Functional data analysis with covariate-dependent mean and covariance structures

Huazhen Lin

Center of Statistical Research
Southwestern University of Finance and Economics

This is joint work with Chenlin Zhang, Li Liu, Jin Liu and Yi Li

ALSPAC data

- The Avon Longitudinal Study of Parents and Children (ALSPAC), known as Children of the 90s, is a birth cohort study based in England.
- Between 1991 and 1992, 14,000 pregnant women were recruited; they, along with their children and their partners, were followed up intensively over two decades.
- The goal of investigating the environmental and genetic factors that affect a persons health and development.

ALSPAC data

- we take the body mass index (BMI) curves of children measured from 0 to 7-24 years as the growth trajectories of children
- the nine covariates include the birth weight, birth length, presence of maternal gestational diabetes, amniocentesis noted during pregnancy, number of children previously delivered by a mother, and method of delivery.
- After conducting quality control and removing the subjects with missing values, we obtain 7,313 individuals for the data analysis.

Model

We Consider the regression of

- functional $Y_i(\cdot) \sim$ vector of covariates $\mathbf{X}_i$.
- Since mean and covariance are the central profiles of the distribution for modelling functional data.
- Thus we consider
  - $E\{Y_i(t)|\mathbf{X}_i\}$ and $\text{cov}\{Y_i(t), Y_i(s)|\mathbf{X}_i\}$.

Existing methods for functional regression model

- Random Effect Model (Morris and Carroll, 2006):

$$\mathbf{Y}(t) = \mathbf{XB}(t) + \mathbf{ZU}(t) + \mathbf{E}(t),$$

where $\mathbf{B}(t), \mathbf{U}(t)$ are the vectors of fixed effect function and random effect functions.

- Single-Index Model (Jiang and Wang, 2010):

$$\mathbf{Y}(t) = \mu\{t, \mathbf{X}(t)^T\beta\} + \varepsilon\{t, Z(t)\}.$$

- Functional Varying-Coefficient Single-Index Model (Li et al., 2017):

$$Y_{ij}(t) = \mathbf{X}_i^T\alpha_j(t) + g_j(Z_i^T\beta_j) + \varepsilon_{ij}(t).$$

These works do not treat functional outcomes as a whole and treated the covariance of functional outcomes as nuisance.

Construct covariance structure under FPCA

Modeling the covariance structure of functional data under the framework of functional principal component analysis (FPCA):

$$Y_i(t) = \mu(t) + \sum_{k=1}^{K} \xi_{ik}\phi_k(t),$$

$$E\{Y_i(t)\} = \mu(t),\ E\{\xi_{ik}\} = 0,\ \ \mathrm{cov}\{Y_i(t), Y_i(s)\} = \sum_{k=1}^{K} \mathrm{var}(\xi_{ik})\phi_k(t)\phi_k(s).$$

Few eigenfunctions $\phi_k(t), k = 1, \cdots, K$ are used to explore the functional responses. However, traditional FPCA did not consider how the functional responses varies with the covariates.

Existing methods under FPCA

Recent works established the dependence of ξ_{ik} on covariates:

- Li et al. (2016); Chen et al. (2019):

$$\xi_{ik} = X_i^T \boldsymbol{\beta}_k + \varepsilon_{ik},$$

- Chiou et al. (2003a,b):

$$E(\xi_{ik}|X_i) = \alpha_k(X_i^T \boldsymbol{\beta}_k),$$

- Backenroth et al. (2018):

$$\mathrm{var}(\xi_{ik}|X_i) = \exp(X_i^T \alpha_k).$$

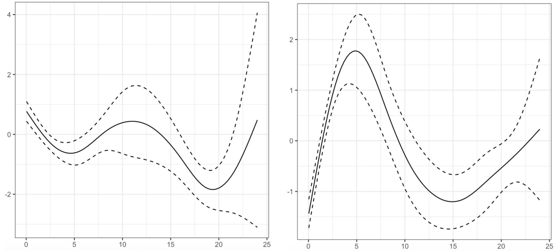
In these works, the eigenfunctions are the same for each individuals. It may be not enough to understand the dependence of $Y_i(t)$ on X_i.

An interesting problem is

- Whether and how the covariance structure, including eigen-functions and its scores, varies with covariates?

Based on the proposed method, we find

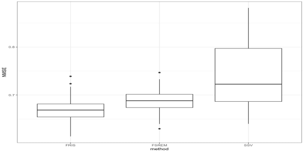
- For the BMI data, the individuals $\mathbf{X}_i'\boldsymbol{\alpha}_2 \geqslant -0.3$ will be expressed only by eigenfunctions $\phi_1(t)$, while the others by both $\phi_1(t)$ and $\phi_2(t)$.



(a) $\phi_1(t)$ (b) $\phi_2(t)$

Based on the proposed method

- Due to the parsimonious representation, prediction and interpretability can be improved.



(c)

Figure 1: PE of FRIS, SSV and FSREM for Avon Longitudinal Study of Parents and Children.

Model

Denote by $\{\mathbf{X}_i, Z_i(\cdot)\}_{(i=1,\cdot,n)}$ n independent and identically distributed (i.i.d.) realizations of random function $\{\mathbf{X}, Z(\cdot)\}$. We propose:

$$Z_i(t) = \mu(t, \mathbf{X}_i'\beta) + \sum_{k=1}^{K_n} \xi_{ik}\phi_k(t), \qquad (1)$$

$$E(\xi_{ik}|\mathbf{X}_i) = 0, \mathrm{var}(\xi_{ik}|\mathbf{X}_i) = \rho_k(\mathbf{X}_i'\alpha_k), \; i = 1, \ldots, n,$$

where μ, ϕ_k and ρ_k are unknown functions. Model (1) is termed functional regression model with individual-specific mean and covariance structures (FRIS).

