BIOMETRICS **, 1–??
DOI: 10.1111/j.1541-0420.2005.00454.x
September 2011

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

Hulin Wu, Hongqi Xue, and Arun Kumar

July 27, 2011.

Appendix A: Technical Proofs

Some assumptions are stated as follows:

(B1) For $\kappa = \max_{0 \leqslant j \leqslant K}(\tau_{j+1} - \tau_j)$, there exists a constant $M > 0$, such that $\kappa / \min_{0 \leqslant j \leqslant K}(\tau_{j+1} - \tau_j) \leqslant M$ and $\max_{0 \leqslant j \leqslant K} |\frac{\tau_{j+1}-\tau_j}{\tau_j - \tau_{j-1}} - 1| = o(1)$.

(B2) For fixed design points, $t_i \in [a, b]$ for $i = 1, \cdots, n$, assume that there exists a distribution function $Q(t)$ with a corresponding positive continuous design density $\rho(t)$ such that, for the empirical distribution of $(t_1, \ldots, t_n)$, $Q_n(t)$, we have $\sup_{t \in [a,b]} |Q_n(t) - Q(t)| = o(K^{-1})$. Moreover, there exist two constants $0 < c_1 < c_2 < \infty$ such that $c_1 \leqslant \rho(t) \leqslant c_2$ for all $t \in [a, b]$.

(B3) The number of knots $K = o(n)$.

(B4) $f(t, x, \boldsymbol{\beta})$ is a continuous function of $\boldsymbol{\beta}$ for $\boldsymbol{\beta} \in \Omega_{\boldsymbol{\beta}}$, where $\Omega_{\boldsymbol{\beta}}$ is a compact subset of R^d.

(B5) For ρ defined in Assumption B2, $\rho(t) \in C[a, b]$.

(B6) $X(t) \in \chi$, where $\chi \subset C^{\nu+1}[a, b]$ with $\nu \geqslant 2$.

(B7) The weight $w(.)$ is a bounded and non-negative function on the interval $[a, b]$. Further, $w(.)$ has a bounded first-order derivative.

(B8) $\mathrm{E}\{w(t)[f(t, X(t), \boldsymbol{\beta}) - f(t, X(t), \boldsymbol{\beta}_0)]^2\} = 0$ if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, where $E[g(t)]$ is the expectation of function $g(t)$ with respect to t in the case of random design and the integral $\int_a^b g(t) dQ(t)$ for function $g(t)$ in the case of fixed design.

(B9) The first and second partial derivatives, $\frac{\partial f(t,x,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$, $\frac{\partial^2 f(t,x,\boldsymbol{\beta})}{\partial x \partial \boldsymbol{\beta}}$, and $\frac{\partial^2 f(t,x,\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}$, exist, are continuous and uniformly bounded for all $t \in [a, b]$, $\boldsymbol{\beta} \in \Omega_{\boldsymbol{\beta}}$, $x \in \chi$, and

$$\Big|\frac{\partial}{\partial \boldsymbol{\beta}} f(t, x_1, \boldsymbol{\beta}) - \frac{\partial}{\partial \boldsymbol{\beta}} f(t, x_2, \boldsymbol{\beta})\Big| \leqslant C|x_1 - x_2|^{\zeta}$$

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

(B10) The first partial derivatives $\frac{\partial}{\partial t} f(t, x, \boldsymbol{\beta})$ and $\frac{\partial}{\partial x} f(t, x, \boldsymbol{\beta})$ are continuous and uniformly bounded for $t \in [a, b]$, $\boldsymbol{\beta} \in \Omega_{\boldsymbol{\beta}}$ and $x \in \chi$.

(B11) $K_q < 1$ with $K_q = (K + \nu + 1 - q)(\lambda \tilde{c}_1)^{1/(2q)} n^{-1/(2q)}$ for some constant $\tilde{c}_1$.

(B12) $K_q \geqslant 1$ for K_q defined in (B11).

(B13) All partial derivatives of $f\{t, \boldsymbol{X}, \boldsymbol{\Upsilon}, \boldsymbol{\beta}\}$ up to order p with respect to t, $\boldsymbol{X}$ and $\boldsymbol{\Upsilon}$ exist and are continuous. The derivatives of input variables $\boldsymbol{\Upsilon}(t)$ up to pth-order with respect to t exist and are continuous.

Some notations are required. Let $\|\boldsymbol{a}\|$ be the Euclidean norm of a vector $\boldsymbol{a}$; $\|\boldsymbol{A}\|_\infty = \max_{1 \leqslant i \leqslant m} \sum_{j=1}^{n} |a_{ij}|$ be the supremum norm of an $m \times n$ matrix $\boldsymbol{A}$, where a_{ij} is the (i, j)-th element of $\boldsymbol{A}$; $\boldsymbol{A}^{\otimes 2} = \boldsymbol{A}\boldsymbol{A}^T$ for a matrix $\boldsymbol{A}$; $C^r[a, b]$ be the class of functions with r-order continuous derivatives on the interval $[a, b]$; $W^q[a, b]$ be the Sobolev space of order q, i.e. $W^q[a, b] = \{f : f$ has $q - 1$ absolutely continuous dirivatives, $\int_a^b \{f^{(q)}\}^2 dx < \infty\}$; $\|f\|_\infty = \sup_t |f(t)|$ be the supremum norm of a function f. Denote $\boldsymbol{G}_{K,n} = \boldsymbol{Z}^T \boldsymbol{Z}/n$, $\boldsymbol{H}_{K,n} = \boldsymbol{G}_{K,n} + \lambda \boldsymbol{D}_q/n$ and $\boldsymbol{H} = \boldsymbol{G} + \lambda \boldsymbol{D}_q/n$ with $\boldsymbol{G} = \int_a^b \boldsymbol{N}_{\nu+1}(t) \boldsymbol{N}_{\nu+1}^T(t) \rho(t) dt$. Denote c as a general constant.

Lemma 1: (i) Under Assumptions B1-B3 and B11, if $X(.) \in C^{\nu+1}[a, b]$ with $\nu \geqslant 1$, then for any $i = 0, 1, \cdots, \nu - 1$,

$$
\begin{aligned}
E[\hat{X}^{(i)}(t)] - X^{(i)}(t) &= b_i(t, \nu + 1) + d_i(t) + o(\kappa^{\nu+1-i}) + o(\lambda n^{-1} \kappa^{-q-i}) \\
&= O(\kappa^{\nu+1-i}) + O(\lambda n^{-1} \kappa^{-q-i})
\end{aligned}
$$

and

$$
\begin{aligned}
\text{Var}[\hat{X}^{(i)}(t)] &= \frac{\sigma^2}{n} \boldsymbol{N}_{\nu+1-i}^T(t) \boldsymbol{\Delta}_i (\boldsymbol{G} + \frac{\lambda}{n} \boldsymbol{D}_q)^{-1} \boldsymbol{G} (\boldsymbol{G} + \frac{\lambda}{n} \boldsymbol{D}_q)^{-1} \boldsymbol{\Delta}_i^T \boldsymbol{N}_{\nu+1-i}(t) \\
&\quad + o(n^{-1} \kappa^{-2i-1}) \\
&= O(n^{-1} \kappa^{-2i-1}).
\end{aligned}
$$

(ii) Under Assumptions B1-B3 and B12, if $X(.) \in W^q[a,b]$ with $q \geqslant 2$, then for any $i = 0, 1, \cdots, q-2$,

$$\begin{aligned} E[\hat{X}^{(i)}(t)] - X^{(i)}(t) &= b_i(t,q) + d_i(t) + o(\kappa^{q-i}) + o((\lambda/n)^{1/2}\kappa^{-i}) \\ &= O(\kappa^{q-i}) + O((\lambda/n)^{1/2}\kappa^{-i}) \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}[\hat{X}^{(i)}(t)] &= \frac{\sigma^2}{n}\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{G}(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{\Delta}_i^T\boldsymbol{N}_{\nu+1-i}(t) \\ &\quad + o(n^{-1}\kappa^{-2i}(\lambda/n)^{-1/(2q)}) \\ &= O(n^{-1}\kappa^{-2i}(\lambda/n)^{-1/(2q)}), \end{aligned}$$

where

$$b_i(t,\nu+1) = -\frac{X^{(\nu+1)}(t)}{(\nu+1-i)!}\sum_{j=0}^{K} I_{[\tau_j,\tau_{j+1}]}(t)(\tau_{j+1}-\tau_j)^{\nu+1-i}B_{\nu+1-i}(\frac{t-\tau_j}{\tau_{j+1}-\tau_j}),$$

and

$$d_i(t) = -\frac{\lambda}{n}\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i(\boldsymbol{G} + \lambda\boldsymbol{D}_q/n)^{-1}\boldsymbol{D}_q\boldsymbol{\xi},$$

with $B_j(.)$ denoting the jth Bernoulli polynomial (see Ghizzetti and Ossicini, 1970), and $\boldsymbol{\xi} = (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{S}_X$ for $\boldsymbol{S}_X = \{S_X(t_1), \cdots, S_X(t_n)\}^T$, where $S_X(t) = \boldsymbol{N}^T_{\nu+1}(t)\boldsymbol{\delta} \in S(\nu+1, \underline{\tau})$ is the best L_∞ approximation to the true value of function $X(t)$.

