Principal Component Models for Sparse Functional Data

GARETH M. JAMES
TREVOR J. HASTIE
CATHERINE A. SUGAR

July 18, 2019

Outline

1. Introduction

2. The Growth Curve Data

3. The Reduced Rank Model

4. Fitting the Reduced Rank Model

5. The Reduced Rank and Mixed Effects Methods Compared

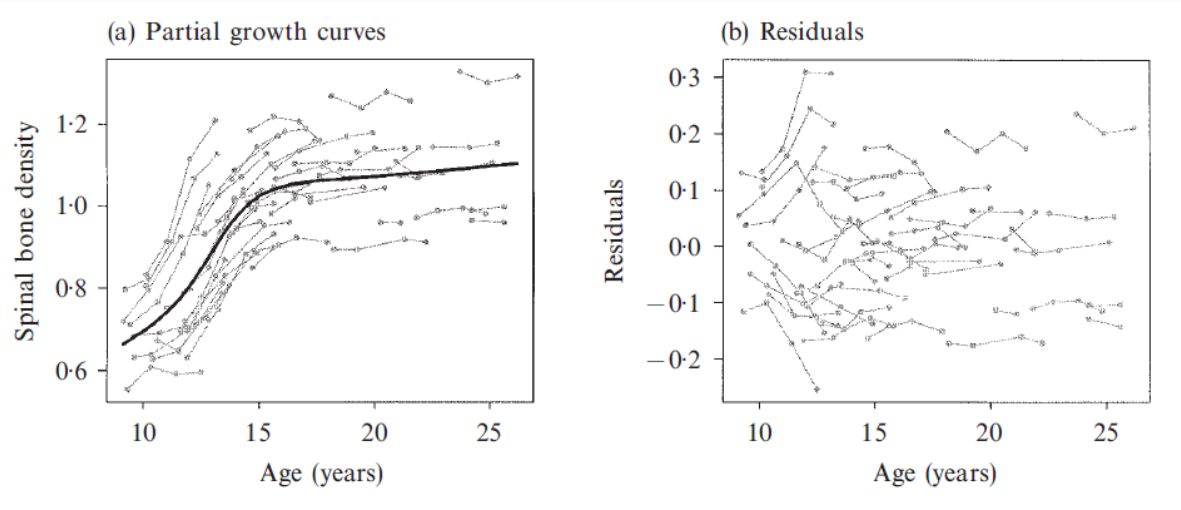
6. Model Selection and Inference

7. Comparison of the Reduced Rank Method and Classical Principal Components

8. Appendix

Introduction

Introduction



- If each curves observed different time points, it is not good way to apply method with equal time points.
- In this paper, present an estimation technique when the data are sparse and measured at the different time points each curves.

Introduction

The direct method

When the curves are not measured at common time points, we estimate the curves by basis expansion and then perform PCA on the estimated curves.

Drawbacks

- If each curves have few observation, there are not existed unique basis coefficients.
- The solution is not optimal ($\because$ perform PCA onto estimated curves)

Mixed effect model

$$Y_i(t) = \mathbf{b}(t)^T \boldsymbol{\beta} + \mathbf{b}(t)^T \boldsymbol{\gamma}_i + \epsilon_i(t), \ i = 1, \ldots, N$$

where $\mathbf{b}(t) = [b_1(t), \ldots, b_q(t)]^T$ is the vector of spline basis functions, $\boldsymbol{\beta}$ is a fixed vector of spline coefficients, $\boldsymbol{\gamma}_i$ is a random vector of spline coefficients with covariance matrix $\boldsymbol{\Gamma}$,

$$\mathbf{Y}_i = \mathbf{B}_i \boldsymbol{\beta} + \mathbf{B}_i \boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i$$

where $\mathbf{Y}_i$ is the n_i-dimensional vector, $\mathbf{B}$ is the $n_i \times q$ spline basis matrix

Fitting the mixed effect model

- `EM` algorithm is used to fit the mixed effect model to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$.
- Given these estimates $\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$, predictions are obtained for the $\boldsymbol{\gamma}_i$'s using best linear unbiased prediction (BLUP)

$$\hat{\boldsymbol{\gamma}}_i = (\hat{\boldsymbol{\Gamma}}^{-1}/\hat{\sigma}^2 + \mathbf{B}_i^T\mathbf{B}_i)^{-1}\mathbf{B}_i^T(\mathbf{Y}_i - \mathbf{B}_i\hat{\boldsymbol{\beta}})$$

- Using the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$, we can estimate the mean and PC curves.
- Using the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$ and the prediction for $\boldsymbol{\gamma}_i$, we can predict the individual curve $\mathbf{Y}_i(t)$.

