Flexible smoothing with P-splines: some applications

Maria Durbán
Department of Statistics, Universidad Carlos III de Madrid, Spain

Joint work with Raymon Carroll, Iain Currie, Paul Eilers, Jareck Harezla and Matt Wand

Department of Economics,Bielefeld University, June 2003

What is this talk about?

What is this talk about?

- Introduction
  - ⋆ Smoothing
  - ⋆ Why P-splines?
  - ⋆ Mixed model representation of P-splines

What is this talk about?

- Introduction
  - ⋆ Smoothing
  - ⋆ Why P-splines?
  - ⋆ Mixed model representation of P-splines

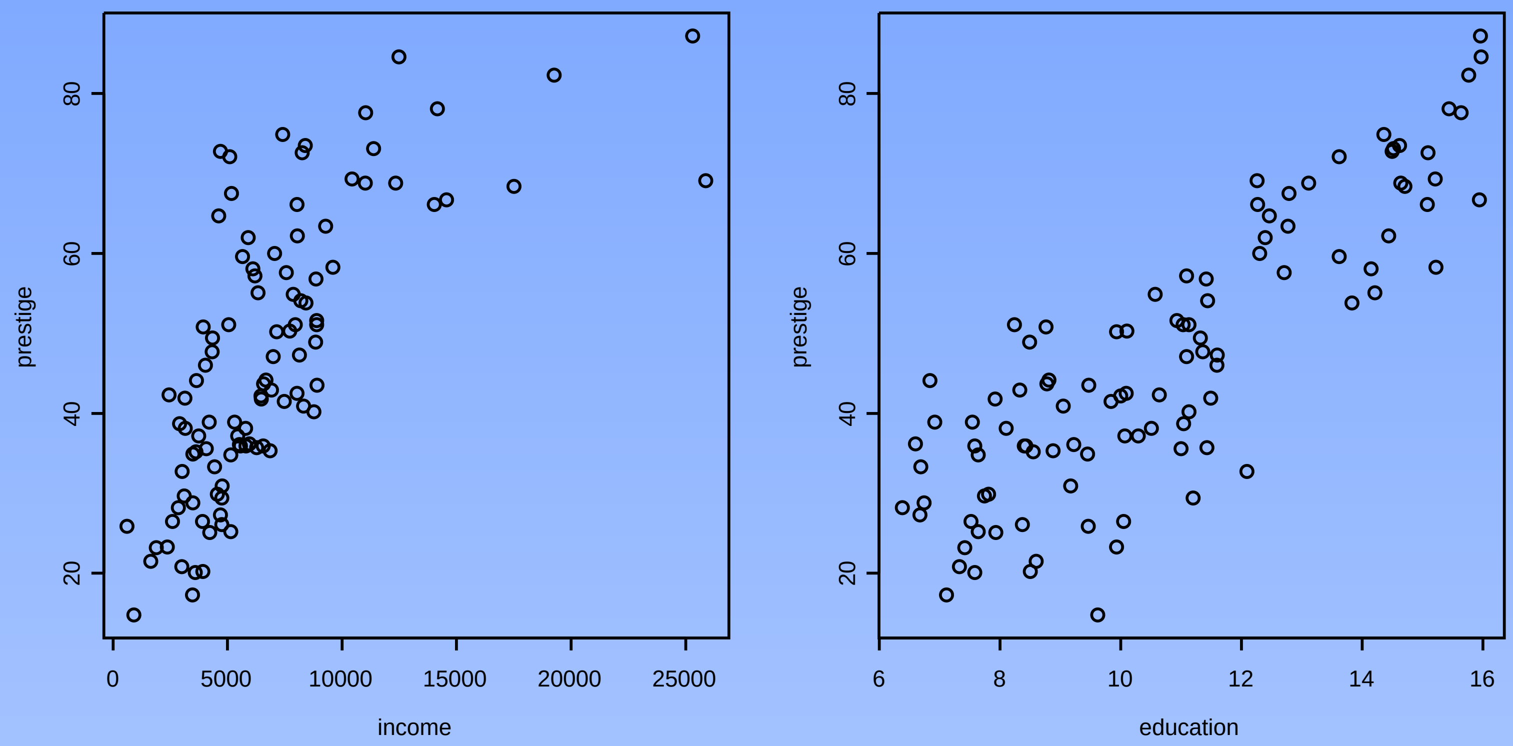
- Applications
  - ⋆ Additive models
  - ⋆ Models with heteroscedastic errors
  - ⋆ Smoothing and correlation
  - ⋆ Generalised additive models

What is this talk about?

- Introduction
  - ⋆ Smoothing
  - ⋆ Why P-splines?
  - ⋆ Mixed model representation of P-splines
- Applications
  - ⋆ Additive models
  - ⋆ Models with heteroscedastic errors
  - ⋆ Smoothing and correlation
  - ⋆ Generalised additive models
- P-splines for longitudinal data

Canadian Occupational Prestige Data (B. Blishen, 1971)

Data consist of prestige scores, average income (in $\$1000$) and education (in years) for 102 occupations.



Smoothing

- Prestige score varies smoothly along the income range

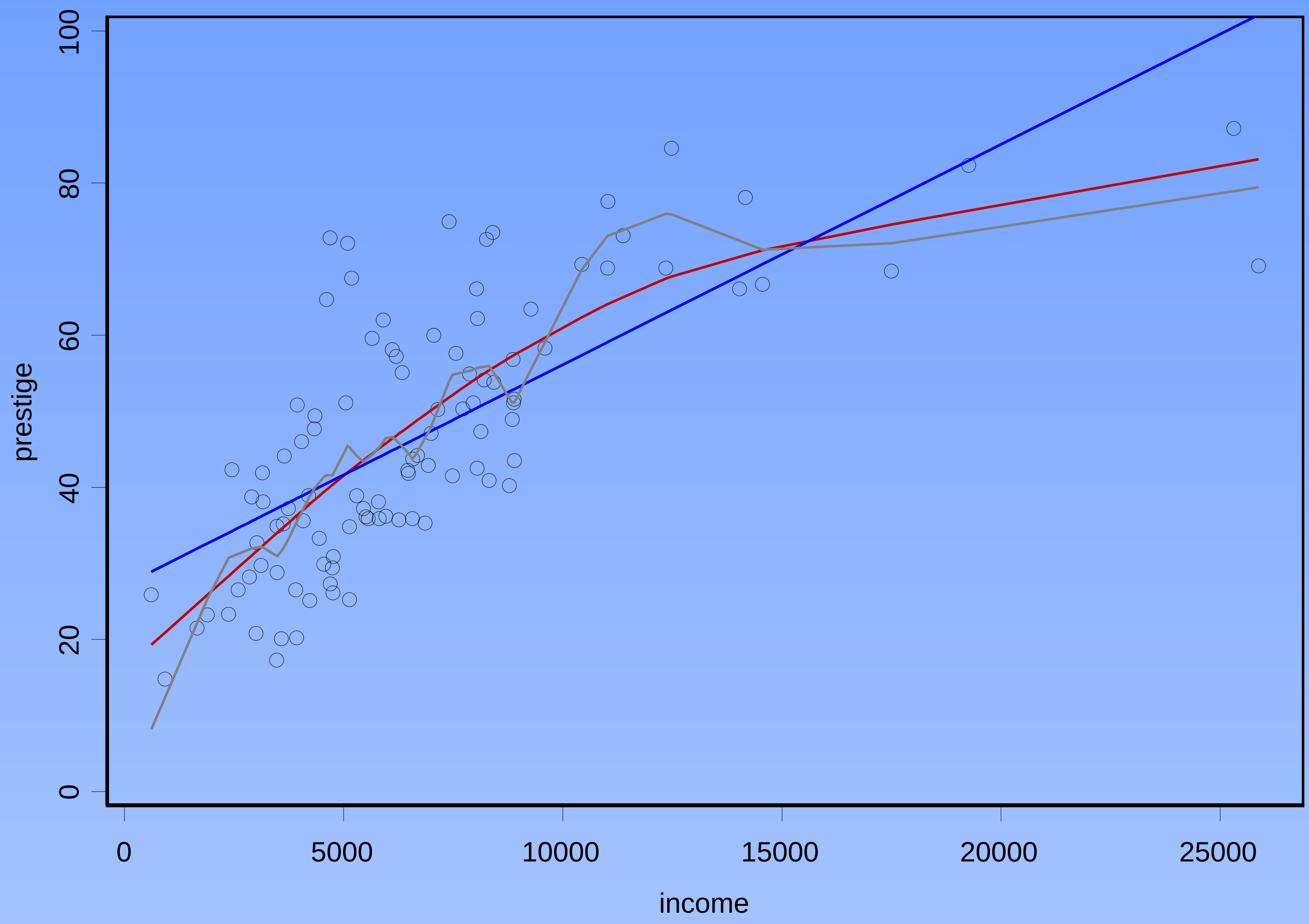
- A suitable model for these data could be:

$$\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{\epsilon}$$

where x is the covariate (income) f is a smooth function of x which depends on $\lambda =$smoothing parameter

- Smoothing methods fall into two groups:

  ⋆ Specified by the fitting procedure: **Kernels**
  ⋆ Solution of a minimisation problem: **Splines**


prestige
0
20
40
60
80
100
0
5000
10000
15000
20000
25000
income

P-spline

- Eilers and Marx, 1996.

- They are a generalisation of ordinary regression.
- Modify the log-likelihood by a penalty on the regression coefficients.

$$\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{\epsilon} \quad f(\boldsymbol{x}) \approx \boldsymbol{Ba} \quad S = (\boldsymbol{y} - \boldsymbol{Ba})'(\boldsymbol{y} - \boldsymbol{Ba}) + \lambda \boldsymbol{a}'\boldsymbol{Pa}$$

$$\hat{\boldsymbol{a}} = (\boldsymbol{B}'\boldsymbol{B} + \lambda \boldsymbol{P})^{-1}\boldsymbol{B}'\boldsymbol{y}$$

P-spline

- Eilers and Marx, 1996.

- They are a generalisation of ordinary regression.
- Modify the log-likelihood by a penalty on the regression coefficients.

$$\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{\epsilon} \quad f(\boldsymbol{x}) \approx \boldsymbol{B}\boldsymbol{a} \quad S = (\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a})'(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a}) + \lambda \boldsymbol{a}'\boldsymbol{P}\boldsymbol{a}$$

$$\hat{\boldsymbol{a}} = (\boldsymbol{B}'\boldsymbol{B} + \lambda \boldsymbol{P})^{-1}\boldsymbol{B}'\boldsymbol{y}$$

P-splines receive also other names:

- Penalised splines
- pseudosplines
- low-rank smoothers

Basis for P-splines

B-splines, truncated polynomial basis, radial basis, etc.

