Equation Matching

Anran Wang

November 9, 2022

Overview

1. Direct Integral Estimator
   - Direct Integral Estimation
   - Direct Integral Estimation of Partially Observed Systems
2. One-step Kernel Estimator
3. numerical discretization-based estimator

Model

The process of interest is usually modelled by the system

$$\begin{cases} \mathbf{x}'(t) = f(\mathbf{x}(t); \boldsymbol{\beta}), t \in [0, 1], \\ \mathbf{x}(0) = \boldsymbol{\xi}, \end{cases} \tag{1}$$

where $\mathbf{x}(t) \in \mathbb{R}^d, \boldsymbol{\xi} \in \Xi \subset \mathbb{R}^d$, and $\boldsymbol{\beta} \in \boldsymbol{N} \subset \mathbb{R}^q$. Given the values of $\boldsymbol{\xi}$ and $\boldsymbol{\beta}$, we denote the solution of the equality by

$$\mathbf{x}(t) = \mathbf{x}(t; \boldsymbol{\beta}, \boldsymbol{\xi}), t \in [0, 1].$$

The aim is to estimate the unknown parameter $\boldsymbol{\beta}$ (and if necessary $\boldsymbol{\xi}$) from noisy observations

$$Y_j(t_i) = x_j(t_i; \boldsymbol{\beta}, \boldsymbol{\xi}) + \varepsilon_{j,i}, i = 1, \ldots, n, j = 1, \ldots, r, \tag{2}$$

where $0 \leq t_1 < \cdots < t_n = T < \infty$ and $\varepsilon_{j,i}$ is the unobserved measurement error for x_j at time t_i.

Overview

1. Direct Integral Estimator
   - Direct Integral Estimation
   - Direct Integral Estimation of Partially Observed Systems
2. One-step Kernel Estimator
3. numerical discretization-based estimator

Integral Matching

Integral Matching is a method that

1. get nonparametric curve $\hat{x}$ that fit to x,
2. then, get the estimate $\hat{\beta}, \hat{\boldsymbol{\xi}}$ of $\beta, \boldsymbol{\xi}$ by minimizing $\|\hat{x}(\cdot) - \boldsymbol{\xi} - \int_0^{\cdot} f(\hat{x}(s); \beta)\, ds\|$, where $\|\cdot\|$ is a suitable norm.

Different between Integral Matching and Gradient Matching:

- The bias of Gradient Matching mainly from estimating $\hat{x}'$, while we do not need to consider the estimator of x' in Integral Matching.
- If the derivative of x does not exist in some positions, we cannot use Gradient Matching but can use Integral Matching.
- We can use discontinuous estimator (e.g. Step function estimator) of x in Integral Matching.

A class of ODEs

In this section, we consider the class of ODEs (3), where θ and ξ are identifiability.

In many applications states and parameters can be separated in the sense that there exist measurable functions $g : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^p$ and $h : N \to \mathbb{R}^p$ such that

$$f(x(t); \beta) = g(x(t))h(\beta) \tag{3}$$

holds. We write $\theta = h(\beta)$, $\theta \in \Theta = h(N) \subset \mathbb{R}^p$, and call it the natural parameter, where β is the parameter of interest.
The class of ODEs (3) is widely used in practice because of interpretability of the natural parameters as rate constants.

Interpretability of θ, ξ

Given the values of ξ and θ the solution of (1), (3) is denoted by

$$x(t) = x(t; \theta, \xi), \quad t \in [0, 1].$$

In the present context identifiability means that knowledge of a solution $x(t), t \in [0, 1]$, for the system (1), (3) yields the values of the parameters ξ and θ. For $\xi = x(0)$ this is obviously true, while identifiability of θ means that

$$\theta' \neq \theta \Rightarrow x\left(\cdot; \theta', \xi\right) \neq x(\cdot; \theta, \xi). \tag{4}$$

- **Under the model** (1) **and** (3), θ **is identifiability, i.e. equation** (4) **holds.**

Example: FitzHugh-Nagumo model

Next, we consider to estimate x by two estimators:

1. Smooth estimator: *Local polynomial estimator*.
2. Step function estimator.

- **First we consider an example using smooth estimator.**

FitzHugh-Nagumo model is used in neurophysiology as an approximation of the observed spike potential and takes the form

$$\begin{cases} x_1'(t) = \gamma \left(x_1(t) - x_1^3(t) + x_2(t) \right), \\ x_2'(t) = -(1/\gamma) \left(x_1(t) - \alpha + \beta x_2(t) \right). \end{cases} \tag{5}$$

The voltage $x_1(t)$ moving across the cell membrane depends on the recovery variable $x_2(t)$.

Example: FitzHugh-Nagumo model

The action of an excitable neuron has the following characteristics that we know from experiments:

- The neuron cell is initially at a resting potential value.
- If we experimentally displace the potential a little bit, it return to the resting value.
- If the perturbation is higher than a threshold value, the potential will shoot up to a very high value. In other words the spike will occur. After the spike the membrane potential will return to its resting value.

Example: FitzHugh-Nagumo model

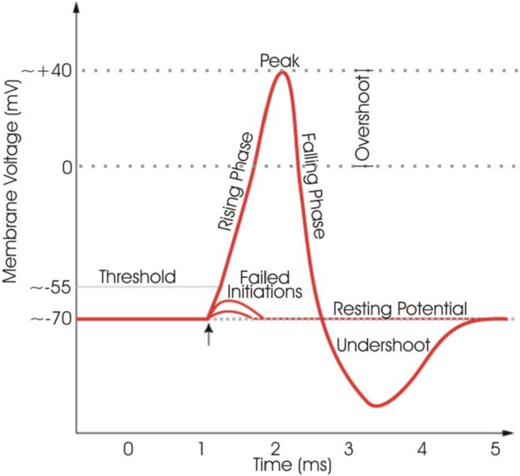

Figure 1: FitzHugh-Nagumo model

Example: Lotka-Volterra model

Lotka-Volterra system is a population dynamics model that describes evolution over time of the populations of two species, predators and their preys. In mathematical terms the Lotka-Volterra model is described by a system consisting of two equations and depending on the parameter $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)^T$. The system takes the form

$$\begin{cases} x_1'(t) = \theta_1 x_1(t) - \theta_2 x_1(t) x_2(t), \\ x_2'(t) = -\theta_3 x_2(t) + \theta_4 x_1(t) x_2(t). \end{cases} \tag{6}$$

Here x_1 represents the size of the prey population and x_2 of the predator population.

Example: Lotka-Volterra model

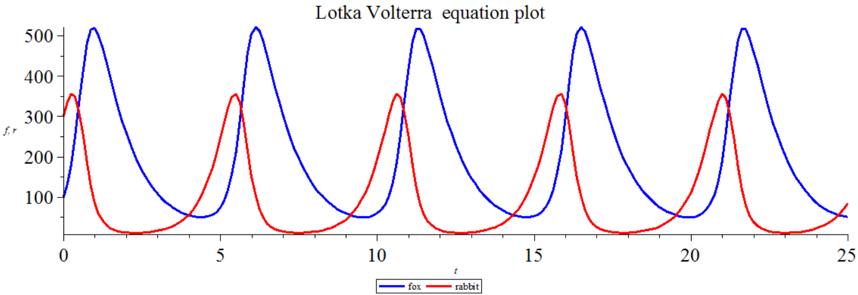

Figure 2: Lotka-Volterra model

A semidefinite inner product

In this subsection, we consider a kind of direct integral estimation.

Let W **be a symmetric** $d \times d$**-matrix of finite signed measures** on $([0,1], \mathcal{B})$ with $\mathcal{B}$ the sigma field of Borel sets, and let $x : [0,1] \to \mathbb{R}^d$ and $y : [0,1] \to \mathbb{R}^d$ be Borel measurable vector valued functions. We assume that W is chosen in such a way that

$$\langle x, y \rangle_W = \int_0^1 x^T(t) \mathrm{d}W(t) y(t) \tag{7}$$

is a semidefinite inner product and

$$\|x\|_W = \langle x, x \rangle_W^{1/2}$$

is the corresponding seminorm.

A semidefinite inner product

Denote by $w(\cdot)$ the $d \times d$-matrix of the Radon-Nikodym derivatives of the signed measures in W with respect to μ. Now it may be rewritten as

$$\langle x, y \rangle_W = \int_0^1 x^T(t) w(t) y(t) \mathrm{d}\mu(t).$$

If $x : [0,1] \to \mathbb{R}^{d \times k}$ and $y : [0,1] \to \mathbb{R}^{d \times \ell}$ are measurable matrix valued functions, then $\langle x, y \rangle_W$ will be interpreted as the $k \times \ell$ matrix of the inner products of the columns of x and of y.

