Statistica Sinica: Supplement

Nonlinear function on function additive model with multiple predictor curves

Xin Qi and Ruiyan Luo

Georgia State University

Supplementary Material

This supplementary material is organized as follows. Section S1 includes additional figures and tables in simulations and application. In Section S2, we provides a set of identifibility conditions for the nonlinear function $F(x, s, t)$ in the model (3.2) in the main manuscript. Section S3 provides additional computational details and the choice of tuning parameters. Section S4 provides details of calculating the estimation error for $F(x, s, t)$ in Simulation 1. The proofs of theorems and technical lemmas are provided in Sections S5 and S6, respectively. In this supplementary material, all the labels of equations, figures, tables, sections, and so on, are prefixed with "S.", such as equations: (S2.1), (S2.2), $\cdots$, sections: S.1.1, and so on. The equation/section/$\cdots$ numbers without prefix "S." are for those in the main manuscript.

S1 Additional figures and tables

S.1.1 Figures

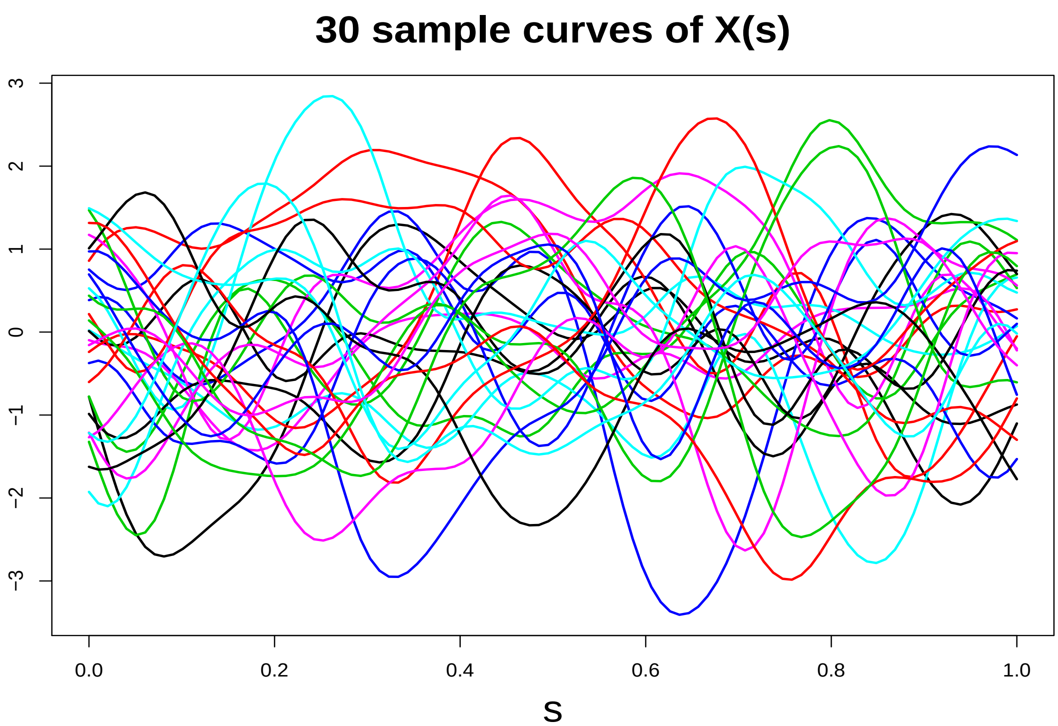

Figure S.1: Thirty sample curves of $X(s)$ in Simulation 1.

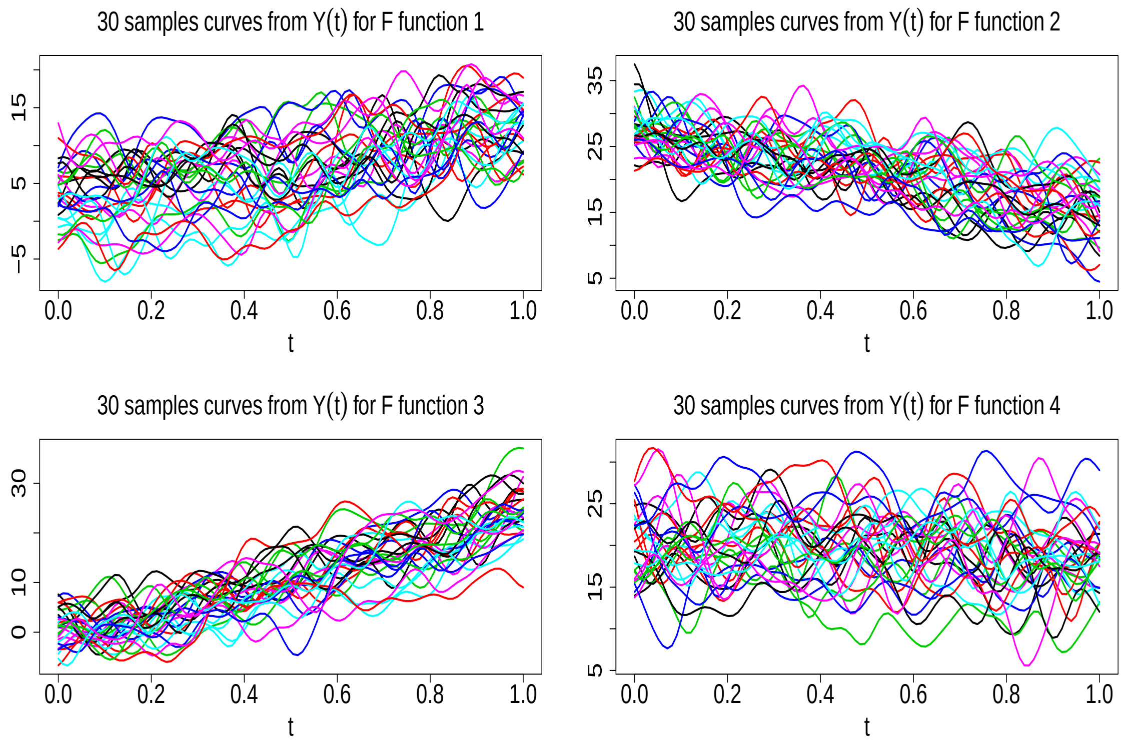

Figure S.2: Thirty sample curves of $Y(t)$ for each of the four $F(x, s, t)$ with $\rho = 0.7$ and $\sigma^2 = 10$ in Simulation 1.

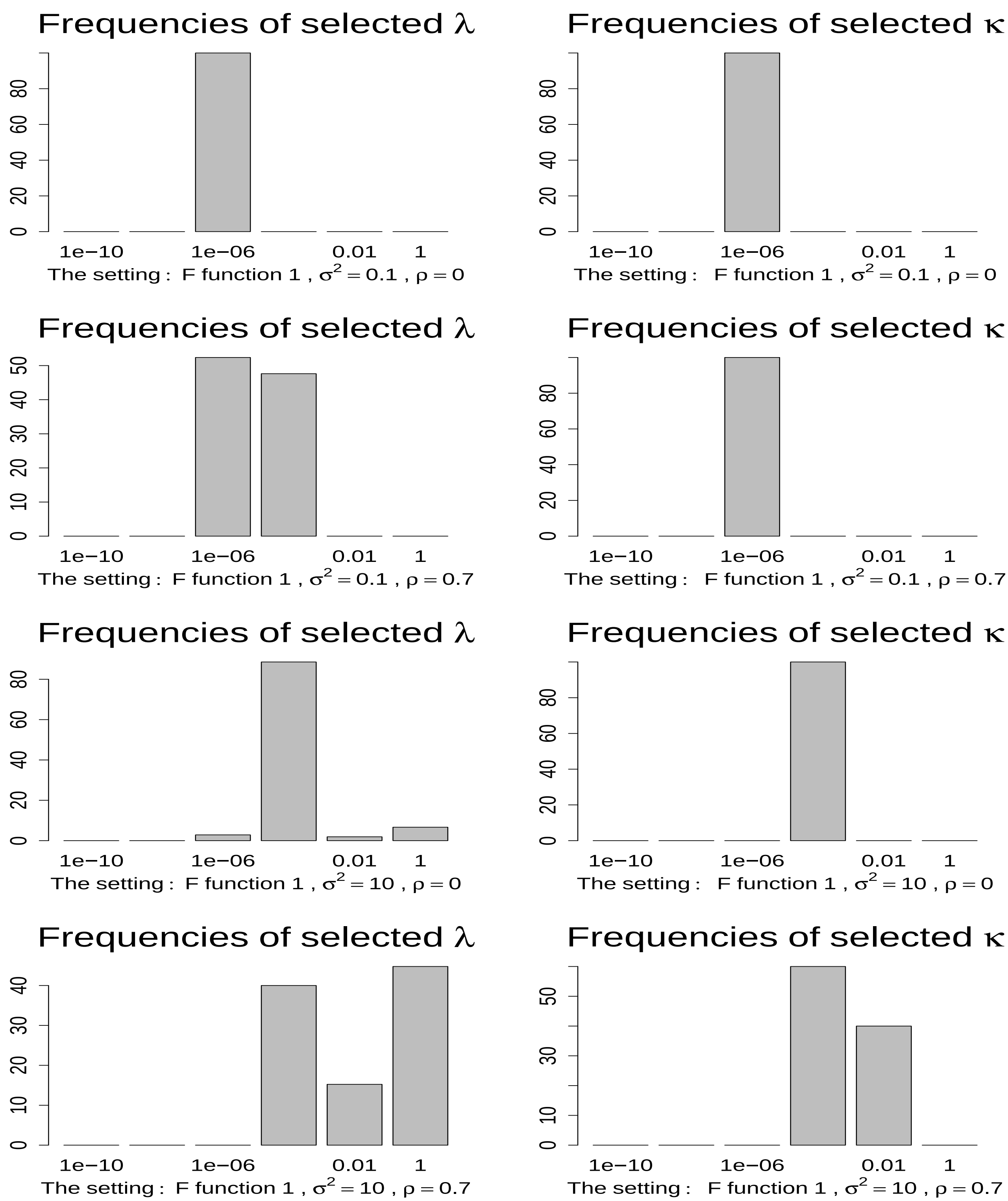

Figure S.3: The frequencies of the selected tuning parameters for the first $F(x, s, t)$ in Simulation 1.

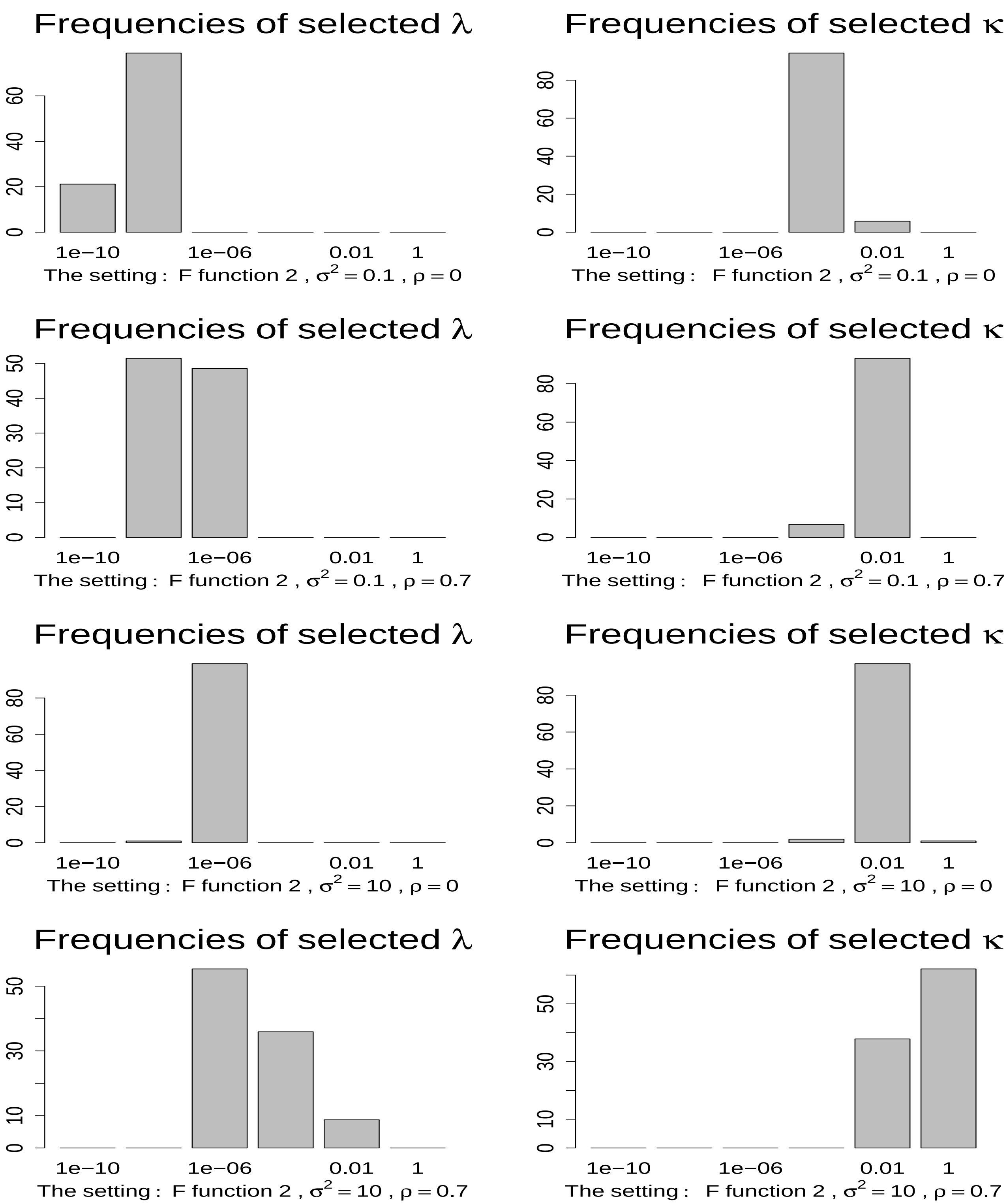

Figure S.4: The frequencies of the selected tuning parameters for the second $F(x, s, t)$ in Simulation 1.

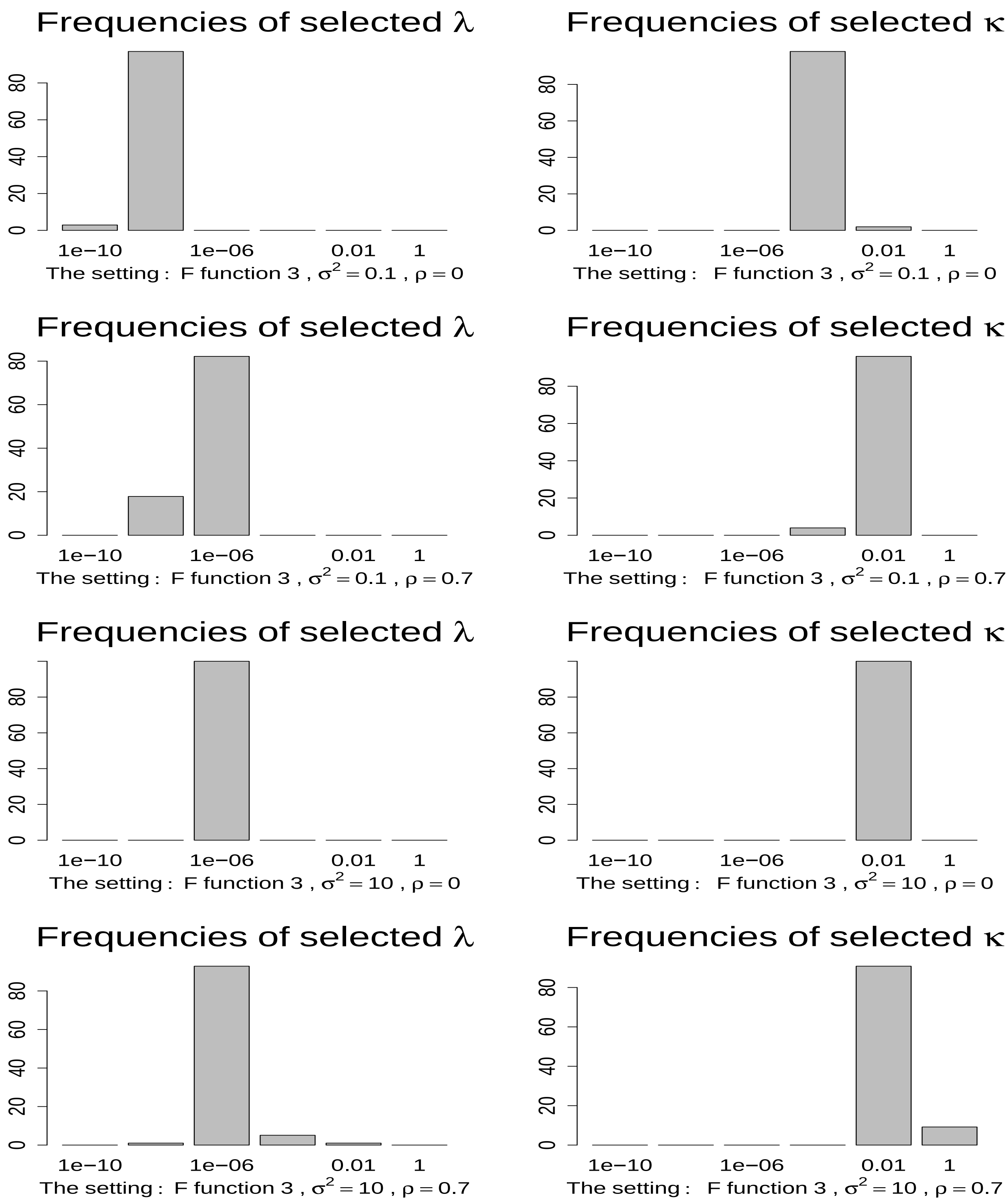

Figure S.5: The frequencies of the selected tuning parameters for the third $F(x, s, t)$ in Simulation 1.

