# 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

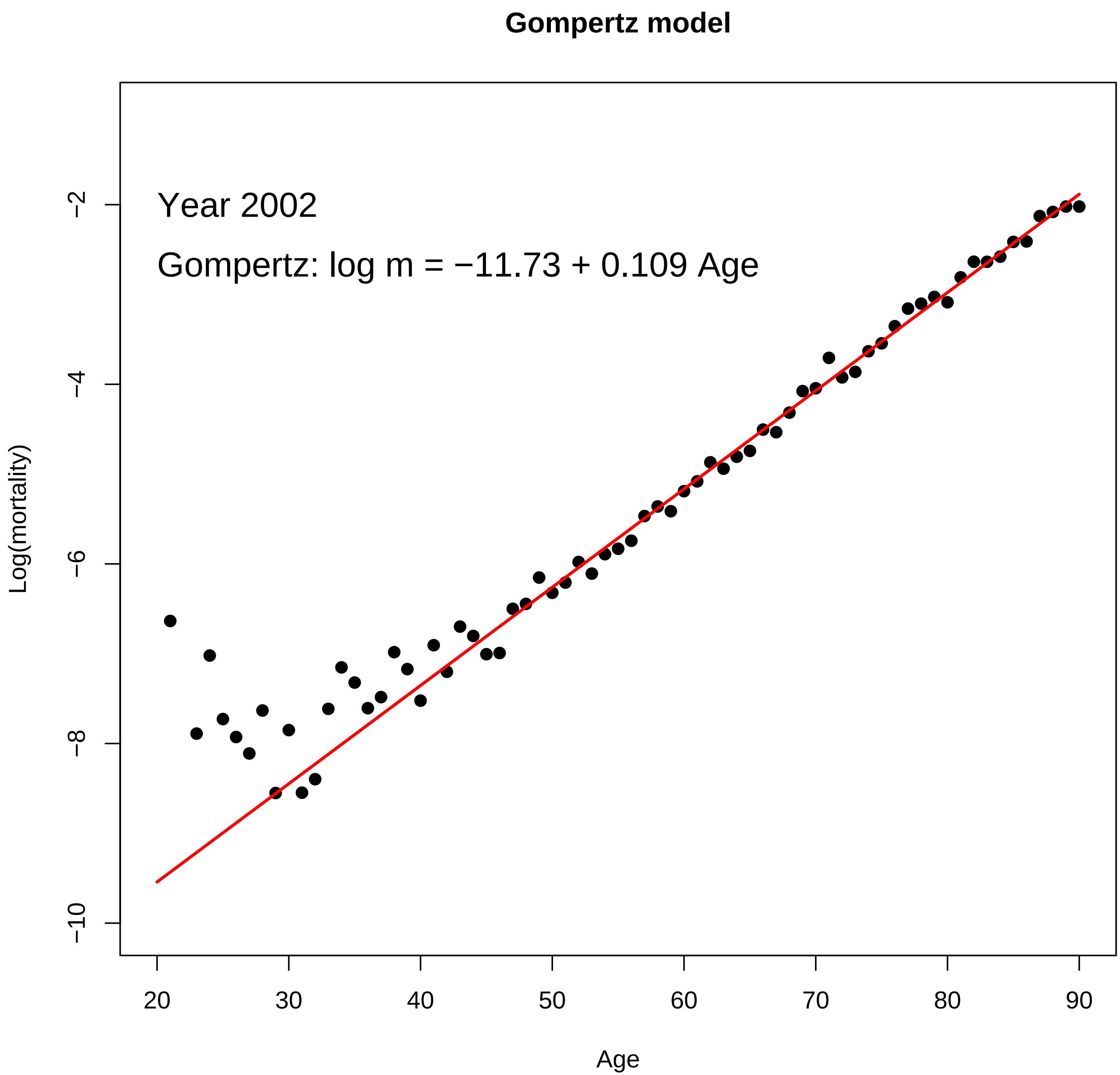
- Data: CMI assured lives
  - Age: 20 to 90
  - Year: 1947 to 2002
- Problem: forecast table to 2046