$\| \cdot \|_W$ is well-defined and positive definite $\Leftrightarrow$ A_W is well-defined with finite entries and positive definite, where I_d is $d \times d$ identity matrix and A_W is $d \times d$ matrix that

$$A_W := \int_0^1 \mathrm{d}W(t) = \langle I_d, I_d \rangle_W .$$

A representation of θ, ξ with $\|\cdot\|_W$

Proposition 1 (prop 1. of Dattner and Klaassen (2015)).

Let $\xi \in \Xi$ *and* $\theta \in \Theta$ *with* Θ *an open subset of* $\mathbb{R}^p$. *Let* $x(t) = x(t; \theta, \xi), t \in [0, 1]$, *satisfy the system* (1), (3) *and write*

$$G(t) = \int_0^t g(x(s))\mathrm{d}s, \quad t \in [0, 1].$$

Let W be a symmetric d × d-matrix of signed measures as in (5) satisfying (8) and having the other properties mentioned above. Assume that the d × p-and p × p-matrices

$$B_W = \langle I_d, G \rangle_W \,, \quad C_W = \langle G, G \rangle_W$$

are well-defined with finite entries.

A representation of θ, ξ with $\|\cdot\|_W$

1. *If C_W is nonsingular then $A_W - B_W C_W^{-1} B_W^T$ is nonsingular and*

$$\xi = \left(A_W - B_W C_W^{-1} B_W^T\right)^{-1} \left\langle I_d - G C_W^{-1} B_W^T, x \right\rangle_W, \tag{8}$$

$$\theta = C_W^{-1} \left(\langle G, x \rangle_W - B_W^T \xi \right) \tag{9}$$

   hold.

2. *Conversely, if knowledge for all $i = 1, \ldots, d$ of $x_i(t)$ for W_{ii}-almost all $t \in [0,1]$ determines θ, then C_W is nonsingular.*

Remarks:

- Item 2 is said that if θ is interpretability i.e. (4), then C_W is nonsingular W.a.e.
- Via (9) we have $\beta = h^{-1}\left(C_W^{-1}\left(\langle G, x\rangle_W - B_W^T \xi\right)\right)$.

Sketch proof of Proposition 1

1. Sketch proof of item 1.
   - $\partial\|x(t) - \boldsymbol{\xi} - \int_0^t g(x(s))ds\ \theta\|_W/\partial\theta = 0 \Rightarrow$ equation (9) holds.
   - Equation (9) and $\partial\|x(t) - \boldsymbol{\xi} - \int_0^t g(x(s))ds\ \theta\|_W/\partial\xi = 0 \Rightarrow$ equation (8) holds.
2. Sketch proof of item 2.
   C_W is singular if and only if there exists a p-vector $\eta \neq 0$ with

$$C_W\eta = \langle G, G\eta\rangle_W = 0,$$

which implies

$$\eta^T C_W\eta = \|G\eta\|_W^2 = 0.$$

Consequently, $G(t)\eta = 0$ for W-almost all $t \in [0, 1]$.
Then $x(t; \theta + \eta, \xi) = x(t; \theta, \xi)$ also satisfy the system (1), (3), which contradicts (4).

□

Methodological approach I

Let $\widehat{x}_n(t), t \in [0,1]$, be an estimator of $x(t;\theta,\xi)$ based on the observations (2). It makes sense to estimate the parameters θ and ξ by minimizing

$$\left\| \widehat{x}_n(t) - \zeta - \int_0^t g\left(\widehat{x}_n(s)\right) \mathrm{d}s\eta \right\|_{W_n}^2 \tag{10}$$

over η and ζ, where W_n is an appropriate $d \times d$-matrix of signed measures on $([0,1], \mathcal{B})$ as in Proposition 1. Denote

$$\begin{aligned} \widehat{G}_n(t) &= \int_0^t g\left(\widehat{x}_n(s)\right) \mathrm{d}s, \quad t \in [0,1], \\ A_n = \langle I_d, I_d \rangle_{W_n}, \quad \widehat{B}_n &= \left\langle I_d, \widehat{G}_n \right\rangle_{W_n}, \quad \widehat{C}_n = \left\langle \widehat{G}_n, \widehat{G}_n \right\rangle_{W_n}. \end{aligned} \tag{11}$$

Methodological approach II

Minimizing the criterion function (10) with respect to ζ and η results in the direct estimators (cf. (9) and (8))

$$\begin{aligned} \widehat{\xi}_n &= \left(A_n - \widehat{B}_n \widehat{C}_n^{-1} \widehat{B}_n^T\right)^{-1} \left\langle I_d - \widehat{G}_n \widehat{C}_n^{-1} \widehat{B}_n^T, \widehat{x}_n \right\rangle_{W_n}, \\ \widehat{\theta}_n &= \widehat{C}_n^{-1} \left(\left\langle \widehat{G}_n, \widehat{x}_n \right\rangle_{W_n} - \widehat{B}_n^T \widehat{\xi}_n \right). \end{aligned} \tag{12}$$

In case the initial value ξ is known, the second equation may be used with $\widehat{\xi}_n$ replaced by ξ.

In order to estimate the parameter of interest ν we choose a distance function $d_n(\cdot, \cdot)$ on $\mathbb{R}^p$ and we choose $\widehat{\beta}_n$ in such a way that

$$d_n\left(h\left(\widehat{\beta}_n\right), \widehat{\theta}_n\right) \leq \inf_{\beta \in N} d_n\left(h(\beta), \widehat{\theta}_n\right) + \frac{1}{n} \tag{13}$$

holds. Of course, if the infimum is attained, we choose $\widehat{\beta}_n$ as the minimizer.

Example: FitzHugh-Nagumo model

First, we consider the FitzHugh-Nagumo model (5).

The process to estimate γ, α, β:

1. Take the model in to the form (3).
   Set $\nu = (\alpha, \beta, \gamma)^T$, then the system takes the form (3) with $\theta = h(\nu) = (\gamma, 1/\gamma, \alpha/\gamma, \beta/\gamma)^T$ and the corresponding matrix g is

$$g(x(t)) = \begin{pmatrix} x_1(t) - x_1^3(t) + x_2(t) & 0 & 0 & 0 \\ 0 & -x_1(t) & 1 & -x_2(t) \end{pmatrix}.$$

2. Do Integral Matching. That is, get local polynomial estimator of x, then solve equation (12).
3. Take suitable $d_n(\cdot, \cdot)$ in equation (13). Then solve equation (13).
   We can take d_n to be the Mahalanobis distance:

$$d_n(x, y) = \sqrt{(x-y)^T \widehat{\Sigma}_n^{-1} (x-y)}, \tag{14}$$

where $\widehat{\Sigma}_n$ is the estimated covariance matrix of $\widehat{\theta}_n$.

Example: FitzHugh-Nagumo model

Estimating $\widehat{\Sigma}_n$ in (14) by a bootstrap as following.

1. Repeat B times the following steps:
   - For each point t_i generate residuals $\widetilde{\varepsilon}_j(t_i) = Y_j(t_i) - \widehat{x}_j(t_i)$.
   - Center the residuals: $\bar{\varepsilon}_j(t_i) = \widetilde{\varepsilon}_j(t_i) - \frac{1}{n}\sum_{i=1}^{n}\widetilde{\varepsilon}_j(t_i)$.
   - Sample n residuals (with replacement) from $\bar{\varepsilon}_j(t_1), \ldots, \bar{\varepsilon}_j(t_n)$ to obtain the bootstrap residuals $\varepsilon_j^{\star}(t_1), \ldots, \varepsilon_j^{\star}(t_n)$.
   - Set $Y_j^{\star}(t_i) = \widehat{x}_j(t_i) + \varepsilon_j^{\star}(t_i), i = 1, \ldots, n$.
2. We then use the bootstrap sample $Y_j^{\star}(t_1), \ldots, Y_j^{\star}(t_n)$ and apply the estimation procedure. Denote the estimator for the vector θ in the bth bootstrap sample by $\widehat{\theta}^*_{n,b}$ and its corresponding average over the B bootstrap samples by $\overline{\theta^*_n}$. Then we define

$$\widehat{\Sigma}_n = \frac{1}{B}\sum_{b=1}^{B}\left\{\left(\widehat{\theta}^*_{n,b} - \overline{\theta^*_n}\right)\left(\widehat{\theta}^*_{n,b} - \overline{\theta^*_n}\right)^T\right\}.$$

Example: FitzHugh-Nagumo model

Comparison of the results in Dattner and Klaassen (2015)(left) and Liang and Wu (2008)(right).

