
Supplementary Material: “Modeling spiky functional data with derivatives of smooth functions in function-on-function regression”

Ruiyan Luo and Xin Qi

Georgia State University

Contents

S1	Proof of Theorems	2
S1.1	Proof of Theorem 1	2
S1.2	Proof of Theorem 2	2
S2	Additional details for computation	14
S2.1	Solving (2.17)	14
S2.2	Solving (2.18)	15
S3	Additional details for numerical illustration	16
S3.1	Simulation 1	16
S3.1.1	Results for settings with $2^7 = 128$ observation points	18
S3.1.2	Additional results with $2^9 = 512$ observation time points	22
S3.2	Simulation 2	28
S3.3	Simulation 3	29
S3.4	Real data	33

In this supplementary material, we provide proofs for the theorems in the main manuscript, additional details for computation, simulations and real data analysis. The labels of equations, figures, tables, sections, and so on, in this supplementary material are all prefixed with ‘‘S’’ , for example, equations: (S1.1), (S1.2), \dots , and section S1.1. Those in the main manuscript are labeled using the original index as in the main manuscript without prefix ‘‘S’’.

S1 Proof of Theorems

S1.1 Proof of Theorem 1

Let $\sum_{k=1}^K \varphi_k^{(opt)}(s)\zeta_k^{(opt)}(t)$ denote the optimal representation of $\mathfrak{B}(s, t)$ defined in Luo and Qi (2017) that minimizes

$$\min_{\substack{\varphi_k(s), \zeta_k(t), \\ 1 \leq k \leq K}} \mathbf{E} \left[\int_0^1 \left(F(t) - \left\{ \mathfrak{U}(t) + \int_0^1 X(s) \sum_{k=1}^K \varphi_k(s) \zeta_k(t) ds \right\} \right)^2 dt \right], \quad (\text{S1.1})$$

where the minimization is over all possible square intergrable functions $\varphi_k(s)$ and $\zeta_k(t)$ for $1 \leq k \leq K$. Theorem 1 in Luo and Qi (2017) characterizes $\varphi_k^{(opt)}(s)$, $1 \leq k \leq K$, as the solution to

$$\begin{aligned} \max_{\varphi} \int_0^1 \int_0^1 \varphi(s) \mathbf{B}(s, s') \varphi(s') ds ds', \quad \text{subject to} \quad \int_0^1 \int_0^1 \varphi(s) \mathbf{\Sigma}(s, s') \varphi(s') ds ds' = 1, \\ \text{and} \quad \int_0^1 \int_0^1 \varphi(s) \mathbf{\Sigma}(s, s') \varphi_l^{(opt)}(s') ds ds' = 0 \quad \text{for all} \quad 1 \leq l \leq k - 1. \end{aligned} \quad (\text{S1.2})$$

Comparing (2.1) to (S1.1), $\sum_{k=1}^K D^{d_1} \phi_k^{(opt)}(s) D^{d_2} \xi_k^{(opt)}(t)$ is an optimal representation of $\mathfrak{B}(s, t)$. So the $D^{d_1} \phi_k^{(opt)}(s)$ satisfy the optimization problem (S1.2), which means that $\phi_k^{(opt)}(s)$ satisfies (2.2). Part (a) is proved.

Because $\sum_{k=1}^K D^{d_1} \phi_k^{(opt)}(s) D^{d_2} \xi_k^{(opt)}(t)$ is an optimal representation of $\mathfrak{B}(s, t)$, part (b) follows from Theorem 1 in Luo and Qi (2017).

S1.2 Proof of Theorem 2

For convenience, we introduce some notations. For any $\psi, \phi \in L^2[0, 1]$, let $\langle \phi, \psi \rangle = \int_0^1 \phi(s) \psi(s) ds$ and $\|\phi\| = \sqrt{\langle \phi, \phi \rangle} = \sqrt{\int_0^1 \phi(s)^2 ds}$ denote the inner product of ϕ and ψ and the L^2 -norm of ϕ , respectively. Recall that $\|\cdot\|_2$ denotes the l_2 norm of a vector. For any square integrable kernel function $\mathbf{K}(s, s')$, where $0 \leq s, s' \leq 1$, we can view it as an integral operator in $L^2[0, 1]$, still denoted by \mathbf{K} , such that for any

$\phi \in L^2[0, 1]$, $(\mathbf{K}\phi)(s) = \int_0^1 \mathbf{K}(s, s')\phi(s')ds'$ is a function in $L^2[0, 1]$. Then $\int_0^1 \int_0^1 \phi(s)\mathbf{K}(s, s')\psi(s')dsds' = \langle \phi, \mathbf{K}\psi \rangle$. Moreover, if the function $\mathbf{K}(s', s)$ is symmetric, i.e., $\mathbf{K}(s, s') = \mathbf{K}(s', s)$, then the operator \mathbf{K} is also symmetric in the sense that $\langle \phi, \mathbf{K}\psi \rangle = \langle \mathbf{K}\phi, \psi \rangle = \langle \psi, \mathbf{K}\phi \rangle$. With these notations, the optimization problem (2.2) in Theorem 1 can be expressed as

$$\begin{aligned} & \max_{\phi} \quad \langle D^{d_1}\phi, \mathbf{B}D^{d_1}\phi \rangle \\ & \text{subject to} \quad \langle D^{d_1}\phi, \mathbf{\Sigma}D^{d_1}\phi \rangle = 1, \quad \langle D^{d_1}\phi, \mathbf{\Sigma}D^{d_1}\phi_l^{(opt)} \rangle = 0, \quad \text{for all } 1 \leq l \leq k-1, \end{aligned}$$

which, for convenience, is rewritten as

$$\begin{aligned} & \max_{\phi} \quad \frac{\langle D^{d_1}\phi, \mathbf{B}D^{d_1}\phi \rangle}{\langle D^{d_1}\phi, \mathbf{\Sigma}D^{d_1}\phi \rangle}, \tag{S1.3} \\ & \text{subject to} \quad \langle D^{d_1}\phi, \mathbf{\Sigma}D^{d_1}\phi \rangle = 1, \quad \langle D^{d_1}\phi, \mathbf{\Sigma}D^{d_1}\phi_l^{(opt)} \rangle = 0, \quad \text{for all } 1 \leq l \leq k-1. \end{aligned}$$

When d_1 is a positive integer, the solution to (S1.3) is not unique. We choose a solution to (S1.3), denoted by $\phi_k^{(opt)}(s)$, $k \geq 1$, and fix it in this proof. Let σ_k^2 be the maximum value of the objective function in (S1.3) for any $k \geq 1$. Similarly, the optimization problem (2.5) can be written as

$$\begin{aligned} & \max_{\phi} \quad \frac{\langle D^{d_1}\phi, \widehat{\mathbf{B}}D^{d_1}\phi \rangle}{\langle D^{d_1}\phi, \widehat{\mathbf{\Sigma}}D^{d_1}\phi \rangle + \lambda\|\phi\|^2 + \lambda\tau\|D^2\phi\|^2} \tag{S1.4} \\ & \text{subject to} \quad \langle D^{d_1}\phi, \widehat{\mathbf{\Sigma}}D^{d_1}\phi \rangle = 1, \quad \langle D^{d_1}\phi, \widehat{\mathbf{\Sigma}}D^{d_1}\widehat{\phi}_l \rangle = 0, \quad \text{for all } 1 \leq l \leq k-1. \end{aligned}$$

Let $\widehat{\phi}_k$ denote the solution to (S1.4) and $\widehat{\sigma}_k^2$ denote the maximum value. Because $\lambda > 0$, the solution to (S1.4) is unique up to a sign as long as $\widehat{\sigma}_k^2$ is positive. We choose the sign of $\widehat{\phi}_k$ such that

$$\langle D^{d_1}\widehat{\phi}_k, \mathbf{\Sigma}D^{d_1}\phi_k^{(opt)} \rangle \geq 0. \tag{S1.5}$$

We next provide some inequalities which will be used in our proof. The following lemma is Lemma A.1 in the supplementary material of Luo and Qi (2017) and is provided here for readers' convenience.

Lemma 1. *Assume that $E[\|X\|^4] < \infty$ and $E[\|\varepsilon\|^2] < \infty$. Then given any $\epsilon > 0$, for each n , there exists an event $\Omega_{n,\epsilon}$ with $P(\Omega_{n,\epsilon}) > 1 - \epsilon$, such that in $\Omega_{n,\epsilon}$, for any $\psi, \phi \in L^2[0, 1]$, we have*

$$\begin{aligned} & \left| \langle \psi, (\widehat{\mathbf{B}} - \mathbf{B})\phi \rangle \right| \leq \frac{c_0}{\sqrt{n}} \|\psi\| \|\phi\|, \quad \left| \langle \psi, (\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})\phi \rangle \right| \leq \frac{c_0}{\sqrt{n}} \|\psi\| \|\phi\|, \quad \|\bar{X}\| \leq \frac{c_0}{\sqrt{n}}, \\ & \|\bar{Y} - \mathfrak{U}\| \leq \frac{c_0}{\sqrt{n}}, \quad \left\| \frac{1}{n} \sum_{i=1}^n \int_0^1 \psi(s)[X_i(s) - \bar{X}(s)]ds \{\varepsilon_i - \bar{\varepsilon}\} \right\| \leq \frac{c_0}{\sqrt{n}} \|\psi\|, \tag{S1.6} \end{aligned}$$

where the constant c_0 depends on ϵ but does not depend on n .

In the rest of this proof, all the calculations are performed in the event $\Omega_{n,\epsilon}$. Hence, the inequalities in Lemma 1 always hold. The following equalities follow from the constraints in (S1.3) and (S1.4):

$$\begin{aligned} \langle D^{d_1} \phi_k^{(opt)}, \Sigma D^{d_1} \phi_k^{(opt)} \rangle &= 1, & \langle D_k^{d_1} \phi^{(opt)}, \Sigma D_l^{d_1} \phi^{(opt)} \rangle &= 0, \\ \langle D^{d_1} \hat{\phi}_k, \widehat{\Sigma} D^{d_1} \hat{\phi}_k \rangle &= 1, & \langle D^{d_1} \hat{\phi}_k, \widehat{\Sigma} D^{d_1} \hat{\phi}_l \rangle &= 0, \quad \forall k \neq l. \end{aligned} \quad (\text{S1.7})$$

By the form of (S1.3), $D^{d_1} \phi_k^{(opt)}$ is indeed the generalized eigengfunction of \mathbf{B} , so we have

$$\mathbf{B} D^{d_1} \phi_k^{(opt)} = \sigma_k^2 \Sigma D^{d_1} \phi_k^{(opt)}, \quad (\text{S1.8})$$

for any $k \geq 1$. We define two vectors $\mathbf{z}_k = (z_{1k}, \dots, z_{nk})^\top$ and $\widehat{\mathbf{z}}_k = (\widehat{z}_{1k}, \dots, \widehat{z}_{nk})^\top$, where

$$z_{ik} = \int_0^1 X_i(s) D^{d_1} \phi_k^{(opt)}(s) ds, \quad \widehat{z}_{ik} = \int_0^1 [X_i(s) - \bar{X}(s)] D^{d_1} \hat{\phi}_k(s) ds. \quad (\text{S1.9})$$

The key step of our proof is to prove the following claim by induction.

Claim: In $\Omega_{n,\epsilon}$, for any $1 \leq k \leq K$, we have

$$\begin{aligned} \|\hat{\phi}_k\|^2 &\leq H_{k,1}, & \|D^2 \hat{\phi}_k\|^2 &\leq H_{k,2}, & \|D^{d_1} \hat{\phi}_k\|^2 &\leq H_{k,d_1}, \\ |\hat{\sigma}_k^2 - \sigma_k^2| &\leq \frac{H_{k,3}}{\sqrt{n}}, & \frac{1}{n} \|\widehat{\mathbf{z}}_k - \mathbf{z}_k\|_2^2 &\leq \frac{H_{k,4}}{\sqrt{n}}, & \|D^{d_2} \hat{\xi}_k - D^{d_2} \xi_k^{(opt)}\|^2 &\leq \frac{H_{k,5}}{\sqrt{n}}, \\ \langle (D^{d_1} \hat{\phi}_k - D^{d_1} \phi_k^{(opt)}), \widehat{\Sigma} (D^{d_1} \hat{\phi}_k - D^{d_1} \phi_k^{(opt)}) \rangle &\leq \frac{H_{k,6}}{\sqrt{n}}, \\ \langle (D^{d_1} \hat{\phi}_k - D^{d_1} \phi_k^{(opt)}), \Sigma (D^{d_1} \hat{\phi}_k - D^{d_1} \phi_k^{(opt)}) \rangle &\leq \frac{H_{k,7}}{\sqrt{n}}, \end{aligned} \quad (\text{S1.10})$$

where $H_{k,1} \sim H_{k,7}$ are constants only depending on $\epsilon, C, C_\kappa, c_\tau, C_\tau, \sigma_k^2, \|\phi_k^{(opt)}\|, \|D^2 \phi_k^{(opt)}\|, \|\xi_k^{(opt)}\|$ and $\|D^2 \xi_k^{(opt)}\|$, $1 \leq k \leq K$, but not on n .

We provide the detailed proof for the claim when $k = 1$. The induction step for $k > 1$ is similar and more tedious. So we skip it.

• Prove the claim for $k = 1$.

We split the proof into several steps.

Step 1: Provide upper bounds for $\|\hat{\phi}_1\|, \|D^{d_1} \hat{\phi}_1\|$ and $\|D^2 \hat{\phi}_1\|$.

We notice that the objective function in (S1.3) is scale-invariant in the sense that if we replace $\phi(s)$ by $c\phi(s)$, where c is any scalar constant, the value of the objective function in (S1.3) is unchanged.

Since $\phi_1^{(opt)}$ is a solution to (S1.3) with $k = 1$, we have

$$\frac{\langle D^{d_1} \widehat{\phi}_1, \mathbf{B} D^{d_1} \widehat{\phi}_1 \rangle}{\langle D^{d_1} \widehat{\phi}_1, \widehat{\Sigma} D^{d_1} \widehat{\phi}_1 \rangle} \leq \frac{\langle D^{d_1} \phi_1^{(opt)}, \mathbf{B} D^{d_1} \phi_1^{(opt)} \rangle}{\langle D^{d_1} \phi_1^{(opt)}, \widehat{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle} = \sigma_1^2, \quad (\text{S1.11})$$

which implies that

$$\langle D^{d_1} \widehat{\phi}_1, \mathbf{B} D^{d_1} \widehat{\phi}_1 \rangle \leq \sigma_1^2 \langle D^{d_1} \widehat{\phi}_1, \widehat{\Sigma} D^{d_1} \widehat{\phi}_1 \rangle. \quad (\text{S1.12})$$

Similarly, because $\widehat{\phi}_1$ is a solution to (S1.4) with $k = 1$, we have

$$\begin{aligned} & \frac{\langle D^{d_1} \phi_1^{(opt)}, \widehat{\mathbf{B}} D^{d_1} \phi_1^{(opt)} \rangle}{\langle D^{d_1} \phi_1^{(opt)}, \widehat{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2} \\ & \leq \widehat{\sigma}_1^2 = \frac{\langle D^{d_1} \widehat{\phi}_1, \widehat{\mathbf{B}} D^{d_1} \widehat{\phi}_1 \rangle}{\langle D^{d_1} \widehat{\phi}_1, \widehat{\Sigma} D^{d_1} \widehat{\phi}_1 \rangle + \lambda \|\widehat{\phi}_1\|^2 + \lambda \tau \|D^2 \widehat{\phi}_1\|^2} \\ & = \frac{\langle D^{d_1} \widehat{\phi}_1, \mathbf{B} D^{d_1} \widehat{\phi}_1 \rangle + \langle D^{d_1} \widehat{\phi}_1, (\widehat{\mathbf{B}} - \mathbf{B}) D^{d_1} \widehat{\phi}_1 \rangle}{1 + \lambda \|\widehat{\phi}_1\|^2 + \lambda \tau \|D^2 \widehat{\phi}_1\|^2} \quad (\text{due to the third equality in (S1.7)}) \\ & \leq \frac{\sigma_1^2 \langle D^{d_1} \widehat{\phi}_1, \widehat{\Sigma} D^{d_1} \widehat{\phi}_1 \rangle + \frac{c_0}{\sqrt{n}} \|D^{d_1} \widehat{\phi}_1\|^2}{1 + \lambda \|\widehat{\phi}_1\|^2 + \lambda \tau \|D^2 \widehat{\phi}_1\|^2} \quad (\text{due to (S1.12) and Lemma 1}) \\ & = \frac{\sigma_1^2 \langle D^{d_1} \widehat{\phi}_1, \widehat{\Sigma} D^{d_1} \widehat{\phi}_1 \rangle + \sigma_1^2 \langle D^{d_1} \widehat{\phi}_1, (\widehat{\Sigma} - \Sigma) D^{d_1} \widehat{\phi}_1 \rangle + \frac{c_0}{\sqrt{n}} \|D^{d_1} \widehat{\phi}_1\|^2}{1 + \lambda \|\widehat{\phi}_1\|^2 + \lambda \tau \|D^2 \widehat{\phi}_1\|^2} \\ & \leq \frac{\sigma_1^2 + \frac{\sigma_1^2 c_0}{\sqrt{n}} \|D^{d_1} \widehat{\phi}_1\|^2 + \frac{c_0}{\sqrt{n}} \|D^{d_1} \widehat{\phi}_1\|^2}{1 + \lambda \|\widehat{\phi}_1\|^2 + \lambda \tau \|D^2 \widehat{\phi}_1\|^2} \quad (\text{due to the first equality in (S1.7) and Lemma 1}) \\ & = \frac{\sigma_1^2 + \frac{(1 + \sigma_1^2) c_0}{\sqrt{n}} \|D^{d_1} \widehat{\phi}_1\|^2}{1 + \lambda \|\widehat{\phi}_1\|^2 + \lambda \tau \|D^2 \widehat{\phi}_1\|^2}. \end{aligned} \quad (\text{S1.13})$$

On the other hand,

$$\begin{aligned} & \frac{\langle D^{d_1} \phi_1^{(opt)}, \widehat{\mathbf{B}} D^{d_1} \phi_1^{(opt)} \rangle}{\langle D^{d_1} \phi_1^{(opt)}, \widehat{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2} \\ & = \frac{\langle D^{d_1} \phi_1^{(opt)}, \mathbf{B} D^{d_1} \phi_1^{(opt)} \rangle + \langle D^{d_1} \phi_1^{(opt)}, (\widehat{\mathbf{B}} - \mathbf{B}) D^{d_1} \phi_1^{(opt)} \rangle}{\langle D^{d_1} \phi_1^{(opt)}, \widehat{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle + \langle D^{d_1} \phi_1^{(opt)}, (\widehat{\Sigma} - \Sigma) D^{d_1} \phi_1^{(opt)} \rangle + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2} \\ & \geq \frac{\langle D^{d_1} \phi_1^{(opt)}, \mathbf{B} D^{d_1} \phi_1^{(opt)} \rangle - \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2}{\langle D^{d_1} \phi_1^{(opt)}, \widehat{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle + \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2 + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2} \quad (\text{due to Lemma 1}) \\ & = \frac{\sigma_1^2 - \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2}{1 + \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2 + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2}, \quad (\text{S1.14}) \end{aligned}$$

where the last equality is due to the first constraint in (S1.6) and $\phi_1^{(opt)}$ is its solution. Combining (S1.13) and (S1.14), we obtain the following inequality,

$$\frac{\sigma_1^2 - \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2}{1 + \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2 + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2} \leq \frac{\sigma_1^2 + \frac{(1+\sigma_1^2)c_0}{\sqrt{n}} \|D^{d_1} \widehat{\phi}_1\|^2}{1 + \lambda \|\widehat{\phi}_1\|^2 + \lambda \tau \|D^2 \widehat{\phi}_1\|^2}. \quad (\text{S1.15})$$

Rearranging it, we obtain the following equality,

$$\begin{aligned} & \|\widehat{\phi}_1\|^2 \left[\lambda \sigma_1^2 - \frac{c_0}{\sqrt{n}} \lambda \|D^{d_1} \phi_1^{(opt)}\|^2 \right] + (\sigma_1^2 - \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2) \lambda \tau \|D^2 \widehat{\phi}_1\|^2 \\ & - \frac{(1 + \sigma_1^2)c_0}{\sqrt{n}} \|D^{d_1} \widehat{\phi}_1\|^2 \left[1 + \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2 + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2 \right] \\ & \leq \frac{(1 + \sigma_1^2)c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2 + \sigma_1^2 \lambda \|\phi_1^{(opt)}\|^2 + \sigma_1^2 \lambda \tau \|D^2 \phi_1^{(opt)}\|^2. \end{aligned} \quad (\text{S1.16})$$

The following lemma follows from the Gagliardo–Nirenberg interpolation inequality in a bounded region.

Lemma 2. *If $\varphi(s)$ is a function in $[0, 1]$ with continuous second derivative, then we have*

$$\|D^{d_1} \varphi\|^2 \leq c_1 \|\varphi\|^2 + c_2 \|D^2 \varphi\|^2, \quad (\text{S1.17})$$

for any $0 \leq d_1 \leq 2$, where c_1 and c_2 are two positive constants not depending on $\varphi(s)$.

