Variable selection for functional linear models with strong heredity constraint

Sanying Feng[a], Menghan Zhang[a] and Tiejun Tong[b]

[a] *School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, China*

[b] *Department of Mathematics, Hong Kong Baptist University, Hong Kong*

Supplementary Material

This is a supplement to the paper "Variable selection for functional linear models with strong heredity constraint", in which it contains the proofs of Theorems 1–2, and Lemmas 1–2 and their proofs.

S1 Appendix A: Proofs of theorems

We provide the proofs of Theorems 1–2 in Appendix A.

S1.1 Proof of Theorem 1

For part (i), a simple calculation yields

$$
\begin{aligned}
&\|\hat{\beta}_j(t)-\beta_j^*(t)\|^2 = \|\sum_{k=1}^{K_j}\hat{\beta}_{jk}\hat{\phi}_{jk}(t)-\sum_{k=1}^{\infty}\beta_{jk}^*\phi_{jk}(t)\|^2 \\
\leq\ & 2\|\sum_{k=1}^{K_j}\hat{\beta}_{jk}\hat{\phi}_{jk}(t)-\sum_{k=1}^{K_j}\beta_{jk}^*\phi_{jk}(t)\|^2+2\|\sum_{k=K_j+1}^{\infty}\beta_{jk}^*\phi_{jk}(t)\|^2 \\
\leq\ & 4\|\sum_{k=1}^{K_j}(\hat{\beta}_{jk}-\beta_{jk}^*)\hat{\phi}_{jk}(t)\|^2+4\|\sum_{k=1}^{K_j}\beta_{jk}^*(\hat{\phi}_{jk}(t)-\phi_{jk}(t))\|^2+2\sum_{k=K_j+1}^{\infty}\beta_{jk}^{*2} \\
\leq\ & 4\sum_{k=1}^{K_j}(\hat{\beta}_{jk}-\beta_{jk}^*)^2+8K_j\sum_{k=1}^{K_j}\beta_{jk}^{*2}\|\hat{\phi}_{jk}(t)-\phi_{jk}(t)\|^2+2\sum_{k=K_j+1}^{\infty}\beta_{jk}^{*2}.
\end{aligned}
$$

Note that

$$\sum_{k=K_j+1}^{\infty} \beta_{jk}^{*2} \le \sum_{k=K_j+1}^{\infty} k^{-2b} = O(K_j^{-(2b-1)}) = O(K^{-(2b-1)}).$$

Moreover, invoking Lemma 1 and condition (A3), we have

$$K_j \sum_{k=1}^{K_j} \beta_{jk}^{*2} \|\hat{\phi}_{jk}(t) - \phi_{jk}(t)\|^2 \le n^{-1} K_j \sum_{k=1}^{K_j} j^{-2b+2} = O_p(n^{-1}K_j) = O_p(n^{-1}K).$$

Hence, invoking Lemma 2, we complete the proof of part (i).

For part (ii), the proof is similar and so is omitted. $\square$

S1.2 Proof of Theorem 2

We first consider $P(\hat{\beta}_j(t) = 0 \text{ for } j \in A_1^c) \to 1$. Suppose that there exists a $k_0 \in A_1^c$ such that $\hat{\beta}_{k_0}(t) \neq 0$, then $\|\hat{\beta}_{k_0}\|_2 > 0$. For such k_0, let $\breve{\beta}$ denote the vector whose entries $\breve{\beta}_j$ equal $\hat{\beta}_j$ except for $j = k_0$ and $\breve{\beta}_{k_0} = 0$. Then,