TABLE 1

Empirical means (standard deviations) for estimating the parameters of the FitzHugh-Nagumo system*. The parameters are estimated by (19)–(20) and (38), with (29) a local polynomial estimator of order $\ell = 1$ and bandwidth $b = n^{-1/3}$. Based on 500 Monte Carlo simulations. The true parameter vector is $(\alpha, \beta, \gamma)^T = (0.34, 0.2, 3)^T$. The two signals are first generated by solving the system at 0.1 time units on the interval $[0, 20]$ $(n = 201)$ and then adding Gaussian measurement errors with zero mean and variances σ_1^2, σ_2^2 respectively*

		Parameters		
σ_1^2	σ_2^2	α	β	γ
0.050	0.050	0.339 (0.004)	0.200 (0.022)	3.005 (0.033)
	0.060	0.340 (0.005)	0.203 (0.024)	3.004 (0.039)
	0.070	0.340 (0.004)	0.202 (0.029)	3.004 (0.043)
	0.080	0.340 (0.005)	0.204 (0.030)	3.004 (0.046)
	0.090	0.340 (0.005)	0.202 (0.034)	3.010 (0.055)
	0.100	0.340 (0.005)	0.206 (0.038)	3.006 (0.059)
0.060	0.050	0.339 (0.005)	0.201 (0.023)	2.999 (0.034)
	0.060	0.340 (0.005)	0.200 (0.026)	3.002 (0.039)
	0.070	0.340 (0.005)	0.202 (0.030)	3.005 (0.047)
	0.080	0.340 (0.005)	0.201 (0.033)	3.008 (0.049)
	0.090	0.340 (0.006)	0.204 (0.036)	3.006 (0.058)
	0.100	0.340 (0.006)	0.202 (0.039)	3.001 (0.064)
0.070	0.050	0.339 (0.005)	0.201 (0.024)	2.998 (0.036)
	0.060	0.339 (0.005)	0.199 (0.027)	2.999 (0.041)
	0.070	0.339 (0.006)	0.200 (0.029)	3.001 (0.047)
	0.080	0.339 (0.006)	0.200 (0.031)	3.003 (0.052)
	0.090	0.340 (0.006)	0.204 (0.037)	3.003 (0.055)
	0.100	0.340 (0.007)	0.206 (0.041)	3.002 (0.065)
0.080	0.050	0.338 (0.006)	0.201 (0.024)	3.001 (0.038)
	0.060	0.339 (0.006)	0.201 (0.028)	2.999 (0.042)
	0.070	0.339 (0.006)	0.204 (0.029)	3.003 (0.048)
	0.080	0.339 (0.006)	0.199 (0.035)	2.997 (0.052)
	0.090	0.339 (0.007)	0.197 (0.035)	3.000 (0.056)
	0.100	0.340 (0.007)	0.206 (0.041)	2.994 (0.064)
0.090	0.050	0.339 (0.006)	0.201 (0.025)	2.998 (0.039)
	0.060	0.339 (0.006)	0.201 (0.031)	3.000 (0.046)
	0.070	0.339 (0.007)	0.201 (0.031)	3.000 (0.051)
	0.080	0.339 (0.007)	0.200 (0.034)	3.004 (0.055)
	0.090	0.339 (0.007)	0.201 (0.038)	2.997 (0.061)
	0.100	0.339 (0.008)	0.202 (0.043)	2.999 (0.066)
0.100	0.050	0.338 (0.007)	0.198 (0.028)	3.000 (0.041)
	0.060	0.339 (0.007)	0.201 (0.030)	2.999 (0.048)
	0.070	0.339 (0.007)	0.202 (0.034)	2.995 (0.051)
	0.080	0.339 (0.008)	0.202 (0.035)	3.000 (0.056)
	0.090	0.338 (0.008)	0.202 (0.041)	2.998 (0.060)
	0.100	0.339 (0.008)	0.198 (0.040)	3.000 (0.063)

			PsLS(s.e., CP)	
σ_1^2	σ_2^2	α	β	γ
0.05	0.05	0.330(0.08,93.4)	0.223(0.12,94.0)	3.076(0.17,93.0)
	0.06	0.336(0.08,97.3)	0.216(0.12,92.0)	3.069(0.17,96.2)
	0.07	0.335(0.09,94.3)	0.213(0.12,97.0)	3.079(0.18,94.5)
	0.08	0.333(0.10,94.0)	0.214(0.13,92.0)	3.086(0.18,94.1)
	0.09	0.344(0.10,98.0)	0.216(0.14,92.0)	3.089(0.18,93.8)
	0.1	0.341(0.10,98.0)	0.220(0.12,94.0)	3.081(0.18,97.3)
0.06	0.05	0.336(0.08,98.1)	0.221(0.10,95.0)	3.038(0.18,93.3)
	0.06	0.338(0.08,96.0)	0.216(0.13,91.0)	3.038(0.17,91.4)
	0.07	0.338(0.09,97.1)	0.217(0.11,97.0)	3.040(0.18,92.1)
	0.08	0.341(0.08,94.9)	0.218(0.14,92.0)	3.037(0.18,94.6)
	0.09	0.337(0.10,96.7)	0.211(0.12,94.2)	3.037(0.19,97.4)
	0.1	0.338(0.11,98.0)	0.209(0.16,92.0)	3.051(0.20,95.3)
0.07	0.05	0.331(0.07,97.2)	0.219(0.10,94.0)	3.017(0.19,93.7)
	0.06	0.334(0.08,97.0)	0.223(0.11,92.0)	3.022(0.17,94.3)
	0.07	0.336(0.10,95.0)	0.219(0.11,93.2)	3.018(0.18,94.3)
	0.08	0.337(0.11,95.0)	0.212(0.11,92.0)	3.018(0.20,95.8)
	0.09	0.336(0.11,95.0)	0.220(0.15,94.0)	3.023(0.19,96.1)
	0.1	0.340(0.11,92.0)	0.218(0.14,93.0)	3.008(0.19,94.8)
0.08	0.05	0.335(0.08,96.0)	0.221(0.10,92.0)	2.984(0.20,94.2)
	0.06	0.336(0.09,94.0)	0.217(0.12,95.0)	2.990(0.18,93.6)
	0.07	0.337(0.09,98.0)	0.213(0.10,94.5)	2.989(0.20,93.6)
	0.08	0.337(0.10,92.0)	0.215(0.12,96.0)	2.975(0.19,93.4)
	0.09	0.339(0.10,94.0)	0.211(0.14,93.0)	2.991(0.19,93.8)
	0.1	0.347(0.10,93.0)	0.212(0.15,92.0)	2.994(0.22,92.3)
0.09	0.05	0.331(0.08,95.8)	0.227(0.12,93.0)	2.976(0.17,92.9)
	0.06	0.336(0.08,96.0)	0.226(0.10,92.0)	2.961(0.21,93.2)
	0.07	0.338(0.10,97.0)	0.219(0.11,98.0)	2.963(0.20,92.9)
	0.08	0.342(0.10,97.0)	0.219(0.13,94.0)	2.983(0.21,95.8)
	0.09	0.338(0.08,96.0)	0.217(0.13,93.2)	2.966(0.21,96.7)
	0.1	0.342(0.10,90.0)	0.210(0.15,94.0)	2.966(0.20,93.2)
0.1	0.05	0.334(0.07,98.2)	0.220(0.11,93.1)	2.951(0.20,93.7)
	0.06	0.338(0.08,96.9)	0.214(0.10,96.0)	2.946(0.20,95.7)
	0.07	0.338(0.08,94.0)	0.221(0.11,92.1)	2.948(0.22,96.3)
	0.08	0.341(0.09,97.0)	0.212(0.13,93.0)	2.936(0.20,93.7)
	0.09	0.344(0.11,93.0)	0.209(0.14,92.1)	2.954(0.22,97.5)
	0.1	0.345(0.10,91.0)	0.211(0.12,92.3)	2.945(0.20,92.8)