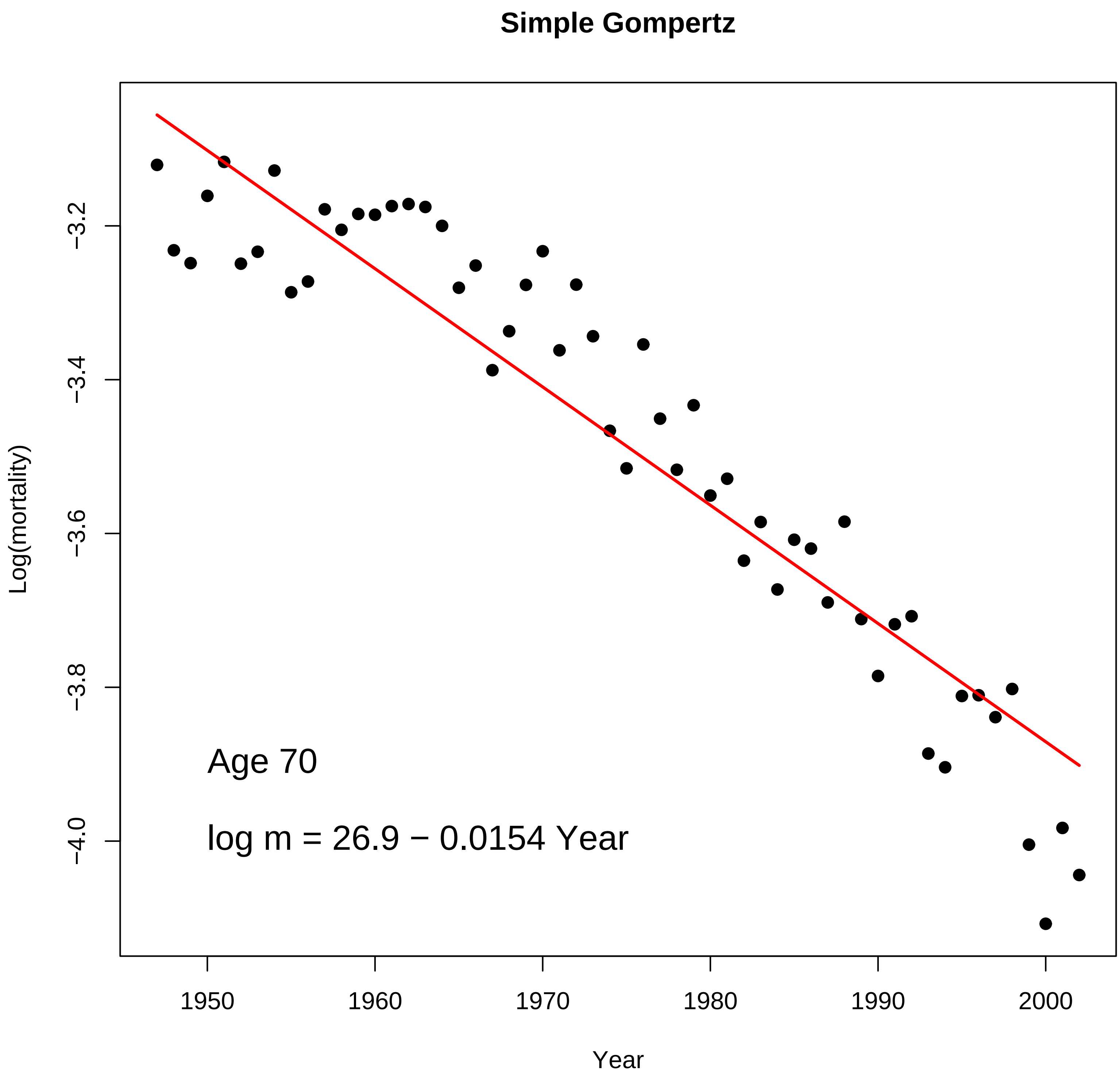
## Plan of talk

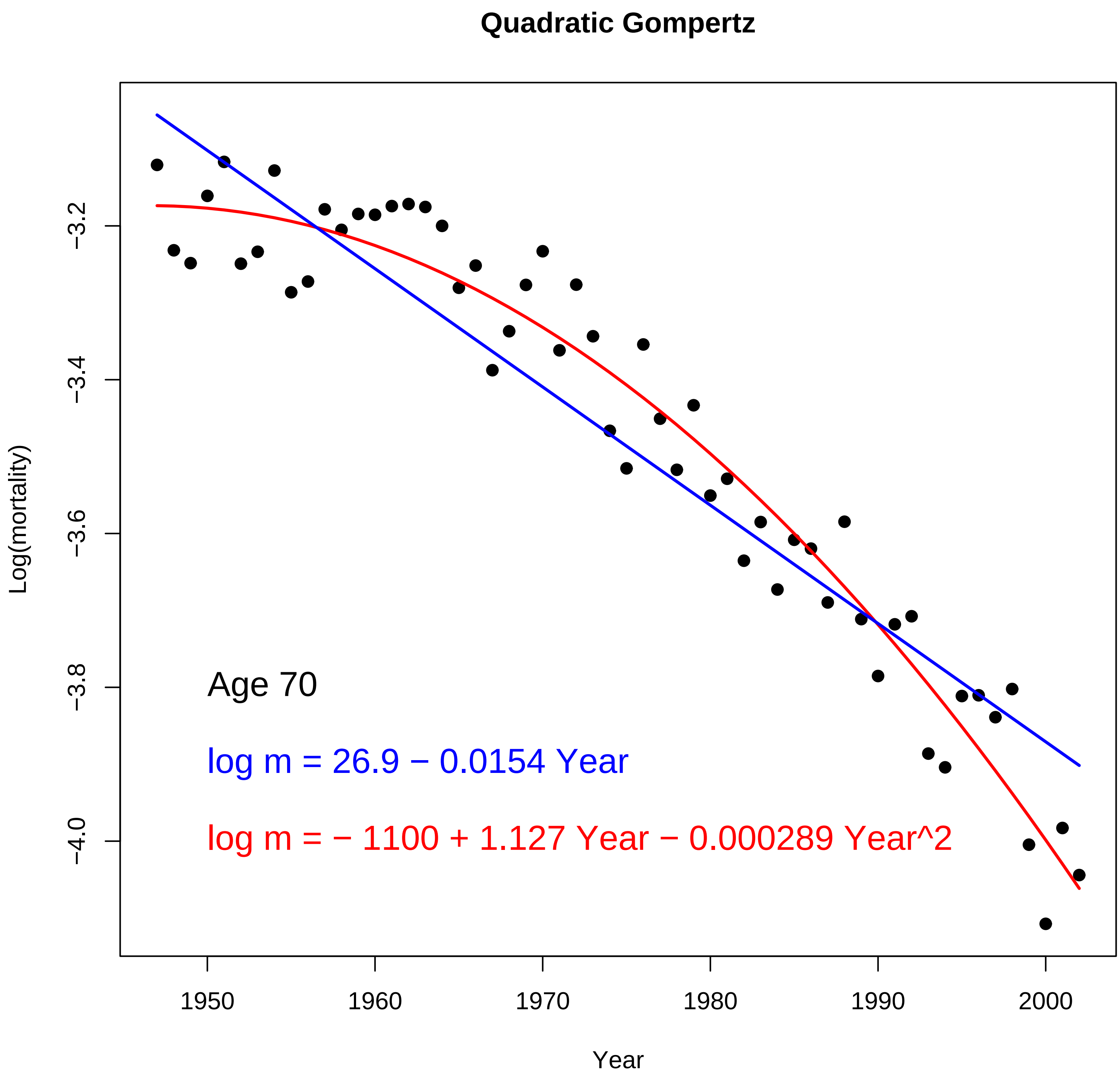
- P-splines in 1-dimension

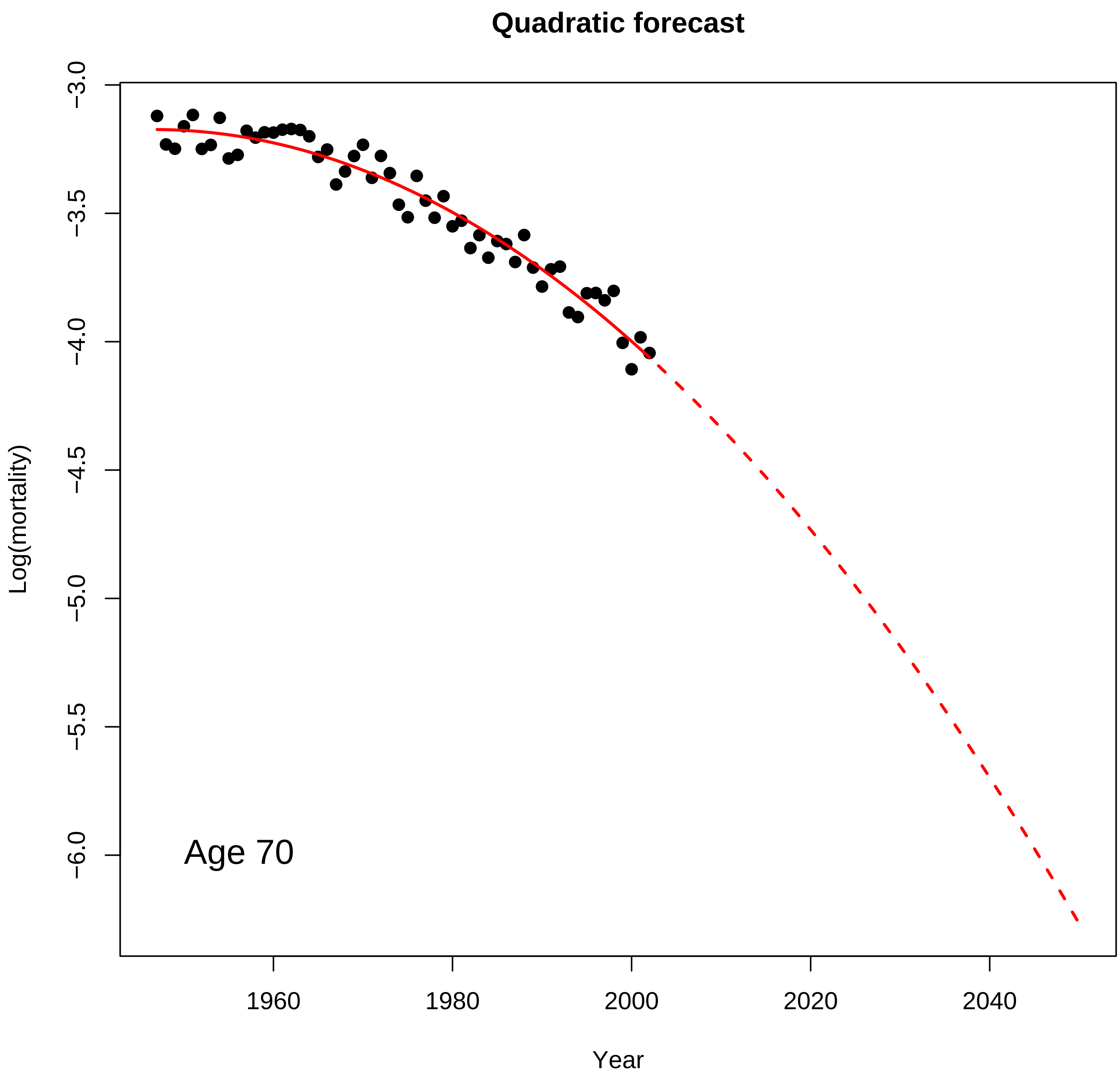
## Plan of talk

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

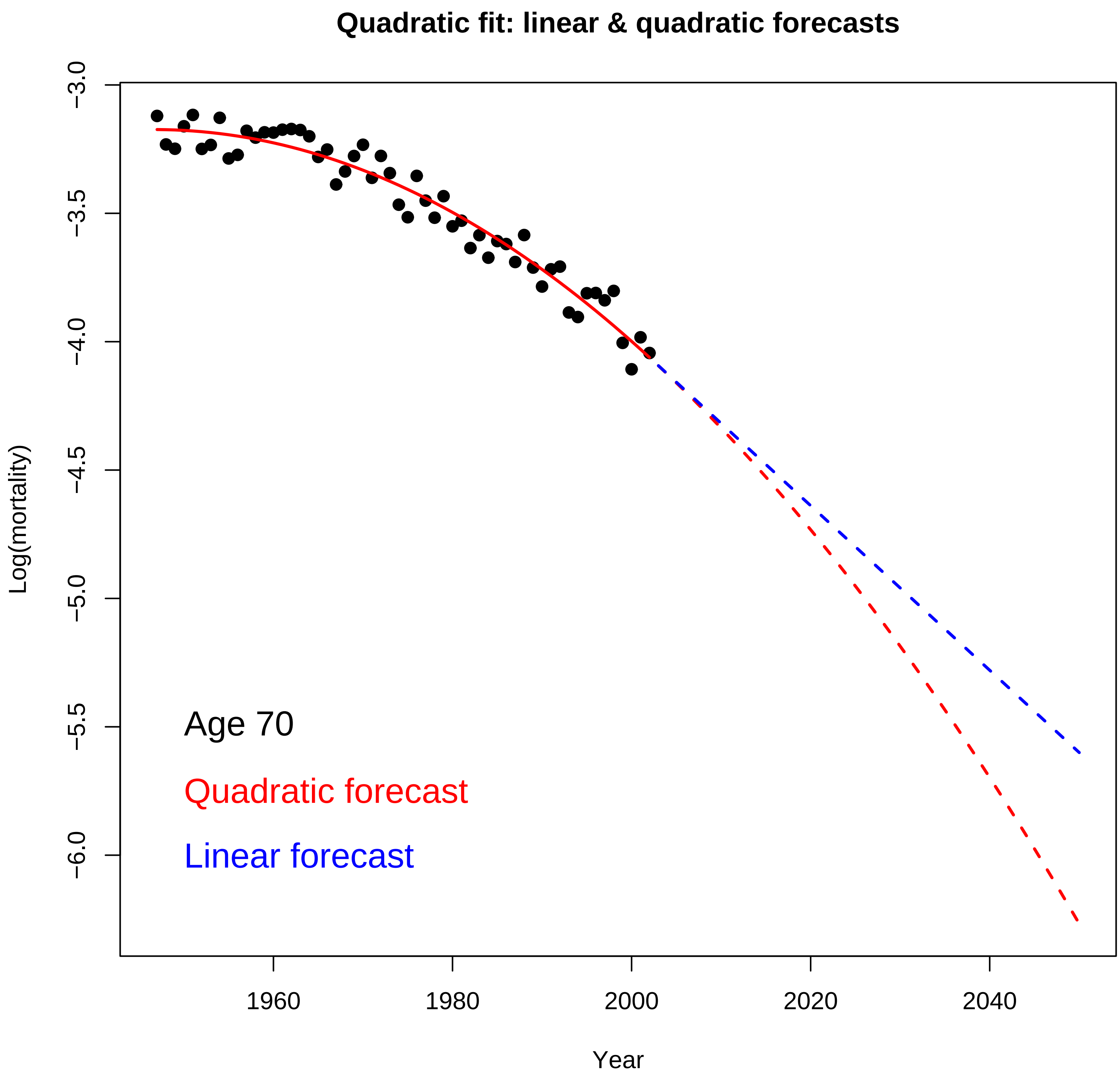

Gompertz model
Year 2002
Gompertz: log m = −11.73 + 0.109 Age
Log(mortality)
−2
−4
−6
−8
−10
20
30
40
50
60
70
80
90
Age


Simple Gompertz
Age 70
log m = 26.9 − 0.0154 Year
Log(mortality)
−3.2
−3.4
−3.6
−3.8
−4.0
1950
1960
1970
1980
1990
2000
Year


Quadratic Gompertz
Log(mortality)
−3.2
−3.4
−3.6
−3.8
−4.0
1950
1960
1970
1980
1990
2000
Year
Age 70
log m = 26.9 − 0.0154 Year
log m = − 1100 + 1.127 Year − 0.000289 Year^2



Quadratic forecast
Log(mortality)
−3.0
−3.5
−4.0
−4.5
−5.0
−5.5
−6.0
1960
1980
2000
2020
2040
Year
Age 70



Quadratic fit: linear & quadratic forecasts
Log(mortality)
−3.0
−3.5
−4.0
−4.5
−5.0
−5.5
−6.0
1960
1980
2000
2020
2040
Year
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

$$\log \mu = a + bx, \qquad \log \hat{\mu} = \hat{a} + \hat{b}x$$

$$\log \mu = a + bx + cx^2, \qquad \log \hat{\mu} = \hat{a} + \hat{b}x + \hat{c}x^2$$

$$\log \mu = a + b_1 x + \ldots + b_p x^p, \qquad \log \hat{\mu} = \hat{a} + \hat{b}_1 x + \ldots + \hat{b}_p x^p$$

# Generalised linear models

Model fitting uses the Generalised Linear Model framework.

$$
\begin{aligned}
& d_x \sim \mathcal{P}(E_x^c \mu_x) \\
\text{and} \quad & \log E_x^c \mu_x = \log E_x^c + a + bx \\
\text{or} \quad & \log E_x^c \mu_x = \log E_x^c + a + bx + cx^2.
\end{aligned}
$$

# 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

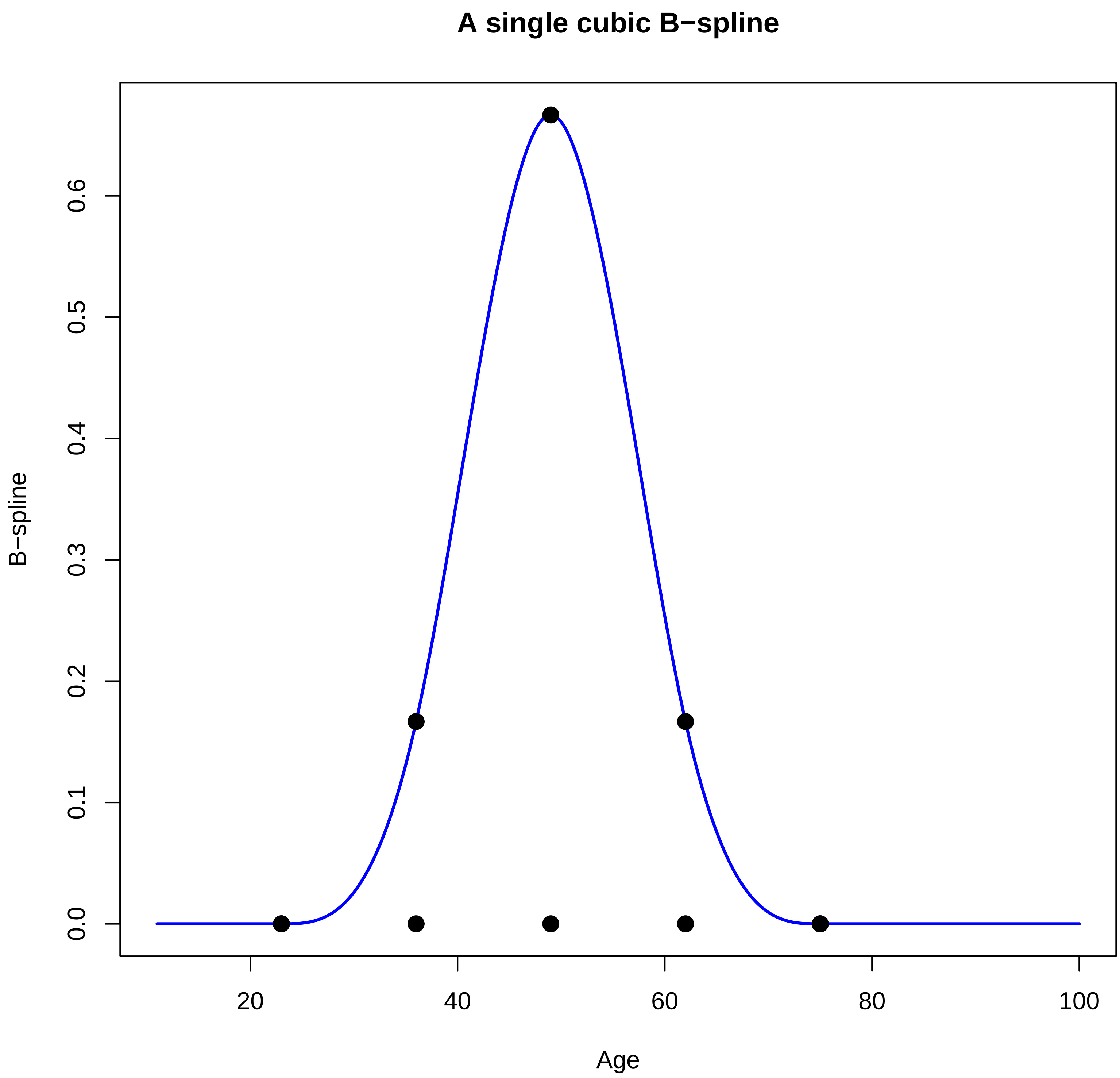
A B-spline regression basis uses local basis functions.

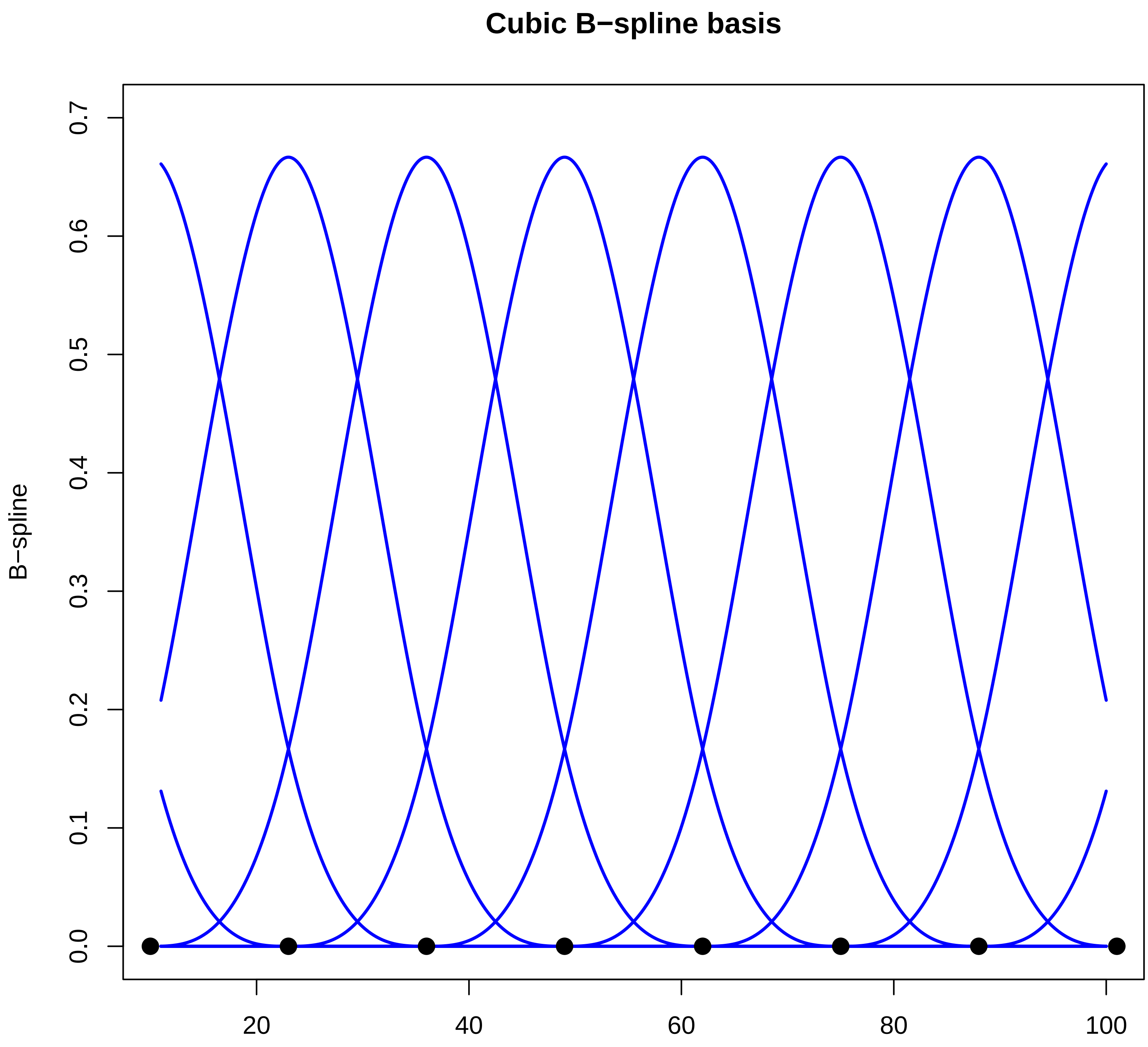
B-spline basis: $\{B_1(x), B_2(x), \ldots, B_p(x)\}$

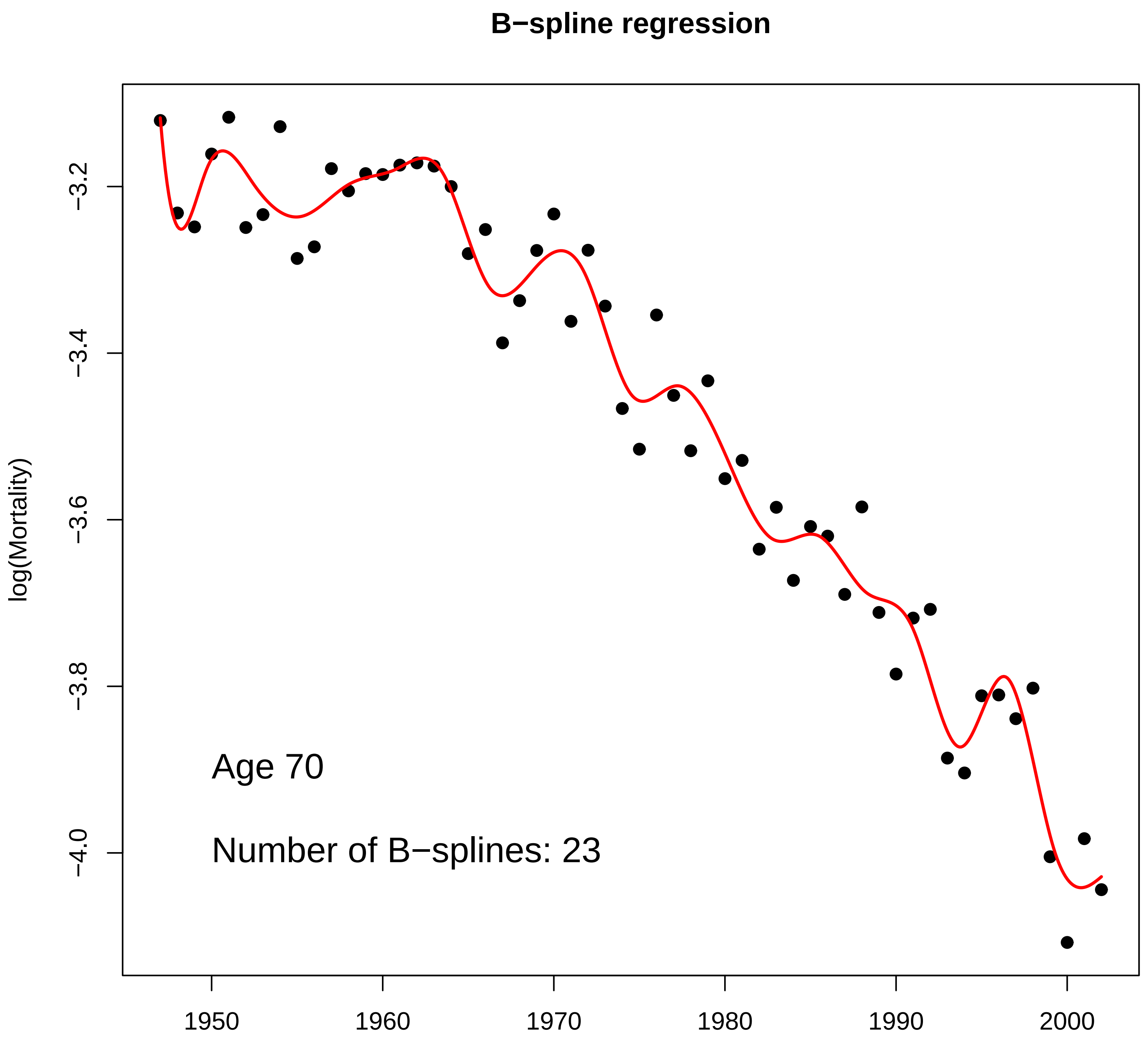
where $B_1(x), B_2(x), \ldots, B_p(x)$ are B-splines.

