BIOMETRICS 000, 1–16 DOI: 000
000 0000

Supporting Information for 'Functional data analysis with covariate-dependent mean and covariance structures' by

Chenlin Zhang[1], Huazhen Lin[1*], Li Liu[2], Jin Liu[3] and Yi Li[4]

[1]Center of Statistical Research and School of Statistics,

Southwestern University of Finance and Economics, Chengdu, China

[2] School of Mathematics and Statistics, Wuhan University, Wuhan, China

[3] Centre for Quantitative Medicine,

Program in Health Services & Systems Research, Duke-NUS Medical School, Singapore

[4] Department of Biostatistics, University of Michigan, Ann Arbor, USA

*email: linhz@swufe.edu.cn

This paper has been submitted for consideration for publication in *Biometrics*

S1. Selection of initial values and tuning parameters

To implement the Algorithm, we need to supply reasonable initial values following four steps.

Step 1. We obtain the initial values $\boldsymbol{\gamma}^{(0)}$ and $\boldsymbol{\beta}^{(0)}$ for $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ by minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{Y_i(t_{ij})-\boldsymbol{\gamma}'\mathbf{B}_n(t_{ij},\mathbf{X}_i'\boldsymbol{\beta})\right\}^2,$$

that ignoring correlations and heterogeneity.

Step 2. Denote $R_i(t_{ij}) = Y_i(t_{ij}) - \boldsymbol{\gamma}^{(0)\prime}\mathbf{B}_n(t_{ij},\mathbf{X}_i'\boldsymbol{\beta}^{(0)})$. Based on a homogeneous variance model, that is, $\mathrm{var}(\xi_{ik}) = \rho_k$, we estimate the eigenfunctions $\phi_k^{(0)}(t) = \boldsymbol{\eta}_k^{(0)\prime}\mathbf{B}_{n1}(t)$ by applying the package of fpca for FPCA to $R_i(\cdot), i = 1,\cdots,n$.

Step 3. We obtain the initial ridge estimator of $\boldsymbol{\zeta}$, $\boldsymbol{\zeta}^{(0)}$, for fixed $\boldsymbol{\beta}^{(0)},\boldsymbol{\gamma}^{(0)}$ and $\phi_k^{(0)}$, by minimizing $-L_n(\boldsymbol{\pi},\boldsymbol{\zeta})+\lambda_{ini}\|\boldsymbol{\zeta}\|_F^2$ with respect to $\boldsymbol{\zeta}$, which is implemented by iteration solution of $\boldsymbol{\zeta}$ as the minimizer of $\frac{1}{2h_{ini}}\|\boldsymbol{\zeta}-\{\widetilde{\boldsymbol{\zeta}}+h_{ini}\dot{L}(\widetilde{\boldsymbol{\zeta}};\boldsymbol{\pi})\}\|_F^2+\lambda_{ini}\|\boldsymbol{\zeta}\|_F^2$, with $\widetilde{\boldsymbol{\zeta}}$ being the estimate of $\boldsymbol{\zeta}$ from the previous step. Then, we obtain $\boldsymbol{\alpha}^{(0)}$ by applying the package of MAVE, and obtain $\boldsymbol{\theta}^{(0)}$ by minimizing $\sum_{i=1}^{n}\sum_{k=1}^{K_n}\left\{\boldsymbol{\zeta}_{ik}-\boldsymbol{\theta}_k'\mathbf{B}_{n2}(\mathbf{X}_i'\boldsymbol{\alpha}_k)\right\}^2$.

Step 4. Finally, let $\boldsymbol{\Sigma}_{n_i}^{(0)} = \mathbf{B}_{n1}(\mathbf{t}_i)'\sum_{k=1}^{K_n}\boldsymbol{\eta}_k^{(0)}\{\boldsymbol{\theta}_k^{(0)\prime}\mathbf{B}_{n2}(\mathbf{X}_i'\boldsymbol{\alpha}_k^{(0)})\}^2\boldsymbol{\eta}_k^{(0)\prime}\mathbf{B}_{n1}(\mathbf{t}_i)$. Then, $\sigma^{2(0)} = \sum_{i=1}^{n}\mathrm{tr}\{(\boldsymbol{\Sigma}_{n_i}^{(0)} - R_i(\mathbf{t}_i)R_i(\mathbf{t}_i)')\}$.

We move on to detail the selection of the step lengths κ, h, ν, the number of FPCs K_n, and the tuning parameter λ. First, following Beck (2014), we take small constants for κ, h, ν to ensure the convergence of the algorithm. Second, as in James et al. (2000); Happ and Greven (2018), we choose K_n by using the proportion of variability explained by each principal component. Particularly, we choose K_n so that $\sum_{k=1}^{K_n}\rho_k^{(0)}/\sum_k\rho_k^{(0)} \geqslant 95\%$, where $\rho_k^{(0)}$ is the initial value for ρ_k, obtained by applying the package fpca to $R_i(\cdot), i = 1,\cdots,n$. We also note, unlike the traditional FPCA, the proposed estimation is less sensitive to the choice of K_n since we will further select FPCs for each individual using individual-specific penalties.

Third, we select λ by minimizing the generalized cross-validation (GCV) error:

$$GCV = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\{Y_i(t_{ij}) - \widehat{Y}_i(t_{ij})\}^2}{N_0(1-\mathrm{df}/N_0)^2},$$

where $N_0 = \sum_{i=1}^{n} n_i$ and $\mathrm{df} = p(1+K_n) + 2K_n m_{n1} + m_{n1}m_{n2} + 1$, the number of parameters in model (6) available to estimate $Y_i(t_{ij})$. We have confirmed the performance of our tuning procedure via simulations in Section 4.

S2. Supplementary materials in simulation and real data analysis

S2.1 *Supplementary results in simulation studies*

Example 1. Figures S1 and S2 plot the estimates of $\phi_k(t)$ and $\rho_k(u), k = 1, 2, 3$, and their 95% pointwise confidence intervals based on 500 bootstrap samples for various settings, where the solid, dashed and dotted lines, respectively, represent the true functions, the average of the estimated functions, and the 95% pointwise confidence bands. It appears that the average of the estimated functions coincides with the truth, with the confidence intervals covering the truth as well. As shown in Figure S3, all of the estimated mean functions overlap with the truth, and the performance improves with increased sample sizes.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

Tables S1 show the proposed estimator $\boldsymbol{\alpha}$ presents smaller biases and variances than SSV for both kinds of distributions.

[Table 1 about here.]

Example 2. Table S2 shows the calculated Type-1 error rates and power levels obtained under significant level 0.05, suggesting that the Type-1 error rate of all parameters is controlled at approximately 0.05, and the power tends to 1 as the sample size increases.

[Table 2 about here.]

Example 3. Table S3 presents the biases and empirical standard deviations for the parametric and non-parametric estimates obtained by FRIS and SSV for the SSV data.

[Table 3 about here.]

Example 4. Table S4 shows the p-values for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by FRIS, which suggest that the mean significantly depends on X_{i1} and X_{i3} and the covariance is associated with X_{i2} under the level of 0.05.

[Table 4 about here.]

Example 5. We consider a multiple index model satisfying

$$E(\mathbf{Y}_i|\mathbf{X}_i) = \boldsymbol{\mu}_i(\mathbf{X}_i) = \mu(\mathbf{t}_i, \mathbf{X}_i'\boldsymbol{\beta}_1, \mathbf{X}_i'\boldsymbol{\beta}_2), \mathrm{cov}(\mathbf{Y}_i|\mathbf{X}_i) = \boldsymbol{\Sigma}_i,$$

where $\boldsymbol{\Sigma}_i = \sum_{k=1}^{3} \phi_k(\mathbf{t}_i)\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_{k1}, \mathbf{X}_i'\boldsymbol{\alpha}_{k2})\phi_k(\mathbf{t}_i)' + \sigma^2\mathbf{I}_{n_i}$. $\mu(t, u_1, u_2) = 10t \cdot \{\exp(u_1) + \exp(u_2)\}$ and $\rho_k(u_1, u_2) = 10^{2-k}(u_1u_2)^2 I(u_1u_2 < 0), k = 1, 2, 3$. Set $\boldsymbol{\beta}_1 = (0.6, 0, 0.8)$, $\boldsymbol{\beta}_2 = (1, 0, 0)$, $\boldsymbol{\alpha}_{k1} = (0, 0.8, 0.6), \boldsymbol{\alpha}_{k2} = (0, 1, 0), k = 1, 2, 3$. Other setting is the same as that in Example 1(1). The setting in Example 5 implies that the covariates (X_{i1}, X_{i3}) are associated with the mean, and the covariates (X_{i2}, X_{i3}) are associated with the covariance of the functional response.

Table S5 summaries the estimates and the p-values for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by FRIS. From the p-values, it can be seen that the covariates X_{i1} and X_{i3} have significant contribution to the mean part, and X_{i2} and X_{i3} have significant association with the covariance under the level of 0.05. Hence, the Type I error is well-controlled for the multiple index model by FRIS.

[Table 5 about here.]

S2.2 *Supplementary results in real data analysis*

In this section, we present supplementary results to the real data. Figure S4 shows the estimated results of the binary function $\mu(t,u)$ of FRIS and SSV in real data. The two functions are very similar, indicating that our method is relatively accurate.

[Figure 4 about here.]

S3. Proof

S3.1 *Notation*

Let $\mathbb{P}_n$ be the empirical measure of $\{(\mathbf{Y}_i,\mathbf{X}_i) : i = 1,2,\ldots,n\}$, and $\mathbb{P}$ be the probability measure of $(\mathbf{Y},\mathbf{X})$. Define the log-likelihood as

$$l(\boldsymbol{\pi}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\{\mathbf{Y}-\boldsymbol{\mu}\}'\boldsymbol{\Sigma}^{-1}\{\mathbf{Y}-\boldsymbol{\mu}\},$$

where $\boldsymbol{\mu} = \mu(\mathbf{t},\mathbf{X}'\boldsymbol{\beta})$, $\boldsymbol{\Sigma} = \sum_{k=1}^{K_n}\phi_k(\mathbf{t})\rho_k(\mathbf{X}'\boldsymbol{\alpha}_k)\phi_k(\mathbf{t})' + \sigma^2\mathbf{I}$. Let $\boldsymbol{\vartheta} = (\boldsymbol{\beta}',\boldsymbol{\alpha}')'$, $\boldsymbol{\psi} = (\sigma^2,\mu,\phi_k,\rho_k)'$ and $\boldsymbol{\pi} = (\boldsymbol{\vartheta}',\boldsymbol{\psi}')$.

