Web Appendix for "Functional Joint Models for Imaging Genetic Data"

Qingzhi Zhong[1], Xinyuan Song[2*], Hongtu Zhu[3,4,5,6*]

[1]*School of Economics, Jinan University, Guangzhou, China.*

[2]*Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, P.R. China.*

Departments of Biostatistics[3], Genetics[4], Computer Science[5], and Statistics[6] University of North Carolina at Chapel Hill, North Carolina, USA

This supplementary document shows the detailed proofs of the theoretical results, the estimation procedure and some additional results in the main paper. In Web Appendix A, we provide detailed estimation equations and implementation algorithms. In Web Appendix B, we prove the theoretical properties, including the nonlinear truncation error of functional data and model identifiability, some Lemmas essential in

*These authors are jointly supervised this work: Xinyuan Song and Hongtu Zhu, Emails for Correspondence: xysong@cuhk.edu.hk and htzhu@email.unc.edu. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`. This work was partially supported by GRF grant 14302519 from the Research Grant Council of the HKSAR for Dr. Song. Zhong's work was supported by National Natural Science Foundation of China (12401349) and China Postdoctoral Science Foundations (2024M751116).

the proof of main theorems, the convergent rates of the estimators of nonlinear FPC scores and spatial varying coefficients, the selection consistency and convergence of the proposed iterative algorithms. For theoretical investigation, we use the standardized Bernstein basis polynomials $S_j(\mathbf{s})/\|S_j(\mathbf{s})\|_2, j = 1, \ldots, J_n$. For the convenience of expression, we continue to use the notations $S_j(\mathbf{s})$ and $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ as the standardized bivariate spline basis functions and the scaled spline coefficients. We use M and C_1 as generic positive constants, which may be different even in the same line. Finally, in Web Appendix C, we provide additive analysis for the ADNI data set, Web Tables 1–11, and Web Figures 1–7.

Web Appendix A: Estimation

Estimation in Nonlinear MFPCA

Considering that $\boldsymbol{\xi}_i$ are unobserved, most existing methods perform spectral decomposition on the covariance function of $\mathbf{X}_i(t)$ and then estimate the latent scores $\boldsymbol{\xi}_i$ under dense observations or Gaussian distribution assumptions (Chiou et al., 2014; Jacques and Preda, 2014; Happ and Greven, 2018; Wong et al., 2019). A key for these methods is the linear structure $f_j\left\{\boldsymbol{\xi}_i^T \boldsymbol{\phi}_j(t)\right\} = M\boldsymbol{\xi}_i^T \boldsymbol{\phi}_j(t)$, where M is a finite constant. When functional variables are nonlinearly related, there exist no closed-form estimators for $\boldsymbol{\xi}_i$ and $\boldsymbol{\phi}_j(\cdot)$, and the FPCA method cannot be directly applied. In this paper, we treat the relation between $X_{ij}(t) - \mu_j(t)$ and $\boldsymbol{\xi}_i^T \boldsymbol{\phi}_j(t)$ as unknown nonparametric functions and the latent scores as parameters. Since $\mu_j(\cdot), \boldsymbol{\phi}_j(\cdot)$, and $f_j(\cdot)$ are unknown, directly estimating them is difficult. We employ some smoothing techniques to address the problem. Specifically, we use B-spline basis functions to approximate $\mu_j(\cdot)$ and $\boldsymbol{\phi}_j(\cdot)$ and the local linear smoother to approximate the link function $f_j(\cdot)$. The definition of B-spline functions can be found in Schumacker (pp. 118, 1981). We use the orthogonal spline functions satisfying $\int_{\mathcal{T}} \mathbf{B}_n(t)\mathbf{B}_n(t)^T dt = \mathbf{I}$ (Zhou et al., 2008; Zhong et al., 2021), which facilitate the implementation of model identification.

Estimation of $\mu_j(\cdot)$

With the spline approximation $\mu_j(t) \approx \mu_{nj}(t) = \mathbf{u}_j^T \mathbf{B}_n(t)$, we can estimate $\mathbf{u}_j$ by minimizing

$$\sum_{i=1}^{n} \sum_{d=1}^{n_{ij}} \left\{ X_{ij}(t_{ijd}) - \mathbf{u}_j^T \mathbf{B}_n(t_{ijd}) \right\}^2, \tag{S0.1}$$

which yields the analytic solutions

$$\widehat{\mathbf{u}}_j = \left\{ \sum_{i=1}^{n} \sum_{d=1}^{n_{ij}} \mathbf{B}_n(t_{ijd}) \mathbf{B}_n(t_{ijd})^T \right\}^{-1} \sum_{i=1}^{n} \sum_{d=1}^{n_{ij}} X_{ij}(t_{ijd}) \mathbf{B}_n(t_{ijd}). \tag{S0.2}$$

Then, we estimate $\mu_j(t)$ by $\widehat{\mu}_j(t) = \widehat{\mathbf{u}}_j^T \mathbf{B}_n(t)$.

Estimation of $f_j(\cdot), \boldsymbol{\phi}_j(\cdot)$, and $\boldsymbol{\xi}_i$

Let $\mathbf{f} = (f_1, \ldots, f_p)^T, \boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p)^T, \boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n)^T$, and $l_1(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi}) = \frac{1}{\sum_{i=1}^n \sum_{j=1}^p n_{ij}} \sum_{i=1}^n \sum_{j=1}^p \sum_{d=1}^{n_{ij}} \left[X_{ij}(t_{ijd}) - \widehat{\mu}_j(t_{ijd}) - f_j \left\{ \boldsymbol{\xi}_i^T \boldsymbol{\phi}_j(t_{ijd}) \right\} \right]^2$. Denote $\mathcal{K}(\cdot)$ as a kernel function, h_1 as bandwidth, and $\mathcal{K}_{h_1}(\cdot) = h_1^{-1} \mathcal{K}(\cdot / h_1)$. Then, we plug the estimated $\widehat{\mu}_j$ into (3.6) in the main paper and solve the resulting minimization by decomposing it into the following three linear regression problems.

Given $\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd})$, applying the local linear approximation to $f_j(\cdot)$, we have $f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd})\right\} \approx f_j(u) + \nabla f_j(u)\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd}) - u\right\}$ for $\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd})$ close to u. Plugging this approximation into (3.6) in the main paper, we obtain a weighted least square function for $(f_j(\cdot), \nabla f_j(\cdot))$,

$$\sum_{i=1}^{n}\sum_{d=1}^{n_{ij}}\left[X_{ij}(t_{ijd}) - \widehat{\mu}_j(t_{ijd}) - f_j(u) - \nabla f_j(u)\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd}) - u\right\}\right]^2 \times \mathcal{K}_{h_1}\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd}) - u\right\}, \tag{S0.3}$$

which leads to the following closed-form estimator,

$$\left\{\begin{array}{c}\widehat{f}_j(u)\\ \widehat{\nabla f}_j(u)\end{array}\right\} = \left[\sum_{i=1}^{n}\sum_{d=1}^{n_{ij}}\mathcal{K}_{h_1}\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd}) - u\right\}\mathbf{Z}_{ijd}^f(u)\mathbf{Z}_{ijd}^f(u)^T\right]^{-1} \times \left[\sum_{i=1}^{n}\sum_{d=1}^{n_{ij}}\mathcal{K}_{h_1}\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd}) - u\right\}\mathbf{Z}_{ijd}^f(u)\left\{X_{ij}(t_{ijd}) - \widehat{\mu}_j(t_{ijd})\right\}\right], \tag{S0.4}$$

where $\mathbf{Z}_{ijd}^f(u) = \left(1, \boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd}) - u\right)^T$.

Given $\{f_j(\cdot), \nabla f_j(\cdot), \boldsymbol{\xi}_i\}$ and $\boldsymbol{\phi}_j^{(o)}(\cdot) = \boldsymbol{\Gamma}_j^{(o)}\mathbf{B}_n(\cdot)$, using B-spline approximation for $\boldsymbol{\phi}_j(t)$ and applying Taylor's expansion to $f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t)\right\}$ at $f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j^{(o)}(t)\right\}$, we have $f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t)\right\} \approx f_j(\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j^{(o)}(t)\} + \nabla f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j^{(o)}(t)\right\}\boldsymbol{\xi}_i^T\left\{\boldsymbol{\Gamma}_j - \boldsymbol{\Gamma}_j^{(o)}\right\}\mathbf{B}_n(t)$ for $\boldsymbol{\Gamma}_j$ close to $\boldsymbol{\Gamma}_j^{(o)}$, where $\boldsymbol{\phi}_j^{(o)}$ and $\boldsymbol{\Gamma}_j^{(o)}$ denote the provided estimators for $\boldsymbol{\phi}_j$ and $\boldsymbol{\Gamma}_j$, respectively, exemplifying the (o)th iteration estimator. Substituting this approximation into (3.6) in the main paper yields a least square function for $\boldsymbol{\Gamma}_j$ as follows:

$$\sum_{i=1}^{n}\sum_{d=1}^{n_{ij}}\Big[X_{ij}(t_{ijd})-\widehat{\mu}_j(t_{ijd})-f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j^{(o)}\mathbf{B}_n(t_{ijd})\right\}+\nabla f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j^{(o)}\mathbf{B}_n(t_{ijd})\right\}$$
$$\times\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j^{(o)}\mathbf{B}_n(t_{ijd})-\nabla f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j^{(o)}\mathbf{B}_n(t_{ijd})\right\}(\boldsymbol{\xi}_i\otimes\mathbf{B}_n(t_{ijd}))^T\vec{\boldsymbol{\Gamma}}_j\Big]^2,$$

where "$\otimes$" is the kronecker product, and $\vec{\boldsymbol{\Gamma}}_j$ denotes the vector formed by concatenating the rows of matrix $\boldsymbol{\Gamma}_j$.

Let $\mathbf{Z}_{ijd}^{\boldsymbol{\Gamma}}=\nabla f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j^{(o)}\mathbf{B}_n(t_{ijd})\right\}\{\boldsymbol{\xi}_i\otimes\mathbf{B}_n(t_{ijd})\}$, we obtain the following explicit expression:

$$\widehat{\vec{\boldsymbol{\Gamma}}}_j=\left\{\sum_{i=1}^{n}\sum_{d=1}^{n_{ij}}\mathbf{Z}_{ijd}^{\boldsymbol{\Gamma}}\mathbf{Z}_{ijd}^{\boldsymbol{\Gamma}\,T}\right\}^{-1}\Bigg(\sum_{i=1}^{n}\sum_{d=1}^{n_{ij}}\mathbf{Z}_{ijd}^{\boldsymbol{\Gamma}}\big[X_{ij}(t_{ijd})-\widehat{\mu}_j(t_{ijd})$$
$$-f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j^{(o)}\mathbf{B}_n(t_{ijd})\right\}+\nabla f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j^{(o)}\mathbf{B}_n(t_{ijd})\right\}\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j^{(o)}\mathbf{B}_n(t_{ijd})\big]\Bigg). \tag{S0.5}$$

Given $(f_j(\cdot),\nabla f_j(\cdot))$ and $\boldsymbol{\phi}_j(\cdot)$, the local linear approximation $f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd})\right\}\approx f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd'})\right\}+\nabla f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd'})\right\}\boldsymbol{\xi}_i^T\left\{\boldsymbol{\phi}_j(t_{ijd})-\boldsymbol{\phi}_j(t_{ijd'})\right\}$ for $\boldsymbol{\phi}_j(t_{ijd})$ close to $\boldsymbol{\phi}_j(t_{ijd'})$, together with (3.6) in the main paper leads to a weighted least square function for $\boldsymbol{\xi}_i$,

$$\sum_{j=1}^{p}\sum_{d'=1}^{n_{ij}}\frac{1}{\sum_{d=1}^{n_{ij}}\omega_{ij,dd'}}\Bigg(\sum_{d=1}^{n_{ij}}\big[X_{ij}(t_{ijd})-\widehat{\mu}_j(t_{ijd})-f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd'})\right\}$$
$$-\nabla f_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd'})\right\}\left\{\boldsymbol{\phi}_j(t_{ijd})-\boldsymbol{\phi}_j(t_{ijd'})\right\}^T\boldsymbol{\xi}_i\big]^2\omega_{ij,dd'}\Bigg), \tag{S0.6}$$

where $\omega_{ij,dd'}=\mathcal{K}_{h_1}\left\{\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd})-\boldsymbol{\xi}_i^T\boldsymbol{\phi}_j(t_{ijd'})\right\}$.

Let $\mathbf{Z}^{\boldsymbol{\xi}}_{ij,dd'} = \nabla f_j \left\{ \boldsymbol{\xi}_i^T \boldsymbol{\phi}_j(t_{ijd'}) \right\} \left\{ \boldsymbol{\phi}_j(t_{ijd}) - \boldsymbol{\phi}_j(t_{ijd'}) \right\}$. Hence, the estimator for $\boldsymbol{\xi}_i$ is

$$\widehat{\boldsymbol{\xi}}_i = \left(\sum_{j=1}^{p} \sum_{d'=1}^{n_{ij}} \sum_{d=1}^{n_{ij}} \frac{\omega_{ij,dd'}}{\sum_{d=1}^{n_{ij}} \omega_{ij,dd'}} \mathbf{Z}^{\boldsymbol{\xi}}_{ij,dd'} \mathbf{Z}^{\boldsymbol{\xi}\,T}_{ij,dd'} \right)^{-1} \tag{S0.7}$$
$$\cdot \left(\sum_{j=1}^{p} \sum_{d'=1}^{n_{ij}} \sum_{d=1}^{n_{ij}} \frac{\omega_{ij,dd'}}{\sum_{d=1}^{n_{ij}} \omega_{ij,dd'}} \mathbf{Z}^{\boldsymbol{\xi}}_{ij,dd'} \left[X_{ij}(t_{ijd}) - \widehat{\mu}_j(t_{ijd}) - f_j \left\{ \boldsymbol{\xi}_i^T \boldsymbol{\phi}_j(t_{ijd'}) \right\} \right] \right).$$

Remark 1. We can also estimate $f_j(\cdot), \boldsymbol{\Gamma}_j$, and $\boldsymbol{\xi}_i$ by fitting a single-index model based on the observations $\left\{ X_{ij}(t_{ijd}) - \widehat{\mu}_j(t_{ijd}), \mathbf{B}_n(t_{ijd}) \right\}, \forall d = 1, \ldots, n_{ij}$, and then perform PCA on $\left(\widehat{\boldsymbol{\xi}_i^T \boldsymbol{\Gamma}_1}, \ldots, \widehat{\boldsymbol{\xi}_i^T \boldsymbol{\Gamma}_p} \right)$. However, this procedure increases the computational cost because it results in np models of k_n dimension and high computation intensity from PCA on an $n \times k_n p$ matrix.

Estimation in FMVCM with Sparse Penalty

To keep the splines smooth across the shared edges of adjoining triangles, we require the spline coefficients to satisfy $\mathbf{H}(\boldsymbol{\alpha}_k^*)^T = \mathbf{0}$ and $\mathbf{H}(\boldsymbol{\beta}^*)^T = \mathbf{0}$. Here, the bivariate spline basis $\mathbf{S}_n^*$ and constraint matrix $\mathbf{H}$ can be constructed based on the R package *BPST* (Wang et al., 2019). We remove the linear constraints $\mathbf{H}(\boldsymbol{\alpha}_k^*)^T = \mathbf{0}$ and $\mathbf{H}(\boldsymbol{\beta}^*)^T = \mathbf{0}$ through QR decomposition $\mathbf{H}^T = \mathbf{Q}\mathbf{R} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{pmatrix} \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{pmatrix}$ to simplify computation, where the submatrix $\mathbf{Q}_1$ is the first $r_{\mathbf{H}}$ columns of $\mathbf{Q}$, with $r_{\mathbf{H}}$ being the rank of $\mathbf{H}$, and $\mathbf{R}_2$ is a matrix of zeros. Denote $\boldsymbol{\alpha}_k^* = \boldsymbol{\alpha}_k \mathbf{Q}_2^T$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta} \mathbf{Q}_2^T$ for some $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}$. Then, we perform reparameterization using

$$\mathbf{g}_{nk}(\mathbf{s}) = \boldsymbol{\alpha}_k \mathbf{S}_n(\mathbf{s}), \quad \boldsymbol{\theta}_n(\mathbf{s}) = \boldsymbol{\beta} \mathbf{S}_n(\mathbf{s}), \tag{S0.8}$$

where $\mathbf{S}_n(\mathbf{s}) = \mathbf{Q}_2^T\mathbf{S}_n^*(\mathbf{s})$ and the reparametrized spline coefficients satisfy $\mathbf{H}\boldsymbol{\alpha}_k^T = \mathbf{0}$ and $\mathbf{H}\boldsymbol{\beta}^T = \mathbf{0}$.

Based on the approximation (S0.8), model (2.5) in the main paper can be considered as a standard multiple-index model. Then, the procedure discussed in Xia (2008) can be used for estimation. However, this procedure leads to considerably intensive computation because it needs to minimize the local linear approximation errors for all approximation points $\sum_{k=1}^{K_n}\mathbf{g}_k(\mathbf{s}_j)\widehat{\zeta}_{ik}, i = 1,\ldots,n, j = 1,\ldots,N_s$ in the estimation of $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}$, where N_s and the dimension of bivariate spline coefficients are relatively large in the imaging data analysis. In addition, this method cannot deal with the problem of group sparsity, being inapplicable to images on high-dimensional scalar regression. Therefore, we develop an estimation procedure to avoid the two issues. The procedure is stated in the subsequent section.

Estimation of $\psi(\cdot), \mathbf{G}(\cdot)$, and $\boldsymbol{\theta}(\cdot)$

Let

$$l_2(\mathbf{G},\boldsymbol{\theta},\psi) = \frac{1}{nN_s}\sum_{i=1}^{n}\sum_{j=1}^{N_s}\left[Y_i(\mathbf{s}_j) - \psi\left\{\sum_{k=1}^{K_n}\mathbf{g}_k(\mathbf{s}_j)\widehat{\zeta}_{ik}\right\} - \boldsymbol{\theta}(\mathbf{s}_j)^T\mathbf{Z}_i\right]^2 + \sum_{k=1}^{K_n}\lambda\frac{\|\mathbf{g}_k\|_2}{\|\widetilde{\mathbf{g}}_k\|_2}.$$

The minimization problem in FMVCM can be decomposed into three linear regression problems and solved by alternatively fixing $\psi(\cdot)$, $\mathbf{G}(\cdot)$, or $\boldsymbol{\theta}(\cdot)$.