Proof of Lemma 1: From the expressions (8) and (9) , we have

$$\hat{X}^{(i)}(t) = \boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i(\boldsymbol{Z}^T\boldsymbol{Z} + \lambda\boldsymbol{D}_q)^{-1}\boldsymbol{Z}^T\boldsymbol{Y} = \boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}_{K,n}\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{Y}.$$

From the definition of $\boldsymbol{G}_{K,n}$ and $\boldsymbol{H}_{K,n}$, we have $\boldsymbol{H}^{-1}_{K,n} - \boldsymbol{G}^{-1}_{K,n} = -\frac{\lambda}{n}\boldsymbol{H}^{-1}_{K,n}\boldsymbol{D}_q\boldsymbol{G}^{-1}_{K,n}$. It follows that

$$\begin{aligned} \hat{X}^{(i)}(t) &= \boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{G}^{-1}_{K,n}\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{Y} + \boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i(\boldsymbol{H}^{-1}_{K,n} - \boldsymbol{G}^{-1}_{K,n})\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{Y} \\ &= \hat{X}^{(i)}_{\mathrm{reg}}(t) - \frac{\lambda}{n}\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}_{K,n}\boldsymbol{D}_q\boldsymbol{G}^{-1}_{K,n}\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{Y} \end{aligned}$$

with $\hat{X}^{(i)}_{\text{reg}}(t) = \boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{G}^{-1}_{K,n}\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{Y}$ from the expression (5) in Zhou and Wolfe (2000). Then

$$\begin{aligned} & E[\hat{X}^{(i)}(t)] - X^{(i)}(t) \\ = \ & [S^{(i)}_X(t) - X^{(i)}(t)] + E[\hat{X}^{(i)}_{\text{reg}}(t) - S^{(i)}_X(t)] \qquad \text{(A.1)} \\ - \ & \frac{\lambda}{n}\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}_{K,n}\boldsymbol{D}_q\boldsymbol{G}^{-1}_{K,n}\frac{1}{n}\boldsymbol{Z}^T(\boldsymbol{X} - \boldsymbol{S}_X + \boldsymbol{S}_X), \end{aligned}$$

where $\boldsymbol{X} = \{X(t_1), \cdots, X(t_n)\}^T$ and $S^{(i)}_X(t)$ is the ith-order derivative of $S_X(t)$ with $S_X(t)$ and $\boldsymbol{S}_X$ defined in Lemma 1.

From Lemma 5.1 and Theorem 3.1 of Zhou and Wolfe(2000), it holds that $X^{(i)}(t) - S^{(i)}_X(t) = -b_i(t, \nu+1) + o(\kappa^{\nu+1-i})$ and $E[\hat{X}^{(i)}_{\text{reg}}(t) - S^{(i)}_X(t)] = o(\kappa^{\nu+1-i})$ for $X(.) \in C^{\nu+1}[a,b]$, and $X^{(i)}(t) - S^{(i)}_X(t) = -b_i(t,q) + o(\kappa^{q-i})$ and $E[\hat{X}^{(i)}_{\text{reg}}(t) - S^{(i)}_X(t)] = o(\kappa^{q-i})$ for $X(.) \in W^q[a,b]$.

Now we consider the third component of (A.1). In the first step, we use similar arguments to those in the proof of Theorem 2 in Claeskens, Krivobokova and Opsomer(2009) and the definition of $\boldsymbol{\xi} = (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{S}_X$ in Proposition 1, then we have

$$\begin{aligned} & -\frac{\lambda}{n}\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}_{K,n}\boldsymbol{D}_q\boldsymbol{G}^{-1}_{K,n}\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{S}_X \\ = \ & -\frac{\lambda}{n}\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}_{K,n}\boldsymbol{D}_q\boldsymbol{\xi} \\ = \ & -\frac{\lambda}{n}\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}_{K,n}\boldsymbol{\Delta}^T_q\boldsymbol{W}\boldsymbol{S}^{(q)}_X(\boldsymbol{\varrho}) \\ = \ & -\frac{\lambda}{n}\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}\boldsymbol{\Delta}^T_q\boldsymbol{W}\boldsymbol{S}^{(q)}_X(\boldsymbol{\varrho}) \\ & -\frac{\lambda}{n}\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i(\boldsymbol{H}^{-1}_{K,n} - \boldsymbol{H}^{-1})\boldsymbol{\Delta}^T_q\boldsymbol{W}\boldsymbol{S}^{(q)}_X(\boldsymbol{\varrho}) \\ = \ & -\frac{\lambda}{n}\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}\boldsymbol{D}_q\boldsymbol{\xi} - \frac{\lambda}{n}\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i(\boldsymbol{H}^{-1}_{K,n} - \boldsymbol{H}^{-1})\boldsymbol{\Delta}^T_q\boldsymbol{W}\boldsymbol{S}^{(q)}_X(\boldsymbol{\varrho}) \\ = \ & d_i(t) - \frac{\lambda}{n}\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i(\boldsymbol{H}^{-1}_{K,n} - \boldsymbol{H}^{-1})\boldsymbol{\Delta}^T_q\boldsymbol{W}\boldsymbol{S}^{(q)}_X(\boldsymbol{\varrho}), \end{aligned}$$

where $\boldsymbol{W} = \text{diag}\{\sum_{l=j}^{j+\nu-q}\int_{\tau_l}^{\tau_{l+1}} N_{j,q}(t)dt\}$ and $\boldsymbol{\varrho} = (\varrho_{-\nu+q}, \cdots, \varrho_K)^T$ with some $\varrho_j \in [\tau_j, \tau_{j+\nu+1-q}]$ for $j = -\nu+q, \cdots, K$.

In the second step, we claim that $-\frac{\lambda}{n}\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i(\boldsymbol{H}^{-1}_{K,n} - \boldsymbol{H}^{-1})\boldsymbol{\Delta}^T_q\boldsymbol{W}\boldsymbol{S}^{(q)}_X(\boldsymbol{\varrho})$ is of negligible asymptotic order for both $K_q < 1$ and $K_q \geqslant 1$. In fact, from Lemma A2 and the proof of Theorem 2 in Claeskens, Krivobokova and Opsomer(2009), we have $\|\boldsymbol{H}_{K,n}^{-1} - \boldsymbol{H}^{-1}\|_\infty = o(\kappa^{-1})$ for $K_q < 1$ and $o(\kappa^{-1}(1 + K_q^{2q})^{-1})$ for $K_q \geqslant 1$, $\|W\|_\infty = O(\kappa)$, $\|\Delta_q\|_\infty = O(\kappa^{-q})$, $\|\Delta_i\|_\infty = O(\kappa^{-i})$, and $\|\boldsymbol{S}_X^{(q)}(\boldsymbol{\varrho})\|_\infty = O(1)$. Then it follows that for $K_q < 1$,

$$\frac{\lambda}{n}\boldsymbol{N}_{\nu+1-i}^T(t)\boldsymbol{\Delta}_i(\boldsymbol{H}_{K,n}^{-1} - \boldsymbol{H}^{-1})\boldsymbol{\Delta}_q^T\boldsymbol{W}\boldsymbol{S}_X^{(q)}(\boldsymbol{\varrho}) = o(\lambda n^{-1}\kappa^{-(q+i)}),$$

and for $K_q \geqslant 1$,

$$\begin{aligned} & \frac{\lambda}{n}\boldsymbol{N}_{\nu+1-i}^T(t)\boldsymbol{\Delta}_i(\boldsymbol{H}_{K,n}^{-1} - \boldsymbol{H}^{-1})\boldsymbol{\Delta}_q^T\boldsymbol{W}\boldsymbol{S}_X^{(q)}(\boldsymbol{\varrho}) \\ = \;& o(\lambda n^{-1}\kappa^{-(q+i)}(1 + K_q^{2q})^{-1}) \\ = \;& o((\lambda/n)^{1/2}\kappa^{-i}K_q^q(1 + K_q^{2q})^{-1}) \\ = \;& o((\lambda/n)^{1/2}\kappa^{-i}), \end{aligned}$$

since $K_q^q(1 + K_q^{2q})^{-1} < 1/2$ for $K_q \geqslant 1$.