Example: FitzHugh-Nagumo model

		Integral			Derivative		
σ_1^2	σ_2^2	α	β	γ	α	β	γ
0.05	0.05	0.99	8.85	0.91	6.21	17.77	16.33
	0.06	1.06	9.90	1.04	7.27	17.36	15.83
	0.07	1.04	11.83	1.13	7.21	20.63	15.66
	0.08	1.15	12.17	1.24	7.17	26.96	14.53
	0.09	1.22	13.60	1.52	7.27	30.60	14.16
	0.10	1.28	15.03	1.57	7.72	24.42	14.08
0.06	0.05	1.13	9.26	0.90	6.70	16.66	18.38
	0.06	1.13	10.29	1.04	7.33	18.00	17.76
	0.07	1.24	11.94	1.23	6.06	20.85	17.27
	0.08	1.26	13.14	1.34	5.75	26.67	16.97
	0.09	1.38	14.45	1.53	7.32	22.79	16.55
	0.10	1.51	15.33	1.68	7.90	29.71	16.07
0.07	0.05	1.24	9.55	0.96	6.44	14.62	19.22
	0.06	1.30	11.15	1.09	7.70	18.72	18.65
	0.07	1.38	11.75	1.25	7.95	17.30	18.59
	0.08	1.41	12.84	1.42	6.66	19.37	18.08
	0.09	1.50	14.86	1.49	8.18	27.57	17.63
	0.10	1.59	16.87	1.73	8.09	29.94	18.14
0.08	0.05	1.36	9.80	1.00	6.28	16.41	20.94
	0.06	1.51	10.95	1.10	6.90	21.51	20.14
	0.07	1.53	11.90	1.30	7.33	18.55	20.07
	0.08	1.47	13.85	1.40	7.95	21.39	20.23
	0.09	1.65	14.31	1.52	7.78	25.05	18.61
	0.10	1.70	16.51	1.73	7.75	30.93	18.86
0.09	0.05	1.48	10.01	1.03	7.31	17.76	21.77
	0.06	1.49	12.25	1.21	7.22	21.76	21.48
	0.07	1.57	12.57	1.35	7.38	15.44	21.18
	0.08	1.64	13.53	1.47	7.38	22.85	20.30
	0.09	1.69	14.99	1.61	7.04	28.70	20.33
	0.10	1.86	17.21	1.76	8.45	29.78	20.39
0.10	0.05	1.61	11.17	1.11	6.42	18.89	22.68
	0.06	1.62	12.04	1.29	6.78	19.33	21.87
	0.07	1.77	13.54	1.37	6.62	22.09	21.79
	0.08	1.79	13.85	1.47	7.80	23.20	22.12
	0.09	1.84	15.87	1.60	8.30	24.40	20.85
	0.10	1.92	15.91	1.68	8.57	26.50	20.99

TABLE 2
Comparison of the empirical relative estimation error (ARE) in FitzHugh-Nagumo system, of the integral approach (Integral, Dattner and Klaassen (2015)) with that of the derivative based two-step method (Derivative, Liang and Wu (2008)). For the Integral estimator with a local polynomial estimator of order $\ell = 1$ and bandwidth $b = n^{-1/3}$. Based on 500 Monte Carlo simulations. The true parameter vector is $(\alpha, \beta, \gamma)^T = (0.34, 0.2, 3)^T$. The two signals are first generated by solving the system at 0.1 time units on the interval $[0, 20]$ $(n = 201)$ and then adding Gaussian measurement errors with zero mean and variances σ_1^2, σ_2^2 respectively.

error (ARE). The ARE of a real-valued parameter a over the M Monte Carlo simulations is defined as

$$ARE(a) = \frac{1}{M} \sum_{m=1}^{M} \frac{|\widehat{a}_m - a|}{|a|} \times 100\%. \quad (15)$$

Example: FitzHugh-Nagumo model

Comparison of the results in Dattner and Klaassen (2015) and Ramsay et al. (2007).

TABLE 3

Comparison of the empirical means (standard deviation) in FitzHugh-Nagumo system, of the integral approach with that of the generalized profiling of [38]. *For the Integral estimator the parameters are estimated by (19)–(20) and (38), with (29) a local polynomial estimator of order $\ell = 1$ and bandwidth $b = n^{-1/3}$. Based on 500 Monte Carlo simulations. The true parameter vector is $(a, b, c)^T = (0.2, 0.2, 3)^T$. The two signals are first generated by solving the system at 0.05 time units on the interval $[0, 20]$ ($n = 401$) and then adding Gaussian measurement errors with zero mean and variances $\sigma_1^2 = \sigma_2^2 = 0.5$*

	a	b	c
Generalized profiling (Ramsay et al. (2007)	0.2005 (0.0149)	0.1984 (0.0643)	2.9949 (0.0264)
Generalized profiling (here)	0.2003 (0.0166)	0.1986 (0.0679)	3.0010 (0.0795)
Integral estimator	0.1906 (0.0307)	0.1859 (0.0905)	2.9249 (0.1216)

Example: FitzHugh-Nagumo model

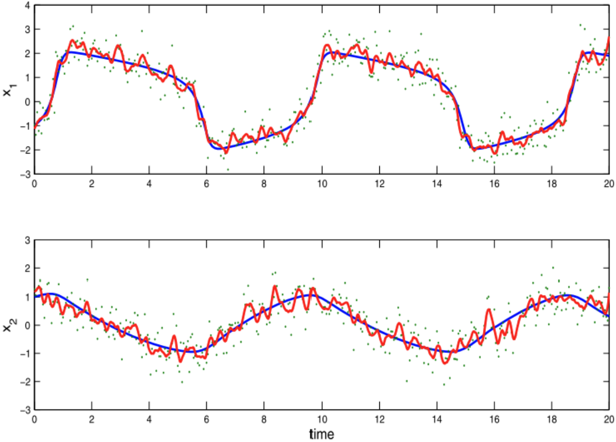

FIG 1. *The FitzHugh-Nagumo system (setup as in Table 3).* ***Top panel.*** *The blue line corresponds to the true solution x_1; the dots correspond to one realization of observations; the red line corresponds to the nonparametric estimator of x_1.* ***Bottom panel.*** *The same as the top but for x_2.*

Step function estimator

- **Next we consider an example using step function estimator.**

While choosing the smoothing parameter in practice may not be trivial. The problem (2) can be avoided in situations like the following repeated measures model,

$$Y^{(j)}(t_i) = x(t_i; \theta, \xi) + \varepsilon^{(j)}(t_i), \quad j = 1, \ldots, J_i, \quad i = 1, \ldots, I, \tag{16}$$

with $t_i = i/I, i = 1, \ldots, I$. Hence, we have $n = \sum_{i=1}^{I} J_i$ observations in total. Within this observation scheme it is natural to estimate $x(t_i)$ by

$$\widehat{x}_n(t_i) = \frac{1}{J_i} \sum_{j=1}^{J_i} Y^{(j)}(t_i),$$

and the step function estimator of $x(t)$ is

$$\widehat{x}_n(t) = \frac{1}{J_i} \sum_{j=1}^{J_i} Y^{(j)}(t_i), \quad (i-1)/I < t \leq i/I, \quad i = 1, \ldots, I,$$

where we complete the definition of $\widehat{x}_n(t)$ on $[0, 1]$ by $\widehat{x}_n(0) = \widehat{x}_n(t_1)$.

Example: Lotka-Volterra model

Then, we consider the Lotka-Volterra model (6).