$$\begin{aligned}
& L(\hat{\beta}, \hat{\alpha}) - L(\breve{\beta}, \hat{\alpha}) \\
= \; & \sum_{i=1}^n (Y_i - \hat{U}_i^\top \hat{\beta} - \hat{W}_i^\top \hat{\alpha})^2 - \sum_{i=1}^n (Y_i - \hat{U}_i^\top \breve{\beta} - \hat{W}_i^\top \hat{\alpha})^2 + n\lambda_{1k_0} \|\hat{\beta}_{k_0}\|_2 \\
= \; & \sum_{i=1}^n (\breve{\beta} - \hat{\beta})^\top \hat{U}_i \hat{U}_i^\top (\breve{\beta} - \hat{\beta}) - 2\sum_{i=1}^n (Y_i - \hat{U}_i^\top \breve{\beta} - \hat{W}_i^\top \hat{\alpha}) \hat{U}_i^\top (\hat{\beta} - \breve{\beta}) + n\lambda_{1k_0} \|\hat{\beta}_{k_0}\|_2 \\
\ge \; & -2\sum_{i=1}^n (\beta^* - \breve{\beta})^\top \hat{U}_i \hat{U}_i^\top (\hat{\beta} - \breve{\beta}) - 2\sum_{i=1}^n (Y_i - \hat{U}_i^\top \beta^* - \hat{W}_i^\top \hat{\alpha}) \hat{U}_i^\top (\hat{\beta} - \breve{\beta}) + n\lambda_{1k_0} \|\hat{\beta}_{k_0}\|_2 \\
\equiv \; & -2D_1 - 2D_2 + n\lambda_{1k_0} \|\hat{\beta}_{k_0}\|_2.
\end{aligned}$$

By Lemma 1, Lemma 2 and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
D_1 \; \le \; & \Big\{ \sum_{i=1}^n (\beta^* - \breve{\beta})^\top \hat{U}_i \hat{U}_i^\top (\beta^* - \breve{\beta}) \Big\}^{1/2} \Big\{ \sum_{i=1}^n (\hat{\beta} - \breve{\beta})^\top \hat{U}_i \hat{U}_i^\top (\hat{\beta} - \breve{\beta}) \Big\}^{1/2} \\
= \; & \Big\{ \sum_{i=1}^n (\beta^* - \hat{\beta} + \hat{\beta} - \breve{\beta})^\top (U_i U_i^\top + O_p(n^{-1/2} K^{1-a/2}))(\beta^* - \hat{\beta} + \hat{\beta} - \breve{\beta}) \Big\}^{1/2} \\
& \times \Big\{ \sum_{i=1}^n (\hat{\beta} - \breve{\beta})^\top (U_i U_i^\top + O_p(n^{-1/2} K^{1-a/2}))(\hat{\beta} - \breve{\beta}) \Big\}^{1/2} \\
= \; & O_p(\sqrt{nK}) \|\hat{\beta}_{k_0}\|_2.
\end{aligned}$$

In addition, a simple calculation yields

$$
\begin{aligned}
D_2 &= \sum_{i=1}^{n}(\varepsilon_i + (U_i - \hat{U}_i)^\top \beta^* + W_i^\top \alpha - \hat{W}_i^\top \hat{\alpha} + R_i)(U_i + o_p(1))^\top (\hat{\beta} - \check{\beta}) \\
&= O_p(n^{1/2} + n^{1/2}K^{1-a/2} + n^{1/2}K^{1-a/2} + n^{1/2}K^{1/2})\|\hat{\beta}_{k_0}\|_2 \\
&= O_p(\sqrt{nK})\|\hat{\beta}_{k_0}\|_2.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
L(\hat{\beta}, \hat{\alpha}) - L(\check{\beta}, \hat{\alpha}) &= O_p(\sqrt{nK})\|\hat{\beta}_{k_0}\|_2 + n\lambda_{1k_0}\|\hat{\beta}_{k_0}\|_2 \\
&= \sqrt{nK}\{O_p(1) + \sqrt{n/K}\lambda_{1k_0}\}\|\hat{\beta}_{k_0}\|_2.
\end{aligned}
$$

Invoking condition (A5), the second term dominates the first term. Consequently, we have

$$
L(\hat{\beta}, \hat{\alpha}) - L(\check{\beta}, \hat{\alpha}) > 0
$$

with probability tending to one, which contradicts to the fact that $(\hat{\beta}, \hat{\alpha})$ is the minimizer of $L(\beta, \alpha)$. This completes the proof of the first part of the theorem.