Define

$$\mathcal{H}_{r,d} = \left\{ f(\cdot) : |\frac{\partial^l f}{\partial x_1^{a_1}\ldots\partial x_d^{a_d}}(x) - \frac{\partial^l f}{\partial y_1^{a_1}\ldots\partial y_d^{a_d}}(y)| \leqslant c\|x-y\|^s, \text{ for any } x,y\in R^d \right\},$$

for $l\in\mathbb{N}_+$, $s\in(0,1]$ with $r = l+s$, for any $a = (a_1,\ldots,a_d)\in\mathbb{N}_+^d$ with $\sum_{j=1}^d a_j = l$, and for a $c>0$.

Let $\mathbf{h} = (1,h_1,\mathbf{h}_2',\mathbf{h}_3')' \in R\times\mathcal{H}_{r,2}\times\prod_{j=1}^{K_n}\mathcal{H}_{r,1}\times\prod_{j=1}^{K_n}\mathcal{H}_{r,1}$ with $\mathbf{h}_2 = (h_{2,1},\ldots,h_{2,K_n})'$ and $\mathbf{h}_3 = (h_{3,1},\ldots,h_{3,K_n})'$. Let $\dot{l}_{\boldsymbol{\alpha}} = (\dot{l}_{\boldsymbol{\alpha}_1}',\ldots,\dot{l}_{\boldsymbol{\alpha}_{K_n}}')'$, $\dot{l}_{\boldsymbol{\phi}}[\mathbf{h}_2] = (\dot{l}_{\phi_1}[h_{2,1}],\ldots,\dot{l}_{\phi_{K_n}}[h_{2,K_n}])'$, and $\dot{l}_{\boldsymbol{\rho}}[\mathbf{h}_3] = (\dot{l}_{\rho_1}[h_{3,1}],\ldots,\dot{l}_{\rho_{K_n}}[h_{3,K_n}])'$. Define $l_1(\boldsymbol{\pi}) = \partial l(\boldsymbol{\pi})/\partial\boldsymbol{\vartheta} = (\dot{l}_{\boldsymbol{\beta}}',\dot{l}_{\boldsymbol{\alpha}}')'$, and $l_2(\boldsymbol{\pi})[\mathbf{h}] =$

$\partial l(\boldsymbol{\vartheta}, \boldsymbol{\psi} + s\mathbf{h})/\partial s\big|_{s=0} = (\dot{l}_{\sigma^2}, \dot{l}_{\mu}[h_1], \dot{l}'_{\boldsymbol{\phi}}[\mathbf{h}_2], \dot{l}'_{\boldsymbol{\rho}}[\mathbf{h}_3])'$, where

$$\begin{aligned}
\dot{l}_{\boldsymbol{\beta}} &= \mathbf{X}\left\{\mathbf{Y}(\mathbf{t}) - \mu(\mathbf{t}, \mathbf{X}'\boldsymbol{\beta})\right\}' \boldsymbol{\Sigma}^{-1} \dot{\mu}_u(\mathbf{t}, \mathbf{X}'\boldsymbol{\beta}), \\
\dot{l}_{\boldsymbol{\alpha}_k} &= \mathbf{X}\rho_k(\mathbf{X}'\boldsymbol{\alpha}_k)\dot{\rho}_k(\mathbf{X}'\boldsymbol{\alpha}_k)\phi_k(\mathbf{t})' \times \\
&\quad \left[-\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\{\mathbf{Y}(\mathbf{t}) - \mu(\mathbf{t}, \mathbf{X}'\boldsymbol{\beta})\}\{\mathbf{Y}(\mathbf{t}) - \mu(\mathbf{t}, \mathbf{X}'\boldsymbol{\beta})\}'\boldsymbol{\Sigma}^{-1}\right] \phi_k(\mathbf{t}), \\
\dot{l}_{\sigma^2} &= tr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{\Sigma}^{-1}\{\mathbf{Y}(\mathbf{t}) - \mu(\mathbf{t}, \mathbf{X}'\boldsymbol{\beta})\}\{\mathbf{Y}(\mathbf{t}) - \mu(\mathbf{t}, \mathbf{X}'\boldsymbol{\beta})\}'\boldsymbol{\Sigma}^{-1}\right], \\
\dot{l}_{\mu}[h_1] &= \{\mathbf{Y}(\mathbf{t}) - \mu(\mathbf{t}, \mathbf{X}'\boldsymbol{\beta})\}'\boldsymbol{\Sigma}^{-1}h_1, \\
\dot{l}_{\phi_k}[h_{2,k}] &= \rho_k(\mathbf{X}'\boldsymbol{\alpha}_k)\phi_k(\mathbf{t})'\left[-\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\{\mathbf{Y}(\mathbf{t}) - \mu(\mathbf{t}, \mathbf{X}'\boldsymbol{\beta})\}\{\mathbf{Y}(\mathbf{t}) - \mu(\mathbf{t}, \mathbf{X}'\boldsymbol{\beta})\}'\boldsymbol{\Sigma}^{-1}\right] h_{2,k}, \\
\dot{l}_{\rho_k}[h_{3,k}] &= \phi_k(\mathbf{t})'\left[-\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\{\mathbf{Y}(\mathbf{t}) - \mu(\mathbf{t}, \mathbf{X}'\boldsymbol{\beta})\}\{\mathbf{Y}(\mathbf{t}) - \mu(\mathbf{t}, \mathbf{X}'\boldsymbol{\beta})\}'\boldsymbol{\Sigma}^{-1}\right] \phi_k(\mathbf{t})h_{3,k},
\end{aligned}$$

for $k = 1, \ldots, K_n$.

Define $\boldsymbol{\Pi}_n^{\delta} = \mathcal{A}^{\delta} \times \boldsymbol{\Theta}_n^{\delta} \times \boldsymbol{\Theta}_{1n}^{\delta} \times \boldsymbol{\Theta}_{2n}^{\delta}$, where

$$\mathcal{A}^{\delta} = \{(\boldsymbol{\beta}', \boldsymbol{\alpha}', \sigma^2) : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \delta, \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| < \delta, \|\sigma^2 - \sigma_0^2\| < \delta, (\boldsymbol{\beta}', \boldsymbol{\alpha}', \sigma^2) \in \mathcal{A}\},$$

$$\boldsymbol{\Theta}_n^{\delta} = \{\mu(t, u) : \|\mu(t, u) - \mu_0(t, u)\| < \delta, \mu \in \boldsymbol{\Theta}\},$$

$$\boldsymbol{\Theta}_{1n}^{\delta} = \left\{\boldsymbol{\phi}(t) : \|\boldsymbol{\phi}(t) - \boldsymbol{\phi}_0(t)\| < \delta, \boldsymbol{\phi}(t) \in \prod_{k=1}^{K_n} \boldsymbol{\Theta}_{1k}\right\},$$

$$\boldsymbol{\Theta}_{2n}^{\delta} = \left\{\boldsymbol{\rho}(u) : \|\boldsymbol{\rho}(u) - \boldsymbol{\rho}_0(u)\| < \delta, \boldsymbol{\rho}(u) \in \prod_{k=1}^{K_n} \boldsymbol{\Theta}_{2k}\right\}.$$

Let $\|\boldsymbol{v}\|$ denote the ℓ_2 norm for a vector $\boldsymbol{v}$ and $\|f\| = (\int_0^T f^2(t)dt)^{1/2}$ for any function $f(\cdot)$ with a domain $[0, T]$; $a \preceq b$ means $c_1 \leqslant a/b \leqslant c_2$, where $c_2 > c_1 > 0$ are constants and $a \ll b$ means $a/b \to 0$.

S3.2 *Proofs of The Main Results*

LEMMA 1: *For $\mu_0 \in \mathcal{H}_{r,2}, \phi_{k0}, \rho_{k0} \in \mathcal{H}_{r,1}$, $1 \leqslant k \leqslant K_n$, there exists smoothing spline $\mu_n, \phi_{nk}, \rho_{nk}$ such that $\|\mu_n - \mu_0\|_\infty = O(m_n^{-r})$, $\|\phi_{nk} - \phi_{k0}\|_\infty = O(m_n^{-r})$, $\|\rho_{nk} - \rho_{k0}\|_\infty = O(m_n^{-r})$. Moreover, we have $\|\boldsymbol{\pi}_n - \boldsymbol{\pi}_0\|_\infty = O(m_n^{-r})$, where $\boldsymbol{\pi}_n = (\boldsymbol{\beta}, \boldsymbol{\alpha}', \sigma^2, \mu_n, \boldsymbol{\phi}_n', \boldsymbol{\rho}_n')' \in \boldsymbol{\Pi}_n$.*

Proof: The proof follows from Schumaker (1981).

LEMMA 2: *Let $N(\varepsilon, \mathcal{F}, d)$ denote the covering number with respect to a semi-metric d of function class $\mathcal{F}$. Under Conditions (C1)-(C3), the covering number of the class $\mathcal{L}_n(\delta) = \{l(\boldsymbol{\pi}) : \boldsymbol{\pi} \in \boldsymbol{\Pi}_n^\delta\}$ satisfies*

$$N(\varepsilon, \mathcal{L}_n(\delta), \|\cdot\|) \preceq (\delta/\varepsilon)^{(K_n m_n + m_n^2)}.$$

Proof: The proof is similar to that in Ma et al. (2015), and is omitted.

LEMMA 3: *Under (C1)-(C4), we have for any $\delta \leqslant 1$,*

$$\sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}_n^\delta} |\mathbb{P}_n l(\boldsymbol{\pi}) - \mathbb{P} l(\boldsymbol{\pi})| \to 0$$

almost surely.

