Generalized Additive Mixed Models

Workshop: Analysis of Longitudinal Data
12th Nov 2024

Jaroslaw Harezlak
Armando Teixeira-Pinto

THE UNIVERSITY OF SYDNEY

Beyond linearity

Material covered so far:

Repeated measures	→ common time design points
Linear mixed models (LMM)	→ linear relationships over time
GLMMs	→ extensions of LMMs to binary/count outcomes

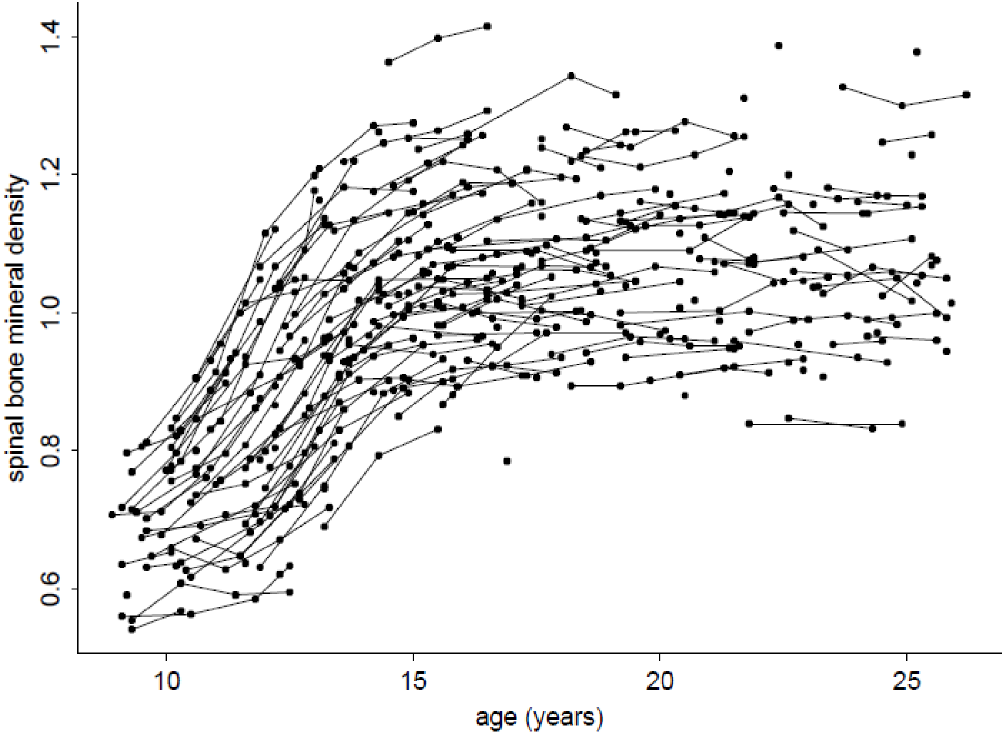
This lecture - GAMMs:

Extensions of (G)LMMs to nonlinear (smooth) relationships between the outcomes and trends over time