Next, we prove $P(\hat{\gamma}_{jm}(s,t) = 0$ for $(j,m) \in A_2^c) \to 1$. For (j,m) where $(j,m) \in A_2^c$ and j, $m \in A_1$: we can prove $P(\hat{\gamma}_{jm}(s,t) = 0) \to 1$ in a similar way. For (j,m) where $(j,m) \in A_2^c$ and either j or m is in A_1^c: without loss of generality, assume that $\|\beta_j^*(t)\| = 0$. Notice that $\|\hat{\beta}_j(t)\| = 0$ implies $\|\hat{\alpha}_{jm}\|_2 = 0$, because if $\|\hat{\alpha}_{jm}\|_2 \neq 0$, then the value of the loss function does not change but the value of the penalty function will increase. Since we already have $P(\hat{\beta}_j(t) = 0) \to 1$, we can conclude $P(\hat{\gamma}_{jm}(s,t) = 0) \to 1$ as well. $\square$

S2 Appendix B: Some lemmas and their proofs

In order to prove Theorems 1–2, we provide Lemmas 1–2 in Appendix B.

Lemma 1. *Assume that conditions (A1)–(A4) hold. Then we have*

(i) $|\hat{\xi}_{ijk} - \xi_{ijk}| = O_p(n^{-1/2}k^{1-a/2})$,

(ii) $\left|\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_{ijk}\hat{\xi}_{ijl} - E(\xi_{ijk}\xi_{ijl})\right| = O_p\left(n^{-1/2}\min\{k^{1-a/2}, l^{1-a/2}\}\right)$,

(iii) $R_i = O_p(K^{-(2b+a-1)/2})$,

(iv) $\quad \|\hat{\phi}_{jk}(t) - \phi_{jk}(t)\| = O(n^{-1/2}k)$,

where $R_i = \sum_{j=1}^{p}\sum_{k=K_j+1}^{\infty}\beta^*_{jk}\xi_{ijk} + \sum\limits_{j<m}\sum_{k=K_j+1}^{\infty}\sum_{l=K_m+1}^{\infty}\gamma^*_{jm,kl}\xi_{ijk}\xi_{iml}$.

Proof. For part (i), by conditions (A1)–(A3) and the similar argument as in the proof of Proposition 1 in Wong et al. (2019), we can obtain that $|\hat{\xi}_{ijk} - \xi_{ijk}| = O_p(n^{-1/2}k^{1-a/2})$.

For part (ii), it can be observed that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_{ijk}\hat{\xi}_{ijl} - E(\xi_{ijk}\xi_{ijl})\\
=\;&\Big(\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_{ijk}\hat{\xi}_{ijl} - \frac{1}{n}\sum_{i=1}^{n}\xi_{ijk}\xi_{ijl}\Big) + \Big(\frac{1}{n}\sum_{i=1}^{n}\xi_{ijk}\xi_{ijl} - E(\xi_{ijk}\xi_{ijl})\Big).
\end{aligned}
$$

Invoking part (i), we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_{ijk}\hat{\xi}_{ijl} - \frac{1}{n}\sum_{i=1}^{n}\xi_{ijk}\xi_{ijl}\\
=\;&\frac{1}{n}\sum_{i=1}^{n}\Big[(\hat{\xi}_{ijk}-\xi_{ijk})\hat{\xi}_{ijl} + \xi_{ijk}(\hat{\xi}_{ijl}-\xi_{ijl})\Big]\\
=\;&\frac{1}{n}\sum_{i=1}^{n}\Big[(\hat{\xi}_{ijk}-\xi_{ijk})(\hat{\xi}_{ijl}-\xi_{ijl}) + \xi_{ijk}(\hat{\xi}_{ijl}-\xi_{ijl}) + (\hat{\xi}_{ijk}-\xi_{ijk})\xi_{ijl}\Big]\\
=\;&O_p\Big(\min\{n^{-1/2}k^{1-a/2}, n^{-1/2}l^{1-a/2}\}\Big).
\end{aligned}
$$

It is obvious that $n^{-1}\sum_{i=1}^{n}\xi_{ijk}\xi_{ijl} - E(\xi_{ijk}\xi_{ijl}) = O_p(n^{-1/2})$. Hence, part (ii) holds.