Proof: Since $\boldsymbol{\pi} \in \boldsymbol{\Pi}_n^\delta$, then $|l(\boldsymbol{\pi})|$ is bounded. Without loss of generality, we assume $\sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}_n^\delta} |l(\boldsymbol{\pi})| \leqslant 1$. It then follows that $\mathbb{P} l^2(\boldsymbol{\pi}) \leqslant \mathbb{P} \sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}_n^\delta} |l(\boldsymbol{\pi})|^2 \leqslant 1$. Let $\max\{(v+\tau)/2, v\} < \phi < 1/2$, $\alpha_n = n^{-1/2+\phi} \log(n)^{1/2}$, and $\varepsilon_n = \varepsilon \alpha_n$ for fixed $\varepsilon > 0$. Then for any $l(\boldsymbol{\pi}) \in \mathcal{L}_n(\delta)$ and sufficiently large n, we have

$$Var\{\mathbb{P}_n l(\boldsymbol{\pi})/(4\varepsilon_n)\} \leqslant \frac{\mathbb{P} l^2(\boldsymbol{\pi})}{16\varepsilon^2 n \alpha_n^2} \ll \frac{1}{16\varepsilon^2 \log n} \leqslant \frac{1}{2}. \tag{1}$$

By inequality (31) and lemma33 in Pollard (1984) and Lemma 2, it yields

$$\begin{aligned}
&P(\sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}_n^\delta} |\mathbb{P}_n l(\boldsymbol{\pi}) - \mathbb{P} l(\boldsymbol{\pi})| > 8\varepsilon_n) \\
&\leqslant 8N(\varepsilon, \mathcal{L}_n(\delta), \|\cdot\|) \exp(-\frac{n\varepsilon_n^2}{128}) P(\sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}_n^\delta} |n^{-1} \sum_{i=1}^n l^2(\boldsymbol{\pi}_n)| \leqslant 64) + P(\sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}_n^\delta} |n^{-1} \sum_{i=1}^n l^2(\boldsymbol{\pi}_n)| > 64) \\
&\preceq \varepsilon_n^{-(K_n m_n + m_n^2)} \exp(-\frac{n\varepsilon^2 \alpha_n^2}{128}) \\
&= \exp\{(K_n m_n + m_n^2) \log(\frac{1}{\varepsilon n^{-1/2+\phi} \log(n)^{1/2}}) - \frac{1}{128} n\varepsilon^2 n^{-1+2\phi} \log n\} \\
&= \exp\left[(K_n m_n + m_n^2)\left\{(\frac{1}{2} - \phi)\log n - \frac{1}{2}\log\log n + \log(1/\varepsilon)\right\} - \frac{1}{128} n\varepsilon^2 n^{-1+2\phi} \log(n)\right] \\
&\preceq \exp(-c^* n^{2\phi} \log n),
\end{aligned} \tag{2}$$

where c^* is a constant. Hence,

$$\sum_{n=1}^{\infty} P(\sup_{\boldsymbol{\pi}\in\boldsymbol{\Pi}_n^{\delta}} |\mathbb{P}_n l(\boldsymbol{\pi}) - \mathbb{P}l(\boldsymbol{\pi})| > 8\varepsilon_n) < \infty. \tag{3}$$

By the Borel-Cantelli lemma, we have

$$P\left[\cap_{k=1}^{\infty} \cup_{n=k}^{\infty} \{\sup_{\boldsymbol{\pi}\in\boldsymbol{\Pi}_n^{\delta}} |\mathbb{P}_n l(\boldsymbol{\pi}) - \mathbb{P}l(\boldsymbol{\pi})| > 8\varepsilon_n\}\right] = 0,$$

meaning that the probability that events $\{\sup_{\boldsymbol{\pi}\in\boldsymbol{\Pi}_n^{\delta}} |\mathbb{P}_n l(\boldsymbol{\pi}) - \mathbb{P}l(\boldsymbol{\pi})| > 8\varepsilon_n\}$ will occur for infinite many times is 0. Since $\varepsilon_n \to 0$, this completes the proof.

S3.2.1 *Proof of Theorem 3.1.* Define $\mathcal{N}_\epsilon = \{\boldsymbol{\pi} : \epsilon_0 > \|\boldsymbol{\pi} - \boldsymbol{\pi}_0\| \geqslant \epsilon, \boldsymbol{\pi} \in \boldsymbol{\Pi}_n^{or}\}$ for some $\epsilon_0 \leqslant 1$ and any $0 < \epsilon < \epsilon_0$. Then we have

$$\begin{aligned} \sup_{\mathcal{N}_\epsilon} \mathbb{P}l(\boldsymbol{\pi}) &= \sup_{\mathcal{N}_\epsilon}\{\mathbb{P}l(\boldsymbol{\pi}) - \mathbb{P}_n l(\boldsymbol{\pi}) + \mathbb{P}_n l(\boldsymbol{\pi})\} \\ &\geqslant -\sup_{\mathcal{N}_\epsilon} |\mathbb{P}_n l(\boldsymbol{\pi}) - \mathbb{P}l(\boldsymbol{\pi})| + \sup_{\mathcal{N}_\epsilon} \mathbb{P}_n l(\boldsymbol{\pi}) = -H_1 + \sup_{\mathcal{N}_\epsilon} \mathbb{P}_n l(\boldsymbol{\pi}), \end{aligned} \tag{4}$$

where $H_1 = \sup_{\mathcal{N}_\epsilon} |\mathbb{P}_n l(\boldsymbol{\pi}) - \mathbb{P}l(\boldsymbol{\pi})|$. For $\widehat{\boldsymbol{\pi}}_n^{or} \in \mathcal{N}_\epsilon$, we have

$$\sup_{\mathcal{N}_\epsilon} \mathbb{P}_n l(\boldsymbol{\pi}) = \mathbb{P}_n l(\widehat{\boldsymbol{\pi}}_n^{or}) \geqslant \mathbb{P}_n l(\boldsymbol{\pi}_0) = -H_2 + \mathbb{P}l(\boldsymbol{\pi}_0), \tag{5}$$

where $H_2 = \mathbb{P}l(\boldsymbol{\pi}_0) - \mathbb{P}_n l(\boldsymbol{\pi}_0)$. Thus, (4) and (5) give

$$\mathbb{P}l(\boldsymbol{\pi}_0) - \sup_{\mathcal{N}_\epsilon} \mathbb{P}l(\boldsymbol{\pi}) \leqslant H_1 + H_2. \tag{6}$$

By Jensen's inequality,

$$\mathbb{P}l(\boldsymbol{\pi}) - \mathbb{P}l(\boldsymbol{\pi}_0) \leqslant \log \mathbb{P}\{\frac{f(\boldsymbol{\pi})}{f(\boldsymbol{\pi}_0)}\} = \log \int \frac{f(\boldsymbol{\pi})}{f(\boldsymbol{\pi}_0)} f(\boldsymbol{\pi}_0) dx = 0,$$

where the equality holds if and only if $\boldsymbol{\pi} = \boldsymbol{\pi}_0$. Let $\delta_\epsilon = \mathbb{P}l(\boldsymbol{\pi}_0) - sup_{\mathcal{N}_\epsilon} \mathbb{P}l(\boldsymbol{\pi}) > 0$. Since $\{\widehat{\boldsymbol{\pi}}_n^{or} \in \mathcal{N}_\epsilon\} \subseteq \{H_1 + H_2 \geqslant \delta_\epsilon\}$, it follows that $H_1 = o(1), H_2 = o(1)$ almost surely, by Lemma 3 and the strong Law of Large Numbers. Therefore, $\|\widehat{\boldsymbol{\pi}}_n^{or} - \boldsymbol{\pi}_0\|_2 = o(1)$ almost surely.

We next conclude the convergence rate by verifying the conditions of Theorem 3.2.5 of van der Varrt and Wellner (1996). To that end, we define a function class

$$\mathcal{L}_\epsilon = \{l(\boldsymbol{\pi}) - l(\boldsymbol{\pi}_0), \|\boldsymbol{\pi} - \boldsymbol{\pi}_0\| \leqslant \epsilon, \boldsymbol{\pi} \in \boldsymbol{\Pi}_n^{or}\}.$$

It can be seen that $\log N_{[]}(\varepsilon, \mathcal{L}_\epsilon, \|\cdot\|) \preceq (K_n m_n + m_n^2)\log(\frac{\epsilon}{\varepsilon})$. Then the bracketing integral

$$J_{[]}(\varepsilon, \mathcal{L}_\epsilon, \|\cdot\|) = \int_0^\epsilon \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{L}_\epsilon, \|\cdot\|)}d\varepsilon \preceq \sqrt{K_n m_n + m_n^2}\epsilon.$$

By Lemma 3.4.2 of van der Varrt and Wellner (1996), we have

$$\begin{aligned} E\Big(\sup_{\mathcal{L}_\epsilon} |\sqrt{n}(\mathbb{P}_n - \mathbb{P})(l(\boldsymbol{\pi}) - l(\boldsymbol{\pi}_0))|\Big) &\preceq J_{[]}(\varepsilon, \mathcal{L}_\epsilon, \|\cdot\|)\Big(1 + \frac{J_{[]}(\varepsilon, \mathcal{L}_\epsilon, \|\cdot\|)}{\epsilon^2\sqrt{n}} M_0\Big) \\ &\preceq O(\sqrt{K_n m_n + m_n^2}\epsilon + \frac{K_n m_n + m_n^2}{\sqrt{n}}). \end{aligned} \tag{7}$$

This shows that the function $\phi_n(\epsilon)$ in Theorem 3.2.5 of van der Varrt and Wellner (1996) is given by $\phi_n(\epsilon) = \sqrt{K_n m_n + m_n^2}\epsilon + \frac{K_n m_n + m_n^2}{\sqrt{n}}$. Obviously, $\phi_n(\epsilon)/\epsilon$ is decreasing in ϵ and $r_n^2\phi_n(\frac{1}{r_n}) = r_n\sqrt{K_n m_n + m_n^2} + r_n^2(K_n m_n + m_n^2)/\sqrt{n} \leqslant \sqrt{n}$ for every n, which implies $r_n \leqslant \sqrt{n}/\sqrt{K_n m_n + m_n^2}$.

Besides, we need to show that $\widehat{\boldsymbol{\pi}}_n^{or}$ satisfies $\mathbb{P}_n l(\widehat{\boldsymbol{\pi}}_n^{or}) \geqslant \mathbb{P}_n l(\boldsymbol{\pi}_0) - O_p(r_n^{-2})$. Note that

$$\begin{aligned} \mathbb{P}_n l(\widehat{\boldsymbol{\pi}}_n^{or}) - \mathbb{P}_n(\boldsymbol{\pi}_0) &= \mathbb{P}_n(l(\widehat{\boldsymbol{\pi}}_n^{or}) - l(\boldsymbol{\pi}_n)) + (\mathbb{P}_n - \mathbb{P})(l(\boldsymbol{\pi}_n) - l(\boldsymbol{\pi}_0)) + \mathbb{P}(l(\boldsymbol{\pi}_n) - l(\boldsymbol{\pi}_0)) \\ &= I_1 + I_2 + I_3, \end{aligned}$$

where $I_1 > 0$. Define $\widetilde{L}(\boldsymbol{\pi}) = \{\frac{l(\boldsymbol{\pi}) - l(\boldsymbol{\pi}_0)}{n^{-rv+\tau/2+\epsilon}}, \boldsymbol{\pi} \in \boldsymbol{\Pi}_n^{or}\}$, which is a P-Donsker class by Lemma 1. Therefore, $I_2 = O_p(n^{-rv+\tau/2+\epsilon}n^{-1/2})$ and $I_3 \geqslant -O_p(n^{-2rv+\tau})$. Since $\epsilon < 1/2 - vr + \tau/2$, it follows $\mathbb{P}_n l(\widehat{\boldsymbol{\pi}}_n^{or}) - \mathbb{P}_n(\boldsymbol{\pi}_0) \geqslant -O_p(n^{-2rv+\tau})$ for $r_n \leqslant n^{vr-\tau/2}$.

Finally, we obtain that

$$\|\widehat{\boldsymbol{\pi}}_n^{or} - \boldsymbol{\pi}_0\| = O_p(r_n^{-1}) = O_p(\delta_n)$$

with $\delta_n = n^{-(1-2v)/2} + n^{-(1-v-\tau)/2} + n^{-(vr-\tau/2)}$ by noting $K_n = n^\tau$. This completes the proof.