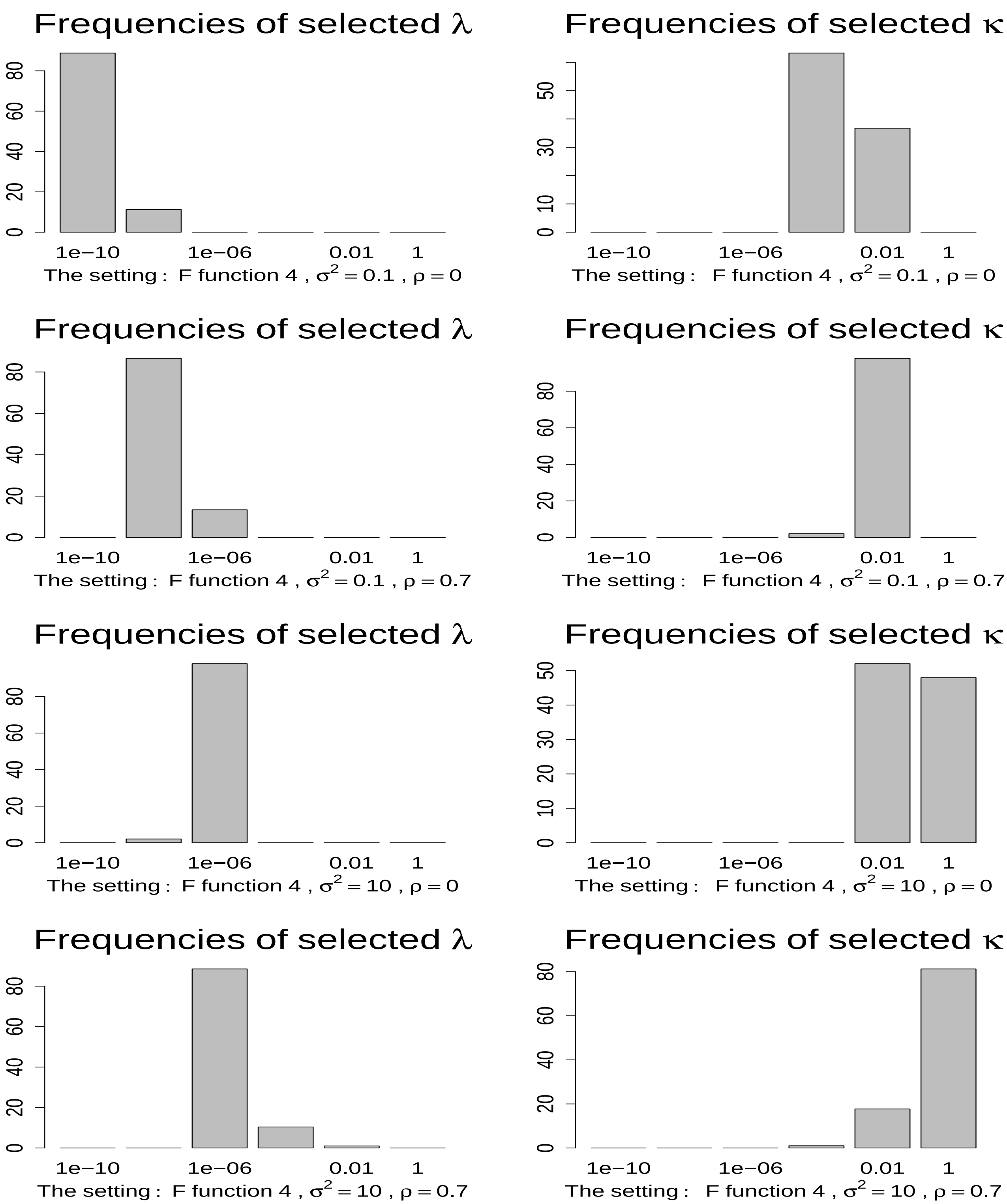

Figure S.6: The frequencies of the selected tuning parameters for the fourth $F(x, s, t)$ in Simulation 1.

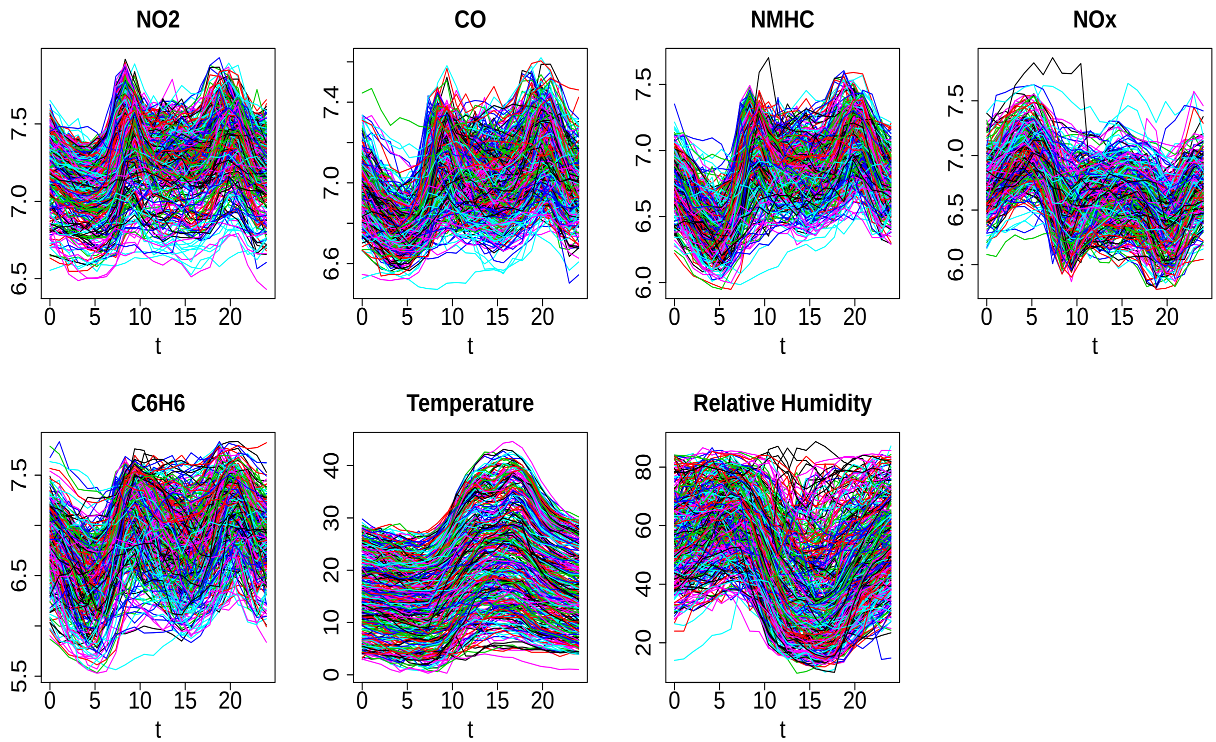

Figure S.7: The 355 sample curves for each of the seven variables in the air quality data.

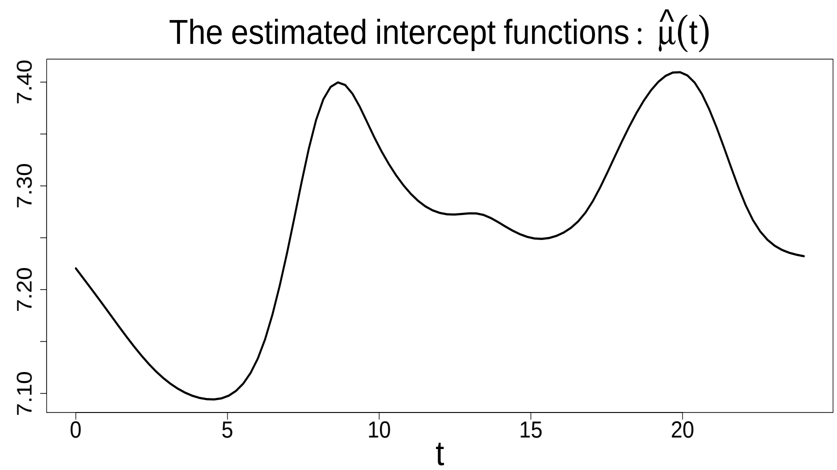

Figure S.8: Estimated intercept function $\widehat{\mu}(t)$ for the daily air quality data.

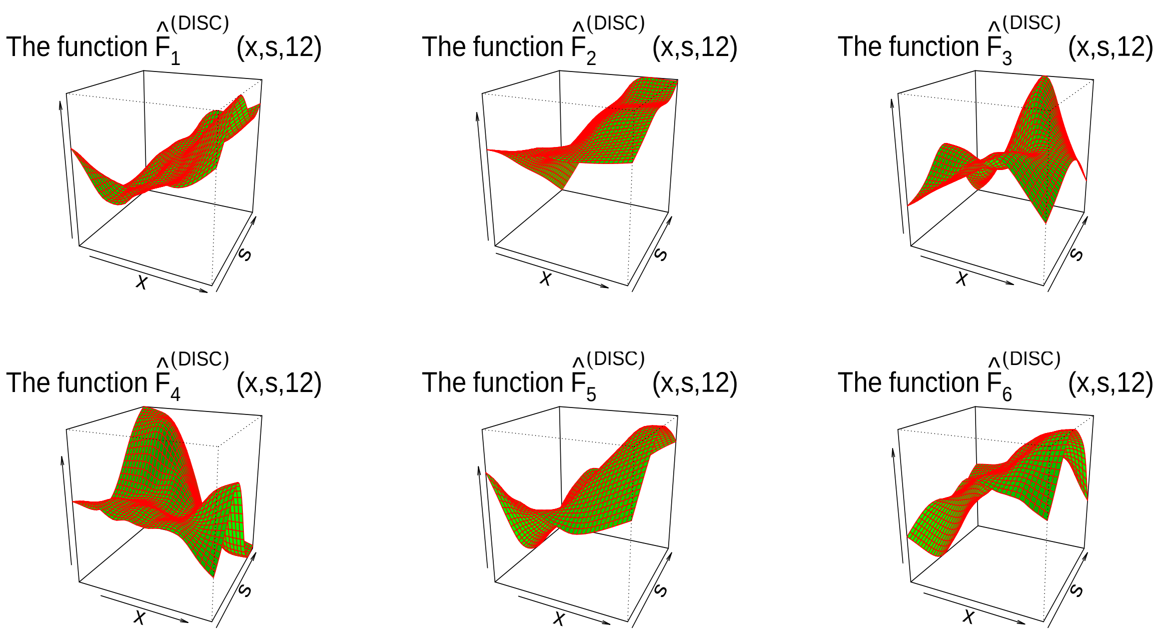

Figure S.9: Estimated functions $\widehat{F}_j^{(DISC)}(x, s, t)$ given $t = 12$, $1 \leq j \leq 6$, for the daily air quality data.

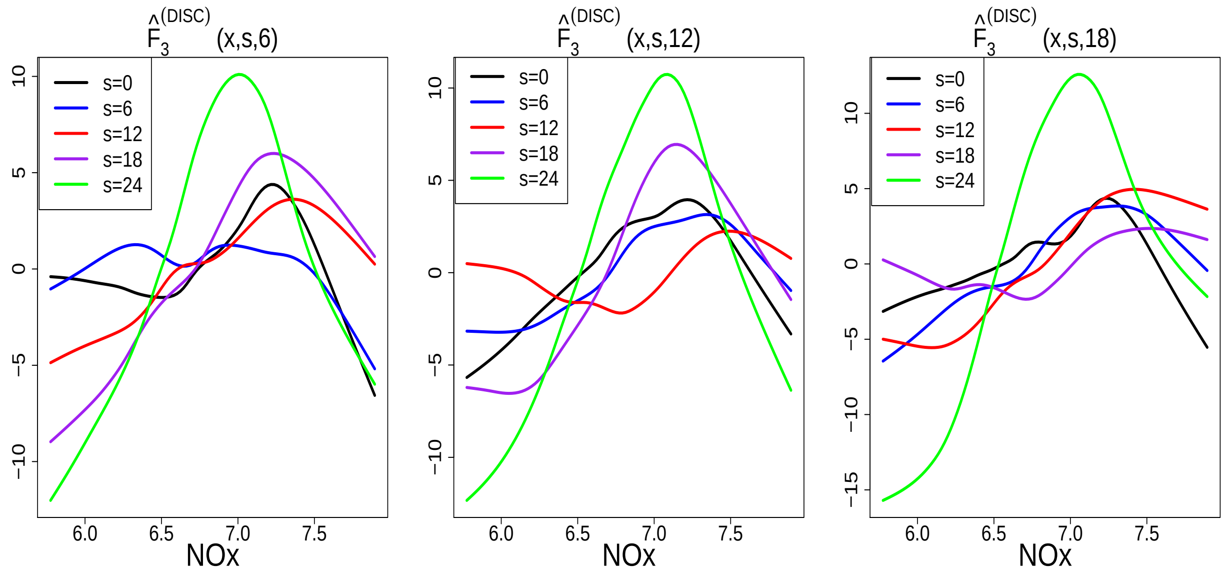

Figure S.10: The curves $\widehat{F}_3^{(DISC)}(x, s, t)$ versus x (predictor NO_x) for different values of s and t in the daily air quality data.

S.1.2 Tables

Table S.1: The averages (and standard deviations) of the number K of the selected components and running time (in seconds) of 100 replicates for Simulation 1. The running time was obtained on a computer cluster with Linux system and Intel(R) Xeon(R) CPU 5160 3.00GHz.

Model	σ^2	ρ	K	time
1	0.1	0	2.95(0.22)	13.717(2.986)
		0.7	2.15(0.36)	13.798(2.880)
	10	0	2.03(0.17)	30.908(6.429)
		0.7	2.05(0.22)	32.480(6.366)
2	0.1	0	2.00(0.00)	10.130(2.187)
		0.7	1.92(0.27)	10.668(2.213)
	10	0	1.82(1.40)	34.958(6.895)
		0.7	2.27(1.18)	35.476(7.023)
3	0.1	0	2.54(0.50)	11.998(2.671)
		0.7	2.75(0.44)	12.298(2.735)
	10	0	2.43(0.62)	28.387(5.461)
		0.7	1.81(1.10)	30.900(6.096)
4	0.1	0	2.18(0.39)	12.916(2.537)
		0.7	2.89(0.31)	12.981(2.688)
	10	0	4.27(2.13)	34.868(6.624)
		0.7	3.03(1.39)	36.010(7.136)

Table S.2: The averages (and standard deviations) of the number K of the selected components and running time (in seconds) of 100 replicates for Simulations 2 and 3.

