Lecture 3:
Forecasting in a smoothing context

Giancarlo Camarda
Institut national d'études démographiques

Ugofilippo Basellini
Max Planck Institute for Demographic Research

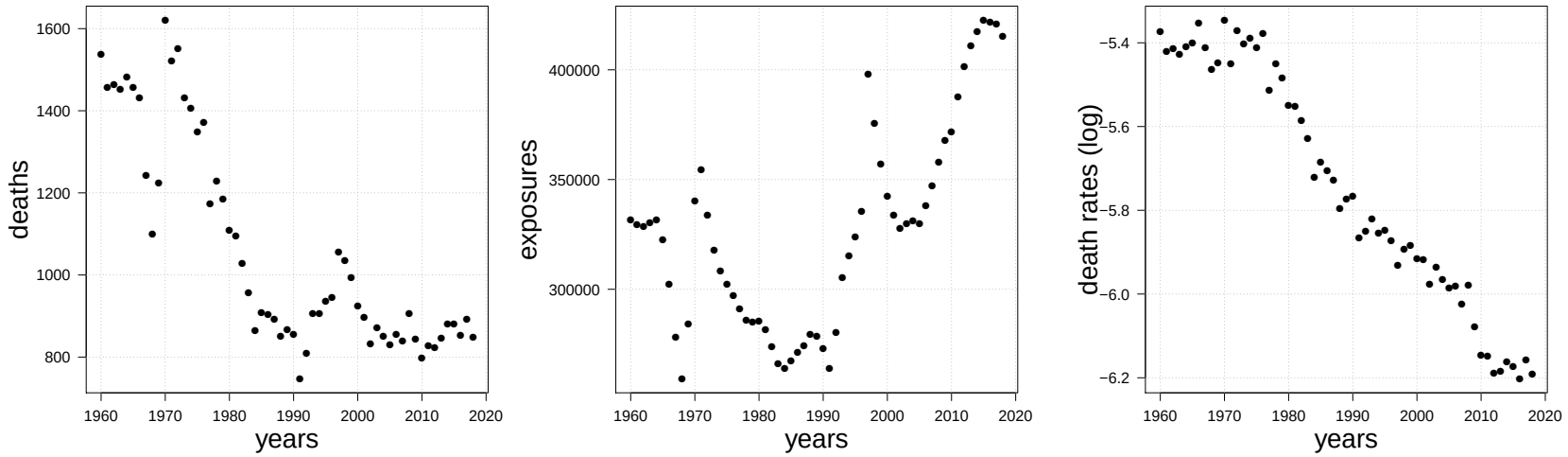
IDEM 117
Advances in Mortality Forecasting

International Advanced Studies in Demography
17 – 21 January, 2022

Introduction

- In the mortality forecasting literature, there are few exceptions with respect to the Lee-Carter model and its numerous variants
- A notable one is the approach proposed by Currie at al. (2004):
  - generalization of the P-splines in two dimensions (age and time)
  - treat future years as not available data
- Model and forecast is achieved simultaneously
- We can borrow all GLM toolbox

Starting from a 1D example



Deaths, exposures and rates (in log scale) over years.
England & Wales, age 50, years 1960-2018.
Source: Human Mortality Database.

- Remainder:

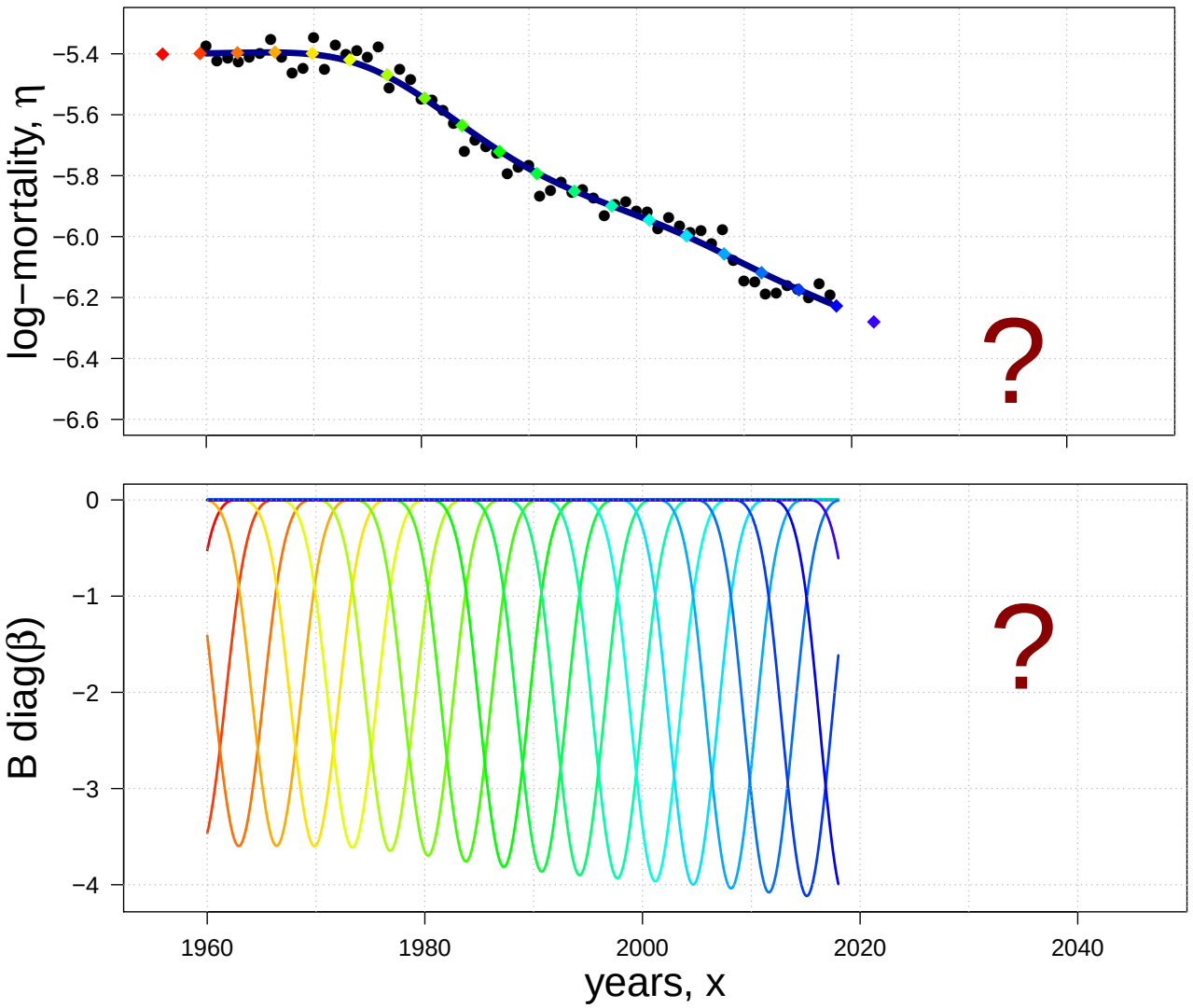
$$\text{data:}\quad \boldsymbol{y} = (y_i) \text{ (deaths) and } \boldsymbol{e} = (e_i) \text{ (exposures)}$$

$$\text{log-mortality:}\quad \ln \boldsymbol{\mu} = \boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\beta}$$

$$\text{estimation:}\quad \tilde{\boldsymbol{\beta}} = (\boldsymbol{B}'\boldsymbol{W}\boldsymbol{B} + \boldsymbol{P})^{-1}\boldsymbol{B}'\boldsymbol{W}\boldsymbol{z}$$

with $\boldsymbol{W} = \texttt{diag}(w_i), w_i = e_i\,\mu_i$ and $z_i = \eta_i + \frac{y_i - e_i\,\mu_i}{e_i\,\mu_i}$

Fitting observed data



Extrapolating with P-splines

- Aim: extrapolate $\boldsymbol{\eta}$ (log-mortality) over years (e.g. up to 2040)
- Augment deaths and exposures with arbitrary values (e.g. 999)

$$\boldsymbol{y}^A = [y_{1960}, \ldots, y_{2018}, \breve{y}_{2019}, \ldots, \breve{y}_{2040}] \qquad \boldsymbol{e}^A = [e_{1960}, \ldots, e_{2018}, \breve{e}_{2019}, \ldots, \breve{e}_{2040}]$$

- We expand the B-spline basis to cover the whole period:

$$\boldsymbol{B}^A = \begin{bmatrix} B_1(1960) & B_2(1960) & \cdot & B_r(1960) & \cdot & B_k(1960) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ B_1(2018) & B_2(2018) & \cdot & B_r(2018) & \cdot & B_k(2018) \\ B_1(2019) & B_2(2019) & \cdot & B_r(2019) & \cdot & B_k(2019) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ B_1(2040) & B_2(2040) & \cdot & B_r(2040) & \cdot & B_k(2040) \end{bmatrix}$$

- New penalty term:

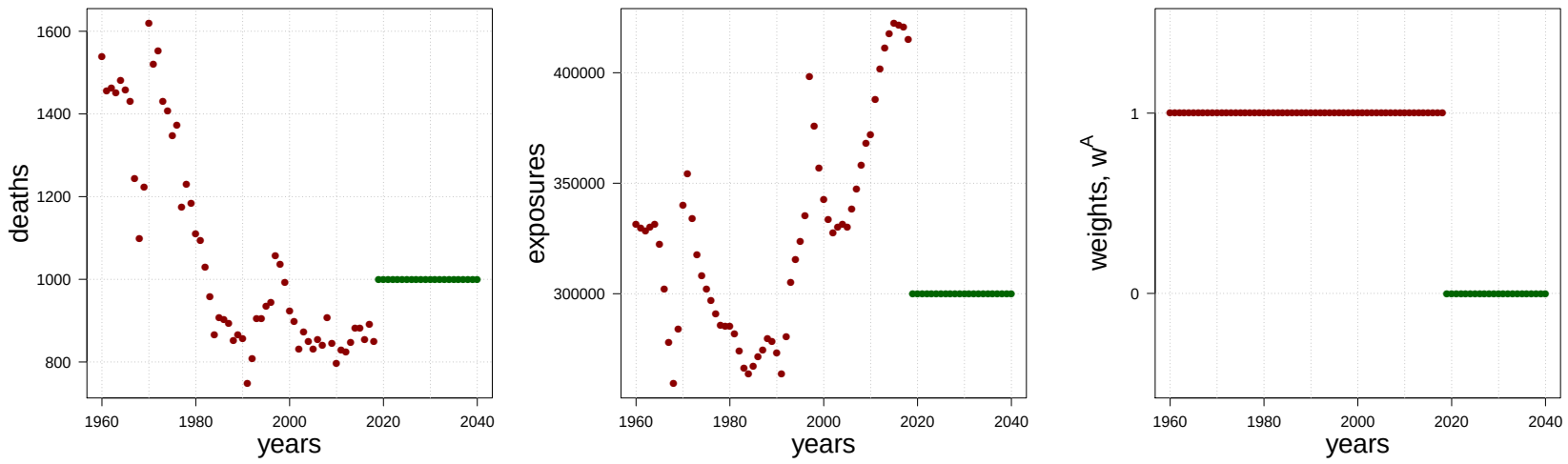
$$\boldsymbol{P} = \lambda\, \boldsymbol{D}'\boldsymbol{D} \quad \text{with } (k-d) \times k \text{ matrix } \boldsymbol{D}$$

- Create new vector of weights:

$$w_i^A = [\underbrace{1, 1, \ldots, 1}_{1960-2018}, \underbrace{0, 0, \ldots, 0}_{2019-2040}]$$

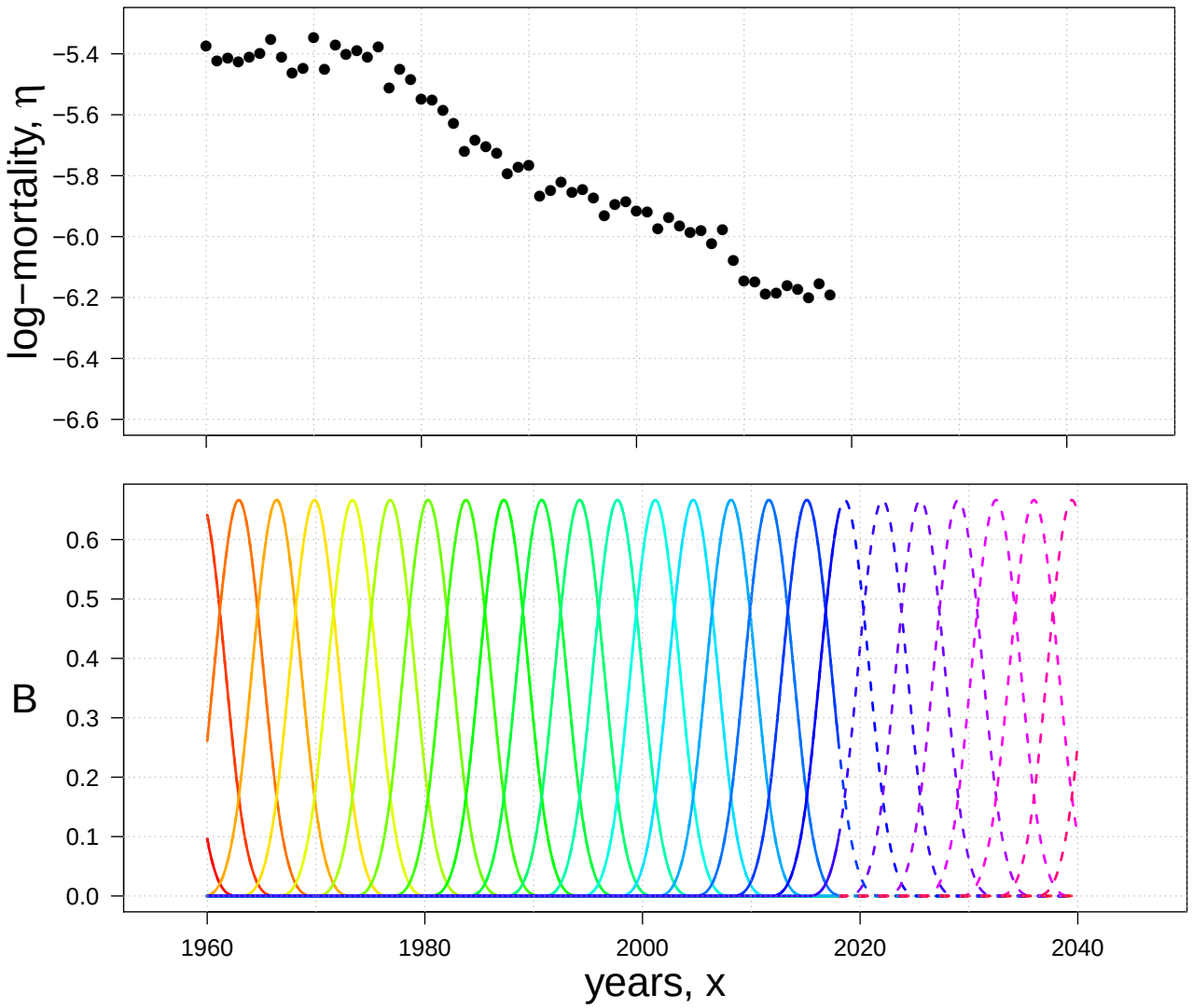
- We include w_i^A in the algorithm:
  - working response: $z_i = w_i^A \left[\eta_i + \frac{y_i - e_i\,\mu_i}{e_i\,\mu_i}\right]$
  - weight matrix: $w_i = w_i^A\, e_i\, \mu_i$