Given $\mathbf{G}(\cdot)$ and $\boldsymbol{\theta}(\cdot)$, for any given $\mathbf{u}$, applying local linear approximation to $\psi(\cdot)$ and plugging the approximation $\psi\left\{\sum_{k=1}^{K_n}\mathbf{g}_k(\mathbf{s})\widehat{\zeta}_{ik}\right\} \approx \psi(\mathbf{u}) + \nabla\psi(\mathbf{u})^T\left\{\sum_{k=1}^{K_n}\mathbf{g}_k(\mathbf{s})\widehat{\zeta}_{ik} - \mathbf{u}\right\}$ into the objective function in the main paper yield the weighted least square function for $\left\{\psi(\mathbf{u}), \nabla\psi(\mathbf{u})^T\right\}^T$, we have

$$\frac{1}{nN_s}\sum_{i=1}^{n}\sum_{j=1}^{N_s}\left[Y_i(\mathbf{s}_j) - \psi(\mathbf{u}) - \nabla\psi(\mathbf{u})^T\left\{\sum_{k=1}^{K_n}\mathbf{g}_k(\mathbf{s}_j)\widehat{\zeta}_{ik} - \mathbf{u}\right\} - \boldsymbol{\theta}(\mathbf{s}_j)^T\mathbf{Z}_i\right]^2$$
$$\times\mathcal{K}_{h_2}(\|\sum_{k=1}^{K_n}\mathbf{g}_k(\mathbf{s}_j)\widehat{\zeta}_{ik} - \mathbf{u}\|), \tag{S0.9}$$

where $\sum_{k=1}^{K_n}\mathbf{g}_k(\mathbf{s}_j)\widehat{\zeta}_{ik}$ is in a small neighborhood of $\mathbf{u}$, and h_2 is a bandwidth. Let $\mathbf{Z}_{ij}^{\psi} = (1, (\sum_{k=1}^{K_n}\mathbf{g}_k(\mathbf{s}_j)\widehat{\zeta}_{ik} - \mathbf{u})^T)^T$. Solving the minimization of (S0.9) leads to the solution,

$$\left\{\begin{array}{c}\widehat{\psi}(\mathbf{u}) \\ \widehat{\nabla\psi}(\mathbf{u})\end{array}\right\} = \left\{\sum_{i=1}^{n}\sum_{j=1}^{N_s}\mathcal{K}_{h_2}(\|\sum_{k=1}^{K_n}\mathbf{g}_k(\mathbf{s}_j)\widehat{\zeta}_{ik} - \mathbf{u}\|)\mathbf{Z}_{ij}^{\psi}\mathbf{Z}_{ij}^{\psi\,T}\right\}^{-1} \tag{S0.10}$$
$$\cdot\left[\sum_{i=1}^{n}\sum_{j=1}^{N_s}\mathcal{K}_{h_2}\left\{\sum_{k=1}^{K_n}\mathbf{g}_k(\mathbf{s}_j)\widehat{\zeta}_{ik} - \mathbf{u}\right\}\left\{Y_i(\mathbf{s}_j) - \boldsymbol{\theta}(\mathbf{s}_j)^T\mathbf{Z}_i\right\}\mathbf{Z}_{ij}^{\psi}\right].$$

Given $\psi(\cdot)$ and $\mathbf{G}(\cdot)$, plugging the bivariate spline approximation $\boldsymbol{\theta}(\mathbf{s}) \approx \boldsymbol{\theta}_n(\mathbf{s}) = \boldsymbol{\beta}\mathbf{S}_n(\mathbf{s})$ into (S0.9) leads to a least square function for $\boldsymbol{\beta}$,

$$\frac{1}{nN_s}\sum_{i=1}^{n}\sum_{j=1}^{N_s}\left[Y_i(\mathbf{s}_j) - \psi\left\{\sum_{k=1}^{K_n}\mathbf{g}_k(\mathbf{s}_j)\widehat{\zeta}_{ik}\right\} - \mathbf{S}_n(\mathbf{s}_j)^T\boldsymbol{\beta}^T\mathbf{Z}_i\right]^2,$$

which yields the estimator of $\boldsymbol{\beta}$ as

$$\widehat{\vec{\boldsymbol{\beta}}} = \left[\sum_{i=1}^{n}\sum_{j=1}^{N_s}\left\{\mathbf{Z}_i\otimes\mathbf{S}_n(\mathbf{s}_j)\right\}\left\{\mathbf{Z}_i\otimes\mathbf{S}_n(\mathbf{s}_j)\right\}^T\right]^{-1} \tag{S0.11}$$
$$\cdot\left(\sum_{i=1}^{n}\sum_{j=1}^{N_s}\left[Y_i(\mathbf{s}_j) - \psi\left\{\sum_{k=1}^{K_n}\mathbf{g}_k(\mathbf{s}_j)\widehat{\zeta}_{ik}\right\}\right]\left\{\mathbf{Z}_i\otimes\mathbf{S}_n(\mathbf{s}_j)\right\}\right).$$

Given $\left\{\psi(\cdot), \nabla\psi(\cdot)^T, \boldsymbol{\theta}(\cdot)\right\}$ and $\mathbf{g}_k^{(o)}(\cdot) = \boldsymbol{\alpha}_k^{(o)}\mathbf{S}_n(\cdot), k = 1, \ldots, K_n$, and let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{K_n})$, $\mathbf{Z}_{ij}^{\boldsymbol{\alpha}} = \left[\nabla\psi\left\{\sum_{k=1}^{K_n}\boldsymbol{\alpha}_k^{(o)}\mathbf{S}_n(\mathbf{s}_j)\widehat{\zeta}_{ik}\right\}\otimes\left\{\widehat{\boldsymbol{\zeta}}_i\otimes\mathbf{S}_n(\mathbf{s}_j)\right\}\right]$, and $\mathbf{Z}_{ij}^{\boldsymbol{\alpha}} = \left[\nabla\psi\left\{\sum_{k=1}^{K_n}\boldsymbol{\alpha}_k^{(o)}\mathbf{S}_n(\mathbf{s}_j)\widehat{\zeta}_{ik}\right\}\right.$

$\otimes \left\{\mathbf{S}_n(\mathbf{s}_j) \otimes \widehat{\boldsymbol{\zeta}}_i^T\right\}\Big]^T$. Using the bivariate spline approximation for $\mathbf{g}_k(\mathbf{s})$ and applying Taylor's expansion to $\psi\left\{\sum_{k=1}^{K_n} \mathbf{g}_k(\mathbf{s}_j)\widehat{\zeta}_{ik}\right\}$ at $\psi\left\{\sum_{k=1}^{K_n} \mathbf{g}_k^{(o)}(\mathbf{s}_j)\widehat{\zeta}_{ik}\right\}$, we have $\psi\left\{\sum_{k=1}^{K_n} \mathbf{g}_k(\mathbf{s}_j)\widehat{\zeta}_{ik}\right\} \approx \psi\left\{\sum_{k=1}^{K_n} \mathbf{g}_k^{(o)}(\mathbf{s}_j)\widehat{\zeta}_{ik}\right\} + \nabla\psi\left\{\sum_{k=1}^{K_n} \mathbf{g}_k^{(o)}(\mathbf{s}_j)\widehat{\zeta}_{ik}\right\}\widehat{\boldsymbol{\zeta}}_i^T\left\{\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_k^{(o)}\right\}\mathbf{S}_n(\mathbf{s}_j)$ for $\boldsymbol{\alpha}_k$ close $\boldsymbol{\alpha}_k^{(o)}$. Substituting this approximation into (S0.9) yields a penalized least square function for $\vec{\boldsymbol{\alpha}}$, we have

$$\frac{1}{nN_s}\sum_{i=1}^{n}\sum_{j=1}^{N_s}\Big[Y_i(\mathbf{s}_j) - \boldsymbol{\theta}(\mathbf{s}_j)^T\mathbf{Z}_i - \psi\Big\{\sum_{k=1}^{K_n}\mathbf{g}_k^{(o)}(\mathbf{s}_j)\widehat{\zeta}_{ik}\Big\} \tag{S0.12}$$

$$+\nabla\psi\Big\{\sum_{k=1}^{K_n}\mathbf{g}_k^{(o)}(\mathbf{s}_j)\widehat{\zeta}_{ik}\Big\}^T\sum_{k=1}^{K_n}\boldsymbol{\alpha}_k^{(o)}\mathbf{S}_n(\mathbf{s}_j)\widehat{\zeta}_{ik} - \vec{\boldsymbol{\alpha}}^T\mathbf{Z}_{ij}^{\boldsymbol{\alpha}}\Big]^2 + \sum_{k=1}^{K_n}\lambda\frac{\|\boldsymbol{\alpha}_k\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|},$$

where $\widetilde{\boldsymbol{\alpha}}_k$ is the last iterative estimator of $\boldsymbol{\alpha}_k$. The minimization of (S0.12) can be solved based on the R package *grpreg*, where λ needn't be given in advance.

Implementation of Algorithms

To implement the estimation procedure, we must set initial values $(\boldsymbol{\Gamma}_j^{(0)}, \boldsymbol{\xi}_i^{(0)}, i = 1, \ldots, n, j = 1, \ldots, p)$ for $(\boldsymbol{\Gamma}_j, \boldsymbol{\xi}_i, i = 1, \ldots, n, j = 1, \ldots, p)$ and $(\boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)})$ for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. We obtain $(\boldsymbol{\Gamma}_j^{(0)}, \boldsymbol{\xi}_i^{(0)}, i = 1, \ldots, n, j = 1, \ldots, p)$ from MFPCA (Happ and Greven, 2018) based on the R package *MFPCA* and $(\boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)})$ based on the R package *MAVE* or from the linear regression of $Y_i(\mathbf{s}_j)$ on $\left[\left\{\widehat{\boldsymbol{\zeta}}_i \otimes \mathbf{S}_n(\mathbf{s}_j)\right\}^T, \{\mathbf{Z}_i \otimes \mathbf{S}_n(\mathbf{s}_j)\}^T\right]^T$ if the MAVE estimator is time-consuming. The MAVE and linear regression estimators are consistent under some regularity conditions (Duan and Li, 1991; Xia et al., 2002). Thus, they can be used as the adaptive weight. Denote $\left\{f_j^{(o-1)}(\cdot), \boldsymbol{\xi}_i^{(o-1)}, \boldsymbol{\Gamma}_j^{(o-1)}\right\}$ and

$\left\{\psi^{(o-1)}(\cdot), \boldsymbol{\alpha}^{(o-1)}, \boldsymbol{\beta}^{(o-1)}\right\}$ as the estimators of $\{f_j(\cdot), \boldsymbol{\xi}_i, \boldsymbol{\Gamma}_j\}$ and $\{\psi(\cdot), \boldsymbol{\alpha}, \boldsymbol{\beta}\}$, respectively, obtained after the $(o-1)$th iteration. The estimation algorithm is implemented as follows.

- Perform nonlinear MFPCA: with the estimator $\widehat{\mu}_j(\cdot)$ based on (S0.2), we obtain the estimators $\{f_j(\cdot), \boldsymbol{\xi}_i, \boldsymbol{\Gamma}_j\}$ by using equations (S0.4), (S0.5), and (S0.7), update $(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi})$ by using the maximum improvement of $l_1(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi})$ (Zhu et al., 2016), and repeat this procedure until convergence. At each step, $\{f_j(\cdot), \boldsymbol{\xi}_i, \boldsymbol{\Gamma}_j\}$ on the right-hand side of the equations are replaced by their most updated value. The convergence is determined by $\|\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\} - \{\mathbf{f}^{(o-1)}, \boldsymbol{\phi}^{(o-1)}, \boldsymbol{\xi}^{(o-1)}\}\| / \|\{\mathbf{f}^{(o-1)}, \boldsymbol{\phi}^{(o-1)}, \boldsymbol{\xi}^{(o-1)}\}\| \leq 10^{-3}$ or $|l_1\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\} - l_1\{\mathbf{f}^{(o-1)}, \boldsymbol{\phi}^{(o-1)}, \boldsymbol{\xi}^{(o-1)}\}| / |l_1\{\mathbf{f}^{(o-1)}, \boldsymbol{\phi}^{(o-1)}, \boldsymbol{\xi}^{(o-1)}\}| \leq 10^{-3}$.

- Transform the scores: obtain the transformed variables $\widehat{\zeta}_{ik} = \Phi\left(\widehat{\lambda}_k^{-1/2}\widehat{\xi}_{ik}\right) - 0.5, k = 1, \ldots, K_n$, where $\widehat{\lambda}_k$ is the kth eigenvalue of $\frac{1}{n}\sum_{i=1}^n \widehat{\boldsymbol{\xi}}_i \widehat{\boldsymbol{\xi}}_i^T$.

- Conduct estimation for FMVCM: treat $\widehat{\boldsymbol{\zeta}}_i$ and $\mathbf{Z}_i$ as covariates, obtain the estimators $(\psi(\cdot), \boldsymbol{\alpha}, \boldsymbol{\beta})$ by repeatedly using equations (S0.10) and (S0.11) and solving (S0.12) through the R package *grpreg*, and replace $(\psi(\cdot), \boldsymbol{\alpha}, \boldsymbol{\beta})$ on the right-hand side of the equations by their most updated value, update $(\mathbf{G}, \boldsymbol{\theta}, \psi)$ by using the maximum improvement of $l_2(\mathbf{G}, \boldsymbol{\theta}, \psi)$ (Zhu et al., 2016), and repeat this procedure until convergence. The convergence is determined by $\|\{\mathbf{G}^{(o)}, \boldsymbol{\theta}^{(o)}, \psi^{(o)}\} -$

$\{\mathbf{G}^{(o-1)}, \boldsymbol{\theta}^{(o-1)}, \psi^{(o-1)}\}\|/\|\{\mathbf{G}^{(o-1)}, \boldsymbol{\theta}^{(o-1)}, \psi^{(o-1)}\}\| \leq 10^{-3}$ or $|l_2\{\mathbf{G}^{(o)}, \boldsymbol{\theta}^{(o)}, \psi^{(o)}\} - l_2\{\mathbf{G}^{(o-1)}, \boldsymbol{\theta}^{(o-1)}, \psi^{(o-1)}\}|/|l_2\{\mathbf{G}^{(o-1)}, \boldsymbol{\theta}^{(o-1)}, \psi^{(o-1)}\}| \leq 10^{-3}$.

In the following, we give a definition of a closed mapping and two assumptions, then establish the convergence of the proposed iterative algorithm.

Definition 1. (David G. Luenberger (2016), page 199) A point-to-set mapping F from $\mathcal{X}$ to $\mathcal{Y}$ is said to be closed at $x \in \mathcal{X}$ if the assumptions 1) $\lim_{k\to\infty} x_k \to x$, 2) $\lim_{k\to\infty} y_k \to y$ and 3) $y_k \in F(x_k)$ imply $y \in F(x)$. Moreover, F is said to be closed over $\mathcal{X}$ if F is closed at every point of $\mathcal{X}$.

Let F_1 be the mapping function of the proposed iterative algorithm for NMF-PCA, which means that the algorithm generates the sequences $\left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\}$ by $\left\{\mathbf{f}^{(o+1)}, \boldsymbol{\phi}^{(o+1)}, \boldsymbol{\xi}^{(o+1)}\right\} = F_1\left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\}$. Let $\mathcal{Q}_1$ be the space of $(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi})$, $\mathcal{Q}_{10} = \left\{(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi}) \in \mathcal{Q}_1 : l_1(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi}) \leq l_1\left(\mathbf{f}^{(0)}, \boldsymbol{\phi}^{(0)}, \boldsymbol{\xi}^{(0)}\right)\right\}$, and $\mathcal{Q}_{1*} = \{$set of local minima of $l_1(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi})$ in the interior of $\mathcal{Q}_{10}\}$. Similarly, we define F_2 as the mapping function of the proposed iterative algorithm for FMVCM, $\mathcal{Q}_2$ as the space of $(\mathbf{G}, \boldsymbol{\theta}, \psi)$, $\mathcal{Q}_{20} = \left\{(\mathbf{G}, \boldsymbol{\theta}, \psi) \in \mathcal{Q}_2 : l_2(\mathbf{G}, \boldsymbol{\theta}, \psi) \leq l_2\left(\mathbf{G}^{(0)}, \boldsymbol{\theta}^{(0)}, \psi^{(0)}\right)\right\}$, and $\mathcal{Q}_{2*} = \{$set of local minima of $l_2(\mathbf{G}, \boldsymbol{\theta}, \psi)$ in the interior of $\mathcal{Q}_{20}\}$.

Then, we assume the following two conditions:

(S1) $\mathcal{Q}_{10}$ and $\mathcal{Q}_{20}$ is compact given the initial value $\left(\mathbf{f}^{(0)}, \boldsymbol{\phi}^{(0)}, \boldsymbol{\xi}^{(0)}\right)$ and $\left(\mathbf{G}^{(0)}, \boldsymbol{\theta}^{(0)}, \psi^{(0)}\right)$, respectively.

(S2) $F_1(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi})$ and $F_2(\mathbf{G}, \boldsymbol{\theta}, \psi)$ is closed over $\mathcal{Q}_{10} \backslash \mathcal{Q}_{1*}$ and $\mathcal{Q}_{20} \backslash \mathcal{Q}_{2*}$, respectively, i.e., the difference of two sets.

Proposition S.1. *If conditions (S1) and (S2) hold, we have*

(i) all the limit points of $\left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\}$ *are local minima of* $l_1(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi})$ *in the space* $\mathcal{Q}_{10}$*, and* $l_1\left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\}$ *converges monotonically to* $L_1^* = l_1(\mathbf{f}^*, \boldsymbol{\phi}^*, \boldsymbol{\xi}^*)$ *for some* $(\mathbf{f}^*, \boldsymbol{\phi}^*, \boldsymbol{\xi}^*) \in \mathcal{Q}_{1*}$.

(ii) all the limit points of $\left\{\mathbf{G}^{(o)}, \boldsymbol{\theta}^{(o)}, \psi^{(o)}\right\}$ *are local minima of* $l_2(\mathbf{G}, \boldsymbol{\theta}, \psi)$ *in the space* $\mathcal{Q}_{20}$*, and* $l_2\left\{\mathbf{G}^{(o)}, \boldsymbol{\theta}^{(o)}, \psi^{(o)}\right\}$ *converges monotonically to* $L_2^* = l_2(\mathbf{G}^*, \boldsymbol{\theta}^*, \psi^*)$ *for some* $(G^*, \boldsymbol{\theta}^*, \psi^*) \in \mathcal{Q}_{2*}$.

Web Appendix B: Proofs of theoretical properties

Let $\nabla_{s_j}^i g(\mathbf{s}) = \frac{\partial^i g(\mathbf{s})}{\partial s_j^i}, \nabla^i g(\mathbf{s}) = \frac{d^i g(\mathbf{s})}{d\mathbf{s}^i}$, and $|g|_{l,\infty} = \max_{i+j=l} \|\nabla_{s_1}^i \nabla_{s_2}^j g(\mathbf{s})\|_\infty$ be the maximum norms of all the lth order derivatives of g over $\mathcal{D}$. Define $\mathcal{W}^{\varpi+1,\infty}(\mathcal{D}) = \{g : |g|_{l,\infty} < \infty, 0 \le l \le \varpi + 1\}$ be the standard Sobolev space, the Hölder space of order $r = l + s$ as $\mathcal{H}_r = \{\, f(\cdot) : |\nabla^l f(t_1) - \nabla^l f(t_2)| \le c|t_1 - t_2|^s, \text{ for any } 0 \le t_1, t_2 \le 1 \,\}$, where l is a non-negative integer, $c > 0$ is a constant, and $s \in (0, 1]$.

Regularity conditions for model identifiability.

(I1) functions $f_j(\cdot), j = 1, \ldots, p$ are continuously differentiable with finite number of zero derivatives, and $\sum_{j=1}^p \int_u f_j(u) du > 0$;

(I2) $\sum_{j=1}^p \int_{\mathcal{T}} \phi_{jk}(t) \phi_{jk'}(t) dt = 0$ if $k' \neq k$, and 1 otherwise, $\lambda_1 > \lambda_2 > \ldots > 0$,

$\int_{\mathcal{T}} \phi_{j1}(t)^2 dt$ and λ_1 are fixed constants, and $\phi_{1k}(c) > 0$, where $c \in \mathcal{T}$ is a fixed constant;

(I3) The derivative $\nabla\psi(\mathbf{G}(\mathbf{s})\boldsymbol{\zeta}_i)$ exists and $E\left[\nabla\psi\left\{\mathbf{G}(\mathbf{s})\boldsymbol{\zeta}_i\right\}\nabla\psi\left\{\mathbf{G}(\mathbf{s})\boldsymbol{\zeta}_i\right\}^T\right]$ is invertible; $\int_{\mathcal{D}} \mathbf{G}(\mathbf{s})\mathbf{G}(\mathbf{s})^T d\mathbf{s} = \mathbf{I}$, $\int_{\mathcal{D}} \mathbf{G}(\mathbf{s})\mathrm{var}(\boldsymbol{\zeta}_i)\mathbf{G}(\mathbf{s})^T d\mathbf{s}$ is a diagonal matrix with decreasing diagonal elements, and the integration of the first nonzero function in each row of $\mathbf{G}(\mathbf{s})$ over $\mathcal{D}$ is positive.

Condition (I1) is a requirement of smoothing and sign identifiable for nonconstant nonparametric functions $f_j(\cdot)$s. Condition (I2) is equivalent to those used in the existing FPCA (Zhu et al., 2014; Happ and Greven, 2018; Wong et al., 2019) with a scale transformation. Condition (I3) imposes restrictions on the derivative of $\psi(\cdot)$ and orthogonality constraints on the varying coefficient vector $\mathbf{G}(\cdot)$; these conditions are all moderate. Without loss of generality, we assume that $\mathrm{var}(\boldsymbol{\zeta}_i) \neq M\mathbf{I}$ with $M > 0$; otherwise, let $\mathbf{G}(\mathbf{s})\mathbf{M}^T$ and $\mathbf{M}\boldsymbol{\zeta}_i$ be the new varying coefficients and transformed scores, where $\mathbf{M}^T\mathbf{M} = \mathbf{I}$ and $\mathbf{M}\mathbf{M}^T \neq \mathbf{I}$.

Regularity conditions for the asymptotic properties.

(C1) Identifiable condition: (I1)–(I3) hold.

(C2) Kernel function: $\mathcal{K}(\cdot)$ is a bounded symmetrical density function with bounded derivatives and satisfies

$$\int u^2\mathcal{K}(u)du \neq 0 \ \text{ and } \ \int |u|^j\mathcal{K}(u)du < \infty, j = 1, 2, \ldots.$$

(C3) Derivatives: all the second-order derivatives of the density functions of $\boldsymbol{\xi}_i^T \boldsymbol{\phi}_{j0}(t)$ and $\mathbf{G}_0(\mathbf{s})\boldsymbol{\zeta}_i$, link functions $f_{j0}(\cdot)$ and $\psi_0(\cdot)$, conditional expectations $E\left\{\boldsymbol{\xi}_i|\boldsymbol{\xi}_i^T \boldsymbol{\phi}_{j0}(t) = \cdot\right\}$ and $E\left\{\boldsymbol{\zeta}_i|\mathbf{G}_0(\mathbf{s})\boldsymbol{\zeta}_i = \cdot\right\}$, and conditional variances $E\left\{\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T|\boldsymbol{\xi}_i^T \boldsymbol{\phi}_{j0}(t) = \cdot\right\}$ and $E\left\{\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^T|\mathbf{G}_0(\mathbf{s})\boldsymbol{\zeta}_i = \cdot\right\}$ are continuous and bounded.

(C4) Bounded eigenvalue: The eigenvalues of $E\left(\mathbf{Z}_i\mathbf{Z}_i^T\right)$ are bounded away from 0 and infinity.

(C5) Smoothness: For $j = 1, \ldots, p, k = 1, \ldots, K_n, l = 1, \ldots, q, l' = 1 \ldots, q_z$, the functions $\mu_{j0}(\cdot)$ and $\phi_{jk,0}(\cdot)$ belong to Hölder space of order $r \geq 2$, and the varying coefficient functions $g_{kl,0}(\cdot)$ and $\theta_{l',0}(\cdot)$ belong to Sobolev space $\mathcal{W}^{\varpi+1,\infty}(\mathcal{D})$ with $\varpi \geq 1$.

(C6) Moment and Errors: $\sum_{j=1}^{p} \int_0^1 E[\{X_{ij}(t) - \mu_j(t)\}^4]dt$ is finite, $\varepsilon_{ij}(t)$s are independent measurement errors and independent of $\epsilon_i(\mathbf{s})$, and there exists a positive constant $M < \infty$, such that

$$\frac{1}{nN_s} \sum_{i=1}^{n} \sum_{i'=1}^{n} \sum_{j=1}^{N_s} \sum_{j'=1}^{N_s} |E\left\{\epsilon_i(\mathbf{s}_j)\epsilon_{i'}(\mathbf{s}_{j'})\right\}| \leq M.$$

(C7) Variation decays: there exist some $c_0 > 1$ and $0 < M < \infty$, such that $\lambda_{k0} - \lambda_{k+1,0} \geq Mk^{-c_0-1}$.