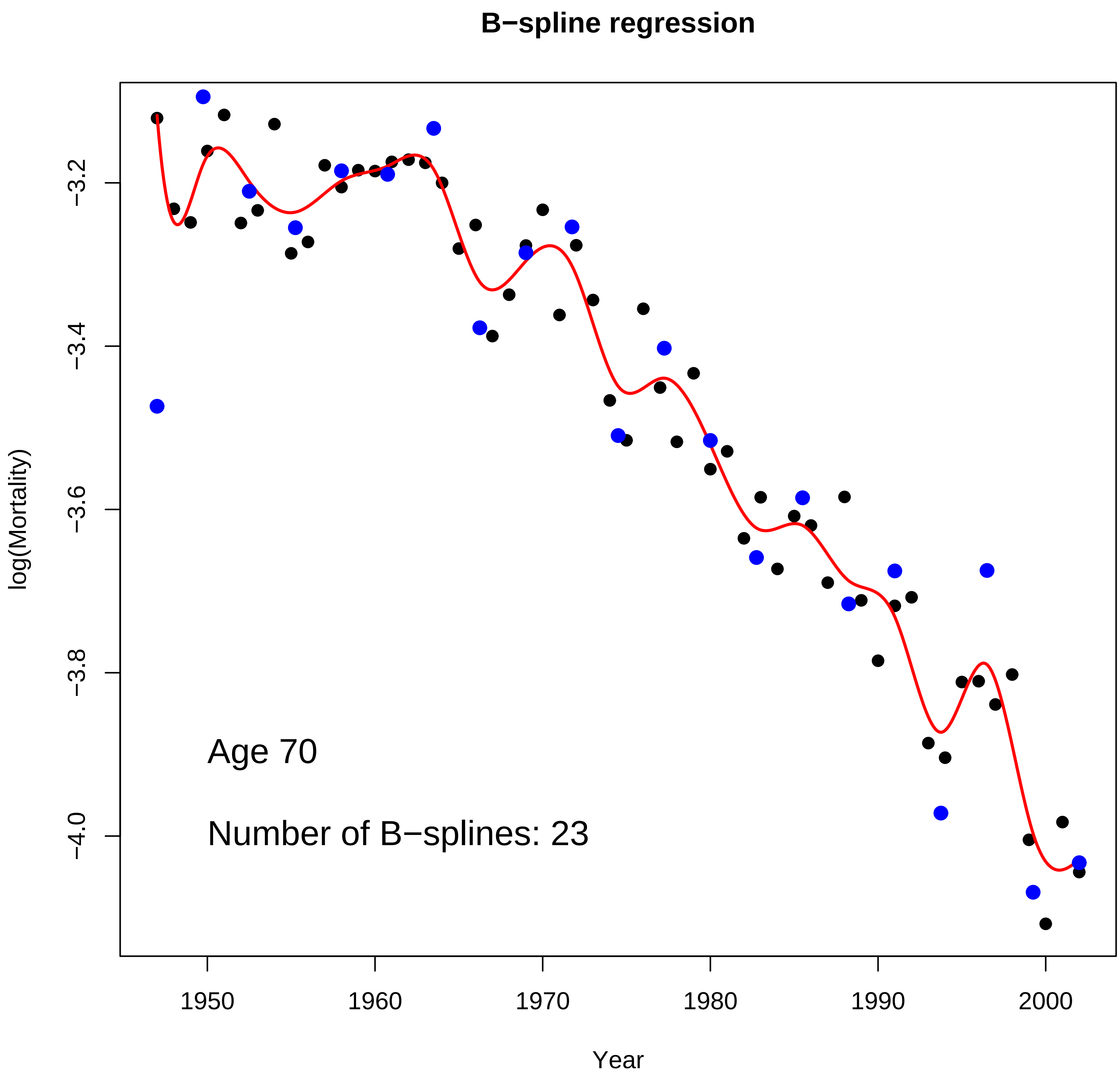
Means and fitted values work the same way

$$\log \mu = \sum_{1}^{p} \theta_j B_j(x), \quad \log \hat{\mu} = \sum_{1}^{p} \hat{\theta}_j B_j(x)$$


A single cubic B−spline
B−spline
0.0
0.1
0.2
0.3
0.4
0.5
0.6
20
40
60
80
100
Age


Cubic B−spline basis
B−spline
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
20
40
60
80
100
Age


B−spline regression
log(Mortality)
−3.2
−3.4
−3.6
−3.8
−4.0
1950
1960
1970
1980
1990
2000
Year
Age 70
Number of B−splines: 23


B−spline regression
log(Mortality)
−3.2
−3.4
−3.6
−3.8
−4.0
1950
1960
1970
1980
1990
2000
Year
Age 70
Number of B−splines: 23

# Penalties

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

$$
\begin{aligned}
P_2(\theta) &= (\theta_1 - 2\theta_2 + \theta_3)^2 + \ldots + (\theta_{p-2} - 2\theta_{p-1} + \theta_p)^2 \\
&= \theta' D_2' D_2 \theta
\end{aligned}
$$

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.

$$\begin{aligned} P_2(\theta) &= (\theta_1 - 2\theta_2 + \theta_3)^2 + \ldots + (\theta_{p-2} - 2\theta_{p-1} + \theta_p)^2 \\ &= \theta' D_2' D_2 \theta \end{aligned}$$

where $D_2$ is a difference matrix.

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

Estimation is via penalised likelihood

$$PL(\theta) = L(\theta) - \tfrac{1}{2}\lambda\theta' D_2' D_2 \theta$$

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

# Penalties

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

$$
\begin{aligned}
P_2(\theta) &= (\theta_1 - 2\theta_2 + \theta_3)^2 + \ldots + (\theta_{p-2} - 2\theta_{p-1} + \theta_p)^2 \\
&= \theta' D_2' D_2 \theta
\end{aligned}
$$

where $D_2$ is a difference matrix.

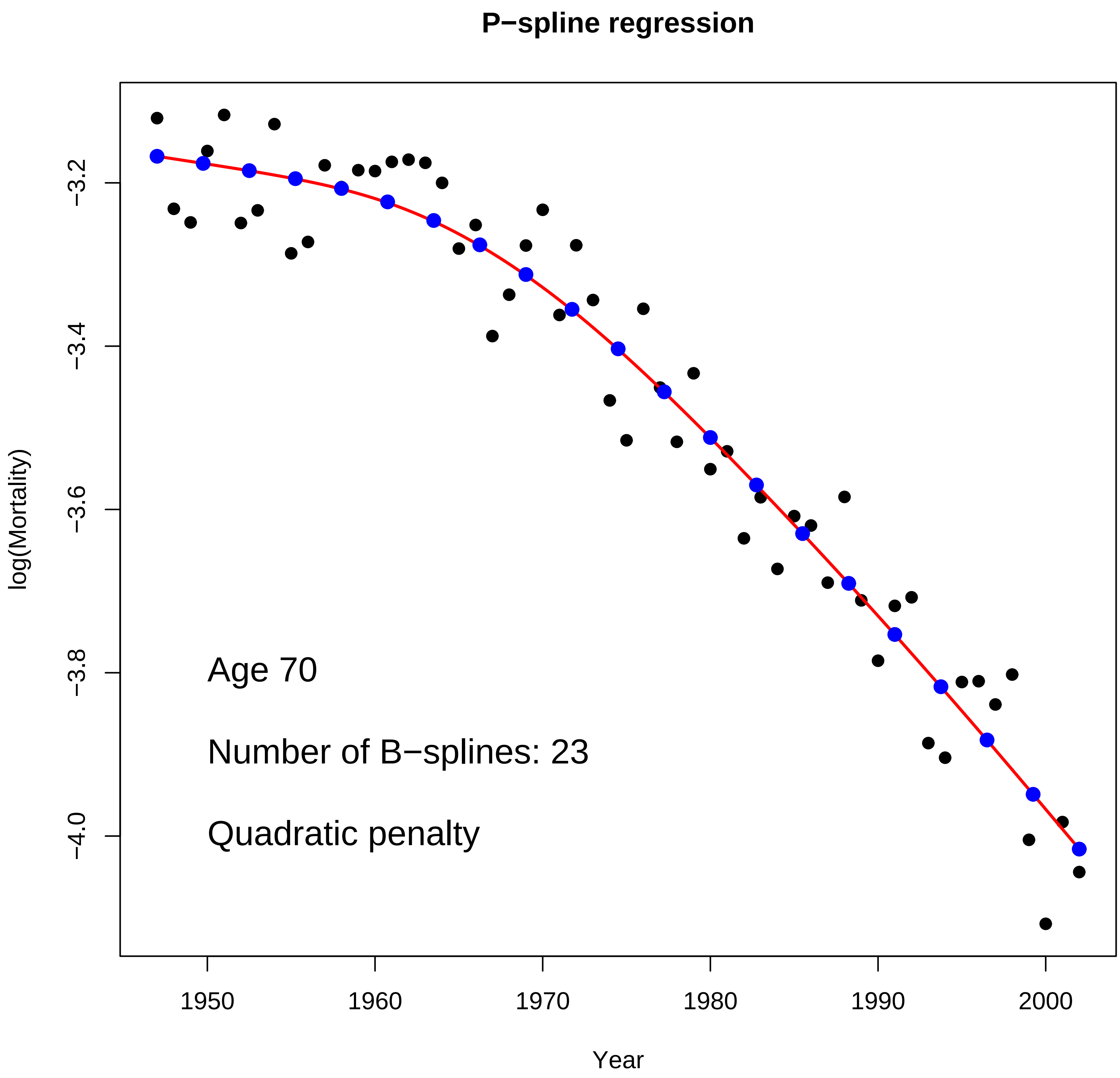
$P_2(\theta)$ is a roughness penalty.

Estimation is via penalised likelihood

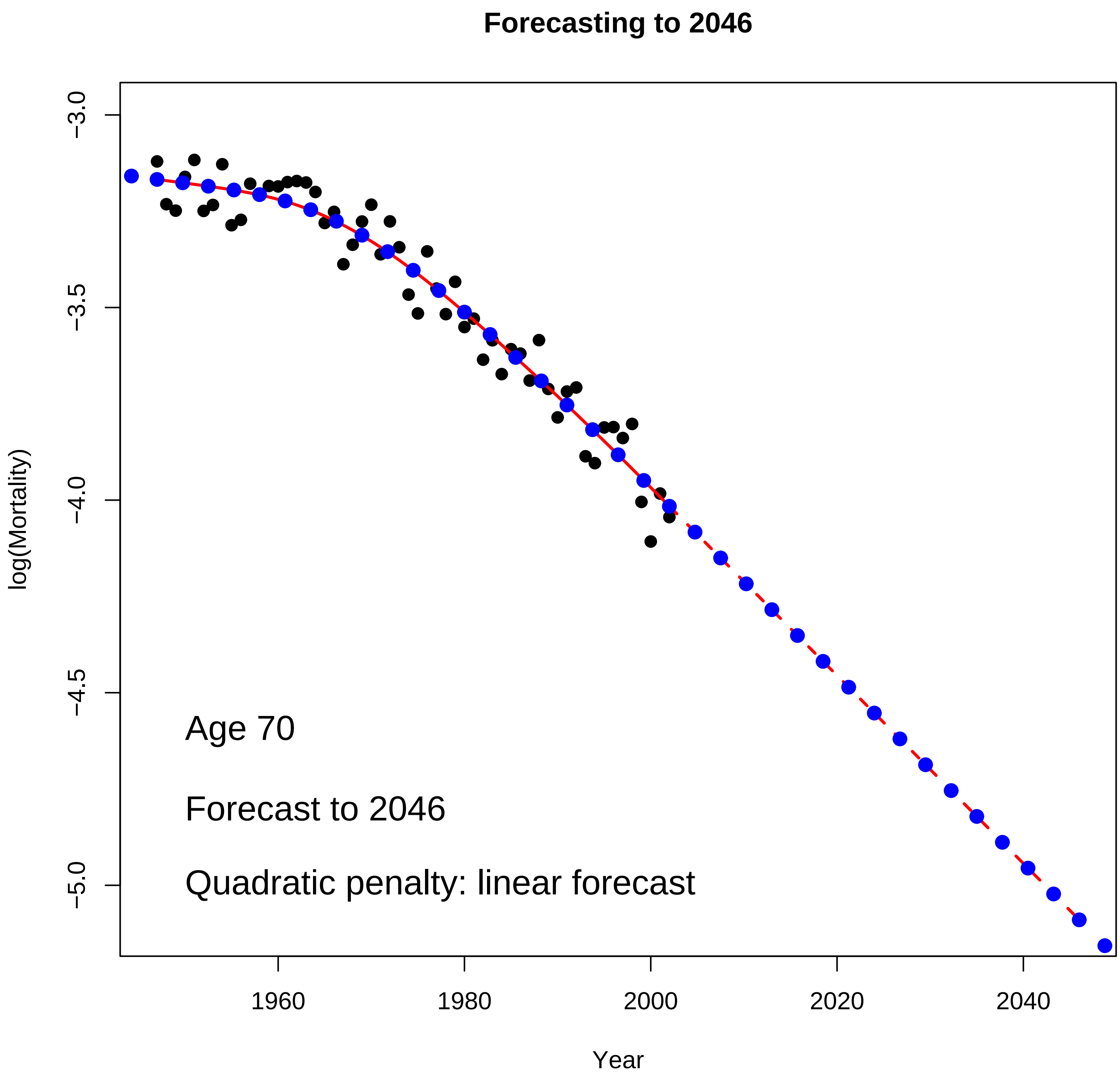
$$PL(\theta) = L(\theta) - \tfrac{1}{2}\lambda\theta' D_2' D_2 \theta$$

