

Functional Linear Models

Statistical Models

So far we have focussed on *exploratory data analysis*

- Smoothing
- Functional covariance
- Functional PCA

Now we wish to examine predictive relationships → generalization of linear models.

Functional Linear Regression

$$y_i = \alpha + \mathbf{x}_i\beta + \epsilon_i$$

Three different scenarios for y_i \mathbf{x}_i

- Functional covariate, scalar response
- Scalar covariate, functional response
- Functional covariate, functional response

We will deal with each in turn.

Scalar Response Models

Generalization of multiple linear regression

$$y_i = \alpha + \sum \beta_j x_i(t_j) + \epsilon_i = \alpha + \mathbf{x}_i \beta + \epsilon_i$$

becomes

$$y_i = \alpha + \int \beta(t) x_i(t) dt + \epsilon_i$$

General trick: functional data model = multivariate model with sums replaced by integrals.

Identification

Problem:

- In linear regression, we must have fewer covariates than observations.
- If I have $y_i, x_i(t)$, there are *infinitely* many covariates.

$$y_i = \alpha + \int \beta(t)x_i(t)dt + \epsilon_i$$

Estimate β by minimizing squared error:

$$\beta(t) = \operatorname{argmin} \sum \left(y_i - \alpha - \int \beta(t)x_i(t)dt \right)^2$$

But I can always make the $\epsilon_i = 0$.

Smoothing

Additional constraints: we want to insist that $\beta(t)$ is smooth.

Add a smoothing penalty:

$$\text{PENSSE}_\lambda(\beta) = \sum_{i=1}^n \left(y_i - \alpha - \int \beta(t) x_i(t) dt \right)^2 + \lambda \int [L\beta(t)]^2 dt$$

Very much like smoothing

Still need to represent $\beta(t)$ – use a basis expansion:

$$\beta(t) = \sum c_i \phi_i(t).$$

Calculation

$$y_i = \alpha + \int \beta(t)x_i(t)dt + \epsilon_i = \alpha + \left[\int \Phi(t)x_i(t)dt \right] \mathbf{c} + \epsilon_i$$

$$= \alpha + \mathbf{x}_i \mathbf{c} + \epsilon_i$$

for $\mathbf{x}_i = \int \Phi(t)x_i(t)dt$. With $Z_i = [1\mathbf{x}_i]$,

$$\mathbf{y} = Z \begin{bmatrix} \alpha \\ \mathbf{c} \end{bmatrix} + \epsilon$$

and with smoothing penalty matrix R_L :

$$[\hat{\alpha} \ \hat{\mathbf{c}}^T]^T = (Z^T Z + \lambda R_L)^{-1} Z^T \mathbf{y}$$

Then

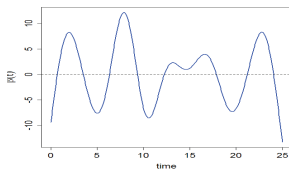
$$\hat{y} = \int \hat{\beta}(t)x_i(t)dt = Z \begin{bmatrix} \hat{\alpha} \\ \hat{\mathbf{c}} \end{bmatrix} = S_{\lambda} \mathbf{y}$$

Choosing Smoothing Parameters

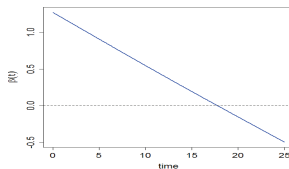
Cross-Validation:

$$CV(\lambda) = \sum \left(\frac{y_i - \hat{y}_i}{1 - S_{ii}} \right)^2$$

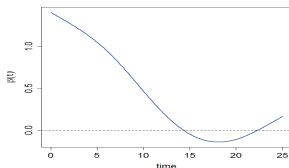
$$\lambda = e^{-1}$$



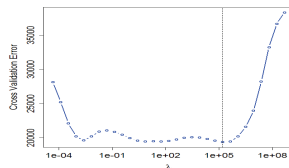
$$\lambda = e^{20}$$



$$\lambda = e^{12}$$



CV Error



Confidence Intervals

Assuming independent

$$\epsilon_i \sim N(0, \sigma_e^2)$$

We have that

$$\text{Var} \begin{bmatrix} \hat{\alpha} \\ \hat{\mathbf{c}} \end{bmatrix} = \left[\left(Z^T Z + \lambda R \right)^{-1} Z^T \right] [\sigma_e^2 \mathbb{I}] \left[Z \left(Z^T Z + \lambda R \right)^{-1} \right]$$