FRIS

- The unknown $\rho_k(\cdot)$ provides an opportunity to identify important eigenfunction for each individual.
- if $|\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)|$ is large, the corresponding component $\phi_k(\cdot)$ is important for individual i to explain the proportion of variation attributable to that direction.
- if $\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k) = 0$, the component $\phi_k(\cdot)$ is not to be selected for individual i, indicating one fewer principal component for $Z_i(\cdot)$.

FRIS

We observe the random functions $Z_i(\cdot)$ with measurement errors, that is,

$$Y_i(t_{ij}) = Z_i(t_{ij}) + \epsilon_i(t_{ij}),\ j = 1, \ldots, n_i;\ i = 1, \ldots, n,$$

where $\epsilon_{ij} = \epsilon_i(t_{ij})$ are independent and identically distributed (iid) measurement errors with $\mathrm{E}(\epsilon_{ij}) = 0$ and $\mathrm{var}(\epsilon_{ij}) = \sigma^2$.

Identification

(IC) $\|\boldsymbol{\beta}\| = 1$, $\|\boldsymbol{\alpha}_k\| = 1$, and the first non-zero elements of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_k$ are positive for $k = 1, \cdots, K_n$. Denote by $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_{K_n}(t))'$; $\phi_k(0) > 0$, $k = 1, \ldots K_n$. We further assume $\int \boldsymbol{\phi}(t)\boldsymbol{\phi}(t)'dt = \mathbf{I}_{K_n}$.

Estimation

Denotes all of the unknown parameters and functions by $\boldsymbol{\pi}$. Maximizing the penalized log quasi-likelihood function,

$$Q_n(\boldsymbol{\pi}) = L_n(\boldsymbol{\pi}) - \sum_{k=1}^{K_n} \sum_{i=1}^{n} p_\lambda(|\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)|), \quad (2)$$

where

$$L_n(\boldsymbol{\pi}) = -\frac{1}{2n} \sum_{i=1}^{n} \log|\boldsymbol{\Sigma}_i| - \frac{1}{2n} \sum_{i=1}^{n} \{\mathbf{Y}_i - \boldsymbol{\mu}_i\}' \boldsymbol{\Sigma}_i^{-1} \{\mathbf{Y}_i - \boldsymbol{\mu}_i\},$$

and $\boldsymbol{\mu}_i = \mu(\mathbf{t}_i, \mathbf{X}_i'\beta)$, $\boldsymbol{\Sigma}_i = \sum_{k=1}^{K_n} \phi_k(\mathbf{t}_i)\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)\phi_k(\mathbf{t}_i)' + \sigma^2 \mathbf{I}_{n_i}$.

Estimation

The unknown functions $\mu(t,u)$, $\phi_k(t)$ and $\rho_k(u)$ can be respectively approximated by

$$\mu(t,u) \approx \boldsymbol{\gamma}'\mathbf{B}_n(t,u),\ \phi_k(t) \approx \boldsymbol{\eta}_k'\mathbf{B}_{n1}(t) \text{ and } \rho_k(u) \approx \left\{\boldsymbol{\theta}_k'\mathbf{B}_{n2}(u)\right\}^2,$$

where $\mathbf{B}_n(t,u) = \mathbf{B}_{n1}(t) \otimes \mathbf{B}_{n2}(u)$, $\otimes$ is the Kronecker product, $\mathbf{B}_{n1}(\cdot)$ and $\mathbf{B}_{n2}(\cdot)$ are two sets of spline basis functions.

Estimation

Remarks on the penalty $p_\lambda(|\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)|)$

- Local sparsity for a function $g(t)$ (James et al., 2009; Zhou et al., 2013; Lin et al., 2017);
- The argument of $\rho_k(\cdot)$ is $\mathbf{X}_i'\boldsymbol{\alpha}_k$, which is individualized and depends on unknown $\boldsymbol{\alpha}_k$. We cannot adapt local sparsity methods to our case.

Optimization via ADMM

Let $\zeta_{ik} = \theta_k' \mathbf{B}_{n2}(\mathbf{X}_i' \alpha_k)$. It follows that maximizing (2) is equivalent to minimizing an augmented Lagrangian objective function:

$$\begin{aligned} \mathcal{L}_n(\pi, \zeta) &= -L_n(\pi, \zeta) + \sum_{k=1}^{K_n} \sum_{i=1}^{n} p_\lambda(|\zeta_{ik}|) \\ &+ \frac{\nu}{2} \sum_{k=1}^{K_n} \sum_{i=1}^{n} \left[\left\{ \zeta_{ik} - \theta_k' \mathbf{B}_{n2}(\mathbf{X}_i' \alpha_k) + \frac{C_{ik}}{\nu} \right\}^2 - c_0 \right], \end{aligned} \tag{3}$$

Optimization via ADMM

To lead to closed-form expressions at each ADMM step, we apply the following Taylor expansions

$$\mathbf{B}_n(\mathbf{t}_i, \mathbf{X}_i'\beta) \approx \mathbf{B}_n(\mathbf{t}_i, \mathbf{X}_i'\tilde{\beta}) + \dot{\mathbf{B}}_n(\mathbf{t}_i, \mathbf{X}_i'\tilde{\beta})\mathbf{X}_i'(\beta - \tilde{\beta}),$$
$$\mathbf{B}_{n2}(\mathbf{X}_i'\alpha_k) \approx \mathbf{B}_{n2}(\mathbf{X}_i'\tilde{\alpha}_k) + \dot{\mathbf{B}}_{n2}(\mathbf{X}_i'\tilde{\alpha}_k)\mathbf{X}_i'(\alpha_k - \tilde{\alpha}_k),$$

where $\tilde{\alpha}_k$, and $\tilde{\beta}$ are the estimators of α_k and β, respectively, from the previous step.