In the third step, we also claim that $-\frac{\lambda}{n}\boldsymbol{N}_{\nu+1-i}^T(t)\boldsymbol{\Delta}_i\boldsymbol{H}_{K,n}^{-1}\boldsymbol{D}_q\boldsymbol{G}_{K,n}^{-1}\frac{1}{n}\boldsymbol{Z}^T(\boldsymbol{X} - \boldsymbol{S}_X)$ is of negligible asymptotic order for both $K_q < 1$ and $K_q \geqslant 1$. In fact, From Result R2 in Claeskens, Krivobokova and Opsomer(2009), it follows that $\boldsymbol{G}_{K,n}^{-1}\frac{1}{n}\boldsymbol{Z}^T(\boldsymbol{X}-\boldsymbol{S}_X) = o(\kappa^{v+1})$ for $X(.) \in C^{\nu+1}[a,b]$ and $o(\kappa^q)$ for $X(.) \in W^q[a,b]$. From Lemma A1 in Claeskens, Krivobokova and Opsomer(2009), $\|\boldsymbol{H}_{K,n}^{-1}\|_\infty = O(\kappa^{-1})$ for $K_q < 1$ and $O(\kappa^{-1}(1 + K_q^{2q})^{-1})$ for $K_q \geqslant 1$. Moreover, we have $\|\boldsymbol{D}_q\|_\infty = O(\kappa^{-2q+1})$ from Lemma 6.2 in Cardot(2000). So for $K_q < 1$, we obtain that

$$\begin{aligned} & \frac{\lambda}{n}\boldsymbol{N}_{\nu+1-i}^T(t)\boldsymbol{\Delta}_i\boldsymbol{H}_{K,n}^{-1}\boldsymbol{D}_q\boldsymbol{G}_{K,n}^{-1}\frac{1}{n}\boldsymbol{Z}^T(\boldsymbol{X} - \boldsymbol{S}_X) \\ = \;& o(\lambda n^{-1}\kappa^{\nu+1-2q-i}) \\ = \;& o(\lambda n^{-1}\kappa^{-(q+i)}\kappa^{\nu+1-q}) \\ = \;& o(\lambda n^{-1}\kappa^{-(q+i)}), \end{aligned}$$

because of $\nu \geqslant q$, and for $K_q \geqslant 1$, we obtain that

$$\frac{\lambda}{n}\boldsymbol{N}_{\nu+1-i}^T(t)\boldsymbol{\Delta}_i\boldsymbol{H}_{K,n}^{-1}\boldsymbol{D}_q\boldsymbol{G}_{K,n}^{-1}\frac{1}{n}\boldsymbol{Z}^T(\boldsymbol{X} - \boldsymbol{S}_X)$$

$$= o(\lambda n^{-1}\kappa^{-q-i}(1+K_q^{2q})^{-1})$$
$$= o((\lambda/n)^{1/2}\kappa^{-i}).$$

Thus, for $K_q < 1$,

$$\begin{aligned} E[\hat{X}^{(i)}(t)] - X^{(i)}(t) &= b_i(t,\nu+1) + d_i(t) + o(\kappa^{\nu+1-i}) + o(\lambda n^{-1}\kappa^{-q-i}) \\ &= O(\kappa^{\nu+1-i}) + O(\lambda n^{-1}\kappa^{-q-i}), \end{aligned}$$

and for $K_q \geqslant 1$,

$$\begin{aligned} E[\hat{X}^{(i)}(t)] - X^{(i)}(t) &= b_i(t,q) + d_i(t) + o(\kappa^{q-i}) + o((\lambda/n)^{1/2}\kappa^{-i}) \\ &= O(\kappa^{q-i}) + O((\lambda/n)^{1/2}\kappa^{-i}). \end{aligned}$$

Now, we consider the variance

$$\mathrm{Var}[\hat{X}^{(i)}(t)] = \frac{\sigma^2}{n}\boldsymbol{N}_{\nu+1-i}^T(t)\boldsymbol{\Delta}_i\boldsymbol{H}_{K,n}^{-1}\boldsymbol{G}_{K,n}\boldsymbol{H}_{K,n}^{-1}\boldsymbol{\Delta}_i^T\boldsymbol{N}_{\nu+1-i}(t).$$

From Result R1 and Lemma A2 in Claeskens, Krivobokova and Opsomer(2009), we have that for $K_q < 1$,

$$\begin{aligned} \mathrm{Var}[\hat{X}^{(i)}(t)] &= \frac{\sigma^2}{n}\boldsymbol{N}_{\nu+1-i}^T(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}\boldsymbol{G}\boldsymbol{H}^{-1}\boldsymbol{\Delta}_i^T\boldsymbol{N}_{\nu+1-i}(t) + o(n^{-1}\kappa^{-2i-1}) \\ &= O(n^{-1}\kappa^{-2i-1}), \end{aligned}$$

and for $K_q \geqslant 1$,

$$\begin{aligned} \mathrm{Var}[\hat{X}^{(i)}(t)] &= \frac{\sigma^2}{n}\boldsymbol{N}_{\nu+1-i}^T(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}\boldsymbol{G}\boldsymbol{H}^{-1}\boldsymbol{\Delta}_i^T\boldsymbol{N}_{\nu+1-i}(t) \\ &\quad + o(n^{-1}\kappa^{-2i-1}(1+K_q^{2q})^{-2}) \\ &= \frac{\sigma^2}{n}\boldsymbol{N}_{\nu+1-i}^T(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}\boldsymbol{G}\boldsymbol{H}^{-1}\boldsymbol{\Delta}_i^T\boldsymbol{N}_{\nu+1-i}(t) \\ &\quad + o(n^{-1}\kappa^{-2i}(\lambda/n)^{-1/(2q)}K_q(1+K_q^{2q})^{-2}) \\ &= \frac{\sigma^2}{n}\boldsymbol{N}_{\nu+1-i}^T(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}\boldsymbol{G}\boldsymbol{H}^{-1}\boldsymbol{\Delta}_i^T\boldsymbol{N}_{\nu+1-i}(t) \\ &\quad + o(n^{-1}\kappa^{-2i}(\lambda/n)^{-1/(2q)}) \\ &= O(n^{-1}\kappa^{-2i}(\lambda/n)^{-1/(2q)}), \end{aligned}$$

since $K_q(1+K_q^{2q})^{-2} \leqslant K_q^q(1+K_q^{2q})^{-1} < 1/2$ for $K_q \geqslant 1$.

Lemma 2: (i) Under Assumptions B1-B3 and B11, if $X(.) \in C^{\nu+1}[a,b]$ with $\nu \geqslant 1$, then for any $i = 0, 1, \cdots, \nu-1$, the mean squared error(MSE), $\text{MSE}[\hat{X}^{(i)}(t)] = E[\hat{X}^{(i)}(t) - X^{(i)}(t)]^2$, satisfies

$$\text{MSE}[\hat{X}^{(i)}(t)] = O(\kappa^{2(\nu+1-i)}) + O(\lambda^2 n^{-2} \kappa^{-2(q+i)}) + O(n^{-1}\kappa^{-2i-1}).$$

Moreover, for $K \sim cn^{1/(2\nu+3)}$ and $\lambda = O(n^\gamma)$ with $\gamma \leqslant (\nu+2-q)/(2\nu+3)$, the P-spline estimator of $X^{(i)}(t)$ attains the optimal rate of convergence for $X(.) \in C^{\nu+1}[a,b]$ with $\text{MSE}[\hat{X}^{(i)}(t)] = O(n^{-2(\nu+1-i)/(2\nu+3)})$.