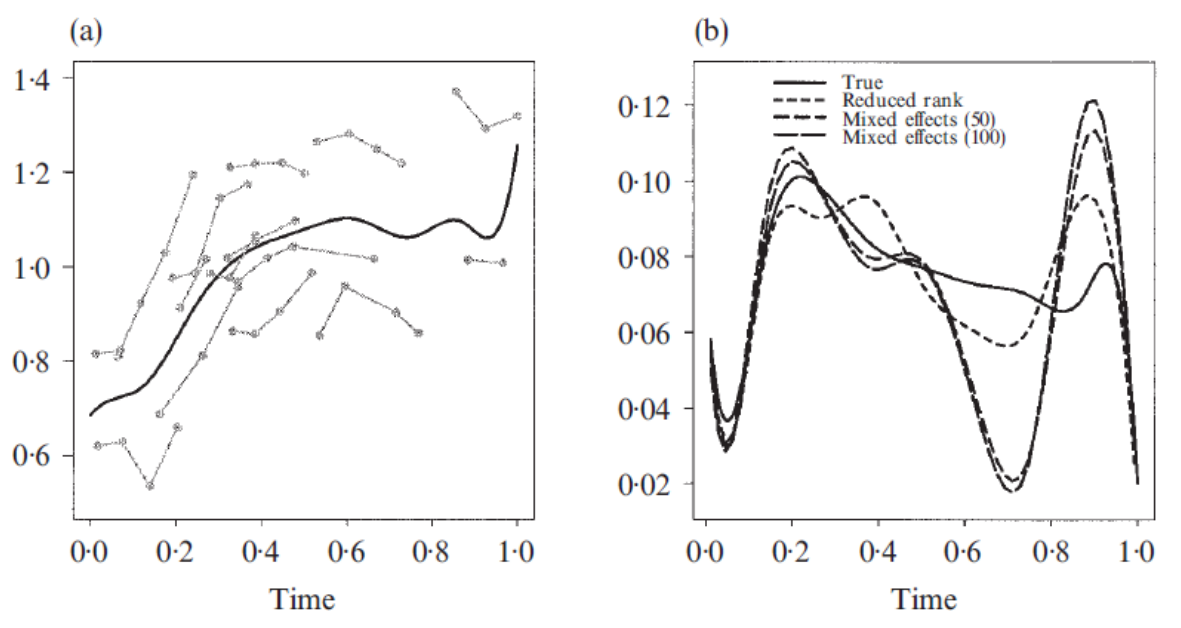
Advantages of the mixed effects model

- When the curve $Y_i(t)$ is insufficient, it can be used for estimating $Y_i(t)$ because of using all observation rather than ith curve.
- Using MLE to estimate $\boldsymbol{\beta}, \boldsymbol{\Gamma}$

Some problems with the mixed effects method

- The dimension of spline basis $= q$
  $\Rightarrow$ # of parameters of $\mathbf{\Gamma} = \frac{q(q+1)}{2}$
- When the data are sparse, the estimate of $\mathbf{\Gamma}$ is highly variable.
  $\Rightarrow$ `EM` algorithm converges to local maximum.

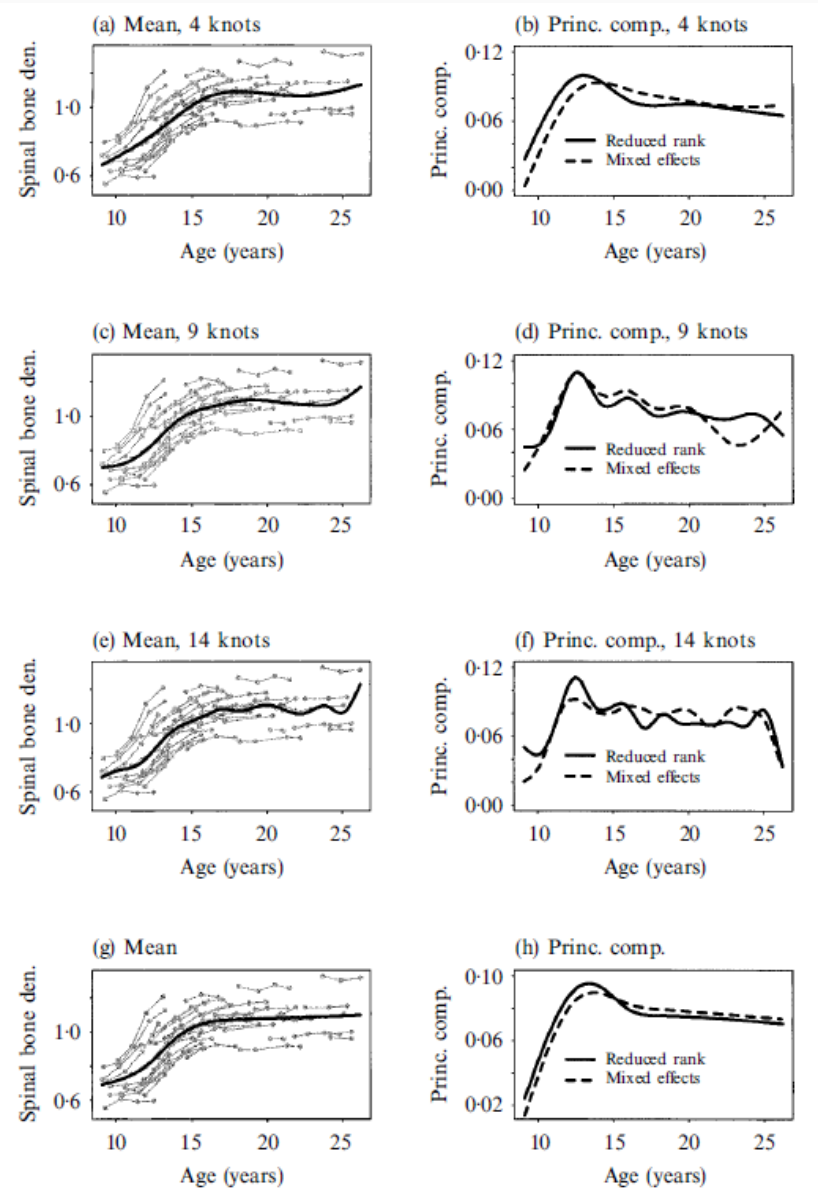
Introduction



- The direct method cannot be applied. ($\because n_i \ll q(=11)$)
- It looks overfitting in the mixed effect model.
- The reduced rank method presented in this paper looks better than the mixed effect method.

The Growth Curve Data

The Growth Curve Data



- The curve is more wiggly as the # of knots increases.
- In the reduced rank model, the peak is looked clear when $Age = 13$.
- In the mixed effect model, the peak is not looked clear.

