Supplementary Material for Function-on-Function Regression Models with Nonlinear Dynamic Effect and Linear Concurrent Effect

Shifan Jia[1], Haolun Shi[2], Tianyu Guan[3]

[1]Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, V5A1S6, BC, Canada.
[2]Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, V5A1S6, BC, Canada.
[3]Department of Mathematics and Statistics, York University, Toronto, M3J1P3, ON, Canada.

1 Proofs of Asymptotic Properties

We will first collect some remarks on notations that are used in the sequel. Boldface symbol is used to denote matrix or vector. If $0 < q < \infty$, L^q is defined as the space of functions $f(t)$ over the interval $[0, 1]$ such that $\int_0^1 |f(t)|^q dt < \infty$. As our estimator is based on B-spline basis, before we dive into proofs of the theorems, we briefly discuss some basic properties of B-spline basis that are used in our proofs. A detailed treatment of B-spline can be found in de Boor (2001). Each B-spline basis function is non-negative and they form a partition of unity, that is, $\sum_{j=1}^{J} \theta_j(t) = 1$ and $\sum_{k=1}^{K} \phi_k(t) = 1$ for all $t \in [0, 1]$, and $\sum_{l=1}^{L} \psi_l(z) = 1$ for all $z \in [0, 1]$. Assume $\beta(t)$ satisfies condition C.2, according to Theorem XII(6) of de Boor (2001), there exists some $\beta_{smooth}(t) = \sum_{j=1}^{J} b_{smooth,j}\theta_j(t)$ with $\boldsymbol{b}_{smooth} = (b_{smooth,1}, \ldots, b_{smooth,J})$, such that $\|\beta(t) - \beta_{smooth}(t)\|_\infty \leq C_1(J - d_1)^{-p_1}$ for some positive constant C_1 and p_1. We use a tensor product of B-spline basis functions to approximate h. Based on the proof of Theorem 2 in Centofanti et al. (2022) and Equation (12.39) on Page 492 of Schumaker (2007), there exist some $h_{smooth}(t, z) = \sum_{k=1}^{K} \sum_{l=1}^{L} \gamma_{smooth,kl}\phi_k(t)\psi_l(z)$ such that $\|h(t, z) - h_{smooth}(t, z)\|_\infty = O((K - d_2)^{-p_2} + (L - d_3)^{-p_2})$ for some positive constant p_2.

Let $\boldsymbol{\eta}_0(t) = \boldsymbol{X}(t)\beta(t) + \int_0^1 h(t, \boldsymbol{A}(s))ds$ and $\boldsymbol{\eta}_{smooth}(t) = \boldsymbol{X}(t)\beta_{smooth}(t) + \int_0^1 h_{smooth}(t, \boldsymbol{A}(s))ds$. Under conditions C.2 and C.3, and given t, this approximation is bounded by

$$
\begin{aligned}
&\|\boldsymbol{\eta}_0(t) - \boldsymbol{\eta}_{smooth}(t)\|_\infty \\
&= \left\| \boldsymbol{X}(t)\beta(t) + \int_0^1 h(t, \boldsymbol{A}(s))ds - \boldsymbol{X}(t)\beta_{smooth}(t) - \int_0^1 h_{smooth}(t, \boldsymbol{A}(s))ds \right\|_\infty \\
&\leq C_2\|\beta(t) - \beta_{smooth}(t)\|_\infty + C_3\|h(t, \boldsymbol{A}(s)) - h_{smooth}(t, \boldsymbol{A}(s))\|_\infty \\
&= O\left((J - d_1)^{-p_1} + (K - d_3)^{-p_2} + (L - d_3)^{-p_2}\right). \qquad (1)
\end{aligned}
$$

1.1 Proof of Theorem 1

We first define the average integral mean square error, conditional on set $\mathcal{A}_n(s) = \{A_1, \ldots, A_n\}$ and $\mathcal{X}_n(t) = \{X_1, \ldots, X_n\}$, as

$$
\text{AIMSE}(\widehat{\boldsymbol{\eta}}(t)|\mathcal{A}_n, \mathcal{X}_n) = \frac{1}{n}\,\text{E}\left\{ \int_{\mathcal{T}} \left(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t)\right)^\top \left(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t)\right) dt | \mathcal{A}_n, \mathcal{X}_n \right\}.
$$

The average integral mean square error can be decomposed into two parts, the variance and bias. We first study the variance.

In Section 4 of the manuscript, we obtain a closed-form solution

$$
\widehat{\boldsymbol{c}} = \left(\int_{\mathcal{T}} \boldsymbol{V}^\top(t)\boldsymbol{V}(t)dt + n\boldsymbol{P}(\kappa, \lambda_t, \lambda_z) \right)^{-1} \int_{\mathcal{T}} \boldsymbol{V}^\top(t)\boldsymbol{Y}(t)dt.
$$