S3.2.2 *Proof of Theorem 3.2.* Denote

$$S_{1n}(\boldsymbol{\pi}) = \mathbb{P}_n l_1(\boldsymbol{\pi}), S_{2n}(\boldsymbol{\pi})[\mathbf{h}] = \mathbb{P}_n l_2(\boldsymbol{\pi})[\mathbf{h}],$$

$$S_1(\boldsymbol{\pi}) = \mathbb{P} l_1(\boldsymbol{\pi}), S_2(\boldsymbol{\pi})[\mathbf{h}] = \mathbb{P} l_2(\boldsymbol{\pi})[\mathbf{h}].$$

From the definitions of $l_1(\boldsymbol{\pi})$ and $l_2(\boldsymbol{\pi})$, it follows

$$S_1(\boldsymbol{\pi}_0) = 0 \quad \text{and} \quad S_2(\boldsymbol{\pi}_0)[\mathbf{h}] = 0.$$

Define

$$\dot{S}_{11}(\boldsymbol{\pi}) = -\mathbb{P} l_1(\boldsymbol{\pi}) l_1(\boldsymbol{\pi})',$$

$$\dot{S}_{12}(\boldsymbol{\pi})[\mathbf{h}] = \dot{S}_{21}'(\boldsymbol{\pi})[\mathbf{h}] - \mathbb{P} l_1(\boldsymbol{\pi}) l_2(\boldsymbol{\pi})[\mathbf{h}],$$

$$\dot{S}_{22}(\boldsymbol{\pi})[\mathbf{h}^*, \mathbf{h}] = -\mathbb{P} l_2(\boldsymbol{\pi})[\mathbf{h}^*] l_2'(\boldsymbol{\pi})[\mathbf{h}]$$

for $\mathbf{h}^*, \mathbf{h} \in R \times \mathcal{H}_{r,2} \times \prod_{j=1}^{K_n} \mathcal{H}_{r,1} \times \prod_{j=1}^{K_n} \mathcal{H}_{r,1}$. To prove the theorem, we need to show

(a) There is an $\mathbf{h}^* \in R \times \mathcal{H}_{r,2} \times \prod_{j=1}^{K_n} \mathcal{H}_{r,1} \times \mathcal{H}_{r,1}^{or}$ such that

$$\dot{S}_{12}(\boldsymbol{\pi}_0)[\mathbf{h}] - \dot{S}_{22}(\boldsymbol{\pi}_0)[\mathbf{h}^*, \mathbf{h}] = 0, \tag{8}$$

and

$$I(\boldsymbol{\vartheta}) = -\dot{S}_{11}(\boldsymbol{\pi}) + \dot{S}_{21}(\boldsymbol{\pi})[\mathbf{h}^*] \tag{9}$$

is the information for estimation of $\boldsymbol{\vartheta}$ with $I^{-1}(\boldsymbol{\vartheta})$ being the information bound for $\mathbf{h} \in R \times \mathcal{H}_{r,2} \times \prod_{j=1}^{K_n} \mathcal{H}_{r,1} \times \prod_{j=1}^{K_n} \mathcal{H}_{r,1}$.

(b) Stochastic equicontinuity, i.e.,

$$\sup_{\|\boldsymbol{\pi}-\boldsymbol{\pi}_0\| \leqslant \delta_n} |\sqrt{n}(S_{1n} - S_1)(\boldsymbol{\pi}) - \sqrt{n}(S_{1n} - S_1)(\boldsymbol{\pi}_0)| = o_p(1), \tag{10}$$

and

$$\sup_{\|\boldsymbol{\pi}-\boldsymbol{\pi}_0\|\leqslant\delta_n} |\sqrt{n}(S_{2n}-S_2)(\boldsymbol{\pi})[\mathbf{h}^*] - \sqrt{n}(S_{2n}-S_2)(\boldsymbol{\pi}_0)[\mathbf{h}^*]| = o_p(1), \tag{11}$$

when $0 < v < 1/4, \tau < \min\{1/2 - v, 2v(r-1), v(2r-1)/2\}$.

(c) In a neighborhood $\{\boldsymbol{\pi} : \|\boldsymbol{\pi} - \boldsymbol{\pi}_0\| \leqslant \delta_n\}$ and for $r > 1$,

$$\begin{aligned} & S_1(\boldsymbol{\pi}) - S_1(\boldsymbol{\pi}_0) - \dot{S}_{11}(\boldsymbol{\pi}_0)[\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0] - \dot{S}_{12}(\boldsymbol{\pi}_0)[\boldsymbol{\psi} - \boldsymbol{\psi}_0] \\ = \;& o(\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\|) + O(\|\boldsymbol{\psi} - \boldsymbol{\psi}_0\|^2), \end{aligned} \tag{12}$$

and

$$\begin{aligned} & S_2(\boldsymbol{\pi})[\mathbf{h}^*] - S_2(\boldsymbol{\pi}_0)[\mathbf{h}^*] - \dot{S}_{21}(\boldsymbol{\pi}_0)[\mathbf{h}^*](\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) - \dot{S}_{22}(\boldsymbol{\pi}_0)[\mathbf{h}^*, \boldsymbol{\psi} - \boldsymbol{\psi}_0] \\ = \;& o(\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\|) + O(\|\boldsymbol{\psi} - \boldsymbol{\psi}_0\|^2). \end{aligned} \tag{13}$$

We first verify (a). In fact,

$$\begin{aligned} \dot{S}_{12}(\boldsymbol{\pi}_0)[\mathbf{h}] - \dot{S}_{22}(\boldsymbol{\pi}_0)[\mathbf{h}^*, \mathbf{h}] &= -\mathbb{P}l_1(\boldsymbol{\pi}_0)l_2(\boldsymbol{\pi}_0)[\mathbf{h}] + \mathbb{P}l_2(\boldsymbol{\pi}_0)[\mathbf{h}^*]l_2(\boldsymbol{\pi}_0)[\mathbf{h}] \\ &= -\mathbb{P}\{l_1(\boldsymbol{\pi}_0) - l_2(\boldsymbol{\pi}_0)[\mathbf{h}^*]\}l_2(\boldsymbol{\pi}_0)[\mathbf{h}], \end{aligned}$$

where $\mathbf{h}^* = E(l_1(\boldsymbol{\pi}_0)|\mathbf{X}, \mathbf{Y})/E(l_2(\boldsymbol{\pi}_0)|\mathbf{X}, \mathbf{Y})$ satisfies $\dot{S}_{12}(\boldsymbol{\pi}_0)[\mathbf{h}] - \dot{S}_{22}(\boldsymbol{\pi}_0)[\mathbf{h}^*, \mathbf{h}] = 0$.

Denote $l^*(\boldsymbol{\pi}) = l_1(\boldsymbol{\pi}) - l_2(\boldsymbol{\pi})[\mathbf{h}^*]$ to be the efficient score for $\boldsymbol{\vartheta}$. Then

$$\begin{aligned} I(\boldsymbol{\vartheta}) &= -\dot{S}_{11}(\boldsymbol{\pi}) + \dot{S}_{21}(\boldsymbol{\pi})[\mathbf{h}^*] \\ &= \mathbb{P}l_1(\boldsymbol{\pi})l_1(\boldsymbol{\pi}) - \mathbb{P}l_2(\boldsymbol{\pi})[\mathbf{h}^*]l_1(\boldsymbol{\pi}) \\ &= \mathbb{P}\{l_1(\boldsymbol{\pi}) - l_2(\boldsymbol{\pi})[\mathbf{h}^*]\}l_1(\boldsymbol{\pi}) \\ &= \mathbb{P}\{l_1(\boldsymbol{\pi}) - l_2(\boldsymbol{\pi})[\mathbf{h}^*]\}\{l_1(\boldsymbol{\pi}) - l_2(\boldsymbol{\pi})[\mathbf{h}^*]\} \\ &= \mathbb{P}\{l^*(\boldsymbol{\pi})\}^{\otimes 2}. \end{aligned}$$

Then for (b), by the definitions of S_{1n}, S_1, S_{2n}, and S_2, we have

$$\sqrt{n}(S_{1n}-S_1)(\boldsymbol{\pi})-\sqrt{n}(S_{1n}-S_1)(\boldsymbol{\pi}_0)=\sqrt{n}(\mathbb{P}_n-\mathbb{P})(l_1(\boldsymbol{\pi})-l_1(\boldsymbol{\pi}_0)),$$

and

$$\sqrt{n}(S_{2n}-S_2)(\boldsymbol{\pi})[\mathbf{h}^*]-\sqrt{n}(S_{2n}-S_2)(\boldsymbol{\pi}_0)[\mathbf{h}^*]=\sqrt{n}(\mathbb{P}_n-\mathbb{P})(l_2(\boldsymbol{\pi})[\mathbf{h}^*]-l_2(\boldsymbol{\pi}_0)[\mathbf{h}^*]).$$

Considering the classes of functions $\{l_1(\boldsymbol{\pi}):\|\boldsymbol{\pi}-\boldsymbol{\pi}_0\|\leqslant\delta_n\}$ and $\{l_2(\boldsymbol{\pi})[\mathbf{h}^*]:\|\boldsymbol{\pi}-\boldsymbol{\pi}_0\|\leqslant\delta_n\}$, it yields

$$E\Big(\sup_{\|\boldsymbol{\pi}-\boldsymbol{\pi}_0\|\leqslant\delta_n}|\sqrt{n}(\mathbb{P}_n-\mathbb{P})(l_1(\boldsymbol{\pi})-l_1(\boldsymbol{\pi}_0))|\Big)\preceq O(\sqrt{K_nm_n+m_n^2}\delta_n+\frac{K_nm_n+m_n^2}{\sqrt{n}})=o(1),$$

when when $0<v<1/4,\tau<\min\{1/2-v,2v(r-1),v(2r-1)/2\}$ for $r>1$. Thus, we have

$$\sup_{\|\boldsymbol{\pi}-\boldsymbol{\pi}_0\|\leqslant\delta_n}|\sqrt{n}(\mathbb{P}_n-\mathbb{P})(l_1(\boldsymbol{\pi})-l_1(\boldsymbol{\pi}_0))|\to 0.$$

Similarly,

$$\sup_{\|\boldsymbol{\pi}-\boldsymbol{\pi}_0\|\leqslant\delta_n}|\sqrt{n}(\mathbb{P}_n-\mathbb{P})(l_2(\boldsymbol{\pi})[\mathbf{h}^*]-l_2(\boldsymbol{\pi}_0)[\mathbf{h}^*])|\to 0.$$