For part (iii), we note that

$$
E(\sum_{k=K_j+1}^{\infty}\beta^*_{jk}\xi_{ijk}) = 0, \quad E\Big(\sum_{k=K_j+1}^{\infty}\sum_{l=K_m+1}^{\infty}\gamma^*_{jm,kl}\xi_{ijk}\xi_{iml}\Big) = 0,\ j<m.
$$

Then by conditions (A1)–(A3), a simple calculation yields

$$
\begin{aligned}
&E(\sum_{k=K_j+1}^{\infty}\beta^*_{jk}\xi_{ijk})^2 = \sum_{k=K_j+1}^{\infty}\beta^{*2}_{jk}\tau_{jk} \le C\sum_{k=K_j+1}^{\infty}k^{-2b-a} = O(K_j^{-(2b+a-1)}),\\
&E(\sum_{k=K_j+1}^{\infty}\sum_{l=K_m+1}^{\infty}\gamma^*_{jm,kl}\xi_{ijk}\xi_{iml})^2 = \sum_{k=K_j+1}^{\infty}\sum_{l=K_m+1}^{\infty}\alpha^{*2}_{jm,kl}\beta^{*2}_{jk}\beta^{*2}_{ml}\tau_{jk}\tau_{ml}\\
\le\;& C\sum_{k=K_j+1}^{\infty}k^{-2b-a}\sum_{l=K_m+1}^{\infty}l^{-2b-a} = O(K_j^{-(2b+a-1)})O(K_m^{-(2b+a-1)}).
\end{aligned}
$$

Further by condition (A4), part (iii) holds.

For part (iv), by formula (5.22) in Hall and Horowitz (2007) we have $\|\hat{\phi}_{jk}(t) - \phi_{jk}(t)\|^2 = O(n^{-1}k^2)$. This verifies part (iv) and so the proof of Lemma 1 is complete. $\square$

Lemma 2. *Assume that conditions (A1)–(A4) hold and let* $\theta^* = (\beta^{*\top}, \alpha^{*\top})^\top$. *Then,*

$$\|\hat{\theta} - \theta^*\|_2 = O_p(\sqrt{K/n} + a_n).$$

Proof. Let $\rho = \sqrt{K/n} + a_n$, $\theta = \theta^* + \rho\delta$ and $\delta = (u^\top, w^\top)^\top$, where $u = (u_1^\top, \cdots, u_p^\top)^\top$, $u_j = (u_{j1}, \cdots, u_{jK_j})^\top$, $w = (w_{12}^\top, \cdots, w_{(p-1)p}^\top)^\top$ and

$$w_{jm} = (w_{jm,11}, \cdots, w_{jm,1K_m}, \cdots, w_{jm,K_jK_m})^\top.$$

Let $\hat{\varsigma}_{ijm} = (\hat{\xi}_{ij1}\hat{\xi}_{im1}, \cdots, \hat{\xi}_{ij1}\hat{\xi}_{imK_m}, \cdots, \hat{\xi}_{ijK_j}\hat{\xi}_{imK_m})^\top$ and

$$\hat{G}_{ijm}^{\beta_j u_m} = \hat{\varsigma}_{ijm} \odot (\beta_j \otimes 1_{K_m}) \odot (1_{K_j} \otimes u_m),$$

where $A \odot B$ and $C \otimes D$ denote the Hadamard product of A and B and the Kronecker product of C and D, respectively. Let also

$$\hat{G}_i^{\beta u} = ((\hat{G}_{i12}^{\beta_1 u_2})^\top, \cdots, (\hat{G}_{i(p-1)p}^{\beta_{p-1} u_p})^\top)^\top.$$

Thus, we have $\hat{W}_i = \hat{G}_i^{\beta\beta}$.