We use numerical integration to approximate the integrals involved in (1). We divide the interval $[0, 1]$ into N subintervals of equal length $\Delta t = 1/N$ using $N+1$ points $t_0, t_1, \ldots, t_N$. Then the integrals can be approximated by

$$
\begin{aligned}
\int_0^1 \boldsymbol{V}^\top(t)\boldsymbol{V}(t)dt &\approx \frac{1}{N}\sum_{q=1}^N \boldsymbol{V}^\top(t_q)\boldsymbol{V}(t_q) = \widetilde{\boldsymbol{V}}^\top\widetilde{\boldsymbol{V}} \quad \text{and} \\
\int_0^1 \boldsymbol{V}^\top(t)\boldsymbol{Y}(t)dt &\approx \frac{1}{N}\sum_{q=1}^N \boldsymbol{V}^\top(t_q)\boldsymbol{Y}(t_q) = \widetilde{\boldsymbol{V}}^\top\widetilde{\boldsymbol{Y}},
\end{aligned}
$$

where $\widetilde{\boldsymbol{V}} = \frac{1}{\sqrt{N}}[\boldsymbol{V}^\top(t_1), \ldots, \boldsymbol{V}^\top(t_N)]^\top$ and $\widetilde{\boldsymbol{Y}} = \frac{1}{\sqrt{N}}[\boldsymbol{Y}^\top(t_1), \ldots, \boldsymbol{Y}^\top(t_N)]^\top$.

Recall $\boldsymbol{P}(\kappa, \lambda_t, \lambda_z) = \begin{pmatrix} \kappa\boldsymbol{R}_\theta & 0 \\ 0 & \lambda_t\boldsymbol{R}_t + \lambda_z\boldsymbol{R}_z \end{pmatrix}$ and let $\lambda\boldsymbol{R}_{tz} = \lambda_t\boldsymbol{R}_t + \lambda_z\boldsymbol{R}_z$. We partition $\left(\int_0^1 \boldsymbol{V}^\top(t)\boldsymbol{V}(t)dt\right)^{-1/2} = \left(\widetilde{\boldsymbol{V}}_1, \widetilde{\boldsymbol{V}}_2\right)$, where $\widetilde{\boldsymbol{V}}_1$ contains the first J columns of $\left(\int_0^1 \boldsymbol{V}^\top(t)\boldsymbol{V}(t)dt\right)^{-1/2}$ and $\widetilde{\boldsymbol{V}}_2$ contains the remaining KL columns. As part of our proof, we will use the spectral decompositions $\widetilde{\boldsymbol{V}}_1\boldsymbol{R}_\theta\widetilde{\boldsymbol{V}}_1^\top = \boldsymbol{U}_1\boldsymbol{D}_1\boldsymbol{U}_1^\top$ and $\widetilde{\boldsymbol{V}}_2\boldsymbol{R}_{tz}\widetilde{\boldsymbol{V}}_2^\top =$

$\boldsymbol{U}_2\boldsymbol{D}_2\boldsymbol{U}_2^\top$. Here $\boldsymbol{D}_1 = \mathrm{diag}(d_1, \ldots, d_J)$ and $\boldsymbol{D}_2 = \mathrm{diag}(d_{J+1}, \ldots, d_{J+KL})$ are diagonal matrices of the two sets of nonincreasing nonnegative eigenvalues and $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are the matrices of the corresponding eigenvectors. Let $r_1 = \mathrm{rank}(\boldsymbol{R}_\theta)$ and $r_2 = \mathrm{rank}(\boldsymbol{R}_{tz})$, we have $d_1 \geq \cdots \geq d_{r_1} > d_{r_1+1} = \cdots = d_J = 0$ and $d_{J+1} \geq \cdots \geq d_{J+r_2} > d_{J+r_2+1} = \cdots = d_{J+KL} = 0$. Therefore

$$\left(\int \boldsymbol{V}^\top \boldsymbol{V}\right)^{-1/2} \boldsymbol{P}(\kappa, \lambda_t, \lambda_z) \left(\int \boldsymbol{V}^\top \boldsymbol{V}\right)^{-1/2} = \kappa \widetilde{\boldsymbol{V}}_1 \boldsymbol{R}_\theta \widetilde{\boldsymbol{V}}_1^\top + \lambda \widetilde{\boldsymbol{V}}_2 \boldsymbol{R}_{tz} \widetilde{\boldsymbol{V}}_2^\top$$
$$= \kappa \boldsymbol{U}_1 \boldsymbol{D}_1 \boldsymbol{U}_1^\top + \lambda \boldsymbol{U}_2 \boldsymbol{D}_2 \boldsymbol{U}_2^\top.$$

Let $\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_1 & 0 \\ 0 & \boldsymbol{U}_2 \end{pmatrix}$ and $\boldsymbol{D} = \begin{pmatrix} \kappa\boldsymbol{D}_1 & 0 \\ 0 & \lambda\boldsymbol{D}_2 \end{pmatrix}$, we have

$$\left(\int_0^1 \boldsymbol{V}^\top(t)\boldsymbol{V}(t)dt\right)^{-1/2} \boldsymbol{P}(\kappa, \lambda_t, \lambda_z) \left(\int_0^1 \boldsymbol{V}^\top(t)\boldsymbol{V}(t)dt\right)^{-1/2} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top.$$

Similarly, we will use the spectral decomposition $\left(\widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{V}}\right)^{-1/2} \boldsymbol{P} \left(\widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{V}}\right)^{-1/2} = \boldsymbol{U}_N \boldsymbol{D}_N \boldsymbol{U}_N^\top$, where $\boldsymbol{D}_N = \mathrm{diag}(d_{N1}, \ldots, d_{N,J+KL})$ is the diagonal matrix of nonincreasing eigenvalues and $\boldsymbol{U}_N$ is the matrix of corresponding eigenvectors. When $N \to \infty$, $d_{Ni} \to d_i$, $i = 1, \ldots, J + KL$.