Estimate

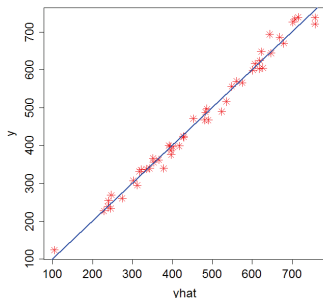
$$\hat{\sigma}_e^2 = SSE / (n - df), \quad df = \text{trace}(S_\lambda)$$

And (pointwise) confidence intervals for $\beta(t)$ are

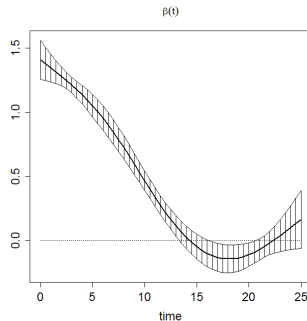
$$\Phi(t)\hat{\mathbf{c}} \pm 2\sqrt{\Phi(t)^T \text{Var}[\hat{\mathbf{c}}] \Phi(t)}$$

Confidence Intervals

$$R^2 = 0.987$$



$$\sigma^2 = 349, \text{ df} = 5.04$$



Extension to multiple functional covariates follows same lines:

$$y_i = \beta_0 + \sum_{j=1}^p \int \beta_j(t) x_{ij}(t) dt + \epsilon_i$$

functional Principal Components Regression

Alternative: principal components regression.

$$x_i(t) = \sum d_{ij} \xi_j(t) \quad d_{ij} = \int x_i(t) \xi_j(t) dt$$

Consider the model:

$$y_i = \beta_0 + \sum \beta_j d_{ij} + \epsilon_i$$

- Reduces to a standard linear regression problem.
- Avoids the need for cross-validation (assuming number of PCs is fixed).

fPCA and Functional Regression Interpretation

$$y_i = \beta_0 + \sum \beta_j d_{ij} + \epsilon_i$$

Recall that $d_{ij} = \int x_i(t)\xi_j(t)dt$ so

$$y_i = \beta_0 + \sum \int \beta_j \xi_j(t) x_i(t) dt + \epsilon_i$$

and we can interpret

$$\beta(t) = \sum \beta_j \xi_j(t)$$

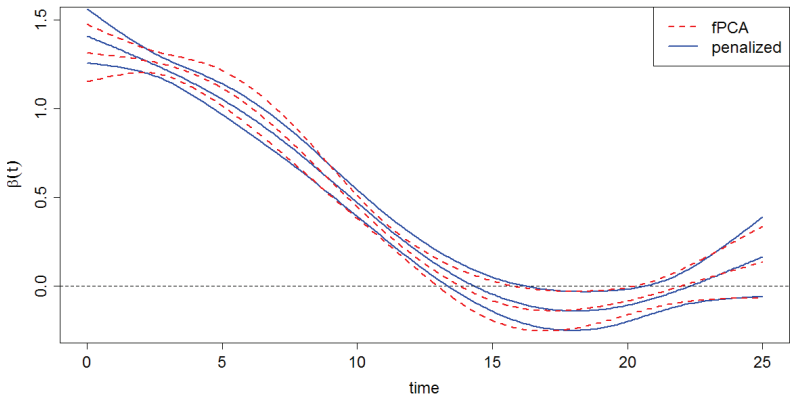
and write

$$y_i = \beta_0 + \int \beta(t) x_i(t) dt + \epsilon_i$$

Confidence intervals derive from variance of the d_{ij} .

A Comparison

Medfly Data: fPCA on 4 components ($R^2 = 0.988$) vs Penalized Smooth ($R^2 = 0.987$)



Advantages of FPCA-based approach

- **Parsimonious** description of functional data as it is the unique linear representation which explains the highest fraction of variance in the data with a given number of components.
- Main attraction is equivalence $X(\cdot) \sim (\xi_1, \xi_2, \dots)$, so that $X(\cdot)$ can be expressed in terms of mean function and the sequence of eigenfunctions and **uncorrelated** FPC scores ξ_k 's.
- For modeling functional regression: Functions $f\{X(\cdot)\}$ have an equivalent function $g(\xi_1, \xi_2, \dots)$
- But need to pay prices
 - FPCs need to be estimated from data (finite sample)
 - Need to choose the number of FPCs

Two Fundamental Approaches

(Almost) all methods reduce to one of

- 1 Perform fPCA and use PC scores in a multivariate method.
- 2 Turn sums into integrals and add a smoothing penalty.