Optimization via ADMM

$$\beta = \left\{\sum_{i=1}^{n} \mathbf{X}_i \gamma' \dot{\mathbf{B}}_n(\mathbf{t}_i, \mathbf{X}_i'\tilde{\beta}) \mathbf{\Sigma}_i^{-1} \dot{\mathbf{B}}_n(\mathbf{t}_i, \mathbf{X}_i'\tilde{\beta})' \gamma \mathbf{X}_i' \right\}^{-1}$$

$$\times \left[\sum_{i=1}^{n} \mathbf{X}_i \gamma' \dot{\mathbf{B}}_n(\mathbf{t}_i, \mathbf{X}_i'\tilde{\beta}) \mathbf{\Sigma}_i^{-1} \left\{\mathbf{Y}_i - \mathbf{B}_n(\mathbf{t}_i, \mathbf{X}_i'\tilde{\beta})'\gamma + \dot{\mathbf{B}}_n(\mathbf{t}_i, \mathbf{X}_i'\tilde{\beta})'\gamma \mathbf{X}_i'\tilde{\beta}\right\}\right], \quad (4)$$

$$\gamma = \left\{\sum_{i=1}^{n} \mathbf{B}_n(\mathbf{t}_i, \mathbf{X}_i'\beta) \mathbf{\Sigma}_i^{-1} \mathbf{B}_n(\mathbf{t}_i, \mathbf{X}_i'\beta)' \right\}^{-1} \sum_{i=1}^{n} \left\{\mathbf{B}_n(\mathbf{t}_i, \mathbf{X}_i'\beta) \mathbf{\Sigma}_i^{-1} Y_i\right\}, \quad (5)$$

$$\theta_k = \left\{\sum_{i=1}^{n} \mathbf{B}_{n2}(\mathbf{X}_i'\alpha_k) \mathbf{B}_{n2}(\mathbf{X}_i'\alpha_k)' \right\}^{-1} \sum_{i=1}^{n} \left\{(\tilde{\zeta}_{ik} + \frac{C_{ik}}{\nu}) \mathbf{B}_{n2}(\mathbf{X}_i'\alpha_k)\right\}, \quad (6)$$

$$\alpha_k = \left[\sum_{i=1}^{n} \mathbf{X}_i \left\{\theta_k' \dot{\mathbf{B}}_{n2}(\mathbf{X}_i'\tilde{\alpha}_k)\right\}^2 \mathbf{X}_i' \right]^{-1}$$

$$\times \left[\sum_{i=1}^{n} \mathbf{X}_i \theta_k' \dot{\mathbf{B}}_{n2}(\mathbf{X}_i'\tilde{\alpha}_k) \left\{\tilde{\zeta}_{ik} + \frac{C_{ik}}{\nu} - \theta_k' \mathbf{B}_{n2}(\mathbf{X}_i'\tilde{\alpha}_k) + \theta_k' \dot{\mathbf{B}}_{n2}(\mathbf{X}_i'\tilde{\alpha}_k) \mathbf{X}_i'\tilde{\alpha}_k \right\}\right]. \quad (7)$$

Optimization via ADMM

With (3), denote $H(\zeta;\boldsymbol{\pi}) = -L_n(\boldsymbol{\pi},\zeta) + \frac{\nu}{2}\|\zeta - \mathbf{W} + \mathbf{C}/\nu\|_F^2$ with $\mathbf{W} = (\boldsymbol{\theta}_k'\mathbf{B}_{n2}(\mathbf{X}_i'\boldsymbol{\alpha}_k))$, and $p_\lambda(|\zeta|) = \sum_{k=1}^{K_n}\sum_{i=1}^{n} p_\lambda(|\zeta_{ik}|)$. We have

$$H(\zeta;\boldsymbol{\pi}) \leq H(\tilde{\zeta};\boldsymbol{\pi}) + \dot{H}(\tilde{\zeta};\boldsymbol{\pi})'(\zeta-\tilde{\zeta}) + \frac{1}{2h}(\zeta-\tilde{\zeta})'(\zeta-\tilde{\zeta}),$$

where h is sufficiently small so that the quadratic term dominates the Hessian of $H(\zeta;\boldsymbol{\pi})$. Then, we update ζ by

$$\zeta = \arg\min_{\zeta} \frac{1}{2h}\|\zeta - (\tilde{\zeta} - h\dot{H}(\tilde{\zeta};\boldsymbol{\pi}))\|_F^2 + p_\lambda(|\zeta|). \tag{8}$$

Finally, we use the gradient descent method to update σ^2 and $\boldsymbol{\eta}_k$.

Algorithm

Algorithm

1: Give initial values $\boldsymbol{\beta}^{(0)\prime}, \boldsymbol{\gamma}^{(0)\prime}, \boldsymbol{\eta}_k^{(0)\prime}, \boldsymbol{\alpha}_k^{(0)\prime}, \boldsymbol{\theta}_k^{(0)\prime}, \sigma^{2(0)}$, and $\zeta_{ik}^{(0)} = \boldsymbol{\theta}_k^{(0)\prime}\mathbf{B}_{n2}(\mathbf{X}_i'\boldsymbol{\alpha}_k^{(0)}), i = 1, \ldots, n; k = 1, \ldots, K_n$, $\mathbf{C}^{(0)} = 0$.

2: Set step-length κ, h, ν, tuning parameters λ and K_n.

3: **while** not converged **do**

4: At $(t+1)$-th iteration, we update $\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\theta}_k, \boldsymbol{\zeta}$ by (4)-(8), where $\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\alpha}}, \boldsymbol{\gamma}, \boldsymbol{\theta}_k, \widetilde{\boldsymbol{\zeta}}$ and $\mathbf{C}$ in the right of (4)-(8) are replaced by the estimators from $t-$th iteration.