Let $\widehat{\boldsymbol{\eta}}(t) = \boldsymbol{V}(t)\widehat{\boldsymbol{c}}$ and denote $\widetilde{\widehat{\boldsymbol{\eta}}} = \left(\widehat{\boldsymbol{\eta}}^\top(t_1), \ldots, \widehat{\boldsymbol{\eta}}^\top(t_N)\right)^\top$. For simplicity, denote $\boldsymbol{P} = \boldsymbol{P}(\kappa, \lambda_t, \lambda_z)$. Writing $\boldsymbol{Q}_N = \widetilde{\boldsymbol{V}} \left(\widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{V}}\right)^{-1/2} \boldsymbol{U}_N$, then

$$\widetilde{\widehat{\boldsymbol{\eta}}} \approx \sqrt{N}\widetilde{\boldsymbol{V}} \left(\widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{V}}\right)^{-1/2} \left\{\boldsymbol{I} + n \left(\widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{V}}\right)^{-1/2} \boldsymbol{P} \left(\widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{V}}\right)^{-1/2}\right\}^{-1} \left(\widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{V}}\right)^{-1/2} \widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{Y}}$$
$$= \sqrt{N}\boldsymbol{Q}_N(\boldsymbol{I} + n\boldsymbol{D}_N)^{-1}\boldsymbol{Q}_N^\top \widetilde{\boldsymbol{Y}}.$$

With $\boldsymbol{Q}_N^\top \boldsymbol{Q}_N = \boldsymbol{U}_N^\top \boldsymbol{U}_N = \boldsymbol{I}_{J+KL}$ and conditional on $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ and $\mathcal{A}_n = \{A_1, \ldots, A_n\}$, the covariance matrix of $\widetilde{\widehat{\boldsymbol{\eta}}}$ is $\mathrm{var}\left(\widetilde{\widehat{\boldsymbol{\eta}}}\right) \approx N\sigma^2 \boldsymbol{Q}_N \left(\boldsymbol{I} + n\boldsymbol{D}_N\right)^{-2} \boldsymbol{Q}_N^\top$. This leads to

$$\frac{1}{n}\mathrm{E}\left\{\int \left[\widehat{\boldsymbol{\eta}}(t) - \mathrm{E}\left(\widehat{\boldsymbol{\eta}}(t)\right)\right]^\top \left[\widehat{\boldsymbol{\eta}}(t) - \mathrm{E}\left(\widehat{\boldsymbol{\eta}}(t)\right)\right] dt\right\}$$
$$= \lim_{N\to\infty} \frac{1}{nN} \sum_{j=1}^N \mathrm{E}\left\{\left[\widehat{\boldsymbol{\eta}}(t_j) - \mathrm{E}\left(\widehat{\boldsymbol{\eta}}(t_j)\right)\right]^\top \left[\widehat{\boldsymbol{\eta}}(t_j) - \mathrm{E}\left(\widehat{\boldsymbol{\eta}}(t_j)\right)\right]\right\}$$
$$= \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^N \sum_{i=1}^n \mathrm{var}\left(\widehat{\eta}_i(t_j)\right) = \lim_{N\to\infty} \frac{1}{nN} \mathrm{tr}\left(\mathrm{var}\left(\widetilde{\widehat{\boldsymbol{\eta}}}\right)\right)$$
$$= \lim_{N\to\infty} \frac{1}{nN} \mathrm{tr}\left(N\sigma^2 \boldsymbol{Q}_N \left(\boldsymbol{I} + n\boldsymbol{D}_N\right)^{-2} \boldsymbol{Q}_N^\top\right)$$

$$
\begin{aligned}
&= \frac{\sigma^2}{n} \lim_{N\to\infty} \operatorname{tr}\left(\left(\boldsymbol{I} + n\boldsymbol{D}_N\right)^{-2} \boldsymbol{Q}_N^\top \boldsymbol{Q}_N \right) = \frac{\sigma^2}{n} \operatorname{tr}\left(\left(\boldsymbol{I} + n\boldsymbol{D}\right)^{-2} \right) \\
&= \frac{\sigma^2}{n} \left(\sum_{l=1}^{J} \frac{1}{(1+n\kappa d_l)^2} + \sum_{l=J+1}^{J+KL} \frac{1}{(1+n\lambda d_l)^2} \right),
\end{aligned} \tag{1}
$$

where $\widehat{\eta}_i(t_j) = \boldsymbol{V}_i(t_j)\widehat{\boldsymbol{c}}$.