				Simulation 2		Simulation 3	
γ	ρ_{curve}	σ^2	ρ	K	time	K	time
2	0	0.1	0	2.00(0.00)	39.705(10.048)	3.000(0.000)	65.814(13.785)
			0.7	2.00(0.00)	40.173(10.242)	3.000(0.000)	68.142(13.880)
		10	0	2.14(0.67)	116.470(28.912)	2.480(1.185)	252.554(46.390)
			0.7	4.29(1.31)	117.427(30.222)	5.090(2.357)	255.804(45.438)
	0.7	0.1	0	2.00(0.00)	40.451(10.572)	3.000(0.000)	64.929(12.247)
			0.7	2.01(0.10)	41.710(10.851)	2.742(0.440)	69.394(12.848)
		10	0	2.00(0.00)	111.726(28.976)	2.309(1.054)	244.877(42.778)
			0.7	4.30(1.66)	116.061(29.490)	3.865(1.822)	246.634(39.248)
1.5	0	0.1	0	2.00(0.00)	39.820(10.067)	3.000(0.000)	67.099(13.627)
			0.7	2.00(0.00)	40.654(10.468)	3.000(0.000)	69.223(14.186)
		10	0	2.03(0.30)	115.534(29.441)	2.400(1.239)	252.579(45.187)
			0.7	4.47(1.31)	118.180(30.605)	5.394(2.113)	253.609(45.759)
	0.7	0.1	0	2.00(0.00)	40.176(10.564)	3.000(0.000)	65.098(12.864)
			0.7	2.00(0.00)	41.643(10.596)	2.639(0.483)	69.132(12.452)
		10	0	2.01(0.10)	112.326(28.705)	2.206(0.676)	242.684(42.064)
			0.7	4.23(1.72)	116.242(29.060)	4.432(1.928)	242.397(35.703)

S2 Conditions for the identifiability of the function $F(x, s, t)$

To ensure the identifiability of $F(x, s, t)$ in model (3.2), we impose a set of conditions on $F(x, s, t)$ and the distribution of $X(s)$ in the following Proposition.

Proposition S.1. *Suppose that the following conditions are satisfied,*

(1). $X(s)$ is a Gaussian process. The covariance function $\Sigma_X(s, s')$ of $X(s)$ is continuous and all the eigenvalues of $\Sigma_X(s, s')$ are positive.

(2). The true function $F(x, s, t)$ satisfies the condition (3.3) in the manuscript, and the partial derivative $\partial_x F(x, s, t)$ is a continuous function on $(x, s, t) \in (-\infty, \infty) \times [0, 1] \times [a, b]$.

If there is another function $\widetilde{F}(x, s, t)$ satisfying the conditions in (2), and $\int_0^1 \widetilde{F}(X(s), s, t)ds = \int_0^1 F(X(s), s, t)ds$ for all $0 \le t \le 1$, then we have $\widetilde{F}(x, s, t) = F(x, s, t)$.

Remark 1. *Different conditions exist for the identifiability of $F(x, s, t)$. But we note that any set of identifiability conditions cannot be verified in practice. Indeed, a necessary condition for identifiability is that all the eigenvalues of the covariance function of $X(s)$ are positive. Indeed, if this condition is not satisfied, let $\beta(s)$ be a nonzero eigenfunction corresponding to the zero eigenvalue. Then we have $\int_0^1 X(s)\beta(s)ds = 0$ and hence the function $F^*(x, s, t) = F(x, s, t) + \{x - E[X(s)]\}\beta(s)$ leads to the same model as the true model. But $F^*(x, s, t)$ satisfies the condition (3.3) and is different from $F(x, s, t)$. Therefore, $F(x, s, t)$ is not identifiable. In practice, we only have a finite number of sample curves and there are* *infinitely many eigenvalues for the covariance function of $X(s)$, we cannot determine if all the eigenvalues are positive. Therefore, any identifiability conditions cannot be verified in practice.*

Proof of Proposition S.1. We prove this proposition by contradiction. Let (Ω, P) denote the probability space. Suppose that there exists another function $\widetilde{F}(x,s,t)$ which is different from $F(x,s,t)$, and satisfies the conditions in (2) in this proposition and $\int_0^1 \widetilde{F}(X(\omega,s),s,t)ds = \int_0^1 F(X(\omega,s),s,t)ds$ for all $\omega \in \Omega$ and $0 \leq t \leq 1$. Then there exists t_0 such that $\widetilde{F}(x,s,t_0) \neq F(x,s,t_0)$. Define $G(x,s) = \widetilde{F}(x,s,t_0) - F(x,s,t_0)$. Then $G(x,s) \neq 0$ and

$$\int_0^1 G(X(\omega,s),s)ds = \int_0^1 \widetilde{F}(X(\omega,s),s,t_0)ds - \int_0^1 F(X(\omega,s),s,t_0)ds = 0, \qquad \text{(S2.1)}$$

for all ω. Then for any two ω and ω', we have

$$0 = \int_0^1 G(X(\omega',s),s)ds - \int_0^1 G(X(\omega,s),s)ds = \int_0^1 \int_{X(\omega,s)}^{X(\omega',s)} \partial_x G(x,s)dxds. \qquad \text{(S2.2)}$$

We show that $\partial_x G(x,s)$ is a nonzero function. Otherwise, if $\partial_x G = 0$, then $G(x,s)$ only depends on s, that is, we have $\widetilde{F}(x,s,t_0) - F(x,s,t_0) = G(x,s) = h(s)$ for some function $h(s)$. By the condition (3.3), $E[\widetilde{F}(X(s),s,t_0)] = 0 = E[F(X(s),s,t_0)]$ for all $0 \leq s \leq 1$, so we have $0 = E[G(X(s),s)] = h(s)$ and hence $G(x,s) = 0$ which contradicts to the fact $G(x,s) \neq 0$. Therefore, $\partial_x G(x,s)$ is a nonzero function. Because $\partial_x G(x,s) = \partial_x \widetilde{F}(x,s,t_0) - \partial_x F(x,s,t_0)$ is nonzero and continuous due to the continuity of both $\partial_x \widetilde{F}(x,s,t_0)$ and $\partial_x F(x,s,t_0)$, we can find a rectangle region $\{(x,s) : x_1 \leq x \leq$

$x_2, s_1 \le s \le s_2\}$, such that $|\partial_x G(x,s)| > \delta$ for some positive constant δ. Without loss of generality, we assume that in this rectangle region, $\partial_x G(x,s) > \delta$, (if $\partial_x G(x,s)$ is negative, the proof is similar). We define two step functions:

$$f_1(s) = \begin{cases} \frac{x_1+x_2}{2} & \text{if } 0 < s < s_1 \text{ or } s_2 < s < 1 \\ x_1 & \text{if } s_1 \le s \le s_2 \end{cases},$$

$$f_2(s) = \begin{cases} \frac{x_1+x_2}{2} & \text{if } 0 < s < s_1 \text{ or } s_2 < s < 1 \\ x_2 & \text{if } s_1 \le s \le s_2 \end{cases}, \tag{S2.3}$$

and plot them in the following Figure S.11.

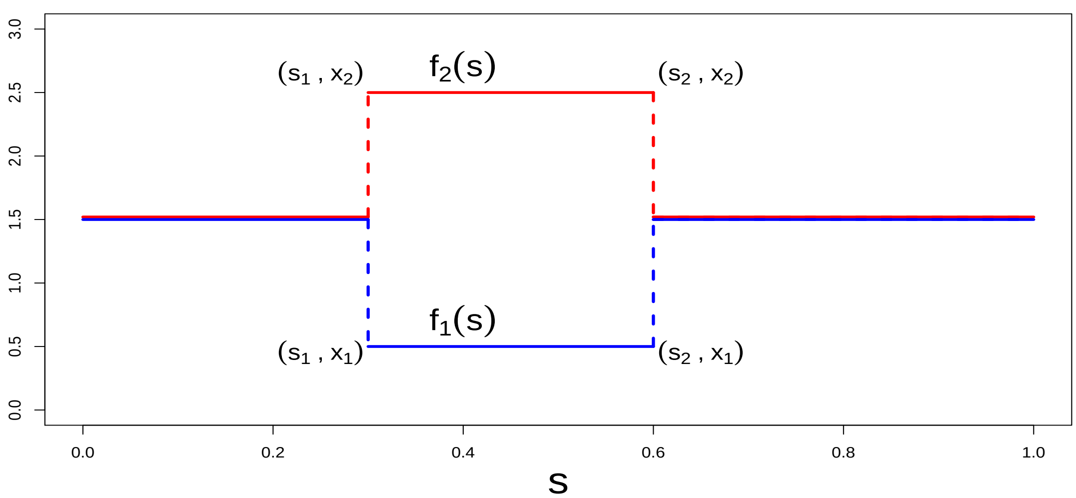

Figure S.11: The plots of $f_1(s)$ and $f_2(s)$ defined in (S2.3).

Because $\partial_x G(x,s) > \delta$ in the region $\{(x,s) : x_1 \le x \le x_2, s_1 \le s \le s_2\}$, we have

$$\int_0^1 G(f_2(s),s)ds - \int_0^1 G(f_1(s),s)ds = \int_0^1 \int_{f_1(s)}^{f_2(s)} \partial_x G(x,s)(x,s)dxds \tag{S2.4}$$
$$= \int\int_{\{x_1\le x\le x_2, s_1\le s\le s_2\}} \partial_x G(x,s)dxds \ge \int\int_{\{x_1\le x\le x_2, s_1\le s\le s_2\}} \delta dxds > \eta > 0,$$

where η is any positive constant satisfying $\eta < \delta(s_2 - s_1)(x_2 - x_1)$ and we will choose a specific η in the following.

Consider the Karhunen-Loève expansion:

$$X(s) = \sum_{k=1}^{\infty} \xi_k \phi_k(s), \tag{S2.5}$$

where $\{\phi_k(s), k \ge 1\}$ is the collection of all the eigenfunctions of $\Sigma_X(s,s')$ and forms an orthonormal basis of the $L^2[0,1]$ space. $\xi_k = \int_0^1 X(s)\phi_k(s)ds$, $k \ge 1$, are uncorrelated random variables and $Var(\xi_k) = \nu_k$, where $\nu_1 \ge \nu_2 \ge \cdots > 0$ are eigenvalues of $\Sigma_X(s,s')$ and they are all positive by the condition (1) in this proposition. Because we have assumed that $X(s)$ is a Gaussian process, $\{\xi_k, k \ge 1\}$ are independent normal variables. Now let

$$f_1(s) = \sum_{k=1}^{\infty} a_k^{(1)} \phi_k(s), \quad f_2(s) = \sum_{k=1}^{\infty} a_k^{(2)} \phi_k(s),$$

be the expansions of $f_1(s)$ and $f_2(s)$ using the orthonormal basis $\{\phi_k(s), k \ge 1\}$. Define an event

$$A_0 = \left\{\omega : \int_0^1 [X(s,\omega) - f_1(s)]^2 ds \le 1\right\} \bigcup \left\{\omega : \int_0^1 [X(s,\widetilde{\omega}) - f_2(s)]^2 ds \le 1\right\}. \tag{S2.6}$$

Let $N_1 = \sup\{|X(s,\omega)| : 0 \le s \le 1, \omega \in A_0\}$ and $N_2 = \sup_{0\le s\le 1} \max\{f_1(s), f_2(s)\}$

$$N = \max\{N_1, N_2\}, \quad C_1 = \sup_{|x|\le N, 0\le s\le 1} |\partial_x G(x,s)|. \tag{S2.7}$$

We choose an integer M large enough such that

$$\sum_{k=M+1}^{\infty} E[\xi_k^2] = \sum_{k=M+1}^{\infty} \nu_k \le \frac{\eta}{32C_1}, \qquad \sum_{k=M+1}^{\infty} \{a_k^{(1)}\}^2 < \frac{\eta}{16C_1}, \qquad \sum_{k=M+1}^{\infty} \{a_k^{(2)}\}^2 < \frac{\eta}{16C_1}. \tag{S2.8}$$

We define two events

$$A_1 = \{\omega : [\xi_k(\omega) - a_k^{(1)}]^2 < \eta/(8C_1M), \forall 1 \le k \le M, \sum_{k=M+1}^{\infty} \xi_k(\omega)^2 \le \eta/(16C_1)\}, \tag{S2.9}$$

$$A_2 = \{\omega : [\xi_k(\omega) - a_k^{(2)}]^2 < \eta/(8C_1M), \forall 1 \le k \le M, \sum_{k=M+1}^{\infty} \xi_k(\omega)^2 \le \eta/(16C_1)\}.$$

Because $\{\xi_k, k \ge 1\}$ are independent normal variables,

$$P(A_1) = \prod_{k=1}^{M} P\left([\xi_k(\omega) - a_k^{(1)}]^2 < \eta/(8C_1M)\right) P\left(\sum_{k=M+1}^{\infty} \xi_k(\omega)^2 \le \eta/(16C_1)\right),$$

where each $P\left([\xi_k(\omega) - a_k^{(1)}]^2 < \eta/(8C_1M)\right)$ is positive, $1 \le k \le M$, and

$$\begin{aligned} P\left(\sum_{k=M+1}^{\infty} \xi_k(\omega)^2 \le \eta/(16C_1)\right) &= 1 - P\left(\sum_{k=M+1}^{\infty} \xi_k^2 > \eta/(16C_1)\right) \\ \ge 1 - \frac{E\left(\sum_{k=M+1}^{\infty} \xi_k^2\right)}{\eta/(16C_1)} &= 1 - \frac{\sum_{k=M+1}^{\infty} E[\xi_k^2]}{\eta/(16C_1)} = 1 - \frac{\sum_{k=M+1}^{\infty} \nu_k}{\eta/(16C_1)} \\ \ge 1 - \frac{\eta/(32C_1)}{\eta/(16C_1)} &= \frac{1}{2}, \end{aligned}$$

where the first inequality in the second line follows from the Markov's inequality, and the first inequality in the last line follows from the first inequality in (S2.8). Therefore, $P(A_1)$ is positive and similarly, $P(A_2)$ is also positive. Pick up an $\omega \in A_1$ and an $\widetilde{\omega} \in A_2$. By

(S2.9) and (S2.8), and noting that $\{\phi_k(s), k \geq 1\}$ is an orthonormal basis, we have

$$\int_0^1 [X(s,\omega) - f_1(s)]^2 ds = \int_0^1 [\sum_{k=1}^{\infty} \xi_k(\omega)\phi_k(s) - \sum_{k=1}^{\infty} a_k^{(1)}\phi_k(s)]^2 ds$$
$$\leq \sum_{k=1}^{M} [\xi_k(\omega) - a_k^{(1)}]^2 + \int_0^1 \left[\sum_{k=M+1}^{\infty} \xi_k(\omega)\phi_k(s) - a_k^{(1)}\phi_k(s)\right]^2 ds$$
$$\leq \sum_{k=1}^{M} [\xi_k(\omega) - a_k^{(1)}]^2 + 2\int_0^1 \left[\sum_{k=M+1}^{\infty} \xi_k(\omega)\phi_k(s)\right]^2 ds + 2\int_0^1 \left[\sum_{k=M+1}^{\infty} a_k^{(1)}\phi_k(s)\right]^2 ds$$
$$\leq \sum_{k=1}^{M} \eta/(8C_1M) + 2\sum_{k=M+1}^{\infty} \xi_k(\omega)^2 + 2\sum_{k=M+1}^{\infty} [a_k^{(1)}]^2$$
$$\leq \frac{\eta}{8C_1} + 2\frac{\eta}{16C_1} + 2\frac{\eta}{16C_1} = \frac{3\eta}{8C_1}. \quad \text{(S2.10)}$$

Similarly, $$\int_0^1 [X(s,\widetilde{\omega}) - f_2(s)]^2 ds \leq \frac{3\eta}{8C_1}. \quad \text{(S2.11)}$$