5: $\boldsymbol{\eta}_k^{(t+1)} = \boldsymbol{\eta}_k^{(t)} - \kappa\partial\mathcal{L}_n(\boldsymbol{\pi}^{(t)}, \boldsymbol{\zeta}^{(t)})/\partial\boldsymbol{\eta}_k$, for $k = 1, \ldots, K_n$,

6: $\sigma^{2(t+1)} = \sigma^{2(t)} - \kappa\partial\mathcal{L}_n(\boldsymbol{\pi}^{(t)}, \boldsymbol{\zeta}^{(t)})/\partial\sigma^2$,

7: $\mathbf{C}^{(t+1)} = \mathbf{C}^{(t)} + \nu(\boldsymbol{\zeta}^{(t+1)} - \mathbf{W}^{(t+1)})$.

8: **end while**

Theoretical properties: Conditions

(C1) Covariate $\mathbf{X}$ are bounded.

(C2) $(\boldsymbol{\beta}_0', \boldsymbol{\alpha}_0', \sigma_0^2)' \in \mathcal{A}$, which is a bounded closed set, and the true functions $(\mu_0, \boldsymbol{\phi}_0', \boldsymbol{\rho}_0')' \in \mathcal{H}_{r,2} \times \prod_{k=1}^{K_n} \mathcal{H}_{r,1} \times \prod_{k=1}^{K_n} \mathcal{H}_{r,1}$ with $r > 1$, where

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

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 $c > 0$.

(C3) Denote by $\triangle_1 = \max_{l+1 \le j \le k_n+l+1} |t_j - t_{j-1}|$ and $\triangle_2 = \min_{l+1 \le j \le k_n+l+1} |t_j - t_{j-1}|$, the maximum and the mimum spacing of knots, respectively. We assume that $\triangle_1 = O(n^{-v})$ with $v \in (0, 0.5)$, and $\triangle_1/\triangle_2$ is bounded.

(C4) The penalty function $p_\lambda(t)$ is non-decreasing and concave on $[0,\infty)$. There exists a constant b such that $p_\lambda(t)$ is a constant for all $t \ge b\lambda$. In addition, $\dot{p}_\lambda(0+) = O(\lambda)$.

(C5) $K_n = n^\tau$ with $\tau \le \min(1-v, 2vr)$.

Theoretical properties

Denote $\rho_{ik} = \rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)$ and $\mathcal{O} = \{(i,k) : \rho_{ik0} \neq 0\}$. Define $\rho_{ik}^{or} = \rho_{ik}$ if $(i,k) \in \mathcal{O}$ and 0 otherwise; The oracle estimator $\hat{\boldsymbol{\pi}}_n^{or}$, the estimator $\hat{\boldsymbol{\pi}}_n$ and the true value $\boldsymbol{\pi}_0$.

Theorem 1 (Consistency and convergence rate of oracle estimators)

Under Conditions (C1)-(C5), we have

$$\|\hat{\boldsymbol{\pi}}_n^{or} - \boldsymbol{\pi}_0\| = O_p(\delta_n),$$

where $\delta_n = n^{-(1-2\nu)/2} + \sqrt{K_n} n^{-(1-\nu)/2} + \sqrt{K_n} n^{-\nu r}$.

Theoretical properties

- The first and second terms in δ_n, corresponding respectively to the estimation error for $\mu(t,u)$ and for $2K_n$ univariate functions $(\phi_k, \rho_k), k = 1, \cdots, K_n$, are related to the spline order n^{ν} and the structural parameter $K_n = n^{\tau}$.
- The last term in δ_n is the approximation error.
- When K_n does not vary with n, i.e., $\tau = 0$, Theorem 1 implies that $\|\widehat{\boldsymbol{\pi}}_n^{or} - \boldsymbol{\pi}_0\| = O_p(n^{-r/(2r+2)})$ with $\nu = 1/(2r+2)$, which is the optimal rate for approximating a nonparametric function (Stone, 1980).

Theoretical properties

Theorem 2 (Asymptotic normality of oracle estimators)

Denote by $I(\vartheta_0) = \mathbb{P}\{l^*(\pi_0)\}^{\otimes 2}$ *and* $\Lambda = \lambda_{\min}(I(\vartheta_0))$, *where* $l^*(\pi_0)$ *is defined as in the Appendix. Under Conditions (C1)-(C5), if* $0 < v < 1/4, \tau < \min\{1/2 - v, 2v(r-1), v(2r-1)/2\}$ *and* $n^{\tau - 1/2}/\Lambda = o_p(1)$ *for* $r > 1$, *and for any vector* $\mathbf{u}$ *with* $||\mathbf{u}|| = 1$, *we have*

$$\sqrt{n}\mathbf{u}' I(\vartheta_0)^{1/2}(\hat{\vartheta}_n^{or} - \vartheta_0) \xrightarrow{d} N(0, 1),$$

as $n \to \infty$.

Theoretical properties

Theorem 3 (Oracle properties)

Under Conditions (C1)-(C5), if $\lambda_{max}\left\{\mathbb{P}\partial^2 L_n(\boldsymbol{\pi}_0)/\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}'\right\}$ *is finite,* $\inf_{(i,k)\in\mathcal{O}} |\rho_{ik0}| \geq b\lambda$ *and* $\lambda \gg \delta_n$ *for some constant* $b > 0$*, we have*

(1) $P(\hat{\boldsymbol{\pi}}_n = \hat{\boldsymbol{\pi}}_n^{or}) \to 1$;

(2) $\|\hat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_0\| = O_p(\delta_n)$, *where* δ_n *is defined as in Theorem 2;*

(3) *Under the conditions in Theorem 2, we have*

$$\sqrt{n}\mathbf{u}'I(\boldsymbol{\vartheta}_0)^{1/2}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{d} N(0,1)$$

for any vector $\mathbf{u}$ *with* $\|\mathbf{u}\| = 1$.