The Reduced Rank Model

The reduced rank model

Generalized additive model

$$Y_i(t) = \mu(t) + \sum_{j=1}^{k} f_j(t)\alpha_{ij} + \epsilon_i(t) = \mu(t) + \boldsymbol{f}(t)^T\boldsymbol{\alpha}_i + \epsilon_i(t)$$

subject to $\int f_j f_l = \delta_{jl},\ the\ Kronecker\ \delta,\ \delta_{jl} = \begin{cases} 1, & j = l \\ 0, & j \neq l \end{cases}$

where $\mu(t)$ is overall mean function, f_j is the jth PC function, $\boldsymbol{f} = (f_1, \ldots, f_k)^T$, $\boldsymbol{\alpha}_i \sim (\mathbf{0}, \Sigma)$, $\epsilon_i(t) \sim (\mathbf{0}, \sigma^2)$ and uncorrelated

In this paper, we let Σ is diagonal.

The reduced rank model

Restrictions when data measured finite # of time points

$$\mu(t) = \mathbf{b}(t)^T \boldsymbol{\theta}_\mu, \quad \boldsymbol{f}(t)^T = \mathbf{b}(t)^T \boldsymbol{\Theta}$$

where $\mathbf{b}(t)$ is a spline basis with dimension q, $\boldsymbol{\Theta}$ is $q \times k$ spline coefficients matrix, $\boldsymbol{\theta}_\mu$ is q-dimensional vector of spline coefficients

Restricted model

$$Y_i(t) = \mathbf{b}(t)^T \boldsymbol{\theta}_\mu + \mathbf{b}(t)^T \boldsymbol{\Theta} \boldsymbol{\alpha}_i + \epsilon_i(t)$$

$$\epsilon_i(t) \sim (\mathbf{0}, \sigma^2), \ \boldsymbol{\alpha}_i \sim (\mathbf{0}, \mathbf{D})$$

subject to

$$\boldsymbol{\Theta}^T \boldsymbol{\Theta} = \mathbf{I}, \ \int \mathbf{b}(t)^T \mathbf{b}(t) dt = 1, \ \int \int \mathbf{b}(t)^T \mathbf{b}(s) dt = 0$$

where $\mathbf{D}$ is diagonal.

The reduced rank model

Reduced rank model

$$\mathbf{Y}_i = \mathbf{B}_i\boldsymbol{\theta}_\mu + \mathbf{B}_i\boldsymbol{\Theta}\boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_i$$

$$\boldsymbol{\Theta}^T\boldsymbol{\Theta} = \mathbf{I},\ \boldsymbol{\epsilon}_i \sim (\mathbf{0}, \sigma^2\mathbf{I}),\ \boldsymbol{\alpha}_i \sim (\mathbf{0}, \mathbf{D})$$

- Parameters to estimate : $\boldsymbol{\theta}_\mu, \boldsymbol{\Theta}, \mathbf{D}, \sigma^2$
- mixed effect model + rank constraint on the covariance matrix (# of PCs)

The reduced rank model

Random vector with unrestricted covariance matrix

$$\gamma_i = [\Theta, \Theta^*] \begin{pmatrix} \alpha_i \\ \alpha_i^* \end{pmatrix}$$

where γ_i is q-dimensional vector in mixed effect model, Θ^* is a $q \times (q-k)$ matrix, α_i is a random vector of length $q-k$ with diagonal covariance matrix.

Mixed effects model

$$\mathbf{Y}_i = \mathbf{B}_i \theta_\mu + \mathbf{B}_i \Theta \alpha_i + \mathbf{B}_i \Theta^* \alpha_i^* + \epsilon_i$$

- The ruduced rank model is a submodel of the mixed effects model.
- Set $\alpha_i^* = 0$, it can be simply fitting the reduced rank model using a different algorithm.

The reduced rank model

Table 1. *Loglikelihoods for the fits in Fig. 3(a)–(f)*

Number of knots	Loglikelihood	
	Constrained	Reduced rank
4	380·63	389·22
9	394·75	409·81
14	399·00	411·36

- The constrained mixed effects model obtained from the mixed effects model after setting the $\boldsymbol{\alpha}_i^* = 0$.
- The reduced rank likelihood is strictly higer than the constrained likelihood.

Fitting the Reduced Rank Model

Fitting the Reduced Rank Model

Goal of FPCA

- Estimate μ and f
- Prediction of the basis coefficients $\boldsymbol{\alpha}_i$

Assuming the spline fit to the functions, it is equivalent to estimate $\boldsymbol{\theta}_\mu$, $\boldsymbol{\Theta}$ and predict the $\boldsymbol{\alpha}_i$

- Parameters to estimate : $\boldsymbol{\theta}_\mu, \boldsymbol{\Theta}, \sigma^2, \mathbf{D}$
- $\mathbf{D}$: variability explained by each PC curve
- σ^2 : variability left unexplained

Two fitting procedures

- Maximum likelihood
- Penalized least squares

Fitting the Reduced Rank Model

Assume that $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ and $\boldsymbol{\alpha}_i \sim N(\mathbf{0}, \mathbf{D})$.

$$\mathbf{Y}_i \sim N(\mathbf{B}_i\boldsymbol{\theta}_\mu, \sigma^2\mathbf{I} + \mathbf{B}_i\boldsymbol{\Theta}\mathbf{D}\boldsymbol{\Theta}^T\mathbf{B}_i^T)$$

The observed likelihood of the $\mathbf{Y}_i$'s is

$$\prod_{i=1}^{N} \frac{1}{(2\pi)^{n_i/2}|\sigma^2\mathbf{I} + \mathbf{B}_i\boldsymbol{\Theta}\mathbf{D}\boldsymbol{\Theta}^T\mathbf{B}_i^T|^{1/2}}$$

$$\times \exp\left\{ -\frac{1}{2}(\mathbf{Y}_i - \mathbf{B}_i\boldsymbol{\theta}_\mu)^T(\sigma^2\mathbf{I} + \mathbf{B}_i\boldsymbol{\Theta}\mathbf{D}\boldsymbol{\Theta}^T\mathbf{B}_i^T)^{-1}(\mathbf{Y}_i - \mathbf{B}_i\boldsymbol{\theta}_\mu)\right\}$$

- But to maximize this likelihood is a nonconvex optimization problem.