Basis for P-splines

B-splines, truncated polynomial basis, radial basis, etc.

B-splines

- B-spline: bell-shaped like Gauss curve
- Polynomial pieces smoothly joining at the knots

Basis for P-splines

B-splines, truncated polynomial basis, radial basis, etc.

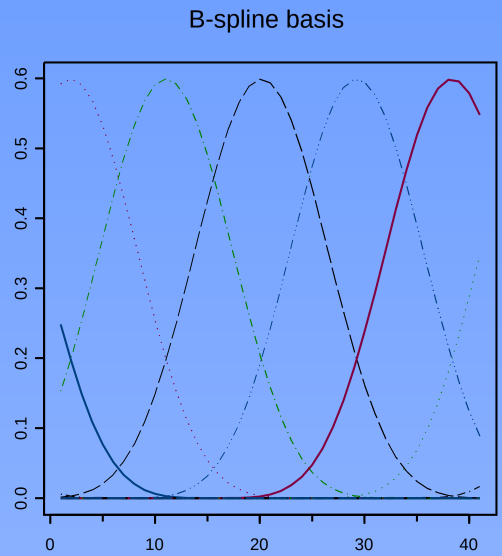
B-splines

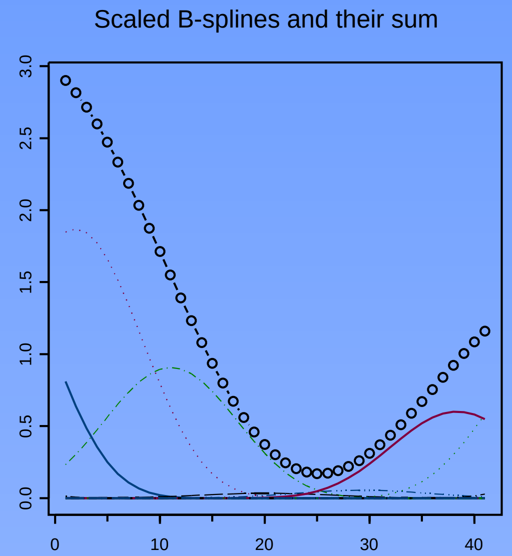
- B-spline: bell-shaped like Gauss curve
- Polynomial pieces smoothly joining at the knots

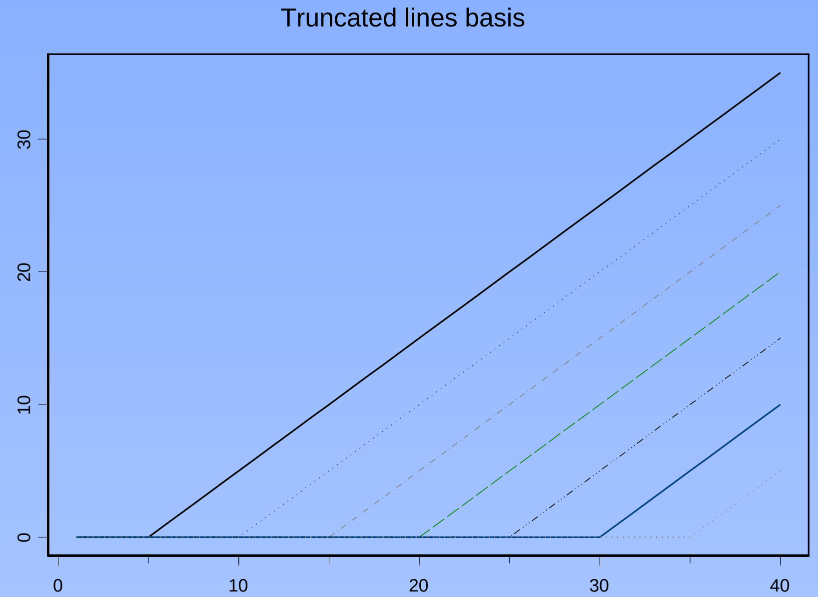
Truncated polynomial

For example: truncated linear basis for knots $\kappa_1, \ldots, \kappa_k$ is:

$$1, \boldsymbol{x}, (\boldsymbol{x} - \kappa_1)_+, \ldots, (\boldsymbol{x} - \kappa_k)_+$$


B-spline basis


Scaled B-splines and their sum


Truncated lines basis

Why P-splines?

- The number of basis functions used to construct the function estimates does not grow with the sample size
- Quite insensitive to the choice of knots (Ruppert, 2000)
- Computationally simpler
- No need for backfitting in the case of additive models
- Easily extended to 2 or more dimesions and non Gaussian errors

Psplines: mixed model approach

Psplines: mixed model approach

$$\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{R})$$

We write $f(\boldsymbol{x}) = \boldsymbol{Ba}$. It can be shown that $\boldsymbol{Ba}$ may be written as

$$\underbrace{\boldsymbol{X\beta}}_{fixed} + \underbrace{\boldsymbol{Zu}}_{random} \quad u \sim N(0, \sigma_u^2 I) \quad \lambda = \sigma^2/\sigma_u^2$$

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Zu} + \boldsymbol{\epsilon} \quad Cov \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \sigma_u^2 \boldsymbol{I} & 0 \\ 0 & \sigma^2 \boldsymbol{R} \end{bmatrix} \quad Cov[\boldsymbol{y}] = \boldsymbol{V} = \boldsymbol{R}\sigma^2 + \boldsymbol{Z}'\boldsymbol{Z}\sigma_u^2$$

Use **REML** for variance parameters

$$l(\boldsymbol{V}) = -\frac{1}{2}\log|\boldsymbol{V}| - \frac{1}{2}\log|\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}| - \boldsymbol{y}'(\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1})\boldsymbol{y},$$

Given $\boldsymbol{R}$, σ^2 and σ_u^2, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{u}}$ are solutions to:

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \lambda\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1} \end{bmatrix} \boldsymbol{y}.$$

Advantages

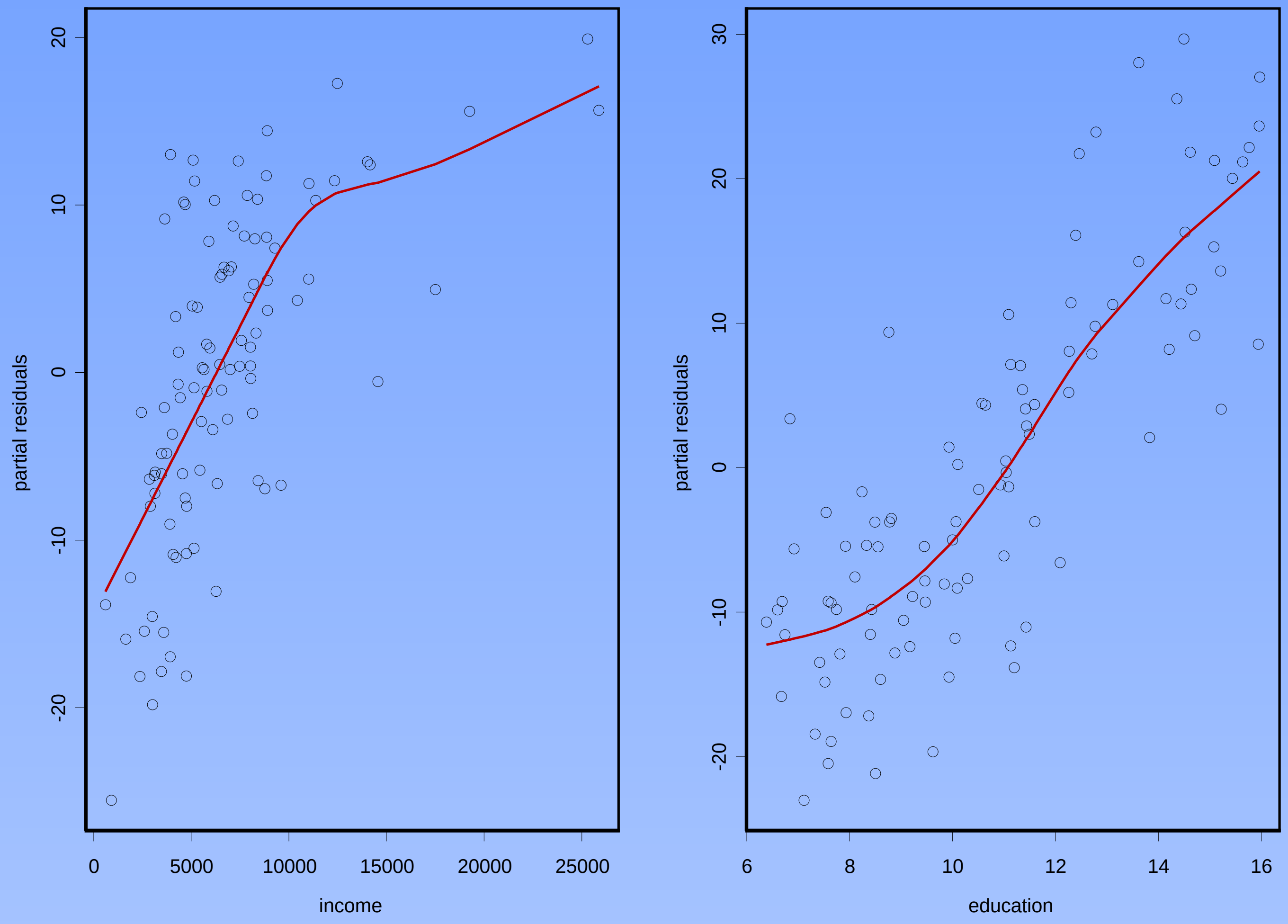
- Unified approach
- Automatic selection of smoothing parameter
- Likelihood ratio test for model selection
- Already implemented in standard sofware: Splus, SAS, R.