Theoretical properties

Theorem 4 (Distribution consistency of bootstrap estimators)

Under Conditions (C1)-(C5) and if $\tau < 1/2 - v$, we have for any k,

$$\sup_{x \in R^p} |P(\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n) \leq x) - P(\sqrt{n}(\hat{\beta}_n - \beta_0) \leq x)| = o_p(1),$$

$$\sup_{x \in R^p} |P(\sqrt{n}(\hat{\alpha}_{nk}^* - \hat{\alpha}_{nk}) \leq x) - P(\sqrt{n}(\hat{\alpha}_{nk} - \alpha_{k0}) \leq x)| = o_p(1),$$

where the inequalities are taken componentwise.

Simulation

- We assess the finite sample performance of the proposed FRIS, by comparing it with
- the functional smooth random effects model (FSREM) (Chiou et al., 2003b)
- the method with a covariate-dependent mean structure and a covariate-independent covariance structure, that is, the same score variance (SSV) across individuals.

Simulation

Example 1. We generate $\mathbf{Y}_i = (Y_i(t_{i1}), \cdots, Y_i(t_{i,n_i}))'$ from a model, satisfying the assumptions of FRIS:

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

where

$$\mu(t, u) = 10 \times (u \cdot \cos(t) + (1-u) \cdot \sin(t)),$$

$$\mathbf{\Sigma}_i = \sum_{k=1}^{3} \phi_k(\mathbf{t}_i)\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)\phi_k(\mathbf{t}_i)' + \sigma^2\mathbf{I}_{n_i}, \sigma^2 = 1.$$

$$\phi_1(t) = \sqrt{2}\cos(\pi t), \phi_2(t) = \sqrt{2}\sin(\pi t), \phi_3(t) = \sqrt{2}\cos(3\pi t),$$

$$\rho_k(u) = 10^{2-k}u^2 I(u < 0), k = 1, 2, 3.$$

We set $\beta = (0.2, 0.8, 0.6)'$, $\boldsymbol{\alpha}_1 = (0.9, 0.1, 0.4)'$, $\boldsymbol{\alpha}_2 = (0.2, 0.6, 0.8)'$, $\boldsymbol{\alpha}_3 = (0.5, 0.8, 0.3)'$.

Simulation

We consider two kinds of distributions for $\mathbf{Y}_i$:

(1) Normal: $\mathbf{Y}_i \sim N(\mu(\mathbf{t}_i, \mathbf{X}_i'\beta), \Sigma_i)$;

(2) Mixture Normal: $\Sigma_i^{-1/2}\{\mathbf{Y}_i - \mu(\mathbf{t}_i, \mathbf{X}_i'\beta)\} \sim \frac{1}{2}N(-\mathbf{1}/2, \mathbf{I}) + \frac{1}{2}N(\mathbf{1}/2, \mathbf{I})$.

Example 1

Table 1: Comparisons of FRIS and SSV under Example 1; presented are bias (sd).

			Normal		Mixture Normal	
			FRIS	SSV	FRIS	SSV
$n_i = 10$	$n = 100$	β_1	0.0010(0.0117)	0.0023(0.0163)	0.0003(0.0147)	0.0036(0.0244)
		β_2	0.0004(0.0069)	0.0008(0.0102)	0.0005(0.0084)	0.0022(0.0152)
		β_3	0.0007(0.0076)	0.0007(0.0118)	0.0009(0.0093)	0.0009(0.0167)
		$\mu(\cdot,\cdot)$	0.0152(0.2844)	0.0247(0.4859)	0.0164(0.4131)	0.0369(0.5704)
		$\rho_1(\cdot)$	0.1435(0.4244)	4.3897(2.2415)	0.1673(0.4384)	4.4381(2.5859)
		$\rho_2(\cdot)$	0.0281(0.1374)	0.4939(0.2575)	0.0346(0.1369)	0.5228(0.2598)
		$\rho_3(\cdot)$	0.0054(0.0212)	0.0442(0.0444)	0.0071(0.0248)	0.0483(0.0481)
	$n = 500$	β_1	0.0010(0.0058)	0.0016(0.0113)	0.0009(0.0066)	0.0011(0.0129)
		β_2	0.0004(0.0031)	0.0005(0.0080)	0.0003(0.0039)	0.0007(0.0086)
		β_3	0.0002(0.0032)	0.0007(0.0085)	0.0002(0.0043)	0.0003(0.0094)
		$\mu(\cdot,\cdot)$	0.0056(0.1204)	0.0163(0.2458)	0.0065(0.1337)	0.0395(0.4699)
		$\rho_1(\cdot)$	0.1409(0.4134)	4.3890(2.2071)	0.1518(0.4239)	3.9318(2.3274)
		$\rho_2(\cdot)$	0.0230(0.1245)	0.4714(0.2311)	0.0231(0.1349)	0.5212(0.2451)
		$\rho_3(\cdot)$	0.0051(0.0207)	0.0436(0.0440)	0.0064(0.0215)	0.0521(0.0495)