(C8) Triangulations: $|\triangle| \to 0$, and the triangulation $\triangle$ is π-quasi-uniform; that is, $(\min_{\tau \in \triangle} R_\tau)^{-1} |\triangle| \leq \pi$ for some positive constant π.

Condition (C1) ensures model identifiability. Conditions (C2) and (C3) are usual requirements in the index model literature (Xia, 2008; Cui et al., 2011). Condition (C4) is a regular condition commonly used in regression analysis (Zhu et al., 2014; Yu et al., 2021). Condition (C5) describes the smoothness requirements on the nonparametric functions, which are frequently used in the nonparametric estimation literature. Condition (C6) imposes weaker restrictions on random errors than Yu et al. (2019, 2021) and Li et al. (2021) that require $\epsilon_i(\mathbf{s}_j)$s to be independent over i and j. Condition (C7) requires the polynomial decay rate of λ_{k0}, the same as Zhu et al. (2014) and Wong et al. (2019). Condition (C8) suggests using more uniform triangulations with smaller shape parameters and is common in the triangulation-based literature (Lai and Schumaker, 2007; Wang et al., 2020).

Proofs of propositions

Proof of proposition 1:

To prove proposition 1, we need to show that:

(i) $\mu_j(t), j = 1, \ldots, p$ are identifiable,

(ii) $f_j(\cdot), j = 1, \ldots, p$ are identifiable,

(iii) For all i, j, k, $\phi_{jk}(\cdot)$ and ξ_{ik} are identifiable.

Proof: (i). Because $E\{X_{ij}(t)\}$ is a unique fixed function, we have $E\{X_{ij}(t)\} = \mu_j(t)$, then $\mu_j(t)$ is identifiable.

(ii). Let $\boldsymbol{\xi}_i^{\infty} = (\xi_{i1}, \ldots, \xi_{i\infty})^T$, $\boldsymbol{\phi}_j^{\infty}(t) = \{\phi_{j1}(t), \phi_{j2}(t), \ldots, \phi_{j\infty}(t)\}^T$, and $\eta_{ij}^{\infty}(t) =$

$\sum_{k=1}^{\infty} \xi_{ik}\phi_{jk}(t)$. Since $f_j\left\{\sum_{k=1}^{\infty} \xi_{ik}\phi_{jk}(t)\right\} = f_j\left\{\eta_{ij}^{\infty}(t)\right\}$, thus, we only need to prove $\eta_{ij}^{\infty}(t)$ is identifiable in terms of location, scale, and sign. From the condition that ξ_{ik}s are zero-mean uncorrelated random variables, we have $E\left\{\eta_{ij}^{\infty}(t)\right\} = 0$, then $\eta_{ij}^{\infty}(t)$ is identifiable in terms of location. Next, we show that $\eta_{ij}^{\infty}(t)$ is identifiable regarding scale and sign. Denote $\bar{\eta}_{ij}^{\infty}(t) = \sum_{k=1}^{\infty} \bar{\xi}_{ik}\bar{\phi}_{jk}(t)$, with $\bar{f}_j(u), \bar{\phi}_{jk}(t)$, and $\bar{\xi}_{ik}$ satisfying Conditions (I1) and (I2), and $f_j\left\{\eta_{ij}^{\infty}(t)\right\} = \bar{f}_j\left\{\bar{\eta}_{ij}^{\infty}(t)\right\}$. Suppose that there exists a constant a_j such that

$$f_j(u/a_j) = \bar{f}_j(u), \tag{S0.13}$$

$$a_j\eta_{ij}^{\infty}(t) = \bar{\eta}_{ij}^{\infty}(t). \tag{S0.14}$$

Then, $a\boldsymbol{\xi}_i^{\infty} = \bar{\boldsymbol{\xi}}_i^{\infty}$ and $a_{j1}\boldsymbol{\phi}_j^{\infty}(t) = \bar{\boldsymbol{\phi}}_j^{\infty}(t)$ with $a \times a_{j1} = a_j$, where $\bar{\boldsymbol{\xi}}_i^{\infty} = \left(\bar{\xi}_{i1}, \ldots, \bar{\xi}_{i\infty}\right)^T$ and $\bar{\boldsymbol{\phi}}_j^{\infty}(t) = \left\{\bar{\phi}_{j1}(t), \bar{\phi}_{j2}(t), \ldots, \bar{\phi}_{j\infty}(t)\right\}^T$.

By Condition (I2), we have $a^2 = a_{j1}^2 = 1$. Furthermore, by conditions $\sum_{j=1}^{p}\int_u f_j(u)du > 0$ and $\phi_{1k}(c) > 0$, a and a_{j1} must be positive. Thus, $a = a_{j1} = 1$.

This proves the identifiability of $f_j(\cdot)$.

(iii). It is sufficient to show that $\sum_{j=1}^{p}\int_{\mathcal{T}} \boldsymbol{\phi}_j^{\infty}(t)\bar{\boldsymbol{\phi}}_j^{\infty,T}(t)dt$ is an identity matrix. By the results of (ii) we have

$$\boldsymbol{\xi}_i^{\infty,T}\boldsymbol{\phi}_j^{\infty}(t) = \bar{\boldsymbol{\xi}}_i^{\infty,T}\bar{\boldsymbol{\phi}}_j^{\infty}(t). \tag{S0.15}$$

Taking right multiplication of $\boldsymbol{\phi}_j^{\infty,T}(t)$ on both sides of (S0.15) with some algebraic calculations yields

$$\boldsymbol{\xi}_i^{\infty,T}\sum_{j=1}^{p}\int_{\mathcal{T}}\boldsymbol{\phi}_j^{\infty}(t)\boldsymbol{\phi}_j^{\infty,T}(t)dt=\bar{\boldsymbol{\xi}}_i^{\infty,T}\sum_{j=1}^{p}\int_{\mathcal{T}}\bar{\boldsymbol{\phi}}_j^{\infty}(t)\boldsymbol{\phi}_j^{\infty,T}(t)dt.$$

By the orthogonality of $\boldsymbol{\phi}_j^{\infty}(t)$, we have

$$\boldsymbol{\xi}_i^{\infty,T}=\bar{\boldsymbol{\xi}}_i^{\infty,T}\sum_{j=1}^{p}\int_{\mathcal{T}}\bar{\boldsymbol{\phi}}_j^{\infty}(t)\boldsymbol{\phi}_j^{\infty,T}(t)dt. \tag{S0.16}$$

Substituting (S0.16) into (S0.15), and taking right multiplication of $\bar{\boldsymbol{\phi}}_j^{\infty,T}(t)$ on both sides, we get

$$\begin{aligned}&\sum_{j=1}^{p}\int_{\mathcal{T}}\bar{\boldsymbol{\phi}}_j^{\infty}(t)\boldsymbol{\phi}_j^{\infty,T}(t)dt\sum_{j=1}^{p}\int_{\mathcal{T}}\boldsymbol{\phi}_j^{\infty}(t)\bar{\boldsymbol{\phi}}_j^{\infty,T}(t)dt\\&=\sum_{j=1}^{p}\int_{\mathcal{T}}\bar{\boldsymbol{\phi}}_j^{\infty}(t)\bar{\boldsymbol{\phi}}_j^{\infty,T}(t)dt\\&=\mathbf{I},\end{aligned} \tag{S0.17}$$

where the second equality holds from the orthogonality of $\bar{\boldsymbol{\phi}}_j^{\infty}(t)$.

The above shows that $\sum_{j=1}^{p}\int_{\mathcal{T}}\boldsymbol{\phi}_j^{\infty}(t)\bar{\boldsymbol{\phi}}_j^{\infty,T}(t)dt$ is an orthogonal matrix, so that its eigenvalues are either 1 or -1. Next, we show that

$\sum_{j=1}^{p}\int_{\mathcal{T}}\boldsymbol{\phi}_j^{\infty}(t)\bar{\boldsymbol{\phi}}_j^{\infty,T}(t)dt$ is an identity matrix. Taking variance and right multiplication of $\sum_{j=1}^{p}\int_{\mathcal{T}}\boldsymbol{\phi}_j^{\infty}(t)\bar{\boldsymbol{\phi}}_j^{\infty,T}(t)dt$ on both sides of (S0.16) yields

$$\mathrm{var}(\boldsymbol{\xi}_i^{\infty})\cdot\sum_{j=1}^{p}\int_{\mathcal{T}}\boldsymbol{\phi}_j^{\infty}(t)\bar{\boldsymbol{\phi}}_j^{\infty,T}(t)dt=\sum_{j=1}^{p}\int_{\mathcal{T}}\boldsymbol{\phi}_j^{\infty}(t)\bar{\boldsymbol{\phi}}_j^{\infty,T}(t)dt\cdot\mathrm{var}(\bar{\boldsymbol{\xi}}_i^{\infty}). \tag{S0.18}$$

This equation implies that $\sum_{j=1}^{p}\int_{\mathcal{T}}\boldsymbol{\phi}_j^{\infty}(t)\bar{\boldsymbol{\phi}}_j^{\infty,T}(t)dt$ is a matrix consisting of eigenvectors of $\mathrm{var}(\boldsymbol{\xi}_i^{\infty})$. Since $\mathrm{var}(\boldsymbol{\xi}_i^{\infty})$ and $\mathrm{var}(\bar{\boldsymbol{\xi}}_i^{\infty})$ are diagonal matrices, combining (S0.17) and (S0.18) yields the desired results.

Proof of proposition 2:

By the fact that $E\{X_{ij}(t)-\mu_j(t)\}=0$ and the mean value theorem,

$$\begin{aligned}
&\sum_{j=1}^{p}\sup_{t\in\mathcal{T}} E\left[X_{ij}(t)-\mu_j(t)-f_j\left\{\sum_{k=1}^{K_n}\xi_{ik}\phi_{jk}(t)\right\}-\varepsilon_{ij}(t)\right]^2\\
&=\sum_{j=1}^{p}\sup_{t\in\mathcal{T}}\operatorname{var}\left\{\dot{f}_j(u_{ij}^*)\sum_{k=K_n+1}^{\infty}\xi_{ik}\phi_{jk}(t)\right\}\\
&\leq M\sum_{j=1}^{p}\sup_{t\in\mathcal{T}}\sum_{k=K_n+1}^{\infty}\lambda_k\phi_{jk}(t)^2\to 0,
\end{aligned}$$

where u_{ij}^* lies between $\sum_{k=1}^{\infty}\xi_{ik}\phi_{jk}(t)$ and $\sum_{k=1}^{K_n}\xi_{ik}\phi_{jk}(t)$, and the last equality holds from the scores $\xi_{i1},\xi_{i2},\ldots$ are uncorrelated, and $\sum_k\lambda_k<\infty$.

Proof of proposition 3:

For $\psi\{\mathbf{G}(\mathbf{s})\boldsymbol{\zeta}_i\}=\bar{\psi}\{\bar{\mathbf{G}}(\mathbf{s})\boldsymbol{\zeta}_i\}$ with $\psi(\cdot),\bar{\psi}(\cdot),\mathbf{G}(\cdot)$, and $\bar{\mathbf{G}}(\cdot)$ satisfying Condition (I3), a direct calculation yields

$$\mathbf{G}(\mathbf{s})^T=\bar{\mathbf{G}}(\mathbf{s})^T\mathbf{A}_1, \tag{S0.19}$$

where $\mathbf{A}_1=E\left[\nabla\bar{\psi}\{\bar{\mathbf{G}}(\mathbf{s})\boldsymbol{\zeta}_i\}\nabla\psi\{\mathbf{G}(\mathbf{s})\boldsymbol{\zeta}_i\}^T\right]\left(E\left[\nabla\psi\{\mathbf{G}(\mathbf{s})\boldsymbol{\zeta}_i\}\nabla\psi\{\mathbf{G}(\mathbf{s})\boldsymbol{\zeta}_i\}^T\right]\right)^{-1}$. From the lines of (iii) in the proof of proposition 1, by Condition (I3), we can show $\mathbf{A}_1=\mathbf{I}$ by firstly proving $\mathbf{A}_1$ is an orthogonal matrix and then proving $\mathbf{A}_1$ is an diagonal matrix.

From (S0.19), we have $\psi\{\mathbf{A}_1^T\bar{\mathbf{G}}(\mathbf{s})\boldsymbol{\zeta}_i\}=\bar{\psi}\{\bar{\mathbf{G}}(\mathbf{s})\boldsymbol{\zeta}_i\}$, which implies that $\mathbf{A}_1$ is a fixed matrix and does not depend on $\mathbf{s}$. Therefore,

$$\begin{aligned}\int_{\mathcal{D}} \mathbf{G}(\mathbf{s})\mathbf{G}(\mathbf{s})^T d\mathbf{s} &= \mathbf{A}_1^T \cdot \int_{\mathcal{D}} \bar{\mathbf{G}}(\mathbf{s})\bar{\mathbf{G}}(\mathbf{s})^T d\mathbf{s} \cdot \mathbf{A}_1, \\ \mathbf{I} &= \mathbf{A}_1^T \mathbf{A}_1.\end{aligned}$$

Similarly, we have

$$\bar{\mathbf{G}}(\mathbf{s})^T = \mathbf{G}(\mathbf{s})^T \mathbf{A}_2 \text{ and } \mathbf{A}_2^T \mathbf{A}_2 = \mathbf{I}. \tag{S0.20}$$

Combining (S0.19) and (S0.20), we get $\mathbf{A}_2 = \mathbf{A}_1^T$ and $\bar{\mathbf{G}}(\mathbf{s})^T = \mathbf{G}(\mathbf{s})^T \mathbf{A}_1^T$. Thus,

$$\int_{\mathcal{D}} \bar{\mathbf{G}}(\mathbf{s})\mathrm{var}(\boldsymbol{\zeta}_i)\bar{\mathbf{G}}(\mathbf{s})^T d\mathbf{s}\mathbf{A}_1 = \mathbf{A}_1 \int_{\mathcal{D}} \mathbf{G}(\mathbf{s})\mathrm{var}(\boldsymbol{\zeta}_i)\mathbf{G}(\mathbf{s})^T d\mathbf{s}, \tag{S0.21}$$

which implies that $\mathbf{A}_1$ is a diagonal matrix from conditions $\int_{\mathcal{D}} \bar{\mathbf{G}}(\mathbf{s})\mathrm{var}(\boldsymbol{\zeta}_i)\bar{\mathbf{G}}(\mathbf{s})^T d\mathbf{s}$ and $\int_{\mathcal{D}} \mathbf{G}(\mathbf{s})\mathrm{var}(\boldsymbol{\zeta}_i)\mathbf{G}(\mathbf{s})^T d\mathbf{s}$ are diagonal matrices with positive decreasing diagonal elements. Therefore, $\mathbf{A}_1$ is an identity matrix. Thus, $\mathbf{G}(\mathbf{s}) = \bar{\mathbf{G}}(\mathbf{s})$ and $\psi(\cdot) = \bar{\psi}(\cdot)$.

Proof of Proposition 4. Let $\mathrm{BIC}(k) = \sum_{k=1}^K \widehat{\lambda}_k - \frac{K\widehat{\lambda}_1}{4n^{1/4}}$. From Lemma 4.3 of Bosq (2000), we have

$$|\widehat{\lambda}_k - \lambda_{k0}| \leq \| \frac{1}{n} \sum_{i=1}^n \widehat{\boldsymbol{\eta}}_i(s)\widehat{\boldsymbol{\eta}}_i(t)^T - E\left\{\boldsymbol{\eta}_{i0}(s)\boldsymbol{\eta}_{i0}(t)^T\right\} \|_2 = O_p(r_{\mathbf{G}}).$$

Then, $\mathrm{BIC}(k) = \sum_{k=1}^K \{\lambda_{k0} + O_p(r_{\mathbf{G}})\} - K \times \frac{\lambda_{10} + O_p(r_{\mathbf{G}})}{4n^{1/4}}$.

By the definition of $\widehat{K}_n = \arg\max_{K \leq K_{\max}} \mathrm{BIC}(k)$, it is sufficiently to prove

$$P\left\{\mathrm{BIC}(\widehat{K}_n) < \mathrm{BIC}(K_n)\right\} \to 1 \text{ if } \widehat{K}_n \neq K_n. \tag{S0.22}$$

We first consider the case of $\widehat{K}_n < K_n$,

$$\begin{aligned}
&\mathrm{BIC}(\widehat{K}_n) - \mathrm{BIC}(K_n) \\
&= -\sum_{k=\widehat{K}_n+1}^{K_n} \{\lambda_{k0} + O_p(r_{\mathbf{G}})\} - (\widehat{K}_n - K_n)\frac{\lambda_{10} + O_p(r_{\mathbf{G}})}{4n^{1/4}} \\
&= -\sum_{k=\widehat{K}_n+1}^{K_n} \lambda_{k0} + (\widehat{K}_n - K_n)O_p(r_{\mathbf{G}}) - (\widehat{K}_n - K_n)\frac{\lambda_{10} + O_p(r_{\mathbf{G}})}{4n^{1/4}} \\
&= -\sum_{k=\widehat{K}_n+1}^{K_n} \lambda_{k0} + o_p(1) < 0,
\end{aligned}$$

where the last equality holds from the eigenvalues $\lambda_{\widehat{K}_n+1,0}, \ldots, \lambda_{K_n,0}$ being positive and conditions $K_n \le K_{\max} = o(n^{1/4})$ and $r_{\mathbf{G}} = o(\frac{1}{n^{1/4}})$.

Next, we consider the case of $\widehat{K}_n > K_n$,

$$\begin{aligned}
&\mathrm{BIC}(\widehat{K}_n) - \mathrm{BIC}(K_n) \\
&= \sum_{k=K_n+1}^{\widehat{K}_n} \{\lambda_{k0} + O_p(r_{\mathbf{G}})\} - (\widehat{K}_n - K_n)\frac{\lambda_{10} + O_p(r_{\mathbf{G}})}{4n^{1/4}} \\
&= (\widehat{K}_n - K_n)O_p(r_{\mathbf{G}}) - (\widehat{K}_n - K_n)\frac{\lambda_{10} + O_p(r_{\mathbf{G}})}{4n^{1/4}},
\end{aligned}$$

where the term $-(K - K_n)\frac{\lambda_{10}}{4n^{1/4}}$ dominates the right hand side because $r_{\mathbf{G}} = o(\frac{1}{n^{1/4}})$. Hence, (S0.22) holds. This completes the proof.

Proof of proposition S.1

We first prove (i). The proof includes five steps as follows:

Step 1: Show that $F_1(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi})$ is a point-to-point mapping function, a special case of the point-to-set mapping. By the explicit iterative equations (S0.4), (S0.5), and (S0.7), we know that $F_1(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi})$ consists of the deterministic combinations and compositions of a series of functions of $(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi})$. Thus, a unique iterative value of $\left\{\mathbf{f}^{(o+1)}, \boldsymbol{\phi}^{(o+1)}, \boldsymbol{\xi}^{(o+1)}\right\}$ can be obtained. That is, there exists a unique $\left\{\mathbf{f}^{(o+1)}, \boldsymbol{\phi}^{(o+1)}, \boldsymbol{\xi}^{(o+1)}\right\}$ such that $\left\{\mathbf{f}^{(o+1)}, \boldsymbol{\phi}^{(o+1)}, \boldsymbol{\xi}^{(o+1)}\right\} = F_1\left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\}$. Therefore, $F_1(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi})$ is a point-to-point mapping function.

Step 2: Show that $l_1(\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi})$ does not increase with respect to the sequence $\left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\}$, i.e.

$$l_1\left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\} \leq l_1\left\{\mathbf{f}^{(o-1)}, \boldsymbol{\phi}^{(o-1)}, \boldsymbol{\xi}^{(o-1)}\right\}. \tag{S0.23}$$

By the principle of parameter update in the implemented algorithms, we obtain (S0.23).

Step 3: Show that if $\left\{\mathbf{f}^{(o-1)}, \boldsymbol{\phi}^{(o-1)}, \boldsymbol{\xi}^{(o-1)}\right\} \notin \mathcal{Q}_{1*}$, then $l_1\left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\} < l_1\left\{\mathbf{f}^{(o-1)}, \boldsymbol{\phi}^{(o-1)}, \boldsymbol{\xi}^{(o-1)}\right\}$ for $\left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\} = F_1\left\{\mathbf{f}^{(o-1)}, \boldsymbol{\phi}^{(o-1)}, \boldsymbol{\xi}^{(o-1)}\right\}$; and if $\left\{\mathbf{f}^{(o-1)}, \boldsymbol{\phi}^{(o-1)}, \boldsymbol{\xi}^{(o-1)}\right\} \in \mathcal{Q}_{1*}$, then $l_1\left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\} = l_1\left\{\mathbf{f}^{(o-1)}, \boldsymbol{\phi}^{(o-1)}, \boldsymbol{\xi}^{(o-1)}\right\}$. It is followed directly by the proof of Step 2 and the definition of $\mathcal{Q}_{1*}$.

Step 4. We show that $l_1\left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\}$ converges monotonically to $L_1^* = l_1(\mathbf{f}^*, \boldsymbol{\phi}^*, \boldsymbol{\xi}^*)$ for some $(\mathbf{f}^*, \boldsymbol{\phi}^*, \boldsymbol{\xi}^*)$.