Fitting the Reduced Rank Model

If $\boldsymbol{\alpha}_i$'s were observed, the joint likelihood is simplified to

$$\prod_{i=1}^{N} \frac{1}{(2\pi)^{(n_i+k)/2}\sigma^{n_i}|\mathbf{D}|^{\frac{1}{2}}}$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{Y}_i-\mathbf{B}_i\boldsymbol{\theta}_\mu-\mathbf{B}_i\boldsymbol{\Theta}\boldsymbol{\alpha}_i)^T(\mathbf{Y}_i-\mathbf{B}_i\boldsymbol{\theta}_\mu-\mathbf{B}_i\boldsymbol{\Theta}\boldsymbol{\alpha}_i)-\frac{1}{2}\boldsymbol{\alpha}_i^T\mathbf{D}^{-1}\boldsymbol{\alpha}_i\right\}$$

- Let $\boldsymbol{\alpha}_i$ is unobserved(missing), we can maximize this likelihood much easier using `EM` algorithm.

Fitting the Reduced Rank Model

We choose $\boldsymbol{\theta}_\mu, \boldsymbol{\Theta}, \boldsymbol{\alpha}_i$'s to minimize the sum of squared residuals,

$$\sum_{i=1}^{N} \left\{ (\mathbf{Y}_i - \mathbf{B}_i\boldsymbol{\theta}_\mu - \mathbf{B}_i\boldsymbol{\Theta}\boldsymbol{\alpha}_i)^T(\mathbf{Y}_i - \mathbf{B}_i\boldsymbol{\theta}_\mu - \mathbf{B}_i\boldsymbol{\Theta}\boldsymbol{\alpha}_i) + \sigma^2 \sum_{j=1}^{K} \frac{\alpha_{ij}^2}{\mathbf{D}_{jj}} \right\}$$

- Minimize SSE $\Leftrightarrow$ Maximize likelihood
- Distribution assumption X
- It can be maximized using `EM` algorithm.

The Reduced Rank and Mixed Effects Methods Compared

The Reduced Rank and Mixed Effects Methods Compared

Study 1

- Generate data from the mean function and PC curve corresponding to reduced rank fit using spline with knots at ages 12, 14, 16 and 18.
- 48 curves were generated using same time points as the growth curve data.

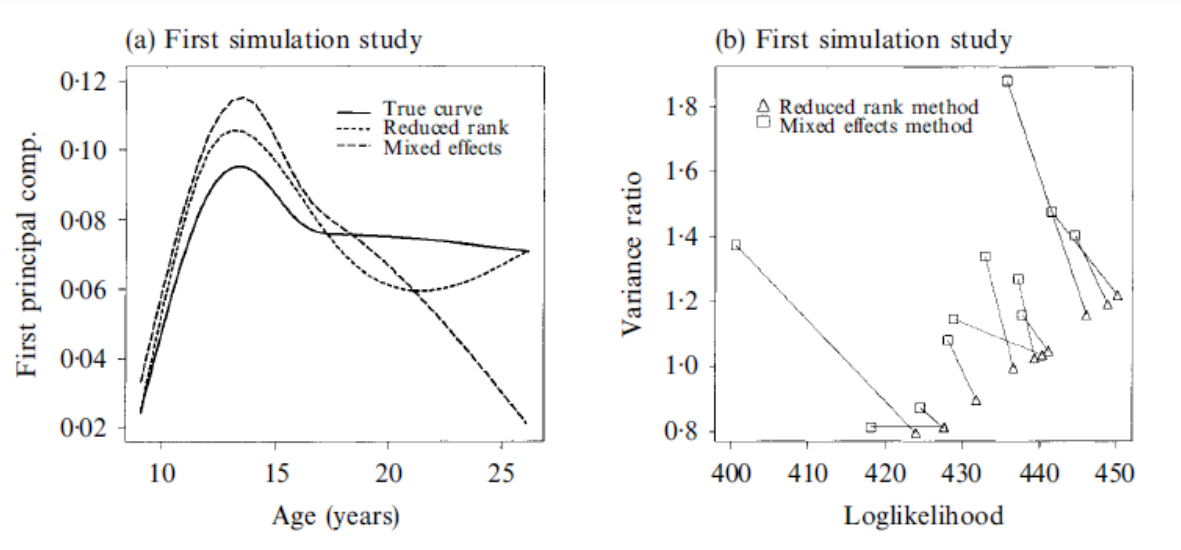
Study 2

- Generate data in the same way as study 1 except that using spline with 7 equally spaced knots were used.
- 16 curves were generated using randomly selected time points of the original 48 growth curves.

In two studies, mixed effects and reduced rank model were fitted to 10 datasets using natural cubic splines.