Example 1

			Normal		Mixture Normal	
			FRIS	SSV	FRIS	SSV
$n_i = 20$	$n = 100$	β_1	0.0008(0.0109)	0.0016(0.0128)	0.0002(0.0149)	0.0017(0.0248)
		β_2	0.0001(0.0054)	0.0004(0.0083)	0.0005(0.0082)	0.0016(0.0142)
		β_3	0.0002(0.0056)	0.0006(0.0099)	0.0010(0.0084)	0.0006(0.0156)
		$\mu(\cdot,\cdot)$	0.0136(0.3247)	0.0215(0.4428)	0.0157(0.3521)	0.0299(0.4783)
		$\rho_1(\cdot)$	0.1429(0.4241)	4.4168(2.2785)	0.1753(0.4607)	4.4328(2.5274)
		$\rho_2(\cdot)$	0.0306(0.1405)	0.4759(0.2023)	0.0257(0.1357)	0.5312(0.2351)
		$\rho_3(\cdot)$	0.0056(0.0215)	0.0466(0.0433)	0.0069(0.0229)	0.0520(0.0495)
	$n = 500$	β_1	0.0006(0.0050)	0.0005(0.0061)	0.0009(0.0070)	0.0009(0.0100)
		β_2	0.0002(0.0029)	0.0003(0.0054)	0.0006(0.0039)	0.0007(0.0068)
		β_3	0.0001(0.0030)	0.0002(0.0069)	0.0005(0.0042)	0.0009(0.0077)
		$\mu(\cdot,\cdot)$	0.0059(0.0958)	0.0133(0.2534)	0.0062(0.1125)	0.0152(0.2814)
		$\rho_1(\cdot)$	0.1408(0.4126)	4.1304(1.9150)	0.1511(0.4431)	4.3713(2.2832)
		$\rho_2(\cdot)$	0.0264(0.1319)	0.4604(0.1676)	0.0232(0.1355)	0.5265(0.2219)
		$\rho_3(\cdot)$	0.0048(0.0205)	0.0438(0.0428)	0.0058(0.0219)	0.0533(0.0481)

Example 1

Table 2: Performance of FRIS for estimating α under Example 1; presented are bias (sd).

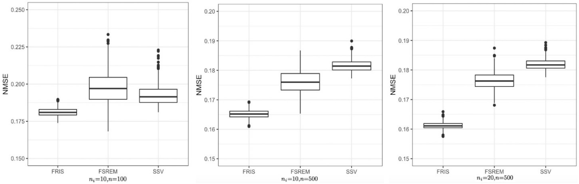
	Normal				Mixture Normal			
	$n_i = 10$		$n_i = 20$		$n_i = 10$		$n_i = 20$	
	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$
α_{11}	0.0014(0.0346)	0.0013(0.0169)	0.0014(0.0314)	0.0012(0.0165)	0.0025(0.0367)	0.0021(0.0172)	0.0027(0.0428)	0.0018(0.0187)
α_{12}	0.0053(0.0937)	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)

Example 1

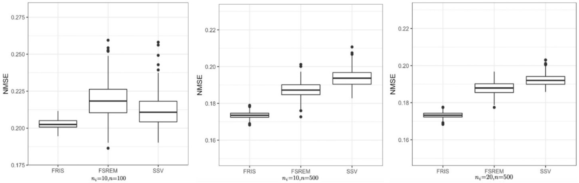
Table 3: The selection results of the eigenfunctions under Example 1; presented are mean (sd).

		Normal				Mixture Normal			
		$n_i = 10$		$n_i = 20$		$n_i = 10$		$n_i = 20$	
		$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$
ϕ_1	FPR	0.0398(0.0837)	0.0502(0.1010)	0.0549(0.0929)	0.0523(0.0963)	0.0475(0.0950)	0.0773(0.1628)	0.0626(0.1030)	0.0630(0.1074)
	FNR	0.0291(0.0291)	0.0398(0.0426)	0.0277(0.0302)	0.0369(0.0408)	0.0282(0.0290)	0.0421(0.0434)	0.0281(0.0297)	0.0413(0.0443)
ϕ_2	FPR	0.1067(0.1635)	0.0959(0.1583)	0.1024(0.1664)	0.1021(0.1643)	0.0920(0.1573)	0.1018(0.1576)	0.0958(0.1591)	0.0987(0.1542)
	FNR	0.0572(0.0635)	0.0641(0.0724)	0.0520(0.0604)	0.0559(0.0647)	0.0634(0.0657)	0.0578(0.0664)	0.0517(0.0598)	0.0572(0.0677)
ϕ_3	FPR	0.0663(0.1425)	0.0708(0.1460)	0.0805(0.1538)	0.0746(0.1471)	0.0834(0.1551)	0.0649(0.1389)	0.0829(0.1574)	0.0656(0.1375)
	FNR	0.0832(0.0881)	0.0885(0.0928)	0.0804(0.0894)	0.0783(0.0907)	0.0797(0.0923)	0.0807(0.0927)	0.0874(0.0991)	0.0806(0.0943)

Example 1



(a) Example 1(1): Normal FRIS data



(b) Example 1(2): Mixed Normal FRIS data

Figure 2: NMSE for FRIS, SSV and FSREM under Example 1.

Simulation

Example 2. To assessed Type 1 error rates and power for β and α from the FRIS, we generate data the same as in Example 1(1) except taking $\beta = (0.6, 0, 0.8)$ and $\alpha_k = (0, 0.8, 0.6), k = 1, 2, 3$.

Example 2

Table 4: Type 1 error rate and power for β and α by FRIS for Example 2.

	Type 1 error rate					Power			
	$n_i = 10$		$n_i = 20$			$n_i = 10$		$n_i = 20$	
	$n = 100$	$n = 500$	$n = 100$	$n = 500$		$n = 100$	$n = 500$	$n = 100$	$n = 500$
β_2	0.0477	0.0518	0.0505	0.0506	β_1	1	1	1	1
α_{11}	0.0451	0.0482	0.0501	0.0489	β_3	1	1	1	1
α_{21}	0.0516	0.0544	0.0463	0.0486	α_{12}	1	1	1	1
α_{31}	0.0468	0.0523	0.0548	0.0511	α_{13}	1	1	1	1
	*	*	*	*	α_{22}	1	1	1	1
	*	*	*	*	α_{23}	1	1	1	1
	*	*	*	*	α_{32}	1	1	1	1
	*	*	*	*	α_{33}	1	1	1	1

Simulation

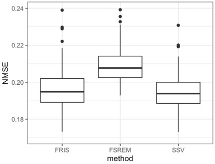
Example 3. We generate data with a common score variance, satisfying the assumption of SSV. Specifically, we generate data the same way as in Example 1(1), except that $\rho_k(u) = \rho_k$ for $k = 1, 2, 3$ and $\rho_1 = 5, \rho_2 = 1, \rho_3 = 0.5$.