Finally, for (c),

$$\begin{aligned}S_1(\boldsymbol{\pi})-S_1(\boldsymbol{\pi}_0) &= \dot S_{11}(\boldsymbol{\pi}_0)[\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0]+\dot S_{12}(\boldsymbol{\pi}_0)[\boldsymbol{\psi}-\boldsymbol{\psi}_0]+\\ &\quad\Big\{(\dot S_{11}(\boldsymbol{\pi})[\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0]+\dot S_{12}(\boldsymbol{\pi})[\boldsymbol{\psi}-\boldsymbol{\psi}_0])-(\dot S_{11}(\boldsymbol{\pi}_0)[\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0]+\dot S_{12}(\boldsymbol{\pi}_0)[\boldsymbol{\psi}-\boldsymbol{\psi}_0])\Big\}\\ &= \dot S_{11}(\boldsymbol{\pi}_0)[\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0]+\dot S_{12}(\boldsymbol{\pi}_0)[\boldsymbol{\psi}-\boldsymbol{\psi}_0]+O(\|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\|^2)+O(\|\boldsymbol{\psi}-\boldsymbol{\psi}_0\|^2).\end{aligned}$$

Similarly,

$$\begin{aligned}& S_2(\boldsymbol{\pi})[\mathbf{h}^*]-S_2(\boldsymbol{\pi}_0)[\mathbf{h}^*]-\dot S_{21}(\boldsymbol{\pi}_0)[\mathbf{h}^*](\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0)-\dot S_{22}(\boldsymbol{\pi}_0)[\mathbf{h}^*,\boldsymbol{\psi}-\boldsymbol{\psi}_0]\\ = {}& O(\|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\|^2)+O(\|\boldsymbol{\psi}-\boldsymbol{\psi}_0\|^2).\end{aligned}$$

Now we are positioned to conclude the normality of $\widehat{\boldsymbol{\vartheta}}_n^{or}$. First, note that

$$S_{1n}(\widehat{\boldsymbol{\pi}}_n^{or})=0, S_{2n}(\widehat{\boldsymbol{\pi}}_n^{or})[\mathbf{h}_n^*]=0 \quad\text{and}\quad S_1(\boldsymbol{\pi}_0)=0, S_2(\boldsymbol{\pi}_0)[\mathbf{h}^*]=0. \tag{14}$$

For $\mathbf{h}_n^* \in R \times \boldsymbol{\Theta} \times \prod_{k=1}^{K_n} \boldsymbol{\Theta}_{1k} \times \boldsymbol{\Theta}_{2n}^{or}$ and $\|\mathbf{h}^* - \mathbf{h}_n^*\| = O(n^{-vr+\alpha/2})$, under the condition in (b) and by some entropy calculation, we have $S_{2n}(\widehat{\boldsymbol{\pi}}_n^{or})[\mathbf{h}^*] = o_p(n^{-1/2})$.

Then we see that

$$\dot{S}_{11}(\widehat{\boldsymbol{\vartheta}}_n^{or} - \boldsymbol{\vartheta}_0) + \dot{S}_{12}[\widehat{\boldsymbol{\psi}}_n^{or} - \boldsymbol{\psi}_0] + o(\|\widehat{\boldsymbol{\vartheta}}_n^{or} - \boldsymbol{\vartheta}_0\|) + O(\|\widehat{\boldsymbol{\psi}}_n^{or} - \boldsymbol{\psi}_0\|^2) - S_{1n}(\boldsymbol{\pi}_0) = o_p(n^{-1/2}),$$
$$\dot{S}_{21}[\mathbf{h}^*](\widehat{\boldsymbol{\vartheta}}_n^{or} - \boldsymbol{\vartheta}_0) + \dot{S}_{12}[\mathbf{h}^*, \widehat{\boldsymbol{\psi}}_n^{or} - \boldsymbol{\psi}_0] + o(\|\widehat{\boldsymbol{\vartheta}}_n^{or} - \boldsymbol{\vartheta}_0\|) + O(\|\widehat{\boldsymbol{\psi}}_n^{or} - \boldsymbol{\psi}_0\|^2) - S_{2n}(\boldsymbol{\pi}_0)[\mathbf{h}^*] = o_p(n^{-1/2}),$$
(1

by (14), (b) and (c).

We then have

$$(\dot{S}_{11} - \dot{S}_{21}[\mathbf{h}^*])(\widehat{\boldsymbol{\vartheta}}_n^{or} - \boldsymbol{\vartheta}_0) + o(\|\widehat{\boldsymbol{\vartheta}}_n^{or} - \boldsymbol{\vartheta}_0\|) = S_{1n}(\boldsymbol{\pi}_0) - S_{2n}(\boldsymbol{\pi}_0)[\mathbf{h}^*] + o_p(n^{-1/2}),$$

by (a) and (15). This yields that

$$\begin{aligned}\sqrt{n}(\dot{S}_{11} - \dot{S}_{21}[\mathbf{h}^*])(\widehat{\boldsymbol{\vartheta}}_n^{or} - \boldsymbol{\vartheta}_0) &= \sqrt{n}\{S_{1n}(\boldsymbol{\pi}_0) - S_{2n}(\boldsymbol{\pi}_0)[\mathbf{h}^*]\} + o_p(1) \\ &= \sqrt{n}\mathbb{P}_n l^*(\boldsymbol{\pi}_0) + o_p(1).\end{aligned}$$

Besides, note that when $n^{\tau-1/2}/\Lambda = o_p(1)$, where $\Lambda = \lambda_{\min}(I(\boldsymbol{\vartheta}_0))$,

$$\mathbf{u}' I(\boldsymbol{\vartheta}_0)^{1/2} \mathbb{P}_n l^*(\boldsymbol{\pi}_0)^{\otimes 2} I(\boldsymbol{\vartheta}_0)^{1/2}\mathbf{u} = 1 + \mathbf{u}' I(\boldsymbol{\vartheta}_0)^{1/2}(\mathbb{P}_n l^*(\boldsymbol{\pi}_0)^{\otimes 2} - I(\boldsymbol{\vartheta}_0)) I(\boldsymbol{\vartheta}_0)^{1/2}\mathbf{u},$$

where the second term is bounded by

$$\|I(\boldsymbol{\vartheta}_0)^{1/2}\|\|\mathbb{P}_n l^*(\boldsymbol{\pi}_0)^{\otimes 2} - I(\boldsymbol{\vartheta}_0)\|\|I(\boldsymbol{\vartheta}_0)^{1/2}\| = \Lambda^{-1/2} O_p(n^{\tau} n^{-1/2})\Lambda^{-1/2} = o_p(1).$$

Therefore, it follows for any vector $\mathbf{u}$ with $\|\mathbf{u}\| = 1$,

$$\sqrt{n}\mathbf{u}' I(\boldsymbol{\vartheta}_0)^{1/2}(\widehat{\boldsymbol{\vartheta}}_n^{or} - \boldsymbol{\vartheta}_0) \to N(0, 1).$$

This completes the proof of Theorem 3.2.

S3.2.3 *Proof of Theorem 3.3.* We show that $\widehat{\boldsymbol{\pi}}_n^{or}$ is a strict minimum of $Q_n(\boldsymbol{\pi})$ for $\boldsymbol{\pi} \in \boldsymbol{\Pi}_n$ with probability approaching 1 through the following two steps.

(1) For any $\boldsymbol{\pi} \in \boldsymbol{\Pi}_n^{\delta_n}$, denote $\boldsymbol{\pi}^* = (\boldsymbol{\beta}', \boldsymbol{\alpha}^{*\prime}, \sigma^2, \mu, \boldsymbol{\rho}^*, \boldsymbol{\phi})' \in \boldsymbol{\Pi}_n^{or}$, where $\boldsymbol{\rho}^* = T(\boldsymbol{\rho})$ and $\rho_{ik}^* = \rho_k^*(\mathbf{X}_i'\boldsymbol{\alpha}_k^*)$,

$$Q_n(\boldsymbol{\pi}^*) \leqslant Q_n(\widehat{\boldsymbol{\pi}}_n^{or}),$$

with the equality only when $\boldsymbol{\pi}^* = \widehat{\boldsymbol{\pi}}_n^{or}$.

(2) Define $\widetilde{\boldsymbol{\Pi}}_n = \{\boldsymbol{\pi} : \|\boldsymbol{\pi} - \widehat{\boldsymbol{\pi}}_n^{or}\| \leqslant t_n, \boldsymbol{\pi} \in \boldsymbol{\Pi}_n^{\delta}\}$, where t_n is a positive sequence. For any $\boldsymbol{\pi} \in \boldsymbol{\Pi}_n^{\delta_n} \bigcup \widetilde{\boldsymbol{\Pi}}_n$,

$$Q_n(\boldsymbol{\pi}) \leqslant Q_n(\boldsymbol{\pi}^*),$$

with the equality only when $\boldsymbol{\pi} = \boldsymbol{\pi}^*$.

We first show (1). Under condition (C4) and $\sup_{(i,k)\in\mathcal{O}} |\rho_{ik0}| > a\lambda$,

$$|\rho_{ik}| \geqslant ||\rho_{ik} - \rho_{ik0}| - |\rho_{ik0}|| \geqslant |\rho_{ik0}| - \delta_n > a\lambda,$$

when $\lambda \gg \delta_n$. Thus,

$$\sum_{i,k} p_\lambda(|\rho_{ik}^*|) = \sum_{(i,k)\in\mathcal{O}} p_\lambda(|\rho_{ik}|) = c_1, \tag{16}$$

where c_1 is a constant not depend on $\boldsymbol{\pi}$. In addition, for any $\boldsymbol{\pi} \in \boldsymbol{\Pi}_n^{\delta_n}$, by the definition of $\widehat{\boldsymbol{\pi}}_n^{or}$, we have $L_n(\widehat{\boldsymbol{\pi}}_n^{or}) \geqslant L_n(\boldsymbol{\pi}^*)$. Therefore, it follows $Q_n(\boldsymbol{\pi}^*) = L_n(\boldsymbol{\pi}^*) + c_1$ and $Q_n(\widehat{\boldsymbol{\pi}}_n^{or}) = L_n(\widehat{\boldsymbol{\pi}}_n^{or}) + c_1$. Hence we get

$$Q_n(\boldsymbol{\pi}^*) \leqslant Q_n(\widehat{\boldsymbol{\pi}}_n^{or}).$$

Next we show (2). For any $\boldsymbol{\pi} \in \boldsymbol{\Pi}_n^{\delta_n} \bigcup \widetilde{\boldsymbol{\Pi}}_n$, by Taylor's expansion, we have

$$\begin{aligned} Q_n(\boldsymbol{\pi}^*) - Q_n(\boldsymbol{\pi}) &= L_n(\boldsymbol{\pi}^*) - L_n(\boldsymbol{\pi}) - \left\{ \sum_{i,k} p_\lambda(|\rho_{ik}^*|) - \sum_{i,k} p_\lambda(|\rho_{ik}|) \right\} \\ &= \left\{ \frac{\partial L_n(\widetilde{\boldsymbol{\pi}})}{\partial \boldsymbol{\rho}} \right\}' (\boldsymbol{\rho}^* - \boldsymbol{\rho}) - \sum_{i,k} \dot{p}_\lambda(|\widetilde{\rho}_{ik}|)(|\rho_{ik}^* - \rho_{ik}|) \\ &= H_1 + H_2, \end{aligned}$$

where $\widetilde{\rho}_{ik}$ lies between ρ^*_{ik} and ρ_{ik}. For H_2, it can be seen that

$$\begin{aligned} H_2 &= -\sum_{i,k} \dot{p}_\lambda(|\widetilde{\rho}_{ik}|)(|\rho^*_{ik} - \rho_{ik}|) \\ &= -\sum_{(i,k)\in\mathcal{O}} \dot{p}_\lambda(|\widetilde{\rho}_{ik}|)(|\rho^*_{ik} - \rho_{ik}|) - \sum_{(i,k)\notin\mathcal{O}} \dot{p}_\lambda(|\widetilde{\rho}_{ik}|)(|\rho^*_{ik} - \rho_{ik}|) \\ &= -\sum_{(i,k)\in\mathcal{O}} \dot{p}_\lambda(|\widetilde{\rho}_{ik}|)(|\rho^*_{ik} - \rho_{nik}|) + \sum_{(i,k)\notin\mathcal{O}} \dot{p}_\lambda(|\widetilde{\rho}_{ik}|)|\rho_{ik}|. \end{aligned}$$