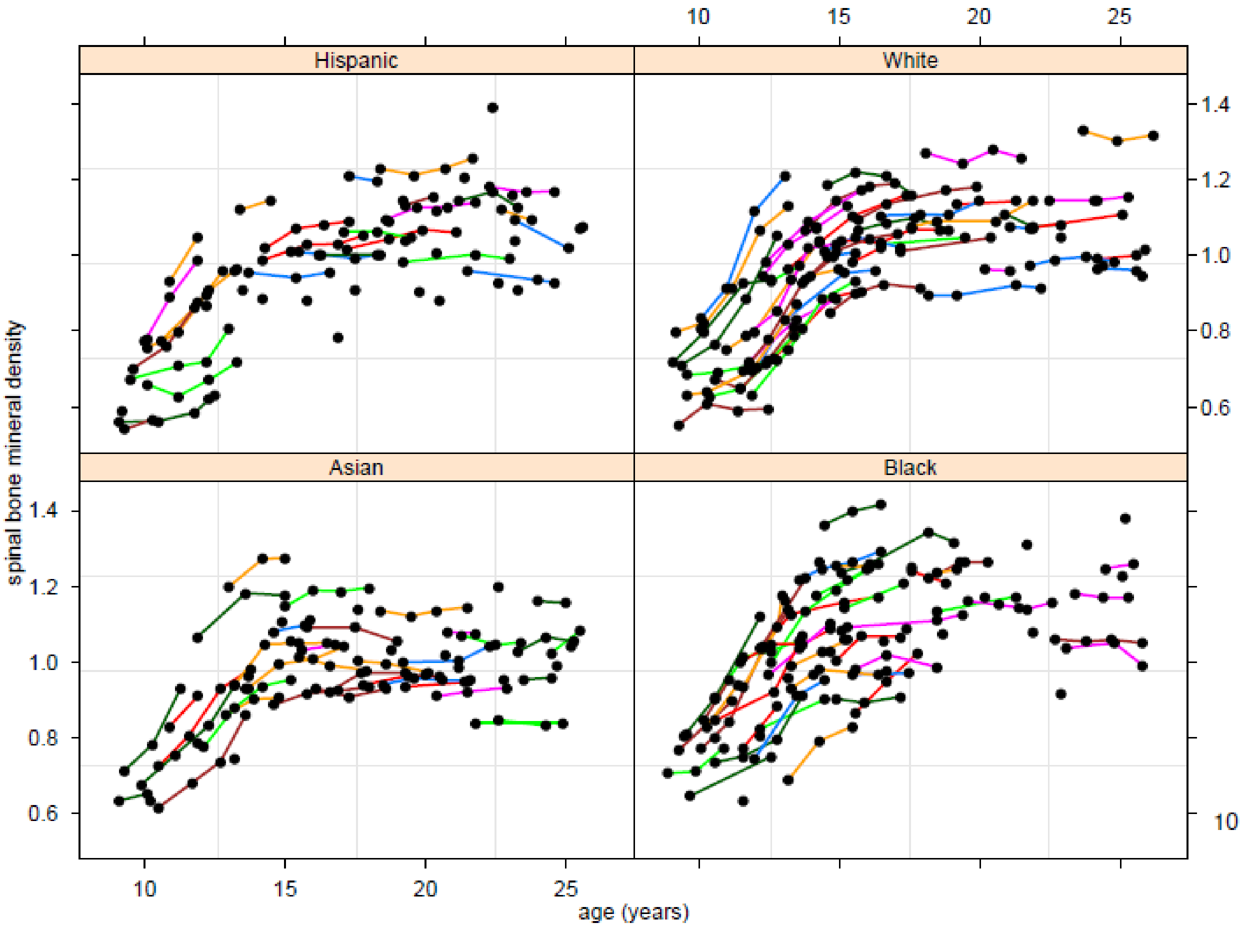
Spinal Bone Mineral Density Data

The following slide shows longitudinal data on spinal bone mineral density (SBMD) for 230 adolescent girls and young women.

Example – bone mineral density



Bachrach, L.K., Hastie, T., Wang, M.-C., Narasimhan, B. and Marcus, R. (1999). Bone mineral acquisition in healthy Asian, Hispanic, Black and Caucasian youth. A longitudinal study. J. Clin. Endocrin. Metab. 84, 4702–12.



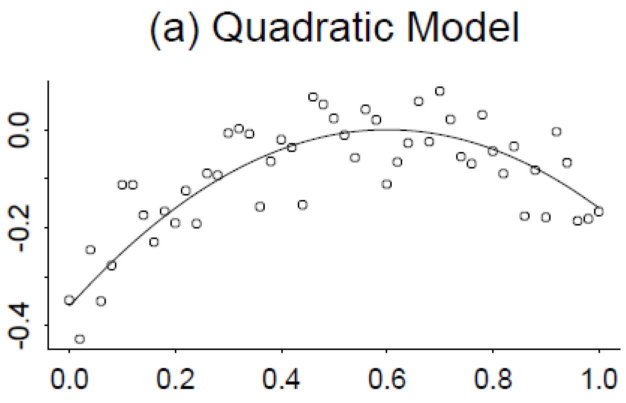
Hispanic
White
Asian
Black
spinal bone mineral density
age (years)
10
15
20
25
0.6
0.8
1.0
1.2
1.4

