Smoothing and Forecasting
Mortality Rates with P-splines

Iain Currie
Heriot Watt University

London, June 2006
Data and problem

- Data: CMI assured lives
  - Age: 20 to 90
  - Year: 1947 to 2002
Data and problem

- Data: CMI assured lives
  - Age: 20 to 90
  - Year: 1947 to 2002
- Problem: forecast table to 2046
Plan of talk

- P-splines in 1-dimension
Plan of talk

- P-splines in 1-dimension
- P-splines in 2-dimensions
  - Lee-Carter model
  - Age-Period-Cohort model
  - 2-d P-splines
Gompertz model

Year 2002

Gompertz: \( \log m = -11.73 + 0.109 \text{ Age} \)
Simple Gompertz

Age 70

\[ \log m = 26.9 - 0.0154 \text{ Year} \]
Age 70

\[ \log m = 26.9 - 0.0154 \text{ Year} \]

\[ \log m = -1100 + 1.127 \text{ Year} - 0.000289 \text{ Year}^2 \]
Quadratic forecast

Age 70

Year

Log(mortality)
Quadratic fit: linear & quadratic forecasts

Age 70

Quadratic forecast

Linear forecast
Regression basis

In ordinary regression we use basis like

Linear basis: \( \{1, x\} \)

Quadratic basis: \( \{1, x, x^2\} \)

Polynomial basis: \( \{1, x, x^2, \ldots, x^p\} \).
Regression basis

In ordinary regression we use basis like

**Linear basis:** \( \{1, x\} \)

**Quadratic basis:** \( \{1, x, x^2\} \)

**Polynomial basis:** \( \{1, x, x^2, \ldots, x^p\} \).

Means and fitted values are linear combinations of the basis functions

\[
\log \mu = a + bx, \quad \hat{\log \mu} = \hat{a} + \hat{b}x \\
\log \mu = a + bx + cx^2, \quad \hat{\log \mu} = \hat{a} + \hat{b}x + \hat{c}x^2 \\
\log \mu = a + b_1 x + \ldots + b_p x^p, \quad \hat{\log \mu} = \hat{a} + \hat{b}_1 x + \ldots + \hat{b}_p x^p
\]
Generalised linear models

Model fitting uses the Generalised Linear Model framework.

\[ d_x \sim \mathcal{P}(E_x^c \mu_x) \]

and \[ \log E_x^c \mu_x = \log E_x^c + a + bx \]

or \[ \log E_x^c \mu_x = \log E_x^c + a + bx + cx^2. \]
B-spline basis

A B-spline regression basis uses local basis functions.

B-spline basis: \{B_1(x), B_2(x), \ldots, B_p(x)\}

where $B_1(x), B_2(x), \ldots, B_p(x)$ are B-splines.
B-spline basis

A B-spline regression basis uses local basis functions.

B-spline basis: \{B_1(x), B_2(x), \ldots, B_p(x)\}

where \(B_1(x), B_2(x), \ldots, B_p(x)\) are B-splines.

Means and fitted values work the same way

\[
\log \mu = \sum_{1}^{p} \theta_j B_j(x), \quad \log \hat{\mu} = \sum_{1}^{p} \hat{\theta}_j B_j(x)
\]
A single cubic B–spline
Cubic B-spline basis

Age

B-spline
B-spline regression

Age 70
Number of B-splines: 23
B-spline regression

Year

log(Mortality)

Age 70

Number of B–splines: 23
Eilers & Marx (1996) imposed penalties on differences between adjacent coefficients.

\[
P_2(\theta) = (\theta_1 - 2\theta_2 + \theta_3)^2 + \ldots + (\theta_{p-2} - 2\theta_{p-1} + \theta_p)^2
= \theta' D_2' D_2 \theta
\]

where \( D_2 \) is a difference matrix.

\( P_2(\theta) \) is a roughness penalty.
Penalties

Eilers & Marx (1996) imposed penalties on differences between adjacent coefficients.

\[ P_2(\theta) = (\theta_1 - 2\theta_2 + \theta_3)^2 + \ldots + (\theta_{p-2} - 2\theta_{p-1} + \theta_p)^2 \]

\[ = \theta' D_2' D_2 \theta \]

where \( D_2 \) is a difference matrix.

\( P_2(\theta) \) is a roughness penalty.

Estimation is via penalised likelihood

\[ PL(\theta) = L(\theta) - \frac{1}{2} \lambda \theta' D_2' D_2 \theta \]

where \( \lambda \) is the smoothing parameter which balances fit and smoothness.
Penalties

Eilers & Marx (1996) imposed penalties on differences between adjacent coefficients.

\[
P_2(\theta) = (\theta_1 - 2\theta_2 + \theta_3)^2 + \ldots + (\theta_{p-2} - 2\theta_{p-1} + \theta_p)^2
= \theta' D_2' D_2 \theta
\]

where \( D_2 \) is a difference matrix.

\( P_2(\theta) \) is a roughness penalty.

Estimation is via penalised likelihood

\[
PL(\theta) = L(\theta) - \frac{1}{2} \lambda \theta' D_2' D_2 \theta
\]

where \( \lambda \) is the smoothing parameter which balances fit and smoothness.