TABLE 4

Gaussian error – empirical means (standard deviation) in Lotka-Volterra system based on 5000 ***Monte Carlo simulations.*** *In each simulation the data consist of* $I = 30$ *noisy observations of* x_1 *and* x_2 *according to measurement error model (33) with* $\sigma_\varepsilon^2 = 0.5$. *The samples were taken at* 0.5 *time units on the interval* $[0, 14.9]$ *for the first parameters setup and at* 1 *time unit on the interval* $[0, 29.9]$ *for the second. At each time point,* J *repeated measures were generated. The parameters are estimated by (19)–(20) with (34) a step function estimator. Last two lines in each block correspond to the empirical means (standard deviation) of the distribution of* $\{\frac{1}{T}\int_0^T \|x(t;\widehat{\theta}_n,\widehat{\xi}_n) - x(t;\theta,\xi)\|^2 \mathrm{d}t\}^{1/2}$ *and* $\|x(\cdot;\widehat{\theta}_n,\widehat{\xi}_n) - x(\cdot;\theta,\xi)\|_\infty$ *respectively*

	Value	$J = 6$	$J = 10$	$J = 15$	$J = 30$
ξ_1	1.000	1.089 (0.143)	1.085 (0.116)	1.085 (0.093)	1.083 (0.065)
ξ_2	0.500	0.446 (0.096)	0.441 (0.075)	0.438 (0.061)	0.436 (0.043)
θ_1	0.500	0.468 (0.077)	0.473 (0.061)	0.474 (0.050)	0.477 (0.035)
θ_2	0.500	0.473 (0.075)	0.477 (0.060)	0.479 (0.048)	0.480 (0.034)
θ_3	0.500	0.500 (0.073)	0.501 (0.057)	0.501 (0.047)	0.501 (0.033)
θ_4	0.500	0.508 (0.073)	0.508 (0.057)	0.509 (0.047)	0.509 (0.033)
		0.214 (0.082)	0.183 (0.066)	0.167 (0.056)	0.148 (0.041)
		0.344 (0.142)	0.290 (0.111)	0.264 (0.094)	0.233 (0.069)
ξ_1	0.500	0.289 (0.160)	0.296 (0.130)	0.301 (0.109)	0.300 (0.076)
ξ_2	1.000	1.000 (0.237)	1.038 (0.185)	1.052 (0.153)	1.070 (0.107)
θ_1	0.200	0.174 (0.040)	0.178 (0.031)	0.180 (0.026)	0.182 (0.019)
θ_2	0.700	0.496 (0.159)	0.525 (0.136)	0.536 (0.115)	0.546 (0.085)
θ_3	0.300	0.305 (0.085)	0.316 (0.069)	0.318 (0.058)	0.320 (0.041)
θ_4	0.500	0.477 (0.126)	0.483 (0.096)	0.481 (0.078)	0.479 (0.054)
		0.443 (0.249)	0.385 (0.189)	0.341 (0.166)	0.284 (0.137)
		1.003 (0.609)	0.890 (0.401)	0.798 (0.365)	0.674 (0.312)

Direct integral estimation of partially observed systems

In this subsection, we think about a kind of direct integral estimation of partially observed systems in Vujačić and Dattner (2018).

In this paper, we consider the estimators of the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ are obtained by minimizing

$$\int_0^T \left\| \widehat{\mathbf{x}}(t) - \boldsymbol{\xi} - \int_0^t f(\widehat{\mathbf{x}}(s)) \mathrm{d}s \boldsymbol{\theta} \right\|^2 \mathrm{d}t$$

with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$, where $\widehat{\mathbf{x}}(t), t \in [0, T]$, is a specific estimator of $\mathbf{x}(t; \boldsymbol{\theta}, \boldsymbol{\xi})$. Here and subsequently, $\| \cdot \|$ denotes the Euclidean norm. It is the case that take $W = \mathbb{I}_d$ in (10). The estimators $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\xi}})$ of the above equation is similar to (12).

Direct integral estimation of partially observed systems

We now describe the construction of the estimator for partially observed systems. Let $\mathcal{M}$ and $\mathcal{U}$ denote the sets of r-dimensional and $(d-r)$-dimensional vector functions on $[0, T]$ that correspond to $\mathbf{m}(\cdot) = \mathbf{m}(\cdot, \boldsymbol{\theta}, \boldsymbol{\xi})$ and $\mathbf{u}^*(\cdot) = \mathbf{u}^*(\cdot, \boldsymbol{\theta}, \boldsymbol{\xi})$, the measured and unmeasured components, respectively. We first construct an estimator $\widehat{\mathbf{m}}_n(\cdot)$ of $\mathbf{m}(\cdot)$ using the observations (1) and for a given $\mathbf{u} \in \mathcal{U}$.

we define

$$\widehat{\mathbf{x}}_{\mathbf{u}}(t) = \left(\widehat{\mathbf{m}}_n(t), \mathbf{u}(t)\right), \quad \widehat{f}_{\mathbf{u}}(t) = \int_0^t f\left(\widehat{\mathbf{x}}_{\mathbf{u}}(s)\right) \mathrm{d}s,$$

$$\widehat{\mathbf{B}}_{\mathbf{u}} = \int_0^T \widehat{f}_{\mathbf{u}}(t) \mathrm{d}t, \qquad \widehat{\mathbf{C}}_{\mathbf{u}} = \int_0^T \widehat{f}_{\mathbf{u}}^T(t) \widehat{f}_{\mathbf{u}}(t) \mathrm{d}t.$$

It is the case that take $W = \mathbb{I}_d$ and treat $\mathbf{u}(t)$ as a variable in (11), and here $\mathbf{A}_{\mathbf{u}} \equiv T$.

Direct integral estimation of partially observed systems

According to (12), the direct integral estimator based on $\widehat{\mathbf{x}}_{\mathbf{u}}(\cdot)$ is

$$\widehat{\boldsymbol{\xi}}_{\mathbf{u}} = \left(T\mathbf{I}_d - \widehat{\mathbf{B}}_{\mathbf{u}}\widehat{\mathbf{C}}_{\mathbf{u}}^{-1}\widehat{\mathbf{B}}_{\mathbf{u}}^{\top}\right)^{-1} \int_0^T \left\{\mathbf{I}_d - \widehat{\mathbf{B}}_{\mathbf{u}}\widehat{\mathbf{C}}_{\mathbf{u}}^{-1}\widehat{f}_{\mathbf{u}}^{\top}(t)\right\} \widehat{\mathbf{x}}_{\mathbf{u}}(t)\mathrm{d}t,$$

$$\widehat{\boldsymbol{\theta}}_{\mathbf{u}} = \widehat{\mathbf{C}}_{\mathbf{u}}^{-1} \int_0^T \widehat{f}_{\mathbf{u}}^{\top}(t) \left\{\widehat{\mathbf{x}}_{\mathbf{u}}(t) - \widehat{\boldsymbol{\xi}}_{\mathbf{u}}\right\} \mathrm{d}t.$$

In case $\mathbf{u}^*$ is measured, we can construct $\widehat{\mathbf{u}}_n(\cdot)$ and the above estimator reduces to the one from (12). Otherwise, for a subset $\mathcal{U}_n \subset \mathcal{U}$ let

$$\widehat{\mathbf{u}}_n := \operatorname{argmin}_{\mathbf{u}\in\mathcal{U}_n} M_n(\mathbf{u}),$$

Direct integral estimation of partially observed systems

where

$$M_n(\mathbf{u}) = \int_0^T \left\| \widehat{\mathbf{x}}_{\mathbf{u}}(t) - \widehat{\boldsymbol{\xi}}_u - \widehat{\mathbf{G}}_{\mathbf{u}}(t)\widehat{\boldsymbol{\theta}}_{\mathbf{u}} \right\|^2 \, \mathrm{d}t$$

The estimators of initial value $\boldsymbol{\xi}$ and the parameter $\boldsymbol{\theta}$ are

$$\begin{aligned} \widehat{\boldsymbol{\xi}}_n &:= \widehat{\boldsymbol{\xi}}_{\widehat{\mathrm{u}}_n}, \\ \widehat{\boldsymbol{\theta}}_n &:= \widehat{\boldsymbol{\theta}}_{\widehat{\mathrm{u}}_n}. \end{aligned} \tag{17}$$

See further about consistency of $\left(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\xi}}_n\right)$ given by (17) under suitable assumptions in Vujačić and Dattner (2018) .

One-step kernel estimator

In this section, we introduce A class of one-step kernel estimators in Hall and Ma (2014).

Features of this method:

- It is one-step parameter estimation: It get the estimators of θ and the parameter (bandwidth h) that estimating x, both from minimize a kind of error about the ODEs.
- It uses the basic kernel method to estimate x, which may bring more bias then Local polynomial estimator.

The error to minimize

To avoid this problem we suggest modifying expression so that the quantity inside the norm in the integrand can be estimated root n consistently. We consider to minimize the expression

$$\int \|\mathbf{T}_t\{f(\hat{x}, \boldsymbol{\beta}) - \hat{x}'\}\|^2 w(t)\mathrm{d}t. \tag{18}$$

We want to choose $\mathbf{T}_t$ that:

- $\mathbf{T}_t$ does not depend on initial value ξ or derivative $x^{(j)}$, but depends on x and $f(\hat{x}, \boldsymbol{\beta})$.
- $\mathbf{T}_t \equiv 0 \Leftrightarrow x$ and $f(\hat{x}, \boldsymbol{\beta})$ satisfy equation (1).