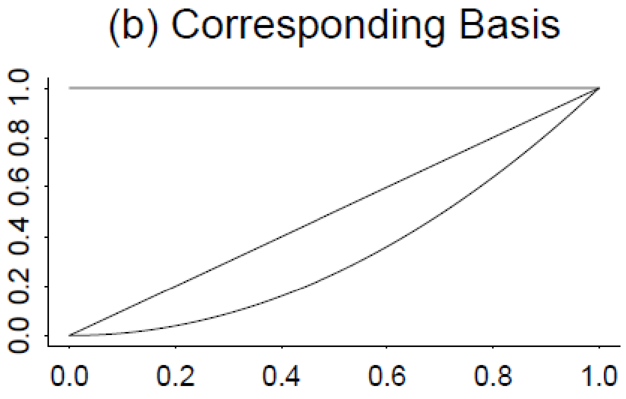
The Univers

Basis expansion idea

- Bases functions for linear models:
  - constant & line

- Bases functions for polynomial p^{th}-degree models:
  - constant, line, quadratic, cubic, …, p^{th} power


(a) Quadratic Model
0.0
-0.2
-0.4
0.0 0.2 0.4 0.6 0.8 1.0

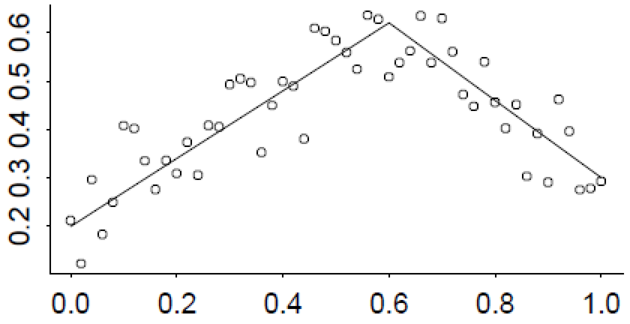

(b) Corresponding Basis
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8 1.0

Basis expansion idea

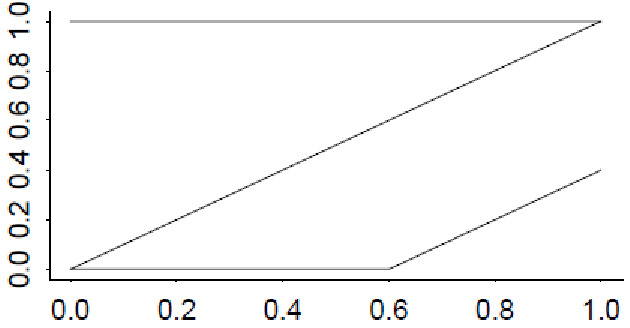
Other basis function families

- Splines (including B-splines, natural cubic splines)
- Fourier expansion (sines and cosines)
- Wavelets (e.g. Haar wavelets)
- other