Applied in functional versions of

- generalized linear models
- generalized additive models
- survival analysis
- mixture regression
- ...

Both methods also apply to functional response models.

Functional Response Models

Functional Response Models

Case 1: Scalar Covariates: $(y_i(t), \mathbf{x}_i)$, most general linear model is

$$y_i(t) = \beta_0(t) + \sum_{j=1}^p \beta_j(t) x_{ij}.$$

Conduct a linear regression at each time t

But we might like to smooth; penalize *integrated squared error*

$$\text{PENSISE} = \sum_{i=1}^n \int (y_i(t) - \hat{y}_i(t))^2 dt + \sum_{j=0}^p \lambda_j \int [L_j \beta_j(t)]^2 dt$$

Usually keep λ_j , L_j all the same.

Concurrent Linear Model

Case 2: functional covariates

Extension of scalar covariate model: response only depends on $x(t)$ at the current time

$$y_i(t) = \beta_0(t) + \beta_1(t)x_i(t) + \epsilon_i(t)$$

- $y_i(t)$, $x_i(t)$ must be measured on same time domain.
- Must be appropriate to compare observations time-point by time-point
- Especially useful if $y_i(t)$ is a derivative of $x_i(t)$

Functional Response, Functional Covariate

General case: $y_i(t), x_i(s)$ - a functional linear regression at each time t :

$$y_i(t) = \beta_0(t) + \int \beta_1(s, t) x_i(s) ds + \epsilon_i(t)$$

- Same identification issues as scalar response models.
- Usually penalize β_1 in each direction separately

$$\lambda_s \int [L_s \beta_1(s, t)]^2 ds dt + \lambda_t \int [L_t \beta_1(s, t)]^2 ds dt$$

- Confidence Intervals etc. follow from same principles.

Summary

Three models

Scalar Response Models ■ Functional covariate implies a functional parameter.
■ Use smoothness of $\beta_1(t)$ to obtain identifiability.

Concurrent Linear Model ■ $y_i(t)$ only depends on $x_i(t)$ at the current time.
■ Scalar covariates = constant functions.

Functional Covariate/Functional Response ■ Most general functional linear model.

Other Topics and Recent Developments

- Inference for functional regression models
- Dependent functional data
 - Multilevel/longitudinal/multivariate designs
- Registratoin
- Dynamics
- FDA for sparse longitudinal data
- More general/flexible regression models

Inference for functional regression models

Testing $H_0 : \beta(t) = 0$ under model

$$Y_i = \beta_0 + \int \beta(t) X_i(t) dt + \epsilon_i$$

- Penalized spline approach
 - $\beta(t) = \sum_{m=1}^M \eta_k B_k(t)$
- FPCA-based approach
 - data reduction: $(\xi_{i1}, \dots, \xi_{iK})$
 - multivariate regression: $Y_i \sim \beta_1 \xi_{i1} + \dots + \beta_K \xi_{iK}$
- Difficulty in inference arising from
 - regularization (smoothing)
 - choices of tuning parameters
 - accounting for uncertainty in two-step procedures

Penalized spline approach

- $H_0 : \eta_0 = \eta_1 = \cdots = \eta_M$
- Use roughness penalty $\lambda \int \beta(t)^2 dt$ to avoid overfitting
- Mixed model equivalence representation

$$Y_i = \beta_0 + \sum_{m=1}^M \eta_m V_{im} + \epsilon_i$$
$$(\eta_1, \cdots, \eta_M) \sim N(0, \sigma^2 \Sigma)$$

- Testing $H_0 : \sigma^2 = 0$
- Restricted LRT proposed in the literature.

Swihart, Goldsmith and Crainiceanu (2014). Restricted likelihood ratio tests for functional effects in the functional linear model. Technometrics, 56:483–493.