By Assumption S1, for $\left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\} \subset \mathcal{Q}_{10}$, we can find a convergent subsequence $\left\{\mathbf{f}^{(o_i)}, \boldsymbol{\phi}^{(o_i)}, \boldsymbol{\xi}^{(o_i)}\right\}$ converging to the limit $(\mathbf{f}^*, \boldsymbol{\phi}^*, \boldsymbol{\xi}^*)$, where $\{o_i, o_i < o_{i+1}, i = 1, 2, \ldots\}$. Since $l_1\{\mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\xi}\}$ is continuous, it induces that

$$\lim_{i\to\infty} l_1\left\{\mathbf{f}^{(o_i)}, \boldsymbol{\phi}^{(o_i)}, \boldsymbol{\xi}^{(o_i)}\right\} = l_1(\mathbf{f}^*, \boldsymbol{\phi}^*, \boldsymbol{\xi}^*)$$

and

$$l_1(\mathbf{f}^*, \boldsymbol{\phi}^*, \boldsymbol{\xi}^*) \leq \lim_{i\to\infty} l_1 \left\{\mathbf{f}^{(o_i)}, \boldsymbol{\phi}^{(o_i)}, \boldsymbol{\xi}^{(o_i)}\right\},$$

with $l_1 \left\{\mathbf{f}^{(o_i)}, \boldsymbol{\phi}^{(o_i)}, \boldsymbol{\xi}^{(o_i)}\right\}$ being monotonic with respect to i. Moreover, we have

$$l_1 \left\{\mathbf{f}^{(o+1)}, \boldsymbol{\phi}^{(o+1)}, \boldsymbol{\xi}^{(o+1)}\right\} \leq l_1 \left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\}, o = 1, 2, \ldots.$$

Then, for $o = 1, 2, \ldots$, there exists an o_i satisfying $o < o_i$ and $l_1(\mathbf{f}^*, \boldsymbol{\phi}^*, \boldsymbol{\xi}^*) \leq l_1 \left\{\mathbf{f}^{(o_i)}, \boldsymbol{\phi}^{(o_i)}, \boldsymbol{\xi}^{(o_i)}\right\}$ $\leq l_1 \left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\}$. That is, $l_1 \left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\}$ is a monotonic and bounded sequence. Since its subsequence $\left[l_1 \left\{\mathbf{f}^{(o_i)}, \boldsymbol{\phi}^{(o_i)}, \boldsymbol{\xi}^{(o_i)}\right\}\right]$ converges to $l_1(\mathbf{f}^*, \boldsymbol{\phi}^*, \boldsymbol{\xi}^*)$, then

$$\lim_{o\to\infty} l_1 \left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\} = l_1(\mathbf{f}^*, \boldsymbol{\phi}^*, \boldsymbol{\xi}^*).$$

Step 5. Finally, we use contradiction approach to show $(\mathbf{f}^*, \boldsymbol{\phi}^*, \boldsymbol{\xi}^*) \in \mathcal{Q}_{1*}$.

Assume that $(\mathbf{f}^*, \boldsymbol{\phi}^*, \boldsymbol{\xi}^*)$ is not in $\mathcal{Q}_{1*}$. We investigate the subsequence $\left\{\mathbf{f}^{(o_i+1)}, \boldsymbol{\phi}^{(o_i+1)}, \boldsymbol{\xi}^{(o_i+1)}\right\}$, where o_i is same as the above one in Step 4. Since all members of this sequence are contained in a compact set, there is a convergent subsequence $\left[\left\{\mathbf{f}^{(o_{i_k}+1)}, \boldsymbol{\phi}^{(o_{i_k}+1)}, \boldsymbol{\xi}^{(o_{i_k}+1)}\right\}\right.$ $\left., i_1 < i_2 < \ldots\right]$ such that

$$\lim_{k\to\infty} \left\{\mathbf{f}^{(o_{i_k}+1)}, \boldsymbol{\phi}^{(o_{i_k}+1)}, \boldsymbol{\xi}^{(o_{i_k}+1)}\right\} = (\mathbf{f}^{**}, \boldsymbol{\phi}^{**}, \boldsymbol{\xi}^{**}).$$

Note that $\left\{\mathbf{f}^{(o_{i_k}+1)}, \boldsymbol{\phi}^{(o_{i_k}+1)}, \boldsymbol{\xi}^{(o_{i_k}+1)}\right\}$ is a subsequence of $\left\{\mathbf{f}^{(o_i+1)}, \boldsymbol{\phi}^{(o_i+1)}, \boldsymbol{\xi}^{(o_i+1)}\right\}$, we have

$$\lim_{k\to\infty} \left\{\mathbf{f}^{(o_{i_k})}, \boldsymbol{\phi}^{(o_{i_k})}, \boldsymbol{\xi}^{(o_{i_k})}\right\} = (\mathbf{f}^{**}, \boldsymbol{\phi}^{**}, \boldsymbol{\xi}^{**}).$$

Using the monotonicity of $l_1 \left\{\mathbf{f}^{(o)}, \boldsymbol{\phi}^{(o)}, \boldsymbol{\xi}^{(o)}\right\}$ with respect to o and $o_{i_k} < o_{i_k}+1 \leq o_{i_{k+1}}$, we obtain

$$l_1\left\{\mathbf{f}^{(o_{i_{k+1}})},\boldsymbol{\phi}^{(o_{i_{k+1}})},\boldsymbol{\xi}^{(o_{i_{k+1}})}\right\}\le l_1\left\{\mathbf{f}^{(o_{i_k}+1)},\boldsymbol{\phi}^{(o_{i_k}+1)},\boldsymbol{\xi}^{(o_{i_k}+1)}\right\}\le l_1\left\{\mathbf{f}^{(o_{i_k})},\boldsymbol{\phi}^{(o_{i_k})},\boldsymbol{\xi}^{(o_{i_k})}\right\},$$

which implies $l_1(\mathbf{f}^*,\boldsymbol{\phi}^*,\boldsymbol{\xi}^*)=l_1(\mathbf{f}^{**},\boldsymbol{\phi}^{**},\boldsymbol{\xi}^{**})$ by the continuity of $l_1(\mathbf{f},\boldsymbol{\phi},\boldsymbol{\xi})$.

Given Condition (S2) and the fact that $\left\{\mathbf{f}^{(o_{i_k}+1)},\boldsymbol{\phi}^{(o_{i_k}+1)},\boldsymbol{\xi}^{(o_{i_k}+1)}\right\}=F_1\left\{\mathbf{f}^{(o_{i_k})},\boldsymbol{\phi}^{(o_{i_k})},\boldsymbol{\xi}^{(o_{i_k})}\right\}$, we can conclude that $(\mathbf{f}^{**},\boldsymbol{\phi}^{**},\boldsymbol{\xi}^{**})=F_1(\mathbf{f}^*,\boldsymbol{\phi}^*,\boldsymbol{\xi}^*)$. However, if $(\mathbf{f}^*,\boldsymbol{\phi}^*,\boldsymbol{\xi}^*)$ is not in $\mathcal{Q}_{1*}$, then according to the results of Step 3, we have $l_1(\mathbf{f}^*,\boldsymbol{\phi}^*,\boldsymbol{\xi}^*)>l_1(\mathbf{f}^{**},\boldsymbol{\phi}^{**},\boldsymbol{\xi}^{**})$. This leads to a contraction, which implies that $(\mathbf{f}^*,\boldsymbol{\phi}^*,\boldsymbol{\xi}^*)$ must belong to $\mathcal{Q}_{1*}$.

This completes the proof of (i). Following the lines of the proof of (i), we can prove (ii).

Lemmas

Lemma 1. *Suppose Conditions (C1)–(C5) hold, we have*

$$\|\widehat{\mu}_j(t)-\mu_{j0}(t)\|_2=O_p\left(k_n^{-r}+\sqrt{\frac{k_n}{\sum_{i=1}^n n_{ij}}}+\frac{1}{\sqrt{n}}\right).$$

Proof: Based on the estimation equation of $\widehat{\mathbf{u}}_j$, we have

$$\begin{aligned}\widehat{\mu}_j(t)-\mu_{j0}(t)&=\mathbf{B}_n(t)^T\widehat{\mathbf{u}}_j-\mu_{j0}(t)\\&=\mathbf{B}_n(t)^T\left\{\sum_{i=1}^n\sum_{d=1}^{n_{ij}}\mathbf{B}_n(t_{ijd})\mathbf{B}_n(t_{ijd})^T\right\}^{-1}\sum_{i=1}^n\sum_{d=1}^{n_{ij}}\Bigg[\mu_{j0}(t_{ijd})\mathbf{B}_n(t_{ijd})+\varepsilon_{ij}(t_{ijd})\mathbf{B}_n(t_{ijd})\\&\quad+f_{j0}\left\{\sum_{k=1}^{\infty}\xi_{ik0}\phi_{jk0}(t_{ijd})\right\}\mathbf{B}_n(t_{ijd})\Bigg]-\mu_{j0}(t).\end{aligned}\qquad(S0.24)$$

Next, we derive the order of right side of (S0.24). By Lemma 6.2 of Cardot (2000), for any $\mu(t) \in \mathcal{H}_r$ we have $\|\frac{1}{\sum_{i=1}^n n_{ij}} \sum_{i=1}^n \sum_{d=1}^{n_{ij}} \mu(t_{ijd})\mathbf{B}_n(t_{ijd}) - \int_0^1 \mu(t)\mathbf{B}_n(t)dt\| = O_p(\frac{\sqrt{k_n}}{\sum_{i=1}^n n_{ij}})$. Then, by the orthogonality of spline basis functions we obtain

$$\left\{\sum_{i=1}^n \sum_{d=1}^{n_{ij}} \mathbf{B}_n(t_{ijd})\mathbf{B}_n(t_{ijd})^T\right\}^{-1} = \frac{1}{\sum_{i=1}^n n_{ij}}\mathbf{I}_{k_n} + O_p\left\{\frac{k_n}{(\sum_{i=1}^n n_{ij})^2}\right\}. \quad \text{(S0.25)}$$

By equation (S0.25), the finite variance of $X_{ij}(t)$, and the independence of samples, we have

$$\begin{aligned}
&E\|\mathbf{B}_n(t)^T \left\{\sum_{i=1}^n \sum_{d=1}^{n_{ij}} \mathbf{B}_n(t_{ijd})\mathbf{B}_n(t_{ijd})^T\right\}^{-1} \sum_{i=1}^n \sum_{d=1}^{n_{ij}} f_{j0}\left\{\sum_{k=1}^\infty \xi_{ik0}\phi_{jk0}(t_{ijd})\right\}\mathbf{B}_n(t_{ijd})\|_2^2 \\
&\leq M \cdot \frac{1}{(\sum_{i=1}^n n_{ij})^2} \sum_{i=1}^n E\|\sum_{d=1}^{n_{ij}} f_{j0}\left\{\sum_{k=1}^\infty \xi_{ik0}\phi_{jk0}(t_{ijd})\right\}\mathbf{B}_n(t_{ijd})\|_2^2 \\
&\leq M \cdot \frac{1}{(\sum_{i=1}^n n_{ij})^2} \sum_{i=1}^n E\left[\sum_{d=1}^{n_{ij}} f_{j0}\left\{\sum_{k=1}^\infty \xi_{ik0}\phi_{jk0}(t_{ijd})\right\}^2\right] \cdot \lambda_{\max}\left\{\sum_{d=1}^{n_{ij}} \mathbf{B}_n(t_{ijd})\mathbf{B}_n(t_{ijd})^T\right\} \\
&= O_p(\frac{1}{n}), \qquad \text{(S0.26)}
\end{aligned}$$

where $\lambda_{\max}(\mathbf{A})$ is the largest eigenvalue of square matrix $\mathbf{A}$. By Lemma 9 in Stone (1985), we have

$$\|\mathbf{B}_n(t)^T \left\{\sum_{i=1}^n \sum_{d=1}^{n_{ij}} \mathbf{B}_n(t_{ijd})\mathbf{B}_n(t_{ijd})^T\right\}^{-1} \sum_{i=1}^n \sum_{d=1}^{n_{ij}} \mu_{j0}(t_{ijd})\mathbf{B}_n(t_{ijd}) - \mu_{j0}(t)\|_2 = k_n^{-r}. \quad \text{(S0.27)}$$

Moreover, by the independence of $\varepsilon_{ij}(t)$ over i, j, and d and the orthogonality of spline basis functions, we have

$$\|\mathbf{B}_n(t)^T \left\{\sum_{i=1}^n \sum_{d=1}^{n_{ij}} \mathbf{B}_n(t_{ijd})\mathbf{B}_n(t_{ijd})^T\right\}^{-1} \sum_{i=1}^n \sum_{d=1}^{n_{ij}} \varepsilon_{ij}(t_{ijd})\mathbf{B}_n(t_{ijd})\|_2 = O_p(\sqrt{\frac{k_n}{\sum_{i=1}^n n_{ij}}} + \frac{1}{\sqrt{n}}). \quad \text{(S0.28)}$$

Combining (S0.24), (S0.26), (S0.27) and (S0.28) yields the desired results.

Lemma 2. *Under Conditions (C1)–(C4), for any* $\boldsymbol{\xi}_i^T \boldsymbol{\Gamma}_j$ *in the parameter space of parameter* $\left(\boldsymbol{\xi}_i^T \boldsymbol{\Gamma}_j, i = 1, \ldots, j = 1, \ldots, p\right)$, *we have*

$$E\left[\left\{\widehat{f}_j(\boldsymbol{\xi}_i^T \boldsymbol{\Gamma}_j \mathbf{B}_n(t)) - f_{j0}(\boldsymbol{\xi}_i^T \boldsymbol{\Gamma}_j \mathbf{B}_n(t))\right\}^2\right]$$
$$= O_p\left(h_1^4 + \frac{1}{\sum_{i=1}^n n_{ij} h_1} + k_n^{-2r} + \frac{k_n}{\sum_{i=1}^n n_{ij}} + \frac{1}{n}\right), \qquad \text{(S0.29)}$$

and for any $\boldsymbol{\alpha}$ *in the parameter space of parameter* $\boldsymbol{\alpha}$,

$$E\left[\left\{\widehat{\psi}(\sum_{k=1}^{K_n} \boldsymbol{\alpha}_k \mathbf{S}_n(\mathbf{s}) \widehat{\zeta}_{ik}) - \psi_0(\sum_{k=1}^{K_n} \boldsymbol{\alpha}_k \mathbf{S}_n(\mathbf{s}) \widehat{\zeta}_{ik})\right\}^2\right]$$
$$= O_p\left(K_n |\triangle|^{2(\varpi+1)} + h_2^4 + \frac{1}{n N_s h_2^q} + \frac{1}{n}\right). \qquad \text{(S0.30)}$$

Proof: Let

$$\widetilde{\mathcal{K}}_{ijd}(u, h_1) = (1, 0)\left[\sum_{i=1}^n \sum_{d=1}^{n_{ij}} \mathcal{K}_{h_1}\left\{\boldsymbol{\xi}_i^T \boldsymbol{\phi}_j(t_{ijd}) - u\right\} \mathbf{Z}_{ijd}^f(u) \mathbf{Z}_{ijd}^f(u)^T\right]^{-1} \mathcal{K}_{h_1}\left\{\boldsymbol{\xi}_i^T \boldsymbol{\phi}_j(t_{ijd}) - u\right\} \mathbf{Z}_{ijd}^f(u).$$

By Condition (C5) and the Corollary 6.21 of Schumacker (1981), there exists $\boldsymbol{\Gamma}_{j0}$ such that

$$\sup_{t \in [0,1]} E|\boldsymbol{\xi}_i^T \boldsymbol{\Gamma}_{j0} \mathbf{B}_n(t) - \eta_{ij}(t)| = O_p\left(k_n^{-r}\right). \qquad \text{(S0.31)}$$

From the estimation equation of $\widehat{f}_j(\cdot)$ and (S0.31), we have

$$\begin{aligned}
&\widehat{f}_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t)\right\} - f_{j0}\left\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t)\right\} \qquad \text{(S0.32)}\\
&= \sum_{i=1}^{n}\sum_{d=1}^{n_{ij}} \widetilde{\mathcal{K}}_{ijd}\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t), h_1\}\left\{X_{ij}(t_{ijd}) - \widehat{\mu}_j(t_{ijd})\right\} - f_{j0}\left\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t)\right\}\\
&= \sum_{i=1}^{n}\sum_{d=1}^{n_{ij}} \widetilde{\mathcal{K}}_{ijd}\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t), h_1\}\left\{\mu_{j0}(t_{ijd}) - \widehat{\mu}_j(t_{ijd})\right\}\\
&+ \sum_{i=1}^{n}\sum_{d=1}^{n_{ij}} \widetilde{\mathcal{K}}_{ijd}\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t), h_1\} f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\} - f_{j0}\left\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t)\right\} + O_p(k_n^{-r})\\
&+ \sum_{i=1}^{n}\sum_{d=1}^{n_{ij}} \widetilde{\mathcal{K}}_{ijd}\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t), h_1\}\varepsilon_{ij}(t_{ijd}).
\end{aligned}$$

By Lemma 1, we have

$$\|\sum_{i=1}^{n}\sum_{d=1}^{n_{ij}} \widetilde{\mathcal{K}}_{ijd}\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t), h_1\}\left\{\mu_{j0}(t_{ijd}) - \widehat{\mu}_j(t_{ijd})\right\}\|_2 = O_p\left(k_n^{-r} + \sqrt{\frac{k_n}{\sum_{i=1}^n n_{ij}}} + \frac{1}{\sqrt{n}}\right). \qquad \text{(S0.33)}$$

By Condition (C6), we have

$$E\|\sum_{i=1}^{n}\sum_{d=1}^{n_{ij}} \widetilde{\mathcal{K}}_{ijd}\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t), h_1\}\varepsilon_{ij}(t_{ijd})\|_2 = O_p\left(\frac{1}{\sqrt{\sum_{i=1}^n n_{ij}h_1}} + \frac{1}{\sqrt{n}}\right). \qquad \text{(S0.34)}$$

Because the link function f_{j0} is differentiable in the neighborhood of $|\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd}) - \boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t_{ijd})| < h_1$, it follows from the Taylor's expansion and Conditions (C2) and (C3) that

$$\begin{aligned}
&E\|\sum_{i=1}^{n}\sum_{d=1}^{n_{ij}} \widetilde{\mathcal{K}}_{ijd}\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t), h_1\} f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\} - f_{j0}\left\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t)\right\}\|_2\\
&= E\|0.5\int u^2\mathcal{K}(u)du \cdot \nabla^2 f_{j0}\left\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t)\right\} h_1^2[1 + o_p(1)]\|_2\\
&= O_p(h_1^2). \qquad \text{(S0.35)}
\end{aligned}$$

Combining (S0.32), (S0.33), (S0.34), and (S0.35) yields (S0.29).

Following a similar proof strategy as in (S0.29), we can establish (S0.30).

Lemma 3. *For any vectors* $\mathbf{A}_1, \mathbf{A}_2$*, and positive constant* M *satisfying* $\mathbf{A}_1^T\mathbf{A}_1 - 2\mathbf{A}_1^T\mathbf{A}_2 \leq M$*, we have* $\|\mathbf{A}_1\|^2 \leq 4\|\mathbf{A}_2\|^2 + 2M$*.*

Proof: By $\mathbf{A}_1^T\mathbf{A}_1 - 2\mathbf{A}_1^T\mathbf{A}_2 \leq M$, we have $2\|\mathbf{A}_1 - \mathbf{A}_2\|^2 \leq 2\|\mathbf{A}_2\|^2 + 2M$ and $\|\mathbf{A}_1\|^2 \leq 2\|\mathbf{A}_1 - \mathbf{A}_2\|^2 + 2\|\mathbf{A}_2\|^2$, thus $\|\mathbf{A}_1\|^2 \leq 4\|\mathbf{A}_2\|^2 + 2M$.

Lemma 4. *Under Condition (C8), there exist positive finite constants* M *and* M_1*, such that*

$$M\|\boldsymbol{\alpha}_k\|^2 \leq \|\boldsymbol{\alpha}_k\mathbf{S}_n(\mathbf{s})\|_2^2 \leq C_1\|\boldsymbol{\alpha}_k\|^2.$$

Proof: Let $\mathbf{S}_n(\mathbf{s})^*$ be the unstandardized spline basis. From Lai and Schumaker (2007), there exist positive finite constants M and C_1, such that

$$M|\triangle|^2\|\boldsymbol{\alpha}_k\|^2 \leq \|\boldsymbol{\alpha}_k\mathbf{S}_n(\mathbf{s})^*\|_2^2 \leq C_1|\triangle|^2\|\boldsymbol{\alpha}_k\|^2. \tag{S0.36}$$

Then, combining $\mathbf{S}_n(\mathbf{s}) = \mathbf{S}_n(\mathbf{s})^*/\|\mathbf{S}_n(\mathbf{s})^*\|_2, \|\mathbf{S}_n(\mathbf{s})^*\|_2 = |\triangle|$ and (S0.36) yields the results of Lemma 4.

Proofs of Theorems

Proof of Theorem 1

Part (a) This proof consists of two steps. The first step proves the consistency of $\widehat{\boldsymbol{\eta}}_i(t)$ to $\boldsymbol{\eta}_{i0}(t)$. The second step derives the convergence rate of $\frac{1}{n}\sum_{i=1}^n \widehat{\boldsymbol{\eta}}_i(s)\widehat{\boldsymbol{\eta}}_i(t)^T$ to

$E\left\{\boldsymbol{\eta}_{i0}(s)\boldsymbol{\eta}_{i0}(t)^T\right\}$.

