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:

\[ y = f(x) + \epsilon \]

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**
P-spline

- Eilers and Marx, 1996.
- They are a generalisation of ordinary regression.
- Modify the log-likelihood by a penalty on the regression coefficients.

\[ y = f(x) + \epsilon \quad f(x) \approx Ba \quad S = (y - Ba)'(y - Ba) + \lambda a'Pa \]
\[ \hat{a} = (B'B + \lambda P)^{-1}B'y \]
P-spline

- Eilers and Marx, 1996.
- They are a generalisation of ordinary regression.
- Modify the log-likelihood by a penalty on the regression coefficients.

\[
y = f(x) + \epsilon \quad f(x) \approx Ba \quad S = (y - Ba)'(y - Ba) + \lambda a'Pa
\]

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

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:

$$1, \mathbf{x}, (\mathbf{x} - \kappa_1)_+, \ldots, (\mathbf{x} - \kappa_k)_+$$
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 dimensions and non Gaussian errors.
Psplines: mixed model approach
Psplines: mixed model approach

\[ y = f(x) + \epsilon \quad \epsilon \sim N(0, \sigma^2 R) \]

We write \( f(x) = B a \). It can be shown that \( B a \) may be written as

\[
\underbrace{X \beta}_{\text{fixed}} + \underbrace{Z u}_{\text{random}} \quad u \sim N(0, \sigma_u^2 I) \quad \lambda = \sigma^2 / \sigma_u^2
\]

\[
y = X \beta + Z u + \epsilon \quad Cov\begin{bmatrix} u \\ \epsilon \end{bmatrix} = \begin{bmatrix} \sigma_u^2 I & 0 \\ 0 & \sigma^2 R \end{bmatrix} \quad Cov[y] = V = R\sigma^2 + Z'Z\sigma_u^2
\]
Use **REML** for variance parameters

\[
l(V) = -\frac{1}{2} \log |V| - \frac{1}{2} \log |X'VX| - y'(V^{-1} - V^{-1}X(X'VX)^{-1}X'V^{-1})y,
\]

Given \( R, \sigma^2 \) and \( \sigma_u^2 \), \( \hat{\beta} \) and \( \hat{u} \) are solutions to:

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

- Unified approach
- Automatic selection of smoothing parameter
- Likelihood ratio test for model selection
- Already implemented in standard software: Splus, SAS, R.
APPLICATIONS
Additive models: Prestige data revisited

\[ y = \underbrace{f(\text{income})}_{X_1 \beta_1 + Z_1 u_1} + \underbrace{f(\text{education})}_{X_2 \beta_2 + Z_2 u_2} + \epsilon \]

\[ = X \beta + Zu + \epsilon \quad Cov \begin{bmatrix} u \\ \epsilon \end{bmatrix} = \begin{bmatrix} \sigma^2_{u_1} I & 0 & 0 \\ 0 & \sigma^2_{u_2} I & 0 \\ 0 & 0 & \sigma^2 I \end{bmatrix} \]

\[ \beta = (\beta'_1, \beta'_2)' \quad u = (u'_1, u'_2)' \quad X = [X_1 : X_2] \quad Z = [Z_1 : Z_2] \]
Partial residuals plot

- Income
  - Partial residuals
  - Income range: 0 to 25000
  - Partial residuals range: -20 to 20

- Education
  - Partial residuals
  - Education range: 6 to 16
  - Partial residuals range: -20 to 20

Plot shows the relationship between income, education, and partial residuals.
Is the model additive?: Conditional plots

[Conditional plots showing the relationship between education and prestige across different conditions.]
Two-dimensional P-splines

Now $y = f(\text{income, education}) + \epsilon = B a + \epsilon$, where

\[ B = B_1 \otimes B_2 \quad P = \lambda_1 P_1 \otimes I_{n_2} + \lambda_2 I_{n_2} \otimes P_1 \]
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.
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 $y = B\alpha + \epsilon$ with $\text{Var}(\epsilon) = \sigma^2 V$ and $V = W^{-1}$, $W = \text{diag}(w_1, \ldots, w_n)$.

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,

\[ f(y|\eta) = \exp \{ y'\eta - 1'b(\eta) + 1'c(y) \} \]

the linear predictor \( \eta = Ba \), using the mixed model representation of \( P \)-splines we rewrite \( Ba = X\beta + Zu \)

\[ f(y|u) = \exp \{ y'(X\beta + Zu) - 1'\exp(X\beta + Zu) - 1'\log(\Gamma(y + 1)) \} \]

and \( u \sim N(0, \sigma_u^2 I) \).