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

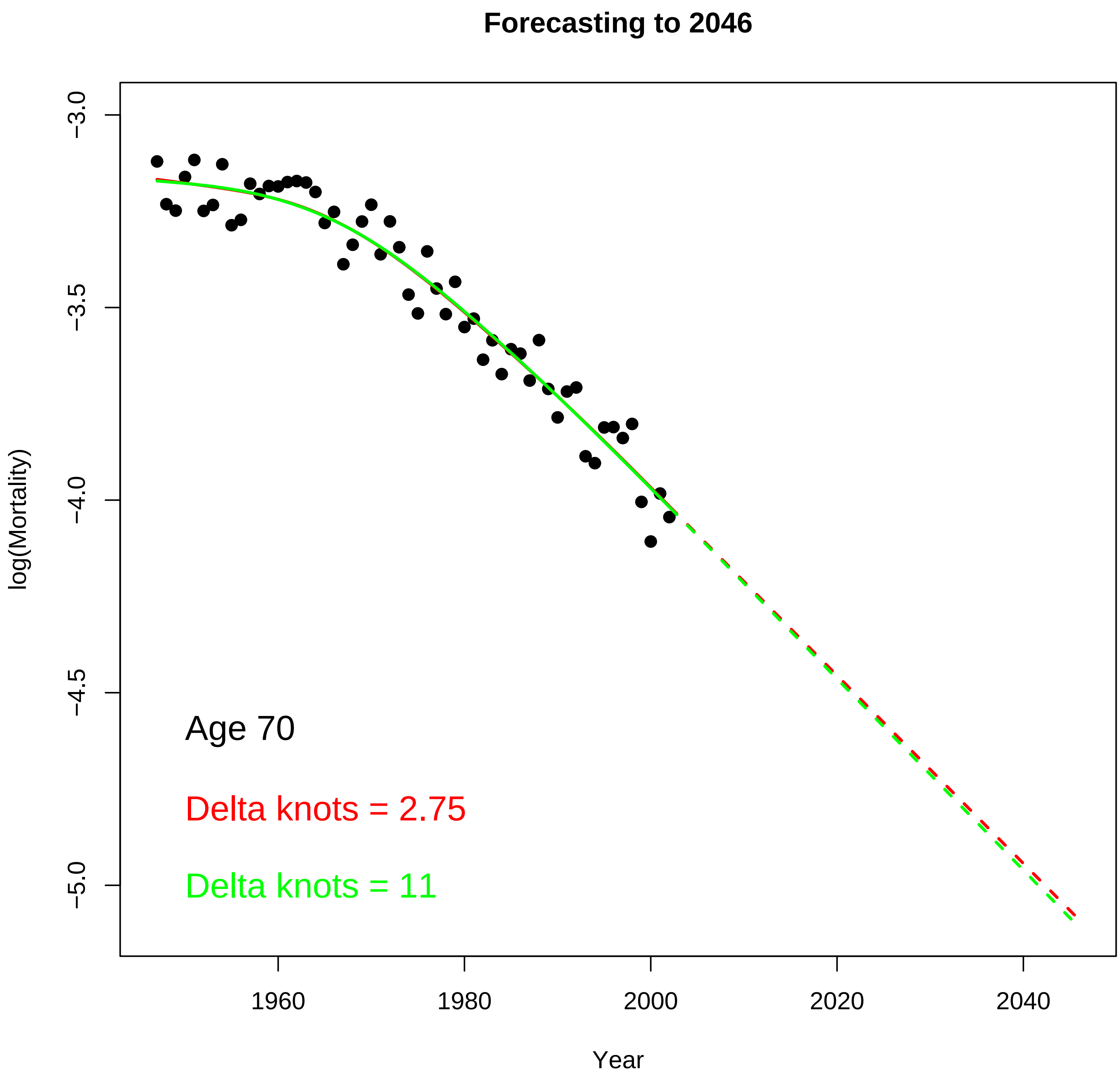
- Interpolation ($\lambda = 0$)
- Linear regression ($\lambda = \infty$)



P−spline regression
log(Mortality)
−3.2
−3.4
−3.6
−3.8
−4.0
1950
1960
1970
1980
1990
2000
Year
Age 70
Number of B−splines: 23
Quadratic penalty



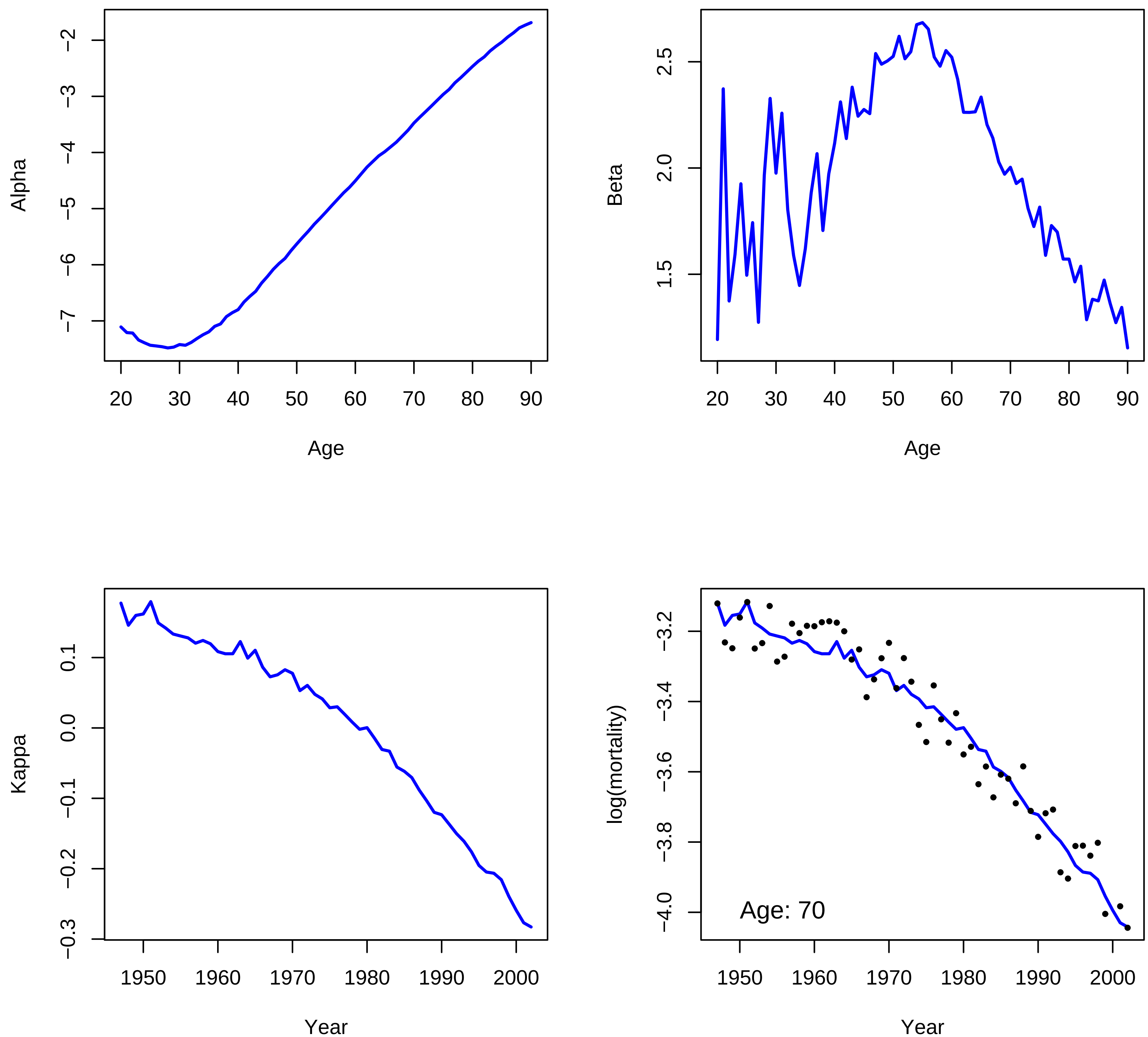
Forecasting to 2046
log(Mortality)
−3.0
−3.5
−4.0
−4.5
−5.0
1960
1980
2000
2020
2040
Year
Age 70
Forecast to 2046
Quadratic penalty: linear forecast

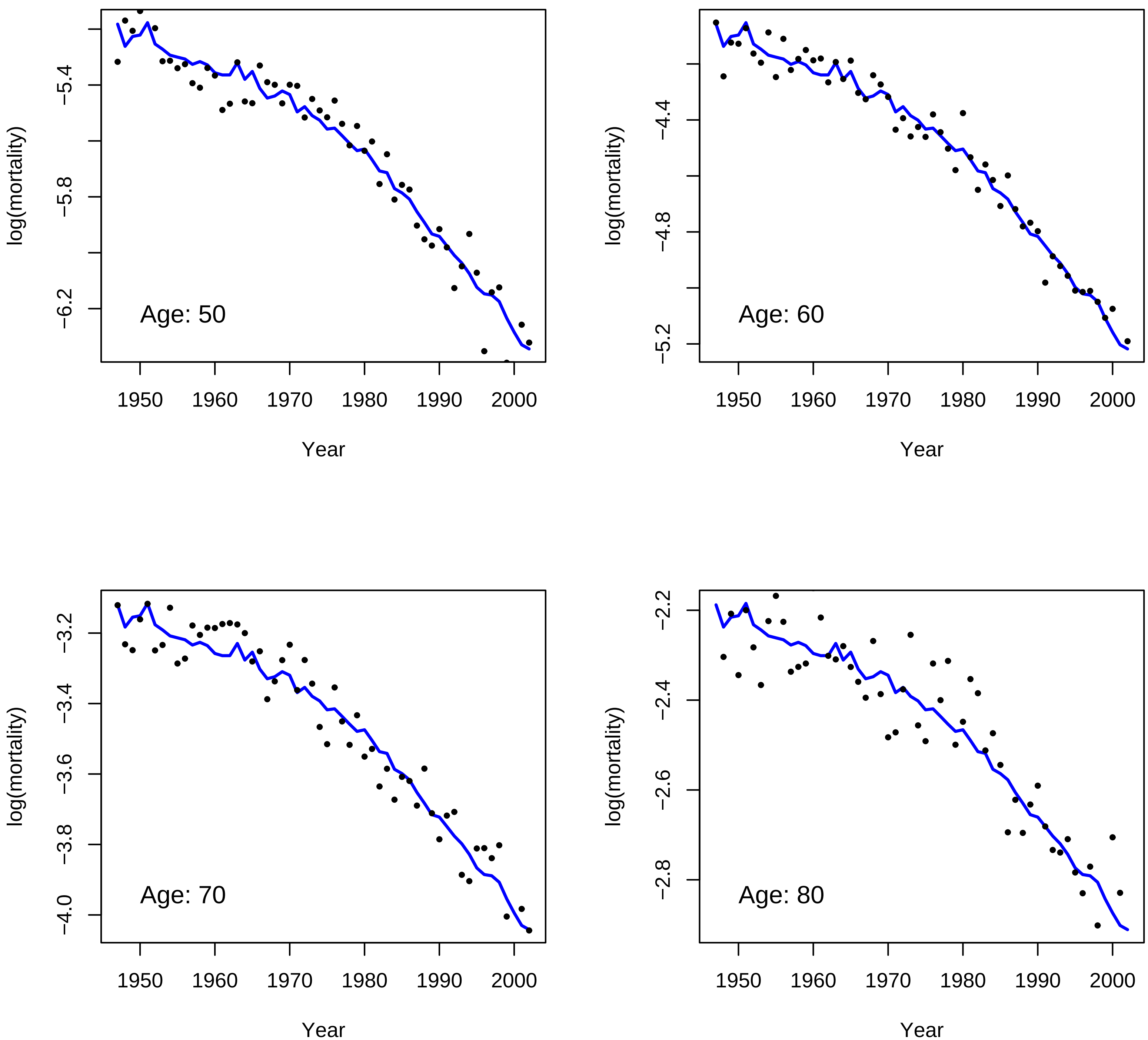

Forecasting to 2046
log(Mortality)
−3.0
−3.5
−4.0
−4.5
−5.0
1960
1980
2000
2020
2040
Year
Age 70
Delta knots = 2.75
Delta knots = 11

# Lee-Carter model

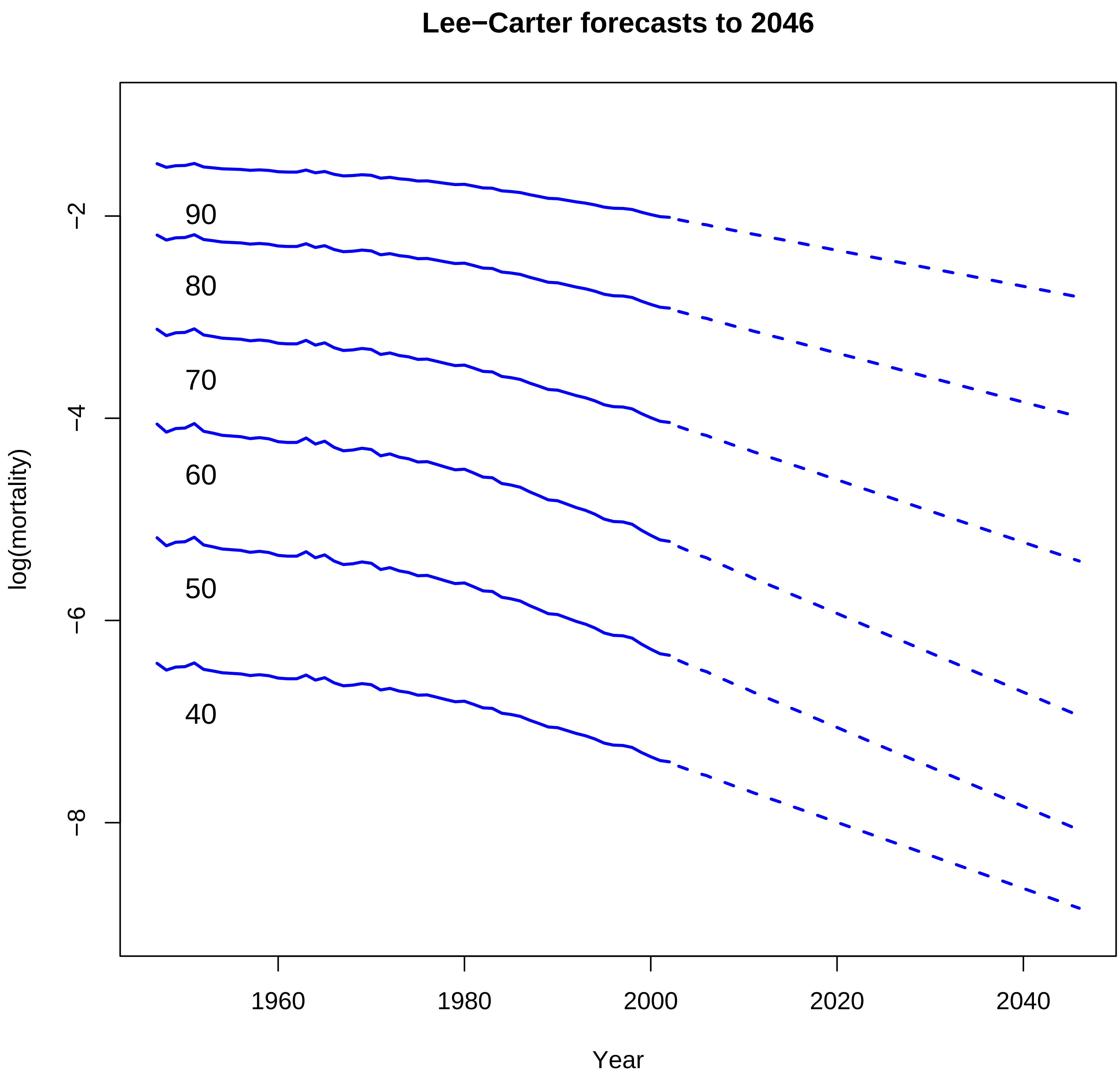
Lee & Carter (1992)

$$\log \mu_{ij} = \alpha_i + \beta_i \kappa_j, \quad 20 \le i \le 90,\ 1947 \le j \le 2002$$


Alpha
−2
−3
−4
−5
−6
−7
20 30 40 50 60 70 80 90
Age
Beta
2.5
2.0
1.5
20 30 40 50 60 70 80 90
Age
Kappa
0.1
0.0
−0.1
−0.2
−0.3
1950 1960 1970 1980 1990 2000
Year
log(mortality)
−3.2
−3.4
−3.6
−3.8
−4.0
Age: 70
1950 1960 1970 1980 1990 2000
Year



Age: 50
Age: 60
Age: 70
Age: 80
log(mortality)
Year
−5.4
−5.8
−6.2
−4.4
−4.8
−5.2
−3.2
−3.4
−3.6
−3.8
−4.0
−2.2
−2.4
−2.6
−2.8
1950
1960
1970
1980
1990
2000



Lee−Carter forecasts to 2046
90
80
70
60
50
40
−2
−4
−6
−8
log(mortality)
1960
1980
2000
2020
2040
Year

# 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

$$\log \boldsymbol{M} = \boldsymbol{\alpha} \mathbf{1}' + \boldsymbol{\beta} \boldsymbol{\kappa}'$$