By (S1.17), the left side of (S1.16) can be bounded from below as follows

$$\begin{aligned} & \|\widehat{\phi}_1\|^2 \left[\lambda \sigma_1^2 - \frac{c_0}{\sqrt{n}} \lambda \|D^{d_1} \phi_1^{(opt)}\|^2 \right] + (\sigma_1^2 - \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2) \lambda \tau \|D^2 \widehat{\phi}_1\|^2 \\ & - \frac{(1 + \sigma_1^2)c_0}{\sqrt{n}} \|D^{d_1} \widehat{\phi}_1\|^2 \left[1 + \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2 + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2 \right] \\ & \geq \|\widehat{\phi}_1\|^2 \left[\lambda \sigma_1^2 - \frac{c_0}{\sqrt{n}} \lambda \|D^{d_1} \phi_1^{(opt)}\|^2 \right] + (\sigma_1^2 - \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2) \lambda \tau \|D^2 \widehat{\phi}_1\|^2 \\ & - \frac{(1 + \sigma_1^2)c_0}{\sqrt{n}} \left(c_1 \|\widehat{\phi}_1\|^2 + c_2 \|D^2 \widehat{\phi}_1\|^2 \right) \left[1 + \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2 + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2 \right] \\ & = \|\widehat{\phi}_1\|^2 \left\{ \lambda \sigma_1^2 - \frac{c_0}{\sqrt{n}} \lambda \|D^{d_1} \phi_1^{(opt)}\|^2 - \frac{(1 + \sigma_1^2)c_0 c_1}{\sqrt{n}} \left[1 + \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2 + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2 \right] \right\} \\ & + \|D^2 \widehat{\phi}_1\|^2 \left\{ \lambda \tau \sigma_1^2 - \frac{c_0}{\sqrt{n}} \lambda \tau \|D^{d_1} \phi_1^{(opt)}\|^2 - \frac{(1 + \sigma_1^2)c_0 c_2}{\sqrt{n}} \left[1 + \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2 + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2 \right] \right\}. \end{aligned} \quad (\text{S1.18})$$

Combining (S1.16) and (S1.18), we obtain

$$\begin{aligned}
& \|\widehat{\phi}_1\|^2 \left\{ \lambda\sigma_1^2 - \frac{c_0}{\sqrt{n}}\lambda\|D^{d_1}\phi_1^{(opt)}\|^2 - \frac{(1+\sigma_1^2)c_0c_1}{\sqrt{n}} \left[1 + \frac{c_0}{\sqrt{n}}\|D^{d_1}\phi_1^{(opt)}\|^2 + \lambda\|\phi_1^{(opt)}\|^2 + \lambda\tau\|D^2\phi_1^{(opt)}\|^2 \right] \right\} \\
& + \|D^2\widehat{\phi}_1\|^2 \left\{ \lambda\tau\sigma_1^2 - \frac{c_0}{\sqrt{n}}\lambda\tau\|D^{d_1}\phi_1^{(opt)}\|^2 - \frac{(1+\sigma_1^2)c_0c_2}{\sqrt{n}} \left[1 + \frac{c_0}{\sqrt{n}}\|D^{d_1}\phi_1^{(opt)}\|^2 + \lambda\|\phi_1^{(opt)}\|^2 + \lambda\tau\|D^2\phi_1^{(opt)}\|^2 \right] \right\} \\
& \leq \sigma_1^2 \frac{c_0}{\sqrt{n}}\|D^{d_1}\phi_1^{(opt)}\|^2 + \sigma_1^2\lambda\|\phi_1^{(opt)}\|^2 + \sigma_1^2\lambda\tau\|D^2\phi_1^{(opt)}\|^2 \\
& \leq \sigma_1^2 \left[\frac{c_0c_1}{\sqrt{n}} + \lambda \right] \|\phi_1^{(opt)}\|^2 + \sigma_1^2 \left[\frac{c_0c_2}{\sqrt{n}} + \lambda\tau \right] \|D^2\phi_1^{(opt)}\|^2, \tag{S1.19}
\end{aligned}$$

where the last inequality follows from Lemma 1. Set $\lambda = C/\sqrt{n}$ and $c_\tau \leq \tau \leq C_\tau$, where C , c_τ and C_τ are constants not depending on n , and require that

$$C > (1 + \sigma_1^2)c_0c_2/\sigma_1^2, \quad C > (1 + \sigma_1^2)c_0c_2/(\tau\sigma_1^2).$$

Then when n is large enough, the terms in the two curly brackets on the left side of (S1.19) are both positive and have order $1/\sqrt{n}$, and the right hand side is equal to $1/\sqrt{n}$ multiplied by a constant as $\phi_1^{(opt)}(s)$ is a fixed function. Hence we obtain the first two inequalities below in (S1.20), and the third inequality in (S1.20) follows from the first two and Lemma 2:

$$\|\widehat{\phi}_1\|^2 \leq H_{1,1}, \quad \|D^2\widehat{\phi}_1\|^2 \leq H_{1,2}, \quad \|D^{d_1}\widehat{\phi}_1\|^2 \leq H_{1,d_1}, \tag{S1.20}$$

where $H_{1,1}$, $H_{1,2}$ and H_{1,d_1} are all constants only depending on $c_0, C, c_\tau, C_\tau, \phi_1^{(opt)}(s), D^{d_1}\phi_1^{(opt)}$ and not depending on n . In the following, we will use $C_{21}, C_{22}, C_2, C_{31}, \dots$, to denote constants only depending on $c_0, C, c_\tau, C_\tau, \phi_1^{(opt)}(s), D^{d_1}\phi_1^{(opt)}$ and not depending on n .

Before we move on to the next step, let us take a close look at the constant $H_{1,2}$. By (S1.19), we have

$$H_{1,2} = \frac{\sigma_1^2 \left[\frac{c_0c_1}{\sqrt{n}} + \lambda \right] \|\phi_1^{(opt)}\|^2 + \sigma_1^2 \left[\frac{c_0c_2}{\sqrt{n}} + \lambda\tau \right] \|D^2\phi_1^{(opt)}\|^2}{\lambda\tau\sigma_1^2 - \frac{c_0}{\sqrt{n}}\lambda\tau\|D^{d_1}\phi_1^{(opt)}\|^2 - \frac{(1+\sigma_1^2)c_0c_2}{\sqrt{n}} \left[1 + \frac{c_0}{\sqrt{n}}\|D^{d_1}\phi_1^{(opt)}\|^2 + \lambda\|\phi_1^{(opt)}\|^2 + \lambda\tau\|D^2\phi_1^{(opt)}\|^2 \right]}.$$

Since λ has the order $1/\sqrt{n}$, when n is large, the denominator of the right hand side of the above equation is dominated by the first term $\lambda\tau\sigma_1^2 - (1+\sigma_1^2)c_0c_2/\sqrt{n}$. Because $\lambda = C/\sqrt{n}$ and $c_\tau \leq \tau \leq C_\tau$, if we require the lower bound c_τ of τ to satisfy $c_\tau \geq 2(1 + \sigma_1^2)c_0c_2/(C\sigma_1^2)$, then we have $\lambda\tau\sigma_1^2 - (1 + \sigma_1^2)c_0c_2/\sqrt{n} \geq \lambda\tau\sigma_1^2/2$. Hence, when n is large and $c_\tau \geq 2(1 + \sigma_1^2)c_0c_2/(C\sigma_1^2)$, we have

$$\begin{aligned}
\|D^2\widehat{\phi}_1\|^2 \leq H_{1,2} & \approx \frac{\sigma_1^2 \left[\frac{c_0c_1}{\sqrt{n}} + \lambda \right] \|\phi_1^{(opt)}\|^2 + \sigma_1^2 \left[\frac{c_0c_2}{\sqrt{n}} + \lambda\tau \right] \|D^2\phi_1^{(opt)}\|^2}{\lambda\tau\sigma_1^2 - \frac{(1+\sigma_1^2)c_0c_2}{\sqrt{n}}} \\
& \leq \frac{1}{\tau} \left[\frac{c_0c_1}{C} + 1 \right] \|\phi_1^{(opt)}\|^2 + \left[\frac{c_0c_2}{\tau C} + 1 \right] \|D^2\phi_1^{(opt)}\|^2, \tag{S1.21}
\end{aligned}$$

which are used in Section 2.3 to explain why our method is not apt to over-smoothing.

Step 2: Provide an upper bound for $|\hat{\sigma}_1^2 - \sigma_1^2|$.

By the first two lines in (S1.13), we have

$$\hat{\sigma}_1^2 = \frac{\langle D^{d_1} \hat{\phi}_1, \hat{\mathbf{B}} D^{d_1} \hat{\phi}_1 \rangle}{\langle D^{d_1} \hat{\phi}_1, \hat{\Sigma} D^{d_1} \hat{\phi}_1 \rangle + \lambda \|\hat{\phi}_1\|^2 + \lambda \tau \|D^2 \hat{\phi}_1\|^2} \quad (\text{S1.22})$$

$$\begin{aligned} &\geq \frac{\langle D^{d_1} \phi_1^{(opt)}, \hat{\mathbf{B}} D^{d_1} \phi_1^{(opt)} \rangle}{\langle D^{d_1} \phi_1^{(opt)}, \hat{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2} \\ &\geq \frac{\sigma_1^2 - \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2}{1 + \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2 + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2} \quad (\text{due to (S1.14)}) \\ &= \sigma_1^2 - \frac{\frac{(1+\sigma_1^2)c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2 + \sigma_1^2 \lambda \|\phi_1^{(opt)}\|^2 + \sigma_1^2 \lambda \tau \|D^2 \phi_1^{(opt)}\|^2}{1 + \frac{c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2 + \lambda \|\phi_1^{(opt)}\|^2 + \lambda \tau \|D^2 \phi_1^{(opt)}\|^2} \\ &\geq \sigma_1^2 - \left(\frac{(1+\sigma_1^2)c_0}{\sqrt{n}} \|D^{d_1} \phi_1^{(opt)}\|^2 + \sigma_1^2 \lambda \|\phi_1^{(opt)}\|^2 + \sigma_1^2 \lambda \tau \|D^2 \phi_1^{(opt)}\|^2 \right) \geq \sigma_1^2 - \frac{C_{21}}{\sqrt{n}}, \end{aligned} \quad (\text{S1.23})$$

where the last inequality follows from $\lambda = C/\sqrt{n}$ and $c_\tau \leq \tau \leq C_\tau$, and the constant $C_{21} = (1 + \sigma_1^2)c_0 \|D^{d_1} \phi_1^{(opt)}\|^2 + \sigma_1^2 C \|\phi_1^{(opt)}\|^2 + \sigma_1^2 C C_\tau \|D^2 \phi_1^{(opt)}\|^2$.

Conversely, by (S1.11),

$$\begin{aligned} \sigma_1^2 &= \frac{\langle D^{d_1} \phi_1^{(opt)}, \mathbf{B} D^{d_1} \phi_1^{(opt)} \rangle}{\langle D^{d_1} \phi_1^{(opt)}, \Sigma D^{d_1} \phi_1^{(opt)} \rangle} \geq \frac{\langle D^{d_1} \hat{\phi}_1, \mathbf{B} D^{d_1} \hat{\phi}_1 \rangle}{\langle D^{d_1} \hat{\phi}_1, \Sigma D^{d_1} \hat{\phi}_1 \rangle} = \frac{\langle D^{d_1} \hat{\phi}_1, \hat{\mathbf{B}} D^{d_1} \hat{\phi}_1 \rangle - \langle D^{d_1} \hat{\phi}_1, (\hat{\mathbf{B}} - \mathbf{B}) D^{d_1} \hat{\phi}_1 \rangle}{\langle D^{d_1} \hat{\phi}_1, \hat{\Sigma} D^{d_1} \hat{\phi}_1 \rangle - \langle D^{d_1} \hat{\phi}_1, (\hat{\Sigma} - \Sigma) D^{d_1} \hat{\phi}_1 \rangle} \\ &= \frac{\hat{\sigma}_1^2 [\langle D^{d_1} \hat{\phi}_1, \hat{\Sigma} D^{d_1} \hat{\phi}_1 \rangle + \lambda \|\hat{\phi}_1\|^2 + \lambda \tau \|D^2 \hat{\phi}_1\|^2] - \langle D^{d_1} \hat{\phi}_1, (\hat{\mathbf{B}} - \mathbf{B}) D^{d_1} \hat{\phi}_1 \rangle}{\langle D^{d_1} \hat{\phi}_1, \hat{\Sigma} D^{d_1} \hat{\phi}_1 \rangle - \langle D^{d_1} \hat{\phi}_1, (\hat{\Sigma} - \Sigma) D^{d_1} \hat{\phi}_1 \rangle} \\ &\quad (\text{due to the first line in (S1.22)}) \\ &= \frac{\hat{\sigma}_1^2 [1 + \lambda \|\hat{\phi}_1\|^2 + \lambda \tau \|D^2 \hat{\phi}_1\|^2] - \langle D^{d_1} \hat{\phi}_1, (\hat{\mathbf{B}} - \mathbf{B}) D^{d_1} \hat{\phi}_1 \rangle}{1 - \langle D^{d_1} \hat{\phi}_1, (\hat{\Sigma} - \Sigma) D^{d_1} \hat{\phi}_1 \rangle} \quad (\text{because } \langle D^{d_1} \hat{\phi}_1, \hat{\Sigma} D^{d_1} \hat{\phi}_1 \rangle = 1) \\ &\geq \frac{\hat{\sigma}_1^2 [1 + \lambda \|\hat{\phi}_1\|^2 + \lambda c_\tau \|D^{d_1} \hat{\phi}_1\|^2] - \frac{c_0}{\sqrt{n}} \|D^{d_1} \hat{\phi}_1\|^2}{1 + \frac{c_0}{\sqrt{n}} \|D^{d_1} \hat{\phi}_1\|^2} \quad (\text{due to Lemma 1}) \\ &\geq \frac{\hat{\sigma}_1^2 - \frac{c_0}{\sqrt{n}} \|D^{d_1} \hat{\phi}_1\|^2}{1 + \frac{c_0}{\sqrt{n}} \|D^{d_1} \hat{\phi}_1\|^2} \geq \hat{\sigma}_1^2 (1 - \frac{C_{22}}{\sqrt{n}}) - \frac{C_{23}}{\sqrt{n}}, \quad (\text{due to (S1.20)}), \end{aligned} \quad (\text{S1.24})$$

where C_{22} and C_{23} are constants. By (S1.24), as n is large enough, we have

$$\hat{\sigma}_1^2 - \sigma_1^2 \leq \frac{C_{22}\sigma_1^2 + C_{23}}{\sqrt{n} - C_{22}} \leq \frac{2C_{22}\sigma_1^2 + 2C_{23}}{\sqrt{n}},$$

which, together with (S1.23), gives

$$|\hat{\sigma}_1^2 - \sigma_1^2| \leq \frac{H_{1,3}}{\sqrt{n}}, \quad (\text{S1.25})$$

where $H_{1,3} = \max\{C_{21}, 2C_{22}\sigma_1^2 + 2C_{23}\}$.

Step 3: Provide upper bounds for $\langle (D^{d_1}\hat{\phi}_1 - D^{d_1}\phi_1^{(opt)}), \Sigma(D^{d_1}\hat{\phi}_1 - D^{d_1}\phi_1^{(opt)}) \rangle$ and $\langle (D^{d_1}\hat{\phi}_1 - D^{d_1}\phi_1^{(opt)}), \hat{\Sigma}(D^{d_1}\hat{\phi}_1 - D^{d_1}\phi_1^{(opt)}) \rangle$.

Let $\tilde{\psi}_1 = D^{d_1}\hat{\phi}_1 - \langle D^{d_1}\hat{\phi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle D^{d_1}\phi_1^{(opt)}$. Then

$$D^{d_1}\hat{\phi}_1 = \langle D^{d_1}\hat{\phi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle D^{d_1}\phi_1^{(opt)} + \tilde{\psi}_1, \quad (\text{S1.26})$$

where $\tilde{\psi}_1(s)$ satisfies $\langle \tilde{\psi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle = 0$ as $\langle D^{d_1}\hat{\phi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle = 1$. Taking the inner products of both sides of (S1.26) with $\Sigma D^{d_1}\hat{\phi}_1$, we have

$$\langle D^{d_1}\hat{\phi}_1, \Sigma D^{d_1}\hat{\phi}_1 \rangle = \langle D^{d_1}\hat{\phi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle^2 + \langle \tilde{\psi}_1, \Sigma D^{d_1}\hat{\phi}_1 \rangle. \quad (\text{S1.27})$$

Now we calculate

$$\begin{aligned} & \langle D^{d_1}\hat{\phi}_1, \mathbf{B}D^{d_1}\hat{\phi}_1 \rangle \quad (\text{S1.28}) \\ &= \langle D^{d_1}\hat{\phi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle^2 \langle D^{d_1}\phi_1^{(opt)}, \mathbf{B}D^{d_1}\phi_1^{(opt)} \rangle + 2\langle D^{d_1}\hat{\phi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle \langle \tilde{\psi}_1, \mathbf{B}D^{d_1}\phi_1^{(opt)} \rangle + \langle \tilde{\psi}_1, \mathbf{B}\tilde{\psi}_1 \rangle \\ & \quad (\text{replacing } D^{d_1}\hat{\phi}_1 \text{ in (S1.28) by the right hand side of (S1.26)}) \\ &= \langle D^{d_1}\hat{\phi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle^2 \sigma_1^2 \langle D^{d_1}\phi_1^{(opt)}, \Sigma D^{d_1}\phi_1^{(opt)} \rangle + 2\langle D^{d_1}\hat{\phi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle \sigma_1^2 \langle \tilde{\psi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle + \langle \tilde{\psi}_1, \mathbf{B}\tilde{\psi}_1 \rangle \\ & \quad (\text{due to (S1.8)}) \\ &= \sigma_1^2 \langle D^{d_1}\hat{\phi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle^2 + \langle \tilde{\psi}_1, \mathbf{B}\tilde{\psi}_1 \rangle \leq \sigma_1^2 \langle D^{d_1}\hat{\phi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle^2 + \sigma_2^2 \langle \tilde{\psi}_1, \Sigma \tilde{\psi}_1 \rangle, \quad (\text{S1.29}) \end{aligned}$$

where the first equality in the last line follows from $\langle D^{d_1}\phi_1^{(opt)}, \Sigma D^{d_1}\phi_1^{(opt)} \rangle = 1$ and $\langle \tilde{\psi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle = 0$, and the last inequality is due to $\langle \tilde{\psi}_1, \mathbf{B}\tilde{\psi}_1 \rangle \leq \sigma_2^2 \langle \tilde{\psi}_1, \Sigma \tilde{\psi}_1 \rangle$ which is explained as following. Let $\tilde{\phi}_1$ be a function satisfying $D^{d_1}\tilde{\phi}_1 = \tilde{\psi}_1 / \sqrt{\langle \tilde{\psi}_1, \Sigma \tilde{\psi}_1 \rangle}$. Then we have

$$\langle D^{d_1}\tilde{\phi}_1, \Sigma D^{d_1}\tilde{\phi}_1 \rangle = 1, \quad (\text{S1.30})$$

and by $\langle \tilde{\psi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle = 0$, we have

$$\langle D^{d_1}\tilde{\phi}_1, \Sigma D^{d_1}\phi_1^{(opt)} \rangle = 0. \quad (\text{S1.31})$$

(S1.30) and (S1.31) imply that $\tilde{\phi}_1$ satisfies the constraints in the optimization problem (S1.3) with $k = 2$. Since the maximum value of (S1.3) with $k = 2$ is equal to σ_2^2 , we have

$$\sigma_2^2 \geq \frac{\langle D^{d_1} \tilde{\phi}_1, \mathbf{B} D^{d_1} \tilde{\phi}_1 \rangle}{\langle D^{d_1} \tilde{\phi}_1, \mathbf{\Sigma} D^{d_1} \tilde{\phi}_1 \rangle} = \frac{\langle \tilde{\psi}_1, \mathbf{B} \tilde{\psi}_1 \rangle}{\langle \tilde{\psi}_1, \mathbf{\Sigma} \tilde{\psi}_1 \rangle},$$

which implies that $\langle \tilde{\psi}_1, \mathbf{B} \tilde{\psi}_1 \rangle \leq \sigma_2^2 \langle \tilde{\psi}_1, \mathbf{\Sigma} \tilde{\psi}_1 \rangle$.