FPCA-based approach

- $Y_i \sim \beta_1 \xi_{i1} + \dots + \beta_K \xi_{iK}$
- Testing $H_0 : \beta_1 = \dots = \beta_K = 0$
- The number of PCs are chosen by
 - Percent of variance explained (PVE): e.g., 95% or 99%
 - Cross Validation
 - AIC, BIC
- Wald test

$$T = \sum_{k=1}^K \frac{\hat{\beta}_k^2}{\text{var}(\hat{\beta}_k)} = \frac{1}{n\hat{\sigma}_\epsilon^2} \sum_{k=1}^K \frac{Y^T \hat{\xi}_k \hat{\xi}_k^T Y}{\hat{\lambda}_k} \sim \chi_K^2$$

- *But is it a good idea to rank based on $X(t)$ only?
And how sensitive is the power to the choice of K ?*

FPCA-based approach

- Under alternative $H_a : \beta_k \neq 0$, where $\beta_k \neq 0$ for some k , it can be shown that $T \sim \chi_K^2(\eta)$, where

$$\eta = \frac{n}{\sigma_\epsilon^2} \sum_{k=1}^K \lambda_k \beta_k^2$$

- The power contribution of the k^{th} component depends on both λ_k and β_k
- We propose a new association-variation index (AVI):
$$AVI_k = \lambda_k \beta_k^2$$
- Propose to rank and choose PCs based on AVI
- Asymptotics involves order statistics of χ_1^2 random variables

Su, Di and Hsu (2014). Hypothesis testing for functional linear models.

Submitted.

FPCA-based approach

An example

Results with FA in RCST				
RCST	p-values		npc	
p(a)ve	AVI	V	AVI	V
0.50	0.0332	0.1007	2	2
0.85	0.0147	0.0637	3	5
0.99	0.0211	0.0035	5	10

- Standard FPCA approach sensitive to tuning parameter
- The new AVI-based approach is more robust

Dependent Functional Data

$$Y_{ij}(t) = X_{ij}(t) + \epsilon_{ij}(t)$$

- i : subject; j : visit
- $Y_{ij}(t)$ is recorded on $\Omega_{ij} = \{t_{ijs} : s = 1, 2, \dots, T_{ij}\}$
- Functions from the same subject are correlated

$$Y_{ij}(t) = \mu(t) + Z_i(t) + W_{ij}(t) + \epsilon_{ij}(t)$$

- $Z_i(t)$'s and $W_{ij}(t)$'s are centered random functions
- Assume $Z_i(t)$ and $W_{ij}(t)$ are uncorrelated

Multilevel FPCA

KL expansion on both levels

$$Z_i(t) = \sum_{k=1}^{N_1} \xi_{ik} \phi_k^{(1)}(t), \quad W_{ij}(t) = \sum_{l=1}^{N_2} \zeta_{ijl} \phi_l^{(2)}(t)$$

- $\phi_k^{(1)}(t), \phi_l^{(2)}(t)$: eigenfunctions
dominating directions of variation at both levels
- ξ_{ik}, ζ_{ijl} : principal component scores
magnitude of variation for each subject/visit
- $\lambda_k^{(1)} = \text{var}(\xi_{ik}), \lambda_l^{(2)} = \text{var}(\zeta_{ijl})$: eigenvalues
the amount of variation explained

Multilevel FPCA

$$Y_{ij}(t) = \mu(t) + Z_i(t) + W_{ij}(t) + \epsilon_{ij}(t)$$

- Between subject level (level 1):

$$K_B(s, t) := \text{cov}\{Z_i(s), Z_i(t)\} = \sum_{k=1}^{\infty} \lambda_k^{(1)} \phi_k^{(1)}(s) \phi_k^{(1)}(t)$$

- Within subject level (level 2):

$$K_W(s, t) := \text{cov}\{W_{ij}(s), W_{ij}(t)\} = \sum_{l=1}^{\infty} \lambda_l^{(2)} \phi_l^{(2)}(s) \phi_l^{(2)}(t)$$

- Total: $K_T(s, t) := K_B(s, t) + K_W(s, t) + \sigma^2 I(t = s)$

Note that

- $\text{cov}\{Y_{ij}(s), Y_{ik}(t)\} = K_B(s, t) + \sigma^2 I(t = s)$
- $\text{cov}\{Y_{ij}(s), Y_{ij}(t)\} = K_B(s, t) + K_W(s, t) + \sigma^2 I(t = s)$