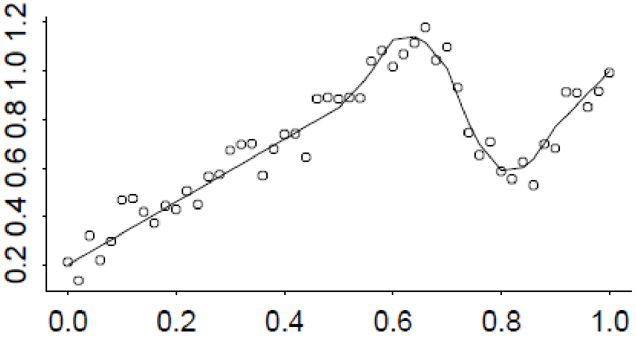
(a) Broken Stick Model



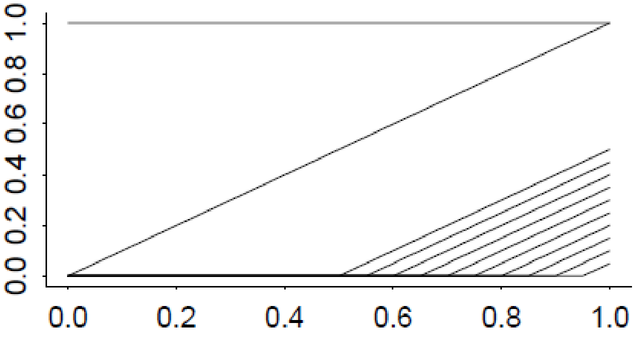
(b) Corresponding Basis



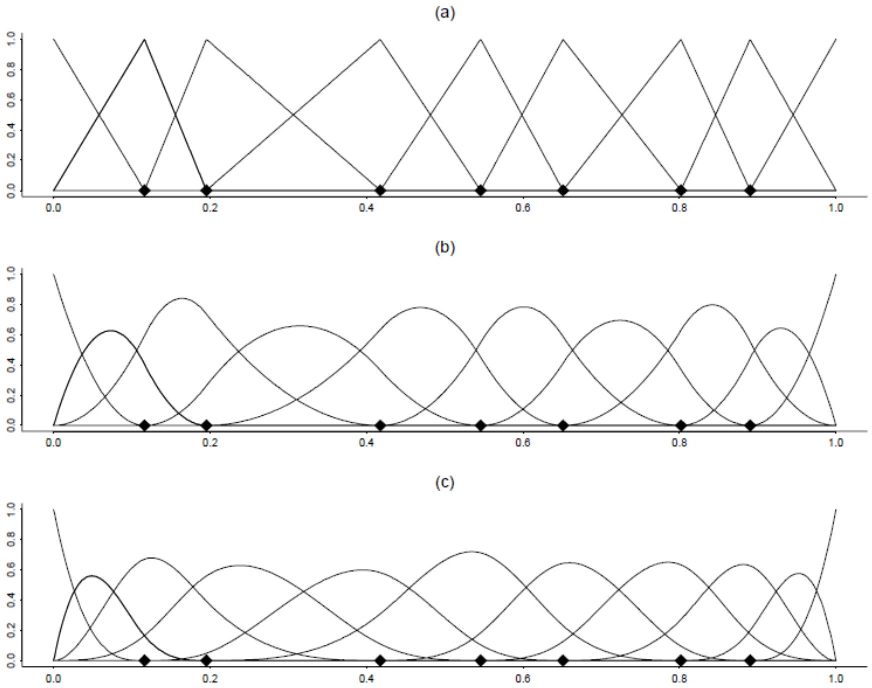
(a) Whip Model



(b) Corresponding Basis



B-spline basis



Smoothing splines

Minimize the criterion:

$$\min_{f \in H_2} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int (f''(x))^2 dx$$

Solution:

Natural cubic splines with knots at all observed values x_i

Regression splines

Minimize the criterion:

$$\min_{b_1,\ldots,b_K} \sum_{i=1}^{n} (y_i - f(x_i))^2,$$

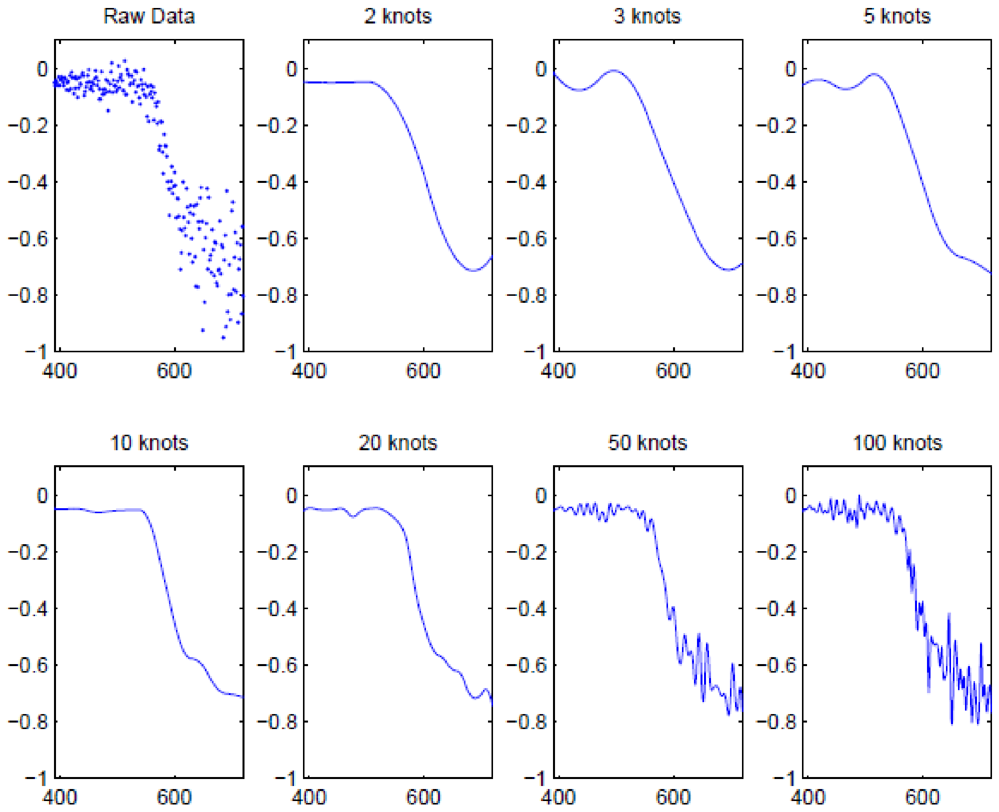
$$\text{where } f(x_i) = \sum_{k=1}^{K} b_k \theta_k(x_i),$$

and $\theta_k(\bullet)$ are basis functions.