(ii) Under Assumptions B1-B3 and B12, if $X(.) \in W^q[a,b]$ with $q \geqslant 2$, then for any $i = 0, 1, \cdots, q-2$,

$$\text{MSE}[\hat{X}^{(i)}(t)] = O(\kappa^{2(q-i)}) + O((\lambda/n)\kappa^{-2i}) + O(n^{-1}\kappa^{-2i}(\lambda/n)^{-1/(2q)}).$$

Moreover, for $\lambda \sim cn^{1/(2q+1)}$ and $K \sim cn^{\varsigma}$ with $\varsigma \geqslant 1/(2q+1)$, the P-spline estimator of $X^{(i)}(t)$ attains the optimal rate of convergence for $X(.) \in W^q[a,b]$ with $\text{MSE}[\hat{X}^{(i)}(t)] = O(n^{-2(q-i)/(2q+1)})$.

Proof of Lemma 2: The order of the mean square errors can be derived from Proposition 1 directly. Then we can verify the optimal rates of convergence by using $\kappa^{\nu+1-i} \geqslant \lambda n^{-1}\kappa^{-q-i}$ and $\kappa^{2(\nu+1-i)} = n^{-1}\kappa^{-2i-1}$ for $K_q < 1$, and $\kappa^{q-i} \leqslant (\lambda/n)^{1/2}\kappa^{-i}$ and $(\lambda/n)\kappa^{-2i} = n^{-1}\kappa^{-2i}(\lambda/n)^{-1/(2q)}$ for $K_q \geqslant 1$.

Lemma 3: (i) Under Assumptions B1-B3 and B11, if $X(.) \in C^{\nu+1}[a,b]$ with $\nu \geqslant 1$, for $K \sim cn^{\varsigma}$ with $\varsigma \geqslant 1/(2\nu+3)$, and $\lambda = O(n^\gamma)$ with $\gamma \leqslant (\nu+2-q)/(2\nu+3)$, then for any fixed $t \in [a,b]$ and any $i = 0, 1, \cdots, \nu-1$,

$$\frac{\hat{X}^{(i)}(t) - X^{(i)}(t) - b_i(t, \nu+1) - d_i(t)}{\sqrt{\text{Var}[\hat{X}^{(i)}(t)]}} \xrightarrow{d} N(0,1).$$

(ii) Under Assumptions B1-B3 and B12, if $X(.) \in W^q[a,b]$ with $q \geqslant 2$, for $\lambda \sim cn^{\gamma}$ with $\gamma \leqslant 1/(2q+1)$, and $K \sim cn^{\varsigma}$ with $\varsigma \geqslant 1/(2q+1)$, then for any fixed $t \in [a,b]$ and any $i = 0, 1, \cdots, q-2$,

$$\frac{\hat{X}^{(i)}(t) - X^{(i)}(t) - b_i(t,q) - d_i(t)}{\sqrt{\mathrm{Var}[\hat{X}^{(i)}(t)]}} \xrightarrow{d} N(0,1).$$

Proof of Lemma 3: For $K_q < 1$, if $K \sim cn^{\varsigma}$ with $\varsigma \geqslant 1/(2\nu+3)$ and $\lambda = O(n^{\gamma})$ with $\gamma \leqslant (\nu+2-q)/(2\nu+3)$, then by Lemma 1(i), for any $i = 0, 1, \cdots, \nu-1$,

$$\|E[\hat{X}^{(i)}(t)] - X^{(i)}(t) - b_i(t,\nu+1) - d_i(t)\|_{\infty} = o(n^{-\frac{\nu+1-i}{2\nu+3}}),$$
$$\sqrt{\mathrm{Var}[\hat{X}^{(i)}(t)]} = O(n^{-\frac{\nu+1-i}{2\nu+3}}).$$

It follows that

$$\frac{E[\hat{X}^{(i)}(t)] - X^{(i)}(t) - b_i(t,\nu+1) - d_i(t)}{\sqrt{\mathrm{Var}[\hat{X}^{(i)}(t)]}} = o(1).$$

Therefore it is enough to show that

$$\frac{\hat{X}^{(i)}(t) - E[\hat{X}^{(i)}(t)]}{\sqrt{\mathrm{Var}[\hat{X}^{(i)}(t)]}} \xrightarrow{d} N(0,1).$$

By (9), we have

$$\hat{X}^{(i)}(t) - E[\hat{X}^{(i)}(t)] = \boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}_{K,n}\boldsymbol{Z}^T\frac{1}{n}\boldsymbol{e} = \sum_{j=1}^{n}\Lambda_j e_j,$$

where $\Lambda_j(t) = \boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}_{K,n}\boldsymbol{N}_{\nu+1}(t_j)/n$. To verify that the Lindeberg-Feller condition holds, it suffices to show that

$$\max_{1\leqslant j\leqslant n}(\Lambda_j^2) = o(\sum_{j=1}^{n}\Lambda_j^2) = o(\mathrm{Var}[\hat{X}^{(i)}(t)]). \tag{A.2}$$

From Lemma A1 in Claeskens, Krivobokova and Opsomer(2009), $\|\boldsymbol{H}^{-1}_{K,n}\|_{\infty} = O(\kappa^{-1})$. Similar to the proof of Theorem 3.3 in Zhou and Wolfe (2000), we have

$$\begin{aligned}
\Lambda_j^2 n^2 &= \boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}_{K,n}\boldsymbol{N}_{\nu+1}(t_j)\boldsymbol{N}^T_{\nu+1}(t_j)\boldsymbol{H}^{-1}_{K,n}\boldsymbol{\Delta}_i^T\boldsymbol{N}_{\nu+1-i}(t)\\
&= \mathrm{tr}[\boldsymbol{N}_{\nu+1-i}(t)\boldsymbol{N}^T_{\nu+1-i}(t)\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}_{K,n}\boldsymbol{N}_{\nu+1}(t_j)\boldsymbol{N}^T_{\nu+1}(t_j)\boldsymbol{H}^{-1}_{K,n}\boldsymbol{\Delta}_i^T]\\
&\leqslant \pi_{\nu+1-i}\mathrm{tr}[\boldsymbol{\Delta}_i^T\boldsymbol{\Delta}_i\boldsymbol{H}^{-1}_{K,n}\boldsymbol{N}_{\nu+1}(t_j)\boldsymbol{N}^T_{\nu+1}(t_j)\boldsymbol{H}^{-1}_{K,n}]\\
&\leqslant \pi_{\nu+1-i}\|\boldsymbol{\Delta}_i^T\|_{\infty}\|\boldsymbol{\Delta}_i\|_{\infty}\mathrm{tr}[\boldsymbol{H}^{-2}_{K,n}\boldsymbol{N}_{\nu+1}(t_j)\boldsymbol{N}^T_{\nu+1}(t_j)]
\end{aligned}$$

$$\begin{aligned}
&\leqslant \pi_{\nu+1-i}\|\boldsymbol{\Delta}_i\|_\infty^2\|\boldsymbol{H}_{K,n}^{-2}\|_\infty \mathrm{tr}[\boldsymbol{N}_{\nu+1}(t_j)\boldsymbol{N}_{\nu+1}^T(t_j)]\\
&\leqslant \pi_{\nu+1-i}O(\kappa^{-2i-2})\sum_{l=1}^{K+\nu+1} N_{l,\nu+1}^2(t_j)\\
&= O(\kappa^{-2i-2}),
\end{aligned}$$

where

$$\pi_{\nu+1-i} = \max_{t\in[a,b]}\{\varphi(t) : \varphi(t) \text{ is the maximum eigenvalue of } \boldsymbol{N}_{\nu+1-i}(t)\boldsymbol{N}_{\nu+1-i}^T(t)\} \leqslant 1.$$

Similarly, for $K_q \geqslant 1$, if $\lambda \sim cn^\gamma$ with $\gamma \leqslant 1/(2q+1)$, and $K \sim cn^\varsigma$ with $\varsigma \geqslant 1/(2q+1)$, from $\|\boldsymbol{H}_{K,n}^{-1}\|_\infty = O(\kappa^{-1}(1+K_q^{2q})^{-1})$, we have

$$\begin{aligned}
\Lambda_j^2 n^2 &= O(\kappa^{-2i-2}(1+K_q^{2q})^{-2})\\
&= O(\kappa^{-2i-1}(\lambda/n)^{-1/(2q)}K_q(1+K_q^{2q})^{-2})\\
&= O(\kappa^{-2i-1}(\lambda/n)^{-1/(2q)}),
\end{aligned}$$

since $K_q(1+K_q^{2q})^{-2} < 1/2$. Hence (A.2) holds for both $K_q < 1$ and $K_q \geqslant 1$, and the proof of Lemma 3 is completed.