Extrapolating in practice

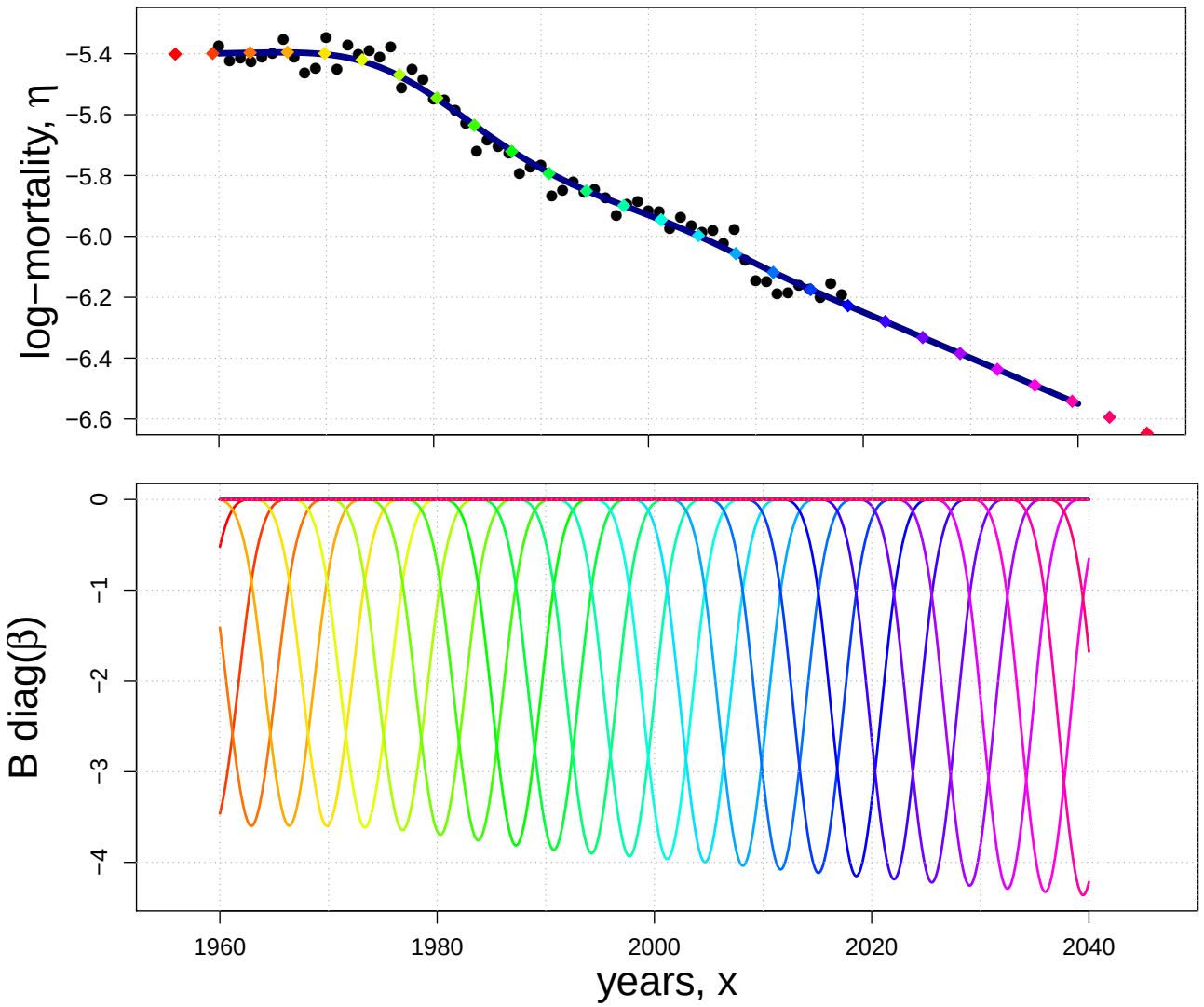


Deaths, exposures and weights used for extrapolating with P-splines.
England & Wales, age 50, years 1960-2018, arbitrary values 2019-2040.
Source: Human Mortality Database.