Similar to (16), $\dot{p}_\lambda(|\widetilde{\rho}_{ik}|)$ in the first term is equal to 0. Thus, the first term in H_2 disappears. For the second term in H_2, noting $|\widetilde{\rho}_{ik}| \leqslant |\rho_{ik}| \leqslant |\rho_{ik} - \widehat{\rho}^{or}_{ik}| + |\widehat{\rho}^{or}_{ik}| \leqslant t_n$ for $(i,k) \notin \mathcal{O}$, we get

$$H_2 \geqslant \sum_{(i,k)\notin\mathcal{O}} \dot{p}_\lambda(t_n)|\rho_{ik}|, \tag{17}$$

by the concavity of $p_\lambda(t)$.

For H_1, by Taylor's expansion, we have

$$\begin{aligned} H_1 &= \sum_{i,k} \left\{\frac{\partial L_n(\widetilde{\boldsymbol{\pi}})}{\partial \rho_{ik}}\right\}' (\rho^*_{ik} - \rho_{ik}) \\ &\leqslant \sum_{i,k} \left\{c\mathbb{P}\frac{\partial L_n(\widetilde{\boldsymbol{\pi}})}{\partial \rho_{ik}}\right\}' (\rho^*_{ik} - \rho_{ik}) \\ &= \sum_{(i,k)\notin\mathcal{O}} \left[c\left\{\mathbb{P}\frac{\partial L_n(\boldsymbol{\pi}_0)}{\partial \rho_{ik}} + \mathbb{P}\frac{\partial^2 L_n(\boldsymbol{\pi}_0)}{\partial \rho^2_{ik}}(\widetilde{\rho}_{ik} - \rho_{ik0})(1 + o_p(1))\right\}\right](-\rho_{ik}) \\ &\leqslant \sum_{(i,k)\notin\mathcal{O}} O_p(t_n + \delta_n)(-\rho_{ik}). \end{aligned}$$

The last inequality is because $|\widetilde{\rho}_{ik} - \rho_{ik0}| \leqslant |\widetilde{\rho}_{ik} - \rho_{ik}| + |\rho_{ik} - \rho_{ik0}| \leqslant t_n + \delta_n$ and $\lambda_{max}\left\{\mathbb{P}\partial^2 L_n(\boldsymbol{\pi}_0)/\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}'\right\} <$ ∞. Thus, $H_1 \geqslant -\sum_{(i,k)\notin\mathcal{O}} O_p(t_n + \delta_n)|\rho_{ik}|$. Combining with (17), we have

$$Q_n(\boldsymbol{\pi}^*) - Q_n(\boldsymbol{\pi}) \geqslant \sum_{(i,k)\notin\mathcal{O}} \left\{\dot{p}_\lambda(t_n) - O_p(t_n + \delta_n)\right\}|\rho_{ik}|. \tag{18}$$

Taking $t_n = o(\lambda)$, we get $\dot{p}_\lambda(t_n) \to \lambda$. Since $\lambda \gg \delta_n$, for sufficiently large n, we have $Q_n(\boldsymbol{\pi}^*) \geqslant Q_n(\boldsymbol{\pi})$. This completes the proof of Theorem 3.3.

S3.2.4 *Proof of Theorem 3.4.* Theorem 3.4 is a standard result of semiparametric M-estimation, which can be referred to Theorem 1 in Cheng and Huang (2010). The proofs of

$\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are similar. So we only consider the proof of $\boldsymbol{\beta}$ in the following by verifying the conditions of Theorem 1 in Cheng and Huang (2010).

We define $\boldsymbol{\Psi} = (\boldsymbol{\alpha}', \sigma^2, \mu, \boldsymbol{\phi}', \boldsymbol{\rho}')'$ is the other parametric which is infinite. Rewrite $\boldsymbol{\pi} = (\boldsymbol{\beta}', \boldsymbol{\Psi}')'$. First, Condition I in Cheng and Huang (2010) is satisfied, which means that the expectation of the derivative of score function and the variance of score function are both nonsingular. In addition, Conditions (W1)-(W5) in Cheng and Huang (2010) are satisfied because it's an exchangeable bootstrap and $c = 1$ under the notation in Cheng and Huang (2010). Therefore, it suffices to verify Conditions (S1)-(S3) and (SB1)-(SB3).

Conditions (S1), (S2), (SB1) and (SB2) follow by Conditions (C1) ad (C2). For Condition (S3), recall that $\|\widehat{\boldsymbol{\Psi}}_n - \boldsymbol{\Psi}_0\| = O_p(\delta_n)$ and $\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| = n^{-1/2}$. If $\tau + \upsilon < 1/2$, there exists $\gamma \in (1/4, 1/2]$ satisfying Condition (S3). For Condition (SB3), we note $E(l(\boldsymbol{\beta}, \boldsymbol{\Psi}) - l(\boldsymbol{\beta}_0, \boldsymbol{\Psi}_0)) \preceq -d^2(\boldsymbol{\Psi}, \boldsymbol{\Psi}_0) - \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2$ and $E(l(\boldsymbol{\beta}, \boldsymbol{\Psi}_0) - l(\boldsymbol{\beta}_0, \boldsymbol{\Psi}_0)) \succeq -\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2$. Thus, we have

$$\begin{aligned} E(l(\boldsymbol{\beta}, \boldsymbol{\Psi}) - l(\boldsymbol{\beta}, \boldsymbol{\Psi}_0)) &\leqslant E(l(\boldsymbol{\beta}, \boldsymbol{\Psi}) - l(\boldsymbol{\beta}_0, \boldsymbol{\Psi}_0)) - E(l(\boldsymbol{\beta}, \boldsymbol{\Psi}_0) - l(\boldsymbol{\beta}_0, \boldsymbol{\Psi}_0)) \\ &\preceq -d^2(\boldsymbol{\Psi}, \boldsymbol{\Psi}_0) + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2. \end{aligned} \tag{19}$$

Denote $V_\delta = \{l(\boldsymbol{\beta}, \boldsymbol{\Psi}) - l(\boldsymbol{\beta}, \boldsymbol{\Psi}_0) : \boldsymbol{\pi} = (\boldsymbol{\beta}', \boldsymbol{\Psi}')' \in \boldsymbol{\Pi}_n^\delta\}$ and its bracketing entropy integral is $J_{[]}(\delta, V_\delta, \|\cdot\|) \preceq \sqrt{K_n m_n + m_n^2}\delta$. Then, similar as (7), we have

$$E\big(\sup_{V_\delta} |\sqrt{n}(\mathbb{P}_n - \mathbb{P})(l(\boldsymbol{\beta}, \boldsymbol{\Psi}) - l(\boldsymbol{\beta}, \boldsymbol{\Psi}_0))|\big)\big) \preceq O(\sqrt{K_n m_n + m_n^2}\delta + \frac{K_n m_n + m_n^2}{\sqrt{n}}), \tag{20}$$

and

$$E\big(\sup_{V_\delta} |\sqrt{n}(\mathbb{P}_n^* - \mathbb{P})(l(\boldsymbol{\beta}, \boldsymbol{\Psi}) - l(\boldsymbol{\beta}, \boldsymbol{\Psi}_0))|\big) \preceq O(\sqrt{K_n m_n + m_n^2}\delta + \frac{K_n m_n + m_n^2}{\sqrt{n}}), \tag{21}$$

where $\mathbb{P}_n^*$ is empirical measure under bootstrap samples satisfying Theorem 3 of Cheng and Huang (2010). So Condition (SB3) follows by (19)-(21). This completes of the proof.

References

Beck, A. (2014). *Introduction to nonlinear optimization Theory, Algorithms, and Applications with MATLAB.* Society for Industrial and Applied Mathematics.

Cheng, G. and Huang, J. (2010). Bootstrap consistency for general semiparametric m-estimation. *The Annals of Statistics* **38,** 2884–2915.

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

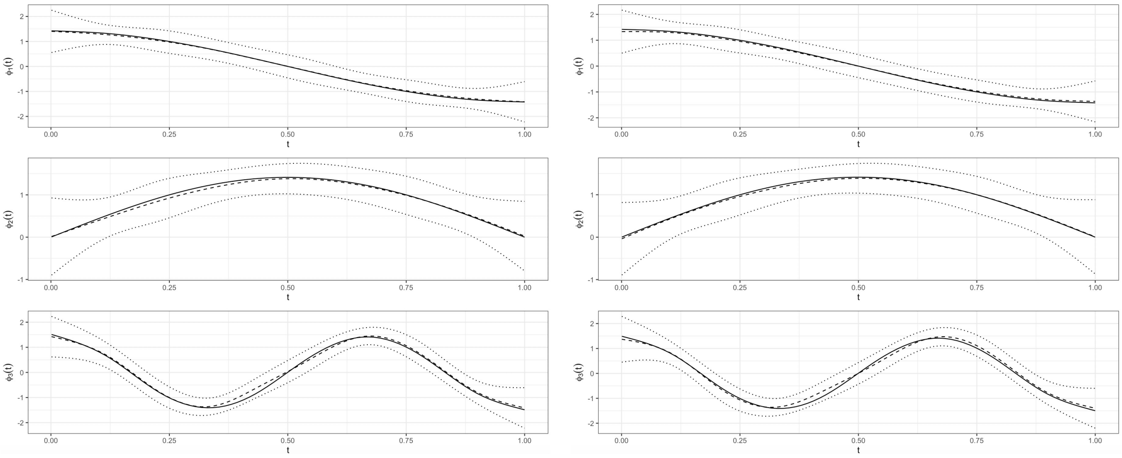
James, G. M., Hastie, T. J., and Suger, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87,** 587–602.