The Reduced Rank and Mixed Effects Methods Compared

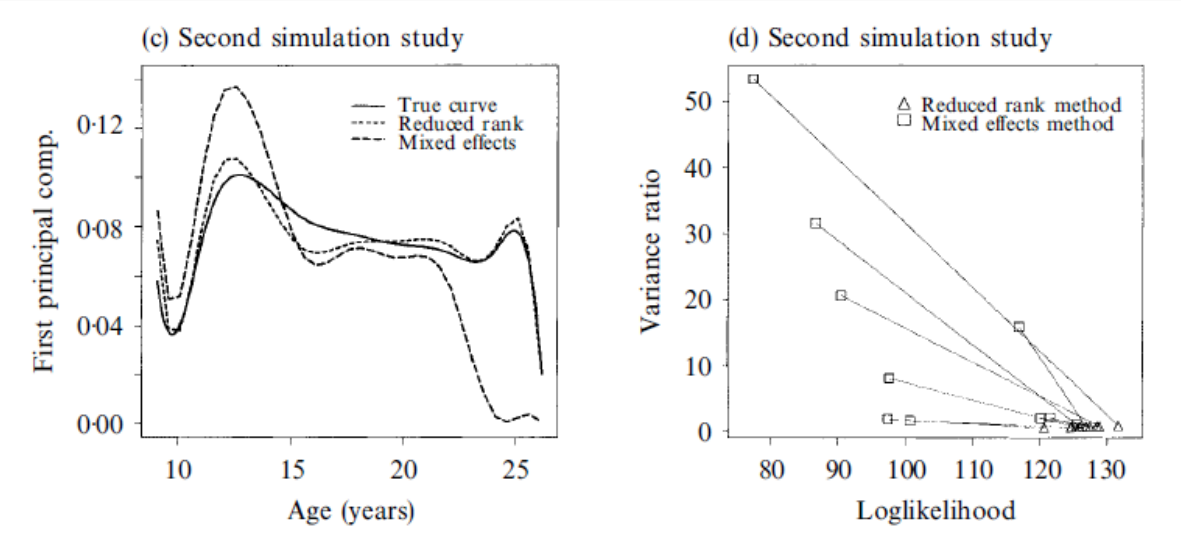
Study 1



- In Figure (a), the reduced rank fit looks better than the mixed effects fit.
- In Figure (b),
  almost all the variance ratios for mixed effects fits > 1
  $\Rightarrow$ mixed effects fit underestimates the variance.(overfitting)

The Reduced Rank and Mixed Effects Methods Compared

Study 2

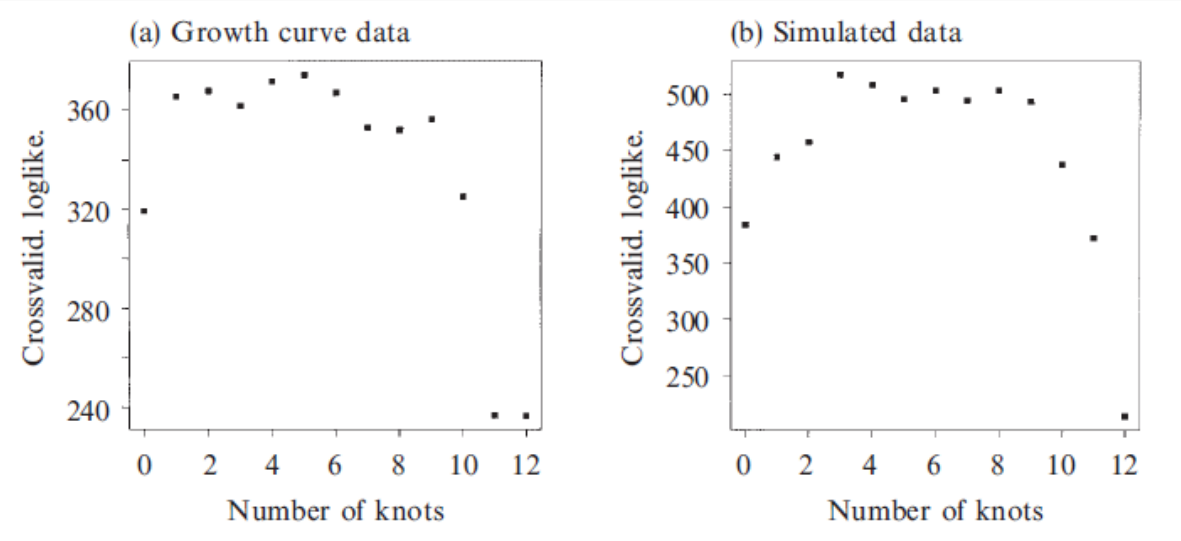


- In Figure (c), the reduced rank fit looks better than the mixed effects fit.
- In Figure (d), overfitting has increased on mixed effects fit.
- The likelihood estmate becomes worse as the variance is underestimated.

Model Selection and Inference

Model Selection and Inference

Select the number using cross validation to maximize the loglikelihood for different numbers of knots.



—

- In Figure (a), the optimal number of knots is between 4 and 6, we opted 4 knots for the law of parsimony.
- In Figure (b), data simulated from a spline with 4 knots but likelihood is maximized for 3 knots.