Example 3

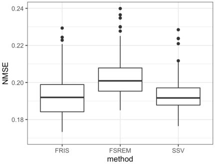
Table 5: Comparisons of FRIS and SSV Under Example 3 for the SSV data; presented are bias (sd).

	$n_i = 10$				$n_i = 20$			
	$n = 100$		$n = 500$		$n = 100$		$n = 500$	
	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)

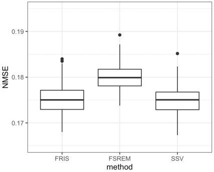
Example 3



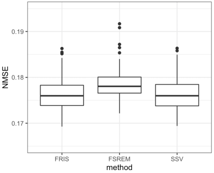
(a) n_i=10,n=100



(b) n_i=20,n=100



(c) n_i=10,n=500



(d) n_i=20,n=500

Figure 3: NMSE of three methods: FRIS, SSV and FSREM in Example 3.

Simulation

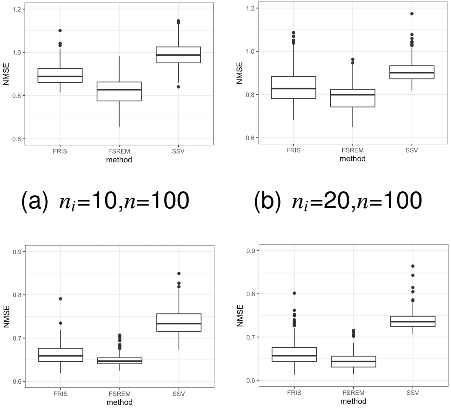
Example 4. We generate data following Chiou et al. (2003b). That is, given covariates $\mathbf{X}_i$, $Y_i(t)$ follows a normal distribution, $Y_i(t) = \mu(t) + \sum_{k=1}^{3} A_{ik}\phi_k(t)$ and assume for the observed random curves, conditional on the covariates,

$$E\{Y_i(t)|\mathbf{X}_i\} = \mu(t) + \sum_{k=1}^{3} E(A_{ik}|\mathbf{X}_i)\phi_k(t),$$

$$\mathrm{cov}\{Y_i(s), Y_i(t)|\mathbf{X}_i\} = \sum_{k=1}^{3} \mathrm{var}(A_{ik}|\mathbf{X}_i)\phi_k(s)\phi_k(t),$$

where $E(A_{ik}|\mathbf{X}_i) = \mu_k(\mathbf{X}_i'\boldsymbol{\beta}_k)$, $\mathrm{var}(A_{ik}|\mathbf{X}_i) = \rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)$, and $\mu(t) = t^2 + 1$, $\mu_1(u) = 1 - \cos(u \cdot \pi)$, $\mu_2(u) = \{1 - \cos(u \cdot \pi)\}/5$, $\mu_3(u) = \{1 - \cos(u \cdot \pi)\}/10$, $\rho_k(u) = \sqrt{\alpha_k(u)}$, $k = 1, 2, 3$, $\phi_k(t)$ is the same as in Example 1 and $\boldsymbol{\beta}_k = (0.8, 0, 0.6)'$, $\boldsymbol{\alpha}_k = (0, 1, 0)'$ for each k.

Example 4



(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 4: NMSE of three methods: FRIS, SSV and FSREM in Example 4.

Analysis of the ALSPAC study

Avon longitudinal Study of Parents and Children (ALSPAC):

- Functional response: body mass index (BMI) curve of children measured from 0 to 24 years;
- Nine covariates: birth weight, birth length, maternal gestational diabetes, amniocentesis noted during pregnancy, the number of children delivered by a mother before, and the method of delivery;
- 7313 individuals.

Analysis of the ALSPAC study

Table 6: Comparisons of the estimates for $\boldsymbol{\beta}$, accounting for covariate-mean relationships and obtained by FRIS, SSV and FSREM; presented are point estimates (Est.) and $p-$values for the ALSPAC study.

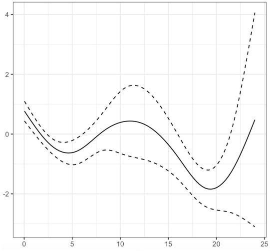
		FRIS		SSV		FSREM ($\boldsymbol{\beta}_1$)		FSREM ($\boldsymbol{\beta}_2$)	
		Est.	p-value	Est.	p-value	Est.	p-value	Est.	p-value
spontaneous	β_0	0.0827	0.7170	-0.0638	0.8208	0.0950	0.8181	0.1241	0.4301
birth weight	β_1	0.3752	0.0000	0.3281	0.0283	0.5348	0.0000	0.7300	0.0000
birth length	β_2	-0.1507	0.0000	-0.1389	0.0608	-0.4040	0.0002	-0.4297	0.0001
diabetes	β_3	0.5470	0.0000	0.7021	0.0007	0.1534	0.1250	0.2731	0.0063
amniocentesis	β_4	-0.3512	0.0002	-0.2938	0.0597	0.0187	0.8517	0.0558	0.5768
# of children	β_5	0.3859	0.0145	0.2068	0.3880	-0.0867	0.3859	-0.1153	0.2489
assisted breech	β_6	-0.1928	0.0346	-0.2540	0.0742	0.1135	0.2564	0.1403	0.1606
Caesarean section	β_7	0.4306	0.0000	0.3955	0.0221	0.3550	0.0004	0.1154	0.2485
forceps delivery	β_8	0.0276	0.4520	0.0519	0.1585	0.2785	0.4151	-0.1227	0.7417
vacuum extraction	β_9	0.1864	0.0019	0.1523	0.0844	0.3665	0.0002	-0.0158	0.8745

Analysis of the ALSPAC study

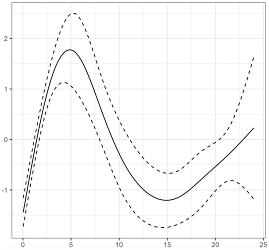
Table 7: Comparisons of the estimates of α, accounting for covariate-covariance relationships; presented are point estimates (Est.) and $p-$values for ALSPAC data.