Lemma 4: (i) Under Assumptions B1-B3 and B11, if $X(.) \in C^{\nu+1}[a,b]$ with $\nu \geqslant 1$, for $K \sim cn^\varsigma$ with $\varsigma \geqslant 1/(2\nu+3)$, and $\lambda = O(n^\gamma)$ with $\gamma \leqslant (\nu+2-q)/(2\nu+3)$, then for any fixed $t \in [a,b]$ and any $i = 0, 1, \cdots, \nu-1$,

$$\|\hat{X}^{(i)}(t) - X^{(i)}(t)\|_\infty = O(\kappa^{\nu+1-i}) + O(\lambda n^{-1}\kappa^{-q-i}) + O(\kappa^{-i-\frac{1}{2}}n^{-\frac{1}{2}}).$$

(ii) Under Assumptions B1-B3 and B12, if $X(.) \in W^q[a,b]$ with $q \geqslant 2$, for $\lambda \sim cn^\gamma$ with $\gamma \leqslant 1/(2q+1)$, and $K \sim cn^\varsigma$ with $\varsigma \geqslant 1/(2q+1)$, then for any fixed $t \in [a,b]$ and any $i = 0, 1, \cdots, q-2$,

$$\|\hat{X}^{(i)}(t) - X^{(i)}(t)\|_\infty = O(\kappa^{q-i}) + O((\lambda/n)^{1/2}\kappa^{-i}) + O(n^{-1/2}\kappa^{-i}(\lambda/n)^{-1/(4q)}).$$

Proof of Lemma 4: From Lemma 3(i), for any $t \in [a,b]$, we have

$$\mathrm{Var}^{-\frac{1}{2}}(\hat{X}^{(i)}(t))[\hat{X}^{(i)}(t) - X^{(i)}(t) - b_i(t,\nu+1) - d_i(t)] = O(1).$$

Then by Lemma 1(i), Lemma 4(i) holds. The proof of Lemma 4(ii) is similar.

Remark A1: Note that Claeskens, Krivobokova and Opsomer(2009) have established Lemma 1 and Lemma 2 for $i = 0$, i.e., the case of the original function $X(t)$ and its estimate $\hat{X}(t)$. The results of Lemmas 1-4 are similar to those for other smoothing methods, such as smoothing splines (Rice and Rosenblatt, 1983), local polynomial (Fan and Gijbels, 1996) and regression splines (Zhou, Shen and Wolfe, 1998; Zhou and Wolfe, 2000).

Remark A2: Lemma 2 shows that the order of the optimal number of knots for P-splines does not depend on i. Such results are similar to that for regression splines given in Remark 1 of Zhou and Wolfe(2000). It provides a clue on how to choose the optimal number of knots for $\hat{X}^{(i)}(t)$.

Remark A3: Similar to the setup for regression splines in Zhou, Shen and Wolfe(1998) and Zhou and Wolfe(2000), in this article we suppose that the knots for P-slines are asymptotically equally-spaced, the design density is continuous, and the order of the spline equals the assumed order of the unknown regression function. These assumptions may be relaxed by the way for regression splines in Huang(2003a,b).

Lemma 5: For all the discretization methods given in Section 2.2, we have

$$F(t_i, X(t_i), X(t_{i+1}), \boldsymbol{\beta}) = f(t_i, X(t_i), \boldsymbol{\beta}) + O(h),\ i = 1, \cdots, n-1.$$

Proof of Lemma 5: We only need to verify this result for the RDB method. For k_2 defined in Section 2.2, under Assumption B10, by the mean value theorem, we have

$$\begin{aligned} k_2 &= f(t_i + h_i/2, X(t_i) + h_i k_1/2, \boldsymbol{\beta}) \\ &= f(t_i, X(t_i), \boldsymbol{\beta}) + \frac{\partial f(t^*, x^*, \boldsymbol{\beta})}{\partial t} \frac{h_i}{2} + \frac{\partial f(t^*, x^*, \boldsymbol{\beta})}{\partial x} \frac{h_i k_1}{2} \\ &= f(t_i, X(t_i), \boldsymbol{\beta}) + O(h_i) \\ &= f(t_i, X(t_i), \boldsymbol{\beta}) + O(h) \end{aligned}$$

where t^* lies between t_i and $t_i + h_i/2$, x^* lies between $X(t_i)$ and $X(t_i) + h_i k_1/2$. Similarly, we have $k_3 = f(t_i, X(t_i), \boldsymbol{\beta}) + O(h)$ and $k_4 = f(t_i, X(t_i), \boldsymbol{\beta}) + O(h)$. Then

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

Lemma 6: Under Assumptions B1-B3 and B11, if $X(.) \in C^{\nu+1}[a,b]$ with $\nu \geqslant 1$, for $K \sim cn^{\varsigma}$ with $\varsigma \geqslant 1/(2\nu+3)$, and $\lambda = O(n^{\gamma})$ with $\gamma \leqslant (\nu+2-q)/(2\nu+3)$, for a continuous function $g(t)$, $\int_a^b g(t)[\hat{X}(t) - X(t)]dt$ is asymptotically normal with mean $\mu_X = \int_a^b g(t)\mathrm{E}[\hat{X}(t) - X(t)]dt = O(\kappa^{v+1}) = o(n^{-1/2})$ and variance $\Sigma_X = \int_a^b \int_a^b g(s)\mathrm{Cov}[\hat{X}(s) - X(s), \hat{X}(t) - X(t)]g^T(t)dsdt = O(1/n)$.

Proof of Lemma 6: By the Delta-method(van der Vaart and Wellner, 1996, p.377) and Lemma 3(i), $\int_a^b g(t)[\hat{X}(t) - X(t)]dt$ is asymptotically normal with mean as

$$\mu_X = \int_a^b g(t)\mathrm{E}[\hat{X}(t) - X(t)]dt,$$

and variance as

$$\Sigma_X = \int_a^b \int_a^b g(s)\mathrm{Cov}[\hat{X}(s) - X(s), \hat{X}(t) - X(t)]g^T(t)dsdt.$$

From Lemma 1(i), $\mu_X = O(\kappa^{v+1}) = o(n^{-1/2})$ and

$$\Sigma_X = \frac{\sigma^2}{n}\int_a^b \int_a^b g(s)\boldsymbol{N}_{\nu+1}^T(s)(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{G}(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{N}_{\nu+1}(t)g^T(t)dsdt.$$

Now, we consider the order of the variance as follows

$$\begin{aligned}
& \|\int_a^b \int_a^b g(s)\boldsymbol{N}_{\nu+1}^T(s)(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{G}(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{N}_{\nu+1}(t)g^T(t)dsdt\|_\infty \\
\leqslant\ & c\|\int_a^b \boldsymbol{N}_{\nu+1}^T(t)(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{G}(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{N}_{\nu+1}(t)dt\|_\infty \\
=\ & c\mathrm{tr}[\int_a^b \boldsymbol{N}_{\nu+1}^T(t)(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{G}(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{N}_{\nu+1}(t)dt] \\
=\ & c\int_a^b \mathrm{tr}[\boldsymbol{N}_{\nu+1}^T(t)(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{G}(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{N}_{\nu+1}(t)]dt \\
=\ & c\int_a^b \mathrm{tr}[(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{G}(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{N}_{\nu+1}(t)\boldsymbol{N}_{\nu+1}^T(t)]dt \\
=\ & c\int_a^b \mathrm{tr}[(\boldsymbol{G}^{-1} + o(\kappa^{-1}))\boldsymbol{N}_{\nu+1}(t)\boldsymbol{N}_{\nu+1}^T(t)]dt \\
\leqslant\ & c\|\boldsymbol{G}^{-1}\|_\infty \mathrm{tr}[\int_a^b \boldsymbol{N}_{\nu+1}(t)\boldsymbol{N}_{\nu+1}^T(t)\rho(t)dt]
\end{aligned}$$

$$\leqslant \quad c\|\boldsymbol{G}^{-1}\|_\infty \|\boldsymbol{G}\|_2 = O(\kappa^{-1})O(\kappa) = O(1),$$

then $\Sigma_X = O(1/n)$.