APPLICATIONS

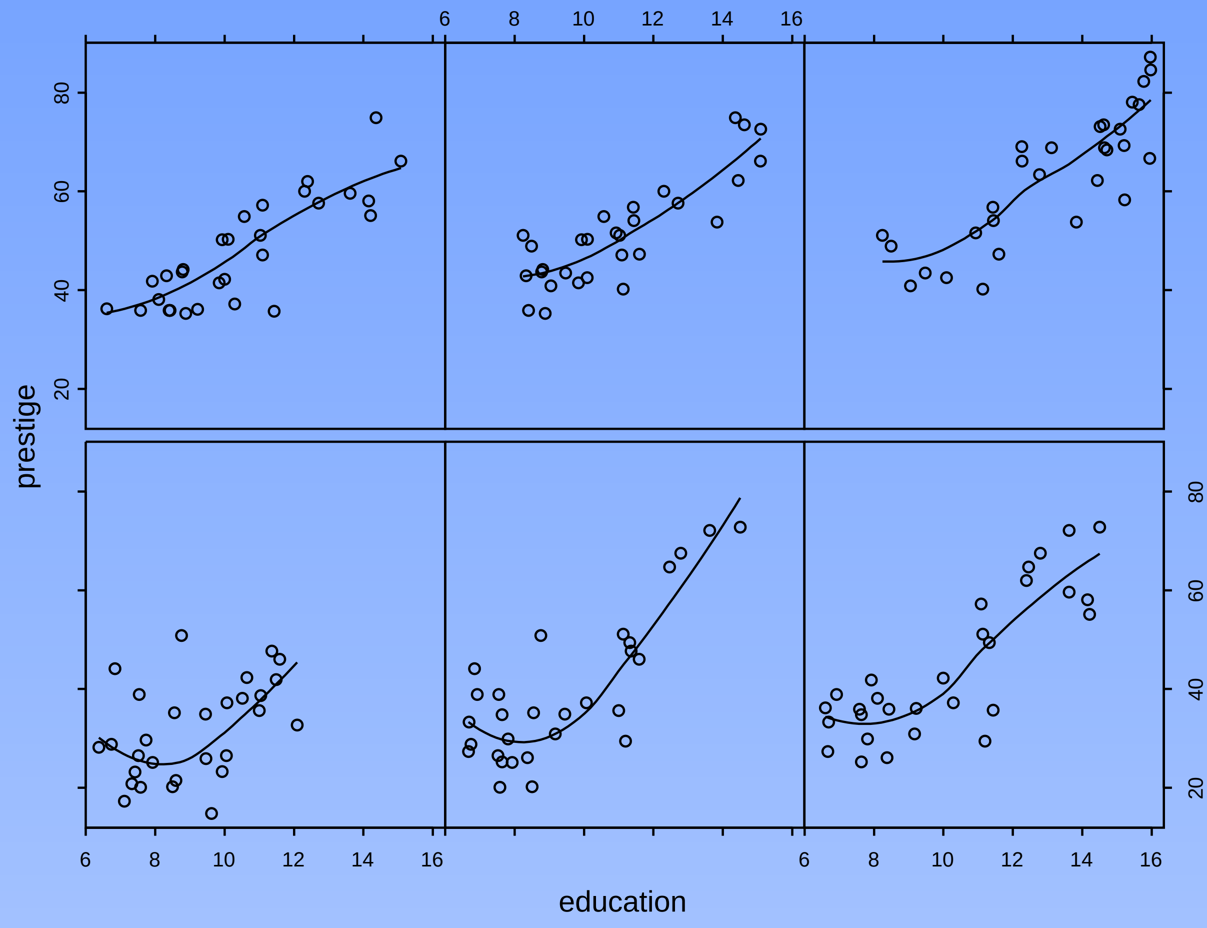
Additive models: Prestige data revisited

$$
\begin{aligned}
\boldsymbol{y} &= \underbrace{f(income)}_{\boldsymbol{X}_1\boldsymbol{\beta}_1+\boldsymbol{Z}_1\boldsymbol{u}_1} + \underbrace{f(education)}_{\boldsymbol{X}_2\boldsymbol{\beta}_2+\boldsymbol{Z}_2\boldsymbol{u}_2} + \boldsymbol{\epsilon} \\
&= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon} \quad Cov\left[\begin{array}{c} \boldsymbol{u} \\ \boldsymbol{\epsilon} \end{array}\right] = \left[\begin{array}{ccc} \sigma^2_{u_1}\boldsymbol{I} & 0 & 0 \\ 0 & \sigma^2_{u_2}\boldsymbol{I} & 0 \\ 0 & 0 & \sigma^2\boldsymbol{I} \end{array}\right] \\
& \boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')' \quad \boldsymbol{u} = (\boldsymbol{u}_1', \boldsymbol{u}_2')' \quad \boldsymbol{X} = [\boldsymbol{X}_1 : \boldsymbol{X}_2] \quad \boldsymbol{Z} = [\boldsymbol{Z}_1 : \boldsymbol{Z}_2]
\end{aligned}
$$

Partial residuals plot



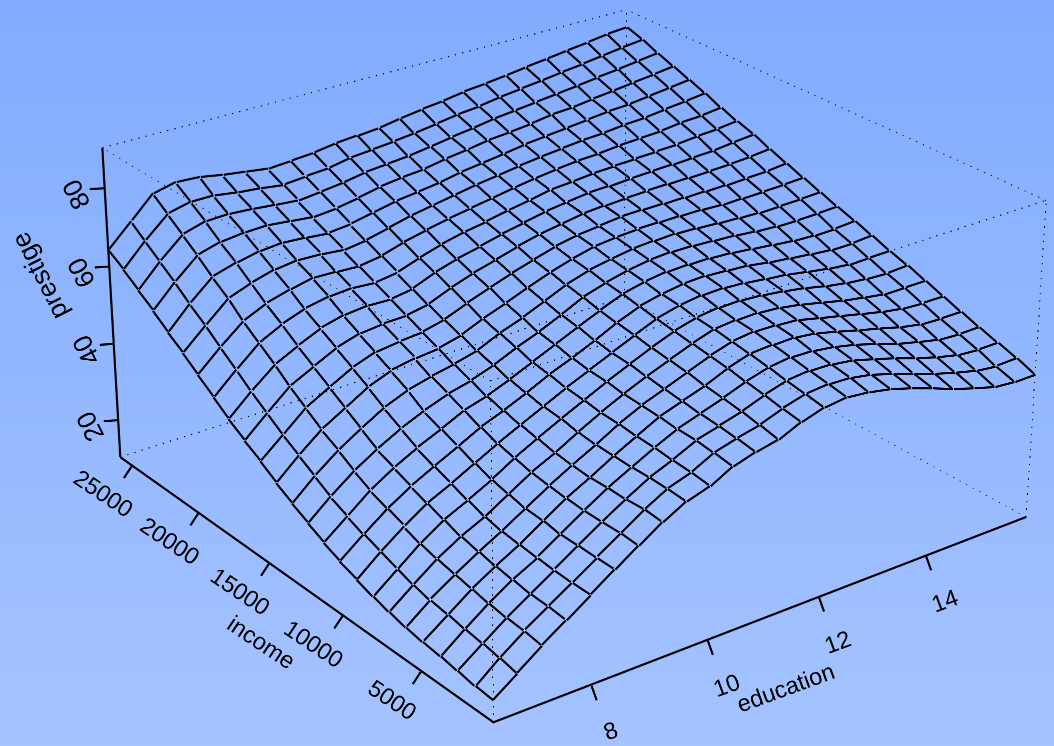
Is the model additive?: Conditional plots



Two-dimensional P-splines

Now $\boldsymbol{y} = f(income, education) + \boldsymbol{\epsilon} = \boldsymbol{Ba} + \boldsymbol{\epsilon}$, where

$$\boldsymbol{B} = \boldsymbol{B}_1 \otimes \boldsymbol{B}_2 \quad \boldsymbol{P} = \lambda_1 \boldsymbol{P}_1 \otimes \boldsymbol{I}_{n_2} + \lambda_2 \boldsymbol{I}_{n_2} \otimes \boldsymbol{P}_1$$



Smoothing and correlation (Currie and Durbán, 2002)

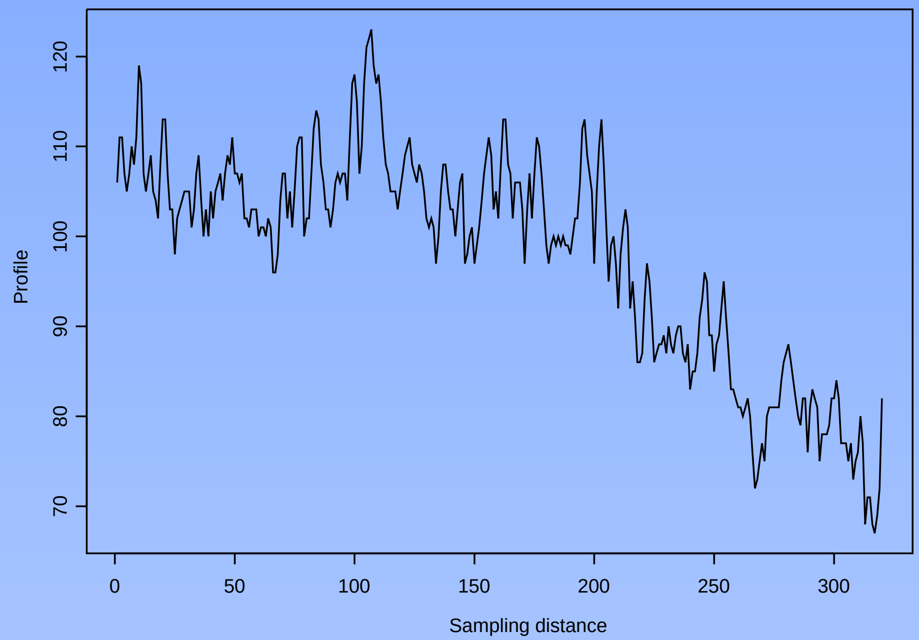
AIC and GCV lead to underestimation of the smoothing parameter in the presence of positive serial correlation. The general approach to modelling with P-splines takes care of this problem.

Smoothing and correlation (Currie and Durbán, 2002)