	FRIS				FSREM			
	α_1		α_2		α_1		α_2	
	Est.	$p-$value	Est.	$p-$value	Est.	$p-$value	Est.	$p-$value
spontaneous	0.6297	0.0000	-0.1166	0.2823	0.3173	0.4020	-0.0982	0.8092
birth weight	0.0588	0.3444	0.2781	0.0000	0.1918	0.0220	0.4033	0.0000
birth length	-0.0014	0.9693	0.2771	0.0000	-0.0807	0.2799	-0.2047	0.0425
diabetes	0.4916	0.0000	0.3391	0.0012	-0.1311	0.7338	-0.4010	0.2268
amniocentesis	-0.3442	0.0006	0.4673	0.0001	0.4095	0.2559	-0.6868	0.0492
# of children	0.3822	0.0000	0.1369	0.1185	-0.5482	0.0400	0.0588	0.7893
assisted breech	0.2492	0.0001	0.2662	0.0813	-0.0472	0.8417	0.0555	0.8450
Caesarean section	-0.1053	0.0248	0.3240	0.0001	0.3765	0.2805	-0.3448	0.3587
forceps delivery	-0.0972	0.1556	0.4059	0.0003	0.4238	0.2242	0.1194	0.7331
vacuum extraction	0.1051	0.0268	0.3737	0.0011	0.2167	0.5644	-0.1167	0.7834

Analysis of the ALSPAC study



(a) $\phi_1(t)$



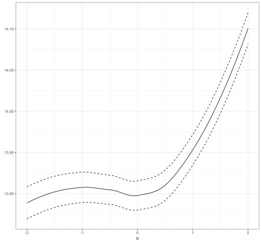
(b) $\phi_2(t)$

Figure 5: Estimates of the eigenfunctions for $\phi_1(t)$ and $\phi_2(t)$ (solid-average of the estimated function; dashed-95% pointwise confident band).

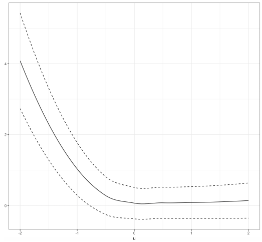
Analysis of the ALSPAC study

- Figure 5(a) for the first eigenfunctions $\phi_1(\cdot)$ shows that the periodicity of variation of the BMI which achieving peaks and troughs roughly at infancy, 5 years old, 12 years old and 18 years old. Figure 5(b) for the second eigenfunctions $\phi_2(\cdot)$ implies that the BMI has large fluctuation at 5 years old and 18 years old.

Analysis of the ALSPAC study



(a) $\rho_1(u)$



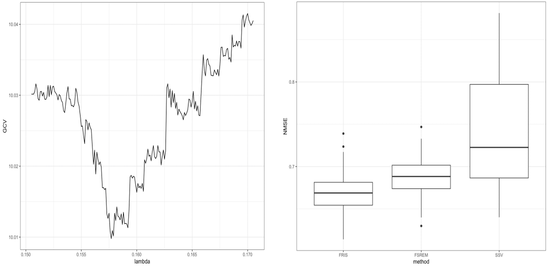
(b) $\rho_2(u)$

Figure 6: Estimates of the score variance functions for $\rho_1(u)$ and $\rho_2(u)$ (solid-average of the estimated function; dashed-95% pointwise confident band).

Analysis of the ALSPAC study

- Figure 6(a) shows that $\rho_1(u)$ is nonzero for all individuals, while Figure 6(b) shows that some individuals may have a zero value for $\rho_2(u)$, suggesting $\phi_2(t)$ is not necessary for all individuals.
- Particularly, those satisfying $\mathbf{X}_i'\alpha_2 \geqslant -0.3$ will be expressed only by eigenfunctions $\phi_1(t)$, while the others by both $\phi_1(t)$ and $\phi_2(t)$.

Analysis of the ALSPAC study



(a) (b)

Figure 7: (a) GCV results of tuning parameters λ; (b) NMSEs of FRIS, SSV and FSREM for Avon Longitudinal Study of Parents and Children.

Thanks For Your Attention!

Reference I

Backenroth, D., Goldsmith, J., Harran, M. D., Cortes, J. C., Krakauer, J. W., and Kitago, T. (2018). Modeling motor learning using heteroskedastic functional principal components analysis. *Journal of the American Statistical Association*, 113:1003–1015.

Chen, X., Li, H., Liang, H., and Lin, H. (2019). Functional response regression analysis. *Journal of Multivariate Analysis*, 169:218–233.

Chiou, J., Müller, H., and Wang, J. (2003a). Functional response models. *Statistica Sinica*, 14:675–693.

Reference II

Chiou, J. M., Müller, H. G., and Wang, J. L. (2003b). Functional quasi-likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society, Series B*, 65:405–423.

Jiang, C. R. and Wang, J. L. (2010). Covariate-adjusted functional principal components analysis for longitudinal data. *The Annals of Statistics*, 38:1194–1226.

Li, G., Shen, H., and Huang, J. (2016). Supervised sparse and functional principal component analysis. *Journal of Computational and Graphical Statistics*, 26:859–878.

Reference III

Li, J., Huang, C., and Zhu, H. (2017). A functional varying-coefficient single-index model for functional response data a functional varying-coefficient single-index model for functional response data. *Journal of the American Statistical Association*, 112(519):1169–1181.

Morris, J. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B*, 68:179–19.

Stone, C. (1980). Optimal rate of convergence for nonparametric estimators. *The Annals of Statistics*, 8:1348–1360.