In what follows, we show that, for any given $\epsilon > 0$, there exists a large constant C_0 such that

$$P\left\{\inf_{\|\delta\|_2 = C_0} L(\theta) > L(\theta^*)\right\} \geq 1 - \epsilon. \tag{B.1}$$

Denote $Q(\theta) = \sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta - (\hat{G}_i^{\beta\beta})^\top \alpha)^2$, then a simple calculation yields

$$
\begin{aligned}
& Q(\theta) - Q(\theta^*) \\
= & \sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta - (\hat{G}_i^{\beta\beta})^\top \alpha)^2 - \sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta^* - (\hat{G}_i^{\beta^*\beta^*})^\top \alpha^*)^2 \\
= & \sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta - (\hat{G}_i^{\beta^*\beta^*})^\top \alpha)^2 - \sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta^* - (\hat{G}_i^{\beta^*\beta^*})^\top \alpha^*)^2 \\
& + \sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta - (\hat{G}_i^{\beta\beta})^\top \alpha)^2 - \sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta - (\hat{G}_i^{\beta^*\beta^*})^\top \alpha)^2 \\
\geq & \sum_{i=1}^{n}(\rho \hat{U}_i^\top u + \rho(\hat{G}_i^{\beta^*\beta^*})^\top w)^2 - 2\rho \sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta^* - (\hat{G}_i^{\beta^*\beta^*})^\top \alpha^*)(\hat{U}_i^\top u + (\hat{G}_i^{\beta^*\beta^*})^\top w) \\
& + \sum_{i=1}^{n}[(\hat{G}_i^{\beta\beta} - \hat{G}_i^{\beta^*\beta^*})^\top \alpha]^2 - 2\sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta - (\hat{G}_i^{\beta^*\beta^*})^\top \alpha)(\hat{G}_i^{\beta\beta} - \hat{G}_i^{\beta^*\beta^*})^\top \alpha.
\end{aligned}
$$

Then let $\Delta_n(\theta) = L(\theta) - L(\theta^*) = L(\theta^* + \rho\delta) - L(\theta^*)$, we have

$$
\begin{aligned}
\Delta_n(\theta) = & \; Q(\theta) - Q(\theta^*) + n\sum_{j=1}^{p}\lambda_{1j}(\|\beta_j\|_2 - \|\beta_j^*\|_2) + n\sum_{j<m}\lambda_{2,jm}(\|\alpha_{jm}\|_2 - \|\alpha_{jm}^*\|_2) \\
\geq & \sum_{i=1}^{n}(\rho \hat{U}_i^\top u + \rho(\hat{G}_i^{\beta^*\beta^*})^\top w)^2 - 2\rho \sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta^* - (\hat{G}_i^{\beta^*\beta^*})^\top \alpha^*)(\hat{U}_i^\top u + (\hat{G}_i^{\beta^*\beta^*})^\top w) \\
& + \sum_{i=1}^{n}[(\hat{G}_i^{\beta\beta} - \hat{G}_i^{\beta^*\beta^*})^\top \alpha]^2 - 2\sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta - (\hat{G}_i^{\beta^*\beta^*})^\top \alpha)(\hat{G}_i^{\beta\beta} - \hat{G}_i^{\beta^*\beta^*})^\top \alpha \\
& + n\sum_{j\in A_1}\lambda_{1j}(\|\beta_j\|_2 - \|\beta_j^*\|_2) + n\sum_{(j,m)\in A_2}\lambda_{2,jm}(\|\alpha_{jm}\|_2 - \|\alpha_{jm}^*\|_2) \\
\geq & \sum_{i=1}^{n}(\rho \hat{U}_i^\top u + \rho(\hat{G}_i^{\beta^*\beta^*})^\top w)^2 - 2\rho \sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta^* - (\hat{G}_i^{\beta^*\beta^*})^\top \alpha^*)(\hat{U}_i^\top u + (\hat{G}_i^{\beta^*\beta^*})^\top w) \\
& + \sum_{i=1}^{n}[(\hat{G}_i^{\beta\beta} - \hat{G}_i^{\beta^*\beta^*})^\top \alpha]^2 - 2\sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta - (\hat{G}_i^{\beta^*\beta^*})^\top \alpha)(\hat{G}_i^{\beta\beta} - \hat{G}_i^{\beta^*\beta^*})^\top \alpha \\
& - n\rho^2\Big\{\sum_{j\in A_1}\|u_j\|_2 + \sum_{(j,m)\in A_2}\|w_{jm}\|_2\Big\} \\
\geq & \sum_{i=1}^{n}(\rho \hat{U}_i^\top u + \rho(\hat{G}_i^{\beta^*\beta^*})^\top w)^2 - 2\rho\delta^\top \sum_{i=1}^{n}\hat{\Omega}_i(Y_i - \hat{\Omega}_i^\top \theta^*) + \sum_{i=1}^{n}[(\hat{G}_i^{\beta\beta} - \hat{G}_i^{\beta^*\beta^*})^\top \alpha]^2 \\
& - 2\sum_{i=1}^{n}(Y_i - \hat{U}_i^\top \beta - (\hat{G}_i^{\beta^*\beta^*})^\top \alpha)(\hat{G}_i^{\beta\beta} - \hat{G}_i^{\beta^*\beta^*})^\top \alpha - n\rho^2\|\delta\|_2 \\
\equiv & \; B_1 - B_2 + B_3 - B_4 - B_5,
\end{aligned}
$$