Now we choose η small enough such that $3\eta/(8C_1) < 1$. Then by the definition of the event A_0 in (S2.6), we have $\omega, \widetilde{\omega} \in A_0$, which, together with (S2.7), implies that for any (x, s) with $0 \leq s \leq 1$ and x between $X(s,\omega)$ and $f_1(s)$, or x between $X(s,\widetilde{\omega})$ and $f_2(s)$, we have

$$|\partial_x G(x,s)| \leq C_1. \quad \text{(S2.12)}$$

Now by (S2.2),

$$0 = \left|\int_0^1 G(X(s,\widetilde{\omega}),s)ds - \int_0^1 G(X(s,\omega),s)ds\right| = \left|\int_0^1 \int_{X(s,\omega)}^{X(s,\widetilde{\omega})} \partial_x G(x,s)dxds\right|$$
$$= \left|\int_0^1 \int_{X(s,\omega)}^{f_1(s)} \partial_x G(x,s)dxds + \int_0^1 \int_{f_1(s)}^{f_2(s)} \partial_x G(x,s)dxds + \int_0^1 \int_{f_2(s)}^{X(s,\widetilde{\omega})} \partial_x G(x,s)dxds\right|$$
$$\geq \left|\int_0^1 \int_{f_1(s)}^{f_2(s)} \partial_x G(x,s)dxds\right| - \left|\int_0^1 \int_{f_1(s)}^{X(s,\omega)} \partial_x G(x,s)dxds\right| - \left|\int_0^1 \int_{f_2(s)}^{X(s,\widetilde{\omega})} \partial_x G(x,s)dxds\right|$$

$$\geq \eta - \left| \int_0^1 \int_{f_1(s)}^{X(s,\omega)} \partial_x G(x,s) dx ds \right| - \left| \int_0^1 \int_{f_2(s)}^{X(s,\widetilde{\omega})} \partial_x G(x,s) dx ds \right| \quad \text{(by (S2.4))}$$

$$\geq \eta - \int_0^1 |X(s,\widetilde{\omega}) - f_1(s)| C_1 ds - \int_0^1 |X(s,\widetilde{\omega}) - f_2(s)| C_1 ds \quad \text{(by (S2.12))}$$

$$\geq \eta - C_1 \int_0^1 |X(s,\widetilde{\omega}) - f_1(s)|^2 ds - C_1 \int_0^1 |X(s,\widetilde{\omega}) - f_2(s)|^2 ds \quad \text{(by Cauchy-Schwarz)}$$

$$\geq \eta - \frac{3C_1\eta}{8C_1} - \frac{3C_1\eta}{8C_1} = \frac{\eta}{4}, \tag{S2.13}$$

where the last inequality follows from (S2.10) and (S2.11). Now (S2.13) implies that $0 \geq \eta/4$, but η is a positive number. So we got a contradiction. Hence, we must have $\widetilde{F}(x,s,t) = F(x,s,t)$. The proposition is proved. □

S3 Computational issue

S.3.1 Solving the optimization problem (3.18)

We first calculate the first term, the integrated sum of squared residuals, in the objective function of (3.18).

$$\frac{1}{n}\int_0^1 \left\| \mathbf{Y}(t) - v_0(t)\mathbf{1}_n - \sum_{k=1}^K \widehat{\mathbf{R}}_k v_k(t) \right\|_2^2 dt = \frac{1}{n}\int_0^1 \|\mathbf{Y}(t)\|_2^2 dt + \int_0^1 v_0(t)^2 dt$$

$$- 2\int_0^1 \frac{1}{n}\{\mathbf{1}_n^\top \mathbf{Y}(t)\} v_0(t) dt - 2\sum_{k=1}^K \int_0^1 \frac{1}{n}\{\widehat{\mathbf{R}}_k^\top \mathbf{Y}(t)\} v_k(t) dt + 2\sum_{k=1}^K \int_0^1 \frac{1}{n}\{\mathbf{1}_n^\top \widehat{\mathbf{R}}_k\} v_k(t) v_0(t) dt$$

$$+ 2\sum_{l=1}^K \sum_{k=1}^K \int_0^1 \frac{1}{n}\{\widehat{\mathbf{R}}_l^\top \widehat{\mathbf{R}}_k\} v_k(t) v_l(t) dt. \tag{S3.14}$$

From the definition of $\widehat{\mathbf{\Sigma}}$ in (3.14), if $j \neq k$,

$$\frac{1}{n}\widehat{\mathbf{R}}_j^\top \widehat{\mathbf{R}}_k = \frac{1}{n}\sum_{l=1}^{n}\left[r_l(\widehat{G}_j) - \overline{r}(\widehat{G}_j)\right]\left[r_l(\widehat{G}_k) - \overline{r}(\widehat{G}_k)\right] = \widehat{\mathbf{\Sigma}}(\widehat{G}_j, \widehat{G}_k) = -\lambda\langle \widehat{G}_j, \widehat{G}_k\rangle_{H^2},$$

where the last equality follows from the constraints in the optimization problem (3.15). Similarly, if $j = k$,

$$\frac{1}{n}\widehat{\mathbf{R}}_k^\top \widehat{\mathbf{R}}_k = \frac{1}{n}\sum_{l=1}^{n}\left[r_l(\widehat{G}_k) - \overline{r}(\widehat{G}_k)\right]\left[r_l(\widehat{G}_k) - \overline{r}(\widehat{G}_k)\right] = \widehat{\mathbf{\Sigma}}(\widehat{G}_k, \widehat{G}_k) = 1 - \lambda\langle \widehat{G}_k, \widehat{G}_k\rangle_{H^2},$$

Because in our asymptotic theory, the tuning parameter λ goes to zero as $n \to \infty$, the terms $-\lambda\langle \widehat{G}_j, \widehat{G}_k\rangle_{H^2}$ are typically small. Hence, we have the approximation $\frac{1}{n}\widehat{\mathbf{R}}_j^\top \widehat{\mathbf{R}}_k \approx 0$, for all $k \neq j$, and $\frac{1}{n}\widehat{\mathbf{R}}_k^\top \widehat{\mathbf{R}}_k \approx 1$. Moreover, for any $1 \leq k \leq K$, $\widehat{\mathbf{R}}_k^\top \mathbf{1}_n = \sum_{l=1}^{n}\left[r_l(\widehat{G}_j) - \overline{r}(\widehat{G}_j)\right] = 0$. Then by (S3.14), we have the following approximation,

$$\begin{aligned}
&\frac{1}{n}\int_0^1 \left\|\mathbf{Y}(t) - v_0(t)\mathbf{1}_n - \sum_{k=1}^{K}\widehat{\mathbf{R}}_k v_k(t)\right\|_2^2 dt \approx \frac{1}{n}\int_0^1 \|\mathbf{Y}(t)\|_2^2 dt + \int_0^1 v_0(t)^2 dt \qquad \text{(S3.15)}\\
&- 2\int_0^1 \bar{y}(t)v_0(t)dt - 2\sum_{k=1}^{K}\int_0^1 \frac{1}{n}\{\widehat{\mathbf{R}}_k^\top \mathbf{Y}(t)\}v_k(t)dt + 2\sum_{k=1}^{K}\int_0^1 v_k(t)^2 dt = \frac{1}{n}\int_0^1 \|\mathbf{Y}(t)\|_2^2 dt\\
&+ \int_0^1 \{v_0(t) - \bar{y}(t)\}^2 dt - \int_0^1 \bar{y}(t)^2 dt + 2\sum_{k=1}^{K}\int_0^1 \{v_k(t) - \widehat{w}_k^{(0)}(t)\}^2 dt - 2\sum_{k=1}^{K}\int_0^1 \widehat{w}_k^{(0)}(t)^2 dt,
\end{aligned}$$

where $\widehat{w}_k^{(0)}(t) = \widehat{\mathbf{R}}_k^\top \mathbf{Y}(t)/n$. Thus, the estimates $\widehat{\mu}(t)$ and $\widehat{\phi}_k(t)$ of $\mu(t), \phi_k(t)$, $1 \leq k \leq K$, can be obtained by solving the following problems separately:

$$\min_{v_0(t)} \left[\int_a^b |v_0(t) - \overline{Y}(t)|^2 dt + \kappa\int_a^b |v_0''(t)|^2 dt\right], \qquad \text{(S3.16)}$$

$$\min_{v_k(t)} \left[\int_a^b |v_k(t) - \frac{1}{n}\widehat{\mathbf{R}}_k^\top \mathbf{Y}(t)|^2 dt + \kappa\int_a^b |v_k''(t)|^2 dt\right], \quad 1 \leq k \leq K. \qquad \text{(S3.17)}$$

We solve (S3.16) and (S3.17) using B-spline basis expansions as in Section 3.2 of Luo and Qi (2017).

S.3.2 Choice of the number of components and tuning parameters

We choose the number of components and the two tuning parameters λ and κ simultaneously based on the following cross-validation procedure. Both λ and κ are chosen from the set $\{10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}, 1\}$. For the ℓ-th value of λ, $1 \le \ell \le 6$, we first determine a maximum number of components, $\widehat{K}_{max,\ell}$. The optimal number of components will be chosen from the integers between 1 and $\widehat{K}_{max,\ell}$. In Theorem 3, we choose the first few components with relatively large σ_k^2. As σ_k^2 can be estimated by $(\widehat{\sigma}_k^{(\ell)})^2$ which is the maximum value of the optimization problem (3.15), we set

$$\widehat{K}_{max,\ell} = \min\left\{ k > 1 : \frac{(\widehat{\sigma}_k^{(\ell)})^2}{(\widehat{\sigma}_1^{(\ell)})^2 + \cdots + (\widehat{\sigma}_k^{(\ell)})^2} \le 0.01 \right\}. \tag{S3.18}$$

Once we have determined all the $\widehat{K}_{max,\ell}$, we use the cross-validation method to determine the tuning parameters (λ, κ) and the optimal number of components, $\widehat{K}_{opt}$, simultaneously. We summarize the details of the procedure in the following algorithm.

Algorithm 1. 1. For the ℓ-th value of the tuning parameter λ, $1 \le \ell \le 6$, we determine $\widehat{K}_{max,\ell}$ using the whole data set and (S3.18).

2. We randomly split the whole data set into five subsets. For each $1 \le v \le 5$, we use the v-th subset as the v-th validation set and all other observations as the v-th training set. Then for the ℓ-th value for λ and the v-th training set,

(a) we estimate $\widehat{G}_k^{(v,\ell)}(x,s)$ for all $1 \le k \le \widehat{K}_{max,\ell}$. Then for the i-th value of κ, $1 \le i \le 6$, we estimate $\widehat{\mu}_0^{(v,\ell,i)}(t)$ and $\widehat{\phi}_k^{(v,\ell,i)}(t)$ for all $1 \le k \le \widehat{K}_{max,\ell}$.

(b) For each $K = 1, \cdots, \widehat{K}_{max,\ell}$, we use $\widehat{\mu}_0^{(v,\ell,i)}(t)$, $\widehat{G}_k^{(v,\ell)}(x,s)$ and $\widehat{\phi}_k^{(v,\ell,i)}(t)$ for $1 \le k \le K$, to obtain the predicted response curves $\{\widehat{y}_j^{(v,\ell,i,K)}(t), 1 \le j \le N_v\}$ for the v-th validation set using (3.19) and then calculate the corresponding validation error $e_{v,\ell,i,K} = \sum_{j=1}^{N_v} \sum_{m=1}^{N_y} (\widehat{y}_j^{(v,\ell,i,K)}(t_m) - y_j^{(v)}(t_m))^2 \delta_m / N_v$, where $\{y_j^{(v)}(t), 1 \le j \le N_v\}$ are the collection of all the response curves in the v-th validation set.

After we repeat (a)-(b) for all $1 \le v \le 5$ and $1 \le \ell \le 6$, we calculate the average validation error, $\bar{e}_{\ell,i,K} = \sum_{v=1}^{5} e_{v,\ell,i,K}/5$.

3. Let $\bar{e}_{\ell_0,i_0,K_0} = \min_{1 \le \ell \le 6, 1 \le i \le 6, 1 \le K \le \widehat{K}_{max,\ell}} \bar{e}_{\ell,i,K}$. Then we choose the ℓ_0-th value for λ, the i_0-th value for κ, and the optimal number of components is $\widehat{K}_{opt} = K_0$.

S4 Details of calculating the estimation error for $F(x,s,t)$ in Simulation 1

In Simulation 1, the functions $F(x,s,t)$ do not satisfy the condition (3.3) and may be unidentifiable. So we consider the following centered function

$$F_0(x,s,t) = F(x,s,t) - E[F(X(s),s,t],$$

and the models in Simulation 1 can be rewritten as

$$Y(t) = \mu_0(t) + \int_0^1 F_0(X(s), s, t)ds + \varepsilon(t),$$

where $\mu_0(t) = \mu(t) + \int_0^1 E[F(X(s), s, t]ds$. To calculate the expectation $E[F(X(s), s, t]$, we note that $X(s)$ is a Gaussian process, and for any s, the marginal distribution of $X(s)$ is the standard normal distribution. So $E[F(X(s), s, t] = \int_{-\infty}^{\infty} e^{-x^2/2} F(x, s, t)dx/\sqrt{2\pi}$ which is calculated using a Riemann sum in practice. Based on Proposition S.1 in Section S2 of this supplementary material, we can show the centered functions $F_0(x, s, t)$ is identifiable. Actually, $F_0(x, s, t)$ satisfies the condition (3.3), and hence satisfies the condition (2) in Proposition S.1. It follows from the results in Section 4.3.1 of the book Rasmussen and Williams (2005) that all the eigenvalues of the covariance function $\mathbf{\Sigma}_X(s, s') = e^{-\{10|s-s'|\}^2}$ are positive. Hence, the condition (1) in Proposition S.1 is also satisfied. It follows from Proposition S.1 that $F_0(x, s, t)$ is identifiable. So we will calculate the estimation error for $F_0(x, s, t)$.

For any method considered in Simulation 1, let $\widehat{F}(x, s, t)$ denote the estimate of $F(x, s, t)$. Then we use $\widehat{F}_0(x, s, t) = \widehat{F}(x, s, t) - \int_{-\infty}^{\infty} e^{-x^2/2} \widehat{F}(x, s, t)dx/\sqrt{2\pi}$ as the estimate of $F_0(x, s, t)$. In practice, we have only a finite number of sample predictor curves. The range only covers a finite region of x. So we cannot obtain any information about $F_0(x, s, t)$ when x is outside of this finite region, although the function $F_0(x, s, t)$ is defined in the unbounded region $(-\infty, \infty) \times [0, 1] \times [0, 1]$. In the following Figure S.12, we plot 100 sample curves of $X(s)$ in the training set of one repeat in Simulation 1. The plot shows that most values of these sample curves fall between -2.5 and 2.5.

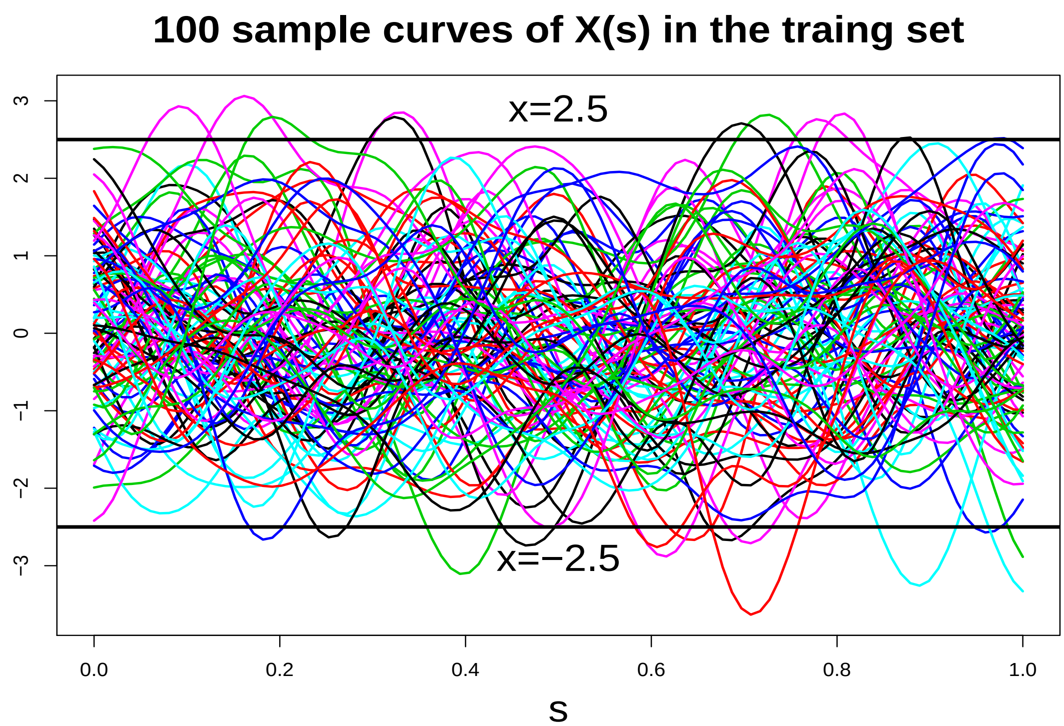

Figure S.12: 100 sample curves of $X(s)$ in the training set of one repeat in Simulation 1.