On the other hand,

$$\begin{aligned} \langle D^{d_1} \hat{\phi}_1, \mathbf{B} D^{d_1} \hat{\phi}_1 \rangle &= \langle D^{d_1} \hat{\phi}_1, \hat{\mathbf{B}} D^{d_1} \hat{\phi}_1 \rangle - \langle D^{d_1} \hat{\phi}_1, (\hat{\mathbf{B}} - \mathbf{B}) D^{d_1} \hat{\phi}_1 \rangle \\ &= \hat{\sigma}_1^2 \left[\langle D^{d_1} \hat{\phi}_1, \hat{\mathbf{\Sigma}} D^{d_1} \hat{\phi}_1 \rangle + \lambda \|\hat{\phi}_1\|^2 + \lambda \tau \|D^2 \hat{\phi}_1\|^2 \right] - \langle D^{d_1} \hat{\phi}_1, (\hat{\mathbf{B}} - \mathbf{B}) D^{d_1} \hat{\phi}_1 \rangle \quad (\text{due to the first line in (S1.22)}) \\ &= \hat{\sigma}_1^2 \left[1 + \lambda \|\hat{\phi}_1\|^2 + \lambda \tau \|D^2 \hat{\phi}_1\|^2 \right] - \langle D^{d_1} \hat{\phi}_1, (\hat{\mathbf{B}} - \mathbf{B}) D^{d_1} \hat{\phi}_1 \rangle \quad (\text{follows from the third equality in (S1.7)}) \\ &\geq \hat{\sigma}_1^2 - \langle D^{d_1} \hat{\phi}_1, (\hat{\mathbf{B}} - \mathbf{B}) D^{d_1} \hat{\phi}_1 \rangle \geq \sigma_1^2 - \frac{C_{21}}{\sqrt{n}} - \frac{c_0}{\sqrt{n}} \|D^{d_1} \hat{\phi}_1\|^2 \quad (\text{due to (S1.23) and Lemma 1}) \\ &\geq \sigma_1^2 - \frac{C_{31}}{\sqrt{n}}, \end{aligned} \tag{S1.32}$$

where the last inequality follows from (S1.20) and the constant $C_{31} = C_{21} + c_0 H_{1,d_1}$. Now combining (S1.29) and (S1.32), we have

$$\begin{aligned} \sigma_1^2 - \frac{C_{31}}{\sqrt{n}} &\leq \langle D^{d_1} \hat{\phi}_1, \mathbf{B} D^{d_1} \hat{\phi}_1 \rangle \leq \sigma_1^2 \langle D^{d_1} \hat{\phi}_1, \mathbf{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle^2 + \sigma_2^2 \langle \tilde{\psi}_1, \mathbf{\Sigma} \tilde{\psi}_1 \rangle \\ &= \sigma_1^2 \langle D^{d_1} \hat{\phi}_1, \mathbf{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle^2 + \sigma_2^2 \{ \langle D^{d_1} \hat{\phi}_1, \mathbf{\Sigma} D^{d_1} \hat{\phi}_1 \rangle - \langle D^{d_1} \hat{\phi}_1, \mathbf{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle \}^2 \\ &\quad (\text{by plugging in } \tilde{\psi}_1) \\ &= (\sigma_1^2 - \sigma_2^2) \langle D^{d_1} \hat{\phi}_1, \mathbf{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle^2 + \sigma_2^2 \langle D^{d_1} \hat{\phi}_1, \mathbf{\Sigma} D^{d_1} \hat{\phi}_1 \rangle \\ &= (\sigma_1^2 - \sigma_2^2) \langle D^{d_1} \hat{\phi}_1, \mathbf{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle^2 + \sigma_2^2 \langle D^{d_1} \hat{\phi}_1, \hat{\mathbf{\Sigma}} D^{d_1} \hat{\phi}_1 \rangle + \sigma_2^2 \langle D^{d_1} \hat{\phi}_1, (\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}) D^{d_1} \hat{\phi}_1 \rangle \\ &\leq (\sigma_1^2 - \sigma_2^2) \langle D^{d_1} \hat{\phi}_1, \mathbf{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle^2 + \sigma_2^2 \langle D^{d_1} \hat{\phi}_1, \hat{\mathbf{\Sigma}} D^{d_1} \hat{\phi}_1 \rangle + \sigma_2^2 \frac{c_0}{\sqrt{n}} \|D^{d_1} \hat{\phi}_1\|^2 \\ &\quad (\text{by Lemma 1}) \\ &= (\sigma_1^2 - \sigma_2^2) \langle D^{d_1} \hat{\phi}_1, \mathbf{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle^2 + \sigma_2^2 + \sigma_2^2 \frac{c_0}{\sqrt{n}} \|D^{d_1} \hat{\phi}_1\|^2 \quad (\text{by (S1.7)}) \\ &\leq (\sigma_1^2 - \sigma_2^2) \langle D^{d_1} \hat{\phi}_1, \mathbf{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle^2 + \sigma_2^2 + \frac{C_{32}}{\sqrt{n}}, \end{aligned} \tag{S1.33}$$

where the last inequality follows from (S1.20) and the constant $C_{32} = \sigma_2^2 c_0 H_{1,d_1}$. By (S1.33), we have $(\sigma_1^2 - \sigma_2^2) \leq (\sigma_1^2 - \sigma_2^2) \langle D^{d_1} \hat{\phi}_1, \mathbf{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle^2 + (C_{31} + C_{32})/\sqrt{n}$, which leads to

$$\langle D^{d_1} \hat{\phi}_1, \mathbf{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle^2 \geq 1 - C_{33}/\sqrt{n}, \quad \text{and hence, } \langle D^{d_1} \hat{\phi}_1, \mathbf{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle \geq 1 - C_{34}/\sqrt{n}, \tag{S1.34}$$

where we use $\langle D^{d_1} \widehat{\phi}_1, \Sigma D^{d_1} \phi_1^{(opt)} \rangle \geq 0$ in (S1.5). It follows from (S1.34) and Lemma 1 that

$$\langle D^{d_1} \widehat{\phi}_1, \widehat{\Sigma} D^{d_1} \phi_1^{(opt)} \rangle = \langle D^{d_1} \widehat{\phi}_1, \Sigma D^{d_1} \phi_1^{(opt)} \rangle - \langle D^{d_1} \widehat{\phi}_1, (\Sigma - \widehat{\Sigma}) D^{d_1} \phi_1^{(opt)} \rangle \geq 1 - C_{35}/\sqrt{n}. \quad (\text{S1.35})$$

Now we calculate

$$\begin{aligned} & \langle (D^{d_1} \widehat{\phi}_1 - D^{d_1} \phi_1^{(opt)}), \Sigma (D^{d_1} \widehat{\phi}_1 - D^{d_1} \phi_1^{(opt)}) \rangle \\ &= \langle D^{d_1} \phi_1^{(opt)}, \Sigma D^{d_1} \phi_1^{(opt)} \rangle - 2 \langle D^{d_1} \widehat{\phi}_1, \Sigma D^{d_1} \phi_1^{(opt)} \rangle + \langle D^{d_1} \widehat{\phi}_1, \Sigma D^{d_1} \widehat{\phi}_1 \rangle \\ &= 1 - 2 \langle D^{d_1} \widehat{\phi}_1, \Sigma D^{d_1} \phi_1^{(opt)} \rangle + \langle D^{d_1} \widehat{\phi}_1, \Sigma D^{d_1} \widehat{\phi}_1 \rangle \quad (\text{due to the first equality in (S1.7)}) \\ &= 1 - 2 \langle D^{d_1} \widehat{\phi}_1, \Sigma D^{d_1} \phi_1^{(opt)} \rangle + \langle D^{d_1} \widehat{\phi}_1, \widehat{\Sigma} D^{d_1} \widehat{\phi}_1 \rangle + \langle D^{d_1} \widehat{\phi}_1, (\Sigma - \widehat{\Sigma}) D^{d_1} \widehat{\phi}_1 \rangle \\ &\leq 1 - 2 \langle D^{d_1} \widehat{\phi}_1, \Sigma D^{d_1} \phi_1^{(opt)} \rangle + 1 + \frac{c_0}{\sqrt{n}} \|D^{d_1} \widehat{\phi}_1\|^2 \quad (\text{by Lemma 1 and the third equality in (S1.7)}) \\ &\leq 2 - 2(1 - C_{35}/\sqrt{n}) + \frac{c_0 H_{1,d_1}}{\sqrt{n}} \quad (\text{due to (S1.34) and (S1.20)}) \\ &= H_{1,7}/\sqrt{n} \end{aligned}$$

where the constant $H_{1,7} = 2C_{35} + c_0 H_{1,d_1}$. Then we have

$$\begin{aligned} & \langle (D^{d_1} \widehat{\phi}_1 - D^{d_1} \phi_1^{(opt)}), \widehat{\Sigma} (D^{d_1} \widehat{\phi}_1 - D^{d_1} \phi_1^{(opt)}) \rangle \\ &\leq \langle (D^{d_1} \widehat{\phi}_1 - D^{d_1} \phi_1^{(opt)}), \Sigma (D^{d_1} \widehat{\phi}_1 - D^{d_1} \phi_1^{(opt)}) \rangle + \langle (D^{d_1} \widehat{\phi}_1 - D^{d_1} \phi_1^{(opt)}), (\widehat{\Sigma} - \Sigma) (D^{d_1} \widehat{\phi}_1 - D^{d_1} \phi_1^{(opt)}) \rangle \\ &\leq \frac{H_{1,7}}{\sqrt{n}} + \frac{c_0}{\sqrt{n}} \|D^{d_1} \widehat{\phi}_1 - D^{d_1} \phi_1^{(opt)}\|^2 \leq \frac{H_{1,7}}{\sqrt{n}} + \frac{c_0}{\sqrt{n}} \{ \|D^{d_1} \widehat{\phi}_1\|^2 + \|D^{d_1} \phi_1^{(opt)}\|^2 \} \leq \frac{H_{1,6}}{\sqrt{n}}, \end{aligned} \quad (\text{S1.36})$$

where $H_{1,6} = H_{1,7} + c_0 H_{1,d_1} + \|D^{d_1} \phi_1^{(opt)}\|^2$.

Step 4: Estimate $\frac{1}{n} \|\widehat{\mathbf{z}}_1 - \mathbf{z}_1\|_2^2$.

In addition to $\widehat{\mathbf{z}}_1$ and \mathbf{z}_1 defined in (S1.9), we define another vector $\widetilde{\mathbf{z}}_1 = (\widetilde{z}_{11}, \dots, \widetilde{z}_{n1})^\top$, where $\widetilde{z}_{ik} = \int_0^1 [X_i(s) - \overline{X}(s)] D^{d_1} \phi_k^{(opt)}(s) ds$. Then we have

$$\frac{1}{n} \|\widehat{\mathbf{z}}_1 - \widetilde{\mathbf{z}}_1\|_2^2 = \langle (D^{d_1} \widehat{\phi}_1 - D^{d_1} \phi_1^{(opt)}), \widehat{\Sigma} (D^{d_1} \widehat{\phi}_1 - D^{d_1} \phi_1^{(opt)}) \rangle \leq H_{1,6}/\sqrt{n}, \quad (\text{S1.37})$$

where the last inequality follows from (S1.36), and

$$\frac{1}{n} \|\mathbf{z}_1 - \widetilde{\mathbf{z}}_1\|_2^2 = \langle \overline{X}, D^{d_1} \phi_1^{(opt)} \rangle^2 \leq \|\overline{X}\|^2 \|D^{d_1} \phi_1^{(opt)}\|^2 \leq c_0 n^{-1} \|D^{d_1} \phi_1^{(opt)}\|^2, \quad (\text{S1.38})$$

where the last inequality follows from (S1.6) in Lemma 1. Combining (S1.37) and (S1.38), we obtain

$$\frac{1}{n} \|\widehat{\mathbf{z}}_1 - \mathbf{z}_1\|_2^2 \leq H_{1,4}/\sqrt{n}. \quad (\text{S1.39})$$

Step 5: Estimate $\|D^{d_2}\widehat{\xi}_1 - D^2\xi_1^{(opt)}\|^2$ **and** $\|D^{d_2}\widehat{\mu} - \mathfrak{U}\|^2$.

Note that $\frac{1}{n}\sum_{i=1}^n \widehat{z}_{ik} = \sum_{i=1}^n \int_0^1 [X_i(s) - \bar{X}(s)] D^{d_1}\widehat{\phi}_k(s) ds = 0$, $\frac{1}{n}\sum_{i=1}^n \widehat{z}_{ik}^2 = \langle D^{d_1}\widehat{\phi}_k, \widehat{\Sigma}D^{d_1}\widehat{\phi}_k \rangle = 1$, $\frac{1}{n}\sum_{i=1}^n \widehat{z}_{ik}\widehat{z}_{il} = \langle D^{d_1}\widehat{\phi}_k, \widehat{\Sigma}D^{d_1}\widehat{\phi}_l \rangle = 0$ for all $k \neq l$, where the second and third equalities follow from the last two equalities in (S1.7). The objective function in the optimization problem (2.6) can be expanded as

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \|Y_i - D^{d_2}\mu - \sum_{k=1}^K \widehat{z}_{ik} D^{d_2}\xi_k\|^2 + \kappa \|D^2\mu\|^2 + \kappa \sum_{k=1}^K \|D^2\xi_k\|^2 \tag{S1.40} \\ &= \frac{1}{n} \sum_{i=1}^n \|Y_i\|^2 + \|D^{d_2}\mu\|^2 + \sum_{k=1}^K \|D^{d_2}\xi_k\|^2 - 2\langle \bar{Y}, D^{d_2}\mu \rangle - 2 \sum_{k=1}^K \langle \widehat{\nu}_k^0, D^{d_2}\xi_k \rangle + \kappa \|D^2\mu\|^2 + \kappa \sum_{k=1}^K \|D^2\xi_k\|^2 \\ &= \frac{1}{n} \sum_{i=1}^n \|Y_i\|^2 + \|D^{d_2}\mu - \bar{Y}\|^2 - \|\bar{Y}\|^2 + \sum_{k=1}^K \|D^{d_2}\xi_k - \widehat{\nu}_k^0\|^2 - \sum_{k=1}^K \|\widehat{\nu}_k^0\|_2^2 + \kappa \|D^2\mu\|^2 + \kappa \sum_{k=1}^K \|D^2\xi_k\|^2, \end{aligned}$$

where $\widehat{\nu}_k^0(t) = \frac{1}{n} \sum_{i=1}^n \widehat{z}_{ik} Y_i(t) = \frac{1}{n} \widehat{\mathbf{z}}_k^T \mathbf{Y}(t)$. So based on the expansion in (S1.40), solving the optimization problem (2.6) is equivalent to solving the following optimization problems separately,

$$\min_{\mu(t)} \left[\|D^{d_2}\mu - \bar{Y}\|^2 + \kappa \|D^2\mu\|^2 \right], \tag{S1.41}$$

$$\min_{\xi_k(t)} \left[\|D^{d_2}\xi_k - \widehat{\nu}_k^0\|^2 + \kappa \|D^2\xi_k\|^2 \right], \quad \text{for } 1 \leq k \leq K. \tag{S1.42}$$

The following inequality follows from Lemma A.4 in supplementary material of Luo and Qi (2017),

$$\|D^{d_2}\xi_1^{(opt)} - \widehat{\nu}_1^0\|^2 \leq \frac{C_{63}}{\sqrt{n}}, \tag{S1.43}$$

because $D^{d_2}\xi_1^{(opt)}$ is the function in the optimal expansion for $\mathfrak{B}(s, t)$, where C_{63} is a constant. Since $\widehat{\xi}_1$ is the solution to (S1.42) with $k = 1$, we have

$$\|D^{d_2}\widehat{\xi}_1 - \widehat{\nu}_1^0\|^2 + \kappa \|D^2\widehat{\xi}_1\|^2 \leq \|D^{d_2}\xi_1^{(opt)} - \widehat{\nu}_1^0\|^2 + \kappa \|D^2\xi_1^{(opt)}\|^2 \leq \frac{C_{63}}{\sqrt{n}} + \kappa \|D^2\xi_1^{(opt)}\|^2, \tag{S1.44}$$

where the last inequality follows from (S1.43). As we choose $\kappa = C_\kappa/\sqrt{n}$, (S1.44) leads to $\|D^{d_2}\widehat{\xi}_1 - \widehat{\nu}_1^0\|^2 \leq C_{64}/\sqrt{n}$, which together with (S1.43) gives

$$\|D^{d_2}\widehat{\xi}_1 - D^{d_2}\xi_1^{(opt)}\|^2 \leq H_{1,5}/\sqrt{n}. \quad \text{and similarly, we can show } \|D^{d_2}\widehat{\mu} - \mathfrak{U}\|^2 \leq H_0/\sqrt{n}. \tag{S1.45}$$

By Steps 1-5, we have proved all the inequalities in the claim (S1.10) for $k = 1$. Then by induction and similar calculation as above, we can show that the equalities in the claim (S1.10) hold for all $1 \leq k \leq K$. We skip the details. The inequalities in (2.8) and (2.9) in the theorem can be obtained from the claim (S1.10) and the second inequality in (S1.45). We next prove the inequality in (2.10).

- Estimate $\frac{1}{n} \int_0^1 \|\widehat{\mathbf{F}}(t) - \mathbf{F}(t)\|_2^2 dt$.

Based on the definitions of $\widehat{\mathbf{z}}_1$ and \mathbf{z}_1 in (S1.9), we have

$$\mathbf{F}(t) = \mathfrak{U}(t)\mathbf{1}_n + \sum_{k=1}^{\infty} \mathbf{z}_k D^{d_2} \xi_k^{(opt)}(t), \quad \widehat{\mathbf{F}}(t) = D^{d_2} \widehat{\mu}(t)\mathbf{1}_n + \sum_{k=1}^K \widehat{\mathbf{z}}_k D^{d_2} \widehat{\xi}_k.$$

Define $\mathbf{F}^K(t) = \mathfrak{U}(t)\mathbf{1}_n + \sum_{k=1}^K \mathbf{z}_k D^{d_2} \xi_k^{(opt)}(t)$ which is the truncation of $\mathbf{F}(t)$ after the first K terms in the series. Then we have

$$\frac{1}{n} \int_0^1 \|\widehat{\mathbf{F}}(t) - \mathbf{F}^K(t)\|_2^2 dt \leq 2\|D^{d_2} \widehat{\mu} - \mathfrak{U}\|^2 + \frac{2}{n} \int_0^1 \left\| \sum_{k=1}^K \widehat{\mathbf{z}}_k D^{d_2} \widehat{\xi}_k - \sum_{k=1}^K \mathbf{z}_k D^{d_2} \xi_k^{(opt)}(t) \right\|_2^2 dt. \quad (\text{S1.46})$$

As

$$\begin{aligned} & \frac{2}{n} \int_0^1 \left\| \sum_{k=1}^K \widehat{\mathbf{z}}_k D^{d_2} \widehat{\xi}_k(t) - \sum_{k=1}^K \mathbf{z}_k D^{d_2} \xi_k^{(opt)}(t) \right\|_2^2 dt \\ & \leq \frac{4}{n} \int_0^1 \left\| \sum_{k=1}^K \widehat{\mathbf{z}}_k \{D^{d_2} \widehat{\xi}_k(t) - D^{d_2} \xi_k^{(opt)}(t)\} \right\|_2^2 dt + \frac{4}{n} \int_0^1 \left\| \sum_{k=1}^K [\widehat{\mathbf{z}}_k - \mathbf{z}_k] D^{d_2} \xi_k^{(opt)}(t) \right\|_2^2 dt \\ & \leq \frac{4}{n} K \sum_{k=1}^K \int_0^1 \left\| \widehat{\mathbf{z}}_k \{D^{d_2} \widehat{\xi}_k(t) - D^{d_2} \xi_k^{(opt)}(t)\} \right\|_2^2 dt + \frac{4}{n} K \sum_{k=1}^K \int_0^1 \left\| [\widehat{\mathbf{z}}_k - \mathbf{z}_k] D^{d_2} \xi_k^{(opt)}(t) \right\|_2^2 dt \\ & = \frac{4K}{n} \sum_{k=1}^K \|\widehat{\mathbf{z}}_k\|_2^2 \|D^{d_2} \widehat{\xi}_k - D^{d_2} \xi_k^{(opt)}\|^2 + \frac{4K}{n} \sum_{k=1}^K \|\widehat{\mathbf{z}}_k - \mathbf{z}_k\|_2^2 \|D^{d_2} \xi_k^{(opt)}\|^2 \\ & = 4K \sum_{k=1}^K \|D^{d_2} \widehat{\xi}_k - D^{d_2} \xi_k^{(opt)}\|^2 + \frac{4K}{n} \sum_{k=1}^K \|\widehat{\mathbf{z}}_k - \mathbf{z}_k\|_2^2 \|D^{d_2} \xi_k^{(opt)}\|^2 \quad (\text{since } \|\widehat{\mathbf{z}}_k\|_2^2 = n) \\ & \leq \left(4K \sum_{k=1}^K H_{k,5} \right) \frac{1}{\sqrt{n}} + \left(4K \sum_{k=1}^K H_{k,4} \|D^{d_2} \xi_k^{(opt)}\|^2 \right) \frac{1}{\sqrt{n}} \quad (\text{by claim (S1.10)}), \end{aligned} \quad (\text{S1.47})$$

combining (S1.46), (S1.45) and (S1.47), we obtain

$$\frac{1}{n} \int_0^1 \|\widehat{\mathbf{F}}(t) - \mathbf{F}^K(t)\|_2^2 dt \leq \frac{M_{K,1}}{\sqrt{n}}. \quad (\text{S1.48})$$

By the inequality after (A.74) in supplementary material of Luo and Qi (2017), we have

$$\frac{1}{n} \int_c^d \|\mathbf{F}(t) - \mathbf{F}^K(t)\|_2^2 dt \leq \frac{M_{K,2}}{\sqrt{n}} + \sum_{k=K+1}^{\infty} \sigma_k^2,$$

which, together with (S1.47), gives

$$\frac{1}{n} \int_0^1 \|\widehat{\mathbf{F}}(t) - \mathbf{F}(t)\|_2^2 dt \leq \frac{M_K}{\sqrt{n}} + \sum_{k=K+1}^{\infty} \sigma_k^2,$$

where the constant $M_K = M_{K,1} + M_{K,2}$.

The inequality for $\mathbf{E} \left[\|\widehat{Y}_{\text{pred}} - Y_{\text{new}}\|^2 \middle| X_i(s), Y_i(t), 1 \leq i \leq n \right]$ can be similarly proved, and we skip the details.

S2 Additional details for computation

S2.1 Solving (2.17)

First we consider the problem

$$\max_{\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^M} \frac{\mathbf{a}^T \Xi \mathbf{a}}{\mathbf{a}^T \mathbf{Q} \mathbf{a}}, \quad \text{subject to } \hat{\mathbf{a}}_l^T \mathbf{H} \mathbf{a} = 0, \quad 1 \leq l \leq k-1, \quad (\text{S2.1})$$

which removes the constraint $\mathbf{a}^T \mathbf{H} \mathbf{a} = 1$ in (2.17). Let $\tilde{\mathbf{a}}$ be a solution to (S2.1). For any nonzero scalar c , $c\tilde{\mathbf{a}}$ is also a solution to (S2.1) because the values of the objective function for $c\tilde{\mathbf{a}}$ and $\tilde{\mathbf{a}}$ are the same, and $c\tilde{\mathbf{a}}$ also satisfies the constraints in (S2.1). In particular, we choose c such that $\hat{\mathbf{a}}_k = c\tilde{\mathbf{a}}$ satisfies $\hat{\mathbf{a}}_k^T \mathbf{H} \hat{\mathbf{a}}_k = 1$. Hence, $\hat{\mathbf{a}}_k$ satisfies all constraints in (2.17). Now let $\hat{\mathbf{a}}$ be any other vector satisfying all constraints in (2.17), and hence $\hat{\mathbf{a}}$ also satisfies the constraints in (S2.1). As described above, $\hat{\mathbf{a}}_k$ is a solution to (S2.1), so we have

$$\frac{\hat{\mathbf{a}}_k^T \Xi \hat{\mathbf{a}}_k}{\hat{\mathbf{a}}_k^T \mathbf{Q} \hat{\mathbf{a}}_k} \geq \frac{\hat{\mathbf{a}}^T \Xi \hat{\mathbf{a}}}{\hat{\mathbf{a}}^T \mathbf{Q} \hat{\mathbf{a}}}.$$

Hence, $\hat{\mathbf{a}}_k$ is a solution to (2.17). To solve (2.17), we just need to solve (S2.1) and then make a scaling.