In the following proofs of Propositions 1-2, for simplicity, we suppose that f does not depend on t and use the notation $f(X(t), \boldsymbol{\beta})$ to replace $f(t, X(t), \boldsymbol{\beta})$. We also remove the dependence in t for F. Of course, the proof is straightforward in the non-autonomous case.

Proof of Proposition 1: If we denote $\epsilon_i = \frac{\hat{X}(t_{i+1})-\hat{X}(t_i)}{t_{i+1}-t_i} - F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}_0)$, then $\frac{1}{n}\sum_{i=1}^{n-1} w(t_i)\epsilon_i^2 = O(a_n)$ with $a_n = b_n^2 + hb_n + h^2$ for $b_n = \kappa^{\nu} + \lambda n^{-1}\kappa^{-q-1} + \kappa^{-\frac{3}{2}}n^{-\frac{1}{2}}$. This result can be verified by the Taylor expansion and the expression (10), i.e., for $0 < h \leqslant 1$, under Assumptions B7 and B13, as follows

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n-1} w(t_i)\epsilon_i^2 &= \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[\frac{\hat{X}(t_{i+1}) - \hat{X}(t_i)}{t_{i+1}-t_i} - F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}_0)]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[\hat{X}'(t_i) + O(h) - \frac{X(t_{i+1}) - X(t_i)}{t_{i+1}-t_i} + O(h^p)]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[\hat{X}'(t_i) + O(h) - X'(t_i) + O(h) + O(h^p)]^2 \qquad \text{(A.3)} \\
&= \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[\hat{X}'(t_i) - X'(t_i) + O(h)]^2 \\
&\leqslant c\|\hat{X}'(t) - X'(t)\|_\infty^2 + O(h)\|\hat{X}'(t) - X'(t)\|_\infty + O(h^2) \\
&= O(b_n^2 + hb_n + h^2) = O(a_n).
\end{aligned}$$

Then we have

$$\begin{aligned}
\frac{1}{n}S_n(\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[\frac{\hat{X}(t_{i+1}) - \hat{X}(t_i)}{t_{i+1}-t_i} - F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta})]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}_0) + \epsilon_i - F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta})]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}_0) - F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta})]^2 \qquad \text{(A.4)} \\
&\quad + \frac{2}{n}\sum_{i=1}^{n-1} w(t_i)\epsilon_i[F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}_0) - F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta})] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)\epsilon_i^2.
\end{aligned}$$

For the second term of (A.4), by the Cauchy-Schwarz inequality, if $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, then

$$\begin{aligned} & \frac{2}{n}\Big|\sum_{i=1}^{n-1} w(t_i)\epsilon_i[F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}_0) - F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta})]\Big| \\ \leqslant\ & c[\sum_{i=1}^{n-1}\epsilon_i^2]^{\frac{1}{2}}\{\frac{1}{n}\sum_{i=1}^{n-1}[F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}_0) - F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta})]^2\}^{\frac{1}{2}}. \end{aligned}$$

So the second term of (A.4) is bounded by a product of a lower term than that of the first term of (A.4) and $o_p(1)$. Now we consider the first term of (A.4), which can be decomposed as

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}_0) - F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta})]^2 \\ =\ & \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}_0) - F(X(t_i), X(t_{i+1}), \boldsymbol{\beta})]^2 \\ & +\frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}) - F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta})]^2 \\ & +\frac{2}{n}\sum_{i=1}^{n-1} w(t_i)[F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}_0) - F(X(t_i), X(t_{i+1}), \boldsymbol{\beta})] \\ & \times[F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}) - F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta})]. \end{aligned} \tag{A.5}$$

By Assumption B10 and Lemma 4(i), we know that the second term of (A.5) is bounded as follows

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}) - F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta})]^2 \\ \leqslant\ & c\|X(t) - \hat{X}(t)\|_\infty^2 \sup_{x\in\mathcal{X}}\Big|\frac{\partial}{\partial x}f(t, x, \boldsymbol{\beta})\Big|^2 \\ \leqslant\ & c'c_n^2 \end{aligned}$$

with $c_n = \kappa^{\nu+1} + \lambda n^{-1}\kappa^{-q} + \kappa^{-\frac{1}{2}}n^{-\frac{1}{2}}$ for some constant c'. By a similar argument, we note that, if $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, the third term of (A.5) is bounded by $O(c_n)$. For the first term of (A.5), from Lemma 5 and the strong law of large number, if $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, we have

$$\begin{aligned} & \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[F(X(t_i), X(t_{i+1}), \boldsymbol{\beta}_0) - F(X(t_i), X(t_{i+1}), \boldsymbol{\beta})]^2 \\ =\ & \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[f(X(t_i), \boldsymbol{\beta}_0) - f(X(t_i), \boldsymbol{\beta})]^2 \\ =\ & \mathrm{E}\{w(t)[f(X(t), \boldsymbol{\beta}_0) - f(X(t), \boldsymbol{\beta})]^2\}, \end{aligned}$$

a.s. Combining all three terms of (A.5), we can see that the first term of (A.4) is dominated by the term $\mathrm{E}\{w(t)[f(X(t),\boldsymbol{\beta}_0) - f(X(t),\boldsymbol{\beta})]^2\}$. Denote $S(\boldsymbol{\beta}) = \mathrm{E}\{w(t)[f(X(t),\boldsymbol{\beta}_0) - f(X(t),\boldsymbol{\beta})]^2\}$. Then we have $\lim\limits_{n\to\infty} \frac{S_n(\boldsymbol{\beta})}{n} - S(\boldsymbol{\beta}) = 0$, a.s.

For any probability measure Q, let $\mathcal{F}_n$ be the set $\{\frac{S_n(\boldsymbol{\beta})}{n} - S(\boldsymbol{\beta}) : \boldsymbol{\beta} \in \Omega_{\boldsymbol{\beta}}\}$ and $N_1(\epsilon, Q, \mathcal{F}_n)$ be the covering number of the class $\mathcal{F}_n$ in the probability measure Q, as given in Pollard(1984,page 25). Under Assumption B4, using the similar arguments to those in the proof of Theorem 3.1 of Xue, Miao and Wu(2010), we have $\sup_Q N_1(\epsilon, Q, \mathcal{F}_n) \leqslant C(\frac{1}{\epsilon})^d$ for $0 < \epsilon < 1$, where C is some constant. Then by Theorem II.37 in Pollard(1984), $\sup_{\boldsymbol{\beta}} |\frac{S_n(\boldsymbol{\beta})}{n} - S(\boldsymbol{\beta})| \to 0$, a.s., under $P_{\boldsymbol{\beta}_0}$.

Next, from Assumption B8, we know that $\boldsymbol{\beta}_0$ is the unique minimum point of $S(\boldsymbol{\beta})$. Since $\boldsymbol{\beta}_0$ is an interior point of Ω_β, it follows that the first-order derivative $\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ of $S(\boldsymbol{\beta})$ at $\boldsymbol{\beta}_0$ equals to zero and the second-order derivative $\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}$ of $S(\boldsymbol{\beta})$ at $\boldsymbol{\beta}_0$ is positive definite. By Assumptions B2 and B9, the second-order derivative of $S(\boldsymbol{\beta})$ in a small neighborhood of $\boldsymbol{\beta}_0$ is bounded away from 0 and ∞. Then the second-order Taylor expansion of $S(\boldsymbol{\beta})$ gives that there exists a constant $0 < C < \infty$ such that $|S(\hat{\boldsymbol{\beta}}_n) - S(\boldsymbol{\beta}_0)| \geqslant C\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|^2$. Moreover,

$$\begin{aligned} 0 \leqslant S(\hat{\boldsymbol{\beta}}_n) - S(\boldsymbol{\beta}_0) &= S(\hat{\boldsymbol{\beta}}_n) - \frac{S_n(\hat{\boldsymbol{\beta}}_n)}{n} + \frac{S_n(\hat{\boldsymbol{\beta}}_n)}{n} - S(\boldsymbol{\beta}_0) \\ &\leqslant S(\hat{\boldsymbol{\beta}}_n) - \frac{S_n(\hat{\boldsymbol{\beta}}_n)}{n} + \frac{S_n(\boldsymbol{\beta}_0)}{n} - S(\boldsymbol{\beta}_0) \\ &\leqslant 2\sup_{\boldsymbol{\beta}} |\frac{S_n(\boldsymbol{\beta})}{n} - S(\boldsymbol{\beta})| \to 0, \text{ a.s.} \end{aligned}$$

Thus $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| \to 0$, a.s., i.e. $\hat{\boldsymbol{\beta}}_n$ strongly converges to $\boldsymbol{\beta}_0$ when n is large.