Model Selection and Inference

Selecting the rank of the covariance matrix $=$ Selecting $\#$ of PCs

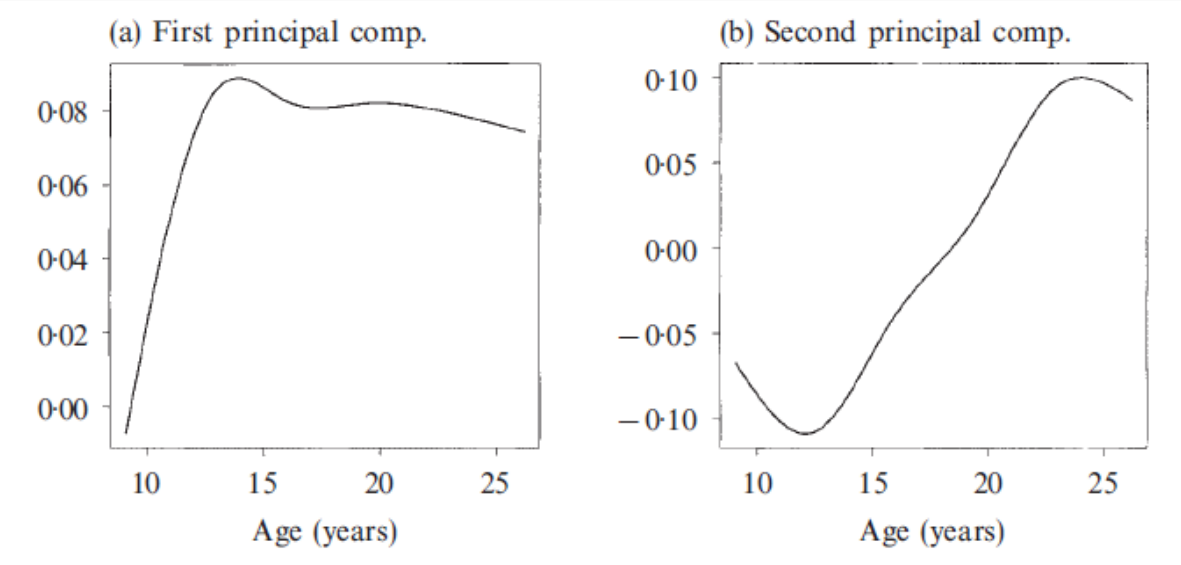
Methods to select the rank, k

- PVE (Proportion of Variance Explained)
- Comparison loglikelihood with different k

PVE

- PVE is difficult to compute directly in FPCA.
- If $\sigma^2 \to 0$ and all curves are measured at similar time points, $Var(\mathbf{Y}_i) \approx Var(\boldsymbol{\alpha}_i) = \mathbf{D}$
- PVE $= \frac{\mathbf{D}_{jj}}{tr(\mathbf{D})}$, $(\because$ $\mathbf{D}$ is diagonal)

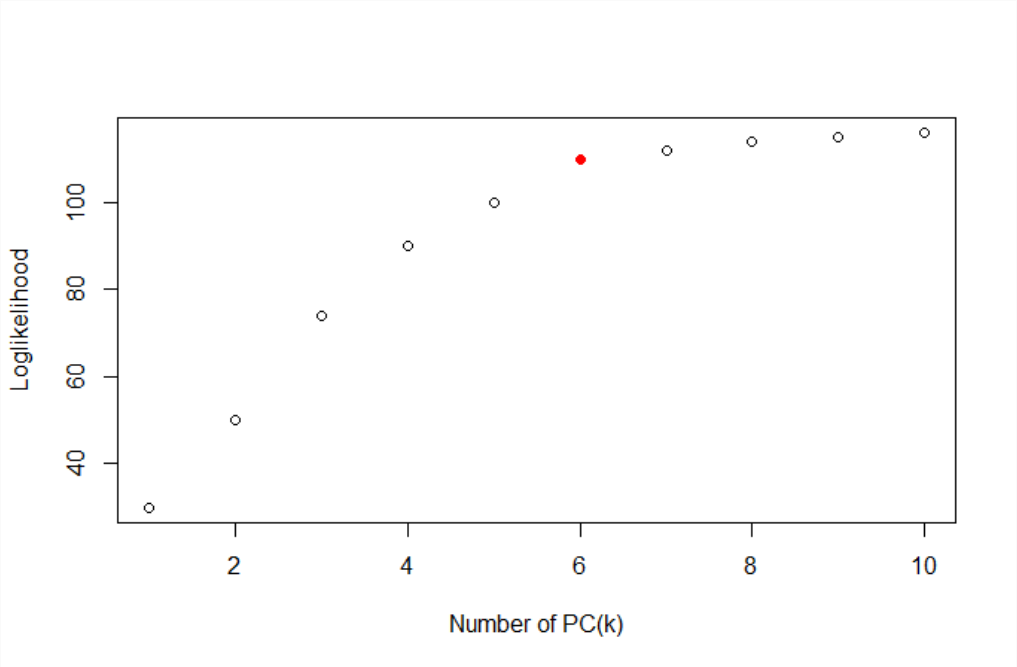
Model Selection and Inference



- In 1st PC, a sharp peak indicates the puberty periods which shows the highest variability with each curves.
- 2nd PC indicates differences in the slopes of individual curves. If the weight of PC2 increases, this curve has slope greater than average.

Comparison loglikelihood with different k

- The loglikelihood increases as k increases.
- Pick the k when the increase levelled off.
- Example of graphical selection



Comparison loglikelihood with different k

- LRT test
  - $H_0 : k = k_0$ vs $H_1 : k = k_1$
  - $LRT = 2(\log L_{k=k_1} - \log L_{k=k_0}) \dot{\sim} \chi^2_{df}$, where $df =$ difference of parameters
- Example
  - $H_0 : k = 1$ vs $H_1 : k = 2$
  - $LRT = 2(\log L_{k=2} - \log L_{k=1}) = 19.28 \dot{\sim} \chi^2_5$, p-value $= 0.002$
  - Reject H_0, we choose $k = 2$.
  - But this dataset is sparse, we shoud caution when using an asymptotic result

Model Selection and Inference

To compute confidence intervals for the overall mean function, PCs and individual curves, we use the bootstrap.