Next, we convert (S2.1) to a generalized eigenvalue problem. Let $\mathbf{C} = [\mathbf{H}\hat{\mathbf{a}}_1, \dots, \mathbf{H}\hat{\mathbf{a}}_{k-1}]$ be an $M \times (k-1)$ matrix, where the l -th column is the vector $\mathbf{H}\hat{\mathbf{a}}_l$, $1 \leq l \leq k-1$. Then the constraints in (S2.1) can be written as $\mathbf{C}^T \mathbf{a} = \mathbf{0}$, that is, \mathbf{a} belongs to the orthogonal complementary subspace spanned by the columns of \mathbf{C} . Let \mathbf{P}^\perp be the orthogonal projection matrix onto the orthogonal complementary subspace spanned by the columns of \mathbf{C} in \mathbb{R}^M . Then we have

$$\mathbf{P}^\perp \mathbf{P}^\perp = \mathbf{P}^\perp, \quad (\mathbf{P}^\perp)^T = \mathbf{P}^\perp, \quad \mathbf{P}^\perp \mathbf{C} = \mathbf{0}. \quad (\text{S2.2})$$

We will show that if $\tilde{\mathbf{b}}$ is a solution to the following problem, then $\tilde{\mathbf{a}} = \mathbf{P}^\perp \tilde{\mathbf{b}}$ is a solution to (S2.1):

$$\max_{\mathbf{0} \neq \mathbf{b} \in \mathbb{R}^M} \frac{\mathbf{b}^T \mathbf{P}^\perp \Xi \mathbf{P}^\perp \mathbf{b}}{\mathbf{b}^T \mathbf{P}^\perp \mathbf{Q} \mathbf{P}^\perp \mathbf{b}}. \quad (\text{S2.3})$$

First, the third property in (S2.2) implies that $\tilde{\mathbf{a}} = \mathbf{P}^\perp \tilde{\mathbf{b}}$ satisfies the constraints in (S2.1). Second, let \mathbf{a}_0 be any vector satisfying the constraint in (S2.1), that is, \mathbf{a} belongs to the orthogonal complementary subspace spanned by the columns of \mathbf{C} . Then we have $\mathbf{P}^\perp \mathbf{a}_0 = \mathbf{a}_0$, and

$$\frac{\tilde{\mathbf{a}}^T \Xi \tilde{\mathbf{a}}}{\tilde{\mathbf{a}}^T \mathbf{Q} \tilde{\mathbf{a}}} = \frac{\tilde{\mathbf{b}}^T \mathbf{P}^\perp \Xi \mathbf{P}^\perp \tilde{\mathbf{b}}}{\tilde{\mathbf{b}}^T \mathbf{P}^\perp \mathbf{Q} \mathbf{P}^\perp \tilde{\mathbf{b}}} \geq \frac{\mathbf{a}_0^T \mathbf{P}^\perp \Xi \mathbf{P}^\perp \mathbf{a}_0}{\mathbf{a}_0^T \mathbf{P}^\perp \mathbf{Q} \mathbf{P}^\perp \mathbf{a}_0} = \frac{\mathbf{a}_0^T \Xi \mathbf{a}_0}{\mathbf{a}_0^T \mathbf{Q} \mathbf{a}_0}, \quad (\text{S2.4})$$

where the inequality is because $\tilde{\mathbf{b}}$ is the solution to (S2.3). Since \mathbf{a}_0 is an arbitrary vector satisfying the constraint in (S2.1), $\tilde{\mathbf{a}} = \mathbf{P}^\perp \tilde{\mathbf{b}}$ is a solution to (S2.1). We note that the solution to (S2.3) is the first eigenvector of the generalized eigenvalue problem

$$\mathbf{P}^\perp \boldsymbol{\Xi} \mathbf{P}^\perp \mathbf{b} = \lambda \mathbf{P}^\perp \mathbf{Q} \mathbf{P}^\perp \mathbf{b}. \quad (\text{S2.5})$$

Therefore, to solve (2.17), we just need to calculate the first eigenvector of the generalized eigenvalue problem (S2.3) which can be obtained by a power method.

S2.2 Solving (2.18)

With the constraints in (2.5), we have that the variables \hat{z}_{ik} are uncorrelated, and $\sum_{i=1}^n \hat{z}_{ik} = 0$. So the penalized least squares problem (2.6) can be decomposed into the following optimization problems:

$$\min_{\mu(t)} \left[-\frac{2}{n} \sum_{i=1}^n \int_0^1 \{Y_i(t) D^{d_2} \mu(t)\}^2 dt + \int_0^1 \{D^{d_2} \mu(t)\}^2 dt + \kappa \int_0^1 \{D^2 \mu(t)\}^2 dt \right], \quad (\text{S2.6})$$

and

$$\min_{\xi_k(t)} \left[-\frac{2}{n} \sum_{i=1}^n \int_0^1 \{Y_i(t) \hat{z}_{ik} D^{d_2} \xi_k(t)\}^2 dt + \int_0^1 \{D^{d_2} \xi_k(t)\}^2 dt + \kappa \int_0^1 \{D^2 \xi_k(t)\}^2 dt \right] \quad (\text{S2.7})$$

for $1 \leq k \leq K$. Let $\boldsymbol{\Pi}(t)$ be a vector of basis functions to expand $\mu(t)$ and $\xi_k(t)$. Solving (S2.6) and (S2.7), we get the estimates

$$\begin{aligned} \hat{\mu}(t) &= \boldsymbol{\Pi}(t)^\top \left[\int_0^1 D^{d_2} \boldsymbol{\Pi}(t) D^{d_2} \boldsymbol{\Pi}(t)^\top dt + \kappa \int_0^1 D^2 \boldsymbol{\Pi}(t) D^2 \boldsymbol{\Pi}(t)^\top dt \right]^{-1} \\ &\quad \times \int_0^1 D^{d_2} \boldsymbol{\Pi}(t) \bar{Y}(t) dt, \end{aligned} \quad (\text{S2.8})$$

$$\begin{aligned} \hat{\xi}_k(t) &= \boldsymbol{\Pi}(t)^\top \left[\int_0^1 D^{d_2} \boldsymbol{\Pi}(t) D^{d_2} \boldsymbol{\Pi}(t)^\top dt + \kappa \int_0^1 D^2 \boldsymbol{\Pi}(t) D^2 \boldsymbol{\Pi}(t)^\top dt \right]^{-1} \\ &\quad \times \int_0^1 \sum_{i=1}^n D^{d_2} \boldsymbol{\Pi}(t) Y_i(t) \hat{z}_{ik} dt / n \end{aligned} \quad (\text{S2.9})$$

for $1 \leq k \leq K$.

S3 Additional details for numerical illustration

S3.1 Simulation 1

Figure S1 shows some sample curves for the two types of $X(s)$ in simulation 1.

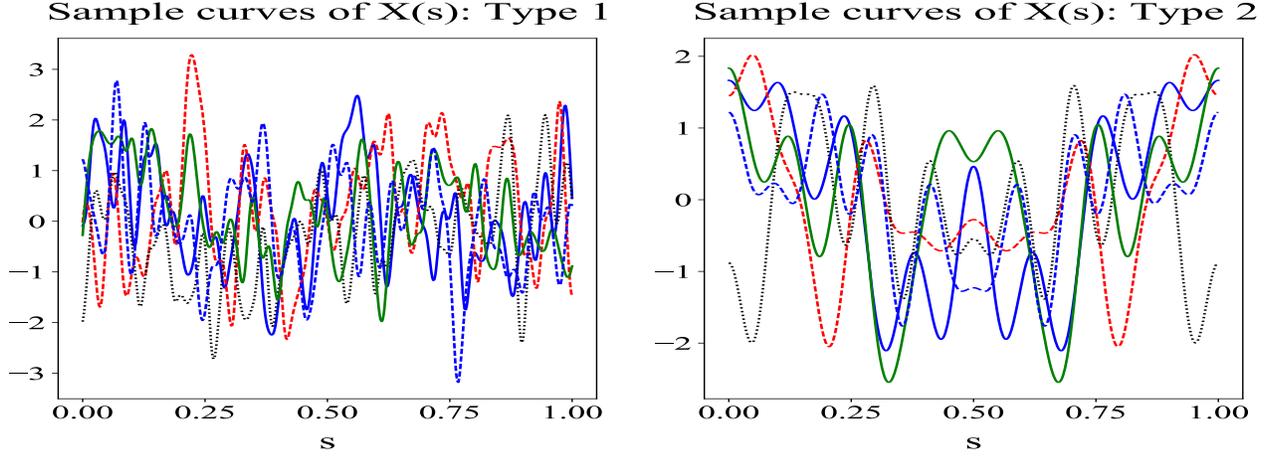


Figure S1: Sample curves for each type of $X(s)$ in Simulation 1.

Coefficient functions are shown in Figure S2. The specific forms of $\mathfrak{B}_1(s, t)$ and $\mathfrak{B}_2(s, t)$ are given below.

- $\mathfrak{B}_1(s, t) = \sum_{i=16}^{30} \{f_i(s)f_i(t)\}/(i - 15)^2$ where $f_i(s) = h(is)$, and $h(s)$, ie, $f_1(s)$, is the triangle wave function shown in the top-left of Figure S3. We also draw the plot of $f_{20}(s) = h(20s)$ in the top-right of Figure S3.
- $\mathfrak{B}_2(s, t) = \sum_{i=21}^{30} g_i(s)g_i(t)/(i - 20)$, where the square wave functions $g_i(s) = (-1)^{d_i(s)}d_i(s)$ with $d_i(s) = \lfloor si \rfloor \% i$. The floor function $\lfloor t \rfloor$ is the largest integer not greater than t , and $a \% b$ is the remainder of the division of the integer a by the integer b . We draw the plots $g_{21}(s)$ and $g_{30}(s)$ in the bottom of Figure S3.

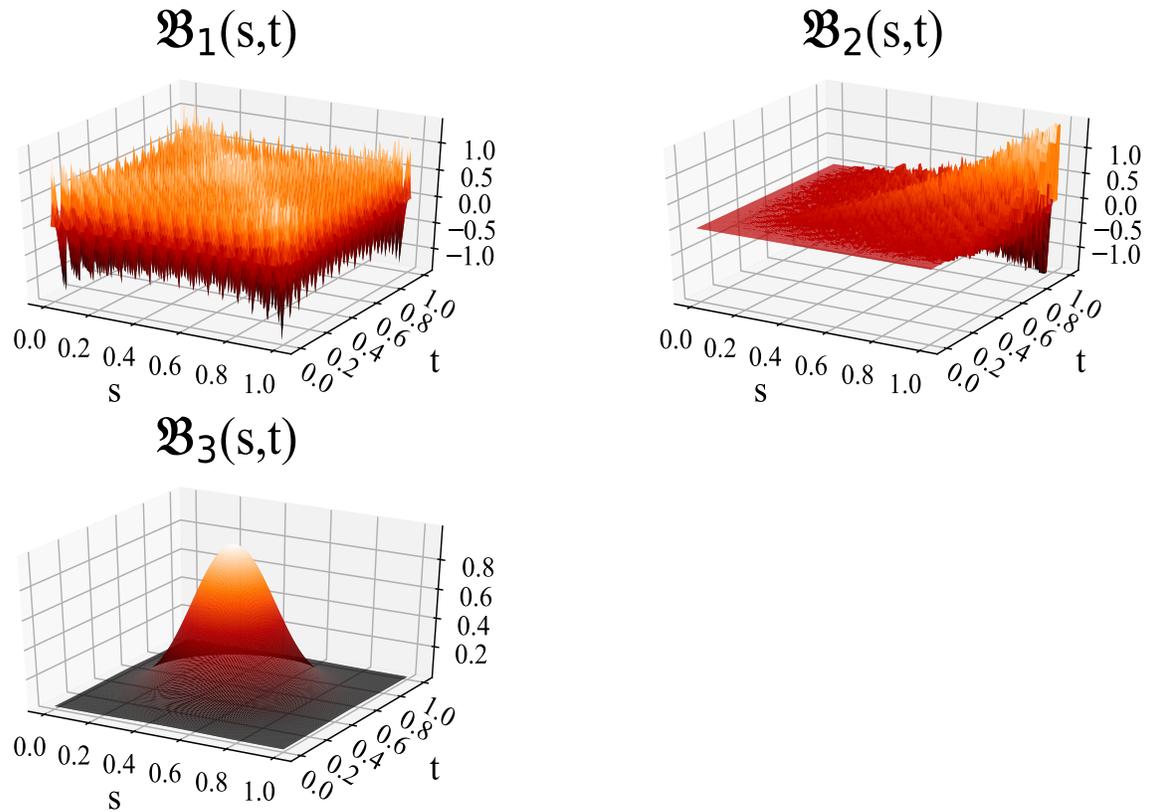
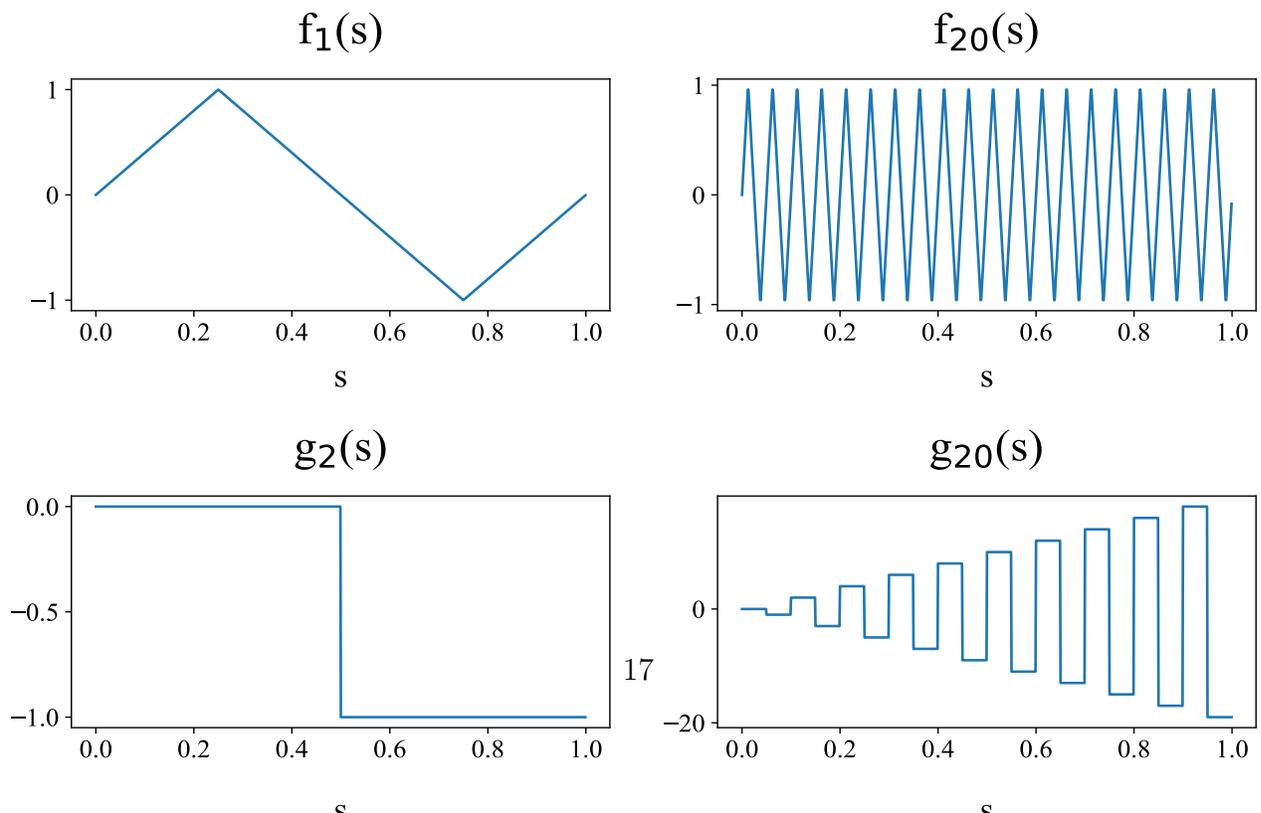


Figure S2: Three types of coefficient functions $\mathfrak{B}_1(s,t) \sim \mathfrak{B}_3(s,t)$ in Simulation 1.



In Simulation 1, we consider two types of intercept coefficient functions: $\mathfrak{U}_1(t) = \sum_{i=2}^{65} g_i(t)$ and $\mathfrak{U}_2(t) = \sin(2\pi t)$, where $g_i(t)$'s are the square wave functions defined above. These two intercept coefficient functions are shown below in Figure S4.

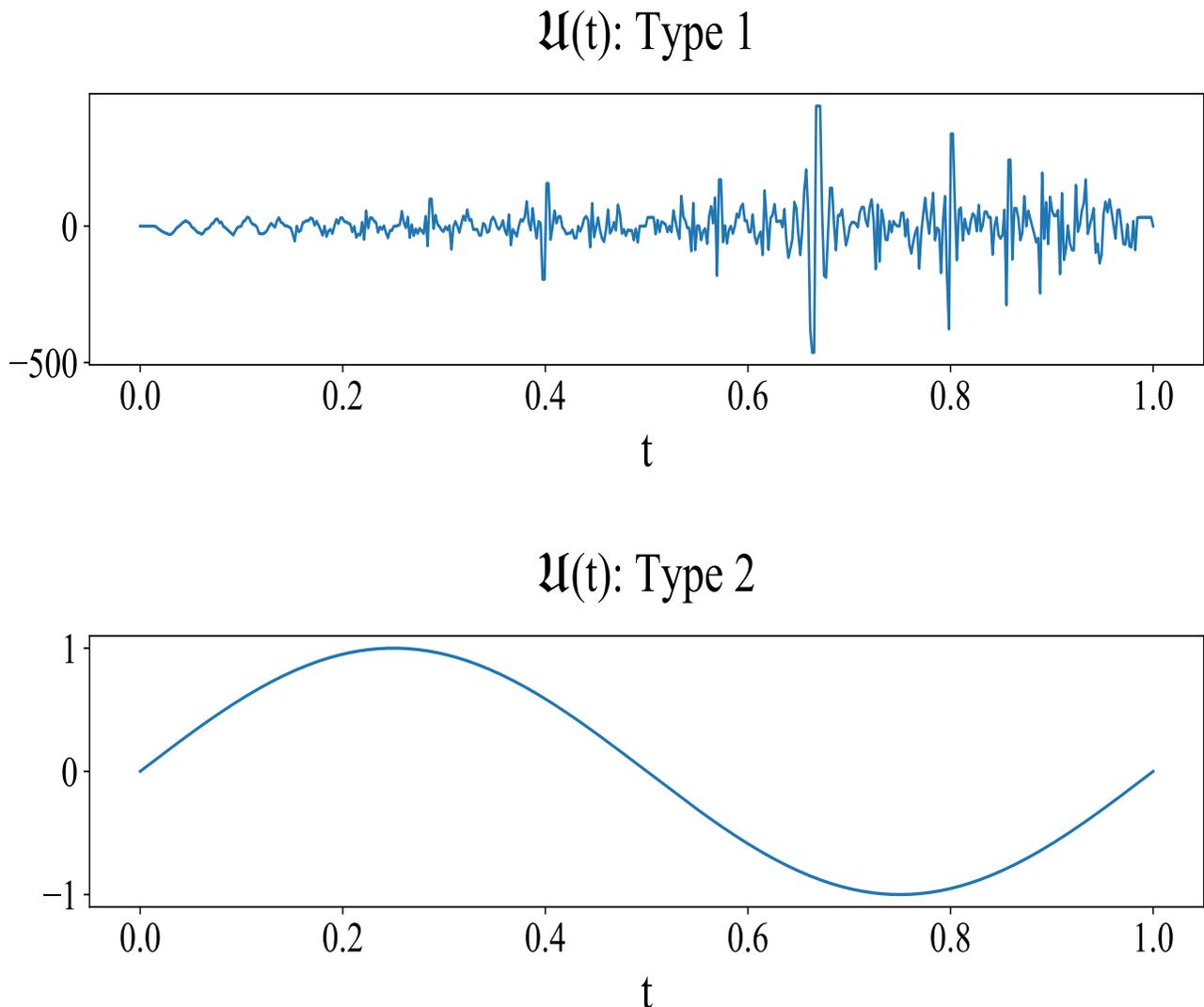


Figure S4: Two types of intercept functions used in Simulation 1.

S3.1.1 Results for settings with $2^7 = 128$ observation points

To run all methods considered in this paper, we consider a setting with $2^7 = 128$ observation time points on each curve. The MISEEs and running times are summarized in Tables S1 and S2, respectively. For $\sigma < 1$, when $\mathfrak{U}(t)$ or $\mathfrak{B}(s, t)$ is spiky, the *fof.deriv* has the lowest MISEEs; when both $\mathfrak{U}(t)$ and $\mathfrak{B}(s, t)$

are smooth (Type 2 of $\mathfrak{U}(t)$ and type 3 of $\mathfrak{B}(s, t)$), all methods except *fdapace* have close MISSEs which are lower than that of *fdapace*. When $\sigma = 1$, the methods *fof.deriv*, *sSigComp* and *wSigComp* perform similarly. The MISSEs of these three methods are lower than *fdapace*, *pffr* and *FDboost* when $\mathfrak{U}(t)$ or $\mathfrak{B}(s, t)$ is spiky, or similar with *pffr* and *FDboost* when both $\mathfrak{U}(t)$ and $\mathfrak{B}(s, t)$ are smooth. In implementation, we use 20 B-spline basis functions of s and 20 B-spline basis functions of t for *pffr*, and 2000 boosting iterations *FDboost*. Increasing these values can slightly decrease the MISSEs of these two methods, but greatly increases the computational times, which have been much larger than other methods.

Table S1: Average (and standard deviation) of MISEEs from 100 replicates for Simulation 1 with $2^7 = 128$ observation time points on each curve.