- Interpolation (\( \lambda = 0 \))
- Linear regression (\( \lambda = \infty \))
P-spline regression

Age 70

Number of B-splines: 23

Quadratic penalty
Forecasting to 2046

Age 70
Forecast to 2046
Quadratic penalty: linear forecast
Forecasting to 2046

Age 70

Delta knots = 2.75

Delta knots = 11
Lee-Carter model

Lee & Carter (1992)

\[ \log \mu_{ij} = \alpha_i + \beta_i \kappa_j, \quad 20 \leq i \leq 90, \quad 1947 \leq j \leq 2002 \]
Lee–Carter forecasts to 2046

The graph shows the Lee–Carter forecasts for various age groups (40, 50, 60, 70, 80, 90) over the years from 1960 to 2040. The y-axis represents the log of mortality, while the x-axis represents the year. The forecast is for the period up to 2046.
Discrete Lee-Carter model

\[ \log \mu_{ij} = \alpha_i + \beta_i \kappa_j, \quad 20 \leq i \leq 90, \quad 1947 \leq j \leq 2002 \]

or in table form

\[ \log M = \alpha' + \beta' \kappa' \]
Discrete Lee-Carter model

\[ \log \mu_{ij} = \alpha_i + \beta_i \kappa_j, \quad 20 \leq i \leq 90, \ 1947 \leq j \leq 2002 \]

or in table form

\[ \log M = \alpha_1' + \beta \kappa' \]

Smooth Lee-Carter

\[ \alpha \to B_{a \ a}, \quad \beta \to B_{a \ b}, \quad \kappa \to B_{y \ k} \]
Age-Period-Cohort model

\[ \log \mu_{ij} = \alpha_i + \beta_j + \gamma_{i-j}, \quad 20 \leq i \leq 90, \ 1947 \leq j \leq 2002 \]
Age-Period-Cohort model

\[ \log \mu_{ij} = \alpha_i + \beta_j + \gamma_{i-j}, \quad 20 \leq i \leq 90, \quad 1947 \leq j \leq 2002 \]

Parameter redundancy

- \(2n_a + 2n_y - 1 = 253\) parameters
- \(2n_a + 2n_y - 4 = 250\) free parameters
Discrete Age-Period-Cohort model

\[ \log \mu_{ij} = \alpha_i + \beta_j + \gamma_{i-j}, \quad 20 \leq i \leq 90, \quad 1947 \leq j \leq 2002 \]
Discrete Age-Period-Cohort model

\[ \log \mu_{ij} = \alpha_i + \beta_j + \gamma_{i-j}, \quad 20 \leq i \leq 90, \quad 1947 \leq j \leq 2002 \]

Smooth Age-Period-Cohort model

\[ \alpha \rightarrow B_a a, \quad \beta \rightarrow B_y b, \quad \gamma \rightarrow B_c c \]
Components of Age

Improvement = 0.015/annum
Components of Year

**Age 90**

Improvement = 0.7/10 years

**Age 20**
Components of Cohort

Improvement = 0.29/10 years
2-d modelling of mortality tables

- Let $B_a$, $71 \times 13$, be a 1-d B-spline basis for modelling a single age.
- Let $B_y$, $56 \times 10$, be a 1-d B-spline basis for modelling a single year.
2-d modelling of mortality tables

- Let $B_a$, $71 \times 13$, be a 1-d B-spline basis for modelling a single age.
- Let $B_y$, $56 \times 10$, be a 1-d B-spline basis for modelling a single year.

Can we combine the marginal bases $B_a$ and $B_y$ to make a 2-d basis?
2-d modelling of mortality tables

Regression matrix

The Kronecker product organises the multiplication of the marginal bases to give the regression matrix

$$B = B_y \otimes B_a, \; 3976 \times 130.$$
2-d modelling of mortality tables

Regression matrix

The Kronecker product organises the multiplication of the marginal bases to give the regression matrix

\[ B = B_y \otimes B_a, \ 3976 \times 130. \]

Penalties in 2-d

- Each regression coefficient is associated with the summit of one of the hills.
- Smoothness is ensured by penalizing the coefficients in rows and columns.
2-d mortality: ages 40 to 90
# Summary of results

<table>
<thead>
<tr>
<th>Method</th>
<th>DEV</th>
<th>TR</th>
<th>BIC</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-d</td>
<td>6832</td>
<td>63</td>
<td>7357</td>
<td>1</td>
</tr>
<tr>
<td>APC smooth</td>
<td>7852</td>
<td>38</td>
<td>8167</td>
<td>2</td>
</tr>
<tr>
<td>LC smooth</td>
<td>8568</td>
<td>30</td>
<td>8815</td>
<td>3</td>
</tr>
<tr>
<td>APC</td>
<td>7002</td>
<td>248</td>
<td>9057</td>
<td>4</td>
</tr>
<tr>
<td>LC</td>
<td>8004</td>
<td>196</td>
<td>9629</td>
<td>5</td>
</tr>
</tbody>
</table>
Summary for age 60

![Graph showing mortality trends from 1960 to 2040, with various lines representing different mortality models: APC, 2-d, LC, SmoothLC. The x-axis represents the year, and the y-axis represents log(mortality).]
Summary for ages 60, 70, 80

Year

log(mortality)

60 70 80

APC 2–d LCSmoothLC
Conclusions

Forecasting a mortality table depends on...
Conclusions

Forecasting a mortality table depends on

- Model choice
Conclusions

Forecasting a mortality table depends on

- Model choice
- Forecasting method
Conclusions

Forecasting a mortality table depends on

- Model choice
- Forecasting method
- Parameter uncertainty
Conclusions

Forecasting a mortality table depends on

- Model choice
- Forecasting method
- Parameter uncertainty
- Stochastic uncertainty
References

Lee-Carter models

Age-Period-Cohort models
Clayton & Schlifflers (1987) Statistics in Medicine, 6, 469-481.

Penalized spline models