Ma, L., Hu, T., and Sun, J. (2015). Sieve maximum likelihood regression analysis of dependent current status data. *Biometrika* **102,** 731–738.

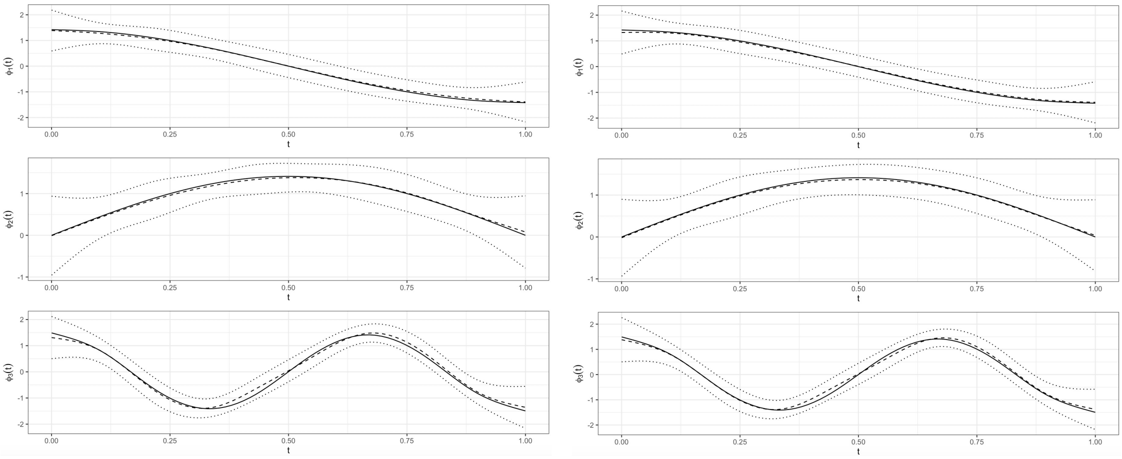
Pollard, D. (1984). *Convergence of Stochastic Procedd.* New York: Spring-Verlag.

Schumaker, L. (1981). *Spline Functions: Basic Theory.* Wiley, New York.

van der Varrt, A. W. and Wellner, J. A. (1996). *Weak convergence and Empirical Processes.* Springer.

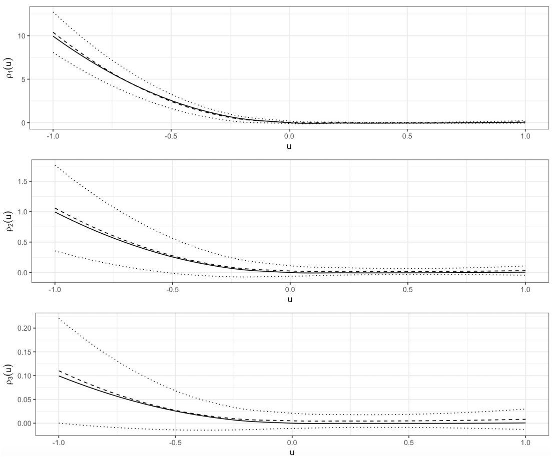


(a) n_i=10,n=100 (b) n_i=20,n=100

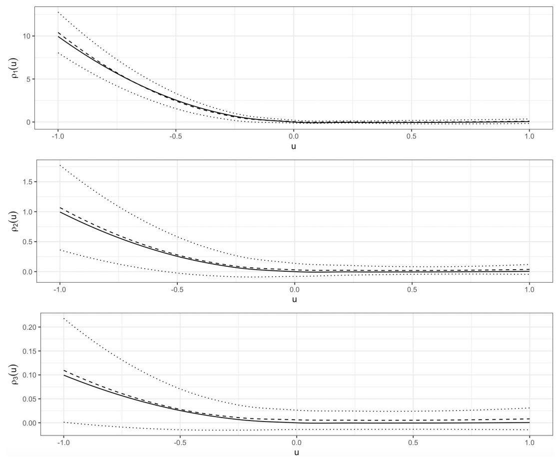


(c) n_i=10,n=500 (d) n_i=20,n=500

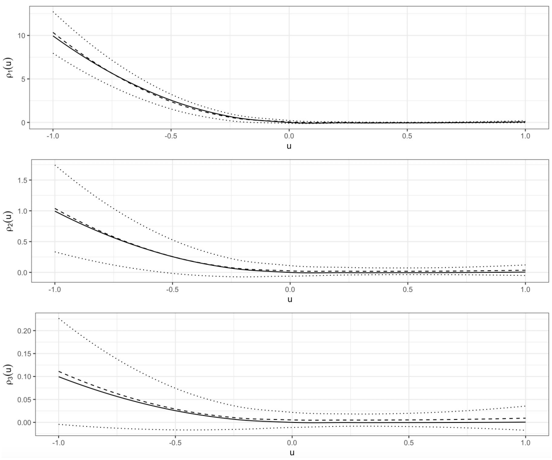
Figure S1: Estimates of the eigenfunctions $\phi_k(t), k = 1, 2, 3$ (solid-true function; dashed-average of the estimated function; dotted-95% pointwise confidence bands) by FRIS in Example 1.



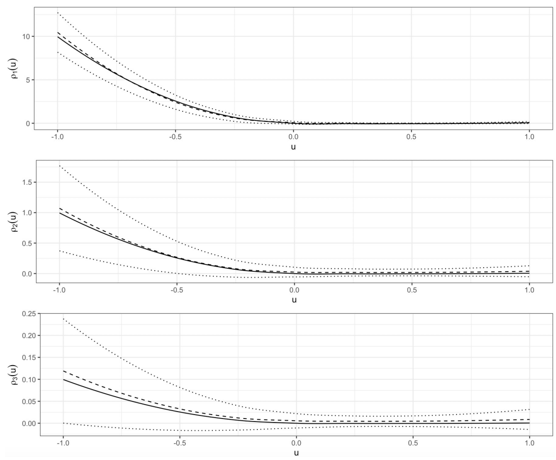
(a) n_i=10,n=100



(b) n_i=20,n=100

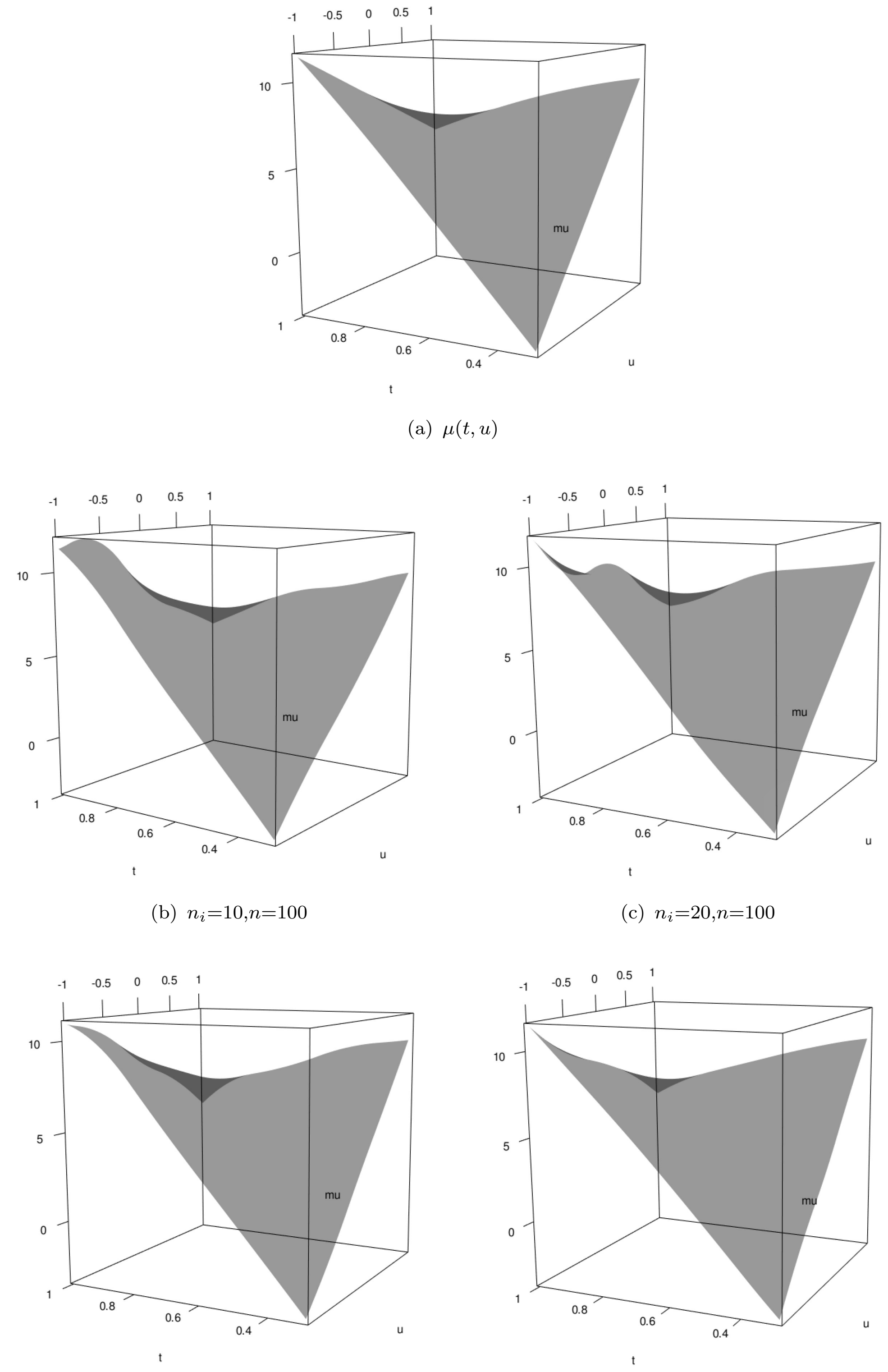


(c) n_i=10,n=500



(d) n_i=20,n=500

Figure S2: Estimates of the score variances $\rho_k(u), k = 1, 2, 3$ (solid-true function; dashed-average of the estimated function; dotted-95% pointwise confidence bands) by FRIS in Example 1.