where $\hat{\Omega}_i = (\hat{U}_i^\top, (\hat{G}_i^{\beta^*\beta^*})^\top)^\top$.

For B_1, by Lemma 1 we have

$$\begin{aligned}
B_1 &= \rho^2 \sum_{i=1}^n \delta^\top \hat{\Omega}_i \hat{\Omega}_i^\top \delta \\
&= \rho^2 \sum_{i=1}^n \delta^\top \Big\{ \Omega_i \Omega_i^\top + (\hat{\Omega}_i - \Omega_i)(\hat{\Omega}_i - \Omega_i)^\top + 2\Omega_i (\hat{\Omega}_i - \Omega_i)^\top \Big\} \delta \\
&= \rho^2 \delta^\top \Big\{ \sum_{i=1}^n \Omega_i \Omega_i^\top \Big\} \delta + \rho^2 O_p(n^{1/2} K^{1-a/2}) \|\delta\|_2^2 \\
&= O_p(n\rho^2)\|\delta\|_2^2 + O_p(\rho^2 n^{1/2} K^{1-a/2})\|\delta\|_2^2. \qquad \text{(B.2)}
\end{aligned}$$

For B_2, note that

$$\begin{aligned}
& \sum_{i=1}^n \delta^\top \hat{\Omega}_i (Y_i - \hat{\Omega}_i^\top \theta^*) \\
= & \sum_{i=1}^n \delta^\top (\hat{\Omega}_i - \Omega_i)(Y_i - \hat{\Omega}_i^\top \theta^*) + \sum_{i=1}^n \delta^\top \Omega_i (Y_i - \hat{\Omega}_i^\top \theta^*) \\
= & \sum_{i=1}^n \delta^\top (\hat{\Omega}_i - \Omega_i)(\varepsilon_i + (\hat{\Omega}_i - \Omega_i)^\top \theta^* + R_i) + \sum_{i=1}^n \delta^\top \Omega_i (\varepsilon_i + (\hat{\Omega}_i - \Omega_i)^\top \theta^* + R_i) \\
= & \sum_{i=1}^n \delta^\top (\hat{\Omega}_i - \Omega_i)\varepsilon_i + \sum_{i=1}^n \delta^\top (\hat{\Omega}_i - \Omega_i)(\hat{\Omega}_i - \Omega_i)^\top \theta^* + \sum_{i=1}^n \delta^\top (\hat{\Omega}_i - \Omega_i) R_i \\
& + \sum_{i=1}^n \delta^\top \Omega_i \varepsilon_i + \sum_{i=1}^n \delta^\top \Omega_i (\hat{\Omega}_i - \Omega_i)^\top \theta^* + \sum_{i=1}^n \delta^\top \Omega_i R_i \\
\equiv & B_{21} + B_{22} + B_{23} + B_{24} + B_{25} + B_{26}.
\end{aligned}$$