Step 1: Under Conditions in Theorem 1, the iterative estimators of $\boldsymbol{\xi}_i$ and $\left(\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_p\right)$ are asymptotic equivalent to the ones from MFPCA on $\boldsymbol{\eta}_i(t)$ which is solution to

$$\min \frac{1}{\sum_{i=1}^n \sum_{j=1}^p n_{ij}} \sum_{i=1}^n \sum_{j=1}^p \sum_{d=1}^{n_{ij}} \left[X_{ij}(t_{ijd}) - \widehat{\mu}_j(t_{ijd}) - \widehat{f}_j\left\{\eta_{ij}(t_{ijd})\right\}\right]^2. \tag{S0.37}$$

Based on the spline approximation $\boldsymbol{\phi}_j(t) \approx \boldsymbol{\phi}_{nj}(t) = \boldsymbol{\Gamma}_j \mathbf{B}_n(t)$, we have

$$\boldsymbol{\eta}_i(t) \approx \left(\boldsymbol{\Gamma}_1^T\boldsymbol{\xi}_i,\ldots,\boldsymbol{\Gamma}_p^T\boldsymbol{\xi}_i\right)^T \mathbf{B}_n(t).$$

Then, from (S0.37), $\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j$ is solution to

$$\min \frac{1}{n_{ij}} \sum_{d=1}^{n_{ij}} \left[X_{ij}(t_{ijd}) - \widehat{\mu}_j(t_{ijd}) - \widehat{f}_j\left\{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j\mathbf{B}_n(t_{ijd})\right\}\right]^2,$$

which implies

$$\frac{1}{n_{ij}} \sum_{d=1}^{n_{ij}} (X_{ij}(t_{ijd}) - \widehat{\mu}_j(t_{ijd}) - \widehat{f}_j(\widehat{\boldsymbol{\xi}}_i^T\widehat{\boldsymbol{\Gamma}}_j\mathbf{B}_n(t_{ijd})))\nabla_{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j}\widehat{f}_j(\widehat{\boldsymbol{\xi}}_i^T\widehat{\boldsymbol{\Gamma}}_j\mathbf{B}_n(t_{ijd}))\mathbf{B}_n(t_{ijd}) = \mathbf{0}.$$

Thus, we have

$$\begin{aligned} &E\left(\left[X_{ij}(t_{ijd}) - \mu_{j0}(t_{ijd}) - f_{j0}\left\{\widehat{\boldsymbol{\xi}}_i^T\widehat{\boldsymbol{\Gamma}}_j\mathbf{B}_n(t_{ijd})\right\}\right]\nabla_{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j} f_{j0}\left\{\widehat{\boldsymbol{\xi}}_i^T\widehat{\boldsymbol{\Gamma}}_j\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t_{ijd})\right) \\ &= o_p(1), \end{aligned}$$

due to Lemmas 1 and 2 and the uniform consistency of sample mean. Hence,

$$\begin{aligned} &E\left(\left[X_{ij}(t_{ijd}) - \mu_{j0}(t_{ijd}) - f_{j0}\left\{\widehat{\boldsymbol{\xi}}_i^T\widehat{\boldsymbol{\Gamma}}_j\mathbf{B}_n(t_{ijd})\right\}\right]\nabla_{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j} f_{j0}\left\{\widehat{\boldsymbol{\xi}}_i^T\widehat{\boldsymbol{\Gamma}}_j\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t_{ijd})\right) \\ &- E\left(\left[X_{ij}(t_{ijd}) - \mu_{j0}(t_{ijd}) - f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\right]\nabla_{\boldsymbol{\xi}_i^T\boldsymbol{\Gamma}_j} f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t_{ijd})\right) \\ &= o_p(1). \end{aligned}$$

Consequently, by Condition (C3) and continuous mapping theorem, we have

$$\widehat{\boldsymbol{\xi}}_i^T \widehat{\boldsymbol{\Gamma}}_j - \boldsymbol{\xi}_{i0}^T \boldsymbol{\Gamma}_{j0} = o_p(1). \tag{S0.38}$$

Combining (S0.31) and (S0.38), we get

$$\|\widehat{\eta}_{ij}(t) - \eta_{ij,0}(t)\|_2 = o_p(1). \tag{S0.39}$$

Step 2: From the consistency of (S0.39) and a Taylor expansion, we have

$$\begin{aligned}
\mathbf{0} = \quad & \frac{1}{n_{ij}} \sum_{d=1}^{n_{ij}} \left[X_{ij}(t_{ijd}) - \widehat{\mu}_j(t_{ijd}) - \widehat{f}_j \left\{ \widehat{\boldsymbol{\xi}}_i^T \widehat{\boldsymbol{\Gamma}}_j \mathbf{B}_n(t_{ijd}) \right\} \right] \\
& \times \nabla \widehat{f}_j \left\{ \widehat{\boldsymbol{\xi}}_i^T \widehat{\boldsymbol{\Gamma}}_j \mathbf{B}_n(t_{ijd}) \right\} \mathbf{B}_n(t_{ijd}) \\
= \quad & \mathbf{R}_1 + \mathbf{R}_2 \left(\widehat{\boldsymbol{\xi}}_i^T \widehat{\boldsymbol{\Gamma}}_j - \boldsymbol{\xi}_{i0}^T \boldsymbol{\Gamma}_{j0} \right)^T,
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{R}_1 \equiv & \frac{1}{n_{ij}} \sum_{d=1}^{n_{ij}} \left[X_{ij}(t_{ijd}) - \widehat{\mu}_j(t_{ijd}) - \widehat{f}_j \left\{ \boldsymbol{\xi}_{i0}^T \boldsymbol{\Gamma}_{j0} \mathbf{B}_n(t_{ijd}) \right\} \right] \\
& \times \nabla \widehat{f}_j \left\{ \boldsymbol{\xi}_{i0}^T \boldsymbol{\Gamma}_{j0} \mathbf{B}_n(t_{ijd}) \right\} \mathbf{B}_n(t_{ijd}), \\
\mathbf{R}_2 \equiv & \frac{1}{n_{ij}} \sum_{d=1}^{n_{ij}} \nabla_{\boldsymbol{\xi}_i^T \boldsymbol{\Gamma}_j} \left[\left\{ X_{ij}(t_{ijd}) - \widehat{\mu}_j(t_{ijd}) - \widehat{f}_j(\boldsymbol{\xi}_i^T \boldsymbol{\Gamma}_j \mathbf{B}_n(t_{ijd})) \right\} \right. \\
& \left. \times \ \nabla \widehat{f}_j \left\{ \boldsymbol{\xi}_i^T \boldsymbol{\Gamma}_j \mathbf{B}_n(t_{ijd}) \right\} \mathbf{B}_n(t_{ijd}) \right] \Big|_{\boldsymbol{\xi}_i^T \boldsymbol{\Gamma}_j = \left(\boldsymbol{\xi}_i^T \boldsymbol{\Gamma}_j \right)^*},
\end{aligned}$$

and $\left(\boldsymbol{\xi}_i^T \boldsymbol{\Gamma}_j \right)^*$ lies on the line that connects $\boldsymbol{\xi}_{i0}^T \boldsymbol{\Gamma}_{j0}$ and $\widehat{\boldsymbol{\xi}}_i^T \widehat{\boldsymbol{\Gamma}}_j$. Thus, we have

$$\widehat{\eta}_{ij}(t) = \eta_{ij,n}(t) - \mathbf{B}_n(t)^T \mathbf{R}_2^{-1} \mathbf{R}_1, \tag{S0.40}$$

where $\eta_{ij,n}(t) = \boldsymbol{\xi}_{i0}^T \boldsymbol{\Gamma}_{j0} \mathbf{B}_n(t)$.

Since $E\left(\left[X_{ij}(t)-\mu_{j0}(t)-f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t)\right\}\right]\nabla f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t)\right\}\mathbf{B}_n(t)\right)=\mathbf{0}$, by some calculations, we have

$$
\begin{aligned}
&\|\widehat{\eta}_{ij}(t)-\eta_{ij,0}(t)\|_2^2=\|\eta_{ij,n}(t)-\eta_{ij,0}(t)-\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{R}_1\|_2^2\\
&\le 2\|\eta_{ij,n}(t)-\eta_{ij,0}(t)\|_2^2+2\|\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{R}_1\|_2^2\\
&\le 4\|\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\frac{1}{n_{ij}}\sum_{d=1}^{n_{ij}}\left[X_{ij}(t_{ijd})-\mu_{j0}(t_{ijd})-f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\right]\\
&\quad\cdot\nabla\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t_{ijd})\|_2^2\\
&\quad+4\|\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\frac{1}{n_{ij}}\sum_{d=1}^{n_{ij}}\left[\mu_{j0}(t_{ijd})-\widehat{\mu}_j(t_{ijd})+f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\right.\\
&\quad\left.-\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\right]\nabla\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t_{ijd})\|_2^2+2\|\eta_{ij,n}(t)-\eta_{ij,0}(t)\|_2^2\\
&\le\frac{4}{n_{ij}}\sum_{d=1}^{n_{ij}}\|\mu_{j0}(t_{ijd})-\widehat{\mu}_j(t_{ijd})+f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}-\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\|_2^2\\
&\quad\times\frac{1}{n_{ij}}\sum_{d=1}^{n_{ij}}\|\nabla\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{B}_n(t_{ijd})\|_2^2+O_p\left(\frac{k_n}{n_{ij}}+k_n^{-2r}\right)\\
&=O_p\left(\|\mu_{j0}(t)-\widehat{\mu}_j(t)\|_2^2+\frac{1}{n_{ij}}\sum_{d=1}^{n_{ij}}\left[f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}-\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\right]^2\right)\\
&\quad+O_p\left(\frac{k_n}{n_{ij}}+k_n^{-2r}\right)\\
&=O_p\left(\frac{k_n}{n_{ij}}+k_n^{-2r}+h_1^4+\frac{1}{\sum_{i=1}^n h_1 n_{ij}}+\frac{1}{n}\right).
\end{aligned}
$$

The third equality holds from the law of large number and (S0.31), and the last equality holds from the results of Lemmas 1 and 2. Hence, by $\|\widehat{\boldsymbol{\eta}}_i(t)-\boldsymbol{\eta}_{i0}(t)\|_2^2=\sum_{j=1}^p\|\widehat{\eta}_{ij}(t)-\eta_{ij,0}(t)\|_2^2$ yields the desired results.

Part (b) From (S0.40), we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\widehat{\eta}_{ij}(s)\widehat{\eta}_{ij}(t)\\
&=\frac{1}{n}\sum_{i=1}^{n}\eta_{ij,0}(s)\eta_{ij,0}(t)+\frac{1}{n}\sum_{i=1}^{n}\eta_{ij,0}(s)\left\{\eta_{ij,n}(t)-\eta_{ij,0}(t)-\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{R}_1\right\}\\
&+\frac{1}{n}\sum_{i=1}^{n}\eta_{ij,0}(t)\left\{\eta_{ij,n}(s)-\eta_{ij,0}(s)-\mathbf{B}_n(s)^T\mathbf{R}_2^{-1}\mathbf{R}_1\right\} \qquad \text{(S0.41)}\\
&+\frac{1}{n}\sum_{i=1}^{n}\left\{\eta_{ij,n}(s)-\eta_{ij,0}(s)-\mathbf{B}_n(s)^T\mathbf{R}_2^{-1}\mathbf{R}_1\right\}\left\{\eta_{ij,n}(t)-\eta_{ij,0}(t)-\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{R}_1\right\}.
\end{aligned}$$

In sequel, we will show the convergence of $\frac{1}{n}\sum_{i=1}^{n}\eta_{ij,0}(s)\eta_{ij,0}(t)$ to $E\{\eta_{ij,0}(s)\eta_{ij,0}(t)\}$ and the last three terms on the right hand side of (S0.41) converge to zero in probability.

We first show that

$$\|\frac{1}{n}\sum_{i=1}^{n}\eta_{ij,0}(s)\eta_{ij,0}(t)-E\left\{\eta_{ij,0}(s)\eta_{ij,0}(t)\right\}\|_2=O_p\left(n^{-1/2}\right). \qquad \text{(S0.42)}$$

By the independence of samples and Condition (C3), it holds

$$\begin{aligned}
&E\left(\int_{\mathcal{T}}\int_{\mathcal{T}}\left[\frac{1}{n}\sum_{i=1}^{n}\eta_{ij,0}(s)\eta_{ij,0}(t)-E\left\{\eta_{ij,0}(s)\eta_{ij,0}(t)\right\}\right]^2 dsdt\right)\\
&=E\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\int_{\mathcal{T}}\int_{\mathcal{T}}\left[\eta_{ij,0}(s)\eta_{ij,0}(t)-E\left\{\eta_{ij,0}(s)\eta_{ij,0}(t)\right\}\right]\right.\\
&\qquad\times\left[\eta_{i'j,0}(s)\eta_{i'j,0}(t)-E\left\{\eta_{i'j,0}(s)\eta_{i'j,0}(t)\right\}\right]dsdt)\\
&=\frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{T}}\int_{\mathcal{T}}E\left(\left[\eta_{ij,0}(s)\eta_{ij,0}(t)-E\left\{\eta_{ij,0}(s)\eta_{ij,0}(t)\right\}\right]^2\right)dsdt\\
&=\frac{1}{n}\int_{\mathcal{T}}\int_{\mathcal{T}}\mathrm{var}\left\{\eta_{ij,0}(s)\eta_{ij,0}(t)\right\}dsdt=O_p\left(n^{-1}\right).
\end{aligned}$$

Thus, it yields (S0.42).

We next show that

$$\|\frac{1}{n}\sum_{i=1}^{n}\eta_{ij,0}(s)\left\{\eta_{ij,n}(t)-\eta_{ij,0}(t)-\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{R}_1\right\}\|_2 \tag{S0.43}$$
$$=O_p\left(k_n^{-r}+h_1^2+\sqrt{\frac{k_n+h_1^{-1}}{\sum_{i=1}^{n}n_{ij}}}+\frac{1}{\sqrt{n}}\right).$$

Using the fact that

$$\|\frac{1}{n}\sum_{i=1}^{n}\eta_{ij,0}(s)\left\{\eta_{ij,n}(t)-\eta_{ij,0}(t)-\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{R}_1\right\}\|_2^2 \tag{S0.44}$$
$$\leq 3\|\frac{1}{n}\sum_{i=1}^{n}\eta_{ij,0}(s)\left\{\eta_{ij,n}(t)-\eta_{ij,0}(t)\right\}\|_2^2$$
$$+3\|\sum_{i=1}^{n}\frac{1}{nn_{ij}}\sum_{d=1}^{n_{ij}}\eta_{ij,0}(s)\varepsilon_{ij}(t_{ijd})\nabla\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{B}_n(t_{ijd})\|_2^2$$
$$+3\|\sum_{i=1}^{n}\frac{1}{nn_{ij}}\sum_{d=1}^{n_{ij}}\eta_{ij,0}(s)\left[\mu_{j0}(t_{ijd})-\widehat{\mu}_j(t_{ijd})+f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\right.$$
$$\left.-\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\right]$$
$$\nabla\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{B}_n(t_{ijd})\|_2^2,$$

it is sufficient to focus on the three terms on the right-hand side of (S0.44). By the Cauchy-Schwarz inequality, we have

$$\|\frac{1}{n}\sum_{i=1}^{n}\eta_{ij,0}(s)\left\{\eta_{ij,n}(t)-\eta_{ij,0}(t)\right\}\|_2^2$$
$$\leq\frac{1}{n}\sum_{i=1}^{n}\|\eta_{ij,0}(s)\|_2^2\frac{1}{n}\sum_{i=1}^{n}\|\eta_{ij,n}(t)-\eta_{ij,0}(t)\|_2^2=O_p\left(k_n^{-2r}\right).$$

By the consistency of kernel estimators and spline estimators, and $\|\widehat{\eta}_{ij}(t)-\eta_{ij,0}(t)\|_2=$

$o_p(1)$, we have

$$
\begin{aligned}
&E\|\sum_{i=1}^{n}\frac{1}{nn_{ij}}\sum_{d=1}^{n_{ij}}\eta_{ij,0}(s)\varepsilon_{ij}(t_{ijd})\nabla\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{B}_n(t_{ijd})\|_2^2\\
&\leq\frac{1}{n^2}\sum_{i=1}^{n}E\left[\frac{1}{n_{ij}}\sum_{d=1}^{n_{ij}}\|\eta_{ij,0}(s)\varepsilon_{ij}(t_{ijd})\|_2^2\frac{1}{n_{ij}}\sum_{d=1}^{n_{ij}}\|\nabla\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{B}_n(t_{ijd})\|_2^2\right]\\
&=O_p\left(1/n\right).
\end{aligned}
$$

Further, from Lemmas 1 and 2, we have

$$
\begin{aligned}
&\|\sum_{i=1}^{n}\frac{1}{nn_{ij}}\sum_{d=1}^{n_{ij}}\eta_{ij,0}(s)\left[\mu_{j0}(t_{ijd})-\widehat{\mu}_j(t_{ijd})+f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}-\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\right]\\
&\qquad\times\nabla\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{B}_n(t_{ijd})\|_2^2\\
&\leq\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n_{ij}}\sum_{d=1}^{n_{ij}}\|\mu_{j0}(t_{ijd})-\widehat{\mu}_j(t_{ijd})+f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}-\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\|_2^2\right.\\
&\qquad\left.\times\frac{1}{n_{ij}}\sum_{d=1}^{n_{ij}}\|\nabla\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{B}_n(t_{ijd})\|_2^2\right]\times\frac{1}{n}\sum_{i=1}^{n}\|\eta_{ij,0}(s)\|_2^2\\
&\leq\left(\|\mu_{j0}(t)-\widehat{\mu}_j(t)\|_2^2+\sum_{i=1}^{n}\sum_{d=1}^{n_{ij}}\frac{1}{nn_{ij}}\left[f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}-\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\right]^2\right)\\
&=O_p\left(k_n^{-2r}+h_1^4+\frac{k_n+h_1^{-1}}{\sum_{i=1}^{n}n_{ij}}+\frac{1}{n}\right).\qquad\text{(S0.45)}
\end{aligned}
$$

Similarly, we have that

$$
\begin{aligned}
&\|\frac{1}{n}\sum_{i=1}^{n}\eta_{ij,0}(t)\left\{\eta_{ij,n}(s)-\eta_{ij,0}(s)-\mathbf{B}_n(s)^T\mathbf{R}_2^{-1}\mathbf{R}_1\right\}\|_2\qquad\text{(S0.46)}\\
&=O_p\left(k_n^{-r}+h_1^2+\sqrt{\frac{k_n+h_1^{-1}}{\sum_{i=1}^{n}n_{ij}}}+\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

Finally, we can show that

$$\begin{aligned}&\|\frac{1}{n}\sum_{i=1}^{n}\left\{\eta_{ij,n}(s)-\eta_{ij,0}(s)-\mathbf{B}_n(s)^T\mathbf{R}_2^{-1}\mathbf{R}_1\right\}\left\{\eta_{ij,n}(t)-\eta_{ij,0}(t)-\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{R}_1\right\}\|_2\\&=O_p\left(\sum_{i=1}^{n}\frac{k_n}{nn_{ij}}+k_n^{-2r}+h_1^4+\frac{1}{\sum_{i=1}^{n}h_1n_{ij}}+\frac{1}{n}\right).\end{aligned}\tag{S0.47}$$

By the Cauchy-Schwarz inequality and the results of part (a), we have

$$\begin{aligned}&\|\frac{1}{n}\sum_{i=1}^{n}\left\{\eta_{ij,n}(s)-\eta_{ij,0}(s)-\mathbf{B}_n(s)^T\mathbf{R}_2^{-1}\mathbf{R}_1\right\}\left\{\eta_{ij,n}(t)-\eta_{ij,0}(t)-\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{R}_1\right\}\|_2\\&\leq\frac{1}{n}\sum_{i=1}^{n}\left(\|\eta_{ij,n}(t)-\eta_{ij,0}(t)\|_2^2+\|\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\mathbf{R}_1\|_2^2\right)\\&\leq\frac{2}{n}\sum_{i=1}^{n}\|\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\frac{1}{n_{ij}}\sum_{d=1}^{n_{ij}}\left[X_{ij}(t_{ijd})-\mu_{j0}(t_{ijd})-f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\right]\\&\quad\times\nabla\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t_{ijd})\|_2^2\\&\quad+\frac{2}{n}\sum_{i=1}^{n}\|\mathbf{B}_n(t)^T\mathbf{R}_2^{-1}\frac{1}{n_{ij}}\sum_{d=1}^{n_{ij}}\left[\mu_{j0}(t_{ijd})-\widehat{\mu}_j(t_{ijd})+f_{j0}\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}-\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\right]\\&\quad\nabla\widehat{f}_j\left\{\boldsymbol{\xi}_{i0}^T\boldsymbol{\Gamma}_{j0}\mathbf{B}_n(t_{ijd})\right\}\mathbf{B}_n(t_{ijd})\|_2^2+O_p\left(k_n^{-2r}\right)\\&=O_p\left(\sum_{i=1}^{n}\frac{k_n}{nn_{ij}}+k_n^{-2r}+h_1^4+\frac{1}{\sum_{i=1}^{n}h_1n_{ij}}+\frac{1}{n}\right).\end{aligned}$$

The arguments for (S0.41), (S0.42), (S0.43), (S0.46), and (S0.47) hold for $\frac{1}{n}\sum_{i=1}^{n}\widehat{\eta}_{ij}(s)\widehat{\eta}_{ij'}(t)^T$ for any $j\neq j'$. Thus, combining (S0.41), (S0.42), (S0.43), (S0.46), and (S0.47) leads to

$$\begin{aligned}&\|\frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{\eta}}_i(s)\widehat{\boldsymbol{\eta}}_i(t)^T-E\left\{\boldsymbol{\eta}_{i0}(s)\boldsymbol{\eta}_{i0}(t)^T\right\}\|_2\\&=O_p\left(\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{p}\frac{k_n}{nn_{ij}}}+k_n^{-r}+h_1^2+\sqrt{\sum_{j=1}^{p}\frac{1}{\sum_{i=1}^{n}h_1n_{ij}}}+\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Part (c) From Lemma 4.3 of Bosq (2000), we have

$$
\begin{aligned}
&\widehat{\xi}_{ik} - \xi_{ik,0} \\
&= O_p(\|\widehat{\boldsymbol{\eta}}_i(t) - \boldsymbol{\eta}_{i0}(t)\|_2 + \delta_k^{-1}\|\boldsymbol{\eta}_{i0}(t)\|_2 \| \frac{1}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{\eta}}_i(s)\widehat{\boldsymbol{\eta}}_i(t)^T - E\left\{\boldsymbol{\eta}_{i0}(s)\boldsymbol{\eta}_{i0}(t)^T\right\} \|_2).
\end{aligned}
$$

Thus, Part (c) follows directly from Condition (C7), the results of parts (a) and (b) in Theorem 1.