Regression splines

- Parametric model with K parameters
- Choice of knots
- Importance of the placement and the number of knots
- Computationally hard problem

Ordinary Least Squares



Penalized regression splines

- Compromise between smoothing splines and regression splines
- Fewer knots than smoothing splines (K << n)
- However, the coefficients are penalized

$$\min_{b_1,\dots,b_K} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \sum_{k=1}^{K} u_k^2$$

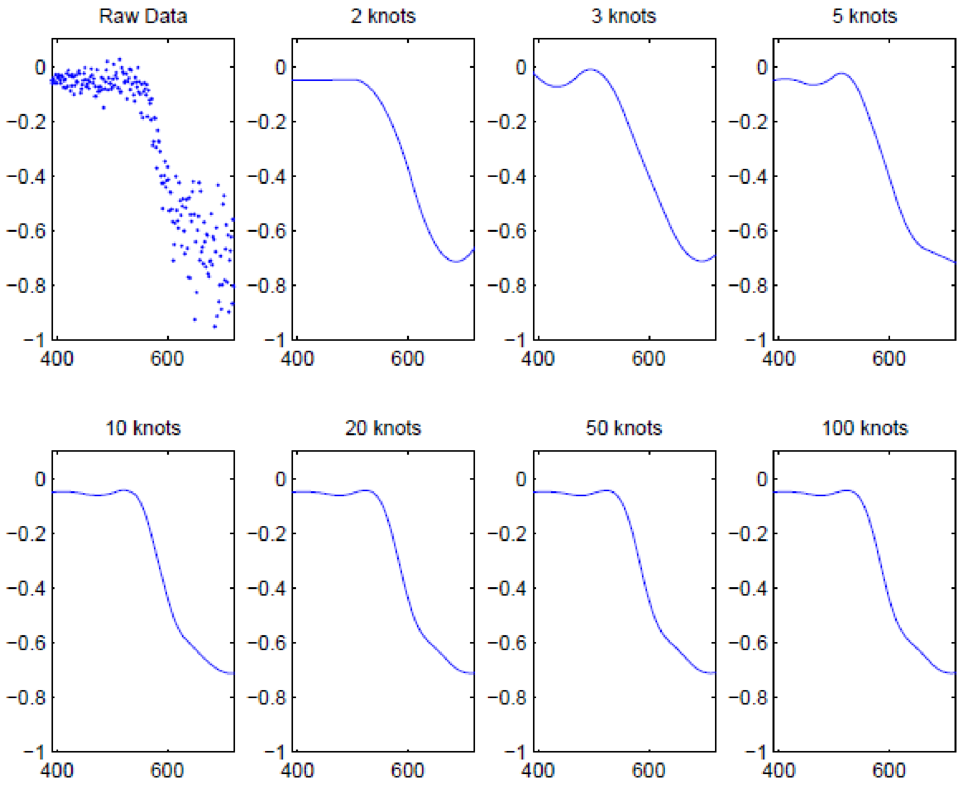
where $f(x_i) = \sum_{k=1}^{K} u_k \theta_k(x_i)$,

and $\theta_k(\bullet)$ are basis functions.