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

In the case of count data \( \lambda = 1/\sigma_u^2 \).
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

down

Model mortality trends overtime and dependence on age.
Additive model: Fitted curves for Ages 34 and 60
Tensor model: Fitted curves for Ages 34 and 60
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 $V = \text{blockdiagonal}(I, 0)$ $I$ is an identity matrix of size $n_y n_a$, $0$ is a square matrix of size $n_f$

- Define a new basis: $\tilde{B} = BV$ and proceed as before
Forecast

Age: 34

Year

-8.5 -8.0 -7.5 -7.0 -6.5

log(mu)

Age: 60

Year

-5.5 -5.0 -4.5 -4.0

log(mu) True

Prediction

C.I.
P-splines for longitudinal data
The data

Objective: Determine the effect of 4 surgical treatments on coronary sinus potassium in dogs

- 36 dogs
- 4 treatments
- 7 measurements per dog
Models for longitudinal data

Basic Model \( y_{ij} = \alpha_0 + \alpha_1 t_{ij} + \beta_i + \epsilon_{ij} \quad 1 \leq j \leq 7 \quad 1 \leq i \leq 36 \)
Models for longitudinal data

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

\[ \Downarrow \quad \text{Relax linearity assumption} \]

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 \[ y_{ij} = \alpha_0 + \alpha_1 t_{ij} + \beta_i + \epsilon_{ij} \quad 1 \leq j \leq 7 \quad 1 \leq i \leq 36 \]

\[ \downarrow \text{Relax linearity assumption} \]

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

\[ \downarrow \text{Add random slope + general covariance matrix} \]

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

\[ \downarrow \text{Subject specific curves} \]

Model C \[ y_{ij} = f_{gr(j)}(t_{ij}) + g_i(t_{ij}) + \epsilon_{ij} \]
The mixed model associated to Model A is:

\[ y = X + Zu + \epsilon \]

\[
Cov \begin{bmatrix}
u \\
\epsilon
\end{bmatrix} =
\begin{bmatrix}
\Sigma_{gr}I & 0 & 0 \\
0 & \sigma^2_{\beta_0} & 0 \\
0 & 0 & \sigma^2I
\end{bmatrix}
\]

\[
X = \begin{bmatrix}
X_{\text{time}} \\
\vdots \\
X_{\text{time}}
\end{bmatrix}
\]

\[
X_{\text{time}} = \begin{bmatrix}
1 & t_1 \\
\vdots & \vdots \\
1 & t_7
\end{bmatrix}
\]

\[
Z = \begin{bmatrix}
Z_1 & \mathbf{1} & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
1 & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & 1 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
Z_4 & 0 & 0 & \ldots & 1
\end{bmatrix}
\]

\[
Z_{gr(i)} = \begin{bmatrix}
Z_{\text{time}} \\
\vdots \\
Z_{\text{time}}
\end{bmatrix}
\]

\[
\Sigma_{gr} = \begin{bmatrix}
\sigma^2_1I & \sigma^2_2I & \sigma^2_3I & \sigma^2_4I
\end{bmatrix}
\]
The mixed model associated to Model B is:

\[ y = X + Zu + \epsilon \quad Cov \begin{bmatrix} u \\ \epsilon \end{bmatrix} = \begin{bmatrix} \Sigma_{gr} & 0 \\ 0 & \text{blockdiag}(\Sigma) \\ 0 & 0 & \sigma^2 I \end{bmatrix} \]

\[ Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_2 \\ \vdots \\ Z_3 \\ \vdots \\ Z_4 \end{bmatrix} \begin{bmatrix} X_{\text{time}} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & X_{\text{time}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_{\text{time}} \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} \]
The mixed model associated to Model C is:

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

\[
Z = \begin{bmatrix}
Z_1 & X_{time} & 0 & \cdots & 0 & Z_{time} & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
Z_2 & 0 & X_{time} & 0 & \cdots & 0 & Z_{time} & 0 & \cdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
Z_3 & 0 & 0 & X_{time} & 0 & \cdots & 0 & Z_{time} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
Z_4 & 0 & 0 & 0 & X_{time} & 0 & 0 & 0 & Z_{time}
\end{bmatrix}
\]
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 software
- Model selection
References
References