An idea is take $\mathbf{T}_t$ as in Integral Matching, and the paper gives another idea.

Differential equations of degree 1

Under this situation, we can take

$$\mathbf{T}_t(x) = \int_c^t \left[\phi'(u,t,c)x(u) + \phi(u,t,c)f\{x(u),\boldsymbol{\beta}\}\right] \mathrm{d}u,$$

for any constant $c \in (0,1)$ and for a function ϕ satisfying

$$\phi(t,t,c) = \phi(c,t,c) = 0, \qquad \text{for all } c,t \in (0,1).$$

Then expression (18) has the form

$$S(\boldsymbol{\beta},h) = \int \left\| \int_c^t \left[\phi'(u,t,c)\hat{x}(u,h) + \phi(u,t,c)f\{\hat{x}(u,h),\boldsymbol{\beta}\}\right] \mathrm{d}u \right\|^2 w(t)\mathrm{d}t, \tag{19}$$

where we can take $\phi(u,t,c) = (t-u)(u-c)$. The notation $\hat{x}(u,h)$ reflects the fact that the estimator $\hat{x}$ depends on a bandwidth h. For a given h we choose $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_h$ to minimize $S(\boldsymbol{\beta},h)$.

Kernel estimator of x

We consider $\hat{x}(t,h) = \Sigma_{1\leqslant i\leqslant n}\omega_i(t,h)\mathbf{Y}_i$ as the kernel estimator of x with a bandwidth h, where $\mathbf{Y}_i$ is the observation that (2) and $\omega_i(t,h)$ can be as follows

$$\omega_i(t,h) = \int_{s_{i-1}}^{s_i} K_h(t-t)\mathrm{d}t,$$

$$\omega_i(t,h) = \frac{K_h\left(t_i - t\right)}{\sum_{1\leqslant i\leqslant n} K_h\left(t_i - t\right)},$$

in the case of Gasser-Müller or Nadaraya-Watson estimators respectively, and

$$\omega_i(t) = \frac{K_h\left(t_i-t\right)\left\{\sum_j K_h\left(t_j-t\right)\left(t_j-t\right)^2 - \left(t_i-t\right)\sum_j K_h\left(t_j-t\right)\left(t_j-t\right)\right\}}{\sum_k K_h\left(t_k-t\right)\left\{\sum_j K_h\left(t_j-t\right)\left(t_j-t\right)^2 - \left(t_k-t\right)\sum_j K_h\left(t_j-t\right)\left(t_j-t\right)\right\}}$$

if $\hat{x}$ is a local linear estimator. Here, K is a kernel function, $K_h(u) = h^{-1}K(u/h)$ and $s_i = \left(t_i + t_{i+1}\right)/2$ for $i = 1, \ldots, n-1$, $s_0 = 0$ and $s_n = 1$.

Further generalizations

Assume that $0 < c < \frac{1}{2}$, let $\psi_1, \psi_2, \ldots$ be a sequence of functions supported on $[c, 1-c]$ and satisfying $\psi_j(c) = \psi_j(1-c) = 0$ for each j, and note that if model (6) holds then

$$\begin{aligned}\mathbf{0} &= \int_c^{1-c} \psi_j(t) \left[x'(t) - f\{x(t), \boldsymbol{\beta}\}\right] \mathrm{d}t \\ &= -\int_c^{1-c} \left[\psi_j'(t)x(t) + \psi_j(t)f\{x(t), \boldsymbol{\beta}\}\right] \mathrm{d}t\end{aligned}$$

for each j.

Further generalizations

This motivates choosing $(\hat{\boldsymbol{\beta}}, h)$ to minimize

$$S_1(\boldsymbol{\beta}, h) = \sum_j w_j \left\| \int_c^{1-c} \left[\psi_j'(t)\hat{x}(t,h) + \psi_j(t) f\{\hat{x}(t,h), \boldsymbol{\beta}\} \right] \mathrm{d}t \right\|^2, \tag{20}$$

where $w_1, w_2, \ldots$ are non-negative weights.
Similar to $S(\boldsymbol{\beta}, h)$ (19) and $S_1(\boldsymbol{\beta}, h)$ (20), consider under the norm $\|\cdot\|_W$ (7), we can get

$$\begin{aligned} S_2(\boldsymbol{\beta}, h) = \sum_{j=1}^{J} \sum_{k=1}^{J} w_{jk} & \left(\int_c^{1-c} \left[\psi_j'(t)\hat{x}(t,h) + \psi_j(t) f\{\hat{x}(t,h), \boldsymbol{\beta}\} \right] \mathrm{d}t \right)^{\mathrm{T}} \\ & \times \left(\int_c^{1-c} \left[\psi_k'(t)\hat{x}(t,h) + \psi_k(t) f\{\hat{x}(t,h), \boldsymbol{\beta}\} \right] \mathrm{d}t \right), \end{aligned} \tag{21}$$

Further generalizations

and

$$S_3(\boldsymbol{\beta}, h) = \sum_{j=1}^{J} \sum_{k=1}^{J} \int \left(\int_c^t \left[\phi_j'(u,t,c)\hat{x}(t,h) + \phi_j(u,t,c) f\{\hat{x}(t,h), \boldsymbol{\beta}\} \right] \mathrm{d}u \right)^{\mathrm{T}} \times \left(\int_c^t \left[\phi_k'(u,t,c)\hat{x}(u,h) + \phi_k(u,t,c) f\{\hat{x}(u,h), \boldsymbol{\beta}\} \right] \mathrm{d}u \right) w_{jk}(t) \mathrm{d}t. \tag{22}$$

Here are some remarks:

- $S_1 = 0$ or $S_2 = 0$ can only guarantee that (1) holds while $t \in [c, 1-c]$. The part $t \in (0, c) \cup (1-c, 1)$ is ignored.
- S, S_1 is special cases of S_3, S_2 that setting $w_{jk} = w_j \mathbb{I}_{\{j=k\}}$.

Simulated examples

The paper Hall and Ma (2014) conducted four Simulated examples:

1. ODE: $x'(t) = \beta x(t)$. Generated n observations from the model $Y = \exp(\beta t) + \varepsilon$.
2. ODE: $x''(t) = -\beta^2 x(t)$, which can be written as $x_1'(t) = \beta x_2(t)$ and $x_2'(t) = -\beta x_1(t)$. Generated data from the model $Y_1 = \cos(\beta t) + \varepsilon_1$ and $Y_2 = -\sin(\beta t) + \varepsilon_2$.
3. ODE: $x'(t) = \beta_1 x(t) + \beta_2$. Generated data from the model $Y = \beta_1^{-1} \exp(\beta_1 t + \beta_2) + \varepsilon$.
4. ODEs: $x_1'(t) = \beta_1 \beta_2 x_2(t)$ and $x_2'(t) = -\beta_2^{-1} \beta_1 x_1(t)$. Generated data from the models $Y_1 = \cos(\beta_1 t) + \varepsilon_1$ and $Y_2 = \beta_2^{-1} \sin(\beta_1 t) + \varepsilon_2$.

Simulated examples

Table 1. Simulation 1: average values of estimators of β, $\bar{\hat{\beta}}$, standard deviations of estimators of β, $\mathrm{sd}(\hat{\beta})$, and average values of estimators of standard deviations of β, $\widehat{\mathrm{sd}}(\hat{\beta})$, for $n = 250, 500, 1000$ and $\sigma = 0.1, 0.2, 0.3$†

	Values for n = 250			*Values for n = 500*			*Values for n = 1000*		
	$\sigma=0.1$	$\sigma=0.2$	$\sigma=0.3$	$\sigma=0.1$	$\sigma=0.2$	$\sigma=0.3$	$\sigma=0.1$	$\sigma=0.2$	$\sigma=0.3$
$\bar{\hat{\beta}}$	0.9999	0.9998	1.0001	0.9996	0.9994	0.9989	1.0001	1.0002	1.0004
$\mathrm{sd}(\hat{\beta})$	0.0259	0.0487	0.0717	0.0166	0.0321	0.0478	0.0108	0.0213	0.0320
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.0283	0.0566	0.0849	0.0176	0.0351	0.0526	0.0113	0.0226	0.0338

†The true value of β was 1.