Penalized regression splines

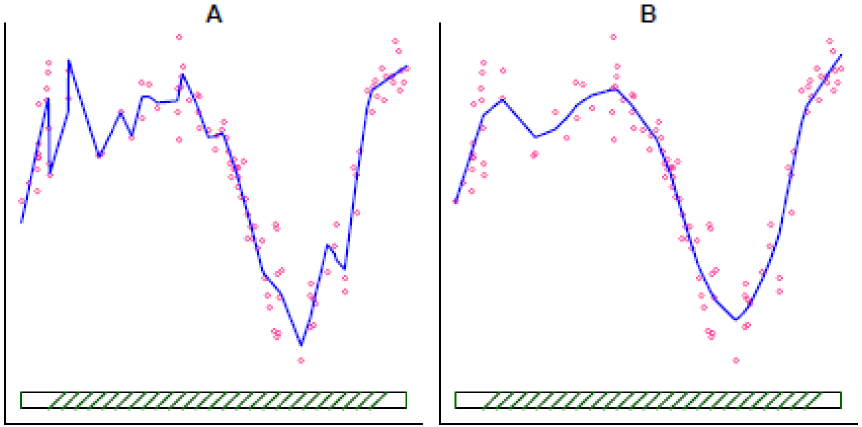
- K has to be large enough to allow estimation of curve f(•) features
- Methodology is easy to extend to longitudinal data, spatial data, additive models, etc.
- Computations can be done using linear mixed model functions (e.g. function lme() in the package "nlme" or functions in the package "mgcv")

Penalized Least Squares



Tricking Mixed Models to do Smoothing

$$y_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} u_k (x_i - \kappa_k)_+ + \varepsilon_i$$



A: u_k's fixed B: u_k i.i.d. $N(0, \sigma_u^2)$

Penalised Splines as Mixed Models

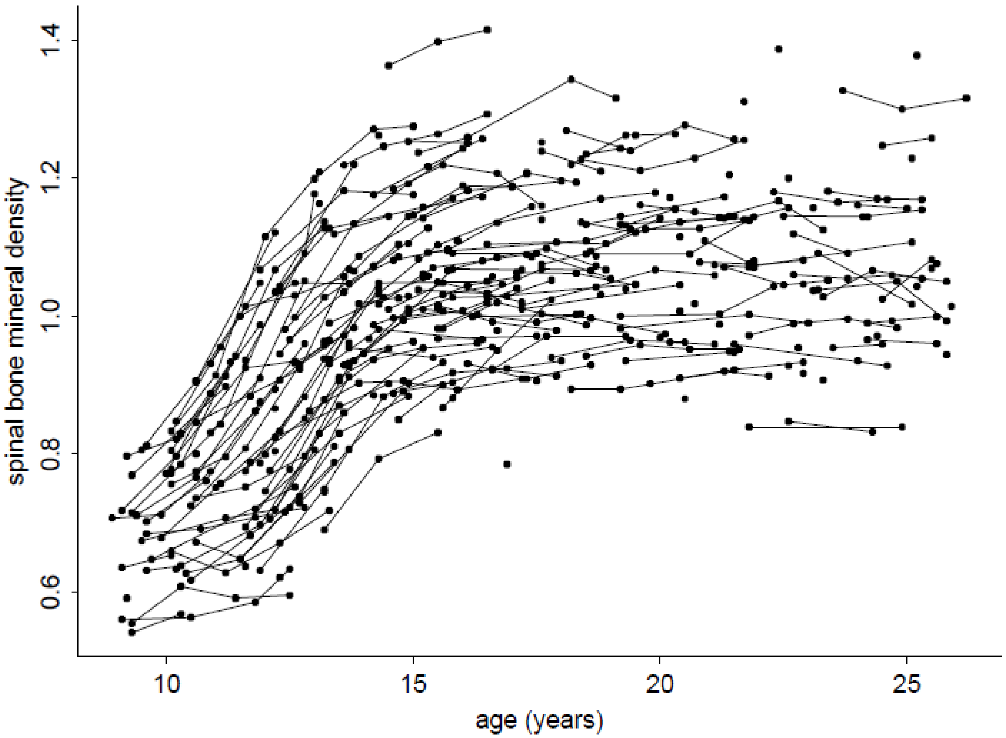
$$y_i = \underbrace{\beta_0 + \beta_1 x_i}_{\text{fixed effects}} + \underbrace{\sum_{k=1}^{K} u_k (x_i - \kappa_k)_+}_{\text{random effects}} + \varepsilon_i$$

