0147	0.0090	0.0218	0.0907	0.0124
AIC	0.0168	0.0147	0.0358	0.0227	0.0123	0.0288	0.1101	0.0263
GCV	0.0162	0.0147	0.0344	0.0223	0.0122	0.0285	0.1094	0.0257
Method II	247.04*	46.004*	886.16*	27.778*	172.66	1.0638	0.1190	0.0546
Method III	0.0669	0.2597	0.0902	0.1085	4.2923	2.3361	3.2438	0.7382
Method IV	0.5208	0.5137	1.1124	1.8757	1.2768	1.7197	0.1987	0.1752
	AVSE Results							
	$\mathscr{F}_a$	$\mathscr{F}_a^c$	$\mathscr{F}_b$	$\mathscr{F}_b^c$	$\mathscr{F}_c$	$\mathscr{F}_c^c$	V	R
BIC	0.0083	0.0151	0.0104	0.0151	0.0062	0.0149	0.0347	0.0195
AIC	0.0330	0.0333	0.0466	0.0395	0.0145	0.0365	0.0470	0.0323
GCV	0.0324	0.0329	0.0454	0.0382	0.0142	0.0355	0.0467	0.0318
Method II	3.7785†	4.2775†	9.1491†	3.7527†	1.9718	0.3603	0.0559	0.0485
Method III	0.0264	0.0264	0.0728	0.0728	0.0363	0.0363	0.0605	0.0414
Method IV	0.1950	0.2271	0.2848	0.5275	0.3052	0.4887	0.0807	0.0897

*: median is reported instead of mean;

†: IQR is reported instead of standard error.

Table 2

The average MISE and AVSE results of the four different estimation methods over 100 repetitions for the Lotka-Volterra model.

	MISE Results									
	$\mathscr{F}_a$	$\mathscr{F}_a^c$	$\mathscr{F}_b$	$\mathscr{F}_b^c$	$\mathscr{F}_c$	$\mathscr{F}_c^c$	$\mathscr{F}_d$	$\mathscr{F}_d^c$	H	L
BIC	0.0003	0.0004	0.0001	0.0001	0.0003	0.0004	0.0003	0.0003	0.0686	0.2256
AIC	0.0003	0.0005	0.0003	0.0003	0.0004	0.0005	0.0005	0.0006	0.0753	0.3286
GCV	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005	0.0006	0.0769	0.3379
Method II	2.0791	0.0243	0.2595	0.0111	0.1354	0.0021	0.0091	0.0014	0.7312	2.4234
Method III	15.092	6.2885	0.6572	0.3224	1.6878	0.8086	0.1592	0.1568	67.461	319.75
Method IV	0.0020	0.0172	0.0005	0.0105	0.0010	0.0018	0.0004	0.0049	1.1003	4.1468
	AVSE Results									
	$\mathscr{F}_a$	$\mathscr{F}_a^c$	$\mathscr{F}_b$	$\mathscr{F}_b^c$	$\mathscr{F}_c$	$\mathscr{F}_c^c$	$\mathscr{F}_d$	$\mathscr{F}_d^c$	H	L
BIC	0.0012	0.0032	0.0013	0.0032	0.0012	0.0035	0.0013	0.0037	0.0095	0.0239
AIC	0.0013	0.0033	0.0013	0.0033	0.0012	0.0035	0.0013	0.0039	0.0096	0.0243
GCV	0.0012	0.0032	0.0015	0.0032	0.0013	0.0038	0.0014	0.0036	0.0098	0.0241
Method II	0.3178	0.0563	0.0974	0.0424	0.0496	0.0137	0.0139	0.0113	0.1648	0.2645
Method III	0.9365	0.9365	0.2106	0.2106	0.3479	0.3479	0.1043	0.1043	0.9214	1.6795
Method IV	0.0124	0.0440	0.0063	0.0309	0.0085	0.0162	0.0056	0.0200	0.1963	0.3553