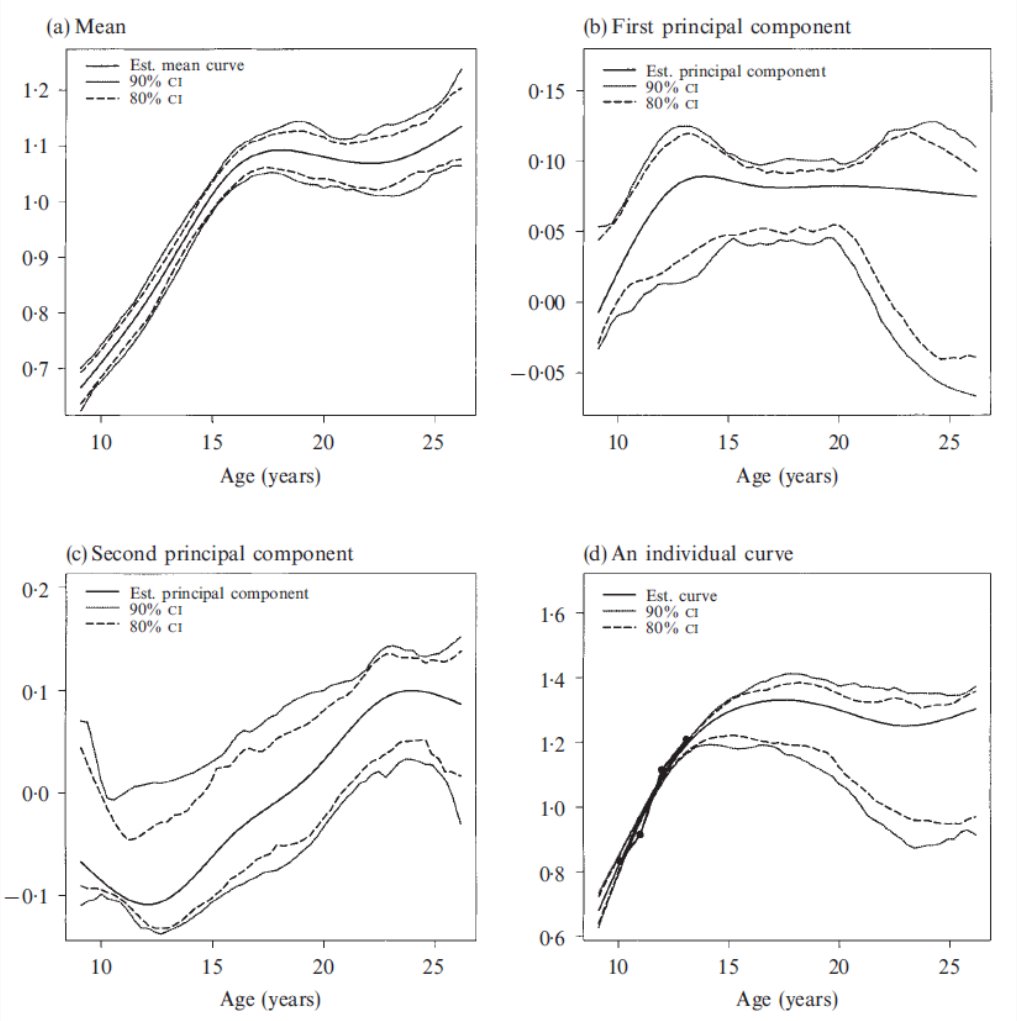
Two methods to bootstrap curve data

- Resampling the individual curves.
  - No parametric assumption
  - Sparse data $\Rightarrow$ performance $\downarrow$
- Resampling the estimated $\boldsymbol{\alpha}_i$'s and residuals and generating new partial curves based on these values.
  - Bootstrap datasets have observations at the same time points as the original dataset.

Confidence intervals on the growth curve data

- Generate 100 bootstrap samples and fit the reduced rank method with $k = 2$ to each
- Using the bootstrap percentile method, we computed 80% and 90% CIs.

Model Selection and Inference



Comparison of the Reduced Rank Method and Classical Principal Components

Comparison of the Reduced Rank Method and Classical Principal Components

The linear model

$$\mathbf{X}_i = \boldsymbol{\theta}_\mu + \boldsymbol{\Theta}\boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_i$$

$$\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}),\ \boldsymbol{\alpha}_i \sim N(\mathbf{0}, \mathbf{D})$$

where $\mathbf{X}_i$ are q-dimensional data vectors, $\boldsymbol{\Theta}$ is an orthogonal matrix

- $\boldsymbol{\Sigma}$ is diagonal $\Rightarrow$ MLE of linear model is equivalent to factor analysis solution.
- $\boldsymbol{\Sigma} = \sigma^2\mathbf{I} \Rightarrow$ if $\sigma^2 \to 0$, then MLE of linear model is equivalent to classical PCA solution with minimizing

$$\sum_{i=1}^{N} \|\mathbf{X}_i - \boldsymbol{\theta}_\mu - \boldsymbol{\Theta}\boldsymbol{\alpha}_i\|^2$$

Comparison of the Reduced Rank Method and Classical Principal Components

The reduced rank model

$$\mathbf{Y}_i = \mathbf{B}_i\boldsymbol{\theta}_\mu + \mathbf{B}_i\boldsymbol{\Theta}\boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_i,\ Cov(\boldsymbol{\epsilon}_i) = \sigma^2\mathbf{I}$$

where $\boldsymbol{\Theta}$ is the PCs and the $\boldsymbol{\alpha}_i$'s are weights for the PC