Proof of Proposition 2: Under Assumption B9, by the Landau-Kolmogorov inequality between different derivatives of a function and Lemma 5, we have

$$\begin{aligned} & \|\frac{\partial F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \frac{\partial f(\hat{X}(t_i), \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\|_\infty \\ \leqslant\ & C\|\frac{\partial^2 F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} - \frac{\partial^2 f(\hat{X}(t_i), \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\|_\infty^{1/2} \|F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta}) - f(\hat{X}(t_i), \boldsymbol{\beta})\|_\infty^{1/2} \\ \leqslant\ & C'\|F(\hat{X}(t_i), \hat{X}(t_{i+1}), \boldsymbol{\beta}) - f(\hat{X}(t_i), \boldsymbol{\beta})\|_\infty^{\frac{1}{2}} \end{aligned}$$

$$= \quad O(h^{\frac{1}{2}})$$

for two constants C and C', where $\|g(\boldsymbol{\beta})\|_\infty = \sup_{\boldsymbol{\beta}} |g(\boldsymbol{\beta})|$ is the supremum norm of a function g . For $\hat{\boldsymbol{\beta}}_n$, it satisfies $\frac{\partial S_n(\hat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}} = 0$. Then from Lemma 5 and the proof of Proposition 1, it follows that

$$\begin{aligned}
&-\frac{1}{n}\frac{\partial S_n(\hat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}} \\
= &\ \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[\frac{\hat{X}(t_{i+1}) - \hat{X}(t_i)}{t_{i+1} - t_i} - F(\hat{X}(t_i), \hat{X}(t_{i+1}), \hat{\boldsymbol{\beta}}_n)]\frac{\partial F(\hat{X}(t_i), \hat{X}(t_{i+1}), \hat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}} \\
= &\ \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)[\hat{X}'(t_i) - f(\hat{X}(t_i), \hat{\boldsymbol{\beta}}_n) + O(h)][\frac{\partial f(\hat{X}(t_i), \hat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}} + O(h^{\frac{1}{2}})] \\
= &\ \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)\frac{\partial f(\hat{X}(t_i), \hat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}}[\hat{X}'(t_i) - X'(t_i) + f(X(t_i), \boldsymbol{\beta}_0) - f(\hat{X}(t_i), \boldsymbol{\beta}_0) \\
&+f(\hat{X}(t_i), \boldsymbol{\beta}_0) - f(\hat{X}(t_i), \hat{\boldsymbol{\beta}}_n)] + O(h^{\frac{1}{2}})o(1) \\
= &\ \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)\frac{\partial f(\hat{X}(t_i), \hat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}}[\hat{X}'(t_i) - X'(t_i) + \frac{\partial f(\tilde{X}_i, \boldsymbol{\beta}_0)}{\partial X}(X(t_i) - \hat{X}(t_i)) \\
&+\frac{\partial f(\hat{X}(t_i), \tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^T}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_n)] + o(h^{\frac{1}{2}}),
\end{aligned}$$

with $\tilde{X}_i$ being some point between $X(t_i)$ and $\hat{X}(t_i)$, and $\tilde{\boldsymbol{\beta}}$ being some point between $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}_n$. $o(h^{\frac{1}{2}}) = o(n^{-1/2})$, since $h = O(n^{-1})$. So we obtain an asymptotic expression for $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0$ as follows

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n-1} w(t_i)\frac{\partial f(\hat{X}(t_i), \hat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}}\frac{\partial f(\hat{X}(t_i), \tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\
= &\ \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)\frac{\partial f(\hat{X}(t_i), \hat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}}\{[\hat{X}'(t_i) - X'(t_i)] \\
&-\frac{\partial f(\tilde{X}_i, \boldsymbol{\beta}_0)}{\partial X}[\hat{X}(t_i) - X(t_i)]\} + o(n^{-\frac{1}{2}}).
\end{aligned}$$

Denote $J_* = \mathrm{E}[w(t)\{\frac{\partial f(X(t), \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\}^{\otimes 2}]$,

$$P_n = \frac{1}{n}\sum_{i=1}^{n-1} w(t_i)\frac{\partial f(\hat{X}(t_i), \hat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}}[(\hat{X}'(t_i) - X'(t_i)) - \frac{\partial f(\tilde{X}_i, \boldsymbol{\beta}_0)}{\partial X}(\hat{X}(t_i) - X(t_i)],$$

and

$$T_n = \int_a^b w(t)\rho(t)\frac{\partial f(X(t), \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}[(\hat{X}'(t) - X'(t)) - \frac{\partial f(X(t), \boldsymbol{\beta}_0)}{\partial X}(\hat{X}(t) - X(t)]dt.$$

From Proposition 1 and Lemma 4(i), we have

$$\frac{1}{n}\sum_{i=1}^{n-1} w(t_i)\frac{\partial f(\hat{X}(t_i),\hat{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}}\frac{\partial f(\hat{X}(t_i),\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^T} \to J_*.$$

For fixed $\hat{X}'(t)$ and $X'(t)$, from the weak law of large number, $P_n - T_n \to 0$ in probability. So $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 = J_*^{-1}T_n + o_p(n^{-1/2})$. Further, applying integration by parts to T_n, we have

$$\begin{aligned} & \int_a^b w(t)\rho(t)\frac{\partial f(X(t),\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}[\hat{X}'(t) - X'(t)]dt \\ = & \ \{w(t)\rho(t)\frac{\partial f(X(t),\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}[\hat{X}(t) - X(t)]\}|_a^b \\ & - \int_a^b \frac{d}{dt}[w(t)\rho(t)\frac{\partial f(X(t),\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}][\hat{X}(t) - X(t)]dt. \end{aligned}$$

If $w(a) = w(b) = 0$, then

$$\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 = -J_*^{-1}\int_a^b A(t)[\hat{X}(t) - X(t)]dt + o_p(n^{-\frac{1}{2}}). \qquad \text{(A.6)}$$

with $A(t) = w(t)\rho(t)\frac{\partial f(X(t),\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\frac{\partial f(X(t),\boldsymbol{\beta}_0)}{\partial X} + \frac{d}{dt}[w(t)\rho(t)\frac{\partial f(X(t),\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}]$. From Lemma 6, the right side of (A.6) is asymptotically normal with mean as

$$\mu_1 = -J_*^{-1}\int_a^b A(t)\mathrm{E}[\hat{X}(t) - X(t)]dt = O(\kappa^{v+1}) = o(n^{-1/2})$$

and variance as

$$\begin{aligned} \Sigma_1 &= J_*^{-1}\{\int_a^b\int_a^b A(s)\mathrm{Cov}[\hat{X}(s) - X(s), \hat{X}(t) - X(t)]A^T(t)dsdt\}J_*^{-1} \\ &= \frac{\sigma^2}{n}J_*^{-1}\Big[\int_a^b\int_a^b A(s)\boldsymbol{N}_{\nu+1}^T(s)(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{G}(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{N}_{\nu+1}(t)A^T(t)dsdt\Big]J_*^{-1} \\ &= O(1/n). \end{aligned}$$

Thus for $w(a) = w(b) = 0$, if $K_q < 1$, $K \sim cn^{\varsigma}$ with $\varsigma \geqslant 1/(2\nu + 3)$, and $\lambda = O(n^{\gamma})$ with $\gamma \leqslant (\nu + 2 - q)/(2\nu + 3)$, then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \Sigma_1^*) \qquad \text{with} \qquad \Sigma_1^* = n\Sigma_1. \qquad \text{(A.7)}$$

Obviously, $\Sigma_1^* = O(1)$, since $\Sigma_1 = O(1/n)$. This rate of convergence is consistent with that of Theorem 4.2 in Bickel and Ritov(2003).