Extrapolating in practice



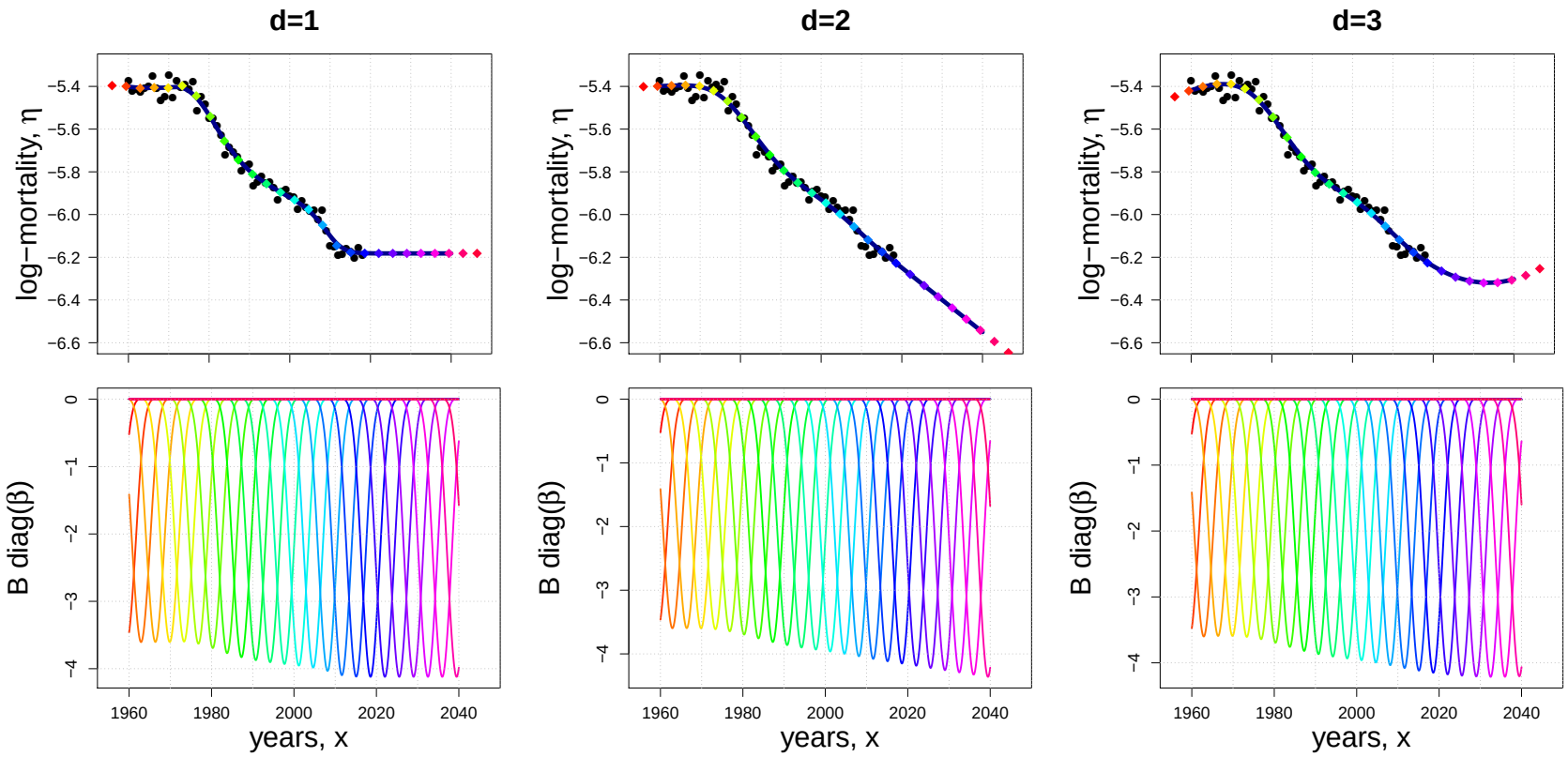
Extrapolating in practice



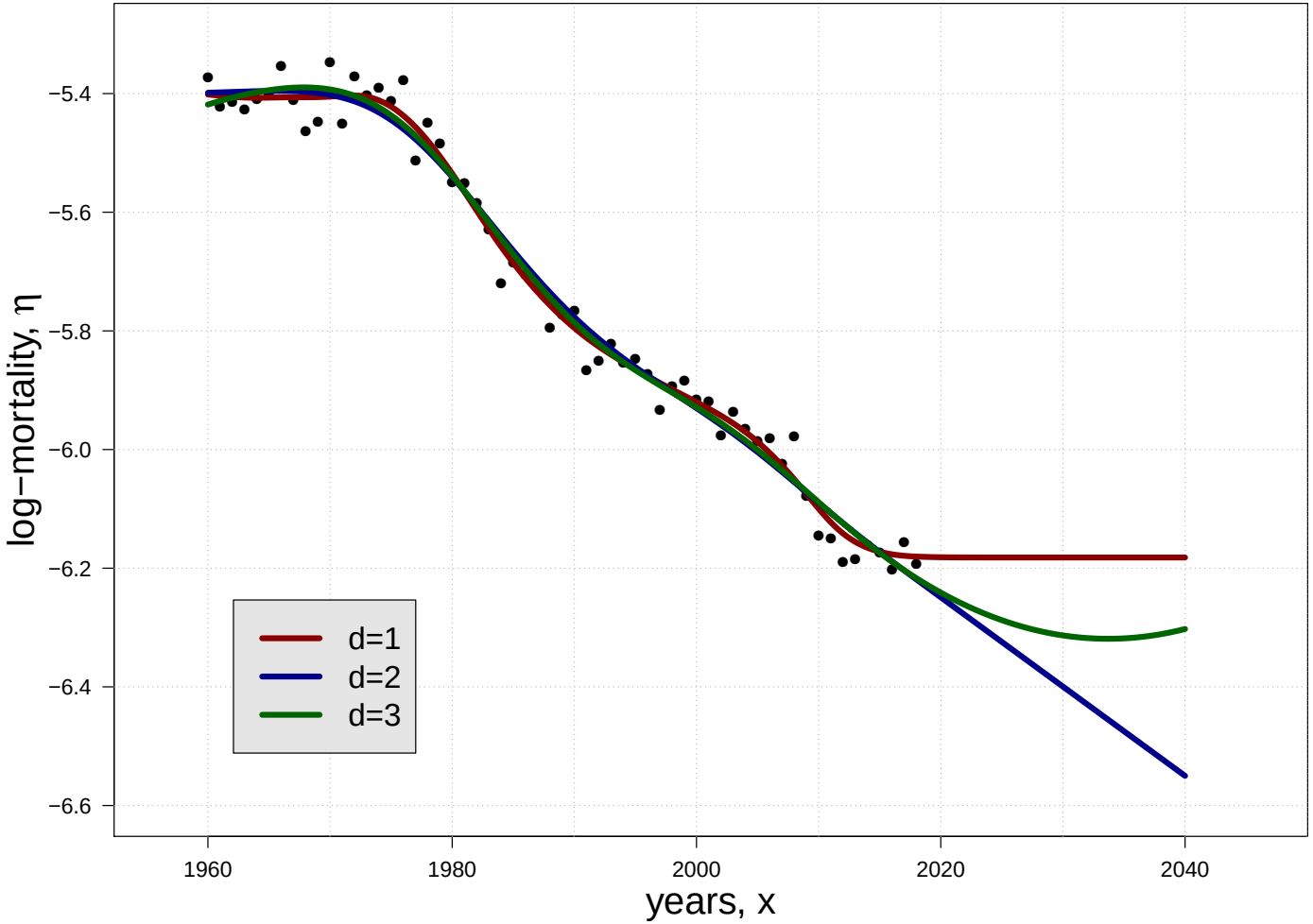
Warning: Order of difference in extrapolating

- Unlike in fitting, the order of difference d is crucial when extrapolating
- In an *unknown land*, P-splines assume the *ultimate* smoothing behaviour at the coefficient level:
  - $d = 1$: keep constant the last fitted coefficient
  - $d = 2$: extrapolate linearly the last two fitted coefficients
  - $d = 3$: extrapolate quadratically the last three fitted coefficients