X	σ	\mathfrak{U}	\mathfrak{B}	<i>fof.deriv</i>	<i>sSigComp</i>	<i>wSigComp</i>	<i>fdapace</i>	<i>pffr</i>	<i>FDboost</i>	
1	.01	1	1	$7.55(1.4) \cdot 10^{-4}$	$2.26(0.6) \cdot 10^{-2}$	$7.66(1.1) \cdot 10^{-3}$	0.64(0.02)	1.60(0.02)	1.55(0.02)	
			2	$7.93(0.4) \cdot 10^{-4}$	$1.18(0.2) \cdot 10^{-2}$	$7.04(5.0) \cdot 10^{-3}$	0.62(0.01)	1.53(0.03)	1.54(0.02)	
			3	$2.48(0.0) \cdot 10^{-5}$	$4.90(0.0) \cdot 10^{-4}$	$4.90(0.0) \cdot 10^{-4}$	0.61(0.00)	0.60(0.00)	0.55(0.00)	
		2	1	$7.45(1.3) \cdot 10^{-4}$	$2.20(0.6) \cdot 10^{-2}$	$8.69(2.1) \cdot 10^{-3}$	0.03(0.02)	1.00(0.02)	1.00(0.03)	
			2	$5.44(0.4) \cdot 10^{-4}$	$1.14(0.2) \cdot 10^{-2}$	$6.06(4.0) \cdot 10^{-3}$	0.02(0.00)	0.93(0.02)	0.97(0.02)	
			3	$7.37(1.6) \cdot 10^{-7}$	$7.51(1.7) \cdot 10^{-7}$	$1.24(0.5) \cdot 10^{-6}$	$4.36(1.9) \cdot 10^{-3}$	$6.25(0.3) \cdot 10^{-6}$	$1.39(0.7) \cdot 10^{-5}$	
	.1	1	1	$2.92(0.4) \cdot 10^{-3}$	$2.39(0.6) \cdot 10^{-2}$	$9.41(1.1) \cdot 10^{-3}$	0.64(0.02)	1.61(0.03)	1.56(0.03)	
			2	$5.01(0.6) \cdot 10^{-3}$	$1.37(0.2) \cdot 10^{-2}$	$8.94(4.8) \cdot 10^{-3}$	0.63(0.01)	1.54(0.03)	1.55(0.03)	
			3	$2.36(0.3) \cdot 10^{-4}$	$7.01(0.3) \cdot 10^{-4}$	$7.68(0.5) \cdot 10^{-4}$	0.62(0.00)	0.61(0.00)	0.56(0.00)	
		2	1	$2.90(0.4) \cdot 10^{-3}$	$2.36(0.6) \cdot 10^{-2}$	$9.36(1.0) \cdot 10^{-3}$	0.04(0.02)	1.00(0.02)	1.01(0.02)	
			2	$4.90(0.7) \cdot 10^{-3}$	$1.36(0.2) \cdot 10^{-2}$	$9.64(1.0) \cdot 10^{-3}$	0.02(0.00)	0.94(0.02)	0.98(0.02)	
			3	$5.34(1.1) \cdot 10^{-5}$	$4.23(1.1) \cdot 10^{-5}$	$1.02(0.3) \cdot 10^{-4}$	$1.28(0.2) \cdot 10^{-2}$	$2.10(0.4) \cdot 10^{-4}$	$1.38(0.4) \cdot 10^{-4}$	
	1	1	1	$6.26(0.6) \cdot 10^{-2}$	0.21(0.05)	$8.23(1.3) \cdot 10^{-2}$	1.47(0.05)	1.60(0.03)	1.56(0.03)	
			2	0.14(0.01)	0.17(0.02)	0.12(0.01)	1.46(0.04)	1.59(0.05)	1.56(0.02)	
			3	$2.11(0.3) \cdot 10^{-2}$	$2.17(0.2) \cdot 10^{-2}$	$2.33(0.3) \cdot 10^{-2}$	1.44(0.03)	0.60(0.01)	0.56(0.01)	
		2	1	$5.58(0.5) \cdot 10^{-2}$	0.18(0.04)	$7.34(1.2) \cdot 10^{-2}$	0.87(0.04)	1.00(0.03)	1.01(0.03)	
			2	0.13(0.01)	0.17(0.02)	0.11(0.01)	0.85(0.04)	0.95(0.03)	0.98(0.02)	
			3	$2.40(0.7) \cdot 10^{-3}$	$3.08(1.2) \cdot 10^{-3}$	$4.22(1.5) \cdot 10^{-3}$	0.83(0.04)	$5.51(0.6) \cdot 10^{-3}$	$1.12(0.1) \cdot 10^{-2}$	
	2	.01	1	1	$4.25(1.5) \cdot 10^{-4}$	$1.22(0.4) \cdot 10^{-3}$	$9.45(2.3) \cdot 10^{-3}$	0.62(0.02)	1.59(0.02)	1.52(0.02)
				2	$1.21(0.1) \cdot 10^{-3}$	$2.06(0.4) \cdot 10^{-3}$	$6.29(0.5) \cdot 10^{-3}$	0.84(0.07)	1.29(0.02)	1.33(0.04)
				3	$2.47(0.0) \cdot 10^{-5}$	$4.90(0.0) \cdot 10^{-4}$	$4.95(0.8) \cdot 10^{-4}$	0.60(0.00)	0.60(0.00)	0.55(0.00)
			2	1	$3.91(1.6) \cdot 10^{-4}$	$7.44(3.7) \cdot 10^{-4}$	$9.84(3.8) \cdot 10^{-3}$	0.02(0.02)	0.99(0.02)	0.95(0.02)
				2	$6.42(1.0) \cdot 10^{-4}$	$1.60(0.4) \cdot 10^{-3}$	$5.78(0.7) \cdot 10^{-3}$	0.25(0.06)	0.68(0.09)	0.74(0.04)
				3	$6.01(1.1) \cdot 10^{-7}$	$6.55(1.4) \cdot 10^{-7}$	$7.45(2.9) \cdot 10^{-7}$	$1.18(0.0) \cdot 10^{-3}$	$3.37(0.2) \cdot 10^{-6}$	$9.65(3.6) \cdot 10^{-6}$
.1		1	1	$1.22(0.2) \cdot 10^{-3}$	$1.99(0.3) \cdot 10^{-3}$	$1.00(0.3) \cdot 10^{-2}$	0.63(0.02)	1.61(0.02)	1.55(0.03)	
			2	$2.18(0.1) \cdot 10^{-3}$	$3.14(0.3) \cdot 10^{-3}$	$6.75(0.5) \cdot 10^{-3}$	0.84(0.06)	1.30(0.03)	1.35(0.03)	
			3	$2.33(0.2) \cdot 10^{-4}$	$6.98(0.2) \cdot 10^{-4}$	$7.02(0.2) \cdot 10^{-4}$	0.61(0.00)	0.61(0.00)	0.56(0.00)	
		2	1	$1.23(0.2) \cdot 10^{-3}$	$1.56(0.4) \cdot 10^{-3}$	$1.01(0.2) \cdot 10^{-2}$	0.03(0.02)	1.01(0.03)	0.98(0.03)	
			2	$1.60(0.1) \cdot 10^{-3}$	$2.67(0.4) \cdot 10^{-3}$	$6.38(0.5) \cdot 10^{-3}$	0.24(0.06)	0.68(0.02)	0.76(0.03)	
			3	$4.96(1.2) \cdot 10^{-5}$	$3.54(0.9) \cdot 10^{-5}$	$7.37(3.1) \cdot 10^{-5}$	$3.09(0.2) \cdot 10^{-3}$	$1.37(0.0) \cdot 10^{-4}$	$1.14(0.0) \cdot 10^{-4}$	
1		1	1	$5.24(0.4) \cdot 10^{-2}$	$5.38(0.4) \cdot 10^{-2}$	$5.32(0.5) \cdot 10^{-2}$	0.82(0.03)	1.54(0.04)	1.54(0.04)	
			2	$6.34(0.5) \cdot 10^{-2}$	$7.57(0.9) \cdot 10^{-2}$	$6.39(0.5) \cdot 10^{-2}$	1.04(0.06)	1.32(0.03)	1.34(0.03)	
			3	$2.08(0.2) \cdot 10^{-2}$	$2.13(0.2) \cdot 10^{-2}$	$2.12(0.2) \cdot 10^{-2}$	0.80(0.02)	0.61(0.01)	0.56(0.01)	
		2	1	$4.66(0.3) \cdot 10^{-2}$	$5.33(0.5) \cdot 10^{-2}$	$4.68(0.4) \cdot 10^{-2}$	0.22(0.02)	1.00(0.02)	0.97(0.03)	
			2	$5.87(0.4) \cdot 10^{-2}$	$6.96(0.8) \cdot 10^{-2}$	$5.88(0.5) \cdot 10^{-2}$	0.43(0.07)	0.71(0.03)	0.75(0.05)	
			3	$2.40(1.0) \cdot 10^{-3}$	$2.87(1.0) \cdot 10^{-3}$	$2.32(0.9) \cdot 10^{-3}$	0.19(0.01)	$3.49(0.1) \cdot 10^{-3}$	$7.36(0.4) \cdot 10^{-3}$	

Table S2: Average (and standard deviation) of running time from 100 replicates for Simulation 1 with $2^7 = 128$ observation time points on each curve.

X	σ	\mathfrak{U}	\mathfrak{B}	<i>fof.deriv</i>	<i>sSigComp</i>	<i>wSigComp</i>	<i>fdapace</i>	<i>pffr</i>	<i>FDboost</i>	
1	.01	1	1	6.45(1.50)	0.76(0.17)	16.84(3.54)	0.27(0.02)	881.03(265.52)	257.20(61.95)	
			2	9.62(2.17)	0.98(0.22)	32.06(7.11)	0.27(0.01)	744.09(279.92)	256.37(61.61)	
			3	3.78(0.93)	0.61(0.15)	4.88(1.20)	0.27(0.03)	473.64(136.99)	260.87(63.81)	
		2	1	6.40(1.44)	0.76(0.16)	16.68(3.64)	0.27(0.01)	913.61(289.19)	256.09(60.81)	
			2	9.70(2.28)	0.98(0.23)	33.26(7.19)	0.27(0.01)	893.20(501.17)	256.10(61.34)	
			3	3.76(0.93)	0.59(0.14)	4.79(1.03)	0.27(0.01)	426.18(118.93)	261.37(64.50)	
	.1	1	1	6.86(1.61)	0.79(0.17)	17.37(4.04)	0.27(0.02)	879.12(266.68)	255.41(61.70)	
			2	9.83(2.47)	0.98(0.24)	33.52(8.11)	0.27(0.01)	718.14(203.36)	255.17(60.94)	
			3	4.04(0.96)	0.62(0.16)	8.47(2.06)	0.27(0.03)	479.81(144.20)	260.42(64.41)	
		2	1	6.66(1.62)	0.77(0.18)	17.59(4.64)	0.27(0.01)	938.39(289.41)	256.18(61.02)	
			2	9.96(2.18)	0.99(0.22)	33.61(7.30)	0.27(0.01)	831.71(394.53)	255.04(60.75)	
			3	4.03(0.96)	0.63(0.16)	9.41(2.13)	0.27(0.01)	395.20(104.82)	261.05(63.57)	
	1	1	1	47.04(1.33)	1.30(0.07)	355.34(63.63)	0.27(0.02)	427.80(134.28)	256.40(61.20)	
			2	45.74(0.88)	1.30(0.05)	306.84(46.83)	0.27(0.01)	645.21(255.56)	255.89(61.74)	
			3	45.99(1.29)	1.34(0.16)	278.86(59.19)	0.27(0.01)	343.97(53.45)	261.12(63.86)	
		2	1	46.72(1.05)	1.29(0.06)	342.95(63.63)	0.27(0.01)	411.02(139.00)	256.50(61.53)	
			2	45.78(0.95)	1.30(0.07)	317.22(46.16)	0.27(0.03)	615.19(251.84)	257.06(61.32)	
			3	46.59(1.10)	1.22(0.05)	305.80(40.81)	0.27(0.01)	334.43(64.00)	261.27(63.78)	
	2	.01	1	1	5.61(1.37)	0.72(0.17)	9.17(2.51)	0.26(0.01)	1199.18(347.90)	254.51(57.20)
				2	6.39(1.45)	0.75(0.17)	10.67(2.19)	0.26(0.01)	596.83(219.97)	255.18(58.12)
				3	3.84(0.92)	0.60(0.15)	3.62(0.77)	0.26(0.01)	561.62(181.14)	258.98(58.81)
			2	1	6.04(1.39)	1.07(0.24)	9.22(1.93)	0.26(0.01)	2042.23(1000.91)	254.60(57.33)
				2	6.43(1.45)	0.76(0.16)	10.74(2.27)	0.27(0.02)	483.22(121.07)	258.29(59.49)
				3	3.84(0.92)	0.59(0.15)	3.61(0.80)	0.26(0.01)	477.68(138.98)	259.41(59.88)
.1		1	1	5.77(1.31)	0.71(0.16)	8.99(2.10)	0.26(0.01)	1161.19(465.40)	254.81(57.03)	
			2	6.56(1.57)	0.77(0.18)	10.93(2.71)	0.26(0.01)	552.98(173.10)	254.45(57.48)	
			3	4.02(0.96)	0.63(0.15)	5.83(1.52)	0.26(0.01)	548.53(152.72)	258.63(58.28)	
		2	1	5.78(1.39)	0.73(0.21)	8.56(2.30)	0.26(0.01)	1927.10(938.49)	254.57(57.02)	
			2	6.57(1.44)	0.77(0.16)	10.60(2.14)	0.26(0.01)	510.83(183.51)	258.08(59.11)	
			3	3.96(0.95)	0.62(0.15)	6.11(1.74)	0.26(0.01)	465.04(132.89)	258.82(58.94)	
1		1	1	18.85(0.59)	0.94(0.04)	60.23(19.61)	0.27(0.02)	857.33(195.60)	254.37(57.14)	
			2	18.62(0.57)	0.91(0.05)	65.18(15.15)	0.26(0.01)	586.18(181.00)	254.68(57.86)	
			3	19.42(0.43)	0.93(0.05)	56.34(17.88)	0.26(0.01)	522.34(32.48)	259.05(59.20)	
		2	1	18.75(0.60)	1.06(0.14)	58.02(12.77)	0.26(0.02)	948.96(362.76)	254.23(56.52)	
			2	18.48(0.43)	0.91(0.07)	65.96(18.53)	0.26(0.02)	537.78(134.72)	256.16(58.55)	
			3	19.42(0.58)	0.93(0.05)	54.79(20.16)	0.26(0.02)	399.37(84.82)	258.64(59.00)	

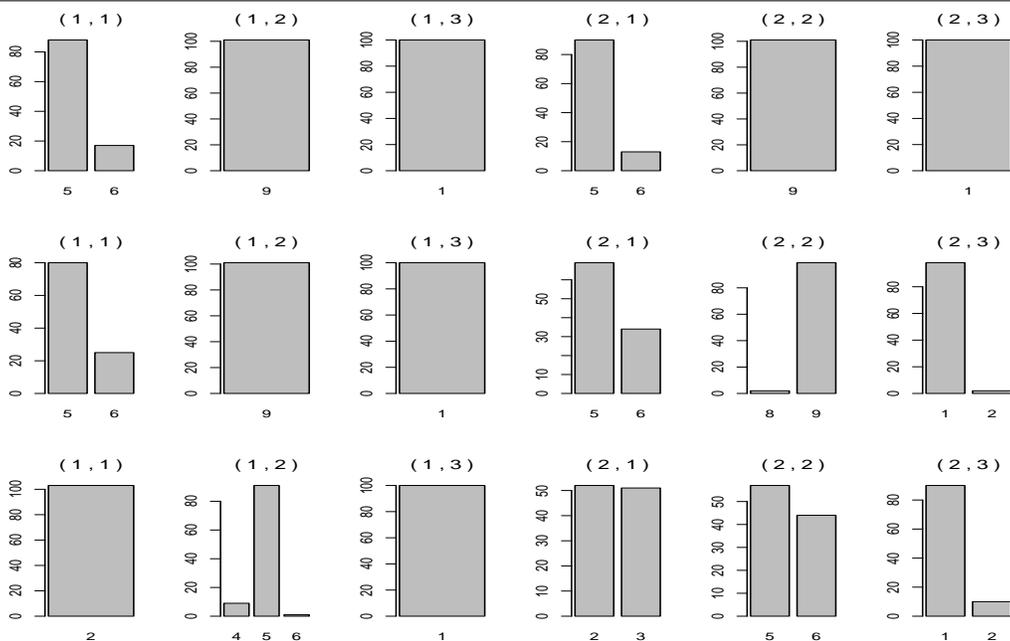
S3.1.2 Additional results with $2^9 = 512$ observation time points

Table S3: Percentage (%) of the selected orders (d_1, d_2) of derivatives in the *fof.deriv* method from 100 replicates in Simulation 1 with $2^9 = 512$ observation time points on each curve.

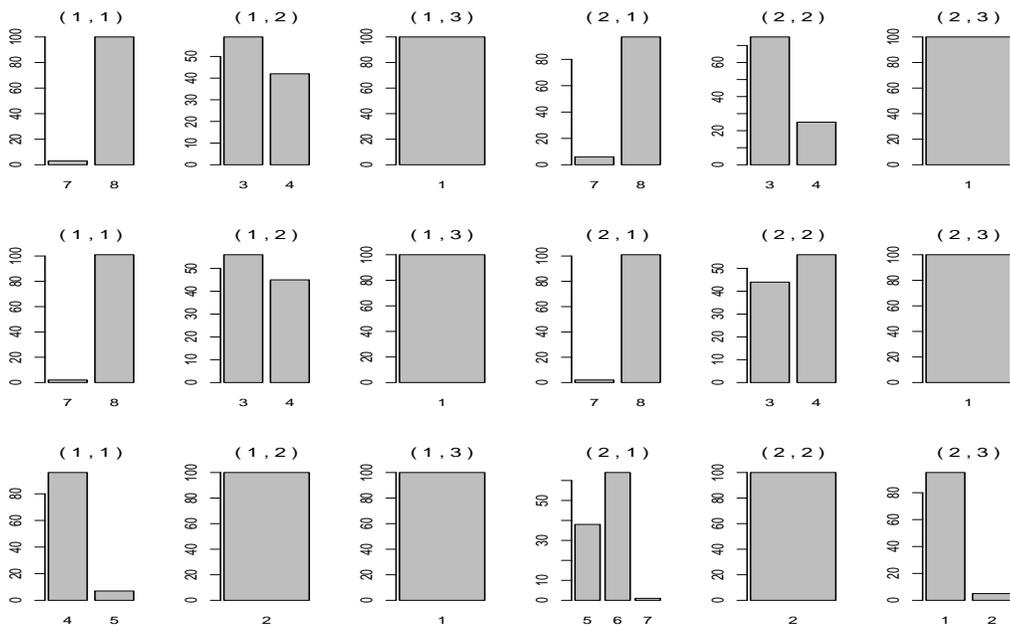
		$\mathfrak{B}_1(s, t)$				$\mathfrak{B}_2(s, t)$				$\mathfrak{B}_3(s, t)$			
		\mathfrak{U}_1		\mathfrak{U}_2		\mathfrak{U}_1		\mathfrak{U}_2		\mathfrak{U}_1		\mathfrak{U}_2	
X	σ	(d_1, d_2)	%	(d_1, d_2)	%	(d_1, d_2)	%	(d_1, d_2)	%	(d_1, d_2)	%	(d_1, d_2)	%
1	0.01	(1,1)	100	(1,2)	93	(1,1)	99	(1,1)	99	(0,1)	100	(0,0)	100
				(1,1)	7	(2,1)	1	(2,1)	1				
	0.1	(1,1)	100	(1,1)	100	(1,1)	100	(1,1)	100	(0,1)	100	(0,0)	100
	1	(0,1)	100	(0,0)	99	(0,1)	51	(0,1)	96	(0,1)	97	(0,0)	95
				(1,0)	1	(0,2)	48	(0,0)	3	(1,1)	3	(1,0)	5
						(1,1)	1	(1,1)	1				
2	0.01	(2,1)	68	(2,1)	49	(2,1)	56	(2,1)	52	(0,1)	100	(0,0)	98
		(1,1)	26	(2,2)	17	(1,1)	44	(1,1)	48			(1,0)	2
		(0,1)	6	(1,1)	16								
	0.1	(2,1)	49	(1,1)	46	(2,1)	67	(2,1)	62	(0,1)	91	(0,0)	87
		(1,1)	39	(2,1)	45	(1,1)	32	(1,1)	36	(1,1)	9	(1,0)	13
		(0,1)	12	(0,1)	9	(0,1)	1	(0,1)	2				
	1	(1,1)	83	(0,0)	60	(1,1)	51	(1,1)	59	(0,1)	91	(0,0)	96
		(1,2)	10	(2,0)	24	(2,1)	47	(2,1)	35	(2,1)	6	(1,0)	2
		(0,1)	5	(1,0)	16	other	2	(0,1)	6	other	3	(2,0)	2

Table S4: Average (and standard deviation) of running time from 100 replicates of settings with Type 1 $\mathfrak{U}(s)$ for Simulation 1 with $2^9 = 512$ observation time points on each curve.