# Discrete Lee-Carter model

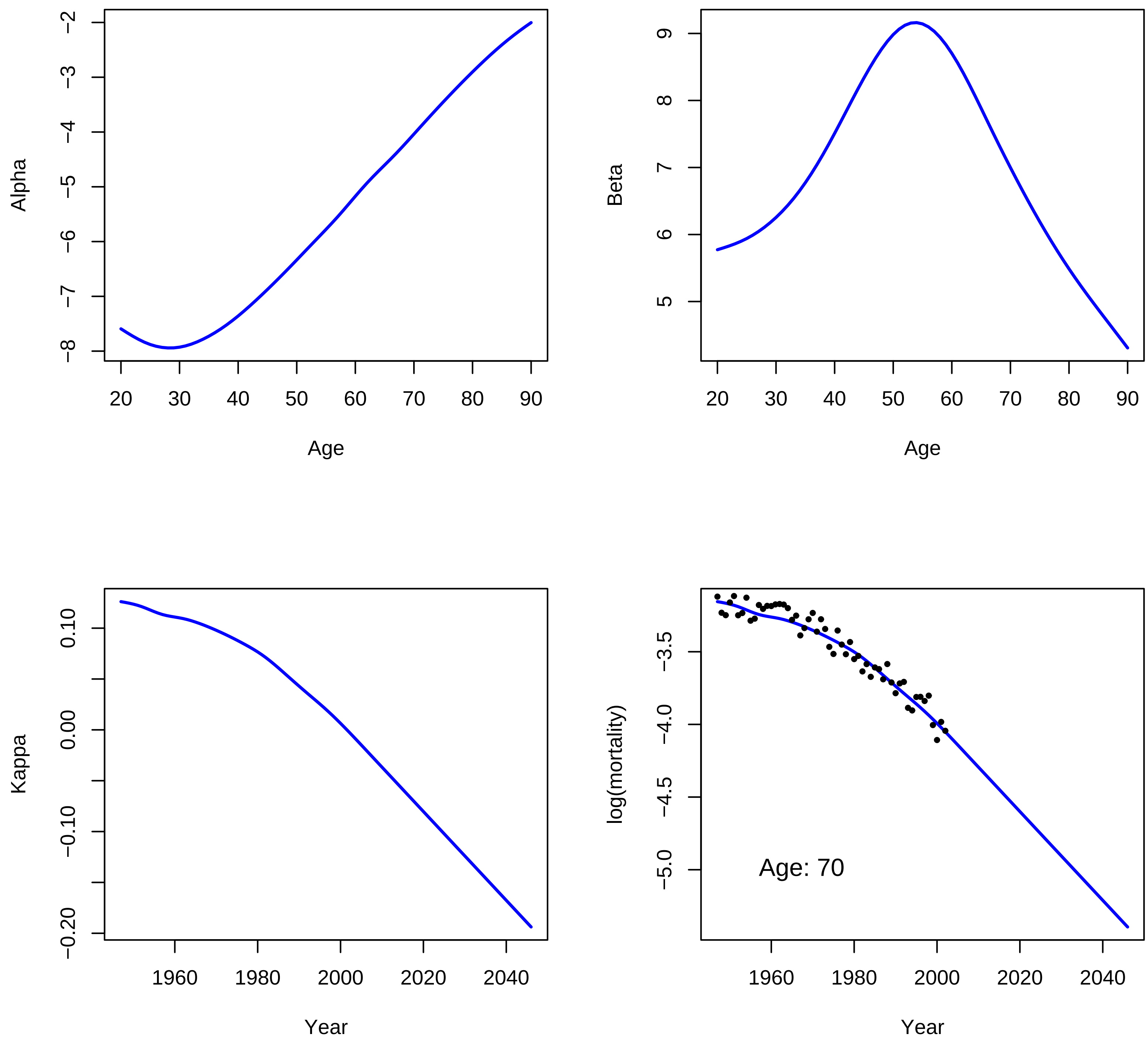
$$\log \mu_{ij} = \alpha_i + \beta_i \kappa_j, \quad 20 \le i \le 90,\ 1947 \le j \le 2002$$

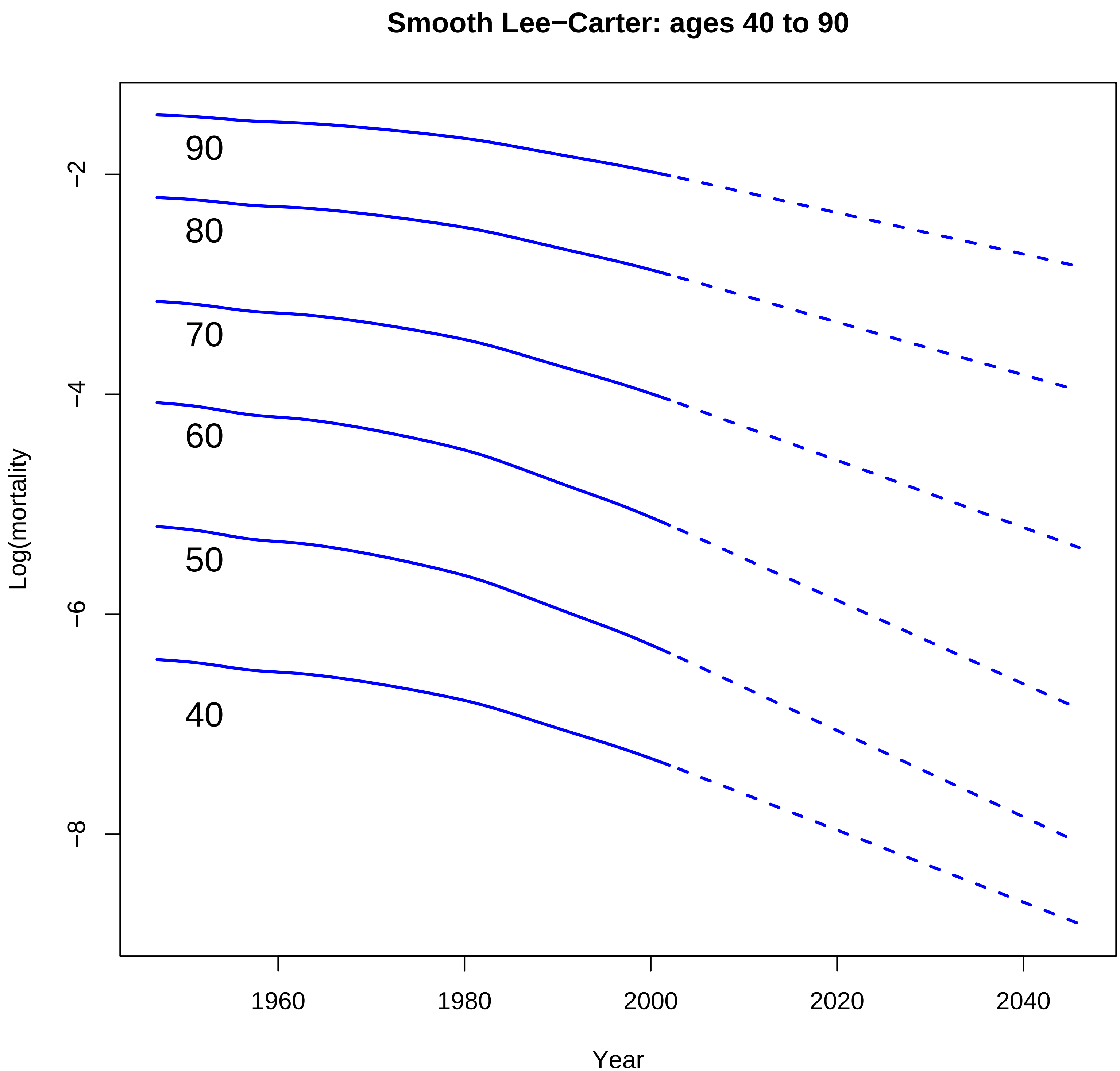
or in table form

$$\log \boldsymbol{M} = \boldsymbol{\alpha} \mathbf{1}' + \boldsymbol{\beta} \boldsymbol{\kappa}'$$

# Smooth Lee-Carter

$$\boldsymbol{\alpha} \rightarrow \boldsymbol{B}_a \boldsymbol{a}, \quad \boldsymbol{\beta} \rightarrow \boldsymbol{B}_a \boldsymbol{b}, \quad \boldsymbol{\kappa} \rightarrow \boldsymbol{B}_y \boldsymbol{k}$$


Alpha
−2
−3
−4
−5
−6
−7
−8
20 30 40 50 60 70 80 90
Age
Beta
9
8
7
6
5
20 30 40 50 60 70 80 90
Age
Kappa
0.10
0.00
−0.10
−0.20
1960 1980 2000 2020 2040
Year
log(mortality)
−3.5
−4.0
−4.5
−5.0
Age: 70
1960 1980 2000 2020 2040
Year



Smooth Lee−Carter: ages 40 to 90
90
80
70
60
50
40
−2
−4
−6
−8
Log(mortality)
1960
1980
2000
2020
2040
Year

# Age-Period-Cohort model

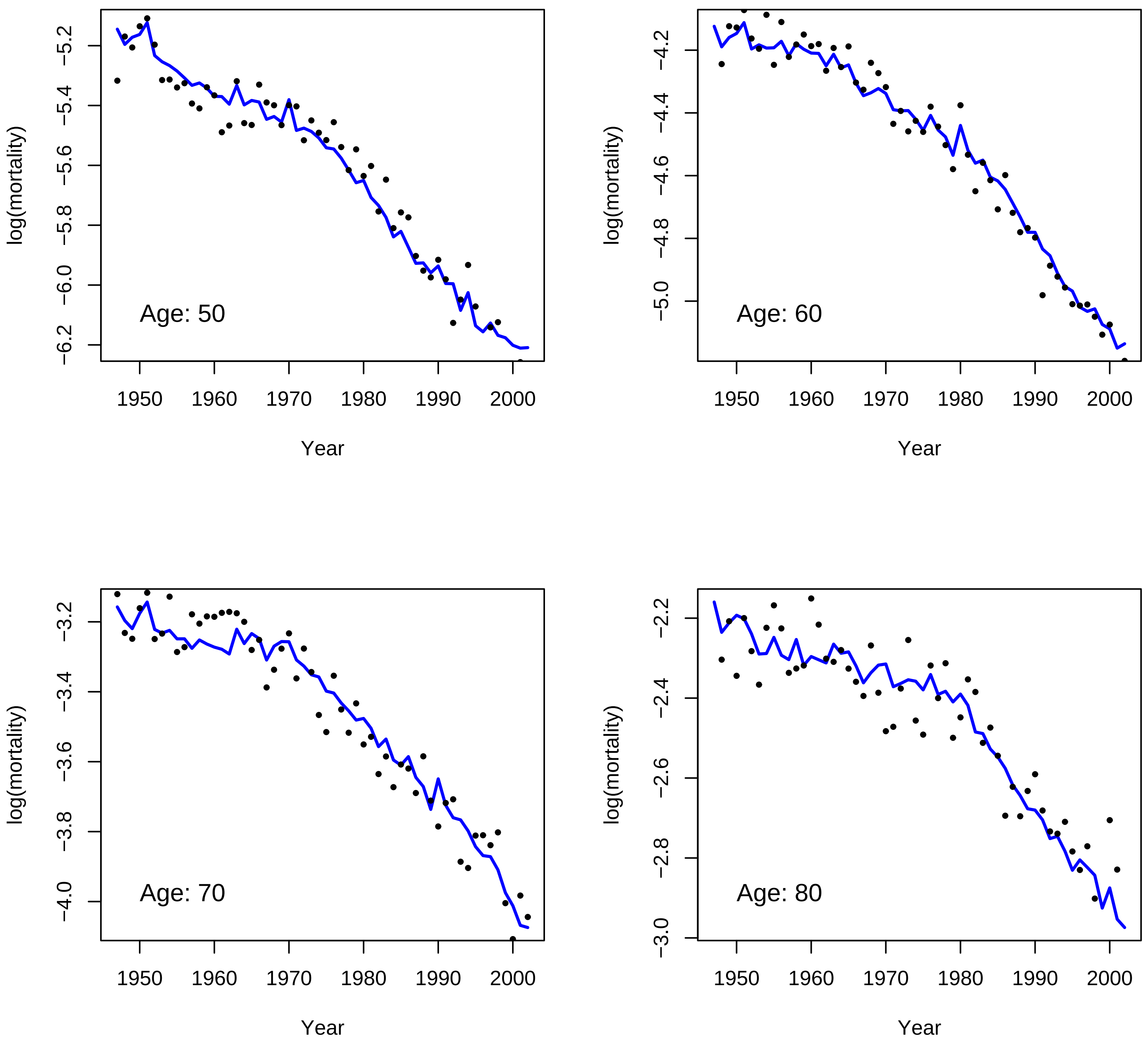
$$\log \mu_{ij} = \alpha_i + \beta_j + \gamma_{i-j}, \quad 20 \leq i \leq 90,\ 1947 \leq j \leq 2002$$

# Age-Period-Cohort model

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

# Parameter redundancy

- $2n_a + 2n_y - 1 = 253$ parameters
- $2n_a + 2n_y - 4 = 250$ free parameters


Age: 50
Age: 60
Age: 70
Age: 80
log(mortality)
Year
1950
1960
1970
1980
1990
2000
−5.2
−5.4
−5.6
−5.8
−6.0
−6.2
−4.2
−4.4
−4.6
−4.8
−5.0
−3.2
−3.4
−3.6
−3.8
−4.0
−2.2
−2.4
−2.6
−2.8
−3.0

# Discrete Age-Period-Cohort model

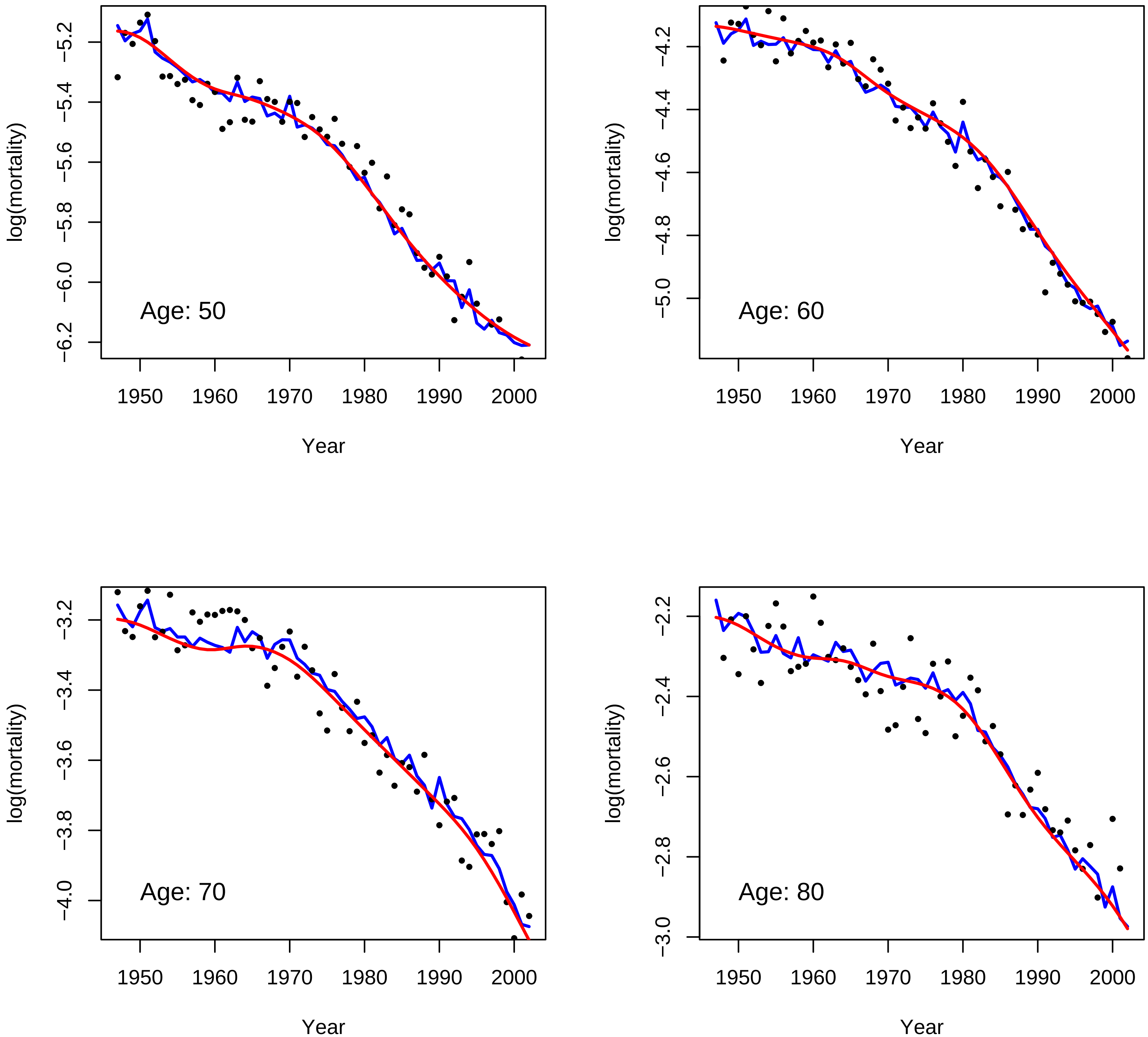
$$\log \mu_{ij} = \alpha_i + \beta_j + \gamma_{i-j}, \quad 20 \leq i \leq 90,\ 1947 \leq j \leq 2002$$