According to Lemma 1, it is easy to derive that $B_{21} = O_p(K^{1-a/2})\|\delta\|_2$, $B_{22} = O_p(K^{2-a})\|\delta\|_2$, $B_{23} = O_p(n^{1/2}K^{-(2b+3a/2-2)})\|\delta\|_2$, $B_{24} = O_p(n^{1/2})\|\delta\|_2$, $B_{25} = O_p(n^{1/2}K^{1-a/2})\|\delta\|_2$, and $B_{26} = O_p(nK^{-(2b+a-1)/2})\|\delta\|_2$. Taken together, we have

$$B_2 = \rho O_p(n^{1/2} + n^{1/2}K^{1-a/2} + nK^{-(2b+a-1)/2})\|\delta\|_2. \qquad \text{(B.3)}$$

For B_3, we have

$$\begin{aligned}
B_3 &= \rho^2\alpha^\top \sum_{i=1}^n \left(\hat{G}_i^{\beta^*u} + \hat{G}_i^{u\beta^*} + \rho\hat{G}_i^{uu}\right)\left(\hat{G}_i^{\beta^*u} + \hat{G}_i^{u\beta^*} + \rho\hat{G}_i^{uu}\right)^\top \alpha \\
&= \rho^2\sum_{i=1}^n \alpha^\top \hat{G}_i^{\beta^*u}(\hat{G}_i^{\beta^*u})^\top\alpha + 2\rho^2\sum_{i=1}^n \alpha^\top \hat{G}_i^{\beta^*u}(\hat{G}_i^{u\beta^*})^\top\alpha + 2\rho^3\sum_{i=1}^n \alpha^\top \hat{G}_i^{\beta^*u}(\hat{G}_i^{uu})^\top\alpha \\
&\quad \rho^2\sum_{i=1}^n \alpha^\top \hat{G}_i^{u\beta^*}(\hat{G}_i^{u\beta^*})^\top\alpha + 2\rho^3\sum_{i=1}^n \alpha^\top \hat{G}_i^{u\beta^*}(\hat{G}_i^{uu})^\top\alpha + \rho^4\sum_{i=1}^n \alpha^\top \hat{G}_i^{uu}(\hat{G}_i^{uu})^\top\alpha \\
&\equiv B_{31} + B_{32} + B_{33} + B_{34} + B_{35} + B_{36}.
\end{aligned}$$

By conditions (A2)–(A4) and Lemma 1, a simple calculation yields

$$\begin{aligned}
B_{31} &= \rho^2\sum_{i=1}^n \alpha^\top(\hat{G}_i^{\beta^*u} - G_i^{\beta^*u} + G_i^{\beta^*u})(\hat{G}_i^{\beta^*u} - G_i^{\beta^*u} + G_i^{\beta^*u})^\top\alpha \\
&= \rho^2\sum_{i=1}^n \alpha^\top(\hat{G}_i^{\beta^*u} - G_i^{\beta^*u})(\hat{G}_i^{\beta^*u} - G_i^{\beta^*u})^\top\alpha \\
&\quad +2\rho^2\sum_{i=1}^n \alpha^\top(\hat{G}_i^{\beta^*u} - G_i^{\beta^*u})(G_i^{\beta^*u})^\top\alpha + \rho^2\sum_{i=1}^n \alpha^\top G_i^{\beta^*u}(G_i^{\beta^*u})^\top\alpha \\
&= O_p(\rho^2K^{2-a})\|u\|_2^2 + O_p(n^{1/2}\rho^2K^{1-a/2})\|u\|_2^2 + O_p(n\rho^2)\|u\|^2 = O_p(n\rho^2)\|u\|_2^2.
\end{aligned}$$