Therefore, one can anticipate that the estimate of $F_0(x, s, t)$ with x outside the interval $[-2.5, 2.5]$ may be less accurate for any method considered in Simulation 1. Therefore, we only consider the estimation error of $F_0(x, s, t)$ in the finite region $[-2.5, 2.5] \times [0, 1] \times [0, 1]$. Specifically, we calculate the following relative mean squared error (RelMSE) for $F_0(x, s, t)$:

$$RelMSE = \frac{\int_{-2.5}^{2.5} \int_0^1 \int_0^1 \left[\widehat{F}_0(x, s, t) - F_0(x, s, t)\right]^2 dxdsdt}{\int_{-2.5}^{2.5} \int_0^1 \int_0^1 F_0(x, s, t)^2 dxdsdt}$$

where the integrals are calculated using Riemann sum over a dense grid in the region $[-2.5, 2.5] \times [0, 1] \times [0, 1]$. We report the averages and standard deviations of the RelMSEs over 100 repeats in Table S.3 for all 16 settings in Simulation 1. Since the output of *pffr.pc* does not provide estimate of $F(x, s, t)$, we do not include this method in Table S.3.

Table S.3: The averages (standard deviations) of RelMSEs of 100 replicates for Simulation 1.

Model	σ^2	ρ	SigComp.nonlin	pffr.nonlin	SigComp.lin	pffr
1	0.1	0	3e-04 (0)	0.0206 (1e-04)	2e-04 (2e-04)	0.0011 (2e-04)
		0.7	0.0011 (2e-04)	0.0213 (4e-04)	6e-04 (2e-04)	0.0778 (0.0462)
	10	0	0.0047 (0.0011)	0.0226 (9e-04)	0.0046 (0.0018)	0.0051 (9e-04)
		0.7	0.0208 (0.0083)	0.1445 (0.0405)	0.0219 (0.008)	2.384 (1.8198)
2	0.1	0	6e-04 (2e-04)	0.002 (7e-04)	0.8878 (0.2913)	55640.0044 (76699.4426)
		0.7	0.0026 (9e-04)	0.0035 (0.0012)	0.9418 (0.7316)	52977.5195 (65548.8397)
	10	0	0.0077 (0.0021)	0.0059 (0.0017)	0.8904 (0.3283)	12748.2956 (60587.9511)
		0.7	0.0284 (0.0105)	0.1024 (0.0364)	0.9038 (0.4537)	210.2648 (1094.5857)
3	0.1	0	9e-04 (1e-04)	0.03 (0.0119)	0.3607 (0.0259)	29182.773 (51436.652)
		0.7	0.0026 (0.0014)	0.0327 (0.0092)	0.3944 (0.2519)	36284.5042 (62447.6018)
	10	0	0.015 (0.0038)	0.0282 (0.0091)	0.3865 (0.1979)	5.131 (25.3823)
		0.7	0.0696 (0.0271)	0.2071 (0.0729)	0.3736 (0.0524)	32.7072 (192.5884)
4	0.1	0	0.0015 (3e-04)	0.0374 (0.0143)	1.0404 (0.5115)	127973.2515 (152373.5881))
		0.7	0.0041 (8e-04)	0.0403 (0.0193)	0.9444 (0.037)	173697.5153 (192651.5963)
	10	0	0.0177 (0.005)	0.0383 (0.014)	1.3213 (2.6472)	39966.9049 (95116.1123)
		0.7	0.0556 (0.0132)	0.1617 (0.054)	0.9547 (0.0842)	41087.235 (122414.8317)

S5 Proofs of theorems

S.5.1 Proof of Theorem 1

For $1 \le k \le K$, let $H_k(x,s)$ be any function satisfying that $\int_0^1 H_k(X(s),s)ds$ has a finite second moment and $\varphi_k(t)$ be any function in $L^2[a,b]$. Let

$$Y_{\text{new}}(t) = \mu(t) + \int_0^1 F(X_{\text{new}}(s), s, t)ds + \varepsilon_{\text{new}}$$
$$Y_1(t) = \mu(t) + \int_0^1 \sum_{k=1}^K G_k(X_{\text{new}}(s), s)\phi_k(t)ds$$
$$Y_2(t) = \mu(t) + \int_0^1 \sum_{k=1}^K H_k(X_{\text{new}}(s), s)\varphi_k(t)ds$$

denote the response function for $X_{\text{new}}(s)$, the predicted response function based on the partial sum $\sum_{k=1}^K G_k(x,s)\phi_k(t)$, and the predicted response function based on $\sum_{k=1}^K H_k(x,s)\varphi_k(t)$, respectively. The mean squared prediction error for $\sum_{k=1}^K G_k(x,s)\phi_k(t)$ is

$$E\left[\|Y_{\text{new}} - Y_1\|_{L^2}^2\right] \tag{S5.19}$$
$$= E\left[\left\|\mu + \int_0^1 F(X_{\text{new}}(s), s, \cdot)ds + \varepsilon_{\text{new}} - \mu - \int_0^1 \sum_{k=1}^K G_k(X_{\text{new}}(s), s)\phi_k ds\right\|_{L^2}^2\right]$$
$$= E\left[\left\|\int_0^1 F(X_{\text{new}}(s), s, \cdot)ds - \int_0^1 \sum_{k=1}^K G_k(X_{\text{new}}(s), s)\phi_k ds\right\|_{L^2}^2\right] + E\left[\|\varepsilon_{\text{new}}\|_{L^2}^2\right]$$

(because $X_{\text{new}}(s)$ and $\varepsilon_{\text{new}}(t)$ are independent) (S5.20)

$$= E\left[\left\|\int_0^1 F(X(s), s, \cdot)ds - \int_0^1 \sum_{k=1}^K G_k(X(s), s)\phi_k ds\right\|_{L^2}^2\right] + E\left[\|\varepsilon\|_{L^2}^2\right]$$

(because $X_{\text{new}}(s)$ and $\varepsilon_{\text{new}}(t)$ have the same distributions as $X(t)$ and $\varepsilon(t)$)

$$=E\left[\left\|S-\sum_{k=1}^{K}\mathbf{r}_k\phi_k\right\|_{L^2}^2\right]+E\left[\|\boldsymbol{\varepsilon}\|_{L^2}^2\right] \qquad \text{(by the definition of } S(t) \text{ and (3.6)).}$$

Similarly, the mean squared prediction error for the $\mu(t)+\sum_{k=1}^{K}H_k(x,s)\varphi_k(t)$ is

$$E\left[\|Y_{\text{new}}-Y_2\|_{L^2}^2\right]=E\left[\left\|S-\sum_{k=1}^{K}\mathbf{q}_k\varphi_k\right\|_{L^2}^2\right]+E\left[\|\boldsymbol{\varepsilon}\|_{L^2}^2\right]$$

where $\mathbf{q}_k=\int_0^1 H_k(X(s),s)ds$. A nice property of the KL expansion is that $\sum_{k=1}^{K}\mathbf{r}_k\phi_k(t)$ is the best K-dimensional random approximation of $S(t)$ and has the smallest approximation error among all K-dimensional random approximation of $S(t)$. So $E\left[\left\|S-\sum_{k=1}^{K}\mathbf{r}_k\phi_k\right\|_{L^2}^2\right]\leq E\left[\left\|S-\sum_{k=1}^{K}\mathbf{q}_k\varphi_k\right\|_{L^2}^2\right]$ and hence $E\left[\|Y_{\text{new}}-Y_1\|_{L^2}^2\right]\leq E\left[\|Y_{\text{new}}-Y_2\|_{L^2}^2\right]$. Because $H_k(x,s)$ and $\varphi_k(t)$ are arbitrary, the partial sum $\sum_{k=1}^{K}G_k(x,s)\phi_k(t)$ has the smallest prediction error.

Moreover, because $\{\mathbf{r}_1,\mathbf{r}_2,\cdots\}$ are uncorrelated random variables with mean zero and variance one and $\{\phi_1,\phi_2,\cdots\}$ are orthogonal functions with $\|\phi_k\|_{L^2}=\sigma_k$, we have

$$E\left[\|S-\sum_{k=1}^{K}\mathbf{r}_k\phi_k\|_{L^2}^2\right]=E\left[\|\sum_{k=K+1}^{\infty}\mathbf{r}_k\phi_k\|_{L^2}^2\right]=\sum_{k=K+1}^{\infty}E\left[\|\mathbf{r}_k\phi_k\|_{L^2}^2\right]=\sum_{k=K+1}^{\infty}\sigma_k^2,$$

which, together with (S5.19), leads to

$$E\left[\|Y_{\text{new}}-Y_1\|_{L^2}^2\right]=\sum_{k=K+1}^{\infty}\sigma_k^2+E[\|\varepsilon\|_{L^2}^2].$$

The inequalities in (3.8) in this theorem have been proved. Under Condition 1 in Section

3.4 of asymptotic theory, a straightforward calculation leads to the inequalities in (3.9).

S.5.2 Proof of Theorem 2

We only prove the theorem for $k = 1$. The proof of the theorem for $k > 1$ is very similar to that for $k = 1$ and hence is omitted. Here is the outline of the proof for $k = 1$. We first show that for any $G(x, s)$ satisfying the constraint $\mathbf{\Sigma}(G, G) = 1$, we have $\mathbf{\Lambda}(G, G) \leq \sigma_1^2$. Then we prove that $G_1(x, s)$ satisfies $\mathbf{\Sigma}(G_1, G_1) = 1$ and $\mathbf{\Lambda}(G_1, G_1) = \sigma_1^2$, which implies that $G_1(x, s)$ is the solution to (3.11) and the maximum value of (3.11) is σ_1^2.

Before providing the details of the proof, we recall some definitions and notations. Recall that for any $K \geq 1$, $\sum_{k=1}^K \mathbf{r}(G_k)\phi_k(t) = \sum_{k=1}^K \mathbf{r}_k\phi_k(t)$ is the best K-dimensional approximation to $S(t)$, and

$$S(t) = \sum_{k=1}^{\infty} \mathbf{r}(G_k)\phi_k(t), \text{where} \tag{S5.21}$$

$$\mathbf{r}(G_1), \mathbf{r}(G_2), \cdots, \text{ are uncorrelated random variables with mean 0 and variance 1,} \tag{S5.22}$$

$$\text{and } \phi_k(t) = \sigma_k\widetilde{\phi}_k, \text{ where } \widetilde{\phi}_1, \widetilde{\phi}_2, \cdots, \text{ are orthonormal eigenfunctions of } S(t). \tag{S5.23}$$

Recall the definitions of $\mathbf{\Lambda}(G, G)$ and $\mathbf{\Sigma}(G, G)$:

$$\mathbf{\Lambda}(G, G) = \int_a^b E\left[S(t)\mathbf{r}(G)\right]^2 dt, \quad \mathbf{\Sigma}(G, G) = E\left[(\mathbf{r}(G))^2\right]. \tag{S5.24}$$

Now let $G(x, s)$ be any function satisfying the following constraint in the optimization problem (3.11):

$$1 = \mathbf{\Sigma}(G, G) = E\left[(\mathbf{r}(G))^2\right]. \tag{S5.25}$$

We have

$$\mathbf{\Lambda}(G, G) = \int_a^b \{E[S(t)\mathbf{r}(G)]\}^2\, dt = \int_a^b \left\{\sum_{k=1}^{\infty} E[\mathbf{r}(G_k)\mathbf{r}(G)]\, \phi_k(t)\right\}^2 dt \quad \text{(by (S5.21))}$$

$$= \sum_{k=1}^{\infty} \{E[\mathbf{r}(G_k)\mathbf{r}(G)]\}^2 \|\phi_k\|_{L^2}^2 \qquad \text{(because } \phi_1(t), \cdots, \phi_K(t) \text{ are orthogonal)}$$

$$= \sum_{k=1}^{\infty} \{E[\mathbf{r}(G_k)\mathbf{r}(G)]\}^2 \sigma_k^2 \qquad \text{(because } \|\phi_k\|_{L^2} = \sigma_k \text{ by (S5.23))}$$

$$\leq \sum_{k=1}^{\infty} \{E[\mathbf{r}(G_k)\mathbf{r}(G)]\}^2 \sigma_1^2 \leq \sigma_1^2 E\left[\mathbf{r}(G)^2\right] \qquad \text{(by (S5.22))}$$

$$= \sigma_1^2, \qquad \text{(by (S5.25))}.$$

Therefore, the maximum value of the optimization problem (3.11) is less than or equal to σ_1^2. On the other hand, by (S5.22),

$$\mathbf{\Sigma}(G_1, G_1) = E\left[(\mathbf{r}(G_1))^2\right] = 1.$$

Therefore, $G_1(x, s)$ satisfies the constraint in (3.11) for $k = 1$. Moreover, by (S5.24) and (S5.21),

$$\mathbf{\Lambda}(G_1, G_1) = \int_a^b \{E[S(t)\mathbf{r}(G_1)]\}^2\, dt = \int_a^b \left\{\sum_{k=1}^{\infty} E[\mathbf{r}(G_k)\mathbf{r}(G_1)]\, \phi_k(t)\right\}^2 dt$$

$$= \sum_{k=1}^{\infty} \{E[\mathbf{r}(G_k)\mathbf{r}(G_1)]\}^2 \|\phi_k\|_{L^2}^2 \qquad \text{(because } \phi_1(t), \cdots, \phi_K(t) \text{, are orthogonal to each other)}$$

$=\sigma_1^2 E\left[\mathbf{r}(G_1)^2\right] \qquad$ (because (S5.22) and $\|\phi_k\|_{L^2} = \sigma_k$)

$=\sigma_1^2, \qquad$ (by (S5.22)).

Therefore, $G_1(x,s)$ is the solution to (3.11) and the maximum value is σ_1^2.

S.5.3 Proof of Theorem 3

For convenience, we first recall some notations in the main manuscript. We use L^2 and H^2 to denote L^2 and Sobolev space in $[0,1]\times[0,1]$, respectively, and both of them are Hilbert spaces. Given a function $G(x,s)$ defined in $[0,1]\times[0,1]$, let

$$\|G\|_{L^2} = \sqrt{\int_0^1\int_0^1 G(x,s)^2 dxds}, \quad \|G\|_{H^2} = \sqrt{\|G\|_{L^2}^2 + \|\partial_{xx}G\|_{L^2}^2 + \|\partial_{xs}G\|_{L^2}^2 + \|\partial_{ss}G\|_{L^2}^2},$$

be the L^2 norm and the Sobolev norm, respectively. For two functions $G(x,s)$ and $\widetilde{G}(x,s)$ defined in $[0,1]\times[0,1]$, the inner products in the two Hilbert spaces are denoted by

$$\langle G,\widetilde{G}\rangle_{L^2} = \int_0^1\int_0^1 G(x,s)\widetilde{G}(x,s)dxds, \quad \langle G,\widetilde{G}\rangle_{H^2} = \int_0^1\int_0^1 \Big\{G(x,s)\widetilde{G}(x,s)$$
$$+\partial_{xx}G(x,s)\partial_{xx}\widetilde{G}(x,s) + \partial_{xs}G(x,s)\partial_{xs}\widetilde{G}(x,s) + \partial_{ss}G(x,s)\partial_{ss}\widetilde{G}(x,s)\Big\}\,dxds,$$

respectively.