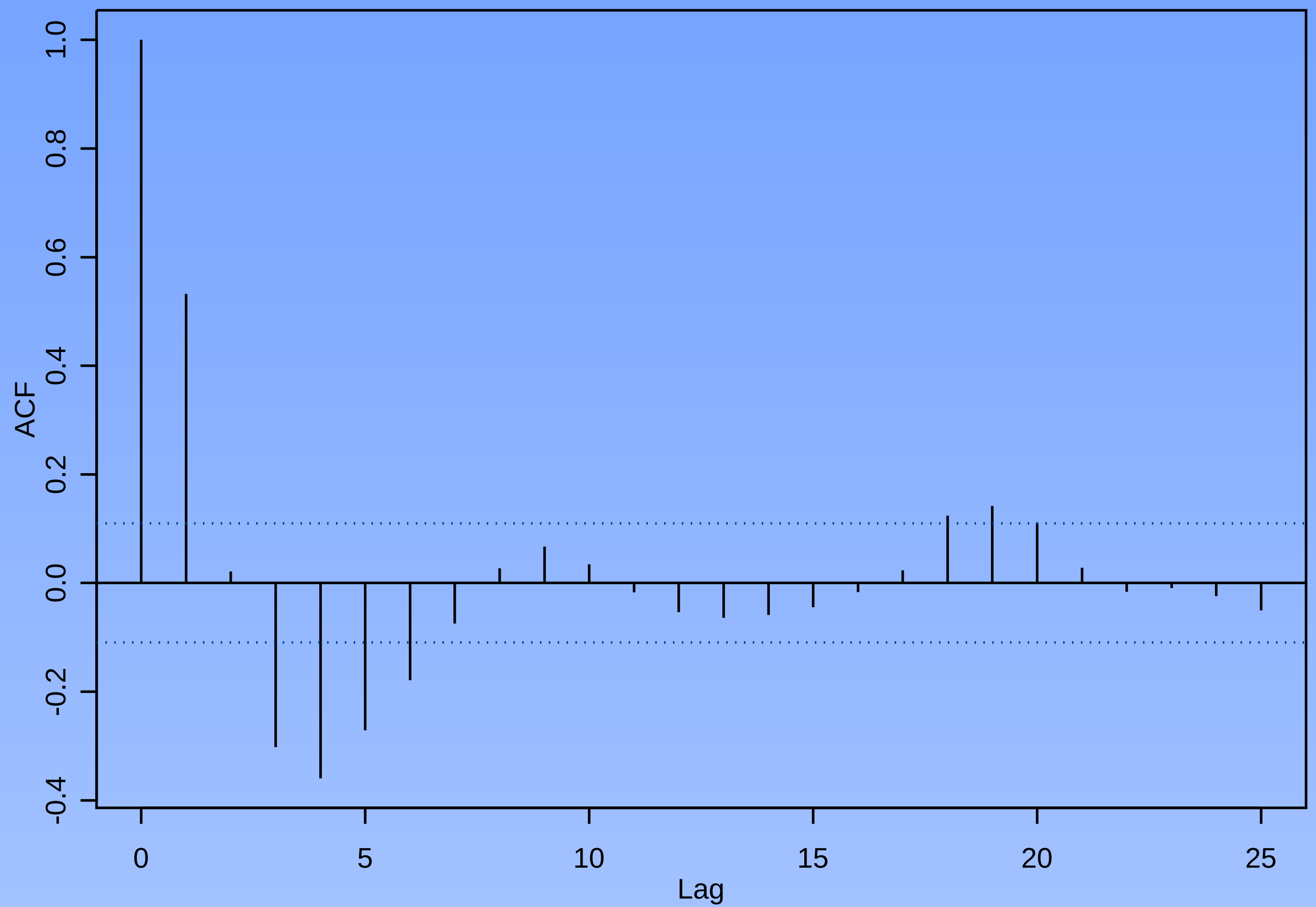
AIC and GCV lead to underestimation of the smoothing parameter in the presence of positive serial correlation. The general approach to modelling with P-splines takes care of this problem.

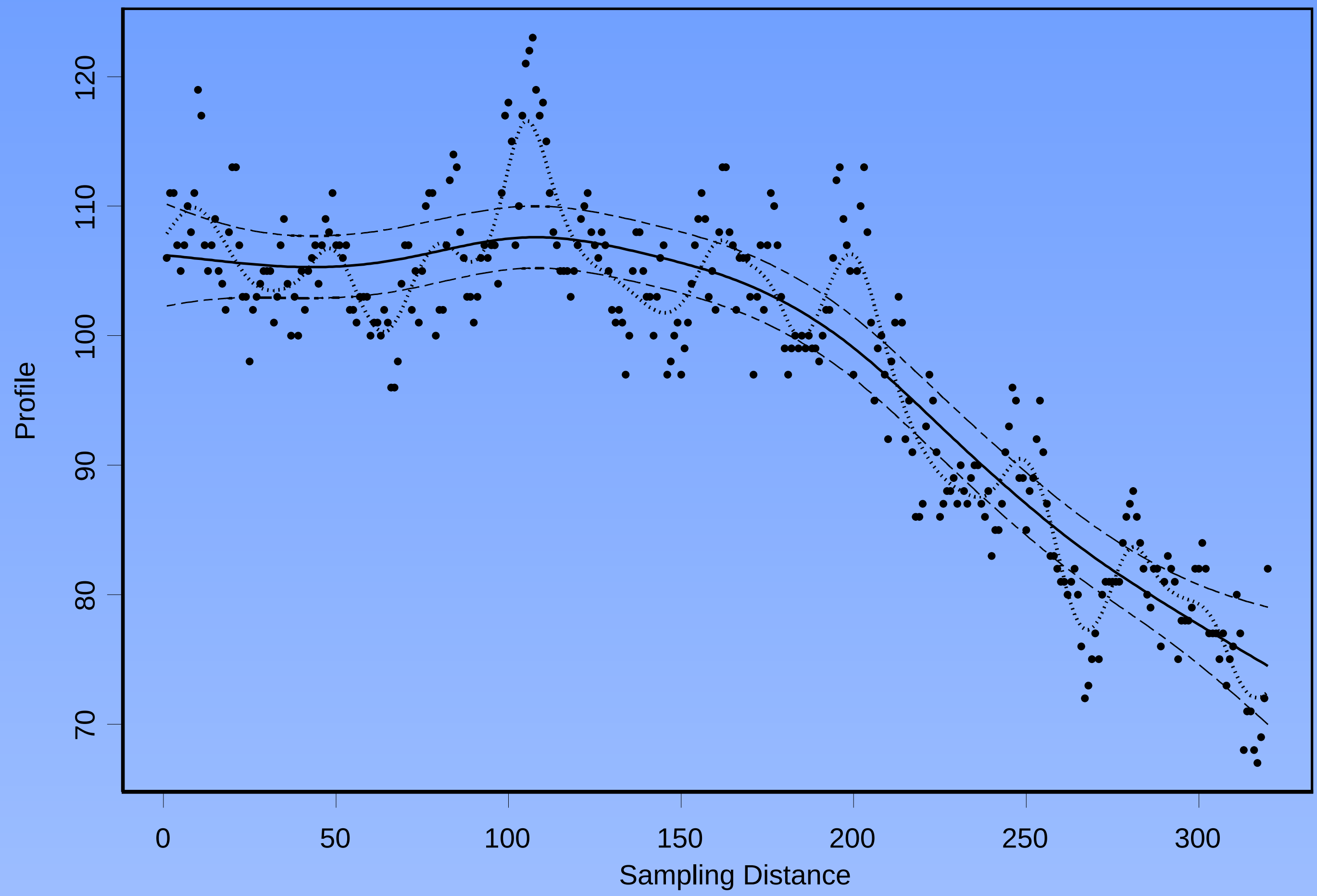
Wood profile data

320 measurements of the profile of a block of wood subject to grinding.

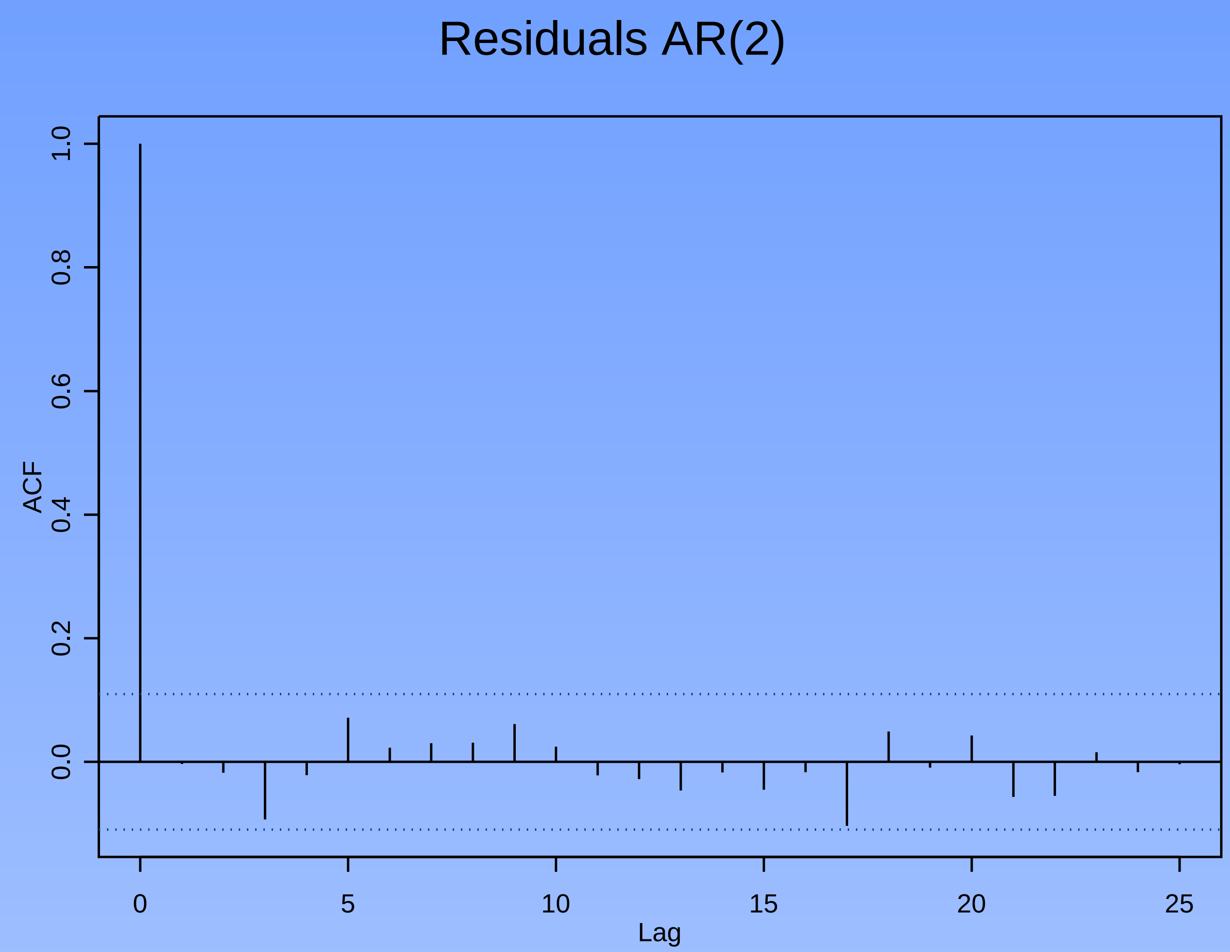


Residuals AR(1)