(a) $\mu(t,u)$

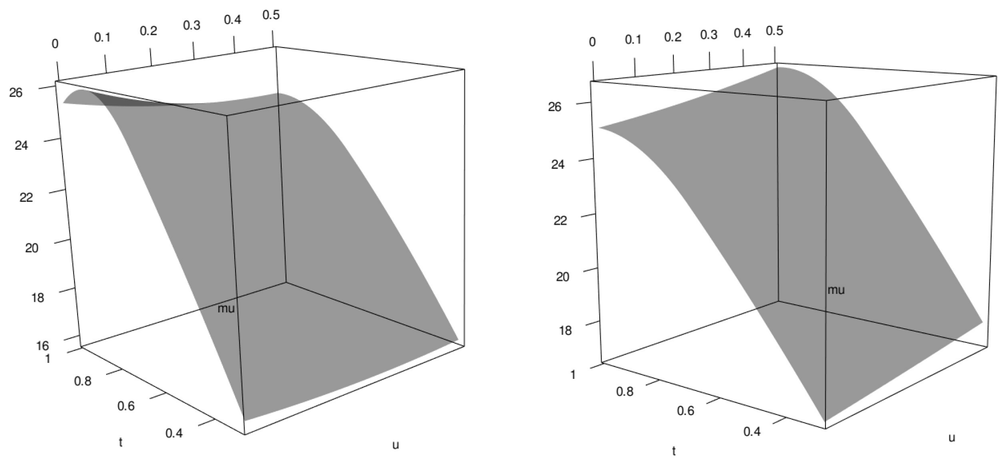
(b) n_i=10,n=100

(c) n_i=20,n=100

(d) n_i=10,n=500

(e) n_i=20,n=500

Figure S3: (a) True mean function $\mu(t,u)$; (b-e)Estimated mean function $\mu(t,u)$ by FRIS in Example 1.



(a) FRIS (b) SSV

Figure S4: Estimates of mean function $\mu(t, u)$ by FRIS and SSV for Avon Longitudinal Study of Parents and Children.

Table S1: Performance of FRIS for estimating $\boldsymbol{\alpha}$ under Example 1; presented are bias (sd).

	Normal				Mixture Normal			
	$n_i = 10$		$n_i = 20$		$n_i = 10$		$n_i = 20$	
	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$
α_{11}	0.0014(0.0346)	0.0013(0.0169)	0.0014(0.0314)	0.0012(0.0165)	0.0025(0.0367)	0.0021(0.0172)	0.0027(0.0428)	0.0018(0.0187)
α_{12}	0.0053(0.0937)	0.0042(0.0635)	0.0051(0.0798)	0.0039(0.0682)	0.0082(0.1063)	0.0063(0.0815)	0.0094(0.1024)	0.0058(0.0908)
α_{13}	0.0031(0.0778)	0.0011(0.0353)	0.0032(0.0806)	0.0013(0.0352)	0.0048(0.0867)	0.0021(0.0334)	0.0062(0.1210)	0.0022(0.0356)
α_{21}	0.0034(0.0936)	0.0030(0.0902)	0.0033(0.1147)	0.0025(0.0868)	0.0089(0.1212)	0.0057(0.0928)	0.0072(0.1283)	0.0070(0.0934)
α_{22}	0.0035(0.0700)	0.0017(0.0458)	0.0038(0.1011)	0.0021(0.0462)	0.0057(0.0981)	0.0039(0.0545)	0.0056(0.1062)	0.0038(0.0493)
α_{23}	0.0025(0.0522)	0.0013(0.0378)	0.0019(0.0511)	0.0010(0.0374)	0.0030(0.0600)	0.0016(0.0436)	0.0022(0.0592)	0.0015(0.0440)
α_{31}	0.0120(0.1059)	0.0046(0.0842)	0.0062(0.0316)	0.0042(0.0280)	0.0152(0.1388)	0.0072(0.0825)	0.0171(0.1466)	0.0073(0.0889)
α_{32}	0.0077(0.0881)	0.0042(0.0549)	0.0080(0.0921)	0.0053(0.0499)	0.0103(0.1310)	0.0065(0.0608)	0.0084(0.1061)	0.0063(0.0534)
α_{33}	0.0122(0.0942)	0.0061(0.0833)	0.0087(0.1027)	0.0065(0.0829)	0.0103(0.1237)	0.0075(0.0876)	0.0108(0.1216)	0.0081(0.0834)

Table S2: Type 1 error rate and power for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by FRIS for Example 2.

	Type 1 error rate							Power					
	$n_i = 10$			$n_i = 20$				$n_i = 10$			$n_i = 20$		
	$n = 50$	$n = 100$	$n = 500$	$n = 50$	$n = 100$	$n = 500$		$n = 50$	$n = 100$	$n = 500$	$n = 50$	$n = 100$	$n = 500$
β_2	0.0466	0.0519	0.0505	0.0525	0.0511	0.0506	β_1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
α_{11}	0.0451	0.0477	0.0482	0.0540	0.0472	0.0483	β_3	0.8864	0.9843	0.9999	0.9062	0.9980	0.9999
α_{21}	0.0526	0.0511	0.0520	0.0514	0.0524	0.0503	α_{12}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
α_{31}	0.0547	0.0561	0.0525	0.0465	0.0510	0.0495	α_{13}	0.8774	0.9616	0.9916	0.8354	0.9697	0.9908
	*	*	*	*	*	*	α_{22}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	*	*	*	*	*	*	α_{23}	0.7659	0.9384	0.9806	0.8237	0.9311	0.9795
	*	*	*	*	*	*	α_{32}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	*	*	*	*	*	*	α_{33}	0.8904	0.9351	0.9737	0.8020	0.9567	0.9857

Table S3: Comparisons of FRIS and SSV Under Example 3 for the SSV data; presented are bias (sd).

	$n_i = 10$				$n_i = 20$			
	$n = 100$		$n = 500$		$n = 100$		$n = 500$	
	FRIS	SSV	FRIS	SSV	FRIS	SSV	FRIS	SSV
β_1	0.0019(0.0245)	0.0015(0.0191)	0.0012(0.0098)	0.0004(0.0072)	0.0015(0.0248)	0.0007(0.0128)	0.0007(0.0086)	0.0007(0.0069)
β_2	0.0006(0.0201)	0.0003(0.0149)	0.0003(0.0072)	0.0003(0.0058)	0.0004(0.0156)	0.0008(0.0091)	0.0006(0.0058)	0.0003(0.0039)
β_3	0.0014(0.0240)	0.0006(0.0178)	0.0004(0.0085)	0.0003(0.0077)	0.0005(0.0172)	0.0006(0.0100)	0.0005(0.0071)	0.0002(0.0049)
$\mu(\cdot,\cdot)$	0.0359(0.6306)	0.0344(0.6050)	0.0186(0.2738)	0.0187(0.2645)	0.0371(0.5952)	0.0366(0.5834)	0.0267(0.2599)	0.0238(0.2579)
$\rho_1(\cdot)$	0.0273(0.3972)	0.0257(0.3540)	0.0128(0.2791)	0.0101(0.2766)	0.0226(0.3542)	0.0178(0.3246)	0.0088(0.2690)	0.0022(0.2451)
$\rho_2(\cdot)$	0.0168(0.1632)	0.0085(0.1525)	0.0049(0.0993)	0.0043(0.0992)	0.0182(0.1622)	0.0088(0.1558)	0.0060(0.0905)	0.0027(0.0878)
$\rho_3(\cdot)$	0.0058(0.0575)	0.0018(0.0548)	0.0042(0.0144)	0.0018(0.0142)	0.0034(0.0535)	0.0021(0.0502)	0.0032(0.0138)	0.0006(0.0131)

Table S4: Estimates and p-value for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ of FRIS Under Example 4 for the FSREM data.

		$n_i = 10$				$n_i = 20$			
		$n = 100$		$n = 500$		$n = 100$		$n = 500$	
		Est.	p-value	Est.	p-value	Est.	p-value	Est.	p-value
	β_1	0.8006	0.0000	0.7959	0.0000	0.7990	0.0000	0.7960	0.0000
$\boldsymbol{\beta}$	β_2	0.0001	0.9994	0.0074	0.7486	-0.0005	0.9935	0.0094	0.7241
	β_3	0.5992	0.0000	0.6053	0.0000	0.6013	0.0000	0.6052	0.0000
	α_{11}	0.1082	0.7186	0.1279	0.6786	0.1184	0.7000	0.1100	0.7243
$\boldsymbol{\alpha}_1$	α_{12}	0.9918	0.0000	0.9884	0.0000	0.9895	0.0000	0.9925	0.0000
	α_{13}	0.0679	0.7729	0.0818	0.7123	0.0829	0.6972	0.0536	0.8085
	α_{21}	0.1189	0.6991	0.1317	0.6465	0.1183	0.7164	0.1110	0.7204
$\boldsymbol{\alpha}_2$	α_{22}	0.9894	0.0000	0.9886	0.0000	0.9901	0.0001	0.9920	0.0000
	α_{23}	0.0832	0.7068	0.0724	0.7200	0.0758	0.7578	0.0606	0.7763
	α_{31}	0.1215	0.6776	0.1127	0.6675	0.0883	0.7769	0.1450	0.6178
$\boldsymbol{\alpha}_3$	α_{32}	0.9905	0.0000	0.9897	0.0000	0.9947	0.0000	0.9857	0.0000
	α_{33}	0.0651	0.7636	0.0738	0.7139	0.0536	0.8083	0.0853	0.6869

Table S5: Estimates and p-value for β and α of FRIS Under Example 4.

		$n_i = 10$				$n_i = 20$			
		$n = 100$		$n = 500$		$n = 100$		$n = 500$	
		Est.	p-value	Est.	p-value	Est.	p-value	Est.	p-value
β	β_1	0.8976	0.0000	0.8936	0.0000	0.8985	0.0000	0.8921	0.0000
	β_2	-0.0008	0.7778	-0.0010	0.3321	0.0015	0.6513	-0.0018	0.1437
	β_3	0.4407	0.0000	0.4489	0.0000	0.4389	0.0000	0.5994	0.0000
α_1	α_{11}	0.0020	0.9676	-0.0012	0.9794	0.0067	0.9129	-0.0116	0.8627
	α_{12}	0.8208	0.0000	0.8034	0.0000	0.1103	0.0000	0.1109	0.0000
	α_{13}	0.5713	0.0000	0.5955	0.0000	0.5813	0.0000	0.4427	0.0000
α_2	α_{21}	-0.0092	0.0554	-0.0119	0.2147	-0.0164	0.1829	-0.0162	0.2679
	α_{22}	0.8112	0.0000	0.8003	0.0000	0.8007	0.0000	0.7930	0.0000
	α_{23}	0.5847	0.0000	0.5995	0.5995	0.5988	0.0000	0.6090	0.0000
α_3	α_{31}	-0.0032	0.9519	0.0042	0.9267	-0.0065	0.9166	-0.0038	0.9152
	α_{32}	0.8238	0.0000	0.8076	0.0000	0.8318	0.0000	0.8232	0.0000
	α_{33}	0.5669	0.0000	0.5897	0.0000	0.5550	0.0000	0.5678	0.0000