Next, we study the bias. We start by denoting the non-penalized least squares estimate $\widehat{\boldsymbol{\eta}}_u(t) = \boldsymbol{V}(t)\left(\int_{\mathcal{T}} \boldsymbol{V}^\top(t)\boldsymbol{V}(t)dt\right)^{-1} \int_{\mathcal{T}} \boldsymbol{V}^\top(t)\boldsymbol{Y}(t)dt$. We observe that $\mathrm{E}(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t)) = \mathrm{E}(\widehat{\boldsymbol{\eta}}(t) - \widehat{\boldsymbol{\eta}}_u(t)) + \mathrm{E}(\widehat{\boldsymbol{\eta}}_u(t) - \boldsymbol{\eta}_0(t))$. Denote $\widetilde{\widehat{\boldsymbol{\eta}}_u} = [\widehat{\boldsymbol{\eta}}_u(t_1)^\top, \ldots, \widehat{\boldsymbol{\eta}}_u(t_N)^\top]^\top$ and $\boldsymbol{\eta}_0 = [\boldsymbol{\eta}_0(t_1)^\top, \ldots, \boldsymbol{\eta}_0(t_N)^\top]^\top$, we have

$$
\begin{aligned}
\mathrm{E}\left(\widetilde{\widehat{\boldsymbol{\eta}}} - \widetilde{\widehat{\boldsymbol{\eta}}}_u \right) &= \lim_{N\to\infty} \mathrm{E}\left(\sqrt{N} \widetilde{\boldsymbol{V}} \left[(\widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{V}} + n\boldsymbol{P})^{-1} - (\widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{V}})^{-1} \right] \widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{Y}} \right) \\
&= -n \lim_{N\to\infty} \sqrt{N} \boldsymbol{Q}_N \left(\boldsymbol{I} + n\boldsymbol{D}_N\right)^{-1} \boldsymbol{D}_N \boldsymbol{Q}_N^\top \boldsymbol{\eta}_0
\end{aligned}
$$

Since $\boldsymbol{Q}_N^\top \boldsymbol{\eta}_0 = O_p(1)$, we can obtain

$$
\begin{aligned}
&\frac{1}{n} \int_0^1 \mathrm{E}\left[\widehat{\boldsymbol{\eta}}(t) - \widehat{\boldsymbol{\eta}}_u(t)\right]^\top \mathrm{E}\left[\widehat{\boldsymbol{\eta}}(t) - \widehat{\boldsymbol{\eta}}_u(t)\right] \\
&= \lim_{N\to\infty} \frac{1}{nN} \left\| \mathrm{E}\left(\widetilde{\widehat{\boldsymbol{\eta}}} - \widetilde{\widehat{\boldsymbol{\eta}}}_u \right) \right\|^2 \\
&= n \lim_{N\to\infty} \operatorname{tr}\left[\boldsymbol{D}_N \left(\boldsymbol{I} + n\boldsymbol{D}_N\right)^{-1} \boldsymbol{Q}_N^\top \boldsymbol{Q}_N \left(\boldsymbol{I} + n\boldsymbol{D}_N\right)^{-1} \boldsymbol{D}_N \right] \\
&= O_p\left\{ \frac{\kappa^2}{n} \sum_{l=1}^{J} \frac{(nd_l)^2}{(1+n\kappa d_l)^2} + \frac{\lambda^2}{n} \sum_{l=J+1}^{J+KL} \frac{(nd_l)^2}{(1+n\lambda d_l)^2} \right\}.
\end{aligned} \tag{2}
$$

Recall that we write $\boldsymbol{\eta}(t) = \boldsymbol{V}(t)\boldsymbol{c}$. Denote $\widetilde{\boldsymbol{\eta}} = [\boldsymbol{\eta}(t_1)^\top, \ldots, \boldsymbol{\eta}(t_N)^\top]^\top$, the approximating error for $\widetilde{\widehat{\boldsymbol{\eta}}}_u$ is

$$
\begin{aligned}
\mathrm{E}\left(\widetilde{\widehat{\boldsymbol{\eta}}}_u \right) - \widetilde{\boldsymbol{\eta}}_0 &= \lim_{N\to\infty} \widetilde{\boldsymbol{V}} \left(\widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{V}} \right)^{-1} \widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{\eta}}_0 - \widetilde{\boldsymbol{\eta}}_0 \\
&= \lim_{N\to\infty} \widetilde{\boldsymbol{V}} \left(\widetilde{\boldsymbol{V}}^\top \widetilde{\boldsymbol{V}} \right)^{-1} \widetilde{\boldsymbol{V}}^\top \left(\widetilde{\boldsymbol{V}}\boldsymbol{c} + \widetilde{\boldsymbol{\eta}}_0 - \widetilde{\boldsymbol{V}}\boldsymbol{c} \right) - \widetilde{\boldsymbol{\eta}}_0 \\
&= \lim_{N\to\infty} \left(\boldsymbol{I} - \boldsymbol{Q}_N \boldsymbol{Q}_N^\top \right) \left(\widetilde{\boldsymbol{\eta}} - \widetilde{\boldsymbol{\eta}_0} \right).
\end{aligned}
$$