- **Step 1: Show that $\boldsymbol{\Sigma}(\cdot,\cdot)$, $\boldsymbol{\Lambda}(\cdot,\cdot)$, $\widehat{\boldsymbol{\Sigma}}(\cdot,\cdot)$, and $\widehat{\boldsymbol{\Lambda}}(\cdot,\cdot)$ are all bounded bilinear functions in H^2**

For any $G(x,s) \in H^2$, by the definition of $\mathbf{r}(G)$ in (3.5), we have

$$
\begin{aligned}
|\mathbf{r}(G)| &= \left| \int_0^1 G(X(s),s)ds - E\left[\int_0^1 G(X(s),s)ds\right]\right| \\
&= \left| \int_0^1 \int_0^{X(s)} \partial_x G(x,s)dxds + \int_0^1 G(0,s)ds - E\left[\int_0^1 \int_0^{X(s)} \partial_x G(x,s)dxds + \int_0^1 G(0,s)ds\right]\right| \\
&= \left| \int_0^1 \int_0^{X(s)} \partial_x G(x,s)dxds - E\left[\int_0^1 \int_0^{X(s)} \partial_x G(x,s)dxds\right]\right| \\
&\leq \int_0^1 \int_0^{X(s)} |\partial_x G(x,s)|dxds + E\left[\int_0^1 \int_0^{X(s)} |\partial_x G(x,s)|dxds\right] \leq 2\int_0^1 \int_0^1 |\partial_x G(x,s)|dxds \\
&= 2\int_0^1 \int_0^1 \left| \int_0^x \partial_{xx} G(y,s)dy + \partial_x G(0,s)\right| dxds \\
&\leq 2\int_0^1 \int_0^1 \int_0^x |\partial_{xx} G(y,s)|\, dydxds + 2\int_0^1 \int_0^1 \int_0^x |\partial_x G(0,s)|\, dydxds \\
&\leq 2\int_0^1 \int_0^1 \int_0^1 |\partial_{xx} G(y,s)|\, dydxds + 2\int_0^1 \int_0^1 \int_0^1 |\partial_x G(0,s)|\, dydxds \\
&\leq 2\int_0^1 \int_0^1 |\partial_{xx} G(y,s)|\, dyds + 2\int_0^1 |\partial_x G(0,s)|\, ds \\
&\leq 2\int_0^1 \int_0^1 |\partial_{xx} G(y,s)|\, dyds + 2K\int_0^1 \left[\int_0^1 \{|G(y,s)| + |\partial_{xx} G(y,s)|\}dy\right] ds \qquad \text{(S5.26)} \\
&= (2K+2)\int_0^1 \left[\int_0^1 \{|G(y,s)| + |\partial_{xx} G(y,s)|\}dy\right] ds \\
&\leq (2K+2)\sqrt{\int_0^1 \int_0^1 \{G(y,s)^2 + \partial_{xx} G(y,s)^2\}dyds}\sqrt{\int_0^1 \int_0^1 2dyds} \qquad \text{(S5.27)} \\
&\leq (2K+2)\sqrt{2}\|G\|_{H^2}, \qquad \text{(S5.28)}
\end{aligned}
$$

where the inequality in (S5.26) follows from the interpolation inequality (Lemma 5.4 in Adams and Fournier (2003)) and the inequality in (S5.27) follows from the Cauchy-

Schwarz inequality. By (S5.28) and the definition of $\mathbf{\Sigma}(\cdot,\cdot)$ in (3.10), for any $G(x,s), \widetilde{G}(x,s) \in H^2$,

$$|\mathbf{\Sigma}(G,\widetilde{G})| = \left|E[\mathbf{r}(G)\mathbf{r}(\widetilde{G})]\right| \leq 2(2K+2)^2 \|G\|_{H^2} \|\widetilde{G}\|_{H^2}. \tag{S5.29}$$

(S5.29) implies that $\mathbf{\Sigma}(\cdot,\cdot)$ is a bounded bilinear function in H^2. By Theorem 12.8 in Rudin (1991), $\mathbf{\Sigma}(\cdot,\cdot)$ defines a bounded linear operator which, for simplicity, is still denoted by $\mathbf{\Sigma}$: for any $G(x,s), \widetilde{G}(x,s) \in H^2$, $\mathbf{\Sigma}(G,\widetilde{G}) = \langle G, \mathbf{\Sigma}\widetilde{G}\rangle_{H^2}$. Similarly, we can show that $\mathbf{\Lambda}(\cdot,\cdot)$, $\widehat{\mathbf{\Sigma}}(\cdot,\cdot)$, and $\widehat{\mathbf{\Lambda}}(\cdot,\cdot)$ are all bounded bilinear functions in H^2. They all define bounded operators in H^2 (still denoted by $\mathbf{\Lambda}$, $\widehat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$):

$$\widehat{\mathbf{\Sigma}}(G,\widetilde{G}) = \langle G, \widehat{\mathbf{\Sigma}}\widetilde{G}\rangle_{H^2}, \quad \mathbf{\Lambda}(G,\widetilde{G}) = \langle G, \mathbf{\Lambda}\widetilde{G}\rangle_{H^2}, \quad \widehat{\mathbf{\Lambda}}(G,\widetilde{G}) = \langle G, \widehat{\mathbf{\Lambda}}\widetilde{G}\rangle_{H^2}.$$

Then the optimization problem (3.11) can be expressed as

$$\max_{G\in H^2} \quad \langle G, \mathbf{\Lambda} G\rangle_{H^2}, \tag{S5.30}$$
$$\text{subject to} \quad \langle G, \mathbf{\Sigma} G\rangle_{H^2} = 1 \quad \text{and} \quad \langle G_{k'}, \mathbf{\Sigma} G\rangle_{H^2} = 0, \quad 1 \leq k' \leq k-1.$$

The optimization problem (3.15) can be expressed as

$$\max_{G\in H^2} \quad \langle G, \widehat{\mathbf{\Lambda}} G\rangle_{H^2},$$
$$\text{subject to} \quad \langle G, \widehat{\mathbf{\Sigma}} G\rangle_{H^2} + \lambda\|G\|_{H^2}^2 = 1 \quad \text{and} \quad \langle \widehat{G}_{k'}, \widehat{\mathbf{\Sigma}} G\rangle_{H^2} + \lambda\langle \widehat{G}_{k'}, G\rangle_{H^2} = 0,$$

for all $1 \leq k' \leq k-1$. This problem can be further expressed as

$$\max_{G\in H^2} \quad \langle G, \widehat{\mathbf{\Lambda}} G\rangle_{H^2}, \tag{S5.31}$$
$$\text{subject to} \quad \langle G, (\widehat{\mathbf{\Sigma}} + \lambda\mathbf{I})G\rangle_{H^2} = 1 \quad \text{and} \quad \langle \widehat{G}_{k'}, (\widehat{\mathbf{\Sigma}} + \lambda\mathbf{I})G\rangle_{H^2} = 0,$$

where $\mathbf{I}$ is the identity operator in H^2.

Step 2: Transform (S5.30) *and* (S5.31) *to eigenvalue problems in* H^2*.*

Because $\mathbf{\Sigma}$ and $(\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})$ are all positive definite (that is, all their eigenvalues are non-negative) and symmetric, their symmetric square root operators uniquely exist (Theorem 12.33 in Rudin (1991)). Let $\mathbf{\Sigma}^{1/2}$ and $(\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})^{1/2}$ denote the symmetric square root operator of $\mathbf{\Sigma}$ and $\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I}$, respectively. Since all the eigenvalues of $\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I}$ are positive (and greater than λ), it is invertible and $(\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})^{-1/2}$ is the symmetric square root operator of $(\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})^{-1}$.

Lemma S.1. *There exist two bounded operators,* $\widetilde{B}$ *and* $\widetilde{H}$*, in* H^2 *such that*

$$\widehat{\mathbf{\Lambda}} = \widehat{\mathbf{\Sigma}}\widetilde{B}\widehat{\mathbf{\Sigma}} + \widetilde{H}, \quad \mathbf{\Lambda} = \mathbf{\Sigma}\widetilde{B}\mathbf{\Sigma}. \tag{S5.32}$$

Let $\eta_k = \mathbf{\Sigma}^{1/2}G_k$ and $\widehat{\eta}_k = (\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})^{1/2}\widehat{G}_k$. Then the optimization problem (S5.30) can be transformed to

$$\max_{\eta\in H^2} \ \langle \eta, \Gamma\eta\rangle_{H^2}, \quad \text{s.t.} \quad \|\eta\|^2_{H^2}=1, \quad \langle \eta_{k'}, \eta\rangle_{H^2}=0, \ \text{for all } 1\le k'\le k-1, \tag{S5.33}$$

where $\Gamma = \mathbf{\Sigma}^{1/2}\widetilde{B}\mathbf{\Sigma}^{1/2}$ and its solutions are η_k's. The optimization problem (S5.31) can be transformed to

$$\max_{\eta\in H^2} \ \langle \eta, \widehat{\Gamma}\eta\rangle_{H^2}, \quad \text{s.t.} \quad \|\eta\|^2_{H^2}=1, \quad \langle \widehat{\eta}_{k'}, \eta\rangle_{H^2}=0, \ \text{for all } 1\le k'\le k-1, \tag{S5.34}$$

where $\widehat{\Gamma} = (\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})^{-1/2}(\widehat{\mathbf{\Sigma}}\widetilde{B}\widehat{\mathbf{\Sigma}}+\widetilde{H})(\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})^{-1/2} = (\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})^{-1/2}\widehat{\mathbf{\Sigma}}\widetilde{B}\widehat{\mathbf{\Sigma}}(\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})^{-1/2} + (\widehat{\mathbf{\Sigma}}+$

$\lambda\mathbf{I})^{-1/2}\widetilde{H}(\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})^{-1/2}$, and $\widehat{\eta}_k$'s are the solutions. Based on the two transformations, we provide the convergence rate in the next step.

Step 3: Provide the convergence rate for $\frac{1}{n}\sum_{l=1}^{n}\left\|\int_0^1 \widehat{F}(X_l(s),s,\cdot)ds-\int_0^1 F(X_l(s),s,\cdot)ds\right\|_{L^2}^2$.

For simplicity, we assume that $E[F(X(s),s,t)]=0$ for all s,t. Similar to (3.7), the signal function can be written as

$$S_l(t)=\int_0^1 F(X_l(s),s,t)ds=\sum_{k=1}^{\infty} r_l(G_k)\phi_k(t), \tag{S5.35}$$

$$\begin{aligned}\widehat{S}_l(t)&=\int_0^1 \widehat{F}(X_l(s),s,t)ds=\sum_{k=1}^{K}\left\{\int_0^1 \widehat{G}_k(X_l(s),s)ds-\int_0^1 \overline{\widehat{G}}_k(s)ds\right\}\widehat{\phi}_k(t)\\&=\sum_{k=1}^{K}\{r_l(\widehat{G}_k)-\bar{r}(\widehat{G}_k)\}\widehat{\phi}_k(t).\end{aligned} \tag{S5.36}$$

By (S5.35) $\sim$ (S5.36),

$$\begin{aligned}\frac{1}{n}\sum_{l=1}^{n}\|\widehat{S}_l-S_l\|_{L^2}^2&=\frac{1}{n}\sum_{l=1}^{n}\left\|\sum_{k=1}^{K}[r_l(\widehat{G}_k)-\bar{r}(\widehat{G}_k)]\widehat{\phi}_k-\sum_{k=1}^{\infty}r_l(G_k)\phi_k\right\|_{L^2}^2\\&\leq\frac{3}{n}\sum_{l=1}^{n}\left\|\sum_{k=1}^{K}[r_l(\widehat{G}_k)-\bar{r}(\widehat{G}_k)]\widehat{\phi}_k^{(0)}-\sum_{k=1}^{\infty}[r_l(G_k)-\bar{r}(G_k)]\phi_k\right\|_{L^2}^2\\&\quad+\frac{3}{n}\sum_{l=1}^{n}\left\|\sum_{k=1}^{K}[r_l(\widehat{G}_k)-\bar{r}(\widehat{G}_k)][\widehat{\phi}_k-\widehat{\phi}_k^{(0)}]\right\|_{L^2}^2+\frac{3}{n}\sum_{l=1}^{n}\left\|\sum_{k=1}^{\infty}\bar{r}(G_k)\phi_k\right\|_{L^2}^2,\end{aligned} \tag{S5.37}$$

where $\widehat{\phi}_k^{(0)}(t)=\widehat{\mathbf{R}}_k^{\top}\mathbf{Y}(t)/n$ is the function in the problem (S3.17) to which $\widehat{\phi}_k$ is the solution. By the definition (3.7), we have $Y_l(t)=\mu(t)+\int_0^1 F(X_l(s),s,t)ds+\varepsilon_l(t)=$

$\mu(t)+\int_0^1 \{\sum_{k=1}^{\infty} G_k(X_l(s),s)\phi_k(t)\}\, ds+\varepsilon_l(t) = \mu(t)+\sum_{k=1}^{\infty} r_l(G_k)\phi_k(t)+\varepsilon_l(t)$. Therefore,

$$\widehat{\phi}_k^{(0)}(t) = \widehat{\mathbf{R}}_k^{\top}\mathbf{Y}(t)/n = \sum_{l=1}^{n}\left[r_l(\widehat{G}_k)-\overline{r}(\widehat{G}_k)\right]Y_l(t)/n \tag{S5.38}$$

$$= \sum_{l=1}^{n}\left[r_l(\widehat{G}_k)-\overline{r}(\widehat{G}_k)\right][Y_l(t)-\bar{Y}(t)]/n$$

$$= \sum_{l=1}^{n}\sum_{j=1}^{\infty}\frac{1}{n}\left[r_l(\widehat{G}_k)-\overline{r}(\widehat{G}_k)\right][r_l(G_j)-\overline{r}(G_j)]\,\phi_j(t)+\frac{1}{n}\sum_{l=1}^{n}\left[r_l(\widehat{G}_k)-\overline{r}(\widehat{G}_k)\right][\varepsilon_l(t)-\bar{\varepsilon}(t)]$$

$$= \sum_{j=1}^{\infty}\langle\widehat{G}_k,\widehat{\mathbf{\Sigma}}G_j\rangle_{H^2}\phi_j(t)+(\widehat{\Xi}\widehat{G}_k)(t),$$

where $\widehat{\Xi}$ is a operator from H^2 to $L^2[0,1]$ such that for any $G\in J^2$, $(\widehat{\Xi}G)(t) = \sum_{l=1}^{n}\left[r_l(G)-\overline{r}(G)\right]\left[\varepsilon_l(t)-\bar{\varepsilon}(t)\right]/n$. By (S5.38) and noting that $\phi_k(t)$'s are orthogonal to each other and $\|\phi_k\| = \sigma_k$, the first term on the right hand side of (S5.37) can be expressed as

$$\frac{3}{n}\sum_{l=1}^{n}\left\|\sum_{k=1}^{K}[r_l(\widehat{G}_k)-\bar{r}(\widehat{G}_k)]\widehat{\phi}_k^{(0)}-\sum_{k=1}^{\infty}[r_l(G_k)-\bar{r}(G_k)]\phi_k\right\|_{L^2}^2$$

$$\leq\frac{6}{n}\sum_{l=1}^{n}\left\|\sum_{k=1}^{K}[r_l(\widehat{G}_k)-\bar{r}(\widehat{G}_k)]\sum_{j=1}^{\infty}\langle\widehat{G}_k,\widehat{\mathbf{\Sigma}}G_j\rangle_{H^2}\phi_j-\sum_{k=1}^{\infty}[r_l(G_k)-\bar{r}(G_k)]\phi_k\right\|_{L^2}^2$$

$$+\frac{6}{n}\sum_{l=1}^{n}\left\|\sum_{k=1}^{K}[r_l(\widehat{G}_k)-\bar{r}(\widehat{G}_k)]\widehat{\Xi}\widehat{G}_k\right\|_{L^2}^2$$

$$=6\sum_{j=1}^{\infty}\sigma_j^2\langle\widehat{G}_j^{(a)}-G_j,\widehat{\mathbf{\Sigma}}(\widehat{G}_j^{(a)}-G_j)\rangle_{H^2}+6\sum_{k=1}^{K}\sum_{k'=1}^{K}\langle\widehat{G}_k,\widehat{\mathbf{\Sigma}}\widehat{G}_{k'}\rangle_{H^2}\langle\widehat{G}_k,\widehat{\Xi}^*\widehat{\Xi}\widehat{G}_{k'}\rangle_{H^2}, \tag{S5.39}$$

where $\widehat{G}_j^{(a)} = \sum_{k=1}^{K}\langle\widehat{G}_k,\widehat{\mathbf{\Sigma}}G_j\rangle_{H^2}\widehat{G}_k$ and $\widehat{\Xi}^*$ is the adjoint operator of $\widehat{\Xi}$.