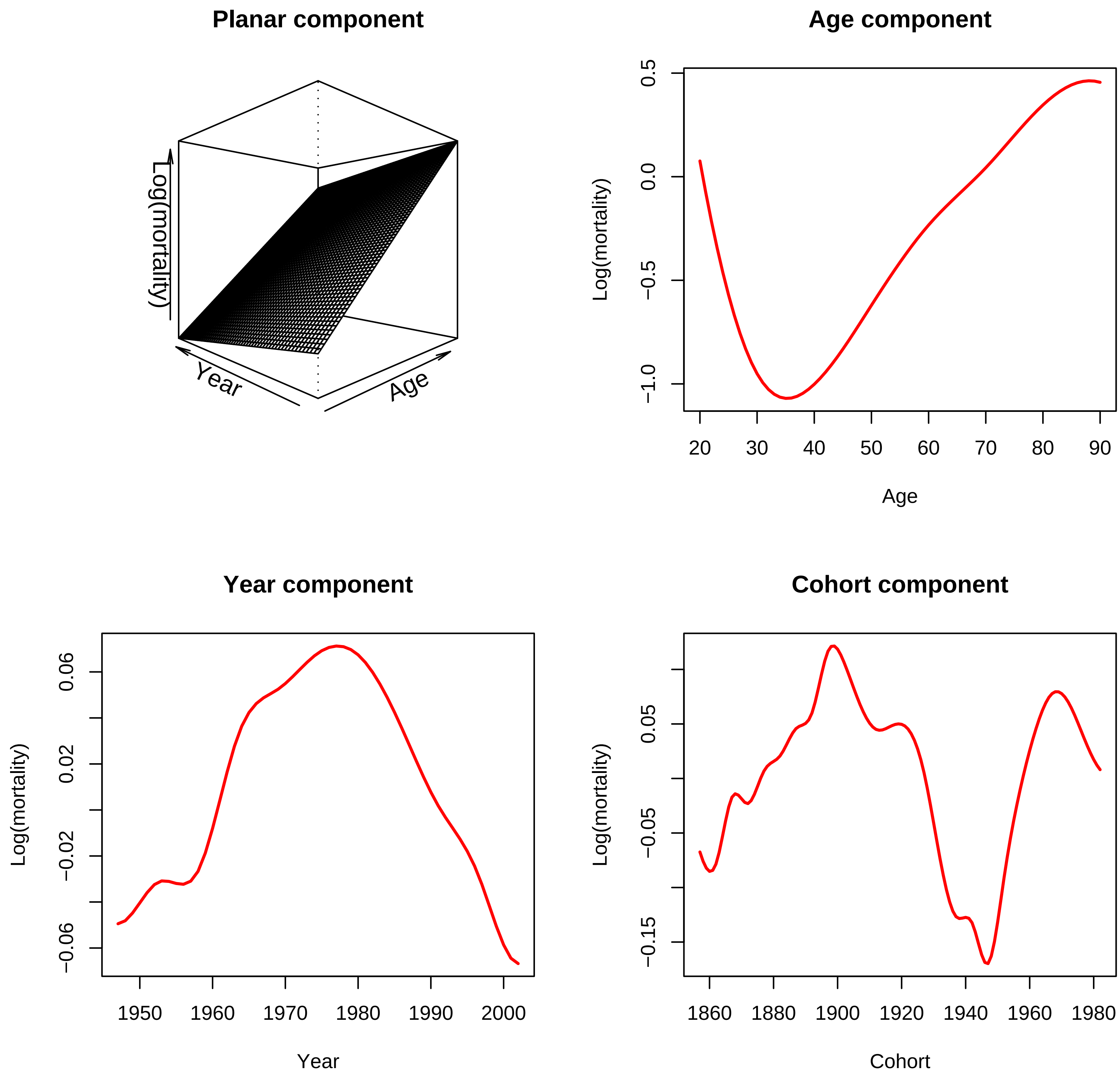
# Discrete Age-Period-Cohort model

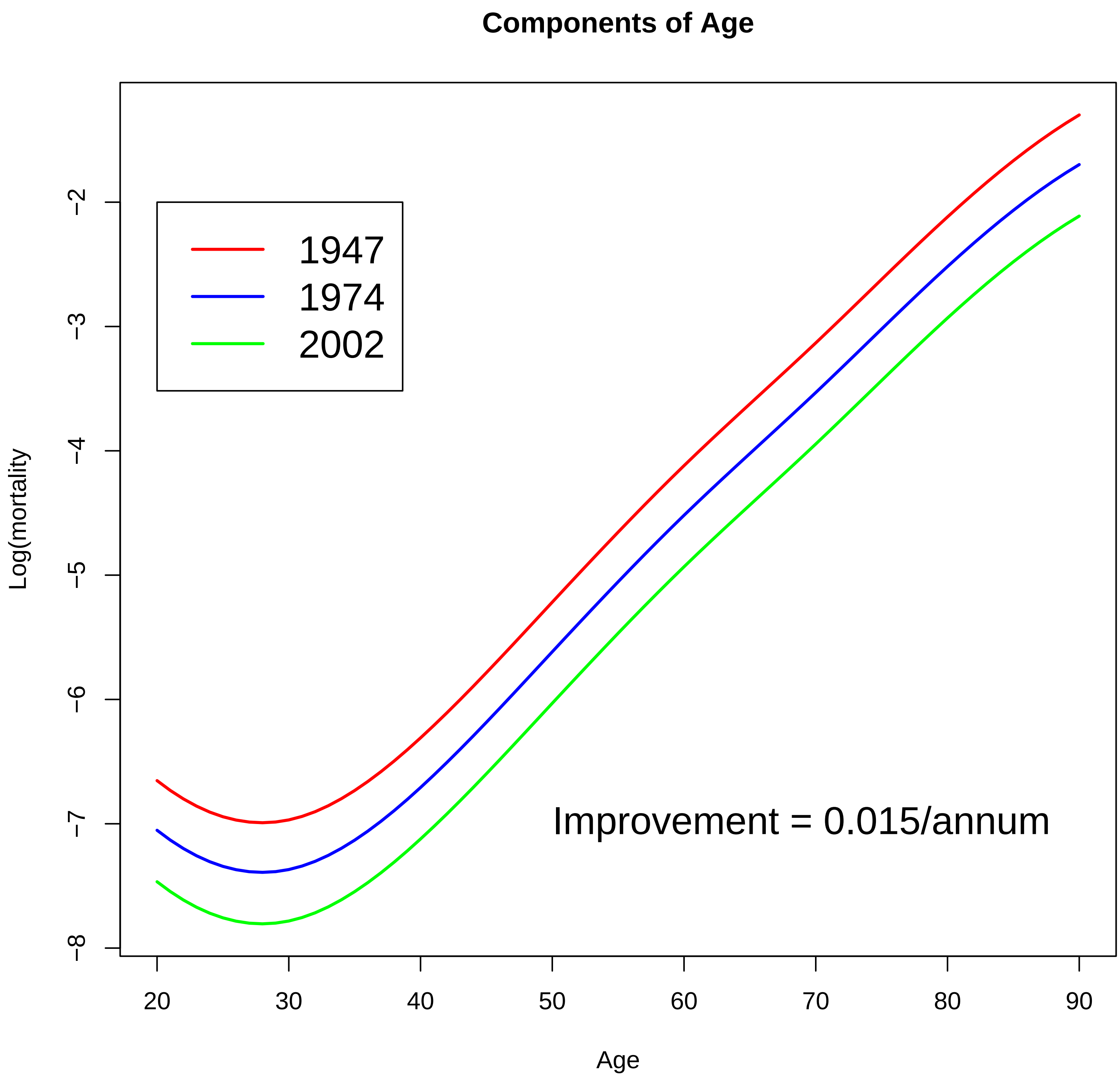
$$\log \mu_{ij} = \alpha_i + \beta_j + \gamma_{i-j}, \quad 20 \leq i \leq 90,\ 1947 \leq j \leq 2002$$

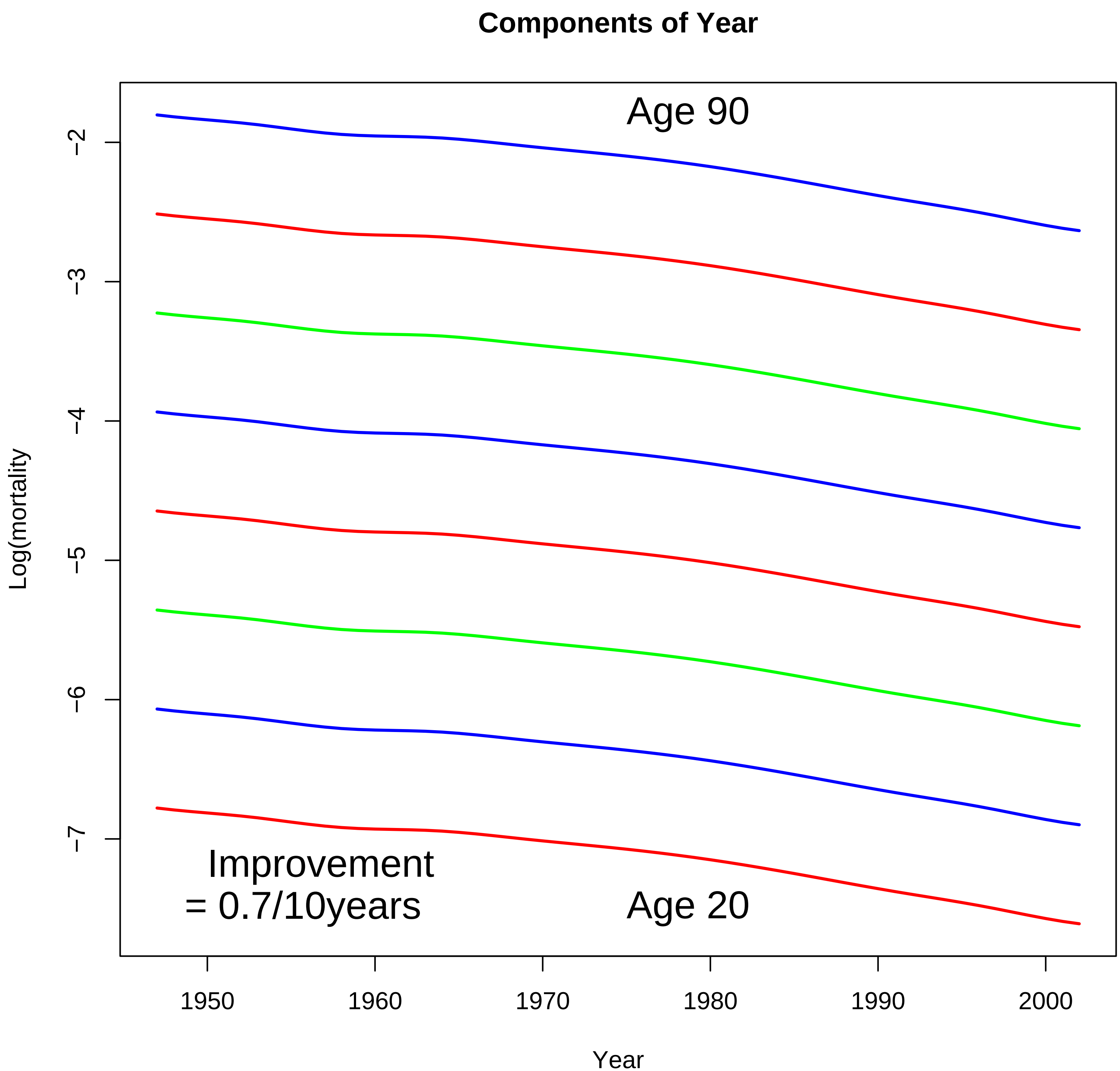
# Smooth Age-Period-Cohort model

$$\boldsymbol{\alpha} \to \boldsymbol{B}_a \boldsymbol{a}, \quad \boldsymbol{\beta} \to \boldsymbol{B}_y \boldsymbol{b}, \quad \boldsymbol{\gamma} \to \boldsymbol{B}_c \boldsymbol{c}$$

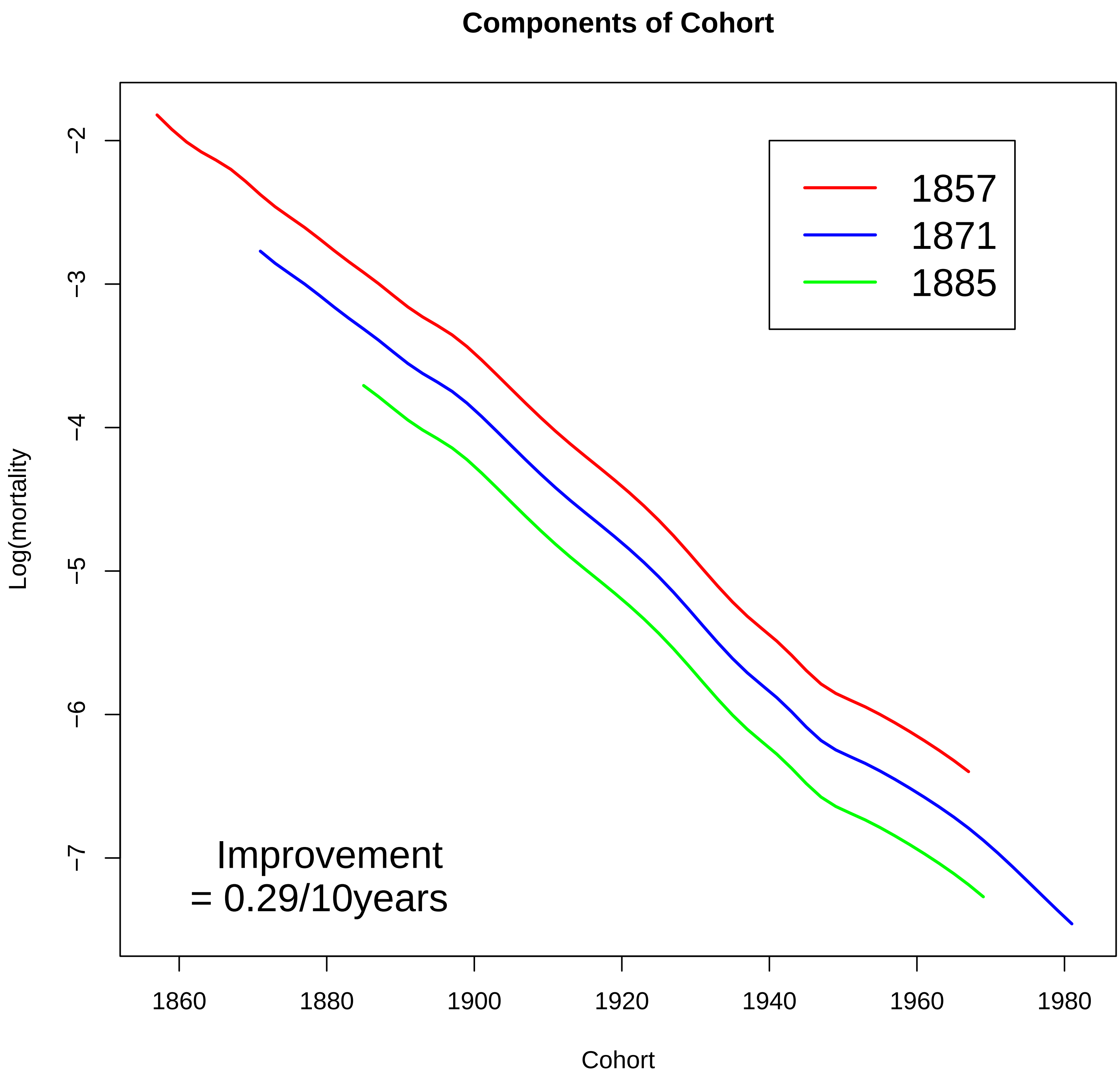

Age: 50
Age: 60
Age: 70
Age: 80
log(mortality)
Year
1950
1960
1970
1980
1990
2000
−5.2
−5.4
−5.6
−5.8
−6.0
−6.2
−4.2
−4.4
−4.6
−4.8
−5.0
−3.2
−3.4
−3.6
−3.8
−4.0
−2.2
−2.4
−2.6
−2.8
−3.0