This yields

$$
\begin{aligned}
&\frac{1}{n} \int_0^1 \mathrm{E}\left[\widehat{\boldsymbol{\eta}}_u(t) - \boldsymbol{\eta}_0(t)\right]^\top \mathrm{E}\left[\widehat{\boldsymbol{\eta}}_u(t) - \boldsymbol{\eta}_0(t)\right] = \lim_{N\to\infty} \frac{1}{nN} \left\| \mathrm{E}(\widetilde{\widehat{\boldsymbol{\eta}}}_u) - \widetilde{\boldsymbol{\eta}}_0 \right\|^2 \\
&= \lim_{N\to\infty} \frac{1}{nN} \operatorname{tr}\left(\left(\widetilde{\boldsymbol{\eta}} - \widetilde{\boldsymbol{\eta}_0} \right)^\top \left(\boldsymbol{I} - \boldsymbol{Q}_N \boldsymbol{Q}_N^\top \right)^2 \left(\widetilde{\boldsymbol{\eta}} - \widetilde{\boldsymbol{\eta}_0} \right) \right) \\
&\le \lim_{N\to\infty} \frac{1}{nN} \operatorname{tr}\left(\boldsymbol{I} - \boldsymbol{Q}_N \boldsymbol{Q}_N^\top \right) \left\| \widetilde{\boldsymbol{\eta}} - \widetilde{\boldsymbol{\eta}_0} \right\|^2
\end{aligned}
$$

$$
\begin{aligned}
&= \lim_{N\to\infty} O_p\left\{\frac{1}{nN}(nN-(J+KL))(J-d_1)^{-p_1}+(K-d_2)^{-p_2}+(L-d_3)^{-p_2}\right\} \\
&= O_p\left((J-d_1)^{-p_1}+(K-d_2)^{-p_2}+(L-d_3)^{-p_2}\right) \qquad (3)
\end{aligned}
$$

Combining (1), (2), and (3), we can prove Theorem 1.

1.2 Proof of Theorem 2

For future predictor trajectories x_0 and a_0, denote

$$
\begin{aligned}
\widehat{\boldsymbol{\eta}}_{x_0,a_0}(t) &= \widehat{\mathrm{E}}\left(Y(t)|\mathcal{X}_n,\mathcal{A}_n,X=x_0,A=a_0\right) = \boldsymbol{v}_0(t)\widehat{\boldsymbol{c}} \\
&= \boldsymbol{v}_0^\top(t)\left(\int_0^1 \boldsymbol{V}^\top(t)\boldsymbol{V}(t)dt/n+\boldsymbol{P}(\kappa,\lambda_t,\lambda_z)\right)^{-1}\left(\int_0^1 \boldsymbol{V}^\top(t)\boldsymbol{Y}(t)dt/n\right), \qquad (4)
\end{aligned}
$$

where $\boldsymbol{v}_0(t)=(x_0(t)\boldsymbol{\theta}^\top(t),\boldsymbol{z}_0^\top(t))$ and $\boldsymbol{z}_0(t)=\mathrm{vec}\left[\phi_k(t)\int_0^1\psi_l(a_0(s))ds\right]$.

Let $\boldsymbol{G}=\int_0^1\boldsymbol{V}^\top(t)\boldsymbol{V}(t)dt/n$ and let $\rho_1=\rho_1(n)$ as the smallest eigenvalue of $\boldsymbol{G}$, then the maximum eigenvalue of $\boldsymbol{G}^{-1}\boldsymbol{P}(\kappa,\lambda_t,\lambda_z)=\boldsymbol{G}^{-1}(\kappa\boldsymbol{P}_1+\lambda\boldsymbol{P}_2)$ is bounded by $(C_4\kappa+C_5\lambda)/\rho_1=o_p(1)$ for some positive constants C_4 and C_5, where $\boldsymbol{P}_1=\begin{pmatrix}\boldsymbol{R}_\theta & 0\\ 0 & 0\end{pmatrix}$ and $\boldsymbol{P}_2=\begin{pmatrix}0 & 0\\ 0 & \boldsymbol{R}_{tz}\end{pmatrix}$. By replacing $\boldsymbol{P}(\kappa,\lambda_t,\lambda_z)$ with $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ in (4), we have

$$
\widehat{\boldsymbol{\eta}}_{x_0,a_0}(t)=\boldsymbol{v}_0^\top(t)\left(\int_0^1\boldsymbol{V}^\top(t)\boldsymbol{V}(t)dt/n+\kappa\boldsymbol{P}_1+\lambda\boldsymbol{P}_2\right)^{-1}\left(\int_0^1\boldsymbol{V}^\top(t)\boldsymbol{Y}(t)dt/n\right),
$$