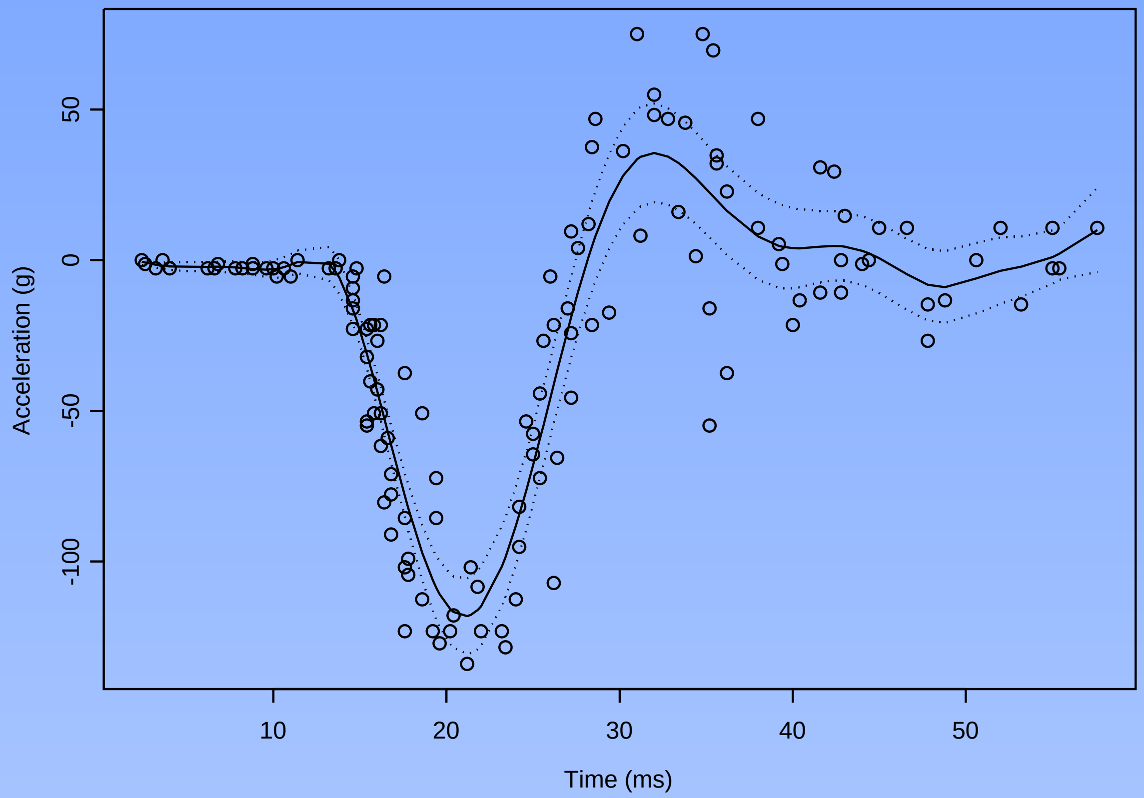
Profile
Sampling Distance
70
80
90
100
110
120
0
50
100
150
200
250
300



Other examples in Durbán and Currie (2003), Computational Statistics.

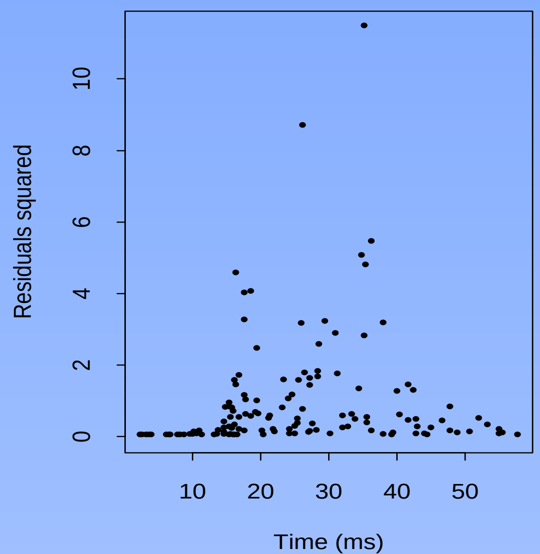
Smoothing and heteroscedasticity (Currie and Durbán (2002)

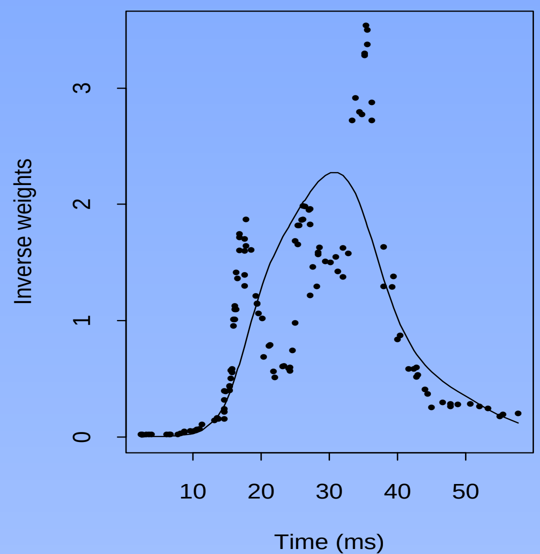
Simulated experiment to test crash helmets, 133 head accelerations and times after impact



Fit $\boldsymbol{y} = \boldsymbol{Ba} + \boldsymbol{\epsilon}$ with $Var(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{V}$ and $\boldsymbol{V} = \boldsymbol{W}^{-1}$, $\boldsymbol{W} = diag(w_1, \ldots, w_n)$.

Use P-splines to smooth $R_i = \log r_i^2$ $r_i^2 = (y_i - \hat{y}_i)^2 / \hat{\sigma}^2$ and $w_i^{-1} \propto exp(\hat{R}_i)$.





Generalised additive models: Count data

The one-parameter exponential family model, with canonical link, has joint density,

$$f(\boldsymbol{y}|\boldsymbol{\eta}) = exp\,\{\boldsymbol{y}'\boldsymbol{\eta} - \mathbf{1}'b(\boldsymbol{\eta}) + \mathbf{1}'c(\boldsymbol{y})\}$$

the linear predictor $\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{a}$, using the mixed model representation of P-splines we rewrite $\boldsymbol{B}\boldsymbol{a} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}$

$$f(\boldsymbol{y}|\boldsymbol{u}) = exp\,\{\boldsymbol{y}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}) - \mathbf{1}'exp(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}) - \mathbf{1}'log(\Gamma(\boldsymbol{y} + \mathbf{1}))\}$$

and $u \sim N(\mathbf{0}, \sigma^{\mathbf{2}}_{\mathbf{u}}\boldsymbol{I})$.

Iterate between penalised quasi-likelihood (PQL) of Breslow (1993) (to estimate $\boldsymbol{\beta}$ and $\boldsymbol{u}$) and REML (to estimate variance components).

In the case of count data $\lambda = 1/\sigma^2_u$.

The data

Male policyholders, source: *Continuous Mortality Investigation Bureau* (CMIB).

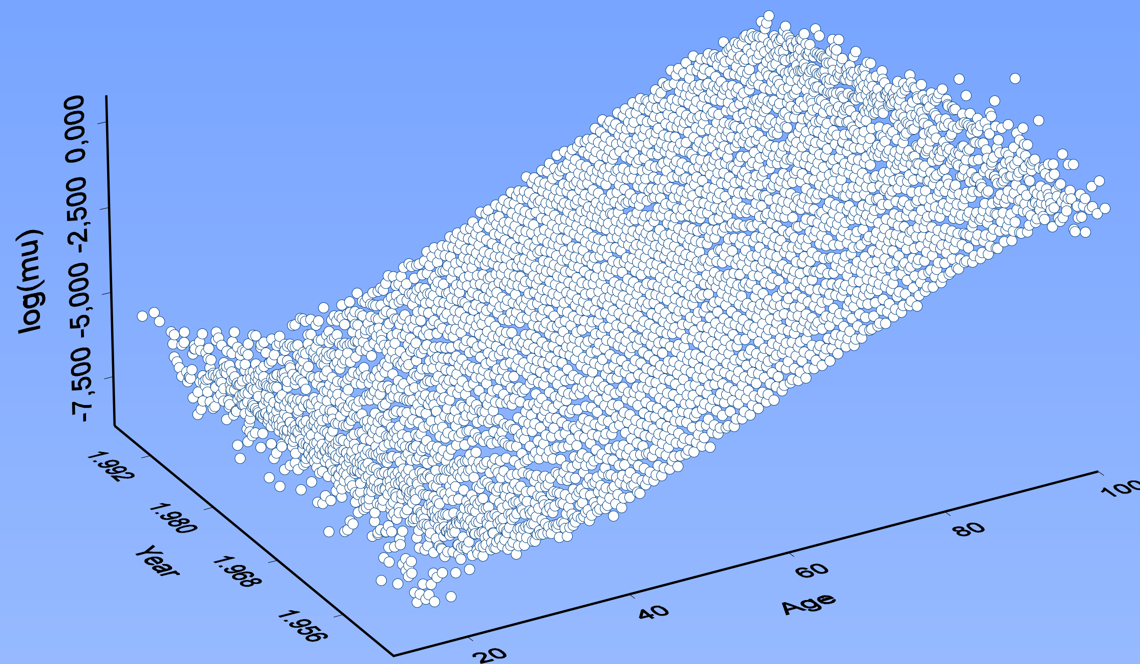
For each calendar year (1947-1999) and each age (11-100) we have:

- Number of years lived (the exposure).
- Number of policy claims (deaths).