Proof of Theorem 2

Since the transformation function $\Phi(\cdot)$ has bounded derivative, using the delta method, we have

$$
\begin{aligned}
\widehat{\zeta}_{ik} - \zeta_{ik,0} &\approx \dot{\Phi}(\lambda_{k0}^{-1/2}\xi_{ik,0})\lambda_{k0}^{-1/2}\left\{\widehat{\xi}_{ik} - \xi_{ik,0} - \frac{1}{2}\xi_{ik,0}\lambda_{k0}^{-1}(\widehat{\lambda}_k - \lambda_{k0})\right\} \\
&= \lambda_{k0}^{-1/2} O_p\left\{\max(r_{\eta,i}, \delta_k^{-1} r_{\mathbf{G}})\right\}.
\end{aligned}
$$

The last equality follows from the results of Theorem 1 and the following equation:

$$
|\widehat{\lambda}_k - \lambda_{k0}| \le \|\frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{\eta}}_i(s)\widehat{\boldsymbol{\eta}}_i(t)^T - E\left\{\boldsymbol{\eta}_{i0}(s)\boldsymbol{\eta}_{i0}(t)^T\right\}\|_2.
$$

Proof of Theorem 3

Let $\widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\zeta}}) = (\widehat{\psi}(\sum_{k=1}^{K_n}\widehat{\boldsymbol{\alpha}}_k\mathbf{S}_n(\mathbf{s}_1)\widehat{\zeta}_{1k}), \ldots, \widehat{\psi}(\sum_{k=1}^{K_n}\widehat{\boldsymbol{\alpha}}_k\mathbf{S}_n(\mathbf{s}_{N_s})\widehat{\zeta}_{nk}))^T$, $\boldsymbol{\alpha}_0 = (\boldsymbol{\alpha}_{10}, \ldots, \boldsymbol{\alpha}_{K_n,0})$, and $\mathbf{Z}^{\boldsymbol{\beta}} = \{\mathbf{Z}_1 \otimes \mathbf{S}_n(\mathbf{s}_1), \ldots, \mathbf{Z}_n \otimes \mathbf{S}_n(\mathbf{s}_{N_s})\}^T$. Denote $\mathbf{P} = \mathbf{I} - \mathbf{Z}^{\boldsymbol{\beta}}(\mathbf{Z}^{\boldsymbol{\beta}T}\mathbf{Z}^{\boldsymbol{\beta}})^{-1}\mathbf{Z}^{\boldsymbol{\beta}T}$ and $\mathbf{Y} = \{Y_1(\mathbf{s}_1), Y_1(\mathbf{s}_2), \ldots, Y_n(\mathbf{s}_{N_s})\}^T$. We first prove that

$$
\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|^2 \le O_p\left\{\frac{K_n}{nN_s|\triangle|^2} + K_n|\triangle|^{2(\varpi+1)}\right\}. \tag{S0.48}
$$

By Condition (C5) and Lemma S.3 in Wang et al. (2020), there exists $\boldsymbol{\alpha}_{k0}$ and $\boldsymbol{\beta}_0$ such that

$$\sup_{\mathbf{s}\in\mathcal{D}} |\boldsymbol{\alpha}_{k0}\mathbf{S}_n(\mathbf{s}) - \mathbf{g}_k(\mathbf{s})| = O\left(|\triangle|^{\varpi+1}\right), \tag{S0.49}$$
$$\sup_{\mathbf{s}\in\mathcal{D}} |\boldsymbol{\beta}_0\mathbf{S}_n(\mathbf{s}) - \boldsymbol{\theta}(\mathbf{s})| = O\left(|\triangle|^{\varpi+1}\right).$$

Furthermore, by the definition of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\alpha}}$, we have

$$\frac{1}{nN_s}\|\mathbf{P}\left\{\mathbf{Y} - \widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\zeta}})\right\}\|^2 + \sum_{k=1}^{K_n} \lambda \frac{\|\widehat{\boldsymbol{\alpha}}_k\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|}$$
$$\leq \frac{1}{nN_s}\|\mathbf{P}\left\{\mathbf{Y} - \widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0, \widehat{\boldsymbol{\zeta}})\right\}\|^2 + \sum_{k=1}^{K_n} \lambda \frac{\|\boldsymbol{\alpha}_{k0}\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|}.$$

It follows that

$$\frac{1}{nN_s}\|\mathbf{P}\left\{\widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\zeta}}) - \widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0, \widehat{\boldsymbol{\zeta}})\right\}\|^2 \tag{S0.50}$$
$$\leq \frac{2}{nN_s}\left\{\mathbf{Y} - \widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0, \widehat{\boldsymbol{\zeta}})\right\}^T \mathbf{P}\left\{\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}, \widehat{\boldsymbol{\zeta}}) - \widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0, \widehat{\boldsymbol{\zeta}})\right\} + \sum_{k\in\mathcal{A}} \lambda \frac{\|\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_{k0}\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|}.$$

By the mean value theorem,

$$\widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\zeta}}) - \widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0, \widehat{\boldsymbol{\zeta}}) = \nabla_{\vec{\boldsymbol{\alpha}}^T}\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}^*, \widehat{\boldsymbol{\zeta}})(\widehat{\vec{\boldsymbol{\alpha}}} - \vec{\boldsymbol{\alpha}}_0),$$

where $\boldsymbol{\alpha}^*$ is between $\widehat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}_0$. Then, (S0.50) changes to

$$\frac{1}{nN_s}\|\mathbf{P}\nabla_{\vec{\boldsymbol{\alpha}}^T}\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}^*, \widehat{\boldsymbol{\zeta}})(\widehat{\vec{\boldsymbol{\alpha}}} - \vec{\boldsymbol{\alpha}}_0)\|^2 \tag{S0.51}$$
$$\leq \frac{2}{nN_s}\left\{\mathbf{Y} - \widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0, \widehat{\boldsymbol{\zeta}})\right\}^T \mathbf{P}\nabla_{\vec{\boldsymbol{\alpha}}^T}\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}^*, \widehat{\boldsymbol{\zeta}})(\widehat{\vec{\boldsymbol{\alpha}}} - \vec{\boldsymbol{\alpha}}_0) + \sum_{k\in\mathcal{A}} \lambda \frac{\|\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_{k0}\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|}.$$

By Lemma 2, Condition (C8), and the results of Theorem 2, we have

$$\begin{aligned}
&\frac{1}{nN_s}\|\mathbf{P}\nabla_{\vec{\boldsymbol{\alpha}}^T}\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})(\vec{\widehat{\boldsymbol{\alpha}}}-\vec{\boldsymbol{\alpha}}_0)\|^2 \\
&\geq (\vec{\widehat{\boldsymbol{\alpha}}}-\vec{\boldsymbol{\alpha}}_0)^T\frac{M}{nN_s}\sum_{i=1}^{n}\sum_{j=1}^{N_s}\left\{\widehat{\boldsymbol{\zeta}}_i\otimes\mathbf{S}_n(\mathbf{s}_j)\right\}\left\{\widehat{\boldsymbol{\zeta}}_i\otimes\mathbf{S}_n(\mathbf{s}_j)\right\}^T(\vec{\widehat{\boldsymbol{\alpha}}}-\vec{\boldsymbol{\alpha}}_0) \\
&\geq C_1\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|^2, \qquad \text{(S0.52)}
\end{aligned}$$

where M and C_1 are positive finite constants.

Combining (S0.51) and (S0.52) and applying Lemma 3 yield that

$$\begin{aligned}
&C_1\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|^2 \\
&\leq \|\frac{2}{nN_s\sqrt{C_1}}\left\{\mathbf{Y}-\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0,\widehat{\boldsymbol{\zeta}})\right\}^T\mathbf{P}\nabla_{\vec{\boldsymbol{\alpha}}^T}\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\|^2+2\sum_{k\in\mathcal{A}}\lambda\frac{\|\widehat{\boldsymbol{\alpha}}_k-\boldsymbol{\alpha}_{k0}\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|} \\
&\leq \frac{M}{n^2N_s^2}\|\left\{\boldsymbol{\psi}_0(\mathbf{G},\boldsymbol{\zeta}_0)-\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0,\widehat{\boldsymbol{\zeta}})\right\}^T\mathbf{P}\nabla_{\vec{\boldsymbol{\alpha}}^T}\boldsymbol{\psi}_0(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\|^2+2\sum_{k\in\mathcal{A}}\lambda\frac{\|\widehat{\boldsymbol{\alpha}}_k-\boldsymbol{\alpha}_{k0}\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|} \\
&\qquad +\frac{M}{n^2N_s^2}\|(\mathbf{Z}\vec{\mathbf{H}}_0)^T\mathbf{P}\nabla_{\vec{\boldsymbol{\alpha}}^T}\boldsymbol{\psi}_0(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\|^2+\frac{M}{n^2N_s^2}\|\boldsymbol{\epsilon}^T\mathbf{P}\nabla_{\vec{\boldsymbol{\alpha}}^T}\boldsymbol{\psi}_0(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\|^2 \\
&= \mathbf{I}_{n1}+\mathbf{I}_{n2}+\mathbf{I}_{n3}+\mathbf{I}_{n4}, \qquad \text{(S0.53)}
\end{aligned}$$

where $\mathbf{H}_0=\{\mathbf{h}_0(\mathbf{s}_1),\ldots,\mathbf{h}_0(\mathbf{s}_{N_s})\}$, $\mathbf{Z}=(\mathbf{Z}_1,\ldots,\mathbf{Z}_n)^T$, and $\boldsymbol{\epsilon}=\{\epsilon_1(\mathbf{s}_1),\ldots,\epsilon_n(\mathbf{s}_{N_s})\}^T$. Next, we derive the bounds for $\mathbf{I}_{n1},\mathbf{I}_{n2},\mathbf{I}_{n3}$, and $\mathbf{I}_{n4}$.

For $\mathbf{I}_{n1}$, by Conditions (C3),(C4), (S0.49), and the result of Lemma 2, we have

$$\begin{aligned}
\mathbf{I}_{n1} &= \frac{M}{n^2N_s^2}\left\{\boldsymbol{\psi}_0(\mathbf{G},\boldsymbol{\zeta}_0)-\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0,\widehat{\boldsymbol{\zeta}})\right\}^T\mathbf{P}\nabla_{\vec{\boldsymbol{\alpha}}^T}\boldsymbol{\psi}_0(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\left\{\mathbf{P}\nabla_{\vec{\boldsymbol{\alpha}}^T}\boldsymbol{\psi}_0(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\right\}^T \\
&\quad\cdot\left\{\boldsymbol{\psi}_0(\mathbf{G},\boldsymbol{\zeta}_0)-\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0,\widehat{\boldsymbol{\zeta}})\right\} \\
&\le \frac{M}{n^2N_s^2}\|\boldsymbol{\psi}_0(\mathbf{G},\boldsymbol{\zeta}_0)-\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0,\widehat{\boldsymbol{\zeta}})\|^2\lambda_{\max}\left[\mathbf{P}\nabla_{\vec{\boldsymbol{\alpha}}^T}\boldsymbol{\psi}_0(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\left\{\mathbf{P}\nabla_{\vec{\boldsymbol{\alpha}}^T}\boldsymbol{\psi}_0(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\right\}^T\right] \\
&= \frac{1}{nN_s}\|\boldsymbol{\psi}_0(\mathbf{G},\boldsymbol{\zeta}_0)-\boldsymbol{\psi}_0(\boldsymbol{\alpha}_0,\boldsymbol{\zeta}_0)+\boldsymbol{\psi}_0(\boldsymbol{\alpha}_0,\boldsymbol{\zeta}_0)-\boldsymbol{\psi}_0(\boldsymbol{\alpha}_0,\widehat{\boldsymbol{\zeta}}) \\
&\quad+\boldsymbol{\psi}_0(\boldsymbol{\alpha}_0,\widehat{\boldsymbol{\zeta}})-\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0,\widehat{\boldsymbol{\zeta}})\|^2O_p(1) \\
&= \left\{O_p\left(K_n|\triangle|^{2(\varpi+1)}+h_2^4+\frac{1}{n^2h_2^{2d}}\right)+\frac{1}{nN_s}\|\nabla_{\boldsymbol{\zeta}}\boldsymbol{\psi}_0(\boldsymbol{\alpha},\boldsymbol{\zeta}^*)(\widehat{\boldsymbol{\zeta}}-\boldsymbol{\zeta}_0)\|^2\right\}O_p(1), \\
&= O_p\left\{\frac{K_n}{nN_s|\triangle|^2}+K_n|\triangle|^{2(\varpi+1)}\right\}, \qquad \text{(S0.54)}
\end{aligned}$$

where $\boldsymbol{\zeta}^*$ is between $\widehat{\boldsymbol{\zeta}}$ and $\boldsymbol{\zeta}_0$.

Similarly, we have

$$\begin{aligned}
\mathbf{I}_{n3} &\le \frac{M}{n^2N_s^2}\|(\mathbf{Z}\vec{\mathbf{H}}_0)^T\mathbf{P}\|^2\lambda_{\max}\left[\mathbf{P}\nabla_{\vec{\boldsymbol{\alpha}}^T}\boldsymbol{\psi}_0(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\left\{\mathbf{P}\nabla_{\vec{\boldsymbol{\alpha}}^T}\boldsymbol{\psi}_0(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\right\}^T\right] \\
&= O_p\left(|\triangle|^{2(\varpi+1)}\right). \qquad \text{(S0.55)}
\end{aligned}$$

For $\mathbf{I}_{n4}$, by Conditions (C3) and (C6), we have

$$\begin{aligned}
\mathbf{I}_{n4} &\le \frac{M}{n^2N_s^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j=1}^{N_s}\sum_{j'=1}^{N_s}\epsilon_i(\mathbf{s}_j)\epsilon_{i'}(\mathbf{s}_{j'})\nabla_{\vec{\boldsymbol{\alpha}}^T}\psi_0\left\{\sum_{k=1}^{K_n}\boldsymbol{\alpha}_k^*\mathbf{S}_n(\mathbf{s}_j)\widehat{\zeta}_{ik}\right\} \\
&\quad\cdot\nabla_{\vec{\boldsymbol{\alpha}}}\psi_0\left\{\sum_{k=1}^{K_n}\boldsymbol{\alpha}_k^*\mathbf{S}_n(\mathbf{s}_{j'})\widehat{\zeta}_{i'k}\right\} \\
&= O_p\left(\frac{K_n}{nN_s|\triangle|^2}\right). \qquad \text{(S0.56)}
\end{aligned}$$

Last, for $\mathbf{I}_{n2}$, we have

$$\mathbf{I}_{n2} \leq 2\lambda\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|\sqrt{\sum_{k\in\mathcal{A}}\|\widetilde{\boldsymbol{\alpha}}_k\|^{-2}}, \tag{S0.57}$$

and

$$\sum_{k\in\mathcal{A}}\|\widetilde{\boldsymbol{\alpha}}_k\|^{-2} \leq |\mathcal{A}|\left(\min_{k\in\mathcal{A}}\|\widetilde{\boldsymbol{\alpha}}_k\|\right)^{-2} = |\mathcal{A}|\left(\frac{\min_{k\in\mathcal{A}}\|\widetilde{\boldsymbol{\alpha}}_k\|}{\min_{k\in\mathcal{A}}\|\boldsymbol{\alpha}_{k0}\|}\right)^{-2}\left(\min_{k\in\mathcal{A}}\|\boldsymbol{\alpha}_{k0}\|\right)^{-2}.$$

Furthermore, by $P\left(\frac{\min_{k\in\mathcal{A}}\|\widetilde{\boldsymbol{\alpha}}_k\|}{\min_{k\in\mathcal{A}}\|\boldsymbol{\alpha}_{k0}\|} < \frac{1}{M}\right) > 1-\epsilon$, we have $\sum_{k\in\mathcal{A}}\|\widetilde{\boldsymbol{\alpha}}_k\|^{-2} = O_p\left(\frac{|\mathcal{A}|}{\min_{k\in\mathcal{A}}\|\boldsymbol{\alpha}_{k0}\|^2}\right) = O_p(\frac{|\mathcal{A}|}{\varrho^2})$. This together with (S0.57) yields

$$\mathbf{I}_{n2} = \lambda\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|O_p\left(\frac{|\mathcal{A}|^{1/2}}{\varrho}\right). \tag{S0.58}$$

Then, (S0.48) is obtained from (S0.53), (S0.54), (S0.58), (S0.55), and (S0.56).

Next, we prove

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 \leq O_p\left\{\frac{K_n}{nN_s|\triangle|^2} + K_n|\triangle|^{2(\varpi+1)}\right\}. \tag{S0.59}$$

From the estimation equation of $\boldsymbol{\beta}$, we have

$$\begin{aligned}
\vec{\widehat{\boldsymbol{\beta}}} - \vec{\boldsymbol{\beta}}_0 &= \left(\mathbf{Z}^{\boldsymbol{\beta}T}\mathbf{Z}^{\boldsymbol{\beta}}\right)^{-1}\mathbf{Z}^{\boldsymbol{\beta}T}\left\{\mathbf{Y} - \widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\zeta}})\right\} - \vec{\boldsymbol{\beta}}_0 \\
&= \left(\mathbf{Z}^{\boldsymbol{\beta}T}\mathbf{Z}^{\boldsymbol{\beta}}\right)^{-1}\mathbf{Z}^{\boldsymbol{\beta}T}\mathbf{Z}\vec{\mathbf{H}}_0 - \vec{\boldsymbol{\beta}}_0 + \left(\mathbf{Z}^{\boldsymbol{\beta}T}\mathbf{Z}^{\boldsymbol{\beta}}\right)^{-1}\mathbf{Z}^{\boldsymbol{\beta}T}\boldsymbol{\epsilon} \\
&\quad + \left(\mathbf{Z}^{\boldsymbol{\beta}T}\mathbf{Z}^{\boldsymbol{\beta}}\right)^{-1}\mathbf{Z}^{\boldsymbol{\beta}T}\left\{\boldsymbol{\psi}_0(\mathbf{G}, \boldsymbol{\zeta}_0) - \widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\zeta}})\right\}.
\end{aligned}$$

By the property of projection matrices, (S0.49), and Condition (C6),

$$\begin{aligned}
\|\left(\mathbf{Z}^{\boldsymbol{\beta}T}\mathbf{Z}^{\boldsymbol{\beta}}\right)^{-1}\mathbf{Z}^{\boldsymbol{\beta}T}\mathbf{Z}\vec{\mathbf{H}}_0 - \vec{\boldsymbol{\beta}}_0\|^2 &= O_p\left(|\triangle|^{2(\varpi+1)}\right), \\
\|\left(\mathbf{Z}^{\boldsymbol{\beta}T}\mathbf{Z}^{\boldsymbol{\beta}}\right)^{-1}\mathbf{Z}^{\boldsymbol{\beta}T}\boldsymbol{\epsilon}\|^2 &= O_p\left(\frac{1}{nN_s|\triangle|^2}\right).
\end{aligned}$$

Following similar calculations as $\mathbf{I}_{n1}$ in (S0.54),

$$
\begin{aligned}
&\| \left(\mathbf{Z}^{\boldsymbol{\beta}T}\mathbf{Z}^{\boldsymbol{\beta}}\right)^{-1} \mathbf{Z}^{\boldsymbol{\beta}T} \left\{\boldsymbol{\psi}_0(\mathbf{G},\boldsymbol{\zeta}_0) - \widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\zeta}})\right\} \|^2 \\
&= \left\{\boldsymbol{\psi}_0(\mathbf{G},\boldsymbol{\zeta}_0) - \widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0,\widehat{\boldsymbol{\zeta}})\right\}^T \mathbf{Z}^{\boldsymbol{\beta}} \left(\mathbf{Z}^{\boldsymbol{\beta}T}\mathbf{Z}^{\boldsymbol{\beta}}\right)^{-1} \left(\mathbf{Z}^{\boldsymbol{\beta}T}\mathbf{Z}^{\boldsymbol{\beta}}\right)^{-1} \mathbf{Z}^{\boldsymbol{\beta}T} \left\{\boldsymbol{\psi}_0(\mathbf{G},\boldsymbol{\zeta}_0) - \widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}_0,\widehat{\boldsymbol{\zeta}})\right\} \\
&\le \frac{1}{nN_s} \|\boldsymbol{\psi}_0(\mathbf{G},\boldsymbol{\zeta}_0) - \widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\zeta}})\|^2 \lambda_{\max}\left[\left\{\frac{1}{nN_s}\mathbf{Z}^{\boldsymbol{\beta}T}\mathbf{Z}^{\boldsymbol{\beta}})^{-1}\right\}\right] \\
&= O_p\left\{\frac{K_n}{nN_s|\triangle|^2} + K_n|\triangle|^{2(\varpi+1)}\right\},
\end{aligned}
$$

where the last equality holds from Conditions (C4), (C8), and the result of (S0.48). To sum up yields the result in (S0.59). Hence, Theorem 3 follows from (S0.48), (S0.59), and Lemma 4.