Similarly, we can obtain that $B_{32} = O_p(n\rho^2)\|u\|_2^2$, $B_{33} = O_p(n\rho^3)\|u\|_2^2$, $B_{34} = O_p(n\rho^2)\|u\|_2^2$, $B_{35} = O_p(n\rho^3)\|u\|_2^2$, and $B_{36} = O_p(n\rho^4)\|u\|_2^2$. Taken together, we have

$$B_3 = O_p(n\rho^2)\|u\|_2^2. \tag{B.4}$$

For B_4, by a similar argument we have

$$\begin{aligned}
B_4 &= \rho\sum_{i=1}^n \left(\varepsilon_i - (\hat{\Omega}_i - \Omega_i)^\top\theta^* - \rho(\hat{\Omega}_i - \Omega_i)^\top\delta - \rho\Omega_i^\top\delta\right)\left(\hat{G}_i^{\beta^*u} + \hat{G}_i^{u\beta^*} + \rho\hat{G}_i^{uu}\right)^\top\alpha \\
&= O_p(n^{1/2}\rho)\|u\|_2 - O_p(n^{1/2}\rho K^{1-a/2})\|u\|_2 - O_p(n\rho^2)\|\delta\|_2\|u\|_2. \qquad (B.5)
\end{aligned}$$

Combining (B.2)—(B.5), it is easy to see that B_1 dominates the rest terms B_2, B_3, B_4 and B_5 uniformly in $\|\delta\|_2 = C_0$. Therefore, by choosing a sufficiently large C_0, (B.1) holds and there exists a local minimizer $\hat{\theta}$ such that $\|\hat{\theta} - \theta^*\|_2 = O_p(\rho)$. This completes the proof of Lemma 2. □

S3 Appendix C: Simulation studies

In this example, we evaluate the performance of the new procedure when the functional predictors are dependent. We consider

$$X_{i3}(t) = \int \eta_1(s,t)X_{i4}(s)ds + \sigma_0 e_{i1}(t)$$

and

$$X_{i4}(t) = \int \eta_2(s,t)X_{i1}(s)ds + \sigma_0 e_{i2}(t),$$

where $\sigma_0 = 0.5$, $\eta_1(s,t) = 0.6st$, $\eta_2(s,t) = 0.4st$, and $e_{i1}(t)$ and $e_{i2}(t)$ are independent Brownian motions on [0, 1]. All other settings remain the same as those for the independent case. We then repeat the simulations and report the variable selection results in Table 7. Comparing with the results in Table 1, we can see that, even though some of the functional predictors are dependent, the proposed variable selection procedure is still able to identify the true model structure with a higher probability.

References

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

[2] Wong, R. K. W., Li, Y., Zhu, Z. (2019). Partially linear functional additive models for multivariate functional data. *Journal of the American Statistical Association*, 114: 406–418.

Table 7: Variable selection results for dependent case.

		Proposed method			Group SCAD			Group Lasso		
	n	C_M	C_I	C_Z	C_M	C_I	C_Z	C_M	C_I	C_Z
EBIC	100	2.6920	0.8820	5.7140	2.0120	0.7760	4.1020	1.8960	0.7280	3.8420
	200	2.7480	0.9140	5.7940	2.0340	0.7920	5.0960	1.9440	0.7780	4.9620
	300	2.8220	0.9360	5.8280	2.0880	0.8180	5.2640	2.0320	0.8040	5.2280
C-EBIC	100	2.6840	0.8940	5.7420	2.0180	0.7840	4.0980	1.9040	0.7260	3.9060
	200	2.7620	0.9120	5.7960	2.0320	0.8020	5.1040	1.9680	0.7720	5.0140
	300	2.8180	0.9380	5.8320	2.0720	0.8220	5.2820	2.0180	0.8060	5.2460
	n	UF	CF	OF	UF	CF	OF	UF	CF	OF
EBIC	100	0.0920	0.8020	0.1060	0.3600	0.2720	0.3680	0.3820	0.0920	0.5260
	200	0.0620	0.8760	0.0620	0.3340	0.3020	0.3640	0.4180	0.0980	0.4840
	300	0.0320	0.9120	0.0560	0.2940	0.3280	0.3780	0.4520	0.1040	0.4440
C-EBIC	100	0.0840	0.8040	0.1120	0.3840	0.2680	0.3480	0.3780	0.0740	0.5480
	200	0.0520	0.8720	0.0760	0.3300	0.3040	0.3660	0.4320	0.0940	0.4740
	300	0.0280	0.9140	0.0580	0.3660	0.3260	0.3080	0.5160	0.0980	0.3860