If $w(a) \neq 0$ or $w(b) \neq 0$, we have

$$\begin{aligned}
& \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 \\
= \; & J_*^{-1}\{w(t)\rho(t)\frac{\partial f(X(t),\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}[\hat{X}(t) - X(t)]\}|_a^b + O_p(n^{-\frac{1}{2}}) \\
= \; & J_*^{-1}\{B(b)[\hat{X}(b) - X(b)] - B(a)[\hat{X}(a) - X(a)]\} + O_p(n^{-\frac{1}{2}}) \\
= \; & J_*^{-1}[B(b)\boldsymbol{N}_{\nu+1}^T(b) - B(a)\boldsymbol{N}_{\nu+1}^T(a)](\boldsymbol{Z}^T\boldsymbol{Z} + \lambda\boldsymbol{D}_q)^{-1}\boldsymbol{Z}^T\boldsymbol{Y} + O_p(n^{-\frac{1}{2}}) \\
= \; & \frac{1}{n}J_*^{-1}M\boldsymbol{H}_{K,n}^{-1}\boldsymbol{Z}^T\boldsymbol{Y} + O_p(n^{-\frac{1}{2}})
\end{aligned}$$

with $B(t) = w(t)\rho(t)\frac{\partial f(X(t),\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}$ and $M = B(b)\boldsymbol{N}_{\nu+1}^T(b) - B(a)\boldsymbol{N}_{\nu+1}^T(a)$. Similar to the proof of Lemma 3(i), $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0$ is asymptotic normal with mean as

$$\mu_2 = J_*^{-1}\Big\{B(b)\mathrm{E}[\hat{X}(b) - X(b)] - B(a)\mathrm{E}[\hat{X}(a) - X(a)]\Big\}$$

and variance as

$$\Sigma_2 = \frac{\sigma^2}{n}J_*^{-1}M(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}\boldsymbol{G}(\boldsymbol{G} + \frac{\lambda}{n}\boldsymbol{D}_q)^{-1}M^TJ_*^{-1}.$$

Now, we consider the orders of this mean and variance. From Lemma 1(i), $\mu_2 = O(\kappa^{\nu+1})$ and $\Sigma_2 = O(n^{-1}\kappa^{-1})$. In order to ensure that the asymptotic bias of $\hat{\boldsymbol{\beta}}_n$ is zero, the assumption $Kn^{-1/(2\nu+3)} \to \infty$ is required. Thus for $w(a) \neq 0$ or $w(b) \neq 0$, if $K_q < 1$, $K \sim cn^\varsigma$ with $\varsigma > 1/(2\nu+3)$, and $\lambda = O(n^\gamma)$ with $\gamma \leqslant (\nu+2-q)/(2\nu+3)$, then

$$\sqrt{n\kappa}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \Sigma_2^*) \qquad \text{with} \qquad \Sigma_2^* = n\kappa\Sigma_2. \tag{A.8}$$

Obviously, $\Sigma_2^* = O(1)$, since $\Sigma_2 = O(n^{-1}\kappa^{-1})$.

References

Bickel, P. J. and Ritov, Y. (2003). Nonparametric estimators which can be "plugged-in". *Ann. Statist.*, **31**, 1033–1053.

Cardot, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *J. Nonpar. Statist.*, **12**, 503–538.

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

Ghizzetti, A. and Ossicini, A. (1970). *Quadrature Formulae.* New York: Academic Press.

Huang, J. Z. (2003a). Local asymptotic for polynomial spline regression. *Ann. Statist.*, **31**, 1600–1635.

Huang, J. Z. (2003b). Asymptotics for polynomial spline regression under weak conditions. *Statist. Prob. Lett.*, **65**, 207–216.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag, New York.

Rice, J. and Rosenblatt, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *Ann. Statist.*, **11**, 141–156.

van der Vaart, A. W. and Wellner, J. A. (1996). *W*eak Convergence and Empirical Processes. Springer-Verlag, New York.

Zhou, S. and Wolfe, D. (2000). On derivative regression in spline estimation. *Statistica Sinica*, **10**, 93–108.

Appendix B: Additional Simulation Studies and Results

In all the simulations in the main text, the weight function $w(t_i)$ in the objective function (15) is assumed to be a constant. To evaluate the effect of the weight function, we performed additional simulations based on Model (19) in Simulation Example II by assuming $w(t_i) = \sin(\pi t_i/5)$. This weight function is selected to satisfy the boundary conditions for the asymptotic results as suggested by Brunel (2008). Similarly we used parameter values $(a, b, c) = (1.5, 1, 2)$ and initial values $(R(0), P(0)) = (0, 1)$ to generate the observations at every time interval of 0.2 in [0,5] giving 26 observations. The measurement error standard deviations were taken as $(\sigma_1, \sigma_2) \in \{(0.02, 0.01), (0.05, 0.03)\}$. For each simulation case, 500 runs were replicated. We report the AREs for the estimates of different methods in Table 1 below.

The results show considerable improvement in the ARE for all the methods. Comparing different methods, the EDB method again has the largest ARE. The LW method performed better than the TDB method for the cases of small noise. For the larger noise cases, the TDB method has the smallest ARE. The RDB method was slightly worse than the LW and the TDB method for most of the cases.

[Table 1 about here.]

We performed more simulations for the case of unequally-distributed observation time points suggested by one of the referees. Here we report the results from one case that is also based on the non-linear ODE model (19). We used parameter values $(a, b, c) = (1.5, 1, 2)$ and initial values $(R(0), P(0)) = (0, 1)$ to generate the observations at unequally-spaced intervals: early frequent measurements with every interval of 0.2 in [0,2.5] and sparse measurements with every interval of 0.4 in [2.5,5]. The measurement error standard deviations were taken as $(\sigma_1, \sigma_2) \in \{(0.02, 0.01), (0.05, 0.03)\}$. For each simulation case, 500 runs were replicated. We report the AREs for the estimates of different methods for different parameters in Table 2 below.

From this table, we can see a similar trend as in the case of equally-spaced observations (Example II in the main text). The results confirm that the EDB method produces the largest AREs for all the simulation cases. The LW method seems to be somewhere in between the EDB and the TDB. The RDB method was worse than the TDB method and was better than the LW method in some cases and worse in others. The TDB method again gave the best results and has been stable in all the simulations.

[Table 2 about here.]

Table 1

Evaluation of the weight effect: Average relative errors in percentage (sample mean± sample standard deviation) for the estimates of the parameters obtained from 500 replications (note that true $a = 1.5, b = 1,$ and $c = 2$ and $n = 26$ for Example II).

Parameter	(σ_1,σ_2)	LW	EDB	TDB	RDB
a	(0.02,0.01)	3.17	9.57	3.39	3.71
		(1.50±0.06)	(1.36±0.06)	(1.50±0.06)	(1.53±0.07)
	(0.05,0.03)	8.33	13.88	7.53	10.27
		(1.44±0.14)	(1.40±0.23)	(1.46±0.13)	(1.57±0.21)
b	(0.02,0.01)	3.73	9.70	3.97	4.25
		(1.00±0.05)	(0.90±0.04)	(1.00±0.05)	(1.02±0.05)
	(0.05,0.03)	9.78	15.57	8.81	11.88
		(0.95±0.11)	(0.93±0.18)	(0.97±0.10)	(1.05±0.16)
c	(0.02,0.01)	0.63	1.58	0.69	0.66
		(2.00±0.02)	(1.97±0.02)	(2.00±0.02)	(2.00±0.02)
	(0.05,0.03)	1.85	1.93	1.84	1.78
		(2.00±0.05)	(1.98±0.04)	(1.99±0.05)	(2.01±0.04)

Table 2

Simulation Results for Unequally-Distributed Observation Time Points: Average relative errors in percentage (sample mean± sample standard deviation) for the estimates of the parameters obtained from 500 replications (note that true a=1.5, b=1, and c=2).

par	(σ_1,σ_2)	LW	EDB	TDB	RDB
a	(0.02,0.01)	10.64	16.37	7.05	11.58
		(1.36±0.12)	(1.25±0.08)	(1.42±0.09)	(1.33±0.09)
	(0.05,0.03)	20.29	20.45	14.88	17.66
		(1.24±0.23)	(1.21±0.17)	(1.32±0.19)	(1.27±0.19)
b	(0.02,0.01)	12.81	17.69	8.52	13.24
		(0.89±0.09)	(0.82±0.07)	(0.94±0.08)	(0.87±0.07)
	(0.05,0.03)	24.26	23.09	17.89	20.82
		(0.79±0.18)	(0.79±0.15)	(0.86±0.16)	(0.82±0.16)
c	(0.02,0.01)	3.36	9.97	1.27	6.99
		(1.93±0.03)	(1.80±0.02)	(1.98±0.02)	(1.86±0.02)
	(0.05,0.03)	6.07	11.25	3.34	8.24
		(1.89±0.08)	(1.77±0.05)	(1.94±0.06)	(1.84±0.06)