Mortality of male policyholders has improved rapidly over the last 30 years

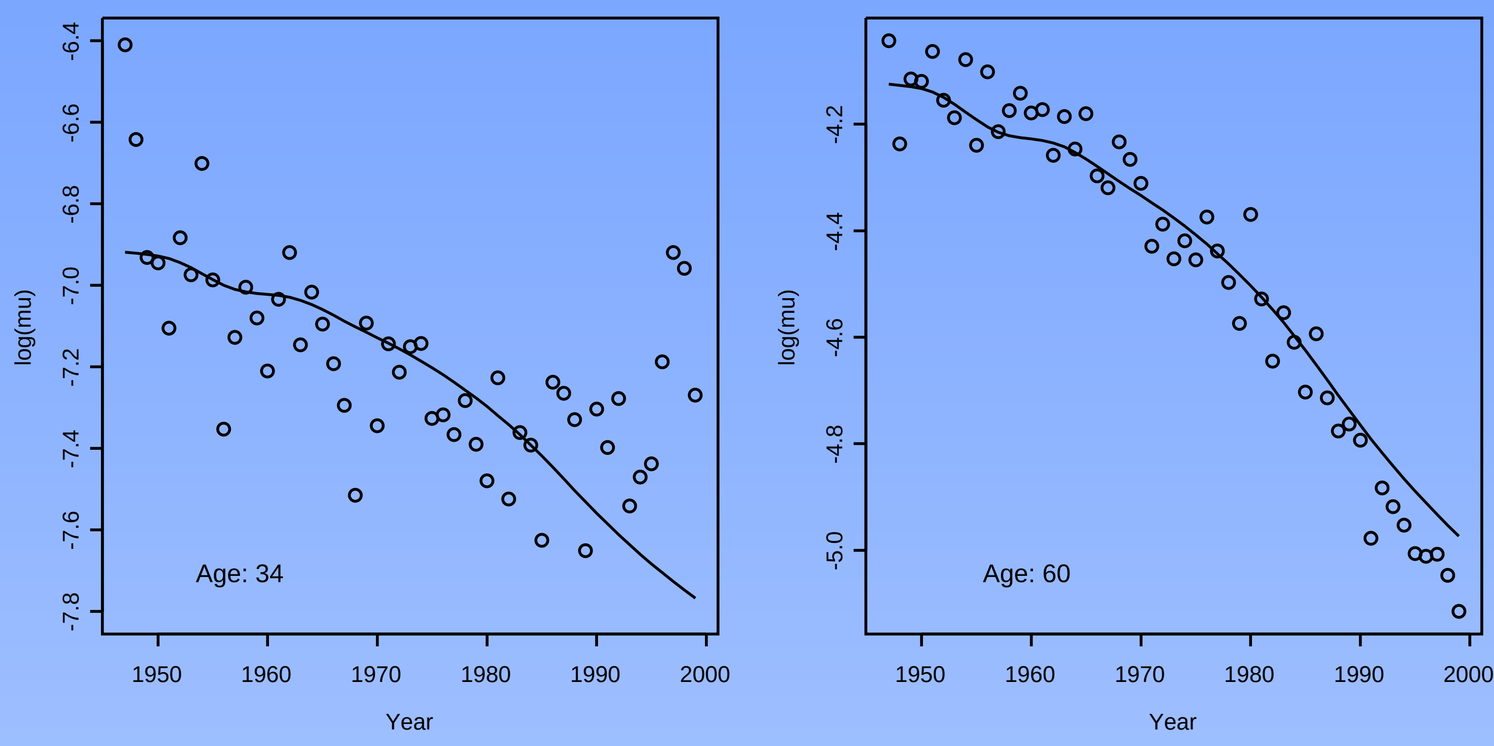
$$\Downarrow$$

Model mortality trends overtime and dependence on age.

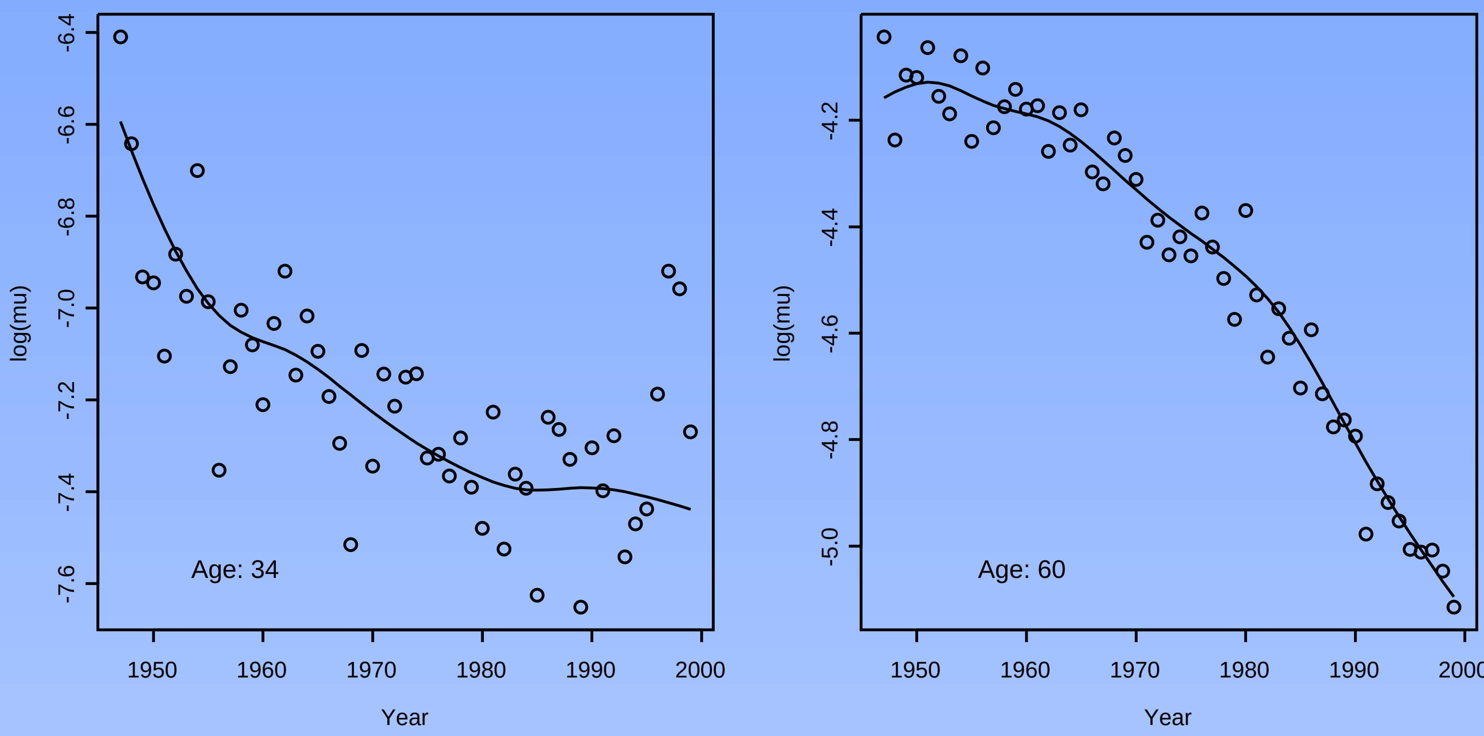


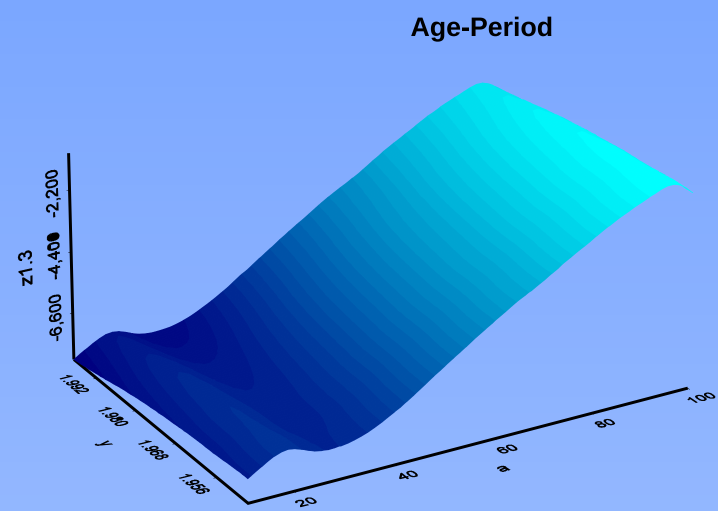
log(mu)
-7,500 -5,000 -2,500 0,000
Year
1.992
1.980
1.968
1.956
Age
20
40
60
80
100

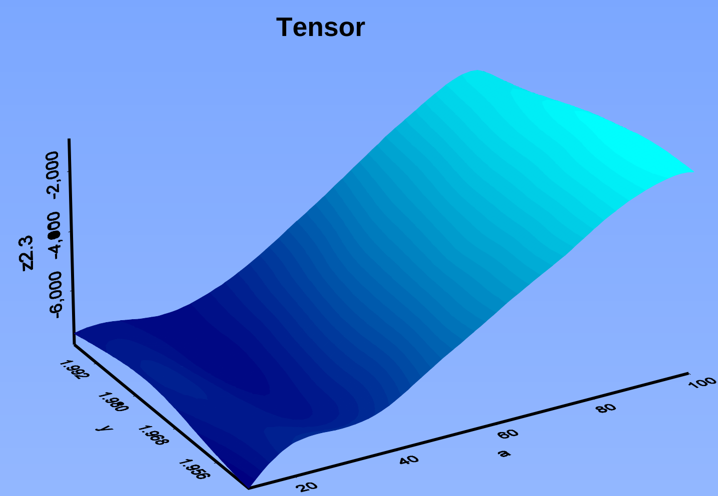
Additive model: Fitted curves for Ages 34 and 60



Tensor model: Fitted curves for Ages 34 and 60




Age-Period
z1.3
-2,200
-4,400
-6,600
1,992
1,980
1,968
1,956
y
20
40
60
80
100
a

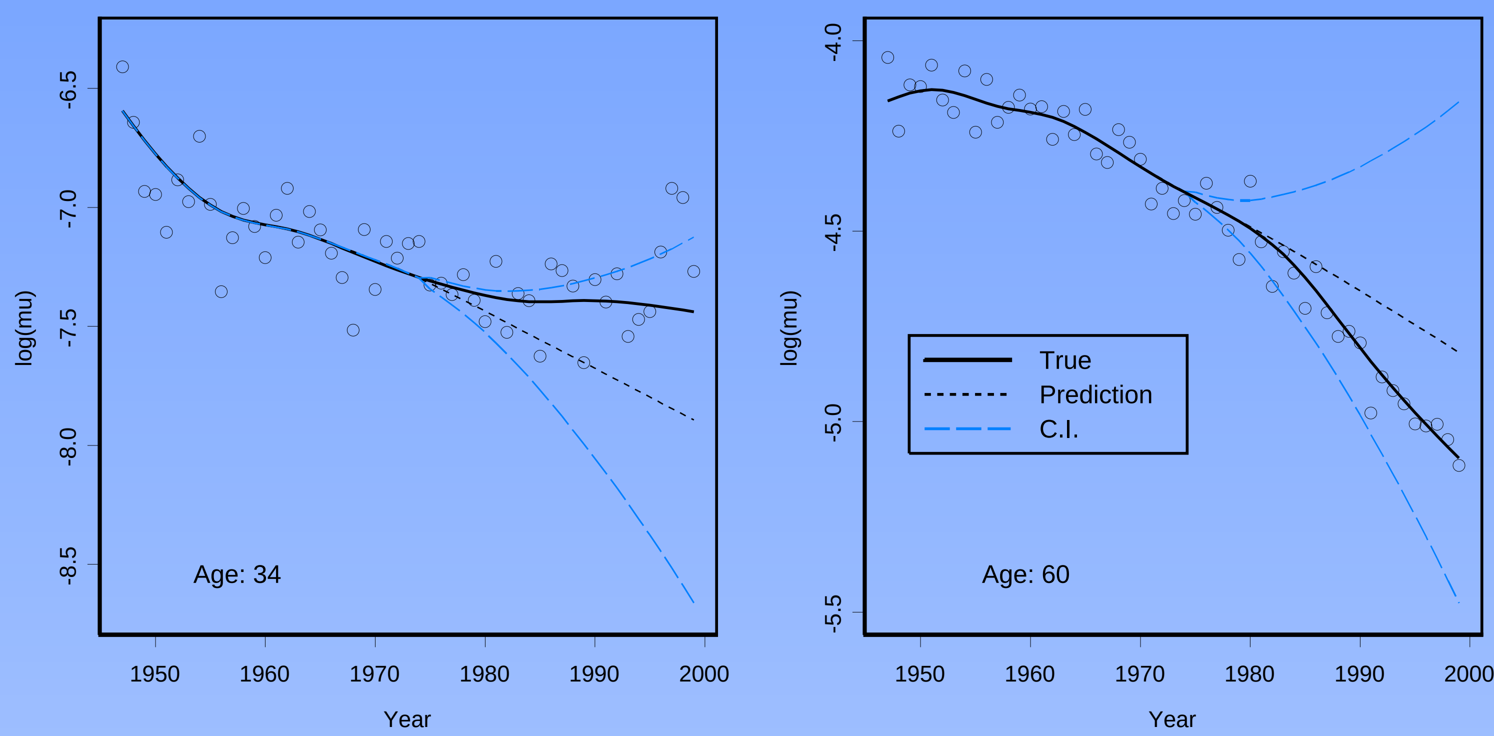

Tensor
z2.3
-2,000
-4,000
-6,000
1,992
1,980
1,968
1,956
y
20
40
60
80
100
a

Forecasting with P-splines

Treat the forecasting of future values as a missing value problem.