By applying Taylor expansion at $\kappa=0$ and $\lambda=0$, we have that for some $\xi_\lambda\in[0,\lambda]$ and $\xi_\kappa\in[0,\kappa]$,

$$
\widehat{\boldsymbol{\eta}}_{x_0,a_0}(t)=\boldsymbol{v}_0^\top(t)\left(I-\kappa\boldsymbol{G}^{-1}\boldsymbol{P}_1-\lambda\boldsymbol{G}^{-1}\boldsymbol{P}_2\right)\boldsymbol{G}^{-1}\left(\int_0^1\boldsymbol{V}^\top(t)\boldsymbol{Y}(t)dt/n\right)+R_n, \qquad (5)
$$

where R_n is the remainder. Let $\boldsymbol{G}_\xi=\int_0^1\boldsymbol{V}^\top(t)\boldsymbol{V}(t)\,dt/n+\xi_\kappa\,\boldsymbol{P}_1+\xi_\lambda\,\boldsymbol{P}_2$, we can express the remainder as $R_n=\frac{1}{2}\boldsymbol{v}_0^\top(t)\,\boldsymbol{G}_\xi^{-1}\left(R_\kappa+R_\lambda+2\,R_{\kappa,\lambda}\right)\boldsymbol{G}_\xi^{-1}\left(\int_0^1\boldsymbol{V}^\top(t)\boldsymbol{Y}(t)dt/n\right)$, where $R_\kappa=\kappa^2\boldsymbol{P}_1\boldsymbol{G}_\xi^{-1}\boldsymbol{P}_1$, $R_\lambda=\lambda^2\boldsymbol{P}_2\boldsymbol{G}_\xi^{-1}\boldsymbol{P}_2$, and $R_{\kappa,\lambda}=2\kappa\lambda\left(\boldsymbol{P}_1\boldsymbol{G}_\xi^{-1}\boldsymbol{P}_2+\boldsymbol{P}_2\boldsymbol{G}_\xi^{-1}\boldsymbol{P}_1\right)$.

Recall $\boldsymbol{\eta}(t)=\boldsymbol{V}(t)\boldsymbol{c}$ and since for a given t, $\|\boldsymbol{\eta}_0(t)-\boldsymbol{\eta}(t)\|_\infty=O_p(1/(J-d_1)^{p_1}+1/(K-d_2)^{p_2}+1/(L-d_3)^{p_2})$, we have $\mathrm{E}\left(\boldsymbol{Y}(t)\right)=\boldsymbol{\eta}_0(t)=\boldsymbol{\eta}(t)(1+O_p(r_\eta))=\boldsymbol{\eta}(t)(1+o_p(1))=\boldsymbol{V}(t)\boldsymbol{c}(1+o_p(1))$ with $r_\eta=1/(J-d_1)^{p_1}+1/(K-d_2)^{p_2}+1/(L-d_3)^{p_2}$. Therefore,

$$
\mathrm{E}\left(\boldsymbol{v}_0^\top(t)\left(\kappa\boldsymbol{G}^{-1}\boldsymbol{P}_1+\lambda\boldsymbol{G}^{-1}\boldsymbol{P}_2\right)\boldsymbol{G}^{-1}\left(\int_0^1\boldsymbol{V}^\top(t)\boldsymbol{Y}(t)dt/n\right)\right)
$$

$$
\begin{aligned}
&= \boldsymbol{v}_0^\top(t)\boldsymbol{G}^{-1}\left(\kappa\boldsymbol{P}_1 + \lambda\boldsymbol{P}_2\right)\boldsymbol{G}^{-1}\left(\int_0^1 \boldsymbol{V}^\top(t)\boldsymbol{V}(t)dt/n\right)\boldsymbol{c}(1+o_p(1))\\
&= \boldsymbol{v}_0^\top(t)\boldsymbol{G}^{-1}\left(\kappa\boldsymbol{P}_1 + \lambda\boldsymbol{P}_2\right)\boldsymbol{c}(1+o_p(1)).
\end{aligned}
$$

Similarly, we can obtain

$$
\begin{aligned}
&\mathrm{E}\left(R_n\right)\\
&= \frac{1}{2}\boldsymbol{v}_0^\top(t)\boldsymbol{G}_\xi^{-1}\left(R_\kappa + R_\lambda + 2R_{\kappa,\lambda}\right)\boldsymbol{G}_\xi^{-1}\left(\int_0^1 \boldsymbol{V}^\top(t)\boldsymbol{V}(t)dt/n\right)\boldsymbol{c}(1+o_p(1))\\
&= \boldsymbol{v}_0^\top(t)\boldsymbol{G}^{-1}(\kappa\boldsymbol{P}_1 + \lambda\boldsymbol{P}_2)\boldsymbol{c}o_p(1)
\end{aligned}
$$

Let $\eta_{0,x_0,a_0}(t) = x_0(t)\beta(t) + \int_0^1 h(t, a_0(s))ds$, the asymptotic bias is

$$
\begin{aligned}
&\mathrm{E}\left(\widehat{\eta}_{x_0,a_0}(t)\right) - \eta_{0,x_0,a_0}(t)\\
&= \mathrm{E}\left(\widehat{\eta}_{u,x_0,a_0}(t)\right) - \eta_{0,x_0,a_0}(t) - \boldsymbol{v}_0^\top(t)\boldsymbol{G}^{-1}\left(\kappa\boldsymbol{P}_1 + \lambda\boldsymbol{P}_2\right)\boldsymbol{c}(1+o_p(1))\\
&= O_p(1/(J-d_1)^{p_1} + 1/(K-d_2)^{p_2} + 1/(L-d_3)^{p_2}))\\
&- \boldsymbol{v}_0^\top(t)\boldsymbol{G}^{-1}\left(\kappa\boldsymbol{P}_1 + \lambda\boldsymbol{P}_2\right)\boldsymbol{c}(1+o_p(1)).
\end{aligned}
$$

The asymptotic bias consists of two components, the approximation bias $O_p(1/(J-d_1)^{p_1} + 1/(K-d_2)^{p_2} + 1/(L-d_3)^{p_2}))$ and the shrinkage bias $-\boldsymbol{v}_0^\top(t)\boldsymbol{G}^{-1}\left(\kappa\boldsymbol{P}_1 + \lambda\boldsymbol{P}_2\right)\boldsymbol{c}(1+o_p(1))$.