Lemma S.2. *For any $\epsilon>0$, there exists an event $\Omega_{n,\epsilon}$ with probability greater than $1-\epsilon$,* *such that in* $\Omega_{n,\epsilon}$*, for any* $1 \le j \le K$*,*

$$\langle \widehat{G}_j^{(a)} - G_j, \widehat{\mathbf{\Sigma}}(\widehat{G}_j^{(a)} - G_j)\rangle_{H^2} \le D_{19}\left[\delta_j^{-2} + \left(\sum_{k=1}^{K}\sigma_k^{-2}\right)^2\right] n^{-1/2},$$

and for any $j > K$*,*

$$\langle \widehat{G}_j^{(a)} - G_j, \widehat{\mathbf{\Sigma}}(\widehat{G}_j^{(a)} - G_j)\rangle_{H^2} \le 1 + D_{20}\left[\delta_K^{-2} + \left(\sum_{k=1}^{K}\sigma_k^{-2}\right)^2\right] n^{-1/2},$$

where $\delta_k = \min_{1\le j\le k}(\sigma_j^2 - \sigma_{j+1}^2)$*, and* D_{19} *and* D_{20} *are two constants which do not depend on* n*.*

In the rest of the proof, we only consider the event $\Omega_{n,\epsilon}$. By Condition 1, we have

$$\delta_j^{-2} = \left\{\min_{1\le l\le j}(\sigma_l^2 - \sigma_{l+1}^2)\right\}^{-2} \le C^{-2} j^{2(\theta+1)}, \quad \sigma_j^2 \le C j^{-\theta}, \quad \sigma_j^{-2} \le C j^{\theta},$$
$$\text{and hence } \left(\sum_{k=1}^{K}\sigma_k^{-2}\right)^2 \le D_{21} K^{2(\theta+1)}. \tag{S5.40}$$

Then by (S5.40) and Lemma S.2,

$$\begin{aligned}
&6\sum_{j=1}^{\infty}\sigma_j^2\langle \widehat{G}_j^{(a)} - G_j, \widehat{\mathbf{\Sigma}}(\widehat{G}_j^{(a)} - G_j)\rangle_{H^2}\\
&\le \left(6D_{19}\sum_{j=1}^{K}\sigma_j^2[\delta_j^{-2} + D_{21}K^{2(\theta+1)}] + 6D_{20}\sum_{j=K+1}^{\infty}\sigma_j^2[\delta_K^{-2} + D_{21}K^{2(\theta+1)}]\right) n^{-1/2} + 6\sum_{j=K+1}^{\infty}\sigma_j^2\\
&\le D_{22}[(K^{\theta+3} + K^{2(\theta+1)})n^{-1/2} + K^{-\theta+1}]\\
&\le D_{23}[K^{2(\theta+1)}n^{-1/2} + K^{-\theta+1}],
\end{aligned} \tag{S5.41}$$

where the inequality in the last line is because $2(\theta+1) > \theta + 3$ when $\theta > 1$. By the definition (S6.61) of $\Omega_{n,\epsilon}$, similar arguments as in the proof of Lemma S.2 lead to an estimate of the second term on the right hand side of (S5.39):

$$6\sum_{k=1}^{K}\sum_{k'=1}^{K}\langle \widehat{G}_k, \widehat{\boldsymbol{\Sigma}}\widehat{G}_{k'}\rangle_{H^2}\langle \widehat{G}_k, \widehat{\Xi}^*\widehat{\Xi}\widehat{G}_{k'}\rangle_{H^2} \le D_{25}n^{-1/2}, \tag{S5.42}$$

Therefore, by (S5.39), (S5.41) and (S5.42), we have

$$\begin{aligned}&\frac{3}{n}\sum_{l=1}^{n}\left\|\sum_{k=1}^{K}[r_l(\widehat{G}_k)-\bar{r}(\widehat{G}_k)]\widehat{\phi}_k^{(0)}-\sum_{k=1}^{\infty}[r_l(G_k)-\bar{r}(G_k)]\phi_k\right\|_{L^2}^2\\&\le D_{23}[K^{2(\theta+1)}n^{-1/2}+K^{-\theta+1}]+D_{25}n^{-1/2}.\end{aligned} \tag{S5.43}$$

By similar arguments, we can obtain the upper bound for the second and third terms on the right hand side of (S5.39),

$$\frac{3}{n}\sum_{l=1}^{n}\left\|\sum_{k=1}^{K}[r_l(\widehat{G}_k)-\bar{r}(\widehat{G}_k)][\widehat{\phi}_k-\widehat{\phi}_k^{(0)}]\right\|_{L^2}^2 \le D_{24}K^{2\theta+3}n^{-1/2}, \tag{S5.44}$$

$$\frac{3}{n}\sum_{l=1}^{n}\left\|\sum_{k=1}^{\infty}\bar{r}(G_k)\phi_k\right\|_{L^2}^2 \le D_{26}n^{-1}, \tag{S5.45}$$

Combining (S5.43), (S5.44) and (S5.45), and noting that $K = C_K n^{1/2(3\theta+2)}$ given in the theorem, we have

$$\frac{1}{n}\sum_{l=1}^{n}\|\widehat{S}_l - S_l\|_{L^2}^2 \le D_1 n^{-(\theta-1)/2(3\theta+2)}. \tag{S5.46}$$

which is the inequality (3.20) in the theorem.

Now we prove the inequality (3.21) in the theorem. Because

$$\begin{aligned}Y_{\text{new}} &= \mu(t) + \int_0^1 F(X_{\text{new}}(s), s, t)ds + \varepsilon_{\text{new}}(t),\\ Y_{\text{pred}} &= \widehat{\mu}(t) + \int_0^1 \widehat{F}(X_{\text{new}}(s), s, t)ds\end{aligned} \tag{S5.47}$$

and $\varepsilon_{\text{new}}(t)$ is independent of $\mathbf{X}(s)$, $\mathbf{Y}(t)$ and $X_{\text{new}}(s)$, we have

$$\begin{aligned}&\mathrm{E}\left[\|Y_{\text{pred}}-Y_{\text{new}}\|_{L^2}^2\,\middle|\,\mathbf{X}(s),\mathbf{Y}(t)\right] \\ =&\left\|\widehat{\mu}(t)+\int_0^1\widehat{F}(X_{\text{new}}(s),s,t)ds-\mu(t)-\int_0^1 F(X_{\text{new}}(s),s,t)ds\right\|_{L^2}^2+\mathrm{E}\left[\|\varepsilon\|_{L^2}^2\right].\end{aligned} \tag{S5.48}$$

The first inequality in (3.21) immediately follows from (S5.48). In order to show the first inequality in (3.21), we just need to show

$$\left\|\widehat{\mu}(t)+\int_0^1\widehat{F}(X_{\text{new}}(s),s,t)ds-\mu(t)-\int_0^1 F(X_{\text{new}}(s),s,t)ds\right\|_{L^2}^2\le D_2 n^{-(\theta-1)/2(3\theta+2)},$$

which can be proved using similar arguments as in the proof of the inequality (S5.46). We skip the details.

S6 Proofs of technical lemmas

S.6.1 The proof of Lemma $S.1$

By the definition (3.7), we have $Y_l(t)=\mu(t)+\int_0^1 F(X_l(s),s,t)ds+\varepsilon_l(t)=\mu(t)+\int_0^1\left\{\sum_{k=1}^\infty G_k(X_l(s),s)\phi_k(t)\right\}ds+\varepsilon_l(t)=\mu(t)+\sum_{k=1}^\infty r_l(G_k)\phi_k(t)+\varepsilon_l(t)$, which, together with the definition (3.14) of $\widehat{\mathbf{\Lambda}}(G,G)$, leads to

$$\begin{aligned}
&\langle G, \widehat{\mathbf{\Lambda}} G\rangle_{H^2} = \widehat{\mathbf{\Lambda}}(G,G) = \frac{1}{n^2}\int_0^1 \left[\sum_{l=1}^n \{r_l(G) - \overline{r}(G)\}\{Y_l(t) - \overline{Y}(t)\}\right]^2 dt \\
=&\frac{1}{n^2}\int_0^1 \left[\sum_{k=1}^\infty \left(\sum_{l=1}^n \{r_l(G) - \overline{r}(G)\}\{r_l(G_k) - \overline{r}(G_k)\}\right)\phi_k(t) + \sum_{l=1}^n \{r_l(G) - \overline{r}(G)\}\{\varepsilon_l(t) - \overline{\varepsilon_l}(t)\}\right]^2 dt \\
=&\frac{1}{n^2}\int_0^1 \left[\sum_{k=1}^\infty n\langle G, \widehat{\mathbf{\Sigma}} G_k\rangle_{H^2}\phi_k(t) + \sum_{l=1}^n \{r_l(G) - \overline{r}(G)\}\{\varepsilon_l(t) - \overline{\varepsilon_l}(t)\}\right]^2 dt \\
=&\sum_{k=1}^\infty \langle G, \widehat{\mathbf{\Sigma}} G_k\rangle_{H^2}^2 \sigma_k^2 + \langle G, \widetilde{H} G\rangle_{H^2} = \langle G, (\widehat{\mathbf{\Sigma}}\widetilde{B}\widehat{\mathbf{\Sigma}} + \widetilde{H})G\rangle_{H^2},
\end{aligned}$$

where $\widetilde{B}$ and $\widetilde{H}$ are two bounded operators. Therefore, we have $\widehat{\mathbf{\Lambda}} = \widehat{\mathbf{\Sigma}}\widetilde{B}\widehat{\mathbf{\Sigma}} + \widetilde{H}$. Similarly, we can show that $\mathbf{\Lambda} = \mathbf{\Sigma}\widetilde{B}\mathbf{\Sigma}$.

S.6.2 The proof of Lemma *S.2*

We use $|||\cdot|||$ to denote the operator norm. We split the proof into several steps.

Step 1: provide an upper bound for $|||\widehat{\Gamma} - \Gamma|||$.

$$\begin{aligned}
&|||\widehat{\Gamma} - \Gamma||| \qquad\qquad (S6.49)\\
=&|||(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2}\widehat{\mathbf{\Sigma}}\widetilde{B}\widehat{\mathbf{\Sigma}}(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2} + (\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2}\widetilde{H}(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2} - \mathbf{\Sigma}^{1/2}\widetilde{B}\mathbf{\Sigma}^{1/2}||| \\
\le&|||(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2}\widehat{\mathbf{\Sigma}}\widetilde{B}\widehat{\mathbf{\Sigma}}(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2} - \mathbf{\Sigma}^{1/2}\widetilde{B}\mathbf{\Sigma}^{1/2}||| + |||(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2}\widetilde{H}(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2}||| \\
\le&|||(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2}\widehat{\mathbf{\Sigma}}||| \cdot \|\widetilde{B}\| \cdot |||(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2}\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}^{1/2}||| \\
&+ |||(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2}\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}^{1/2}||| \cdot \|\widetilde{B}\| \cdot |||\mathbf{\Sigma}^{1/2}||| + |||(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2}\widetilde{H}(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2}|||.
\end{aligned}$$

We first estimate

$$
\begin{aligned}
& |||(\widehat{\boldsymbol{\Sigma}}+\lambda\mathbf{I})^{-1/2}\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}^{1/2}||| \\
\leq & |||(\widehat{\boldsymbol{\Sigma}}+\lambda\mathbf{I})^{-1/2}\widehat{\boldsymbol{\Sigma}}-(\boldsymbol{\Sigma}+\lambda\mathbf{I})^{-1/2}\boldsymbol{\Sigma}|||+|||(\boldsymbol{\Sigma}+\lambda\mathbf{I})^{-1/2}\boldsymbol{\Sigma}-\boldsymbol{\Sigma}^{1/2}||| \\
= & |||f(\widehat{\boldsymbol{\Sigma}})-f(\boldsymbol{\Sigma})|||+|||g(\boldsymbol{\Sigma})|||, \qquad \text{(S6.50)}
\end{aligned}
$$

where $f(z)=(z+\lambda)^{-1/2}z$ and $g(z)=f(z)-z^{1/2}$. Because both $f(z)$ and $g(z)$ are analytic functions in the domain $G=\{z=x+iy: x>-\lambda/2\}$ of the complex plane, it follows the theory in Section 10.26 of Rudin (1991) that

$$
\begin{aligned}
& f(\widehat{\boldsymbol{\Sigma}})-f(\boldsymbol{\Sigma}) \\
= & \int_{C_1+C_2+C_3+C_4} f(z)(z\mathbf{I}-\widehat{\boldsymbol{\Sigma}})^{-1}dz-\int_{C_1+C_2+C_3+C_4} f(z)(z\mathbf{I}-\boldsymbol{\Sigma})^{-1}dz, \qquad \text{(S6.51)}
\end{aligned}
$$

where C_i, $1\leq i\leq 4$, are four segments in the domain G shown in Figure S.13.

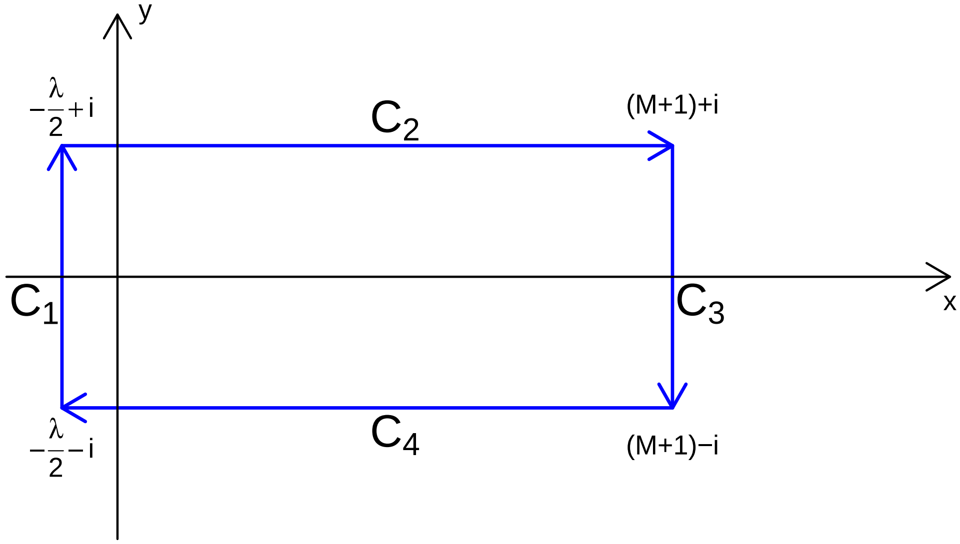

Figure S.13: The contour for the integration in (S6.51) and M is the larger one of the maximum eigenvalues of $\widehat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$.

Now we estimate

$$\left|\int_{C_1} f(z)(z\mathbf{I}-\widehat{\mathbf{\Sigma}})^{-1}dz - \int_{C_1} f(z)(z\mathbf{I}-\mathbf{\Sigma})^{-1}dz\right| = \left|\int_{C_1} f(z)\left[(z\mathbf{I}-\widehat{\mathbf{\Sigma}})^{-1} - (z\mathbf{I}-\mathbf{\Sigma})^{-1}\right]dz\right|$$

$$\leq \int_{C_1} |f(z)| \cdot |||(z\mathbf{I}-\widehat{\mathbf{\Sigma}})^{-1} - (z\mathbf{I}-\mathbf{\Sigma})^{-1}|||dz. \tag{S6.52}$$

Define

$$\Delta = \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}. \tag{S6.53}$$

Then $(z\mathbf{I}-\widehat{\mathbf{\Sigma}})^{-1} - (z\mathbf{I}-\mathbf{\Sigma})^{-1} = (z\mathbf{I}-\widehat{\mathbf{\Sigma}})^{-1}(\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma})(z\mathbf{I}-\mathbf{\Sigma})^{-1} = (z\mathbf{I}-\widehat{\mathbf{\Sigma}})^{-1}\Delta(z\mathbf{I}-\mathbf{\Sigma})^{-1}$.