		Type 1 $X(s)$				Type 2 $X(s)$			
σ	\mathfrak{B}	<i>fof.deriv</i>	<i>sSigComp</i>	<i>wSigComp</i>	<i>fdapace</i>	<i>fof.deriv</i>	<i>sSigComp</i>	<i>wSigComp</i>	<i>fdapace</i>
.01	1	18.3(0.6)	2.3(0.1)	46.5(1.9)	5.35(0.33)	18.2(0.5)	2.3(0.1)	47.8(1.7)	5.20(0.19)
	2	22.6(0.6)	2.6(0.1)	63.5(2.6)	5.31(0.10)	15.7(0.4)	2.2(0.1)	40.3(0.9)	5.18(0.12)
	2	14.3(0.3)	2.1(0.1)	37.2(0.7)	5.29(0.12)	14.5(0.3)	2.1(0.1)	36.4(0.6)	5.21(0.14)
.1	1	18.7(0.7)	2.4(0.1)	47.0(2.1)	5.33(0.21)	18.4(0.5)	2.3(0.1)	47.5(1.6)	5.19(0.10)
	2	22.8(0.6)	2.6(0.1)	63.5(2.5)	5.31(0.12)	15.8(0.4)	2.2(0.1)	40.6(0.9)	5.17(0.13)
	2	14.7(0.2)	2.1(0.1)	39.3(1.0)	5.31(0.12)	14.6(0.2)	2.1(0.1)	38.1(1.0)	5.18(0.14)
1	1	90.1(1.9)	3.5(0.1)	323.5(76.2)	5.30(0.16)	41.7(1.1)	2.8(0.1)	100.1(20.6)	5.17(0.12)
	2	86.9(1.7)	3.4(0.1)	295.0(55.5)	5.30(0.12)	44.6(1.1)	2.8(0.1)	107.8(26.7)	5.16(0.12)
	2	89.1(1.6)	3.3(0.1)	280.9(94.8)	5.33(0.13)	45.1(0.9)	2.8(0.1)	91.8(29.1)	5.18(0.13)

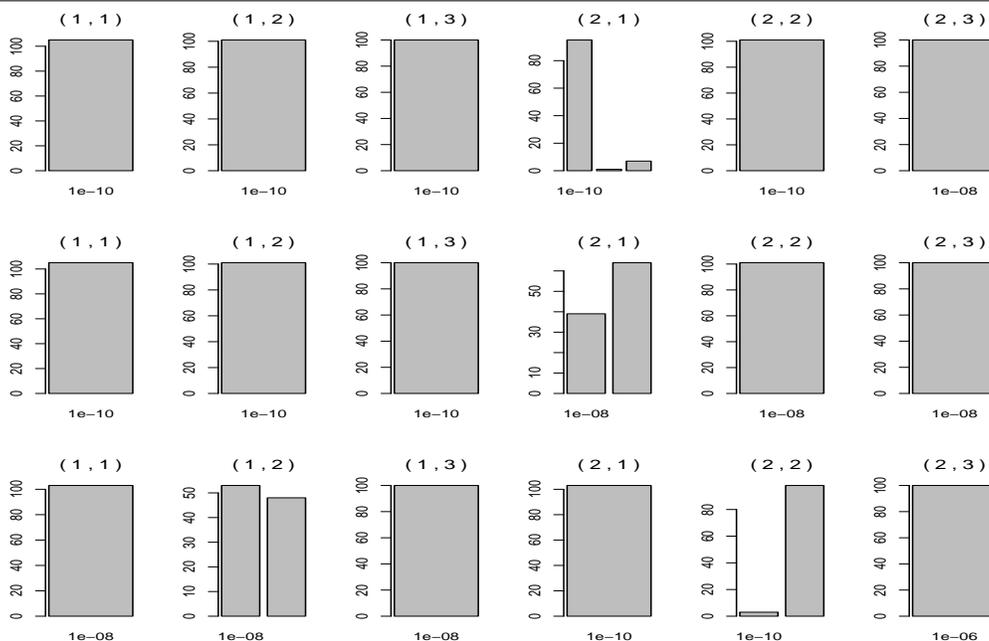


(a) Type 1 $X(s)$

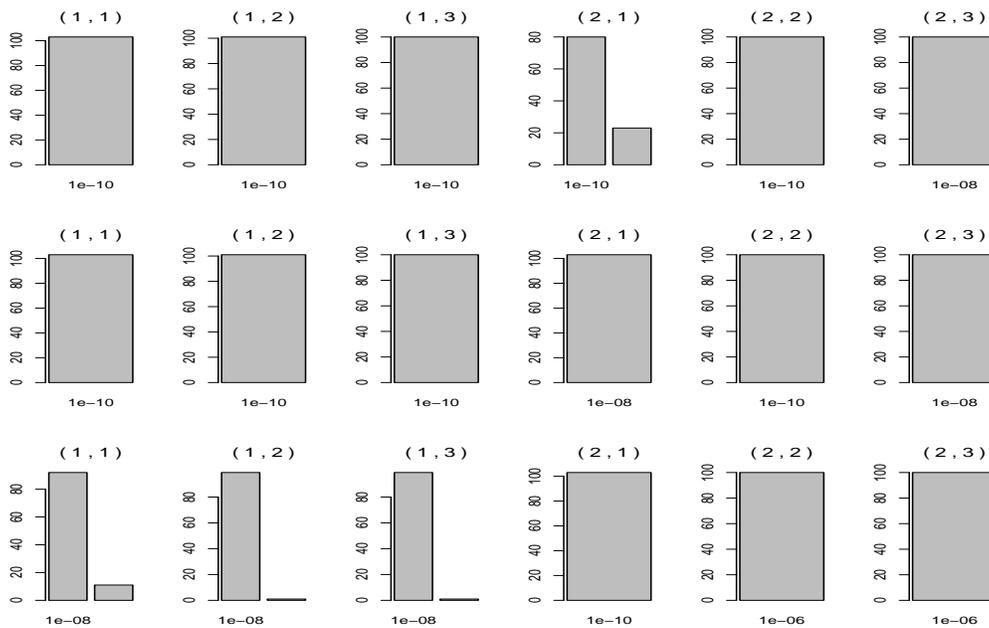


(b) Type 2 $X(s)$

Figure S5: Frequencies of the selected number K of components for each setting of Type 1 (a) and Type 2 (b) of $X(s)$ in Simulation 1 with 512 observation time points on each curve. For each panel, from top to bottom, the three rows show the histograms when $\sigma = 0.01$, 0.1 and 1 , respectively. The two integers on the top of each histogram specify the types of \mathfrak{U} and \mathfrak{B} , respectively.

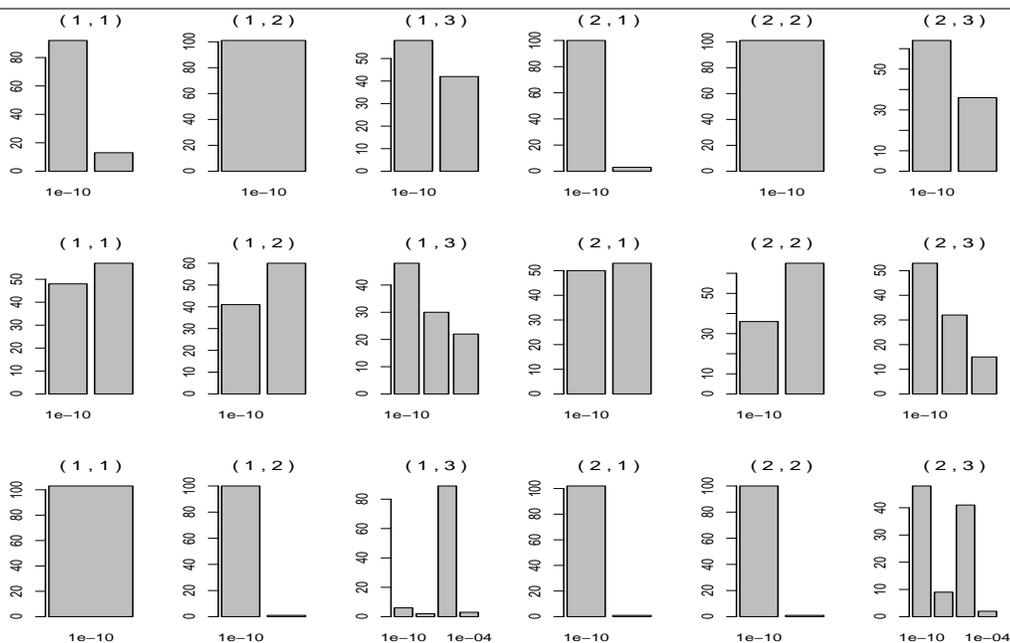


(a) Type 1 X(s)

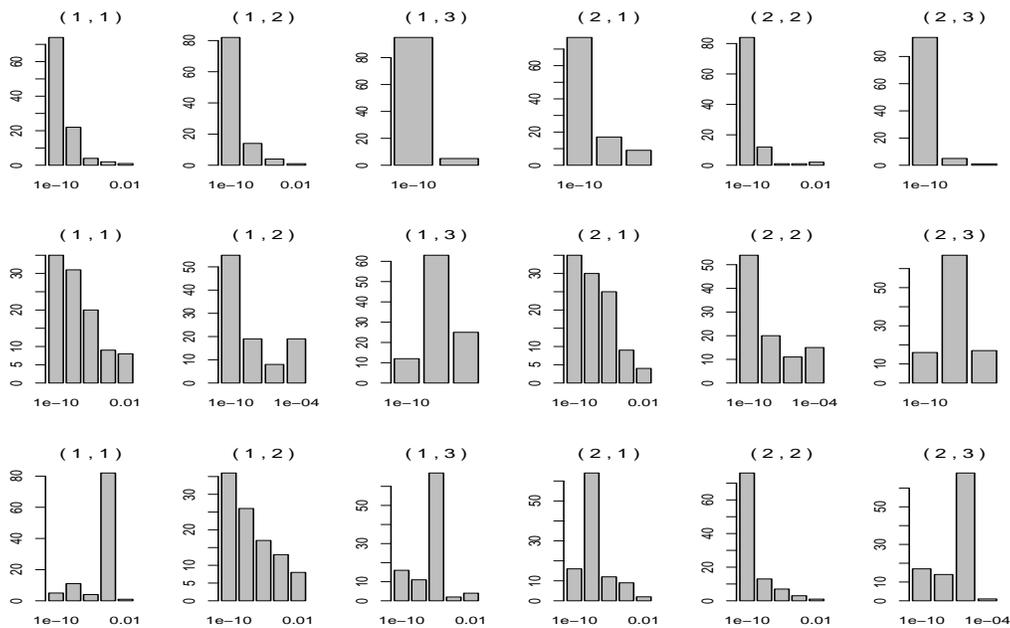


(b) Type 2 X(s)

Figure S6: Frequencies of the selected κ for each setting of Type 1 (a) and Type 2 (b) of $X(s)$ in Simulation 1 with 512 observation time points on each curve. For each panel, from top to bottom, the three rows of panels show the histograms when $\sigma = 0.01, 0.1$ and 1 , respectively. The two integers on the top of each histogram specify the types of \mathfrak{U} and \mathfrak{B} , respectively.

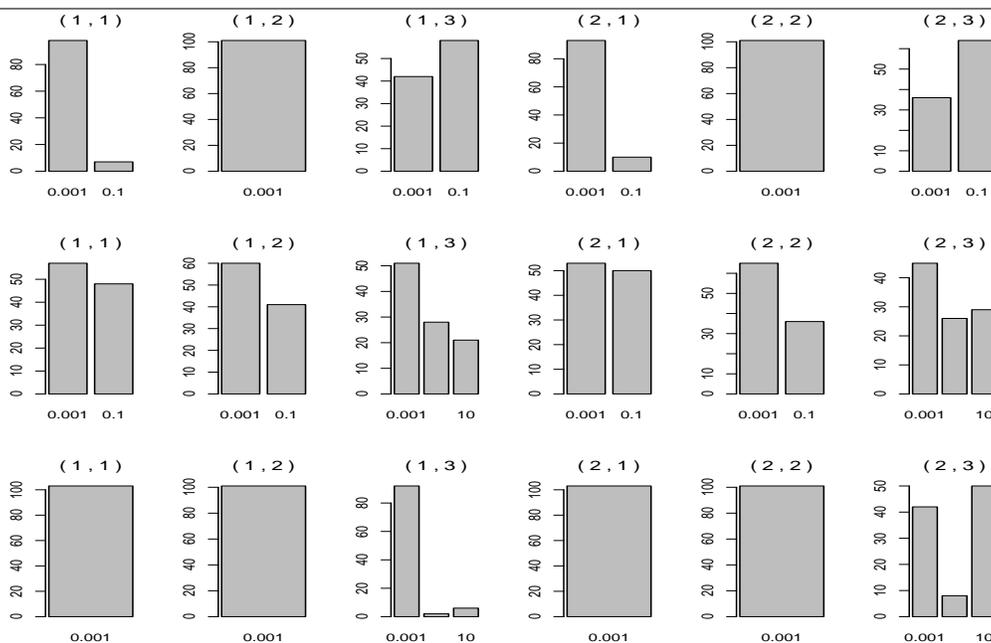


(a) Type 1 $X(s)$

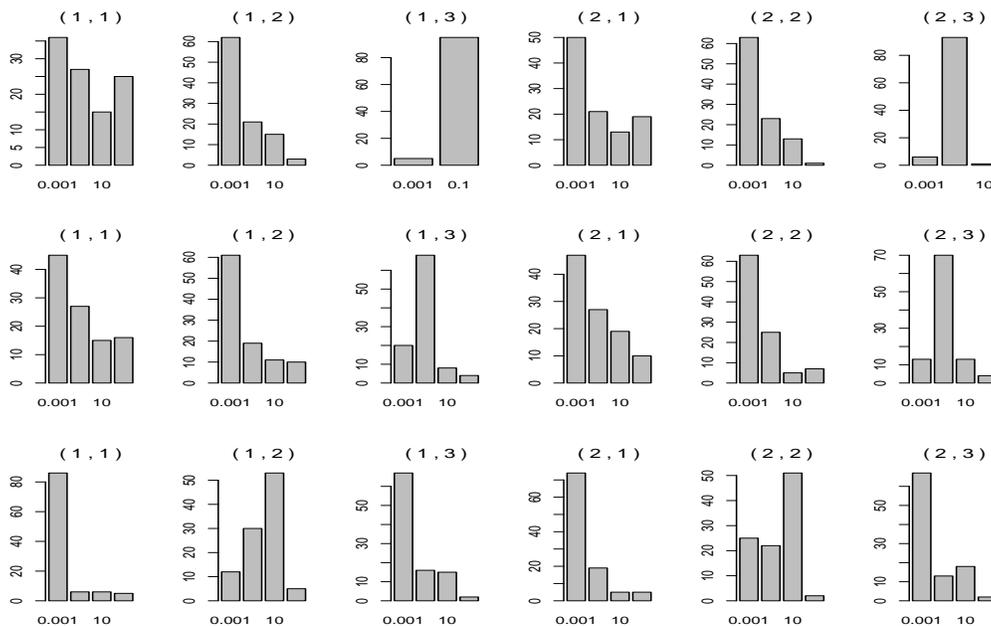


(b) Type 2 $X(s)$

Figure S7: Frequencies of the selected λ for each setting of Type 1 (a) and Type 2 (b) of $X(s)$ in Simulation 1 with 512 observation time points on each curve. For each panel, from top to bottom, the three rows of panels show the histograms when $\sigma = 0.01, 0.1$ and 1 , respectively. The two integers on the top of each histogram specify the types of \mathfrak{U} and \mathfrak{B} , respectively.



(a) Type 1 $X(s)$



(b) Type 2 $X(s)$

Figure S8: Frequencies of the selected τ for each setting of Type 1 (a) and Type 2 (b) of $X(s)$ in Simulation 1 with 512 observation time points on each curve. For each panel, from top to bottom, the three rows of panels show the histograms when $\sigma = 0.01, 0.1$ and 1 , respectively. The two integers on the top of each histogram specify the types of \mathfrak{U} and \mathfrak{B} , respectively.

Table S5: Average (sd) of MISEEs from 100 replicates for Simulation 1 with noisy observations $\tilde{X}(s)$ of Type 1 predictor function, where $\tilde{X}(s) = X(s) + \epsilon_X(s)$, $\epsilon_X(s) \stackrel{\text{iid}}{\sim} N(0, \sigma_X^2)$, $\sigma_X = r\{\sum_{i=1}^T \text{Var}(X(s_i))/T\}^{1/2}$ is $r = 1\%$ or 10% of the variation in $X(s)$, and $T = 512$.

r	σ	\mathfrak{U}	\mathfrak{B}	<i>fof.deriv</i>	<i>sSigComp</i>	<i>wSigComp</i>	<i>fdapace</i>	
1%	.01	1	1	$7.76(1.65) \cdot 10^{-4}$	$2.50(0.54) \cdot 10^{-2}$	$5.50(3.06) \cdot 10^{-3}$	$0.66(0.01)$	
			2	$2.76(0.41) \cdot 10^{-4}$	$1.33(0.19) \cdot 10^{-2}$	$6.26(3.54) \cdot 10^{-3}$	$0.65(0.00)$	
			3	$1.54(0.97) \cdot 10^{-5}$	$3.12(0.00) \cdot 10^{-3}$	$9.68(0.01) \cdot 10^{-4}$	$0.65(0.00)$	
		2	1	$7.80(1.64) \cdot 10^{-4}$	$2.17(0.54) \cdot 10^{-2}$	$5.54(4.55) \cdot 10^{-3}$	$2.07(1.13) \cdot 10^{-2}$	
			2	$2.68(0.39) \cdot 10^{-4}$	$1.03(0.24) \cdot 10^{-2}$	$6.57(4.86) \cdot 10^{-3}$	$1.16(0.27) \cdot 10^{-2}$	
			3	$6.82(0.63) \cdot 10^{-6}$	$6.79(0.62) \cdot 10^{-6}$	$8.90(1.22) \cdot 10^{-6}$	$3.28(1.46) \cdot 10^{-3}$	
	0.1	1	1	$1.78(0.11) \cdot 10^{-3}$	$2.71(0.54) \cdot 10^{-2}$	$1.09(3.09) \cdot 10^{-2}$	$0.67(0.01)$	
			2	$1.98(0.13) \cdot 10^{-3}$	$1.46(0.16) \cdot 10^{-2}$	$9.85(7.01) \cdot 10^{-3}$	$0.66(0.00)$	
			3	$2.23(0.19) \cdot 10^{-4}$	$3.32(0.03) \cdot 10^{-3}$	$1.21(0.03) \cdot 10^{-3}$	$0.65(0.00)$	
		2	1	$1.76(0.14) \cdot 10^{-3}$	$2.41(0.61) \cdot 10^{-2}$	$7.43(7.80) \cdot 10^{-3}$	$2.96(1.24) \cdot 10^{-2}$	
			2	$1.95(0.13) \cdot 10^{-3}$	$1.18(0.18) \cdot 10^{-2}$	$8.82(6.41) \cdot 10^{-3}$	$2.05(0.36) \cdot 10^{-2}$	
			3	$2.48(0.45) \cdot 10^{-5}$	$2.32(0.45) \cdot 10^{-5}$	$3.78(0.88) \cdot 10^{-5}$	$1.21(0.16) \cdot 10^{-2}$	
	1	1	1	$4.94(0.34) \cdot 10^{-2}$	$1.84(0.04) \cdot 10^{-1}$	$5.32(0.35) \cdot 10^{-2}$	$1.54(0.03)$	
			2	$1.01(0.07) \cdot 10^{-1}$	$1.47(0.18) \cdot 10^{-1}$	$1.01(0.06) \cdot 10^{-1}$	$1.52(0.04)$	
			3	$2.02(0.17) \cdot 10^{-2}$	$2.22(0.17) \cdot 10^{-2}$	$2.14(0.18) \cdot 10^{-2}$	$1.52(0.04)$	
		2	1	$2.69(0.28) \cdot 10^{-2}$	$1.13(0.09) \cdot 10^{-1}$	$2.98(0.22) \cdot 10^{-2}$	$0.90(0.04)$	
			2	$6.99(0.43) \cdot 10^{-2}$	$1.21(0.13) \cdot 10^{-1}$	$7.20(0.39) \cdot 10^{-2}$	$0.89(0.04)$	
			3	$6.71(2.18) \cdot 10^{-4}$	$6.62(2.20) \cdot 10^{-4}$	$1.13(0.30) \cdot 10^{-3}$	$0.87(0.04)$	
	10%	.01	1	1	$5.40(0.90) \cdot 10^{-3}$	$2.99(0.56) \cdot 10^{-2}$	$8.85(1.71) \cdot 10^{-3}$	$0.66(0.01)$
				2	$4.19(0.51) \cdot 10^{-3}$	$1.49(0.22) \cdot 10^{-2}$	$8.45(3.79) \cdot 10^{-3}$	$0.65(0.00)$
				3	$6.56(0.69) \cdot 10^{-4}$	$3.75(0.07) \cdot 10^{-3}$	$1.67(0.13) \cdot 10^{-3}$	$0.65(0.00)$
			2	1	$5.27(1.41) \cdot 10^{-3}$	$2.56(0.50) \cdot 10^{-2}$	$7.65(1.44) \cdot 10^{-3}$	$1.75(0.95) \cdot 10^{-2}$
				2	$4.21(0.55) \cdot 10^{-3}$	$1.16(0.22) \cdot 10^{-2}$	$8.30(6.49) \cdot 10^{-3}$	$1.14(0.28) \cdot 10^{-2}$
				3	$6.44(0.76) \cdot 10^{-4}$	$6.35(0.73) \cdot 10^{-4}$	$6.79(0.89) \cdot 10^{-4}$	$3.89(0.99) \cdot 10^{-3}$
0.1		1	1	$6.02(0.95) \cdot 10^{-3}$	$2.95(0.59) \cdot 10^{-2}$	$9.51(1.34) \cdot 10^{-3}$	$0.67(0.01)$	
			2	$5.40(0.53) \cdot 10^{-3}$	$1.65(2.04) \cdot 10^{-2}$	$9.50(2.29) \cdot 10^{-3}$	$0.66(0.00)$	
			3	$8.44(0.61) \cdot 10^{-4}$	$3.93(0.07) \cdot 10^{-3}$	$1.90(0.12) \cdot 10^{-3}$	$0.65(0.00)$	
		2	1	$6.07(1.30) \cdot 10^{-3}$	$2.76(0.58) \cdot 10^{-2}$	$8.25(1.30) \cdot 10^{-3}$	$2.67(0.73) \cdot 10^{-2}$	
			2	$5.45(0.60) \cdot 10^{-3}$	$1.32(0.19) \cdot 10^{-2}$	$8.29(2.07) \cdot 10^{-3}$	$2.07(0.27) \cdot 10^{-2}$	
			3	$6.46(0.59) \cdot 10^{-4}$	$6.42(0.56) \cdot 10^{-4}$	$6.95(0.80) \cdot 10^{-4}$	$1.32(0.14) \cdot 10^{-2}$	
1		1	1	$5.19(0.34) \cdot 10^{-2}$	$1.82(0.34) \cdot 10^{-1}$	$5.63(0.40) \cdot 10^{-2}$	$1.60(0.04)$	
			2	$1.01(0.07) \cdot 10^{-1}$	$1.46(0.16) \cdot 10^{-1}$	$1.01(0.06) \cdot 10^{-1}$	$1.60(0.04)$	
			3	$2.07(0.16) \cdot 10^{-2}$	$2.27(0.16) \cdot 10^{-2}$	$2.21(0.17) \cdot 10^{-2}$	$1.59(0.03)$	
		2	1	$2.95(0.29) \cdot 10^{-2}$	$1.14(0.09) \cdot 10^{-1}$	$3.34(0.25) \cdot 10^{-2}$	$0.96(0.04)$	
			2	$7.17(0.49) \cdot 10^{-2}$	$1.22(0.12) \cdot 10^{-1}$	$7.29(0.36) \cdot 10^{-2}$	$0.96(0.04)$	
			3	$1.28(0.27) \cdot 10^{-3}$	$1.25(0.22) \cdot 10^{-3}$	$1.82(0.36) \cdot 10^{-3}$	$0.95(0.04)$	