Proof of Theorem 4

Under the Conditions in Theorem 4, we get $\widehat{\mathbf{g}}_k \to_p \mathbf{g}_{k0}$ for $\forall k \in \mathcal{A}$, then

$$\lim_{n\to\infty} P\left(\|\widehat{\mathbf{g}}_k\|_2 \ne 0, k \in \mathcal{A}\right) = 1.$$

Thus, to prove the model selection consistency, we only need to show

$$\lim_{n\to\infty} P\left(\|\widehat{\mathbf{g}}_k\|_2 = 0, k \notin \mathcal{A}\right) = 1.$$

By way of contradiction, suppose $\|\widehat{\mathbf{g}}_k\| \ne 0$ for some $k \notin \mathcal{A}$. Denote $\widehat{\boldsymbol{\alpha}}^*$ as $\widehat{\boldsymbol{\alpha}}$ except that $\widehat{\boldsymbol{\alpha}}^*_k = \mathbf{0}$. By some algebras, we have

$$\frac{1}{nN_s}\|\mathbf{P}\left\{\mathbf{Y}-\widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\zeta}})\right\}\|^2+\sum_{k=1}^{K_n}\lambda\frac{\|\widehat{\boldsymbol{\alpha}}_k\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|}-\frac{1}{nN_s}\|\mathbf{P}\left\{\mathbf{Y}-\widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}}^*,\widehat{\boldsymbol{\zeta}})\right\}\|^2-\sum_{k=1}^{K_n}\lambda\frac{\|\widehat{\boldsymbol{\alpha}}^*\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|}$$

$$=\frac{1}{nN_s}\|\mathbf{P}\left\{\widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\zeta}})-\widehat{\boldsymbol{\psi}}\widehat{\boldsymbol{\alpha}}^*,\widehat{\boldsymbol{\zeta}})\right\}\|^2-\frac{2}{nN_s}\left\{\mathbf{Y}-\widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}}^*,\widehat{\boldsymbol{\zeta}})\right\}^T\mathbf{P}\left\{\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha},\widehat{\boldsymbol{\zeta}})-\widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}}^*,\widehat{\boldsymbol{\zeta}})\right\}$$

$$+\lambda\frac{\|\widehat{\boldsymbol{\alpha}}_k\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|}$$

$$=\frac{1}{nN_s}\|\mathbf{P}\nabla_{\boldsymbol{\alpha}_k}\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\widehat{\boldsymbol{\alpha}}_k\|^2-\frac{2}{nN_s}\left\{\mathbf{Y}-\widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}}^*,\widehat{\boldsymbol{\zeta}})\right\}^T\mathbf{P}\nabla_{\boldsymbol{\alpha}_k}\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\widehat{\boldsymbol{\alpha}}_k+\lambda\frac{\|\widehat{\boldsymbol{\alpha}}_k\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|},$$

where $\boldsymbol{\alpha}^*$ is between $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\alpha}}^*$.

By the compatible property of matrix norm, and following the lines of Theorem 3, we have

$$\begin{aligned}
&\|\frac{2}{nN_s}\left\{\mathbf{Y}-\widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}}^*,\widehat{\boldsymbol{\zeta}})\right\}^T\mathbf{P}\nabla_{\boldsymbol{\alpha}_k}\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\widehat{\boldsymbol{\alpha}}_k\|\\
&\leq\|\frac{2}{nN_s}\left\{\mathbf{Y}-\widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}}^*,\widehat{\boldsymbol{\zeta}})\right\}^T\mathbf{P}\nabla_{\boldsymbol{\alpha}_k}\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}^*,\widehat{\boldsymbol{\zeta}})\|\|\widehat{\boldsymbol{\alpha}}_k\|\\
&=O_p\left\{\frac{K_n}{nN_s|\triangle|^2}+K_n|\triangle|^{2(\varpi+1)}\right\}\|\widehat{\boldsymbol{\alpha}}_k\|.
\end{aligned}$$

On the other hand, by the adaptive weight $\max_{k\notin\mathcal{A}}\|\widetilde{\mathbf{g}}_k\|_2=O_p\left(n^{-\delta}\right)$, we have

$$\lambda\frac{\|\widehat{\boldsymbol{\alpha}}_k\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|}=\lambda n^{\delta}\|\widehat{\boldsymbol{\alpha}}_k\|.$$

Thus, by condition $\left\{\frac{K_n}{nN_s|\triangle|^2}+K_n|\triangle|^{2(\varpi+1)}\right\}\left(\lambda n^{\delta}\right)^{-1}\to 0$, we have

$$\begin{aligned}
&\frac{1}{nN_s}\|\mathbf{P}\left\{\mathbf{Y}-\widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}}^*,\widehat{\boldsymbol{\zeta}})\right\}\|^2+\sum_{k=1}^{K_n}\lambda\frac{\|\widehat{\boldsymbol{\alpha}}^*\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|}\\
&<\frac{1}{nN_s}\|\mathbf{P}\left\{\mathbf{Y}-\widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\zeta}})\right\}\|^2+\sum_{k=1}^{K_n}\lambda\frac{\|\widehat{\boldsymbol{\alpha}}_k\|}{\|\widetilde{\boldsymbol{\alpha}}_k\|},
\end{aligned}$$

with probability approaching 1, which is a contradiction.

Web Appendix C: Additional analysis of ADNI data and results

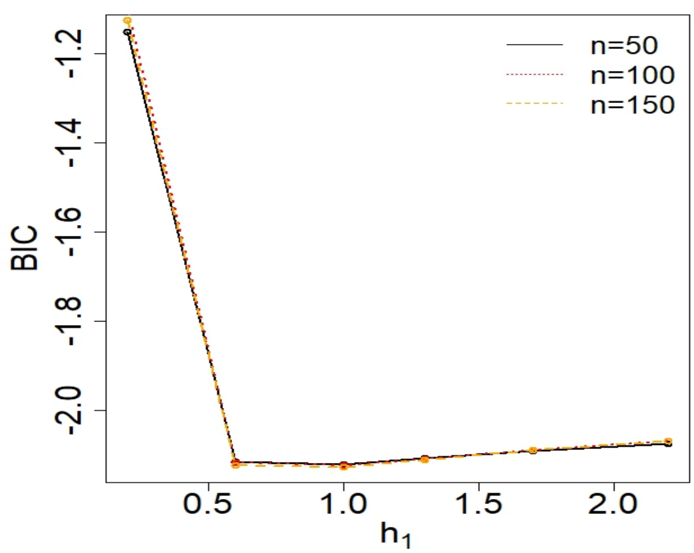
Analysis and Results given APOE-ϵ4 and Disease Status

We repeated the analysis by excluding the SNPs in the 19q13.32 region but including the number of APOE-ϵ4 alleles and baseline diagnosis status as one of the controlling covariates, resulting in $p = 99$ LD blocks to the left and $p = 100$ LD blocks to the right hippocampi. The cytogenetic region 19q13.32 contains the well-known APOE-ϵ4 gene. Because clinical notes provide supplementary information and are considered case-by-case, the effects of the SNPs and demographic variables on the change of cognitive performance may be confounded with the effects of differences in the number of APOE-ϵ4 gene copies and baseline diagnosis. Therefore, we are interested in whether the relationships would alter when adjusting for the number of APOE-ϵ4 alleles and baseline diagnosis status. The number of APOE-ϵ4 gene copies is coded using two dummy variables: APOE1 and APOE2, indicating the number of APOE-ϵ4 alleles; the diagnosis status is coded by dummy variables: MCI and AD. Consequently, $\mathbf{Z}_i$ is 10-dimensional real-valued covariates, including 5 clinical variables, 2 dummy variables for the number of APOE-ϵ4 alleles, 2 dummy variables for the diagnosis status, and 1 intercept term.

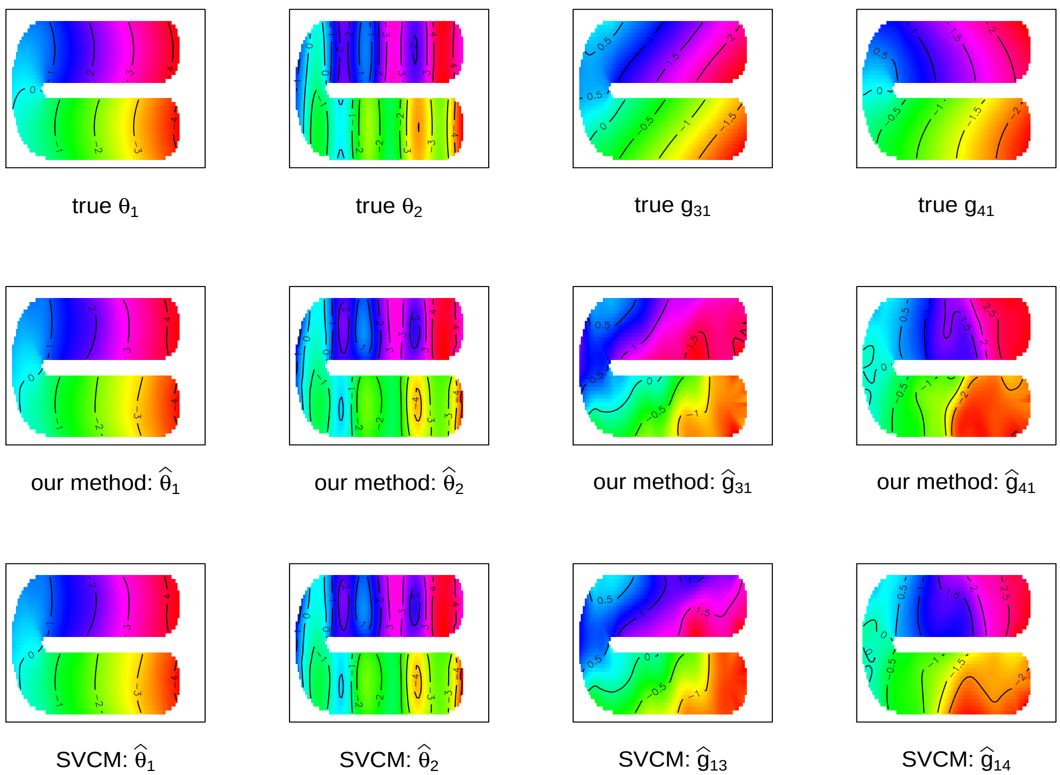
Web Figures 6 and 7 present the estimated coefficient functions, and Table 4 summarizes the prediction error. After introducing the number of APOE-ϵ4 gene copies and baseline diagnosis, the prediction errors decrease, and similar patterns of the coefficient function estimates for clinical variables are observed. The disease status MCI has significant negative effects on the radial distance of both hippocampi, and carrying the APOE-ϵ4 allele negatively affects cognitive function. Compared to MCI and APOE1, AD and APOE2 have more substantial adverse effects on the radial distance of both hippocampi. As a core AD biomarker, an increase in the APOE-ϵ4 alleles increases the risk of AD (Bekris et al., 2010).

Web Tables 1–11 and Figures 1–7

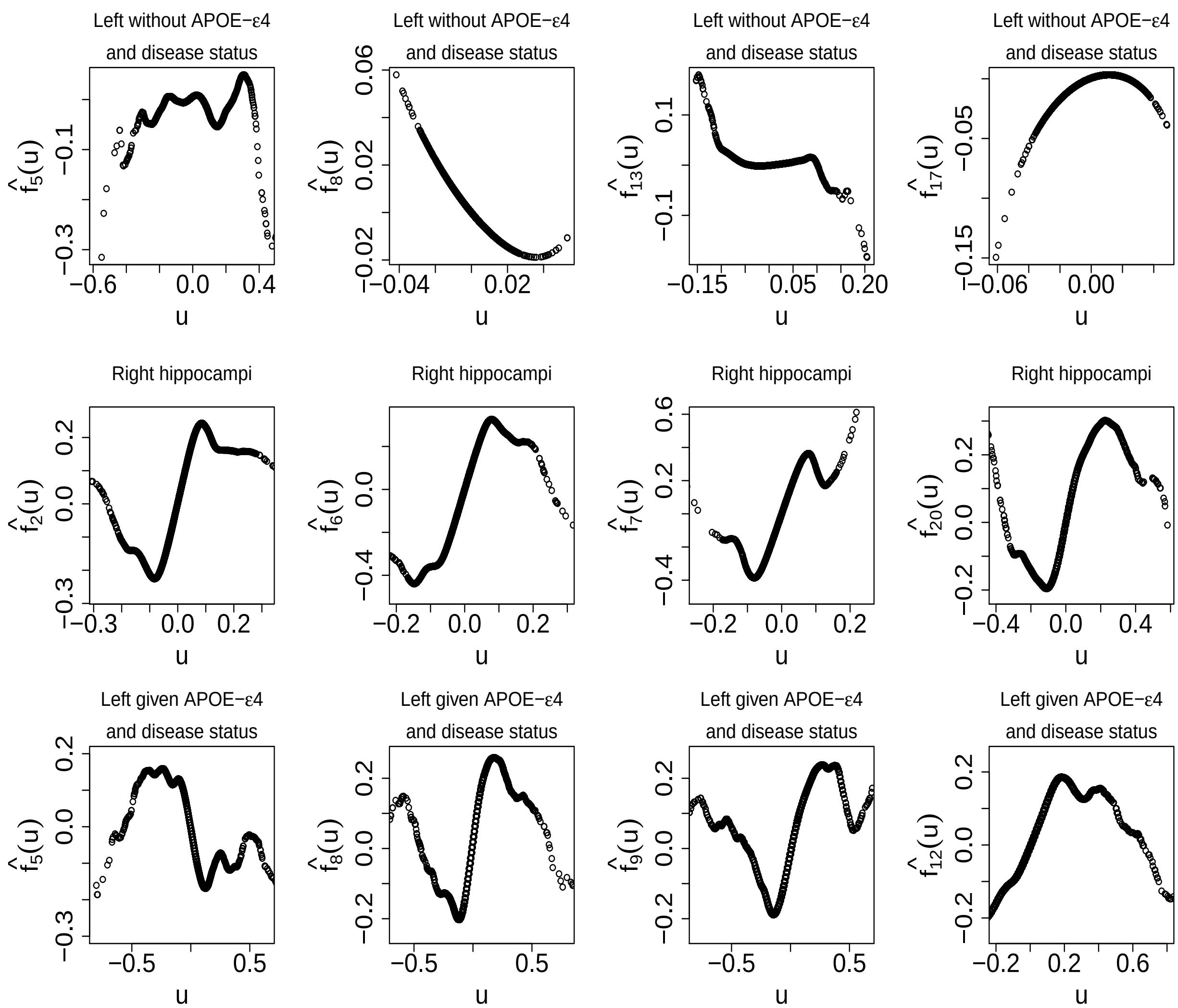
For image-on-scalar regression, we compared FMVCM with SVCM. The standard SVCM is written as $Y_i(\mathbf{s}) = \sum_{k=1}^{K_n} g_{k1}(\mathbf{s})\zeta_{ik} + \boldsymbol{\theta}(\mathbf{s})^T\mathbf{Z}_i + \epsilon_i(\mathbf{s})$.



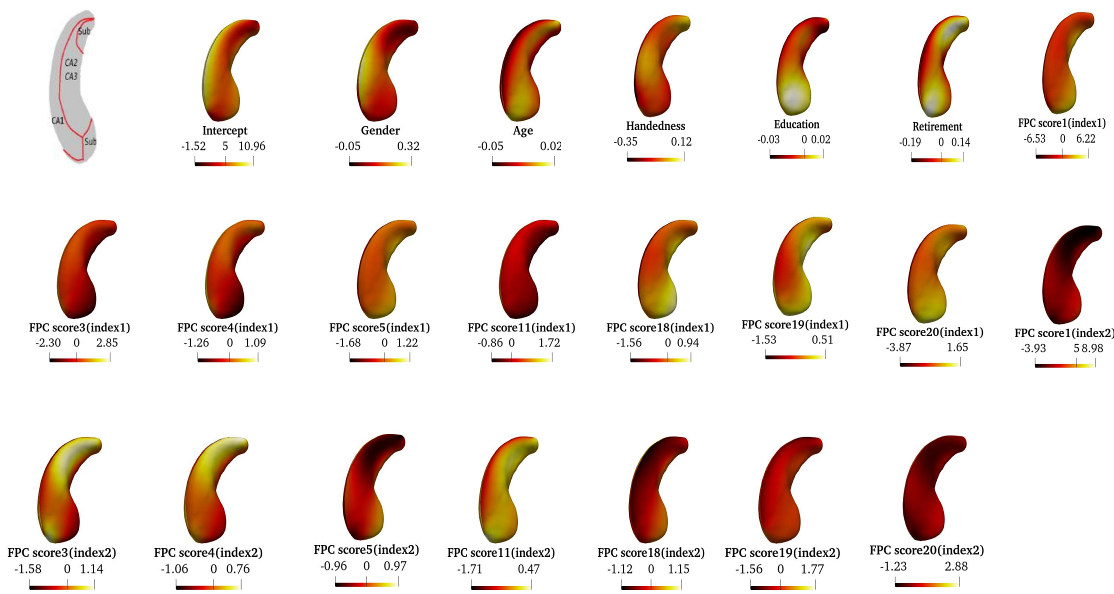
Web Figure 1: The values of BIC(h_1) for Case I in Xmodel with $n = 50, 100$, and 150, $n_{ij} = 80$ and $\sigma^2 = 0.1$.



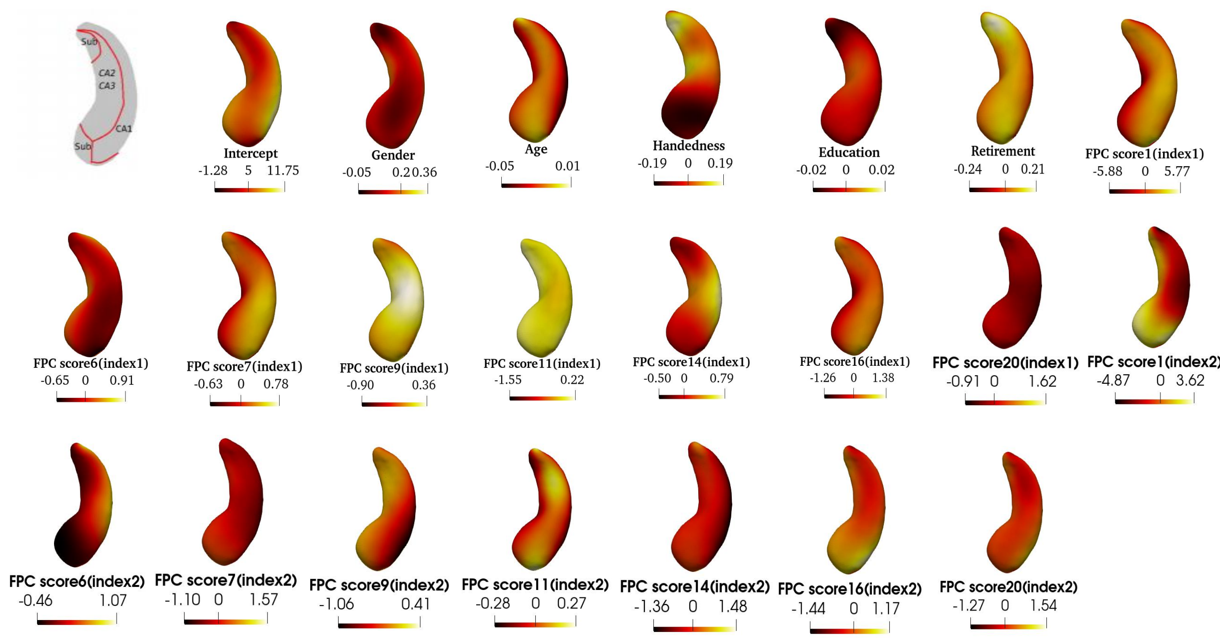
Web Figure 2: The true nonzeo coefficient functions and average estimators for Ymodel II (linear $\psi(\cdot)$) with $n = 100$, and $\sigma_\epsilon^2 = 0.25$.



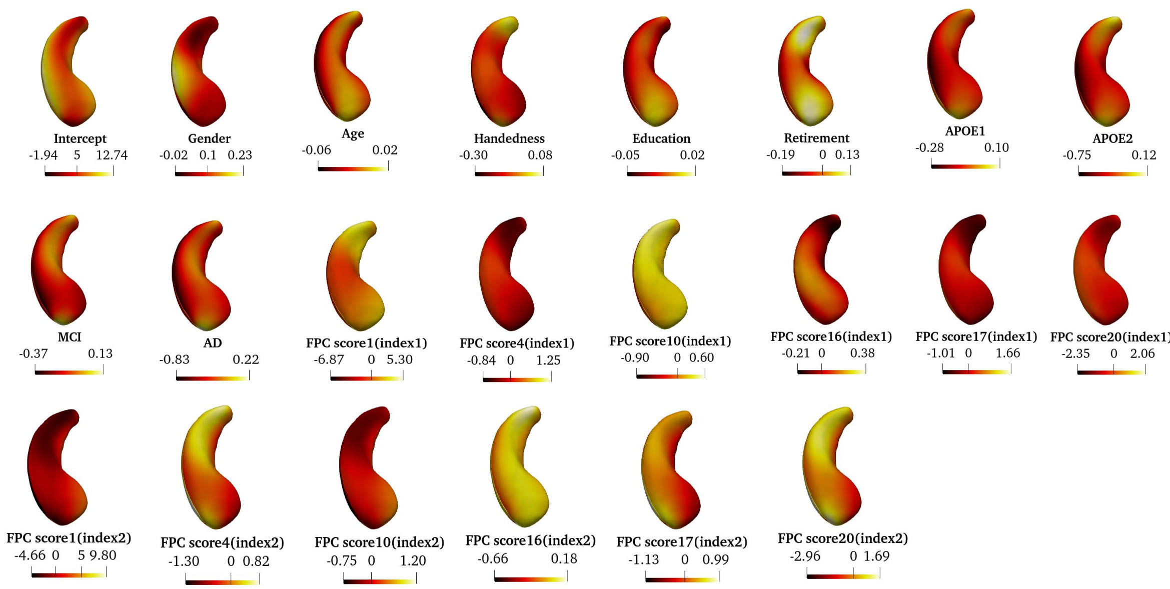
Web Figure 3: Parts of the estimates $\widehat{f}_j(\cdot)$ for the left hippocampus without APOE-ϵ4 and disease status, for the right hippocampus, and for the left hippocampus given APOE-ϵ4 and disease status (from top to bottom). $\widehat{f}_j(\cdot)$ is the same for the right hippocampus with and without APOE-ϵ4 and disease status.