Planar component
Log(mortality)
Year
Age
Age component
Log(mortality)
0.5
0.0
−0.5
−1.0
20
30
40
50
60
70
80
90
Age
Year component
Log(mortality)
0.06
0.02
−0.02
−0.06
1950
1960
1970
1980
1990
2000
Year
Cohort component
Log(mortality)
0.05
−0.05
−0.15
1860
1880
1900
1920
1940
1960
1980
Cohort


Components of Age
1947
1974
2002
Improvement = 0.015/annum
Log(mortality
−2
−3
−4
−5
−6
−7
−8
20
30
40
50
60
70
80
90
Age



Components of Year
Age 90
Improvement
= 0.7/10years
Age 20
Log(mortality
Year
−2
−3
−4
−5
−6
−7
1950
1960
1970
1980
1990
2000


Components of Cohort
Log(mortality)
−2
−3
−4
−5
−6
−7
1860
1880
1900
1920
1940
1960
1980
Cohort
1857
1871
1885
Improvement
= 0.29/10years

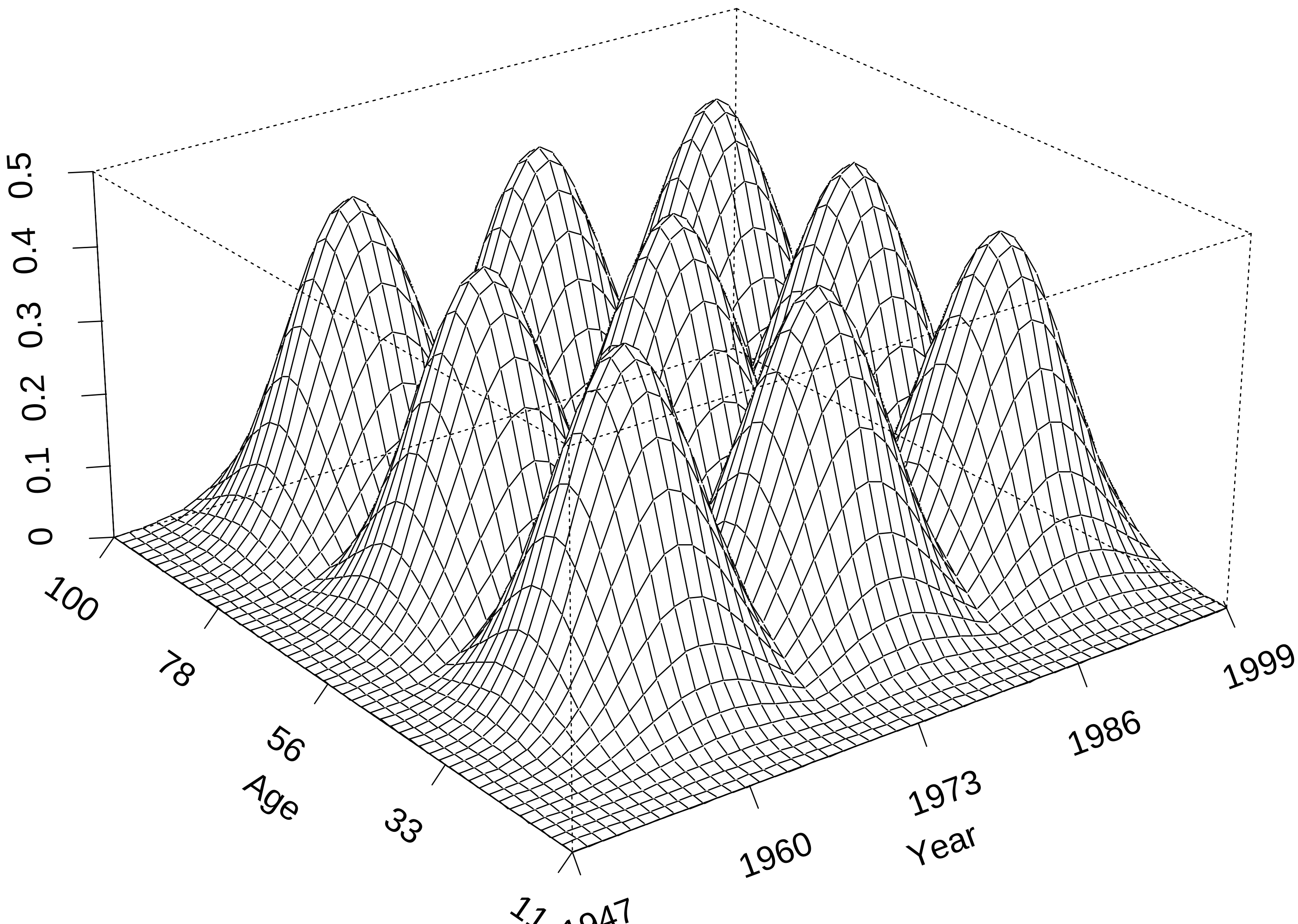
# 2-d modelling of mortality tables

- 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?



0.5
0.4
0.3
0.2
0.1
0
100
78
56
Age
33
11
1947
1960
1973
Year
1986
1999

# 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

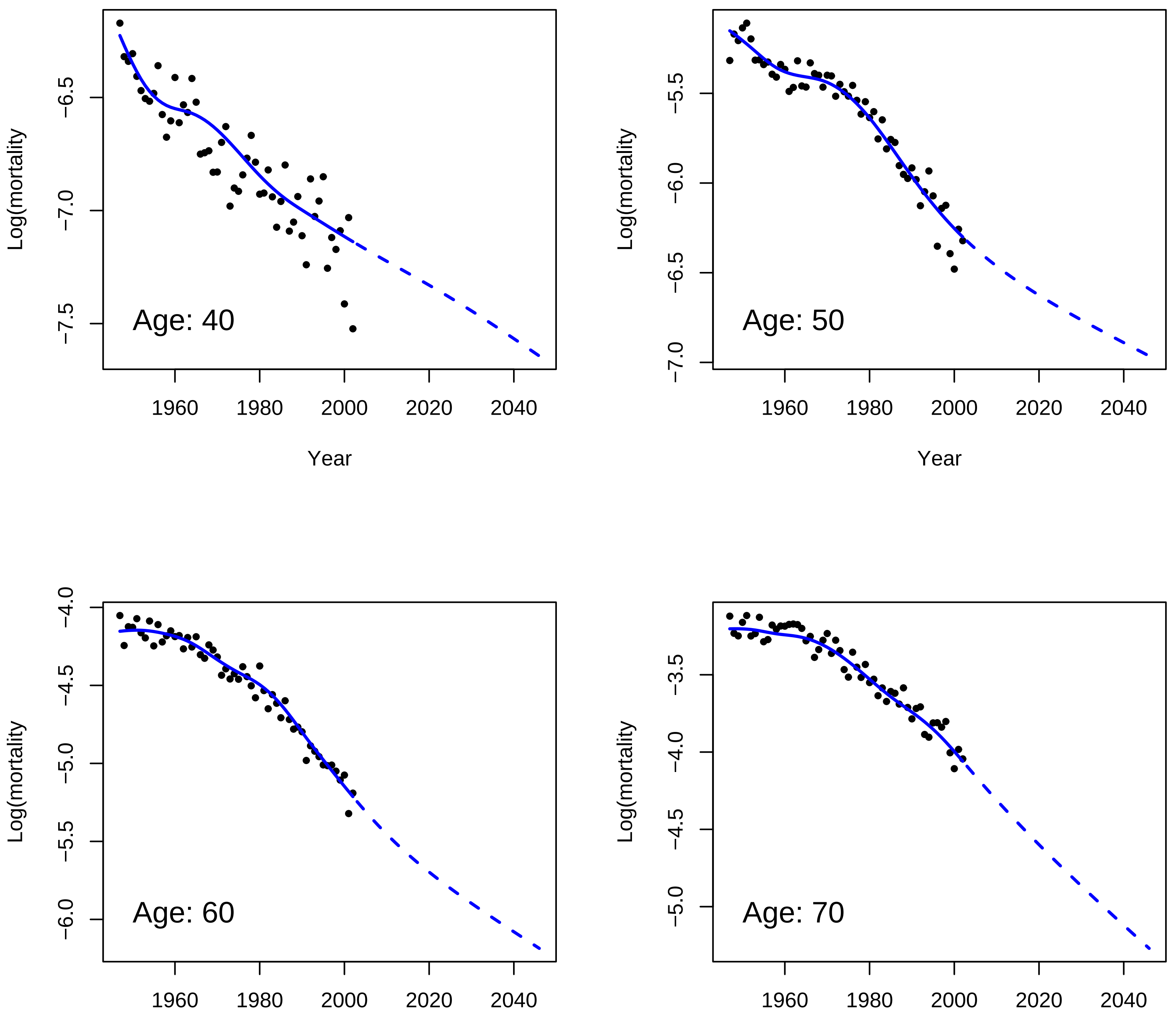
## Regression matrix

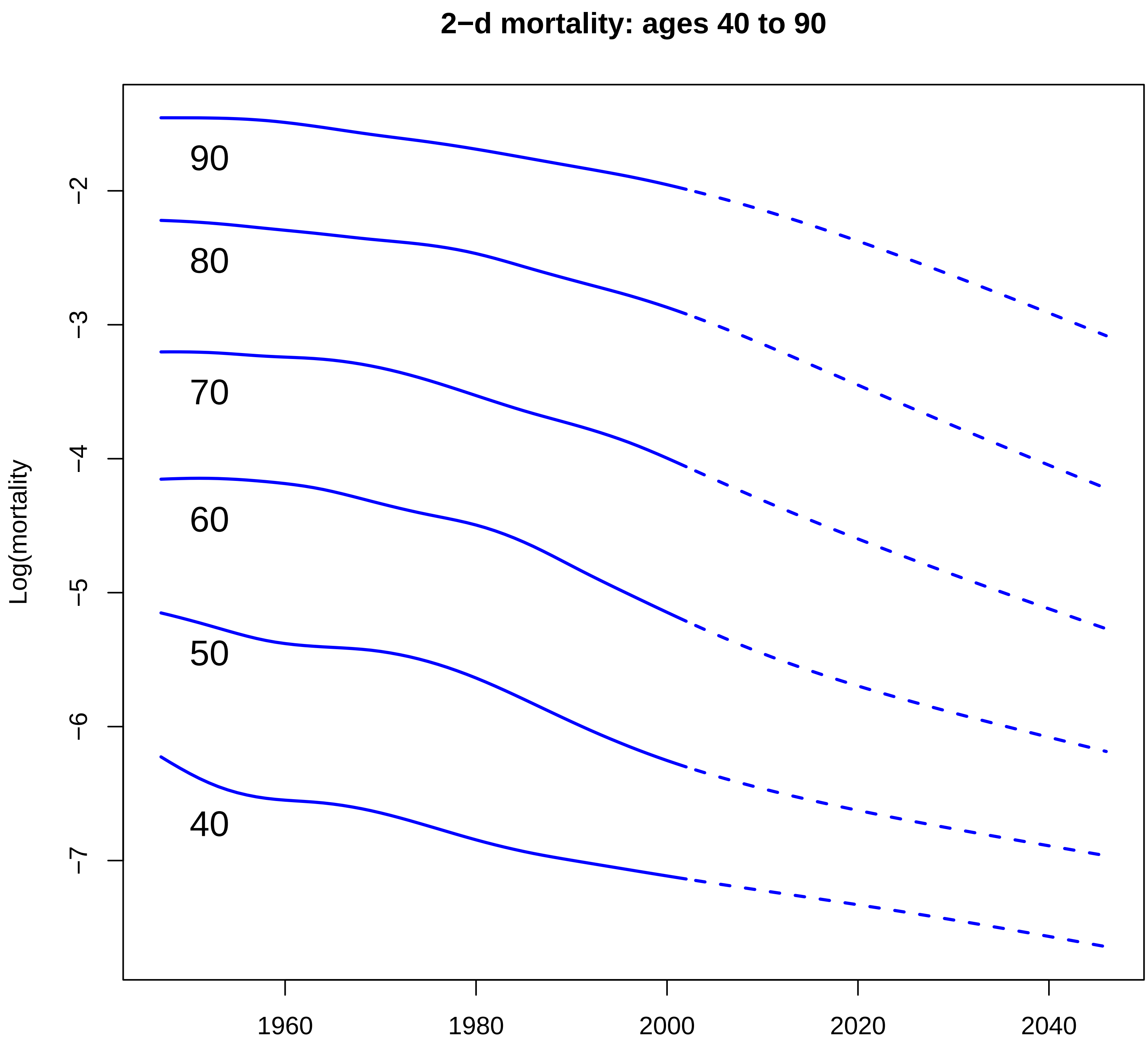
The Kronecker product organises the multiplication of the marginal bases to give the regression matrix