MFPCA Algorithm

- Estimate $\mu(t)$ and $\eta_j(t)$ by univariate smoothing; estimate $K_T(s, t)$ and $K_B(s, t)$ via bivariate smoothing
- Set $\hat{K}_W(s, t) = \hat{K}_T(s, t) - \hat{K}_B(s, t)$
- Eigen-analysis of $\hat{K}_B(s, t)$ and $\hat{K}_W(s, t)$ to obtain $\hat{\lambda}_k^{(1)}$, $\hat{\phi}_k^{(1)}(t)$, $\hat{\lambda}_l^{(2)}$, $\hat{\phi}_l^{(2)}(t)$
- Estimate principal component scores

Note: we use penalized splines with REML for smoothing
R package “SemiPar”

Principal Component Scores

$$Y_{ij}(t) = \mu(t) + \sum_{k=1}^{N_1} \xi_{ik} \phi_k^{(1)}(t) + \sum_{l=1}^{N_2} \zeta_{ijl} \phi_l^{(2)}(t) + \epsilon_{ij}(t)$$

- Estimate scores, $\hat{\xi}_{ik}, \hat{\zeta}_{ijl}$, using *BLUP*
- Dimension reduction
Subject level: $\{Y_{i1}(t), \dots, Y_{iJ}(t)\} \rightarrow (\hat{\xi}_{i1}, \dots, \hat{\xi}_{iN_1})$
- Predict individual curve $\hat{Y}_{ij}(t)$ with confidence bands
- Predict subject level curve $\hat{Z}_i(t)$ with confidence bands

Other extensions

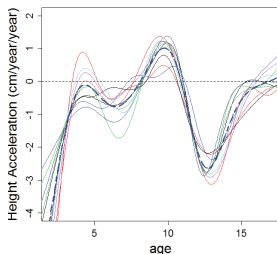
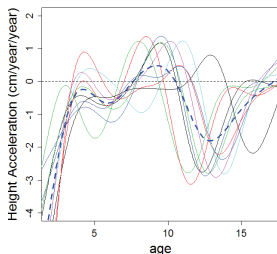
- Multilevel Functional Regression (Crainiceanu et al. 2009)
- Longitudinal/multivariate FPCA (more flexible correlations)

The Registration Problem

Most analyses only account for variation in *amplitude*.

Frequently, observed data exhibit features that vary in *time*.

Berkeley Growth Acceleration
Observed Aligned



- Mean of unregistered curves has smaller peaks than any individual curve.
- Aligning the curves reduces variation by 25%

Defining a Warping Function

Requires a transformation of *time*.

Seek

$$s_i = w_i(t)$$

so that

$$\tilde{x}_i(t) = x_i(s_i)$$

are well aligned.

$w_i(t)$ are *time-warping* (also called *registration*) functions.

Landmark registration

For each curve $x_i(t)$ we choose points

$$t_{i1}, \dots, t_{iK}$$

We need a reference (usually one of the curves)

$$t_{01}, \dots, t_{0K}$$

so these define constraints

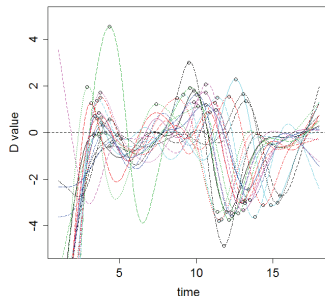
$$w_i(t_{ij}) = t_{0j}$$

Now we define a smooth function to go between these.

Identifying Landmarks

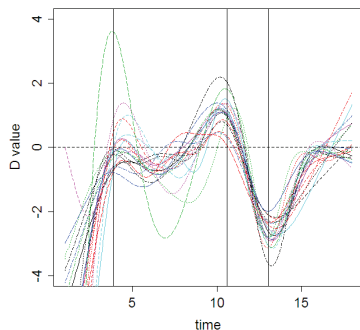
Major landmarks of interest:

- where $x_i(t)$ crosses some value
- location of peaks or valleys
- location of inflections

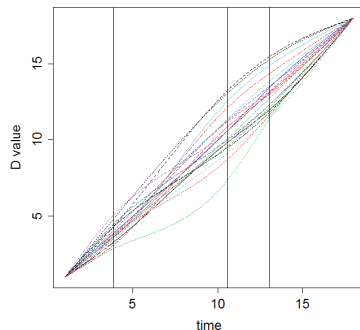


Results of Warping

Registered Acceleration



Warping Functions



Dynamics: Relationships between derivatives

Access to derivatives of functional data allows new models.