Based on Taylor expansion (5), the asymptotic variance can be obtained as

$$
\begin{aligned}
\mathrm{Var}\left(\widehat{\eta}_{x_0,a_0}(t)\right) = \frac{\sigma^2}{n}\Big[&\boldsymbol{v}_0^\top(t)\boldsymbol{G}^{-1}\boldsymbol{v}_0(t) - 2\kappa\boldsymbol{v}_0^\top(t)\boldsymbol{G}^{-1}\boldsymbol{P}_1\boldsymbol{G}^{-1}\boldsymbol{v}_0(t)(1+o_p(1))\\
&- 2\lambda\boldsymbol{v}_0^\top(t)\boldsymbol{G}^{-1}\boldsymbol{P}_2\boldsymbol{G}^{-1}\boldsymbol{v}_0(t)(1+o_p(1))\Big].
\end{aligned}
\tag{6}
$$

The maximum eigenvalue of $\boldsymbol{G}^{-1}(\kappa\boldsymbol{P}_1 + \lambda\boldsymbol{P}_2)$ is bounded by $o_p(1)$, meaning that the second and third terms (the higher-order terms) in (6) reflects a reduction in the variance. Now rewrite the first term (the leading term variance) as $\mathrm{var}_0(t) = \sigma^2/n\boldsymbol{v}_0^\top(t)\boldsymbol{G}^{-1}\boldsymbol{v}_0(t)$. For arbitrary $x_0(t)$ and $a_0(s)$, we can obtain

$$
\begin{aligned}
\boldsymbol{v}_0^\top(t)\boldsymbol{v}_0(t) &= x_0^2(t)\sum_{j=1}^J \theta_j^2(t) + \sum_{k=1}^K\sum_{l=1}^L\left(\phi_k(t)\int_0^1 \psi_l(a_0(s))ds\right)^2\\
&\leq x_0^2(t)\sum_{j=1}^J \theta_j(t) + \sum_{k=1}^K \phi_k^2(t)\sum_{l=1}^L\left(\int_0^1 \psi_l(a_0(s))ds\right)^2 \leq x_0^2(t) + 1.
\end{aligned}
$$

Since $\mathrm{E}\left\|X\right\|_2^2 < \infty$ by Condition C.1, the maximum eigenvalue of $\boldsymbol{G}$ denoted by $\rho_{\max}(\boldsymbol{G}) = \mathrm{tr}(\boldsymbol{G}) \leq 1 + \sum_{i=1}^n \left\|X_i\right\|_2^2/n$. This implies that $O_p\left(\sigma^2 n^{-1}\right) \leq \mathrm{var}_0(t) \leq O_p\left(\sigma^2(n\rho_1)^{-1}\right)$.

Next, we discuss the requirements for the asymptotic normality to hold. As $n \to \infty$, conditional on the designs $\mathcal{A}_n$ and $\mathcal{X}_n$, the approximation bias must become asymptotically negligible. This requires that $\min\left\{(J-d_1)^{2p_1}, (K-d_2)^{2p_2}, (L-d_3)^{2p_2}\right\}\mathrm{var}_0(t) \to \infty$, which is sufficient to assume that $\min\left\{(J-d_1)^{2p_1}, (K-d_2)^{2p_2}, (L-d_3)^{2p_2}\right\}/n \to \infty$. To ensure that the shrinkage bias does not dominate $\mathrm{var}_0(t)$, it must satisfy $-\boldsymbol{v}_0^\top(t)\boldsymbol{G}^{-1}(\kappa\boldsymbol{P}_1+\lambda\boldsymbol{P}_2)\boldsymbol{c}(1+o_p(1))/\mathrm{var}_0^2(t) = O_p(1)$. These conditions can be met if $n\max\left\{\kappa^2,\lambda^2\right\}/\rho_1^2 = O_p(1)$. Since $\mathrm{var}_0(t)$ and $\mathrm{Var}(\widehat{\eta}_{x_0,a_0}(t))$ are asymptotically equivalent, to check the Lindeberg–Feller condition for the central limit theorem, it is sufficient to show that $\dfrac{\widehat{\eta}_{x_0,a_0}(t)-E(\widehat{\eta}_{x_0,a_0}(t))}{\sqrt{\mathrm{Var}(\widehat{\eta}_{x_0,a_0}(t))}}$ converges to a standard normal distribution. Let η_{0i} be the ith elements of $\boldsymbol{\eta}_0(t)$, we have