$$B = B_y \otimes B_a, \; 3976 \times 130.$$

## 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.

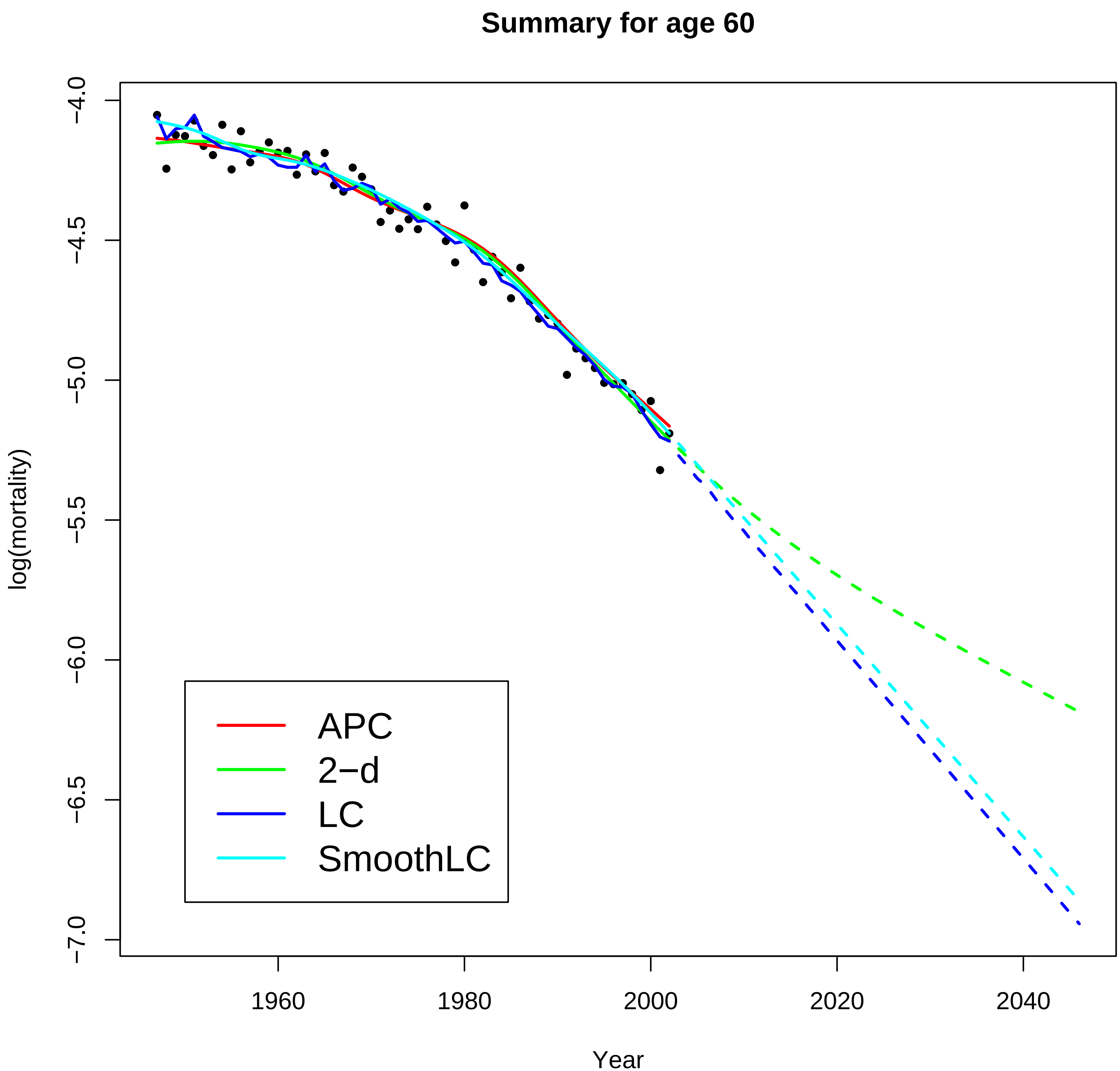

Age: 40
Age: 50
Age: 60
Age: 70
Log(mortality
Log(mortality
Log(mortality
Log(mortality
Year
Year
Year
Year
−6.5
−7.0
−7.5
−5.5
−6.0
−6.5
−7.0
−4.0
−4.5
−5.0
−5.5
−6.0
−3.5
−4.0
−4.5
−5.0
1960
1980
2000
2020
2040



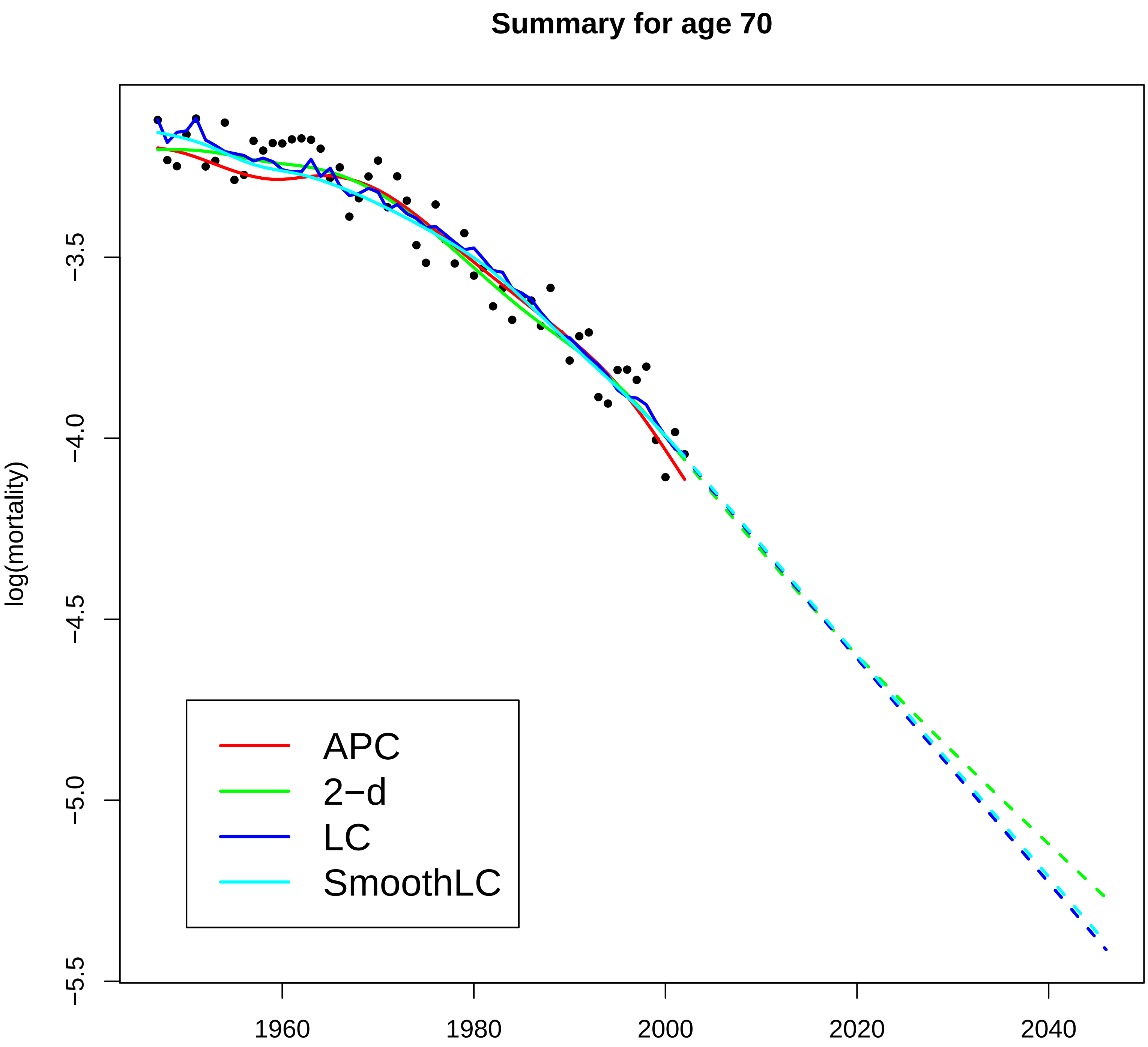
2−d mortality: ages 40 to 90
90
80
70
60
50
40
−2
−3
−4
−5
−6
−7
Log(mortality
1960
1980
2000
2020
2040
Year

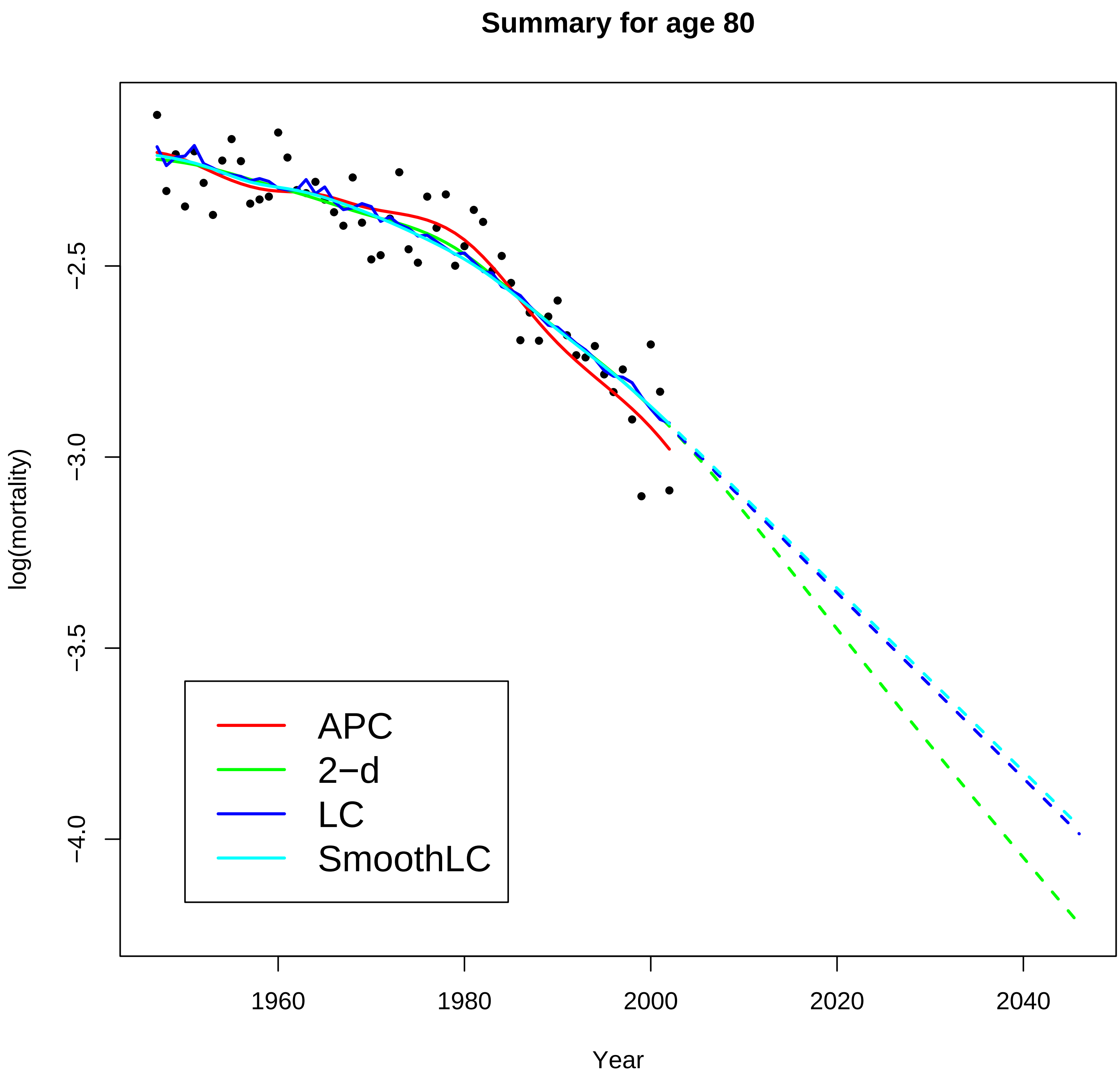
## Summary of results

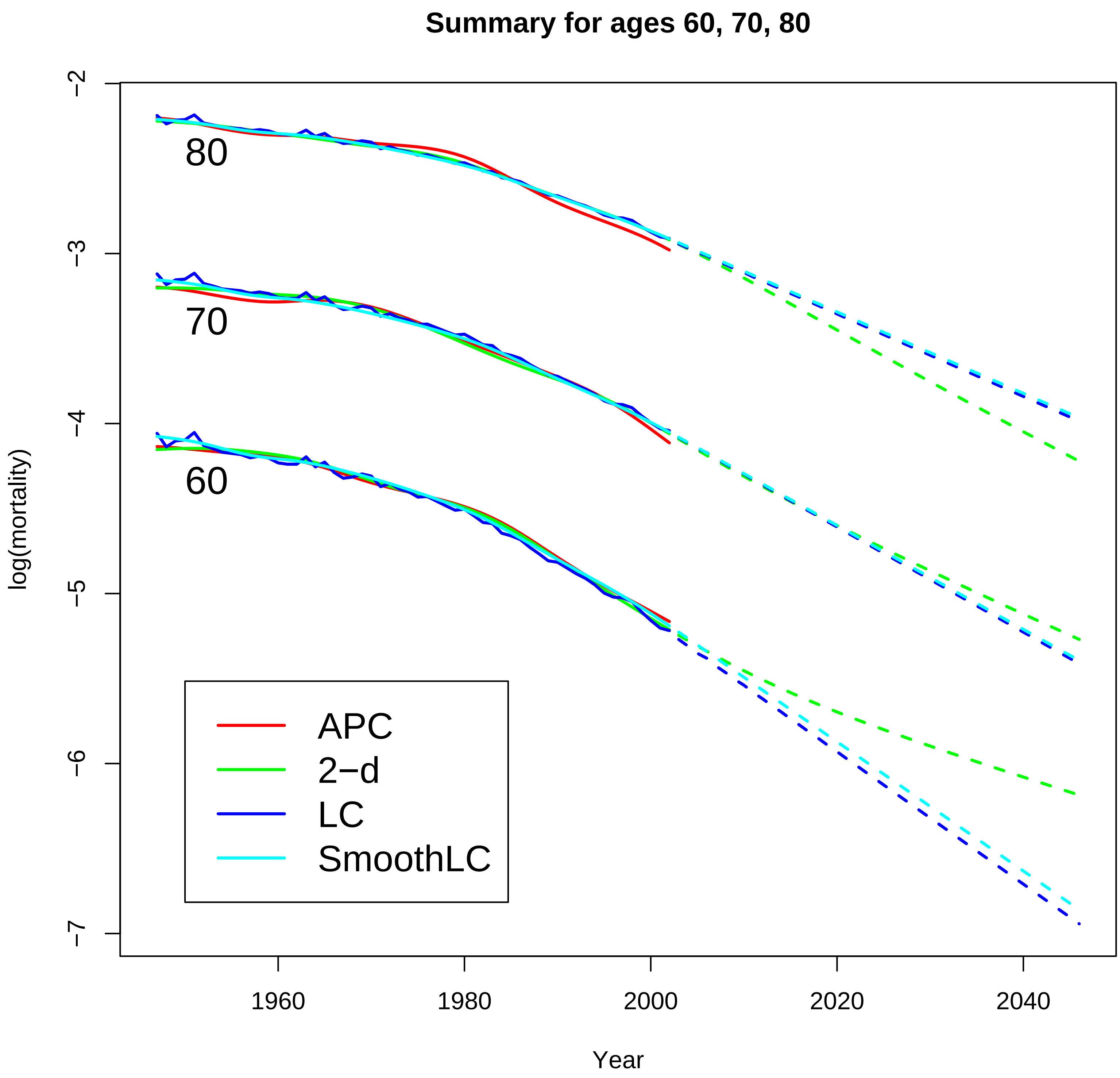
| | DEV | TR | BIC | Rank |
|---|---|---|---|---|
| 2-d | 6832 | 63 | 7357 | 1 |
| APC smooth | 7852 | 38 | 8167 | 2 |
| LC smooth | 8568 | 30 | 8815 | 3 |
| APC | 7002 | 248 | 9057 | 4 |
| LC | 8004 | 196 | 9629 | 5 |



Summary for age 60
log(mortality)
−4.0
−4.5
−5.0
−5.5
−6.0
−6.5
−7.0
1960
1980
2000
2020
2040
Year
APC
2−d
LC
SmoothLC


Summary for age 70
log(mortality)
−3.5
−4.0
−4.5
−5.0
−5.5
1960
1980
2000
2020
2040
Year
APC
2−d
LC
SmoothLC


Summary for age 80
log(mortality)
−2.5
−3.0
−3.5
−4.0
1960
1980
2000
2020
2040
Year
APC
2−d
LC
SmoothLC



Summary for ages 60, 70, 80
log(mortality)
−2
−3
−4
−5
−6
−7
80
70
60
APC
2−d
LC
SmoothLC
1960
1980
2000
2020
2040
Year

# 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

Lee & Carter (1992) J American Statistical Association, 87, 659-675.

Brouhns, Denuit & Vermunt (2002) Insurance: Mathematics & Economics, 31, 373-393.

## Age-Period-Cohort models

Clayton & Schlifflers (1987) Statistics in Medicine, 6, 449-467.

Clayton & Schlifflers (1987) Statistics in Medicine, 6, 469-481.

## Penalized spline models

Eilers & Marx (1996) Statistical Science, 11, 758-783.

Currie, Durban & Eilers (2004) Statistical Modelling, 4, 279-298.

Currie, Durban & Eilers (2006) Journal of the Royal Statistical Society, Series B, 68, 259-280.