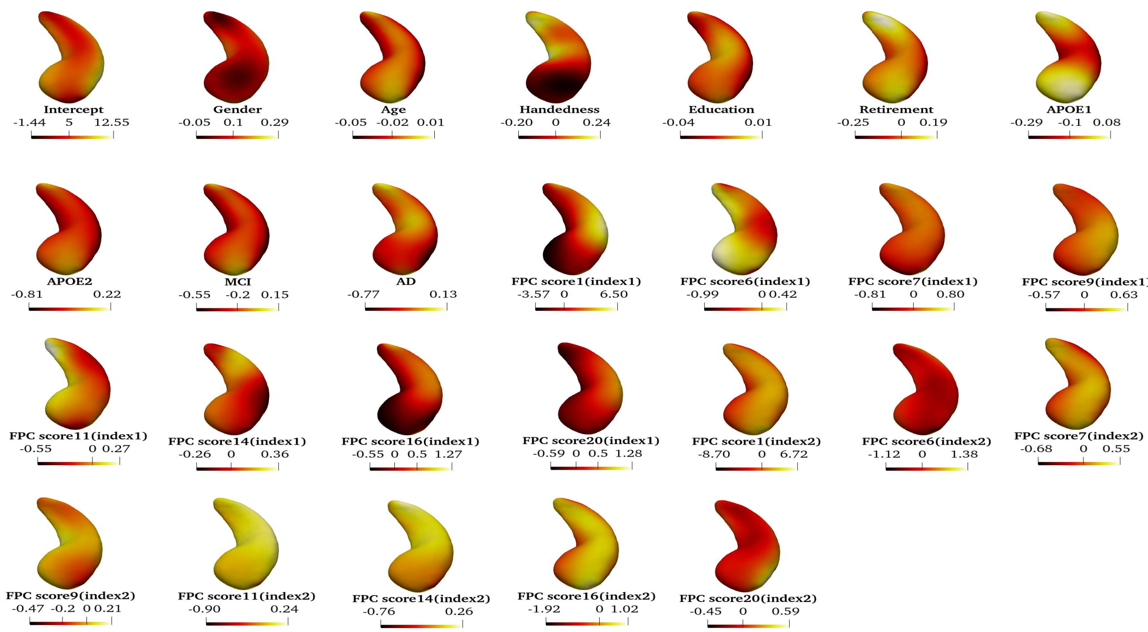
Web Figure 4: The proposed estimates of coefficient functions with two indices for the left hippocampus, excluding APOE-ϵ4 and disease status.



Web Figure 5: The proposed estimates of coefficient functions with two indices for the right hippocampus without APOE-ϵ4, and disease status.



Web Figure 6: The proposed estimates of coefficient functions with two indices for the left hippocampus given APOE-$\epsilon 4$, and disease status.



Web Figure 7: The proposed estimates of coefficient functions with two indices for the right hippocampus given APOE-ϵ4, and disease status.

Web Table 1: The estimated MSEs of kth score ξ_k and jth functional variable $X_j(\cdot)$, using nonlinear MFPCA (the proposed) and MFPCA in Case II (linear f_j) of Xmodel, with various sample sizes, observation levels, and noise levels.

		MFPCA						NMFPCA					
	(n, n_{ij})	$MSE(\widehat{\xi}_1)$	$MSE(\widehat{\xi}_2)$	$MSE(\widehat{\xi}_3)$	$MSE(\widehat{\xi}_4)$	$MSPE(\widehat{X}_1)$	$MSPE(\widehat{X}_2)$	$MSE(\widehat{\xi}_1)$	$MSE(\widehat{\xi}_2)$	$MSE(\widehat{\xi}_3)$	$MSE(\widehat{\xi}_4)$	$MSPE(\widehat{X}_1)$	$MSPE(\widehat{X}_2)$
$\sigma^2 = 0$	(50,10)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01
	(50,40)	2E-5	3E-5	2E-5	4E-6	2E-4	2E-4	4E-4	7E-4	1E-4	1E-4	1E-3	1E-3
	(50,80)	2E-6	3E-5	4E-6	6E-7	2E-4	2E-4	3E-5	3E-4	2E-5	7E-6	4E-4	3E-4
	(100,10)	0.01	3E-3	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	(100,40)	1E-5	2E-5	1E-5	3E-6	2-4	2E-4	7E-5	2E-4	3E-5	2E-5	5E-4	5E-4
	(100,80)	2E-6	2E-5	3E-6	7E-7	1E-4	1E-4	2E-5	2E-4	1E-5	4E-6	3E-4	3E-4
	(150,10)	0.01	3E-3	0.01	0.01	0.01	0.01	2E-3	0.01	4E-3	2E-3	4E-3	4E-3
	(150,40)	1E-5	1E-5	1E-5	3E-6	2E-4	2E-4	4E-5	1E-4	2E-5	1E-5	4E-4	4E-4
	(150,80)	2E-6	1E-5	3E-6	8E-7	1E-4	1E-4	1E-5	1E-4	6E-6	3E-6	2E-4	2E-4
$\sigma^2 = 0.5$	(50,10)	22.37	10.15	5.60	3.67	3.09	3.14	0.36	0.25	0.18	0.14	0.36	0.36
	(50,40)	2.17	0.83	0.47	0.33	0.57	0.58	0.23	0.19	0.13	0.12	0.11	0.11
	(50,80)	0.10	0.09	0.09	0.07	0.09	0.10	0.04	0.06	0.06	0.04	0.04	0.04
	(100,10)	25.17	10.90	6.77	4.49	3.31	3.61	0.34	0.25	0.17	0.12	0.33	0.33
	(100,40)	1.56	0.80	0.51	0.35	0.57	0.57	0.21	0.18	0.12	0.12	0.10	0.10
	(100,80)	0.06	0.07	0.09	0.09	0.10	0.10	0.022	0.05	0.05	0.04	0.04	0.04
	(150,10)	20.81	11.39	6.64	4.53	3.32	3.56	0.33	0.23	0.16	0.11	0.31	0.30
	(150,40)	1.44	0.78	0.54	0.39	0.58	0.56	0.19	0.17	0.12	0.11	0.08	0.08
	(150,80)	0.07	0.07	0.09	0.10	0.11	0.11	0.02	0.05	0.05	0.04	0.03	0.03
$\sigma^2 = 1.5$	(50,10)	71.18	30.27	17.05	11.03	8.43	8.82	0.38	0.28	0.19	0.14	0.72	0.73
	(50,40)	6.12	2.45	1.40	0.89	1.29	1.30	0.29	0.19	0.13	0.12	0.26	0.26
	(50,80)	0.62	0.30	0.20	0.15	0.27	0.29	0.15	0.17	0.11	0.08	0.11	0.11
	(100,10)	71.19	32.40	20.30	13.53	9.36	9.60	0.37	0.26	0.19	0.14	0.64	0.66
	(100,40)	4.75	2.42	1.58	1.07	1.19	1.17	0.28	0.19	0.13	0.12	0.23	0.23
	(100,80)	0.47	0.33	0.21	0.17	0.28	0.28	0.13	0.16	0.11	0.08	0.09	0.09
	(150,10)	60.72	33.44	20.79	14.10	9.53	9.96	0.36	0.25	0.18	0.12	0.60	0.61
	(150,40)	4.39	2.45	1.66	1.14	1.15	1.11	0.27	0.19	0.12	0.12	0.21	0.21
	(150,80)	0.51	0.35	0.23	0.18	0.30	0.30	0.13	0.15	0.10	0.08	0.09	0.09

Web Table 2: Bias and SD of the estimated number of FPCs.

(n, n_{ij})	(50,10)	(50,40)	(50,80)	(100,10)	(100,40)	(100,80)	(150,10)	(150,40)	(150,80)
Bias	0.04	0.03	0.01	0.03	0.02	0.01	0.03	0.01	0.01
SD	0.28	0.25	0.21	0.24	0.19	0.19	0.2	0.15	0.13

Web Table 3: MSEs of nonzero coefficient functions and prediction errors, using FMVCM (the proposed) and SVCM, for Ymodel II (linear $\psi(\cdot)$).

		FMVCM					SVCM				
σ_ϵ^2	n	$MSPE(Y)$	MSE($\widehat{\theta}_1$)	MSE($\widehat{\theta}_2$)	MSE($\widehat{g}_{31}$)	MSE($\widehat{g}_{41}$)	$MSPE(Y)$	MSE($\widehat{\theta}_1$)	MSE($\widehat{\theta}_2$)	MSE($\widehat{g}_{31}$)	MSE($\widehat{g}_{41}$)
0.1	50	1.01	0.01	0.02	0.09	0.07	1.09	4E-3	0.02	0.09	0.06
	100	0.67	3E-3	0.02	0.08	0.04	0.84	3E-3	0.02	0.07	0.02
	150	0.18	4E-3	0.01	0.04	0.03	0.35	2E-3	0.01	0.03	0.02
0.25	50	1.14	0.01	0.03	0.11	0.10	1.23	0.02	0.02	0.10	0.08
	100	0.84	0.01	0.02	0.10	0.07	1.04	0.01	0.02	0.09	0.05
	150	0.33	0.01	0.02	0.07	0.04	0.49	0.01	0.02	0.10	0.03

Web Table 4: Results of the selection percentage for $\mathbf{g}_k$, #correct nonzero (the number of nonzero coefficient functions correctly identified as nonzero functions), and #correct zero (the number of zero coefficient functions correctly identified as zero functions), based on 100 repetitions.

		Ymode I						Ymode II					
		selected percentages				#correct	#correct	selected percentages				#correct	#correct
σ_ϵ^2	n	$\widehat{\mathbf{g}}_1$	$\widehat{\mathbf{g}}_2$	$\widehat{\mathbf{g}}_3$	$\widehat{\mathbf{g}}_4$	nonzero	zero	$\widehat{\mathbf{g}}_1$	$\widehat{\mathbf{g}}_2$	$\widehat{\mathbf{g}}_3$	$\widehat{\mathbf{g}}_4$	nonzero	zero
0.1	50	3	2	95	100	95	97	2	2	100	92	92	98
	100	0	0	96	100	96	100	0	0	100	94	94	100
	150	0	0	98	100	98	100	0	0	100	95	95	100
0.25	50	7	7	86	100	86	93	4	4	100	90	90	96
	100	0	0	92	100	92	100	0	0	100	93	93	100
	150	0	0	95	100	95	100	0	0	100	93	93	100

Web Table 5: Relative frequency of the selected number of index q using the BIC in the main paper based on 100 repetitions.

		$q=1$	$q=2$	$q=3$	$q=4$	$q=5$
Ymodel I	$n=50$	0.00	0.96	0.02	0.02	0.00
	$n=100$	0.00	0.96	0.01	0.00	0.03
	$n=150$	0.00	0.98	0.01	0.01	0.00
Ymodel II	$n=50$	0.97	0.00	0.01	0.00	0.02
	$n=100$	0.99	0.00	0.01	0.00	0.00
	$n=150$	0.99	0.00	0.01	0.00	0.00

Web Table 6: MSE and SD (in parenthesis) of the proposed and MFPCA + FMVCM estimators.

	$MSPE(Y)$	$MSE(\widehat{\theta}_1)$	$MSE(\widehat{\theta}_2)$	$MSE(\widehat{g}_{31})$	$MSE(\widehat{g}_{41})$	$MSE(\widehat{g}_{32})$	$MSE(\widehat{g}_{42})$
	(sd)	(sd)	(sd)	(sd)	(sd)	(sd)	(sd)
MFPCA + FMVCM	1.92	0.01	0.02	0.45	0.97	1.89	1.22
	(0.01)	(1E-3)	(2E-3)	(0.61)	(1.17)	(2.24)	(1.64)
FJM	1.44	0.01	0.03	0.13	0.40	0.60	0.39
	(3E-3)	(4E-4)	(1E-3)	(0.01)	(0.02)	(0.02)	(0.01)
FMVCM + T	1.01	0.01	0.02	0.09	0.29	0.63	0.32
	(2E-3)	(3E-4)	(1E-3)	(0.01)	(0.03)	(0.02)	(0.02)
FMVCM + TI	0.89	4E-3	0.01	0.10	0.31	0.61	0.30
	(2E-3)	(3E-4)	(1E-3)	(0.01)	(0.03)	(0.02)	(0.01)

Web Table 7: MSE and SD (in parenthesis) of the proposed estimators when K_n is misspecified. Results are presented for Ymodel I with $n = 100, \sigma_\epsilon^2 = 0.1$, and true $K_n = 4$. "-" means that the algorithm does not involve this estimator under the corresponding data setting.

$\widehat{K}_n$	$MSE(\widehat{\xi}_1)$	$MSE(\widehat{\xi}_2)$	$MSE(\widehat{\xi}_3)$	$MSE(\widehat{\xi}_4)$	$MSE(\widehat{X}_1)$	$MSE(\widehat{X}_2)$	$MSPE(Y)$	$MSE(\widehat{\theta}_1)$	$MSE(\widehat{\theta}_2)$	$MSE(\widehat{g}_{31})$	$MSE(\widehat{g}_{41})$	$MSE(\widehat{g}_{32})$	$MSE(\widehat{g}_{42})$
	(sd)	(sd)	(sd)	(sd)	(sd)	(sd)	(sd)	(sd)	(sd)	(sd)	(sd)	(sd)	(sd)
2	0.06	0.15	-	-	0.03	0.07	3.38	0.03	0.04	-	-	-	-
	(0.01)	(0.03)	(-)	(-)	(0.01)	(3E-3)	(0.01)	(0.01)	(3E-3)	(-)	(-)	(-)	(-)
4	0.01	0.02	0.01	0.01	0.03	0.01	1.44	0.01	0.03	0.13	0.40	0.60	0.39
	(4E-3)	(0.01)	(0.01)	(0.01)	(0.01)	(3E-3)	(3E-3)	(4E-4)	(1E-3)	(0.01)	(0.02)	(0.02)	(0.01)
7	0.01	0.02	0.01	0.01	0.02	0.02	1.45	0.01	0.02	0.12	0.42	0.60	0.38
	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(1E-3)	(2E-3)	(0.01)	(0.03)	(0.03)	(0.02)
12	0.02	0.03	0.01	0.01	0.02	0.03	1.45	0.02	0.02	0.14	0.41	0.60	0.39
	(0.05)	(0.05)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(1E-3)	(1E-3)	(0.01)	(0.03)	(0.03)	(0.02)

Web Table 8: Results for the simulation setting mimics the ADNI data analysis.

	MFPCA		NMFPCA		FMVCM			SVCM	
σ^2	$MSE(\widehat{\boldsymbol{\xi}})$	$MSPE(\widehat{\mathbf{X}})$	$MSE(\widehat{\boldsymbol{\xi}})$	$MSPE(\widehat{\mathbf{X}})$	$MSPE(Y)$	MSE($\widehat{\boldsymbol{\theta}}$)	MSE($\widehat{\mathbf{G}}$)	$MSPE(Y)$	MSE($\widehat{\boldsymbol{\theta}}$)
0.1	2.13	0.07	0.01	0.02	1.36	0.03	0.6	4.05	0.35
0.5	2.20	0.29	0.03	0.14	1.45	0.06	0.7	4.74	0.49

Web Table 9: Results of the selection percentage for $\mathbf{g}_k$ in the simulation setting mimics the ADNI data analysis.

	selected percentages								#correct	#correct
σ^2	$\widehat{\mathbf{g}}_1$	$\widehat{\mathbf{g}}_3$	$\widehat{\mathbf{g}}_4$	$\widehat{\mathbf{g}}_5$	$\widehat{\mathbf{g}}_{11}$	$\widehat{\mathbf{g}}_{18}$	$\widehat{\mathbf{g}}_{19}$	$\widehat{\mathbf{g}}_{20}$	nonzero	zero
0.1	100	97	94	98	98	94	96	96	94	95
0.5	93	92	93	92	95	93	92	93	92	94

Web Table 10: Computing time for the left and right hippocampi (sec.).

	without APOE-ϵ4 and disease status		given APOE-ϵ4 and disease status	
	left hippocampi	right hippocampi	left hippocampi	right hippocampi
NMFPCA	3983.63	5086.91	4957.96	5086.91
FMVCM	7892.88	7338.12	9754.79	8901.01

Web Table 11: Selected FPCs for the left and right hippocampi.

	MFPCA + FMVCM		NMFPCA + FMVCM	
	left hippocampi	right hippocampi	left hippocampi	right hippocampi
without APOE-ϵ4 and disease status	FPC 1,3,4,6,11, 13,14,16,18,20	FPC 1,2,8,11,13, 14,15,16,20	FPC 1,3,4,5,11,18, 19,20	FPC 1,6,7,9,11,14, 16,20
given APOE-ϵ4 and disease status	FPC 1,2,3,6,8,10,11,12, 13,15,16,17,18,19,20	FPC 1,2,3,5,6,7,8, 12,13,14,17,18,20	FPC 1,4,10,16, 17,20	FPC 1,6,7,9,11,14, 16,20

Web Table 12: Number of negative entries for the estimated coefficient functions of clinical variables.

		intercept	Gender	Hand	Edu	retire	age	MCI	AD	APOE1	APOE2
without APOE-ϵ4 and	left hippocampi	243	157	13,429	11,313	4,709	13,972				
disease status	right hippocampi	251	929	5,079	10,247	1,615	13,897				
given APOE-ϵ4 and	left hippocampi	253	99	13,273	13,664	5,138	14,134	14,076	14,169	13,970	14,143
disease status	right hippocampi	285	1,639	4,839	13,703	3,446	13,917	13,997	14,172	11,003	14,104

Bibliography

Bekris, L., C. Yu, T. Bird, and D. Tsuang (2010). Genetics of alzheimer disease. *Journal of Geriatric Psychiatry and Neurology 23*(4), 213–227.

Bosq, D. (2000). *Linear processes in function spaces: theory and applications*. Springer, New York.

Cardot, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *Journal of Nonparametric Statistics 12*(4), 503–538.

Chiou, J.-M., Y.-T. Chen, and Y.-F. Yang (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica 24*(4), 1571–1596.

Cui, X., W. K. Härdle, and L. Zhu (2011). The efm approach for single-index models. *The Annals of Statistics 39*(3), 1658–1688.

David G. Luenberger, Y. Y. (2016). *Linear and Nonlinear Programming*. Springer International Publishing.

Duan, N. and K.-C. Li (1991). Slicing regression: a link-free regression method. *The Annals of Statistics 19*(2), 505–530.

Happ, C. and S. Greven (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association 113*(522), 649–659.

Jacques, J. and C. Preda (2014). Model-based clustering for multivariate functional data. *Computational Statistics and Data Analysis 71*(Special I), 92–106.

Lai, M.-J. and L. L. Schumaker (2007). *Spline functions on triangulations*. New York: Cambridge University Press.

Li, X., L. Wang, H. J. Wang, and A. D. N. Initiative (2021). Sparse learning and structure identification for ultrahigh-dimensional image-on-scalar regression. *Journal of the American Statistical Association 116*(536), 1994–2008.

Schumacker, L. (1981). *Spline functions: basic theory*. Wiley, New York.

Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics 13*(2), 689–705.

Wang, G., L. Wang, and M.-J. Lai (2019). Package 'BPST'. *R package Version 1.0*.

Wang, L., G. Wang, M.-J. Lai, and L. Gao (2020). Efficient estimation of partially linear models for data on complicated domains by bivariate penalized splines over triangulations. *Statistica Sinica 30*(1), 347–369.

Wong, R. K., Y. Li, and Z. Zhu (2019). Partially linear functional additive models for multivariate functional data. *Journal of the American Statistical Association 114*(525), 406–418.

Xia, Y. (2008). A multiple-index model and dimension reduction. *Journal of the American Statistical Association 103*(484), 1631–1640.

Xia, Y., H. Tong, W. K. Li, and L.-X. Zhu (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B 64*(3), 363–410.

Yu, S., G. Wang, L. Wang, C. Liu, and L. Yang (2019). Estimation and inference for generalized geoadditive models. *Journal of the American Statistical Association 115*(530), 761–774.

Yu, S., G. Wang, L. Wang, and L. Yang (2021). Multivariate spline estimation and inference for image-on-scalar regression. *Statistica Sinica 31*(3), 1463–1487.

Zhong, Q., H. Lin, and Y. Li (2021). Cluster non-gaussian functional data. *Biometrics 77*(3), 852–865.

Zhou, L., J. Z. Huang, and R. J. Carroll (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika 95*(3), 601–619.

Zhu, H., F. Yao, and H. H. Zhang (2014). Structured functional additive regression in reproducing kernel hilbert spaces. *Journal of the Royal Statistical Society: Series B 76*(3), 581–603.

Zhu, Y., X. Shen, and C. Ye (2016). Personalized prediction and sparsity pursuit in latent factor models. *Journal of the American Statistical Association 111*(513), 241–252.