- We have data for n_y years and n_a ages and wish to forecast n_f years
- Define a weight matrix $\boldsymbol{V} = blockdiagonal(\boldsymbol{I}, \mathbf{0})$ $\boldsymbol{I}$ is an identity matrix of size $n_y n_a$, $\mathbf{0}$ is a square matrix of size n_f
- Define a new basis: $\tilde{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{V}$ and proceed as before

Forecast

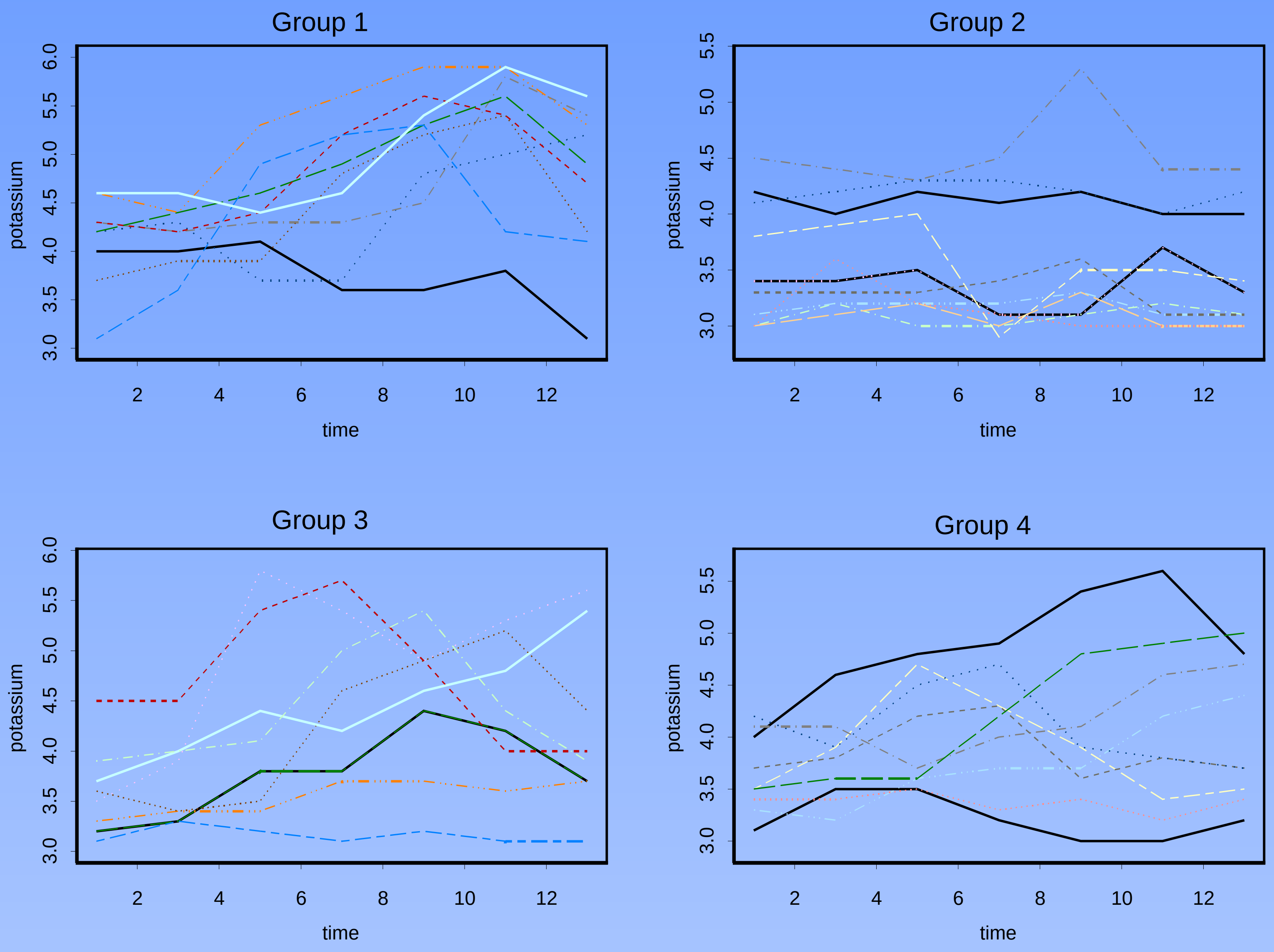

log(mu)
-6.5
-7.0
-7.5
-8.0
-8.5
Age: 34
1950
1960
1970
1980
1990
2000
Year
log(mu)
-4.0
-4.5
-5.0
-5.5
True
Prediction
C.I.
Age: 60
1950
1960
1970
1980
1990
2000
Year

P-splines for longitudinal data

The data

Objetive: Determine the effect of 4 surgical treatments on coronary sinus potasium in dogs

- 36 dogs
- 4 treatments
- 7 measurements per dog



Group 1
Group 2
Group 3
Group 4
potassium
time
3.0
3.5
4.0
4.5
5.0
5.5
6.0
2
4
6
8
10
12

Models for longitudinal data

Basic Model $y_{ij} = \alpha_0 + \alpha_1 t_{ij} + \beta_{i0} + \epsilon_{ij} \quad 1 \leq j \leq 7 \quad 1 \leq i \leq 36$

Models for longitudinal data

Basic Model $y_{ij} = \alpha_0 + \alpha_1 t_{ij} + \beta_{i0} + \epsilon_{ij} \quad 1 \leq j \leq 7 \quad 1 \leq i \leq 36$

$\Downarrow$ Relax linearity assumption

Model A $y_{ij} = f_{gr(i)}(t_{ij}) + \beta_{i0} + \epsilon_{ij} \quad 1 \leq gr(i) \leq 4$

Models for longitudinal data

Basic Model $y_{ij} = \alpha_0 + \alpha_1 t_{ij} + \beta_{i0} + \epsilon_{ij} \quad 1 \leq j \leq 7 \quad 1 \leq i \leq 36$

$\Downarrow$ Relax linearity assumption

Model A $y_{ij} = f_{gr(i)}(t_{ij}) + \beta_{i0} + \epsilon_{ij} \quad 1 \leq gr(i) \leq 4$

$\Downarrow$ Add random slope + general covariance matrix

Model B $y_{ij} = f_{gr(i)}(t_{ij}) + \beta_{i0} + \beta_{i1} t_{ij} + \epsilon_{ij}$

Models for longitudinal data

Basic Model $y_{ij} = \alpha_0 + \alpha_1 t_{ij} + \beta_{i0} + \epsilon_{ij} \quad 1 \leq j \leq 7 \quad 1 \leq i \leq 36$

$\Downarrow$ Relax linearity assumption

Model A $y_{ij} = f_{gr(i)}(t_{ij}) + \beta_{i0} + \epsilon_{ij} \quad 1 \leq gr(i) \leq 4$

$\Downarrow$ Add random slope + general covariance matrix

Model B $y_{ij} = f_{gr(i)}(t_{ij}) + \beta_{i0} + \beta_{i1} t_{ij} + \epsilon_{ij}$

$\Downarrow$ Subject specific curves

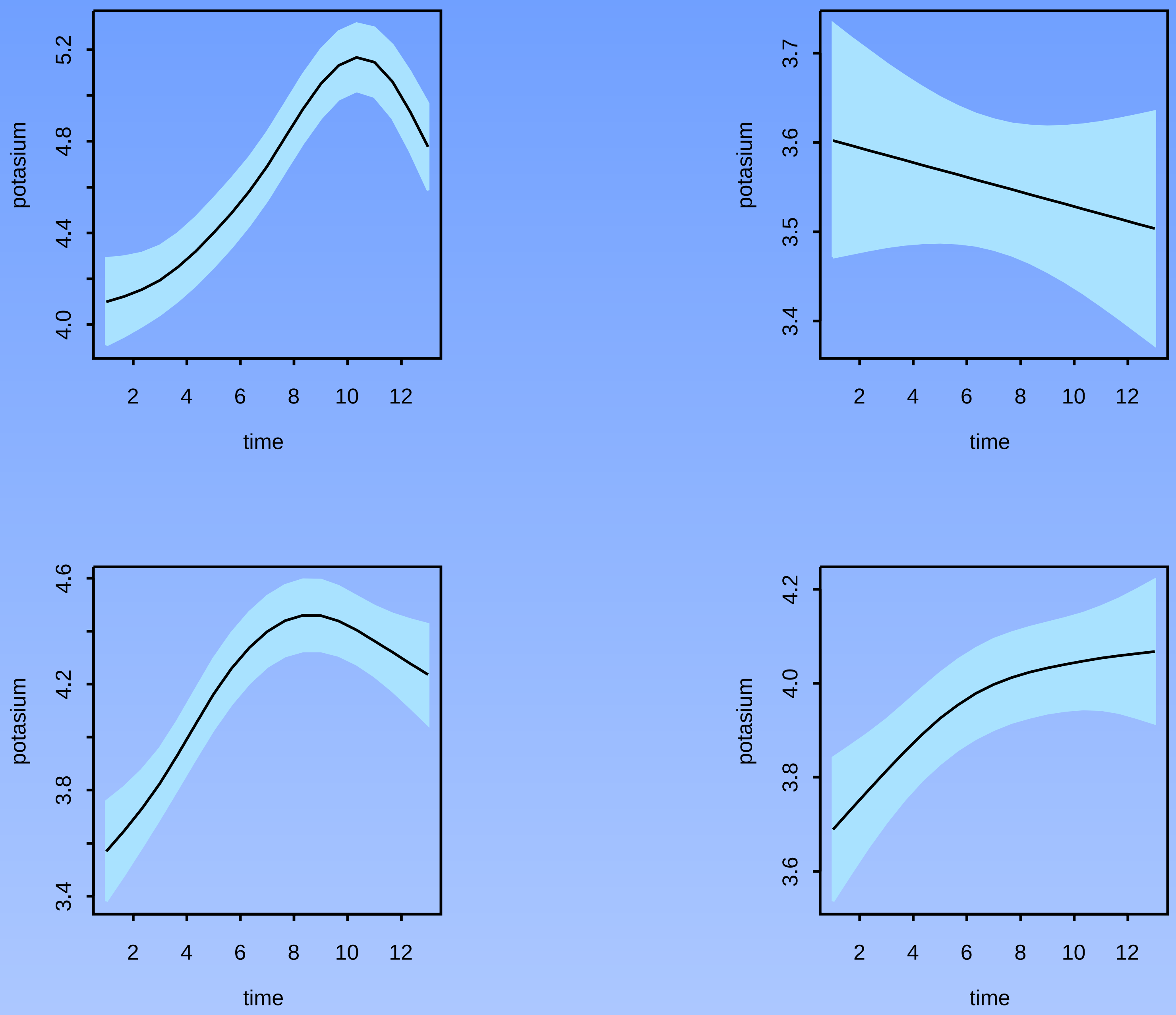
Model C $y_{ij} = f_{gr(j)}(t_{ij}) + g_i(t_{ij}) + \epsilon_{ij}$