Extrapolating in practice

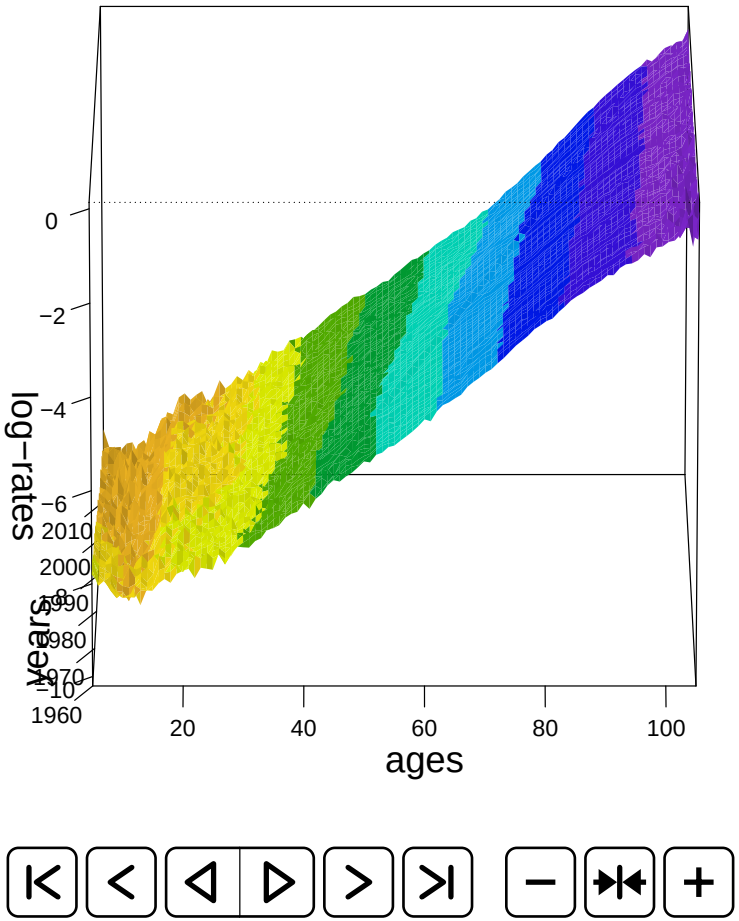


Extrapolating in practice



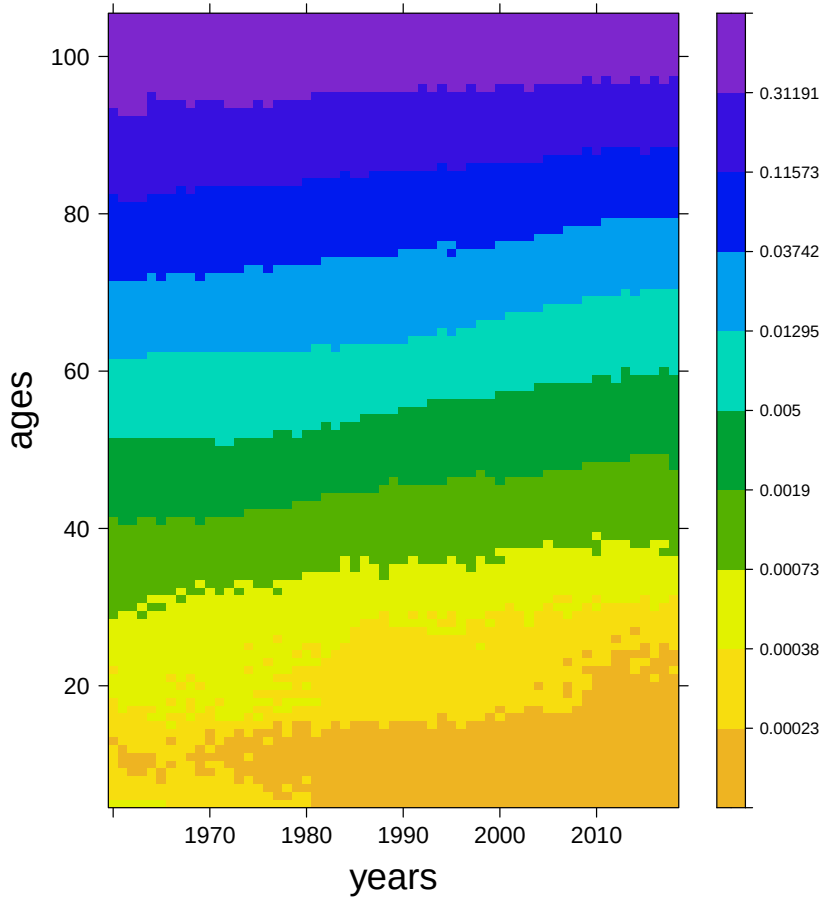
Log-mortality, estimated and forecast with P-splines with different d for the penalty term.
England & Wales, age 50, years 1960-2018
Source: Human Mortality Database.

Mortality over age and year



Age- and year-specific death rates over ages and years, log scale.
England & Wales, females, ages 5-105, 1960-2018.
Source: Human Mortality Database.

Mortality over age and year



Age- and year-specific death rates over ages and years, log scale.
England & Wales, females, ages 5-105, 1960-2018.
Source: Human Mortality Database.

A model for mortality surface

- Data are provided in a rectangular grid over m ages and n years
- We vectorize both deaths (response) and exposures (offset):

$$\begin{aligned} \boldsymbol{y} &= \texttt{vec}(\boldsymbol{Y}) \\ \boldsymbol{e} &= \texttt{vec}(\boldsymbol{E}) \end{aligned}$$

- A possible model for the surface as a whole:

$$\mathbb{E}[\ln(\boldsymbol{y})] = \ln(\boldsymbol{e}) + \boldsymbol{B}\,\boldsymbol{\beta}$$