- $Cov(\boldsymbol{\epsilon}_i) = \sigma^2\mathbf{I} \Rightarrow$ Solution of reduced rank model is equivalent to generalized PCA solution.
- On penalized least squares objective function, if $\sigma^2 \to 0$, then the procedure for fitting the reduced rank model simply minimizes

$$\sum_{i=1}^{N} \|\mathbf{Y}_i - \mathbf{B}_i\boldsymbol{\theta}_\mu - \mathbf{B}_i\boldsymbol{\Theta}\boldsymbol{\alpha}_i\|^2$$

Comparison of the Reduced Rank Method and Classical Principal Components

Let $\hat{\boldsymbol{\gamma}}_i = (\mathbf{B}_i^T\mathbf{B}_i)^{-1}\mathbf{B}_i^T\mathbf{Y}_i$, LSE of the spline coefficients for ith curve, then objective equation is

$$\sum_{i=1}^{N} \|\mathbf{Y}_i - \mathbf{B}_i\hat{\boldsymbol{\gamma}}_i\|^2 + \sum_{i=1}^{N} \|\mathbf{B}_i\hat{\boldsymbol{\gamma}}_i - \mathbf{B}_i\boldsymbol{\theta}_\mu - \mathbf{B}_i\boldsymbol{\Theta}\boldsymbol{\alpha}_i\|^2$$

$$= C(\mathbf{Y}) + \sum_{i=1}^{N} \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\theta}_\mu - \boldsymbol{\Theta}\boldsymbol{\alpha}_i\|^2_{\mathbf{B}_i^T\mathbf{B}_i}$$

where $C(\mathbf{Y})$ is a constant with the parameters.

Therefore, we minimize

$$\sum_{i=1}^{N} \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\theta}_\mu - \boldsymbol{\Theta}\boldsymbol{\alpha}_i\|^2_{\mathbf{B}_i^T\mathbf{B}_i}$$

Comparison of the Reduced Rank Method and Classical Principal Components

- If all curves are measured at the same time points, then $\mathbf{B}_i = \mathbf{B}$.
- WLOG, assume $\mathbf{B}'\mathbf{B} = \mathbf{I}$, minimizing the objective equation is equivalent to performing standard PCA on the spline coefficients. $\Rightarrow$ the direct method
- Similarly to classical PCA, When all curves are measured at the same time points, the reduced rank method finds the best plane using the Euclidean metric.
- Also it finds the best plane when the curves are not measured at the same time points, but the distance between the plane and each data point is measured relative to the metric $\mathbf{B}_i^T\mathbf{B}_i$.

Appendix

Appendix

The mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon}$$

$\boldsymbol{u} \sim (\mathbf{0}, \boldsymbol{G})$, $\boldsymbol{\epsilon} \sim (\mathbf{0}, \boldsymbol{R})$, $Var(\mathbf{y}) = \boldsymbol{V} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T + \boldsymbol{R}$, $Cov(\boldsymbol{u}, \boldsymbol{\epsilon}) = \mathbf{0}$

Assume $\mathbf{y}|\boldsymbol{u} \sim N(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}, \boldsymbol{R})$ and $\boldsymbol{u} \sim N(\mathbf{0}, \boldsymbol{G})$,
then likelihood function is

$$-2\log f(\mathbf{y}|\boldsymbol{u}) = N\log 2\pi + \log|\boldsymbol{R}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{u})^T \boldsymbol{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{u})$$

$$-2\log f(\boldsymbol{u}) = q\log 2\pi + \log|\boldsymbol{G}| + \boldsymbol{u}^T \boldsymbol{G}^{-1}\boldsymbol{u}$$

Maximizing $l(\boldsymbol{\beta}, \boldsymbol{u}|\mathbf{y}) = \log f(\mathbf{y}, \boldsymbol{u})$ is equivalent to minimizing the sum of the two equations.

Appendix

Differentiating the sum of the two equations

$$\frac{\partial[-2l(\boldsymbol{\beta}, \boldsymbol{u}|\mathbf{y})]}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T\boldsymbol{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{u})$$
$$\frac{\partial[-2l(\boldsymbol{\beta}, \boldsymbol{u}|\mathbf{y})]}{\partial \boldsymbol{u}} = -2\boldsymbol{Z}^T\boldsymbol{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{u}) + 2\boldsymbol{G}^{-1}\boldsymbol{u}$$

equating them to 0 gives the following system

$$\mathbf{X}^T\boldsymbol{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z}\hat{\boldsymbol{u}} = \mathbf{X}^T\boldsymbol{R}^{-1}\mathbf{y}$$

$$\boldsymbol{Z}^T\boldsymbol{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + (\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})\hat{\boldsymbol{u}} = \boldsymbol{Z}^T\boldsymbol{R}^{-1}\mathbf{y}$$

called Henderson's mixed model equations (HMME)

Appendix

The matrix form of HMME is

$$\begin{pmatrix} \mathbf{X}^T\boldsymbol{R}^{-1}\mathbf{X} & \mathbf{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\mathbf{X} & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z}+\boldsymbol{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\boldsymbol{R}^{-1}\mathbf{y} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\mathbf{y} \end{pmatrix}$$

Solve this equation,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{y} = \hat{\boldsymbol{\beta}}^{GLS}$$
$$\hat{\boldsymbol{u}} = (\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z}+\boldsymbol{G}^{-1})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})$$

where $\hat{\boldsymbol{\beta}}$ is BLUE(Best Linear Unbiased Estimator) and $\hat{\boldsymbol{u}}$ is BLUP(Best Linear Unbiased Prediction)

Reference

Reference

GARETH M. JAMES, TREVOR J. HASTIE, CATHERINE A. SUGAR

Principal component models for sparse functional data

Biometrika, 87(3):587–602, 2000.

J.O. Ramsay, B.W. Silverman.

Functional Data Analysis 2nd edition.

Springer, 2005.