S3.2 Simulation 2

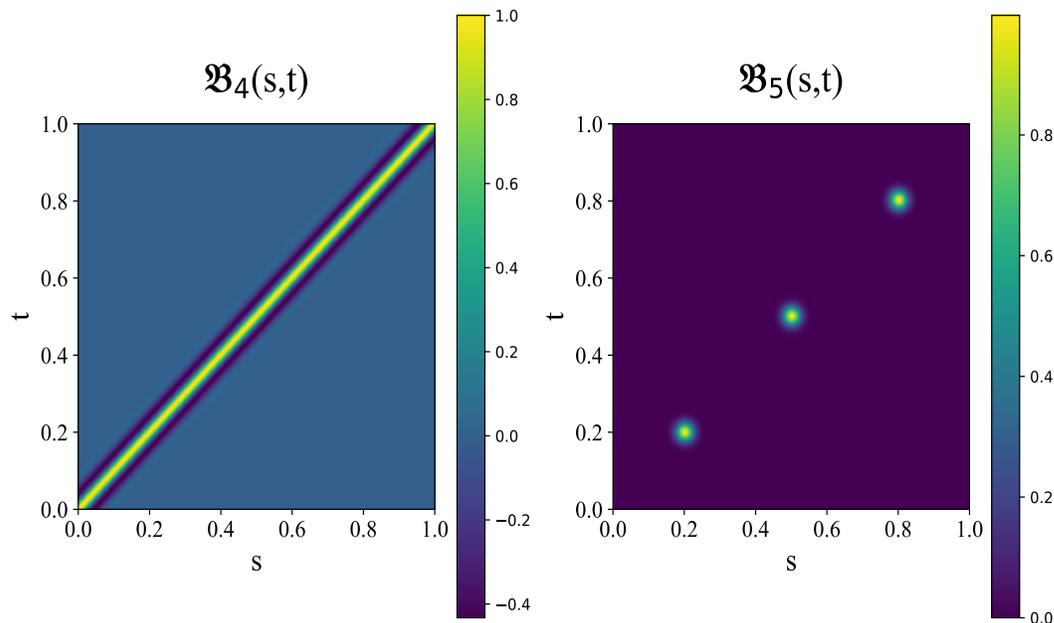


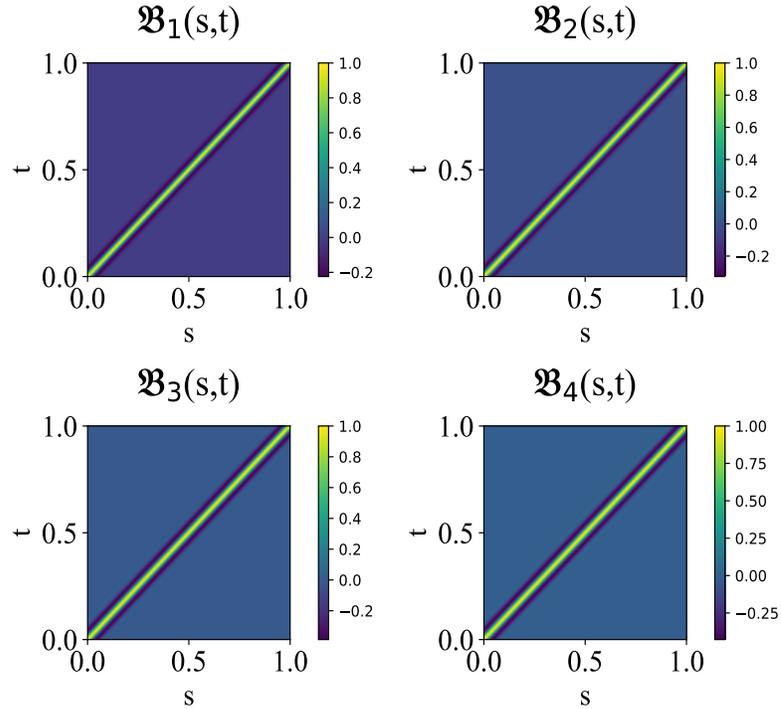
Figure S9: Image plots of the coefficient functions $\mathfrak{B}_4(s, t)$ and $\mathfrak{B}_5(s, t)$ in Simulation 2.

Table S6: Percentage (%) of the selected orders (d_1, d_2) of derivatives in the *fof.deriv* method from 100 replicates in Simulation 2.

T	$2^9 = 512$				$2^{10} = 1024$			
	$\mathfrak{B}_4(s, t)$		$\mathfrak{B}_5(s, t)$		$\mathfrak{B}_4(s, t)$		$\mathfrak{B}_5(s, t)$	
σ	(d_1, d_2)	%						
0.01	(1,0)	96	(1,1)	100	(1,0)	95	(1,1)	100
	(0,0)	2			(0,0)	3		
	(2,0)	2			(2,0)	2		
0.1	(0,0)	99	(0,0)	100	(0,0)	94	(0,0)	89
	(1,0)	1			(1,0)	6	(1,0)	11
0	(0,0)	100	(0,0)	100	(0,0)	100	(0,0)	100

Table S7: Average (and standard deviation) of running time for Simulation 2.

T	$2^9 = 512$					$2^{10} = 1024$			
σ	\mathfrak{B}	<i>fof.deriv</i>	<i>sSigComp</i>	<i>wSigComp</i>	<i>fdapace</i>	<i>fof.deriv</i>	<i>sSigComp</i>	<i>wSigComp</i>	<i>fdapace</i>
.01	4	53.2(1.7)	2.8(0.1)	235.2(11.4)	5.55(0.33)	138.2(2.7)	10.8(0.2)	652.8(28.8)	28.5(1.7)
	5	16.6(0.4)	2.1(0.1)	42.5(2.0)	5.52(0.26)	73.3(1.7)	9.6(0.3)	280.5(16.3)	28.5(0.9)
.1	4	53.1(1.1)	2.7(0.1)	234.2(11.4)	5.51(0.25)	138.5(2.9)	10.7(0.3)	640.2(25.8)	29.3(1.9)
	5	16.7(0.4)	2.1(0.1)	43.6(2.0)	5.51(0.27)	73.5(1.6)	9.5(0.2)	281.1(14.3)	28.8(1.5)
1	4	67.7(1.2)	2.8(0.1)	419.5(27.7)	5.50(0.25)	163.8(3.8)	10.9(0.3)	955.6(56.3)	28.6(0.8)
	5	68.2(1.3)	2.7(0.1)	234.8(43.2)	5.53(0.26)	165.8(3.4)	10.7(0.3)	613.9(117.6)	28.6(0.9)

Figure S10: Image plots of the coefficients $\mathfrak{B}_j(s, t)$ for $1 \leq j \leq 4$ of Type 1 in Simulation 3.

S3.3 Simulation 3

In this simulation, we consider the model (1.1) with four functional predictors.

- (1). We generate the functional predictors $(X_1(s), \dots, X_4(s))$ as follows. Let $W_\ell(s)$, $0 \leq s \leq 1$ and $1 \leq \ell \leq m+3$, be independent Gaussian processes with covariance function $\exp\{-2500(s-s')^2\}$, where

$m = 4$ or 6 . We take $X_j(s) = \sum_{j'=0}^{m-1} W_{j+j'}(s)/\sqrt{m}$ for $1 \leq j \leq m$. Sample curves of each $W_\ell(s)$ are spiky, so are those of the $X_j(s)$. Moreover, a larger value of m implies a stronger correlation between the four functional predictors.

(2). We set $\mathfrak{U}(t) = 0$, and consider two types of $(\mathfrak{B}_1(s, t), \dots, \mathfrak{B}_4(s, t))$. The first type (Figure S10) has $\mathfrak{B}_j(s, t) = e^{-800(s-t)^2/\sqrt{j}} \cos(20\pi(s-t))$ for $1 \leq j \leq 4$, each of which have a ridge along the diagonal line. The second type (Figure S11) has $\mathfrak{B}_j(s, t) = \sum_{j'=1}^2 \xi_{jj'}(s)\zeta_{jj'}(t)$, where $\xi_{jj'}(s)$ and $\zeta_{jj'}(s)$, $1 \leq j \leq 4$ and $1 \leq j' \leq 2$, are independently generated from the following AR(1) time series model,

$$Z(s_j) = 0.1^{8/T} Z(s_{j-1}) + \epsilon_{j-1}, \quad 2 \leq j \leq T. \quad (\text{S3.1})$$

In (S3.1), $0 = s_1 < s_1 < \dots < s_T = 1$ are $T = 512$ equally spaced observation time points on $[0,1]$, $Z(s_1)$ and ϵ_j , $1 \leq j \leq T - 1$, are all independent standard normal random variables. Figure S12 shows four curves generated from this AR(1) model.

(3). The random error $\epsilon(t)$ is generated in the same way as in Simulations 1 and 2 with three noise levels, $\sigma = 0.01, 0.1, 1$. As before, when generating the response, we scale the $\mathfrak{B}_j(s, t)$ with the same scaling factor such that the signal to noise ratio is 1 when $\sigma = 1$.

(4). All samples curves are observed at 2^9 equally spaced points in $[0, 1]$.

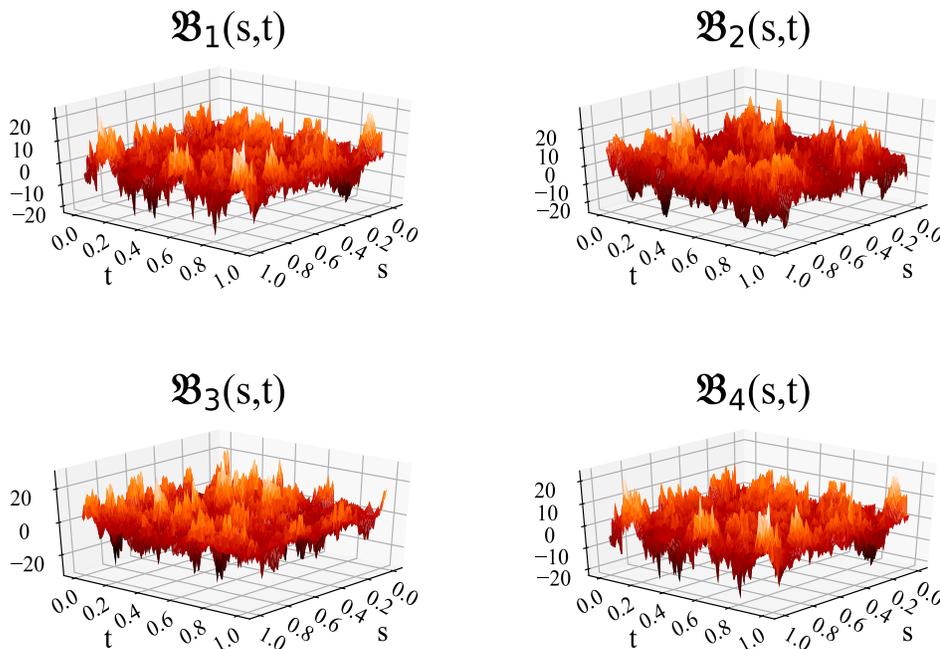


Figure S11: Coefficients $\mathfrak{B}_j(s, t)$ for $1 \leq j \leq 4$ of Type 2 in Simulation 3.

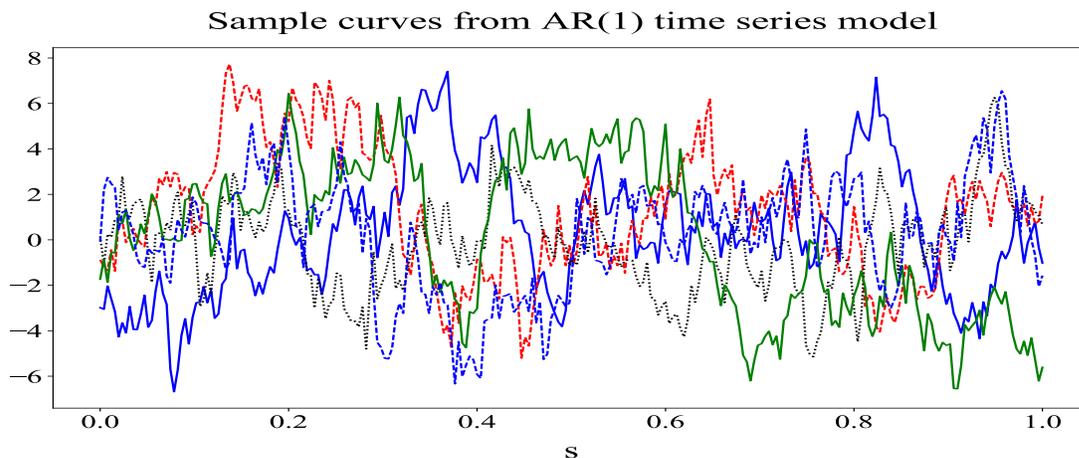


Figure S12: Five time series curves generated from the AR(1) model (S3.1) in Simulation 3.

As summarized in Table S8, in all settings, the new method *fof.deriv* has the lowest average MISEE. All methods have larger MISEEs when the correlation between the four functional predictors gets

stronger (larger m). The $wSigComp$ performs better than $sSigComp$ for the first type of $\mathfrak{B}_j(s, t)$, which decays to zero when $|s - t|$ gets larger at different exponential rate for different $1 \leq j \leq m$, and hence introduces certain amount of sparsity in wavelet coefficients. For the second type of $\mathfrak{B}_j(s, t)$, $wSigComp$ does not show advantage over $sSigComp$ when $\sigma = 0.01, 0.1$. The $fdapace$ has the second lowest MISEEs for the first type of $\mathfrak{B}_j(s, t)$ when $\sigma = 0.01$ or 0.1 , but the highest in other settings.

Table S8: Average (and standard deviation) of MISEE from 100 replicates for Simulation 3.

m	σ	\mathfrak{B}	<i>fof.deriv</i>	<i>sSigComp</i>	<i>wSigComp</i>	<i>fdapace</i>
4	0.01	1	0.032(0.003)	0.192(0.009)	0.133(0.011)	0.034(0.003)
		2	0.052(0.010)	0.078(0.014)	0.087(0.011)	0.144(0.016)
	0.1	1	0.033(0.003)	0.193(0.009)	0.135(0.013)	0.044(0.003)
		2	0.054(0.012)	0.081(0.017)	0.089(0.013)	0.153(0.016)
	1	1	0.145(0.006)	0.323(0.011)	0.190(0.009)	1.028(0.024)
		2	0.131(0.013)	0.165(0.023)	0.154(0.014)	1.140(0.030)
6	0.01	1	0.054(0.004)	0.223(0.010)	0.147(0.011)	0.059(0.004)
		2	0.068(0.013)	0.095(0.019)	0.108(0.018)	0.178(0.019)
	0.1	1	0.055(0.003)	0.224(0.009)	0.149(0.012)	0.070(0.004)
		2	0.069(0.014)	0.097(0.023)	0.110(0.016)	0.185(0.017)
	1	1	0.181(0.008)	0.366(0.015)	0.202(0.008)	1.033(0.022)
		2	0.145(0.016)	0.185(0.027)	0.169(0.018)	1.147(0.031)

Table S9 displays the frequencies of the selected order of derivatives by *fof.deriv*. With $\mathfrak{B}_j(s, t)$, $1 \leq j \leq 4$, generated in the same mechanism for each type, the pattern of selected (d_1, d_2) does not change with m . For example, for the first type of $\mathfrak{B}_j(s, t)$, $d_1 = d_2 = 2$ is the most frequently (97%) selected order when $\sigma = 0.01$, and $(2, 0)$ is selected 100% when σ gets larger, for both $m = 4$ and 6 . For the second type of $\mathfrak{B}_j(s, t)$, $(1, 2)$ is the most frequently selected order when $\sigma < 1$, and $(1, 0)$ is selected 100% when $\sigma = 1$, for both $m = 4$ and 6 . We again observe that for spiky functional coefficients, positive orders of derivatives are selected when the noise is small, and zero order derivative can be chosen when the noise is large as local features can be masked by large noises. The running time summarized in Table S10 shows similar patterns as in Simulations 1 and 2.

Table S9: Percentage (%) of the selected orders (d_1, d_2) of derivatives in the *fof.deriv* method from 100 replicates in Simulation 3.

	Type 1 \mathfrak{B}_j				Type 2 \mathfrak{B}_j			
	$m = 4$		$m = 6$		$m = 4$		$m = 6$	
σ	(d_1, d_2)	%	(d_1, d_2)	%	(d_1, d_2)	%	(d_1, d_2)	%
.01	(2,2)	97	(2,2)	97	(1,2)	90	(1,2)	96
	(2,1)	3	(2,1)	3	(1,1)	10	(1,1)	4
0.1	(2,0)	100	(2,0)	100	(1,2)	90	(1,2)	96
					(1,1)	10	(1,1)	4
1	(2,0)	100	(2,0)	100	(1,0)	100	(1,0)	100

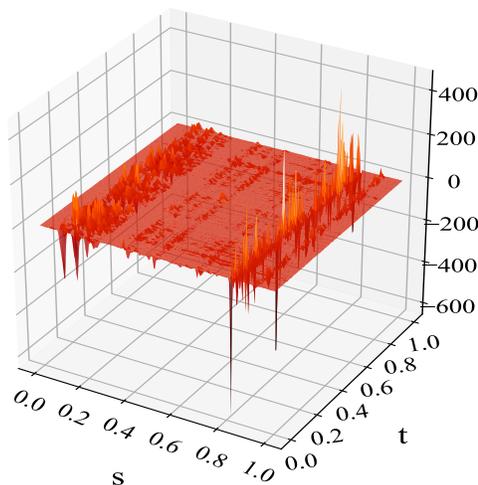
Table S10: Average (and standard deviation) of running time for Simulation 3.

σ	\mathfrak{B}	Type 1 $X(s)$				Type 2 $X(s)$			
		<i>fof.deriv</i>	<i>sSigComp</i>	<i>wSigComp</i>	<i>fdapace</i>	<i>fof.deriv</i>	<i>sSigComp</i>	<i>wSigComp</i>	<i>fdapace</i>
.01	1	205.1(8.2)	11.8(0.5)	545.4(39.7)	68.1(16.9)	205.9(8.1)	12.0(0.5)	555.6(36.3)	67.7(16.1)
	2	51.1(2.4)	5.8(0.3)	117.8(9.6)	67.5(16.2)	52.4(2.5)	5.9(0.3)	122.1(8.8)	67.6(16.4)
.1	1	206.4(8.3)	11.8(0.6)	547.5(38.7)	68.0(16.7)	207.6(8.0)	12.0(0.5)	559.5(40.9)	67.8(15.9)
	2	52.0(3.0)	5.8(0.3)	118.9(9.8)	67.7(16.5)	53.2(2.7)	5.9(0.3)	123.1(9.7)	67.7(16.3)
1	1	256.9(10.6)	13.1(0.6)	976.5(72.9)	67.6(16.2)	258.6(9.1)	13.3(0.5)	1016.2(64.7)	67.6(16.1)
	2	240.5(8.6)	12.1(0.5)	624.2(78.3)	67.8(16.8)	242.6(8.5)	12.3(0.5)	666.8(78.4)	67.5(16.4)

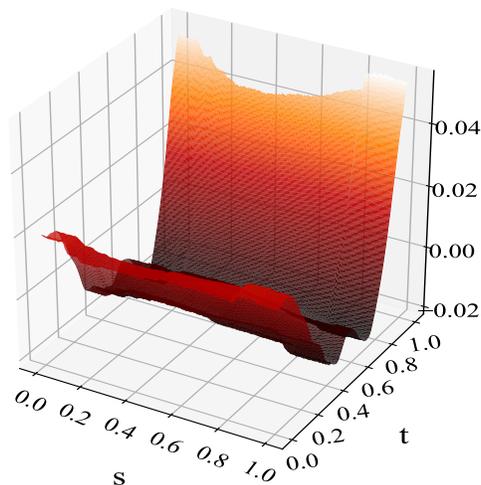
S3.4 Real data

This section provides the estimated coefficients and the corresponding auxiliary smooth functions for Models 2 ~ 7 of the HPLC-PDA data.

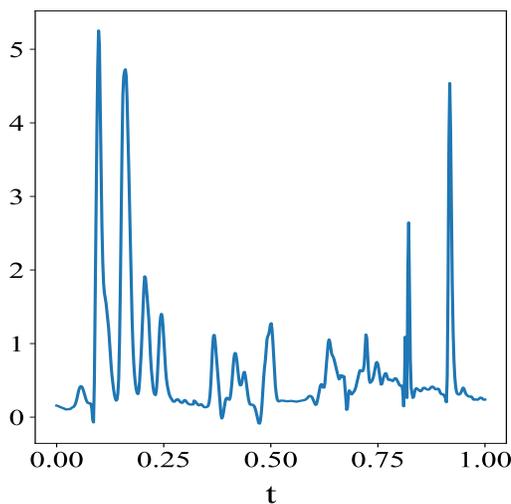
Estimated $\hat{\mathfrak{B}}(s,t)=D_s^2 D_t \hat{\beta}(s,t)$



Estimated $\hat{\beta}(s,t)$



Estimated intercept $\hat{\mathfrak{U}}(t)=D\hat{\mu}(t)$



Estimated $\hat{\mu}(t)$

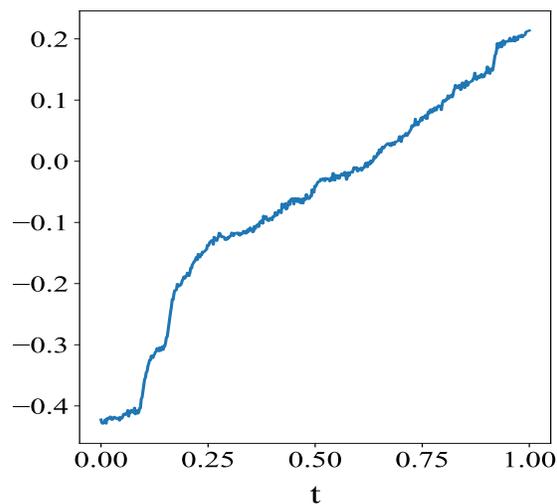
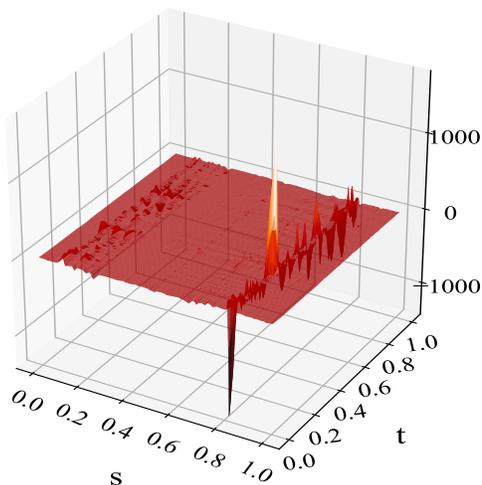
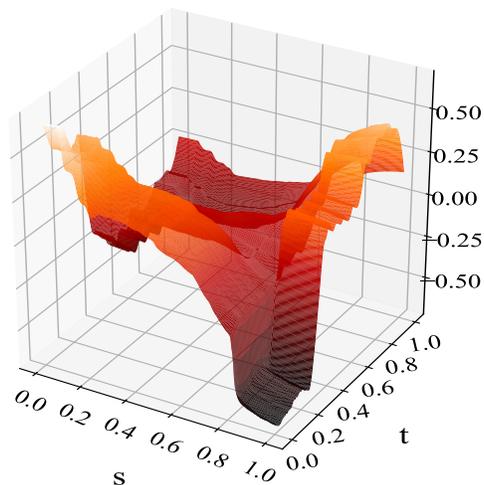


Figure S13: Estimated functions from the **second** model of the HPLC-PDA data, where $X(s)$ and $Y(t)$ are the chromatogram curves at wavelengths 332 and 368 nm, respectively. Top: the estimated coefficient surface $\hat{\mathfrak{B}}(s,t)$ (left) and its corresponding auxiliary smooth function $\hat{\beta}(s,t)$ (right) where $\hat{\mathfrak{B}}(s,t) = D_s^2 D_t \hat{\beta}(s,t)$; Bottom: the estimated intercept function $\hat{\mathfrak{U}}(t)$ (left) and the corresponding auxiliary function $\hat{\mu}(t)$ (right) with $\hat{\mathfrak{U}}(t) = D\hat{\mu}(t)$.

