Variable Selection for Multiple Function-on-Function Linear Regression

Xiong Cai, Liugen Xue and Jiguo Cao

College of Applied Sciences, Beijing University of Technology

Department of Statistics and Actuarial Science, Simon Fraser University

Supplementary Material

S1. Additional Simulation Study

In this section, we present additional simulation study when the function-on-function linear model has 20 functional predictors, i.e., $p = 20$. We generate data using the similar settings of Section 5 in the main manuscript. Specifically, we first generate $p+2$ independent curves using 50 basis functions. For each $j \in \{0, \ldots, p+1\}$, we take $W_j(s) = \sum_{k=1}^{50} \xi_{jk}\psi_k(s)$, where $\xi_{jk} \sim N(0, 16(2k-1)^{-2})$, $\psi_1(s) \equiv 1$ and $\psi_k(s) = \sqrt{2}\cos\{(k-1)\pi s\}$ for $k \geq 2$. For each $j \in \{1, \ldots, 10\}$, the functional predictors are defined through the linear transformations

$$X_j(s) = W_j(s) + \frac{\tau}{2}\{W_{j-1}(s) + W_{j+1}(s)\},$$

and $X_j(s) = W_{j+1}(s)$ for all $j \in \{11, \ldots, p\}$. Then the first 10 functional predictors are correlated with τ tuning the strength of correlations, and the other $p-10$ functional predictors are independent. The bivariate functional coefficients are set as

$$\beta_j(t,s) = \sum_{k,l=1}^{50} (-1)^{k+l}(k+j)^{-1}l^{-1}\psi_k(t)\psi_l(s)$$

for $j \in \{1,\ldots,6\}$ and $\beta_j(t,s) = 0$ for $j \in \{7,\ldots,p\}$. The error term $\epsilon(t)$ is generated from a Gaussian process with covariance function $\Sigma_\epsilon(t_1,t_2) = \sigma^2\rho^{10|t_1-t_2|}$, where σ^2 is the variance of $\epsilon(t)$ and ρ controls the correlation between $\epsilon(t_1)$ and $\epsilon(t_2)$, for all $t_1, t_2 \in [0,1]$. We consider $\sigma = 0.1$ or 0.5, $\rho = 0, 0.5$ or 0.8 and $\tau = 0$ or 0.6. In each Monte Carlo experiment, we generate 200 independent samples $\{X_{ij}, Y_i, j = 1,\ldots,p, i = 1,\ldots 200\}$ for the training set and another 200 samples $\{X^*_{ij}, Y^*_i, j = 1,\ldots,p, i = 1,\ldots 200\}$ for the test set.

Table 1 summaries the simulation results for cases $\sigma = 0.1, 0.5$, $\rho = 0, 0.5, 0.8$ and $\tau = 0, 0.6$ with sample size of $n = 200$. These results coincide with the results in Table 2 in the main paper. The noise level, the within-function correlation in $\epsilon(t)$ and the correlation level between different functional predictors have a significant effect on the estimation errors as well as the noncausal selection rate. The proposed method tends to be more accurate when the noise level σ, the within-function correlation ρ and dependence level τ are low. Compared to noise level, the within-function correlation seems to have milder effects. The estimation errors become larger when the correlation between different functional predictors increases. Given a noise level σ, the proposed method appears to be the best for the Gaussian white noise and independent functional predictor scenario in which $\rho = 0$ and $\tau = 0$.

Table 1: The positive selection rate (PSR), the noncausal selection rate (NSR), the averages and standard deviations (provided inside brackets) of the integrated squared error (ISE) and the relative prediction error (RPE) based on 100 Monte Carlo replicates for the cases $\sigma = 0.1, 0.5$, $\rho = 0, 0.5, 0.8$ and $\tau = 0, 0.6$ with sample size of $n = 200$.

σ	ρ	τ	PSR	NSR	ISE	RPE
0.1	0	0	1.00	1.00	0.027 (0.004)	0.107 (0.025)
		0.6	1.00	1.00	0.031 (0.005)	0.094 (0.015)
	0.5	0	1.00	0.99	0.028 (0.003)	0.119 (0.025)
		0.6	1.00	1.00	0.035 (0.006)	0.100 (0.017)
	0.8	0	1.00	0.98	0.030 (0.004)	0.132 (0.027)
		0.6	1.00	1.00	0.034 (0.005)	0.107 (0.025)
0.5	0	0	1.00	0.97	0.048 (0.005)	0.324 (0.021)
		0.6	1.00	0.96	0.064 (0.008)	0.236 (0.021)
	0.5	0	1.00	0.99	0.153 (0.021)	0.408 (0.039)
		0.6	1.00	0.94	0.238 (0.044)	0.295 (0.031)
	0.8	0	1.00	0.94	0.182 (0.036)	0.467 (0.064)
		0.6	1.00	0.87	0.272 (0.060)	0.337 (0.047)