Table 2. Simulation 2: average values of estimators of β, $\bar{\hat{\beta}}$, standard deviations of estimators of β, $\mathrm{sd}(\hat{\beta})$, and average values of estimators of standard deviations of β, $\widehat{\mathrm{sd}}(\hat{\beta})$, for $n = 250, 500, 1000, 1500$ and $\sigma = 0.1, 0.2, 0.3$†

	Values for n = 250			*Values for n = 500*			*Values for n = 1000*			*Values for n = 1500*		
	$\sigma=0.1$	$\sigma=0.2$	$\sigma=0.3$	$\sigma=0.1$	$\sigma=0.2$	$\sigma=0.3$	$\sigma=0.1$	$\sigma=0.2$	$\sigma=0.3$	$\sigma=0.1$	$\sigma=0.2$	$\sigma=0.3$
$\bar{\hat{\beta}}$	1.0002	0.9989	0.9993	0.9995	0.9987	0.9991	0.9997	0.9994	0.9995	1.0001	1.0002	0.9999
$\mathrm{sd}(\hat{\beta})$	0.0405	0.0781	0.1172	0.0264	0.0527	0.0786	0.0168	0.0336	0.0501	0.0134	0.0269	0.0405
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.0460	0.0921	0.1381	0.0285	0.0570	0.0855	0.0183	0.0366	0.0549	0.0143	0.0286	0.0429

†The true value of β was 1.

Simulated examples

Table 3. Simulation 3: average values of estimators of β, $\bar{\hat{\beta}}$, standard deviations of estimators of β, $\mathrm{sd}(\hat{\beta})$, and average values of estimators of standard deviations of β, $\widehat{\mathrm{sd}}(\hat{\beta})$, for $n=250, 500, 1000, 1500, 2000$ and $\sigma=0.1, 0.2, 0.3$†

	Values for $\sigma=0.1$		*Values for $\sigma=0.2$*		*Values for $\sigma=0.3$*	
	β_1	β_2	β_1	β_2	β_1	β_2
$n=250$						
$\bar{\hat{\beta}}$	0.9958	1.0031	0.9926	1.0069	0.9918	1.0087
$\mathrm{sd}(\hat{\beta})$	0.2675	0.1731	0.3781	0.2539	0.4520	0.3114
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.3187	0.2078	0.6383	0.4170	0.9611	0.6285
$n=500$						
$\bar{\hat{\beta}}$	1.0055	0.9968	1.0081	0.9959	1.0061	0.9977
$\mathrm{sd}(\hat{\beta})$	0.1610	0.1042	0.2556	0.1684	0.3169	0.2135
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.1794	0.1174	0.3590	0.2352	0.5395	0.3536
$n=1000$						
$\bar{\hat{\beta}}$	1.0051	0.9971	1.0120	0.9938	1.0147	0.9928
$\mathrm{sd}(\hat{\beta})$	0.0954	0.0622	0.1834	0.1206	0.2484	0.1654
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.1073	0.0705	0.2147	0.1411	0.3224	0.2118
$n=1500$						
$\bar{\hat{\beta}}$	1.0010	0.9995	1.0033	0.9985	1.0062	0.9973
$\mathrm{sd}(\hat{\beta})$	0.0756	0.0491	0.1481	0.0967	0.2103	0.1375
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.0811	0.0534	0.1622	0.1068	0.2435	0.1604
$n=2000$						
$\bar{\hat{\beta}}$	1.0026	0.9983	1.0054	0.9965	1.0063	0.9961
$\mathrm{sd}(\hat{\beta})$	0.0618	0.0408	0.1221	0.0806	0.1785	0.1180
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.0670	0.0442	0.1340	0.0884	0.2011	0.1326

†The true value of β was $(1, 1)^{\mathrm{T}}$.

Table 4. Simulation 4: average values of estimators of β, $\bar{\hat{\beta}}$, standard deviations of estimators of β, $\mathrm{sd}(\hat{\beta})$, and average values of estimators of standard deviations of β, $\widehat{\mathrm{sd}}(\hat{\beta})$, for $n=250, 500, 1000, 1500, 2000$ and $\sigma=0.1, 0.2, 0.3$†

	Values for $\sigma=0.1$		*Values for $\sigma=0.2$*		*Values for $\sigma=0.3$*	
	β_1	β_2	β_1	β_2	β_1	β_2
$n=250$						
$\bar{\hat{\beta}}$	1.0004	1.0019	0.9981	1.0030	0.9943	1.0055
$\mathrm{sd}(\hat{\beta})$	0.0496	0.0501	0.0982	0.1011	0.1454	0.1496
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.0578	0.0576	0.1168	0.1169	0.1778	0.1794
$n=500$						
$\bar{\hat{\beta}}$	1.0013	1.0009	1.0021	1.0020	1.0018	1.0031
$\mathrm{sd}(\hat{\beta})$	0.0324	0.0320	0.0638	0.0639	0.0954	0.0965
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.0360	0.0357	0.0721	0.0717	0.1088	0.1084
$n=1000$						
$\bar{\hat{\beta}}$	1.0001	1.0002	0.9995	0.9998	0.9994	1.0003
$\mathrm{sd}(\hat{\beta})$	0.0218	0.0216	0.0437	0.0432	0.0656	0.0652
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.0232	0.0230	0.0465	0.0462	0.0699	0.0695
$n=1500$						
$\bar{\hat{\beta}}$	0.9996	1.0005	0.9986	1.0006	0.9977	1.0011
$\mathrm{sd}(\hat{\beta})$	0.0164	0.0176	0.0328	0.0349	0.0491	0.0526
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.0181	0.0180	0.0363	0.0361	0.0545	0.0543
$n=2000$						
$\bar{\hat{\beta}}$	1.0002	1.0004	1.0001	1.0006	1.0002	1.0009
$\mathrm{sd}(\hat{\beta})$	0.0149	0.0149	0.0297	0.0300	0.0447	0.0451
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.0153	0.0152	0.0306	0.0304	0.0460	0.0457

†The true value of β was $(1, 1)^{\mathrm{T}}$.

Example: FitzHugh–Nagumo model

Next, we consider the FitzHugh–Nagumo model (5)**.**

The method in Hall and Ma (2014) compares with other methods:

- The author applied their estimator in the setting of Ramsay et al. (2007) and found that their estimation standard errors for the parameters (0.0390, 0.1573 and 0.5522) which larger then the one in Ramsay et al. (2007) (0.0149, 0.0643 and 0.0264).
  Since the competing method is substantially more difficult to implement than ours and uses a smoothing parameter that requires significantly more skill to compute.
- They also compared our approach with the method of Liang and Wu (2008), in the same setting as theirs, and obtained the estimation standard deviations 0.0541, 0.1446 and 0.4168. In comparison with the corresponding results of 0.08, 0.12 and 0.17 in Liang and Wu (2008).

Example: FitzHugh–Nagumo model

Table 5. Simulation 5: average values of estimators of β, $\bar{\hat{\beta}}$, standard deviations of estimators of β, $\mathrm{sd}(\hat{\beta})$, and average values of estimators of standard deviations of β, $\widehat{\mathrm{sd}}(\hat{\beta})$, for $n=250, 500, 1000, 1500, 2000$ and $\sigma=0.1, 0.2, 0.3$†

	Values for $\sigma=0.1$			*Values for $\sigma=0.2$*			*Values for $\sigma=0.3$*		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
$n=250$									
$\bar{\hat{\beta}}$	4.9363	0.9851	0.4949	4.9222	0.9772	0.4903	4.7183	0.9392	0.4868
$\mathrm{sd}(\hat{\beta})$	2.2596	0.5110	0.0218	3.3960	0.7688	0.0283	3.8731	0.8806	0.0323
$\widehat{\mathrm{sd}}(\hat{\beta})$	3.9317	0.8933	0.0116	7.8449	1.7930	0.0233	11.7125	2.6881	0.0351
$n=500$									
$\bar{\hat{\beta}}$	5.0006	1.0034	0.4984	4.9666	0.9946	0.4953	4.9124	0.9816	0.4928
$\mathrm{sd}(\hat{\beta})$	0.9943	0.2258	0.0076	1.6072	0.3665	0.0125	1.9398	0.4437	0.0163
$\widehat{\mathrm{sd}}(\hat{\beta})$	1.1844	0.2687	0.0059	2.2795	0.5215	0.0119	3.4008	0.7796	0.0179
$n=1000$									
$\bar{\hat{\beta}}$	5.0050	1.0020	0.4991	4.9993	1.0006	0.4970	4.9886	0.9971	0.4956
$\mathrm{sd}(\hat{\beta})$	0.4435	0.1016	0.0038	0.7755	0.1785	0.0071	1.0266	0.2371	0.0097
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.4322	0.0991	0.0035	0.8698	0.1998	0.0070	1.3179	0.3033	0.0105
$n=1500$									
$\bar{\hat{\beta}}$	5.0005	1.0006	0.4991	4.9975	0.9992	0.4977	4.9974	0.9981	0.4963
$\mathrm{sd}(\hat{\beta})$	0.2709	0.0624	0.0028	0.5033	0.1164	0.0052	0.7101	0.1646	0.0071
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.2726	0.0627	0.0026	0.5549	0.1279	0.0053	0.8477	0.1955	0.0080
$n=2000$									
$\bar{\hat{\beta}}$	4.9868	0.9962	0.4993	4.9485	0.9887	0.4981	4.9284	0.9837	0.4971
$\mathrm{sd}(\hat{\beta})$	0.2104	0.0487	0.0022	0.3916	0.0909	0.0042	0.5699	0.1326	0.0058
$\widehat{\mathrm{sd}}(\hat{\beta})$	0.2084	0.0480	0.0022	0.4254	0.0982	0.0044	0.6471	0.1496	0.0066

†The true value of β was $(5, 1, 0.5)^{\mathrm{T}}$.