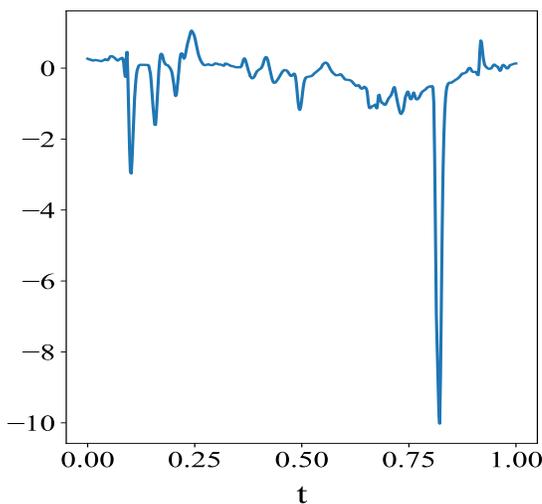
Estimated $\hat{\mathfrak{B}}(s,t)=D_s D_t \hat{\beta}(s,t)$



Estimated $\hat{\beta}(s,t)$



Estimated intercept $\hat{\mathfrak{U}}(t)=D\hat{\mu}(t)$



Estimated $\hat{\mu}(t)$

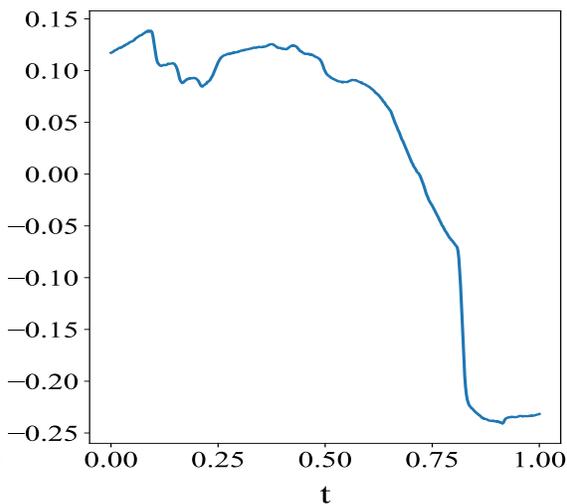
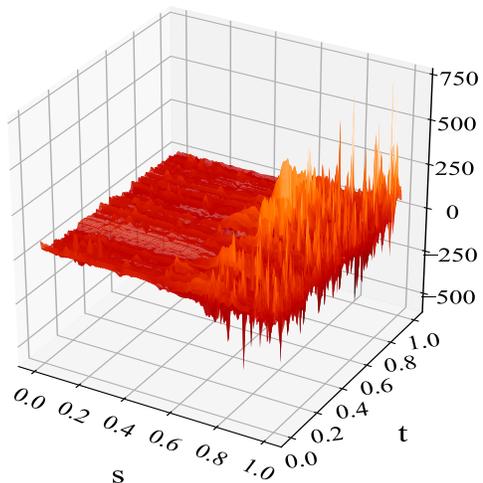
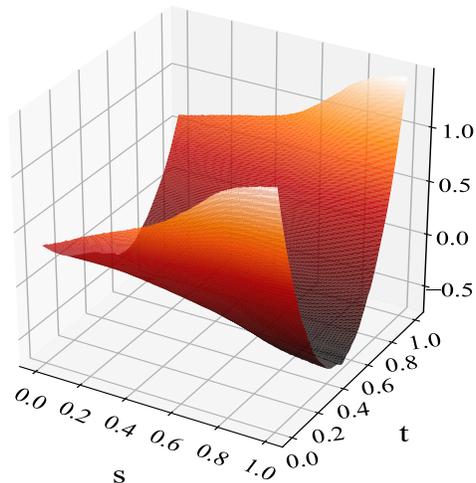


Figure S14: Estimated functions from the **third** model of the HPLC-PDA data, where $X(s)$ and $Y(t)$ are the chromatogram curves at wavelengths 368 and 404 nm, respectively. Top: the estimated coefficient surface $\hat{\mathfrak{B}}(s,t)$ (left) and its corresponding auxiliary smooth function $\hat{\beta}(s,t)$ (right) where $\hat{\mathfrak{B}}(s,t) = D_s D_t \hat{\beta}(s,t)$; Bottom: the estimated intercept function $\hat{\mathfrak{U}}(t)$ (left) and the corresponding auxiliary function $\hat{\mu}(t)$ (right) with $\hat{\mathfrak{U}}(t) = D\hat{\mu}(t)$.

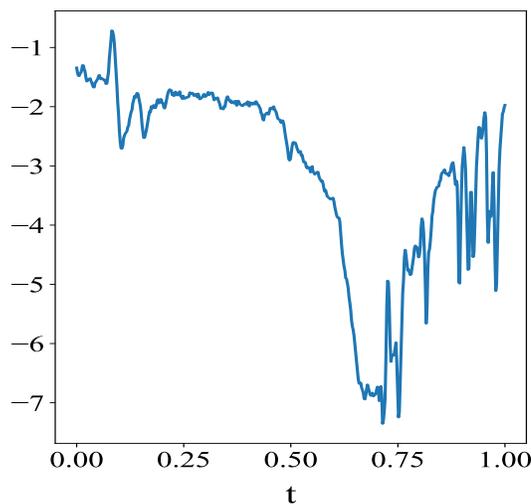
Estimated $\hat{\mathfrak{B}}(s,t)=D_s^2 D_t \hat{\beta}(s,t)$



Estimated $\hat{\beta}(s,t)$



Estimated intercept $\hat{\mathfrak{U}}(t)=D\hat{\mu}(t)$



Estimated $\hat{\mu}(t)$

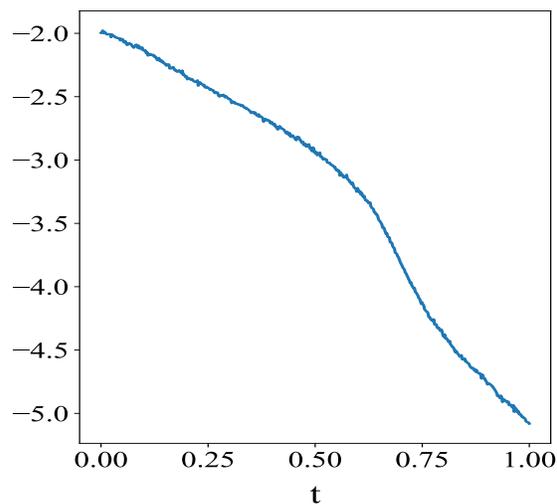
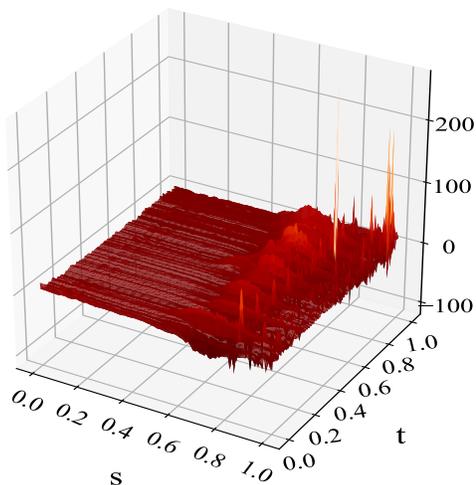
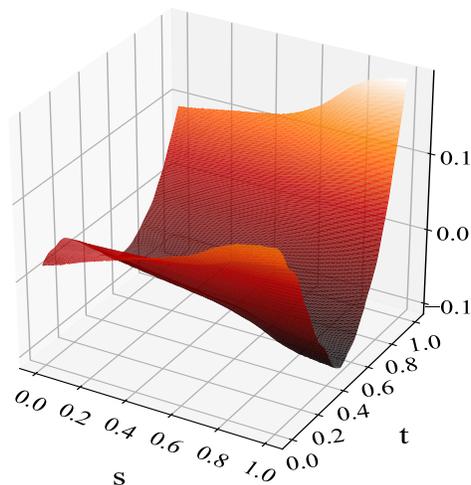


Figure S15: Estimated functions from the **fourth** model of the HPLC-PDA data, where $X(s)$ and $Y(t)$ are the chromatogram curves at wavelengths 404 and 440 nm, respectively. Top: the estimated coefficient surface $\hat{\mathfrak{B}}(s,t)$ (left) and its corresponding auxiliary smooth function $\hat{\beta}(s,t)$ (right) where $\hat{\mathfrak{B}}(s,t) = D_s^2 D_t \hat{\beta}(s,t)$; Bottom: the estimated intercept function $\hat{\mathfrak{U}}(t)$ (left) and the corresponding auxiliary function $\hat{\mu}(t)$ (right) with $\hat{\mathfrak{U}}(t) = D\hat{\mu}(t)$.

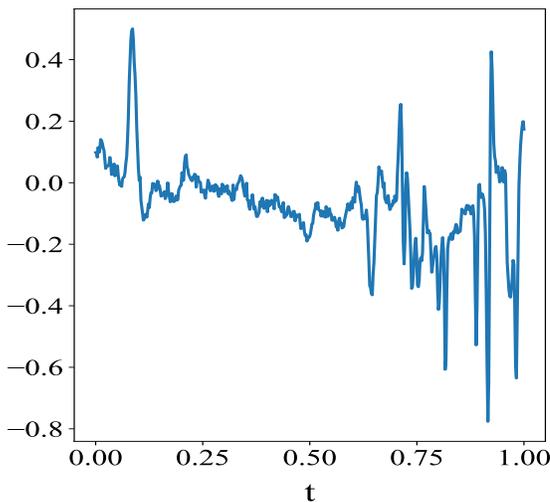
Estimated $\hat{\mathfrak{B}}(s,t)=D_s^2 D_t \hat{\beta}(s,t)$



Estimated $\hat{\beta}(s,t)$



Estimated intercept $\hat{\mathfrak{U}}(t)=D\hat{\mu}(t)$



Estimated $\hat{\mu}(t)$

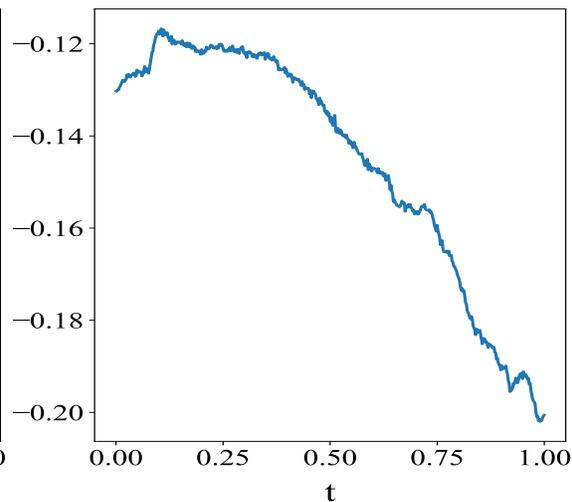
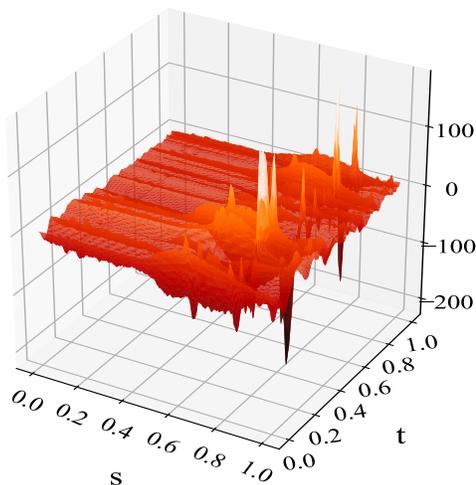
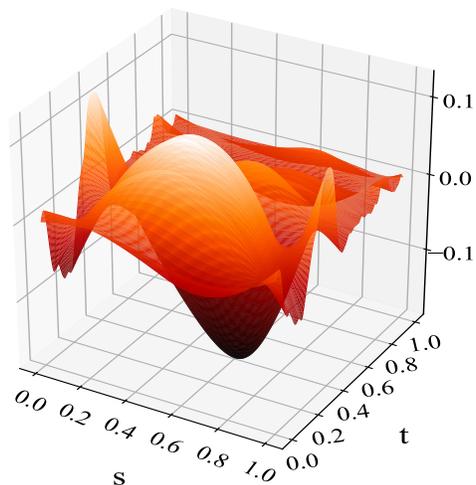


Figure S16: Estimated functions from the **fifth** model of the HPLC-PDA data, where $X(s)$ and $Y(t)$ are the chromatogram curves at wavelengths 440 and 476 nm, respectively. Top: the estimated coefficient surface $\hat{\mathfrak{B}}(s,t)$ (left) and its corresponding auxiliary smooth function $\hat{\beta}(s,t)$ (right) where $\hat{\mathfrak{B}}(s,t) = D_s^2 D_t \hat{\beta}(s,t)$; Bottom: the estimated intercept function $\hat{\mathfrak{U}}(t)$ (left) and the corresponding auxiliary function $\hat{\mu}(t)$ (right) with $\hat{\mathfrak{U}}(t) = D\hat{\mu}(t)$.

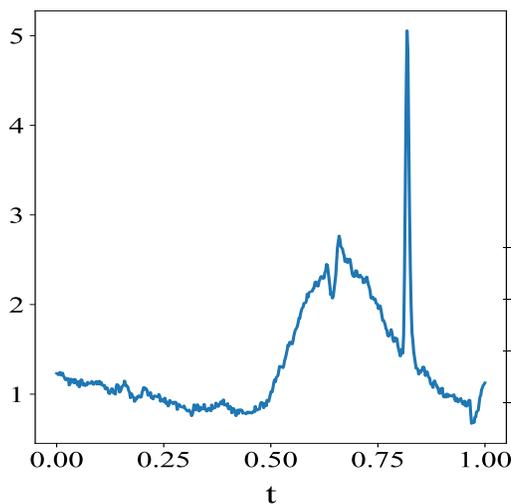
Estimated $\hat{\mathfrak{B}}(s,t)=D_s D_t^2 \hat{\beta}(s,t)$



Estimated $\hat{\beta}(s,t)$



Estimated intercept $\hat{\mathfrak{U}}(t)=D^2 \hat{\mu}(t)$



Estimated $\hat{\mu}(t)$

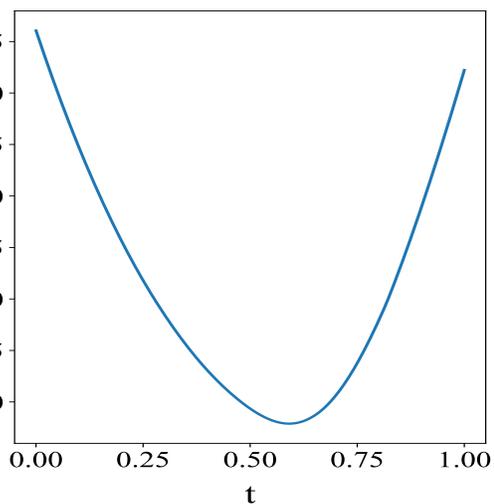
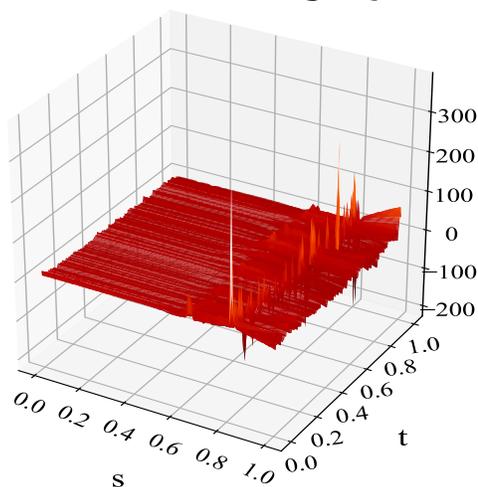
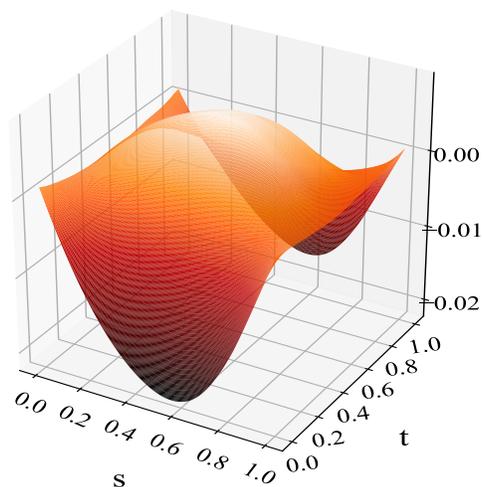


Figure S17: Estimated functions from the **sixth** model of the HPLC-PDA data, where $X(s)$ and $Y(t)$ are the chromatogram curves at wavelengths 476 and 512 nm, respectively. Top: the estimated coefficient surface $\hat{\mathfrak{B}}(s,t)$ (left) and its corresponding auxiliary smooth function $\hat{\beta}(s,t)$ (right) where $\hat{\mathfrak{B}}(s,t) = D_s D_t^2 \hat{\beta}(s,t)$; Bottom: the estimated intercept function $\hat{\mathfrak{U}}(t)$ (left) and the corresponding auxiliary function $\hat{\mu}(t)$ (right) with $\hat{\mathfrak{U}}(t) = D^2 \hat{\mu}(t)$.

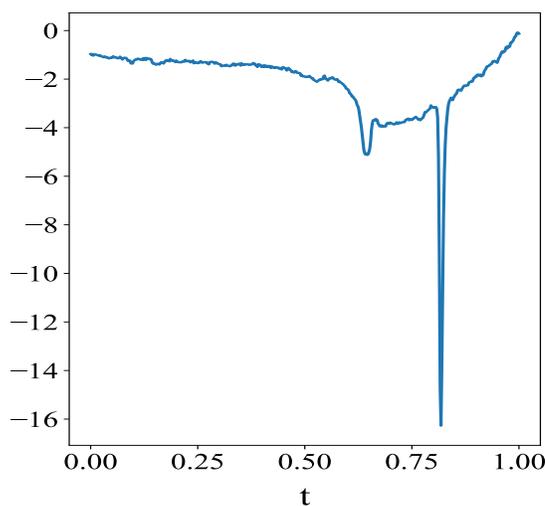
Estimated $\hat{\mathfrak{B}}(s,t)=D_s^2 D_t^2 \hat{\beta}(s,t)$



Estimated $\hat{\beta}(s,t)$



Estimated intercept $\hat{\mathfrak{U}}(t)=D^2 \hat{\mu}(t)$



Estimated $\hat{\mu}(t)$

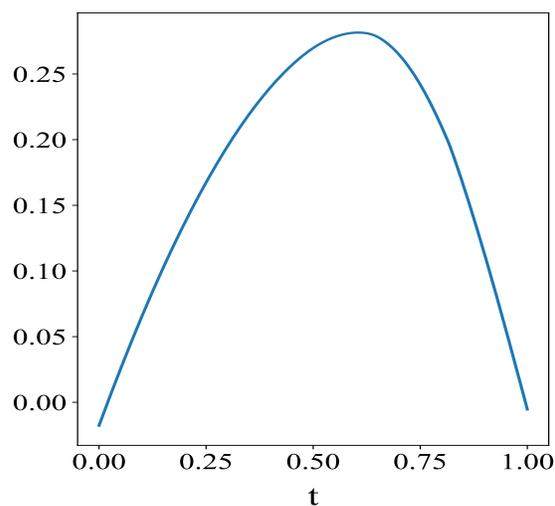


Figure S18: Estimated functions from the **seventh** model of the HPLC-PDA data, where $X(s)$ and $Y(t)$ are the chromatogram curves at wavelengths 512 and 548 nm, respectively. Top: the estimated coefficient surface $\hat{\mathfrak{B}}(s,t)$ (left) and its corresponding auxiliary smooth function $\hat{\beta}(s,t)$ (right) where $\hat{\mathfrak{B}}(s,t) = D_s^2 D_t^2 \hat{\beta}(s,t)$; Bottom: the estimated intercept function $\hat{\mathfrak{U}}(t)$ (left) and the corresponding auxiliary function $\hat{\mu}(t)$ (right) with $\hat{\mathfrak{U}}(t) = D^2 \hat{\mu}(t)$.

References

- Luo, R. and Qi, X. (2017) Function-on-function linear regression by signal compression. *Journal of the American Statistical Association*, **112**, 690–705.