For any $z \in C_1$,

$$|||(z\mathbf{I} - \widehat{\mathbf{\Sigma}})^{-1} - (z\mathbf{I} - \mathbf{\Sigma})^{-1}||| \leq |||(z\mathbf{I} - \widehat{\mathbf{\Sigma}})^{-1}||| \cdot |||\Delta||| \cdot |||(z\mathbf{I} - \mathbf{\Sigma})^{-1}||| \leq |z|^{-2}|||\Delta|||.$$

where the last inequality follows from the fact that $|||(z\mathbf{I} - \mathbf{\Sigma})^{-1}|||$ is less than or equal to the largest one of $|(z-\mu_1)^{-1}|, |(z-\mu_2)^{-1}|, \cdots$, and $|||(z\mathbf{I} - \widehat{\mathbf{\Sigma}})^{-1}|||$ is less than or equal to the largest one of $|(z-\widehat{\mu}_1)^{-1}|, |(z-\widehat{\mu}_2)^{-1}|, \cdots$, (see inequality (3) in Section 12.24 in Rudin (1991)). μ_k's and $\widehat{\mu}_k$'s are the eigenvalues of $\mathbf{\Sigma}$ and $\widehat{\mathbf{\Sigma}}$, and all of them are nonnegative. Therefore, for any $z \in C_1$, $|(z-\mu_k)^{-1}|$'s and $|(z-\widehat{\mu}_k)^{-1}|$'s are all less than or equal to $|z|^{-1}$. Then by (S6.52),

$$\begin{aligned} &\left| \int_{C_1} f(z)(z\mathbf{I} - \widehat{\mathbf{\Sigma}})^{-1} dz - \int_{C_1} f(z)(z\mathbf{I} - \mathbf{\Sigma})^{-1} dz \right| \\ &\leq \int_{C_1} |f(z)| \cdot |z|^{-2} \cdot |dz| \cdot |||\Delta||| \leq D_3 \lambda^{-1/2} |||\Delta|||, \end{aligned}$$

where D_3 is a constant. By a similar argument, we can obtain

$$\left| \int_{C_2+C_3+C_4} f(z)(z\mathbf{I} - \widehat{\mathbf{\Sigma}})^{-1} dz - \int_{C_2+C_3+C_4} f(z)(z\mathbf{I} - \mathbf{\Sigma})^{-1} dz \right| \leq D_4 |||\Delta|||.$$

Therefore, by (S6.51), we have

$$|||(\widehat{\mathbf{\Sigma}} + \lambda\mathbf{I})^{-1/2}\widehat{\mathbf{\Sigma}} - (\mathbf{\Sigma} + \lambda\mathbf{I})^{-1/2}\mathbf{\Sigma}||| = \left| f(\widehat{\mathbf{\Sigma}}) - f(\mathbf{\Sigma}) \right| \leq D_5 \lambda^{-1/2} |||\Delta|||, \tag{S6.54}$$

By the inequality (3) in Section 12.24 in Rudin (1991), we estimate the second term in (S6.50),

$$\begin{aligned} &|||(\mathbf{\Sigma} + \lambda\mathbf{I})^{-1/2}\mathbf{\Sigma} - \mathbf{\Sigma}^{1/2}||| = |||g(\mathbf{\Sigma})||| \leq \sup_{k\geq 1} |g(\mu_k)| \leq \sup_{k \geq 1} |f(\mu_k) - \mu_k^{1/2}| \qquad \text{(S6.55)} \\ &= \sup_{k\geq 1} \left| \frac{\lambda\sqrt{\mu_k}}{\left(\sqrt{\mu_k + \lambda} + \sqrt{\mu_k}\right)\sqrt{\mu_k + \lambda}} \right| \leq \lambda^{1/2}. \end{aligned}$$

(S6.50), (S6.54) and (S6.55) lead to

$$|||(\widehat{\mathbf{\Sigma}} + \lambda\mathbf{I})^{-1/2}\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}^{1/2}||| \leq D_5 \lambda^{-1/2} |||\Delta||| + \lambda^{1/2}. \tag{S6.56}$$

By the same argument as in (S6.55), we have

$$|||(\mathbf{\Sigma} + \lambda \mathbf{I})^{-1/2}||| \leq \sup_{k \geq 1} \left| \frac{1}{\sqrt{\mu_k + \lambda}} \right| \leq \lambda^{-1/2}. \tag{S6.57}$$

By the central limit theorem in Hilbert space, we have

$$E\left[|||\Delta|||^2\right] = E\left[|||\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}|||^2\right] \leq \frac{D_1}{n}, \quad E\left[|||\widehat{\Xi}|||^2\right] \leq \frac{D_2}{n}. \tag{S6.58}$$

With similar arguments as above and (S6.58), we have

$$E\left[|||(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2} \widetilde{H} (\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2}|||\right] \leq D_6 \left[\lambda^{-1/2} n^{-1/2} + \lambda^{-1} n^{-1}\right], \tag{S6.59}$$

where D_1, D_2 and D_6 are constants and $\widehat{\Xi}$ is the operator defined in the last line of (S5.38). Now by (S6.49), (S6.56) $\sim$ (S6.59),

$$\begin{aligned} & E[|||\widehat{\Gamma} - \Gamma|||] \\ \leq & E[|||(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2} \widehat{\mathbf{\Sigma}}||| \cdot \|\widetilde{B}\| \cdot |||(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2} \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}^{1/2}|||] \\ & + E[|||(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2} \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}^{1/2}||| \cdot \|\widetilde{B}\| \cdot |||\mathbf{\Sigma}^{1/2}|||] + E[|||(\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2} \widetilde{H} (\widehat{\mathbf{\Sigma}} + \lambda \mathbf{I})^{-1/2}|||] \\ \leq & D_7 \left(\lambda^{-1/2} n^{-1/2} + \lambda^{1/2} + \lambda^{-1/2} n^{-1/2} + \lambda^{-1} n^{-1}\right) \leq D_8 n^{-1/4}. \end{aligned} \tag{S6.60}$$

where we use the condition $\lambda = C_\lambda n^{-1/2}$. For any $\epsilon > 0$, define the event

$$\begin{aligned} \Omega_{n,\epsilon} = \{ & |||\widehat{\Gamma} - \Gamma||| \leq D_8 \epsilon^{-1} n^{-1/4}/3, \quad |||\Delta||| \leq \sqrt{D_1} \epsilon^{-1} n^{-1/2}/3, \\ & \text{and} \quad \|\widehat{\Xi}\| \leq \sqrt{D_2} \epsilon^{-1} n^{-1/2}/3 \}. \end{aligned} \tag{S6.61}$$

Then by (S6.58), (S6.60) and the Markov inequality, the probability of $\Omega_{n,\epsilon}$ is greater than $1 - \epsilon$. In the rest of the proof of the lemma, we only consider the event $\Omega_{n,\epsilon}$.

Step 3: provide upper bounds for $\|\widehat{\eta}_k - \eta_k\|$, $k \geq 1$.

We first note that the eigenvalues of Γ are the same as the eigenvalues $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq 0$ of the problem (S5.30) (which is equivalent to the problem (3.11) in the main manuscript). As the inequalities in (5.2) in Hall et al. (2007), by the inequality (2.1) in Bhatia et al. (1983) and (S6.60), we have

$$\sup_{k\geq 1}\delta_k\|\widehat{\eta}_k-\eta_k\|\leq\sqrt{8}|||\widehat{\Gamma}-\Gamma|||\leq D_9 n^{-1/4}, \tag{S6.62}$$

where we recall that $\delta_k=\min_{1\leq j\leq k}(\sigma_j^2-\sigma_{j+1}^2)$, the last inequality follows from the definition (S6.61) of $\Omega_{n,\epsilon}$ and $D_9=\sqrt{8}D_8\epsilon^{-1}/3$. Then it follows that

$$\sup_{k\geq 1}\delta_k^2|1-\langle\widehat{\eta}_k,\eta_k\rangle|=\frac{1}{2}\sup_{k\geq 1}\delta_k^2\|\widehat{\eta}_k-\eta_k\|^2\leq\frac{1}{2}D_9^2 n^{-1/2} \tag{S6.63}$$

For any $1\leq k\leq m$, let $\widehat{\mathbf{P}}$ be the orthogonal projection operator onto the subspace spanned by $\{\widehat{\eta}_{k'}:k'\in\mathcal{A}\}$, where $\mathcal{A}$ is any subset of positive integers but does not include k. Then by the inequality (2.1) in Bhatia et al. (1983), we have

$$\delta_k\|\widehat{\mathbf{P}}\eta_k\|\leq|||\widehat{\Gamma}-\Gamma|||\leq D_9 n^{-1/4}. \tag{S6.64}$$

Step 4: provide upper bounds for $\sigma_k^2\|G_k\|_{H^2}^2$ and $\sigma_k^2\|\widehat{G}_k\|_{H^2}^2$, $k\geq 1$.

Because for any $k\geq 1$, we have $G_k(x,s)=\int_0^1 F(x,s,t)\phi_k(t)dt/\sigma_k^2$,

$$\begin{aligned}&\|G_k\|_{L^2}^2=\int_0^1\int_0^1 G_k(x,s)^2dxds=\sigma_k^{-4}\int_0^1\int_0^1\left[\int_0^1 F(x,s,t)\phi_k(t)dt\right]^2dxds\\&\leq\sigma_k^{-4}\int_0^1\int_0^1\int_0^1 F(x,s,t)^2dtdxds\int_0^1\phi_k(t)^2dt=\sigma_k^{-2}\int_0^1\int_0^1\int_0^1 F(x,s,t)^2dtdxds\\&=\sigma_k^{-2}\|F\|_{L^2}^2.\end{aligned}$$

Similarly, we have $\|\partial_{xx}G_k\|_{L^2}^2\leq\sigma_k^{-2}\|\partial_{xx}F\|_{L^2}^2$, $\|\partial_{xs}G_k\|_{L^2}^2\leq\sigma_k^{-2}\|\partial_{xs}F\|_{L^2}^2$ and $\|\partial_{ss}G_k\|_{L^2}^2\leq\sigma_k^{-2}\|\partial_{ss}F\|_{L^2}^2$. Hence, we have

$$\sigma_k^2\|G_k\|_{H^2}^2=\sigma_k^2[\|G_k\|_{L^2}^2+\|\partial_{xx}G_k\|_{L^2}^2+\|\partial_{xs}G_k\|_{L^2}^2+\|\partial_{ss}G_k\|_{L^2}^2]\leq D_{10}, \tag{S6.65}$$

for all $k\geq 1$, where $D_{10}=\|F\|_{L^2}^2+\|\partial_{xx}F\|_{L^2}^2+\|\partial_{xs}F\|_{L^2}^2+\|\partial_{ss}F\|_{L^2}^2$.

To provide the upper bound for $\sigma_1^2\|\widehat{G}_1\|^2$, we consider the following inequality

$$\frac{\langle G_1, \widehat{\mathbf{\Lambda}} G_1\rangle_{H^2}}{\langle G_1, (\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})G_1\rangle_{H^2}} \leq \frac{\langle \widehat{G}_1, \widehat{\mathbf{\Lambda}} \widehat{G}_1\rangle_{H^2}}{\langle \widehat{G}_1, (\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})\widehat{G}_1\rangle_{H^2}} = \langle \widehat{G}_1, \widehat{\mathbf{\Lambda}} \widehat{G}_1\rangle_{H^2}, \tag{S6.66}$$

where the inequality is because $\widehat{G}_1$ is the solution to (S5.31) and the equality is because of $\langle \widehat{G}_1, (\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})\widehat{G}_1\rangle = 1$. By a tedious calculation, it follows from (S6.66) that in the event $\Omega_{n,\epsilon}$, we have $\sigma_1^2\|\widehat{G}_1\|^2 \leq D_{11}$, where D_{11} is a constant. For a general k, we can similarly obtain

$$\sigma_k^2\|\widehat{G}_k\|^2 \leq D_{11}, \tag{S6.67}$$

<u>*Step 5: prove the inequalities in the lemma.*</u>

Because $\eta_k = \mathbf{\Sigma}^{1/2}G_k$ and $\widehat{\eta}_k = (\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})^{1/2}\widehat{G}_k$, we have

$$\widehat{\mathbf{\Sigma}}^{1/2}\widehat{G}_k = \widehat{\eta}_k + \widehat{q}_k, \quad \widehat{\mathbf{\Sigma}}^{1/2}G_k = \eta_k + q_k, \tag{S6.68}$$

where $\widehat{q}_k = \left[(\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})^{1/2} - \widehat{\mathbf{\Sigma}}^{1/2}\right]\widehat{G}_k$ and $q_k = \left[\widehat{\mathbf{\Sigma}}^{1/2} - \mathbf{\Sigma}^{1/2}\right]G_k$. By similar arguments as in Step 2, we can show that $|||(\widehat{\mathbf{\Sigma}}+\lambda\mathbf{I})^{1/2} - \widehat{\mathbf{\Sigma}}^{1/2}||| \leq D_{12}n^{-1/4}$ and $|||\widehat{\mathbf{\Sigma}}^{1/2} - \mathbf{\Sigma}^{1/2}||| \leq D_{13}n^{-1/4}$. By (S6.65) and (S6.67),

$$\|\widehat{q}_k\| \leq D_{14}\sigma_k^{-2}n^{-1/4}, \quad \|q_k\| \leq D_{15}\sigma_k^{-2}n^{-1/4}. \tag{S6.69}$$

For any $1 \leq j \leq K$,

$$\langle \widehat{G}_j^{(a)} - G_j, \widehat{\mathbf{\Sigma}}(\widehat{G}_j^{(a)} - G_j)\rangle = \|\widehat{\mathbf{\Sigma}}^{1/2}\widehat{G}_j^{(a)} - \widehat{\mathbf{\Sigma}}^{1/2}G_j\|^2 = \|\sum_{k=1}^{K}\langle \widehat{G}_k, \widehat{\mathbf{\Sigma}}G_j\rangle\widehat{\mathbf{\Sigma}}^{1/2}\widehat{G}_k - \widehat{\mathbf{\Sigma}}^{1/2}G_j\|^2$$

$$\leq 2\|\langle \widehat{G}_j, \widehat{\mathbf{\Sigma}}G_j\rangle\widehat{\mathbf{\Sigma}}^{1/2}\widehat{G}_j - \widehat{\mathbf{\Sigma}}^{1/2}G_j\|^2 + 2\|\sum_{k\neq j,k=1}^{K}\langle \widehat{G}_k, \widehat{\mathbf{\Sigma}}G_j\rangle\widehat{\mathbf{\Sigma}}^{1/2}\widehat{G}_k\|^2. \tag{S6.70}$$

By (S6.62), (S6.63), (S6.69), we have

$$\|\langle \widehat{G}_j, \widehat{\mathbf{\Sigma}}G_j\rangle\widehat{\mathbf{\Sigma}}^{1/2}\widehat{G}_j - \widehat{\mathbf{\Sigma}}^{1/2}G_j\|^2 = \|\langle \widehat{\eta}_j + \widehat{q}_j, \eta_j + q_j\rangle(\widehat{\eta}_j + \widehat{q}_j) - (\eta_j + q_j)\|^2$$

$$\leq D_{16}[\delta_j^{-2} + \sigma_j^{-4}]n^{-1/2} \leq 2D_{16}\delta_j^{-2}n^{-1/2} \tag{S6.71}$$

where the last inequality is because $\delta_j = \min_{1\leq l\leq j}(\sigma_l^2 - \sigma_{l+1}^2) \leq \sigma_j^2$. Similarly,

$$\|\sum_{k\neq j,k=1}^{K}\langle\widehat{G}_k, \widehat{\boldsymbol{\Sigma}}G_j\rangle\widehat{\boldsymbol{\Sigma}}^{1/2}\widehat{G}_k\|^2 = \|\sum_{k\neq j,k=1}^{K}\langle\widehat{\eta}_k + \widehat{q}_k, \eta_j + q_j\rangle(\widehat{\eta}_k + \widehat{q}_k)\|^2 \tag{S6.72}$$

$$\leq D_{17}\left[\|\sum_{k\neq j,k=1}^{K}\langle\widehat{\eta}_k, \eta_j\rangle\widehat{\eta}_k\|^2 + (\sum_{k=1}^{K}\|\widehat{q}_k\|)^2 + \|q_j\|^2\right] \leq D_{18}\left[\delta_j^{-2} + \left(\sum_{k=1}^{K}\sigma_k^{-2}\right)^2\right]n^{-1/2}$$

where the last inequality follows from (S6.64), (S6.69) and the fact that $\sum_{k\neq j,k=1}^{K}\langle\widehat{\eta}_k, \eta_j\rangle\widehat{\eta}_k$ is the orthogonal projection of η_j onto the space spanned by $\{\widehat{\eta}_k : 1 \leq k \leq K, k \neq j\}$. Combining (S6.70), (S6.71) and (S6.72) leads to

$$\langle\widehat{G}_j^{(a)} - G_j, \widehat{\boldsymbol{\Sigma}}(\widehat{G}_j^{(a)} - G_j)\rangle \leq D_{19}\left[\delta_j^{-2} + \left(\sum_{k=1}^{K}\sigma_k^{-2}\right)^2\right]n^{-1/2},$$

for any $1 \leq l \leq K$. Similarly, for any $l > K$, we have

$$\langle\widehat{G}_j^{(a)} - G_j, \widehat{\boldsymbol{\Sigma}}(\widehat{G}_j^{(a)} - G_j)\rangle \leq 1 + D_{20}\left[\delta_K^{-2} + \left(\sum_{k=1}^{K}\sigma_k^{-2}\right)^2\right]n^{-1/2}.$$

Bibliography

Adams, R. A. and J. J. Fournier (2003). *Sobolev spaces, Volume 140 of Pure and Applied Mathematics*. Academic Press, Amsterdam.

Bhatia, R., C. Davis, and A. McIntosh (1983). Perturbation of spectral subspaces and solution of linear operator equations. *Linear Algebra and its Applications 52*, 45–67.

Hall, P., J. L. Horowitz, et al. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics 35*(1), 70–91.

Luo, R. and X. Qi (2017). Function-on-function linear regression by signal compression. *Journal of the American Statistical Association 112*(518), 690–705.

Rasmussen, C. E. and C. K. I. Williams (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning series)*, Chapter 4. The MIT Pres.

Rudin, W. (1991). *Functional Analysis*. International series in pure and applied mathematics. McGraw-Hill.