$$
\begin{aligned}
&\widehat{\eta}_{x_0,a_0}(t) - E(\widehat{\eta}_{x_0,a_0}(t))\\
&= \boldsymbol{v}_0^\top(t)\left(\int_0^1 \boldsymbol{V}^\top(t)\boldsymbol{V}(t)dt/n + \kappa\boldsymbol{P}_1 + \lambda\boldsymbol{P}_2\right)^{-1}\int_0^1 \boldsymbol{V}^\top(t)\left(\boldsymbol{Y}(t)-\boldsymbol{\eta}_0(t)\right)dt/n\\
&= \sum_{i=1}^n \int_0^1 a_i(t)\epsilon_i(t)dt,
\end{aligned}
$$

where $a_i(t) = \boldsymbol{v}_0^\top(t)\left(\int_0^1 \boldsymbol{V}^\top(t)\boldsymbol{V}(t)\,dt/n + \kappa\,\boldsymbol{P}_1 + \lambda\,\boldsymbol{P}_2\right)^{-1}\boldsymbol{V}_i^\top(t)/n$ and $\epsilon_i(t) = Y_i(t) - \eta_{0i}$. It is suffice to show that $\max_i\int_0^1 a_i^2(t)dt = o_p(\sum_i^n\int_0^1 a_i^2 dt) = o_p(\mathrm{Var}(\widehat{\eta}_{x_0,a_0}(t)))$. Recall that the maximum eigenvalue of $\boldsymbol{v}_0(t)\boldsymbol{v}_0^\top(t)$ is no greater than $\boldsymbol{v}_0(t)\boldsymbol{v}_0^\top(t) \le x_0^2(t)+1$ for any x_0 and a_0. This implies

$$
\begin{aligned}
&\int_0^1 a_i^2(t)dt\\
&= \int_0^1 \boldsymbol{v}_0^\top(t)\left(\boldsymbol{G}+\kappa\boldsymbol{P}_1+\lambda\boldsymbol{P}_2\right)^{-1}\boldsymbol{V}_i^\top(t)/n\boldsymbol{V}_i(t)/n\left(\boldsymbol{G}+\kappa\boldsymbol{P}_1+\lambda\boldsymbol{P}_2\right)^{-1}\boldsymbol{v}_0 dt\\
&\le n^{-2}\mathrm{tr}\left(\left(\boldsymbol{G}+\kappa\boldsymbol{P}_1+\lambda\boldsymbol{P}_2\right)^{-2}\int_0^1 \boldsymbol{V}_i(t)\boldsymbol{V}_i^\top(t)dt\right)\\
&\le n^{-2}\mathrm{tr}\left(\boldsymbol{G}^{-2}\left(I+\kappa\boldsymbol{P}_1\boldsymbol{G}^{-1}+\lambda\boldsymbol{P}_2\boldsymbol{G}^{-1}\right)^{-2}\int_0^1 \boldsymbol{V}_i(t)\boldsymbol{V}_i^\top(t)dt\right).
\end{aligned}
$$

Since the smallest eigenvalue of $\boldsymbol{G}$ is ρ_1, the maximum eigenvalue of $\boldsymbol{G}^{-1}$ is ρ_1^{-1}. Using the fact that for any two nonnegative-definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\mathrm{tr}(\boldsymbol{AB}) \le \rho(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B})$ with $\rho(\boldsymbol{A})$ being the the mamixum eigenvalue of $\boldsymbol{A}$, we have

$$
\int_0^1 a_i^2(t)dt \le n^{-2}\rho_1^{-2}\mathrm{tr}\left(\left(I+\kappa\boldsymbol{P}_1\boldsymbol{G}^{-1}+\lambda\boldsymbol{P}_2\boldsymbol{G}^{-1}\right)^{-2}\int_0^1 \boldsymbol{V}_i(t)\boldsymbol{V}_i^\top(t)dt\right).
$$

Using a similar Taylor expansion as shown earlier and noting that $\max\{\kappa, \lambda\}/\rho_1 = o_p(1)$, we observe that the bound for $\int_0^1 a_i^2(t)dt$ is bounded in probability by $n^{-2}\rho_1^{-2}\mathrm{tr}\left(\int_0^1 \boldsymbol{V}_i(t)\boldsymbol{V}_i^\top(t)dt\right) \leq C_6 n^{-2}\rho_1^{-2}$ for some positive constant C_6.

Consequently, the conditions $n\max\{\kappa^2, \lambda^2\}/\rho_1^2 = O_p(1)$ and $\min\{\kappa, \lambda\} \to \infty$ imply that $\max_i \int_0^1 a_i^2(t)dt/\mathrm{Var}(\widehat{\eta}_{x_0,a_0}(t)) \leq (n\rho_1)^{-2}/(n^{-1}\sigma^2) \to 0$, thereby completing the proof.

References

Centofanti, F., Fontana, M., Lepore, A., Vantini, S.: Smooth lasso estimator for the function-on-function linear regression model. Comput. Statist. Data Anal. **176**, 107556 (2022) https://doi.org/10.1016/j.csda.2022.107556

Boor, C.: A practical guide to splines. springer-verlag. New York (2001)

Schumaker, L.: Spline Functions: Basic Theory, 3rd edn. Cambridge Mathematical Library. Cambridge University Press, Cambridge (2007)