Example: Lotka-Volterra model

Then, we consider the Lotka-Volterra model (6).

The estimated parameter values in Hall and Ma (2014) are $\hat{\beta} = (4.5273, 1.4418, 4.0914, 0.4172)^{\mathrm{T}}$, with associated standard errors $(0.4452, 0.3365, 3.5567, 3.0131)^{\mathrm{T}}$.

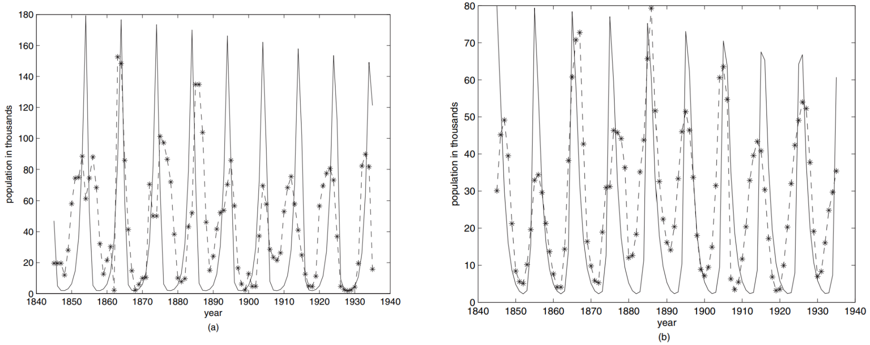

Figure 3: Estimated (solid) and observed (dotted) fluctuation of (a) the hare and (b) the lynx populations.

Numerical discretization-based estimator

In this section we consider a kind of numerical discretization-based estimator in Wu et al. (2012).

It is an estimator within two-steps:

1. Smooth x by spline.
2. Minimize a numerical discretization error between estimator of $\mathbf{x}'(t)$ and $f(\mathbf{x}(t);\boldsymbol{\beta})$.

- **First we use B-spline to smooth x and get $\hat{x}$.**

We approximate $x(t)$ at time t by $x(t) \approx \sum_{j=-\nu}^{K} \delta_j \, N_{j,\nu+1}(t) = \boldsymbol{N}_{\nu+1}^T(t)\boldsymbol{\delta}$, where $\boldsymbol{\delta} = (\delta_{-\nu}, \dots, \delta_K)^T$ is the unknown coefficient vector to be estimated from the data, and $\boldsymbol{N}_{\nu+1}(t) = \{N_{-\nu,\nu+1}(t), \dots, N_{K,\nu+1}(t)\}^T$ is the B-spline basis function vector of degree ν (order $\nu + 1$).

Numerical discretization-based estimator

We get the estimator of δ by minimize

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

Then we can get

$$\hat{\boldsymbol{\delta}} = \left(\boldsymbol{Z}^T \boldsymbol{Z} + \lambda \boldsymbol{D}_q \right)^{-1} \boldsymbol{Z}^T \boldsymbol{Y},$$

and the estimator of x as $\hat{x} = \boldsymbol{N}_{\nu+1}^T(t) \hat{\boldsymbol{\delta}}$.

See chapter 8.2 in James Ramsay (2017) for more details.

Numerical discretization-based estimator

- **Next we Minimize a numerical discretization error.**

we want to get a suitable $F\left(t_i, \hat{X}(t_i), \hat{X}(t_{i+1}), \beta_0\right)$ and minimize the error between $\frac{\hat{X}(t_{i+1})-\hat{X}(t_i)}{t_{i+1}-t_i}$ and $F\left(t_i, \hat{X}(t_i), \hat{X}(t_{i+1}), \beta_0\right)$.

To get F, we consider three different discretization methods:

- For Euler's method, we have

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

with $p = 1$.

Numerical discretization-based estimator

- The trapezoidal rule gives the form

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

with $p = 2$,

- For the fourth-order Runge-Kutta method, we have

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

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

Numerical discretization-based estimator

Then, we propose a numerical discretization-based estimator $\hat{\beta}_n$ of β by minimizing the following weighted least squares criterion:

$$S_n(\boldsymbol{\beta}) = \sum_{i=1}^{n-1} w(t_i) \left[\frac{\hat{X}(t_{i+1}) - \hat{X}(t_i)}{t_{i+1} - t_i} - F\left(t_i, \hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta}\right) \right]^2, \tag{23}$$

and while $m \geq 2$, we use the following objective function:

$$S_n(\boldsymbol{\beta}) = \sum_{i=1}^{n-1} \sum_{j=1}^{m} w_j(t_i) \left[\frac{\hat{X}_j(t_{i+1}) - \hat{X}_j(t_i)}{t_{i+1} - t_i} - F_j\left(t_i, \hat{X}_j(t_i), \hat{X}_j(t_{i+1}), \boldsymbol{\beta}\right) \right]^2. \tag{24}$$

Simulated examples

The paper Wu et al. (2012) conducted two Simulated examples:

1. ODE:
$$V'(t) = \alpha_0 - \alpha_1 V(t).$$
Use parameter values $(\alpha_0, \alpha_1) = (3, 1/3)$ and $V(0) = -1$ to generate the data.
2. ODEs:
$$\begin{cases} \frac{dR}{dt} = \frac{a}{1+\exp(-P)} - bR, \\ \frac{dP}{dt} = 2R - cP. \end{cases}$$
Use parameter values $(a, b, c) = (1.5, 1, 2)$ and initial values $(R(0), P(0)) = (0, 1)$ to generate the observations.

Then we compare the method in Liang and Wu (2008) (LW), and the method in this paper with Euler's method, trapezoidal rule and Runge–Kutta method (EDB, TDB and RDB).

Simulated examples

Table 2

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

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

Simulated examples

Table 3

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

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

Where ARE is defined as (15).

References I

Itai Dattner and Chris A. J. Klaassen. Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters. *Electronic Journal of Statistics*, 9(2):1939 – 1973, 2015. doi: 10.1214/15-EJS1053. URL `https://doi.org/10.1214/15-EJS1053`.

Peter Hall and Yanyuan Ma. Quick and easy one-step parameter estimation in differential equations. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 76(4):735–748, September 2014. ISSN 1369-7412. doi: 10.1111/rssb.12040. Copyright: Copyright 2014 Elsevier B.V., All rights reserved.

Giles Hooker James Ramsay. Dynamic Data Analysis. 2017. doi: 10.1007/978-1-4939-7190-9.

Hua Liang and Hulin Wu. Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of the American Statistical Association*, 103:1570 – 1583, 2008.

References II

J. O. Ramsay, G. Hooker, D. Campbell, and J. Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5):741–796, 2007. doi: https://doi.org/10.1111/j.1467-9868.2007.00610.x.

Ivan Vujačić and Itai Dattner. Consistency of direct integral estimator for partially observed systems of ordinary differential equations. *Statistics and Probability Letters*, 132:40–45, 2018. ISSN 0167-7152. doi: https://doi.org/10.1016/j.spl.2017.08.013.

Hulin Wu, Hongqi Xue, and Arun Kumar. Numerical discretization-based estimation methods for ordinary differential equation models via penalized spline smoothing with applications in biomedical research. *Biometrics*, 68(2):344–352, 2012. ISSN 0006341X, 15410420. URL `http://www.jstor.org/stable/23270435`.