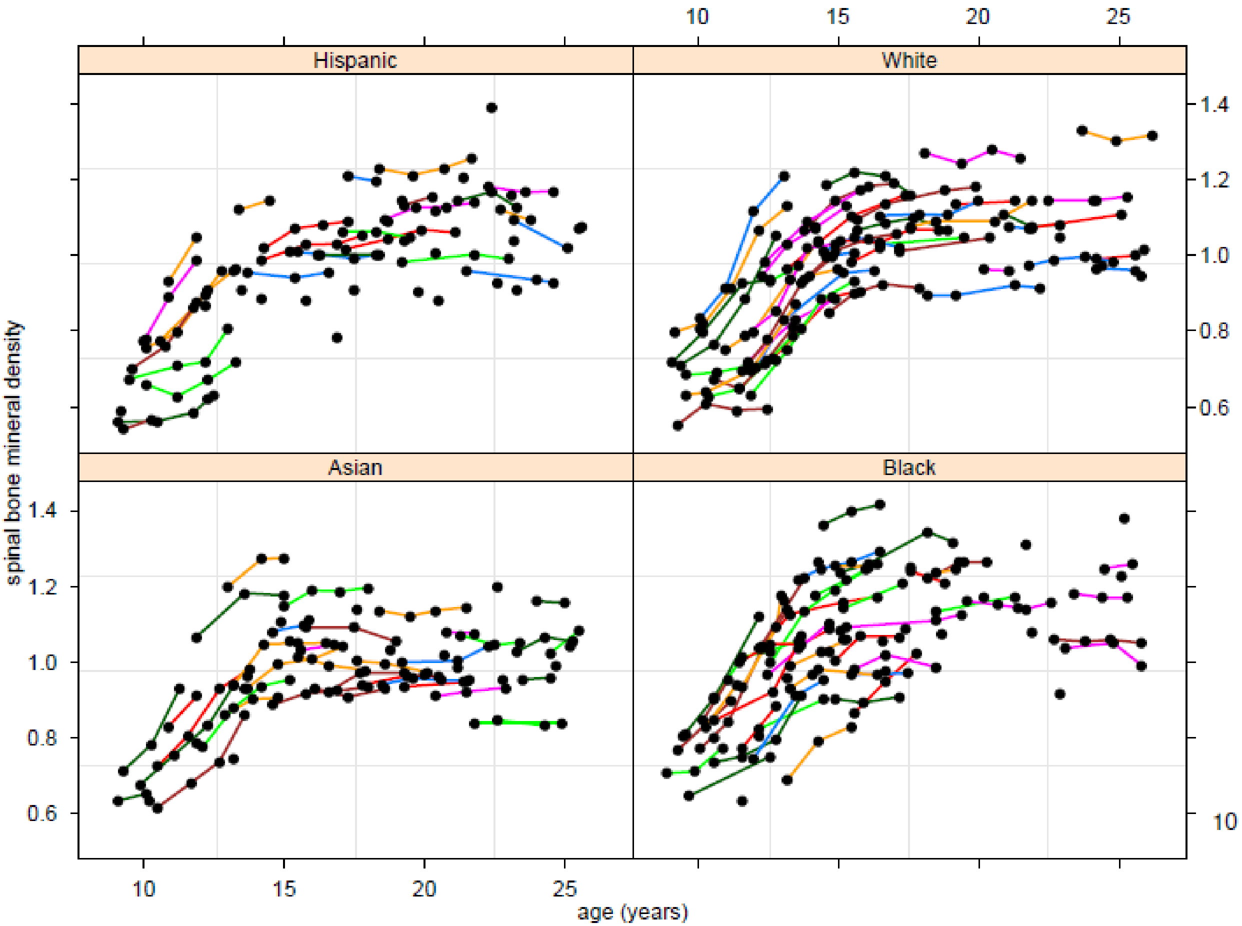
Longitudinal data

- We will go back to the example shown in the second slide

- We will talk about extensions of penalized regression splines to longitudinal data

Example – bone mineral density




Hispanic
White
Asian
Black
spinal bone mineral density
age (years)
10
15
20
25
0.6
0.8
1.0
1.2
1.4

The Univers

Double Subscript Notation

Let

m = number of subjects

n_i = number of measurements on subject i $(1 \leq i \leq m)$

Then

y_{ij} = response for j^{th} measurement on subject i.