The mixed model associated to Model A is:

$$\boldsymbol{y} = \boldsymbol{X} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon} \quad Cov\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \Sigma_{gr}\boldsymbol{I} & 0 & 0 \\ 0 & \sigma^2_{\beta_0} & 0 \\ 0 & 0 & \sigma^2\boldsymbol{I} \end{bmatrix}$$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_{\text{time}} \\ \vdots \\ \boldsymbol{X}_{\text{time}} \end{bmatrix} \quad \boldsymbol{X}_{\text{time}} = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_7 \end{bmatrix} \quad \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & & & & 1 & 0 & \cdots & 0 \\ & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & 1 & 0 & \cdots & 0 \\ & \boldsymbol{Z}_2 & & & 0 & 1 & \cdots & 0 \\ & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & 0 & 1 & \cdots & 0 \\ & & \boldsymbol{Z}_3 & & \vdots & \vdots & \vdots & \vdots \\ & & & & 0 & 0 & \cdots & 1 \\ & & & & \vdots & \vdots & \ddots & \vdots \\ & & & \boldsymbol{Z}_4 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$\boldsymbol{Z}_{\text{gr(i)}} = \begin{bmatrix} \boldsymbol{Z}_{\text{time}} \\ \vdots \\ \boldsymbol{Z}_{\text{time}} \end{bmatrix} \quad \Sigma_{gr} = \begin{bmatrix} \sigma_1^2\boldsymbol{I} & & & \\ & \sigma_2^2\boldsymbol{I} & & \\ & & \sigma_3^2\boldsymbol{I} & \\ & & & \sigma_4^2\boldsymbol{I} \end{bmatrix}$$


potasium
time
4.0
4.4
4.8
5.2
3.4
3.5
3.6
3.7
3.4
3.8
4.2
4.6
3.6
3.8
4.0
4.2
2
4
6
8
10
12

The mixed model associated to Model B is:

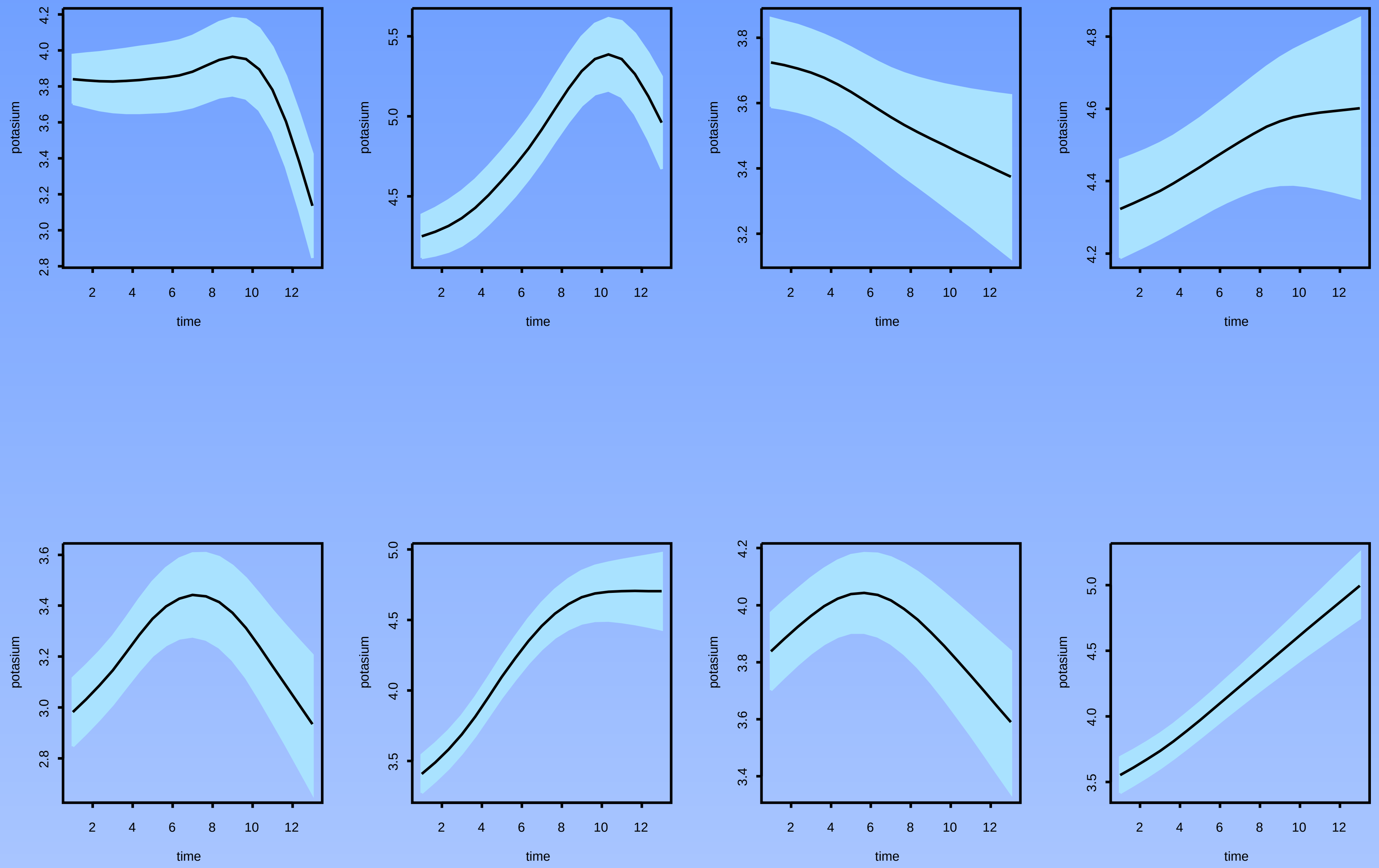
$$\boldsymbol{y} = \boldsymbol{X} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon} \quad Cov \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \Sigma_{gr} & 0 & 0 \\ 0 & blockdiag(\Sigma) & 0 \\ 0 & 0 & \sigma^2 \boldsymbol{I} \end{bmatrix}$$

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & & & & \boldsymbol{X}_{\mathsf{time}} & 0 & \cdots & 0 \\ & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & \boldsymbol{X}_{\mathsf{time}} & 0 & \cdots & 0 \\ & \boldsymbol{Z}_2 & & & 0 & \boldsymbol{X}_{\mathsf{time}} & \cdots & 0 \\ & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & 0 & \boldsymbol{X}_{\mathsf{time}} & \cdots & 0 \\ & & \boldsymbol{Z}_3 & & \vdots & \vdots & \vdots & \vdots \\ & & & & 0 & 0 & \cdots & \boldsymbol{X}_{\mathsf{time}} \\ & & & & \vdots & \vdots & \ddots & \vdots \\ & & & \boldsymbol{Z}_4 & 0 & 0 & \cdots & \boldsymbol{X}_{\mathsf{time}} \end{bmatrix}$$

The mixed model associated to Model C is:

$$\boldsymbol{y} = \boldsymbol{X} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon} \quad Cov\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \Sigma_{gr} & 0 & 0 & 0 \\ 0 & blockdiag(\Sigma) & 0 & 0 \\ 0 & 0 & \sigma_c^2 \boldsymbol{I} & 0 \\ 0 & 0 & 0 & \sigma^2 \boldsymbol{I} \end{bmatrix}$$

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & & & & \boldsymbol{X}_{\text{time}} & 0 & \cdots & 0 & \boldsymbol{Z}_{\text{time}} & 0 & \cdots & 0 \\ & & & & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ & & & & \boldsymbol{X}_{\text{time}} & 0 & \cdots & 0 & \boldsymbol{Z}_{\text{time}} & 0 & \cdots & 0 \\ & \boldsymbol{Z}_2 & & & 0 & \boldsymbol{X}_{\text{time}} & \cdots & 0 & 0 & \boldsymbol{Z}_{\text{time}} & \cdots & 0 \\ & & & & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ & & & & 0 & \boldsymbol{X}_{\text{time}} & \cdots & 0 & 0 & \boldsymbol{Z}_{\text{time}} & \cdots & 0 \\ & & \boldsymbol{Z}_3 & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & 0 & 0 & \cdots & \boldsymbol{X}_{\text{time}} & 0 & 0 & \cdots & \boldsymbol{Z}_{\text{time}} \\ & & & & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ & & & \boldsymbol{Z}_4 & 0 & 0 & \cdots & \boldsymbol{X}_{\text{time}} & 0 & 0 & \cdots & \boldsymbol{Z}_{\text{time}} \end{bmatrix}$$


potasium
time

Conclusions and work in progress

Conclusions and work in progress

- P-splines are useful tool to model data in many situations
- P-splines as mixed models
- Easy to implement in standard sorfware
- Model selection

References

References

- Currie, I. and Durban, M. and Eilers, P. (2003). Smoothing and forecasting mortality rates.
- Currie, I. and Durban, M. (2002). Flexible smoothing with P-splines: a unified approach. *Statistical Modelling* 2.
- Durbán, M. and Currie,I. (2003). A note on P-Spline additive models with correlated errors. *Computational Statistics*, 18.
- Durbán, M., Harezla,J., Carrol, R. and Wand, M. (2003). Simple fitting of subject-specific curves for longitudinal data.
- Eilers, P.H.C. & Marx, B.D. (1996). Flexible smoothing with B-splines ans penalties. *Statist. Sci.* 11.
- Ruppert, D., Wand, M.P., Carroll, R.J. (2003). Semiparametric Regression. Cambridge University Press.
- Wand, M.P. (2003). Smoothing and mixed models. *Comput. Stat.* 18.