Variant on the concurrent linear model: e.g.

$$Dy_i(t) = \beta_0(t) + \beta_1(t)y_i(t) + \beta_2(t)x_i(t) + \epsilon_i(t)$$

Can be estimated like concurrent linear model.

But how do we understand these systems?

Principle Differential Analysis

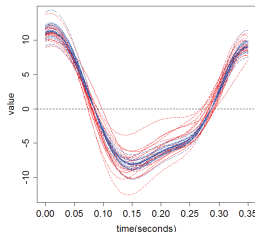
Translate autonomous dynamic model into linear differential operator:

$$Lx = D^2x + \beta_1(t)Dx(t) + \beta_0(t)x(t) = 0$$

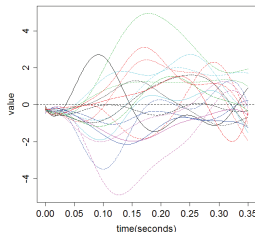
Potential use in improving smooths (theory under development).

We can ask what is smooth? How does the data deviate from smoothness?

Solutions of $Lx(t) = 0$



Observed $Lx(t)$

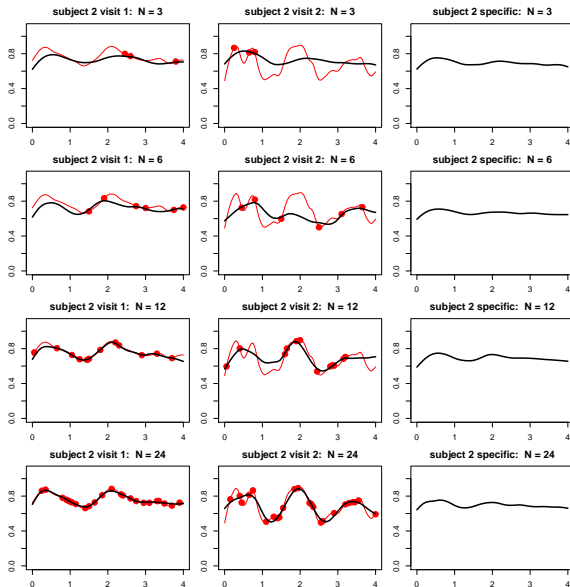


FDA for sparse longitudinal data

$$Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$$

- Data is recorded on sparse and irregular grid points
 $\Omega_i = \{t_{i1}, t_{i2}, \dots, t_{in_i}\}$, n_i is small (bounded)
- But grid points are dense collectively, $\Omega = \cup_i \Omega_i$
- Difficulty of applying FDA techniques (e.g., FPCA)
 - Cannot pre-smooth trajectory for each subject
 - Estimation of FPC needs numerical integration
$$d_{ik} = \int \{x_i(t) - \mu(t)\} \phi_k(t) dt$$
- Solution: Yao et al. (2005)
 - Pool all data, use (bivariate) smoothing
 - Estimate FPC by conditional expectations (BLUPs)

FDA for sparse longitudinal data



More general regression models

- Functional additive models (Muller et al., 2008; McLean et al., 2014)
- Partially functional linear regression (Kong et al., 2015)
- Functional mixture regression (Yao et al. 2011)
- ...

Recommended readings

- Yao, Muller and Wang(2005). Functional data analysis for sparse longitudinal data. JASA, 100: 577-590.
- Reiss and Ogden (2007). Functional Principal Component Regression and Functional Partial Least Squares. JASA, 102: 984–996.
- Ramsay, Hooker, Campbell, and Cao (2007). Parameter estimation for differential equations: a generalized smoothing approach. JRSS-B, 69: 741-796.
- Kneip and Ramsay (2008). Combining registration and fitting for functional models. JASA, 103(483), 1155-1165.
- Di, Crainiceanu, Caffo and Punjabi (2009). Multilevel Functional Principal Component Analysis. AOAS, 3: 458–488.
- Crainiceanu, Staicu and Di (2009). Generalized Multilevel Functional Regression. JASA, 104: 1550–1561.
- Senturk and Muller (2010). Functional varying coefficient models for longitudinal data. JASA, 105: 1256-1264.
- Goldsmith, Greven and Crainiceanu(2013). Corrected confidence bands for functional data using principal components. Biometrics, 69(1), 41-51.