Outcome (SBMD) and predictor (age)

$SBMD_{ij}$ = j^{th} spinal bone mineral density for subject i

age_{ij} = age corresponding to $SBMD_{ij}$ measurement

Reasonable model for SBMD

$$\text{SBMD}_{ij} = \boldsymbol{U}_i + \boldsymbol{f}(\text{age}_{ij}) + \varepsilon_{ij}$$

$\boldsymbol{U}_i \overset{\text{ind.}}{\sim} \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}_u^2)$ is a random subject intercept

Treating Longitudinal and Smoothing Together

$$\mathbf{Z} = \left[\begin{array}{c|c} \text{random} & \text{spline} \\ \text{intercept} & \text{basis} \\ \text{indicators} & \text{functions} \end{array} \right]$$

$$\text{Cov}(\mathbf{u}) = \begin{bmatrix} \sigma_U^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 \mathbf{I} \end{bmatrix}$$

$$\begin{aligned} \sigma_U^2 &= \text{controls between-subject variability} \\ \sigma_u^2 &= \text{controls amount of smoothing in estimation of mean} \end{aligned}$$

Additive Model Extension

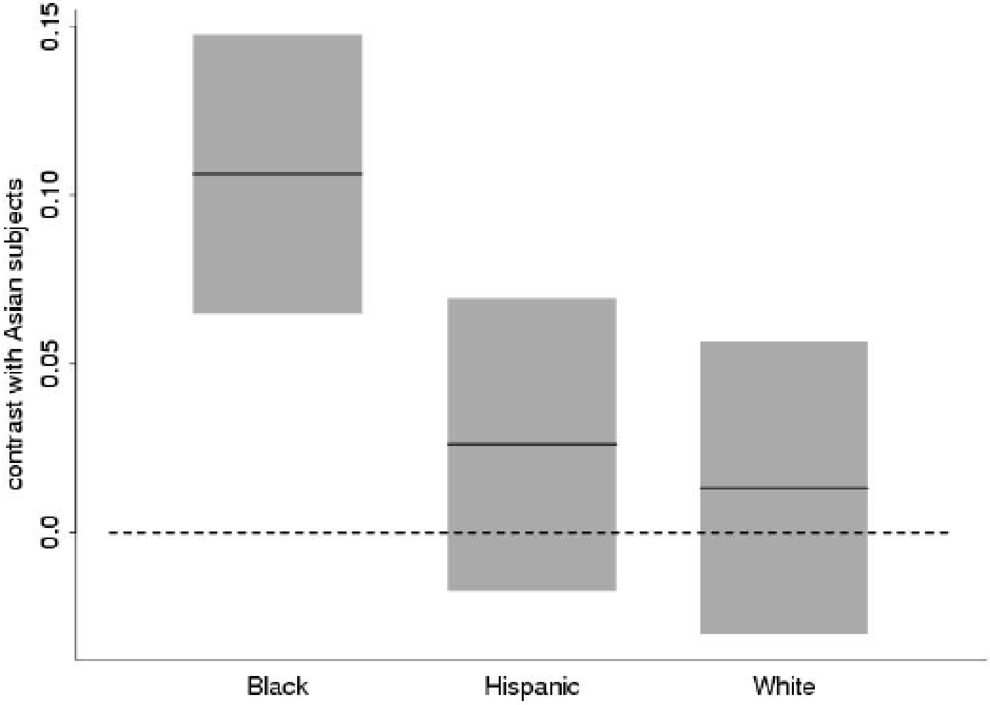
With the Asian ethnicity as a baseline we have

$$\begin{aligned}\text{SBMD}_{ij} &= U_i + f(\text{age}_{ij}) + \beta_1\,\texttt{black}_i + \beta_2\,\texttt{hispanic}_i\\ &\quad +\beta_3\,\texttt{white}_i + \varepsilon_{ij},\end{aligned}$$

Effect of Ethnicity

From the mixed model output:

variable	Value	Approx. Std. Error	z ratio
black	0.1062	0.02066	5.141
hispanic	0.0260	0.02164	1.203
white	0.0131	0.02165	0.6069


0.15
0.10
0.05
0.0
contrast with Asian subjects
Black
Hispanic
White

Fitting penalized spline models in R

1. Create the design matrix X with the fixed effects
2. Matrix Z is created by evaluating the basis functions (e.g. truncated lines) at each time point
3. Use lme() function (in R) with the random structure specification corresponding to the penalty put on the random coefficients
4. Get the BLUEs for the fixed effects and BLUPs for the random coefficients to get an estimate of the smooth curve

OR (much easier)

Use function gamm() in the package “mgcv”

Main Points

– Many longitudinal applications benefit from semiparametric regression.

– Mixed models are the standard vehicle for longitudinal data analysis.

– Penalized splines with mixed model representation allow for a seamless fusion between semiparametric regression and longitudinal data analysis.