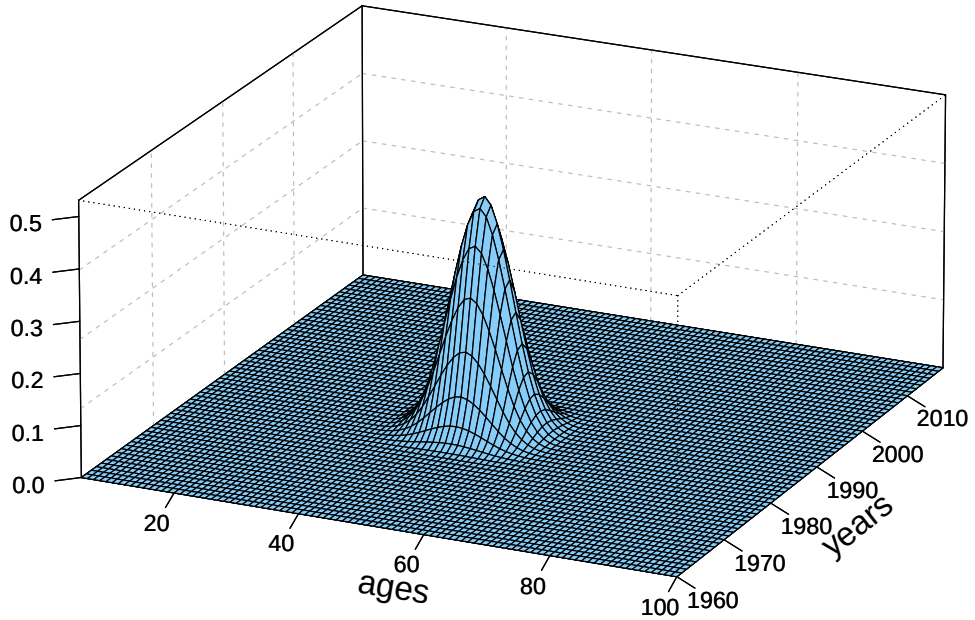
- Note that model description is like in the one-dimensional setting $\Rightarrow$ estimation produce will be the same. But ...
- We need a
  - $\boldsymbol{B}$ that runs over both age and years
  - $\boldsymbol{P}$ (penalty term) that penalizes over both age and years

B-splines over age and years

- Kronecker product of bases:

$$\underbrace{\boldsymbol{B}}_{mn \times k_a k_t} = \underbrace{\boldsymbol{B}_t}_{n \times k_t} \otimes \underbrace{\boldsymbol{B}_a}_{m \times k_a}$$

where $\boldsymbol{B}_a$ and $\boldsymbol{B}_t$ are B-spline basis along ages (a) and years (t)



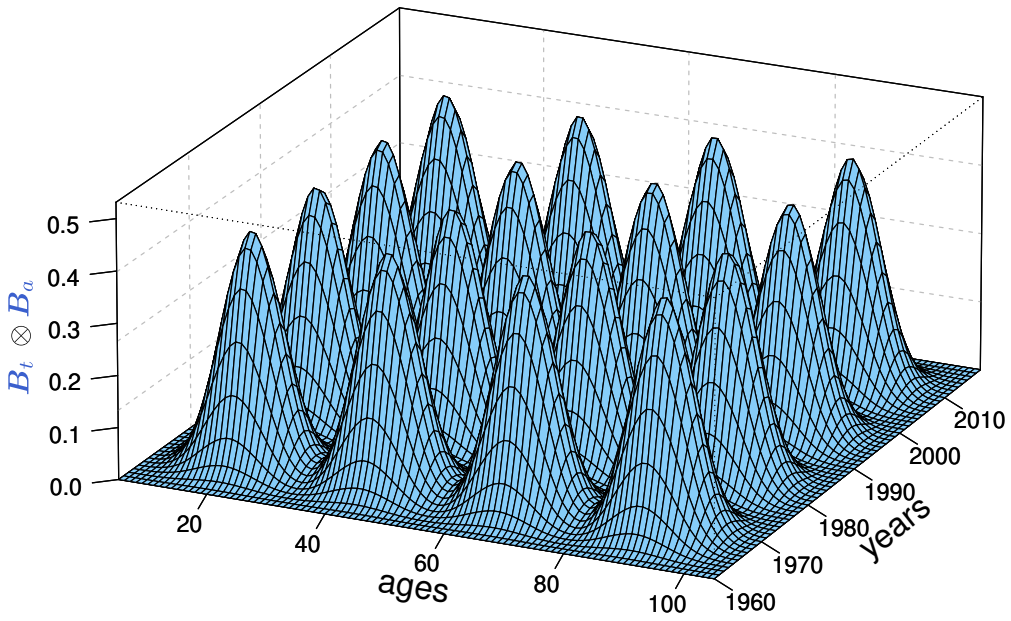
Kronecker product between two B-splines basis over ages (5-105) and years (1960-2018)

B-splines over age and years

- Kronecker product of bases:

$$\underbrace{\boldsymbol{B}}_{mn \times k_a k_t} = \underbrace{\boldsymbol{B}_t}_{n \times k_t} \otimes \underbrace{\boldsymbol{B}_a}_{m \times k_a}$$

where $\boldsymbol{B}_a$ and $\boldsymbol{B}_t$ are B-spline basis along ages (a) and years (t)



"Egg carton": schematic view of Kronecker product between few B-splines over ages (5-105) and years (1960-2018)

Two-dimensional penalty

- In the model $\boldsymbol{\beta}$ is a $k_a k_t$ vector
- Convert $\boldsymbol{\beta}$ into a $k_a \times k_t$ matrix $\mathcal{B}$ (a coefficient for each "egg")
- Generalized the penalty term in the log-likelihood

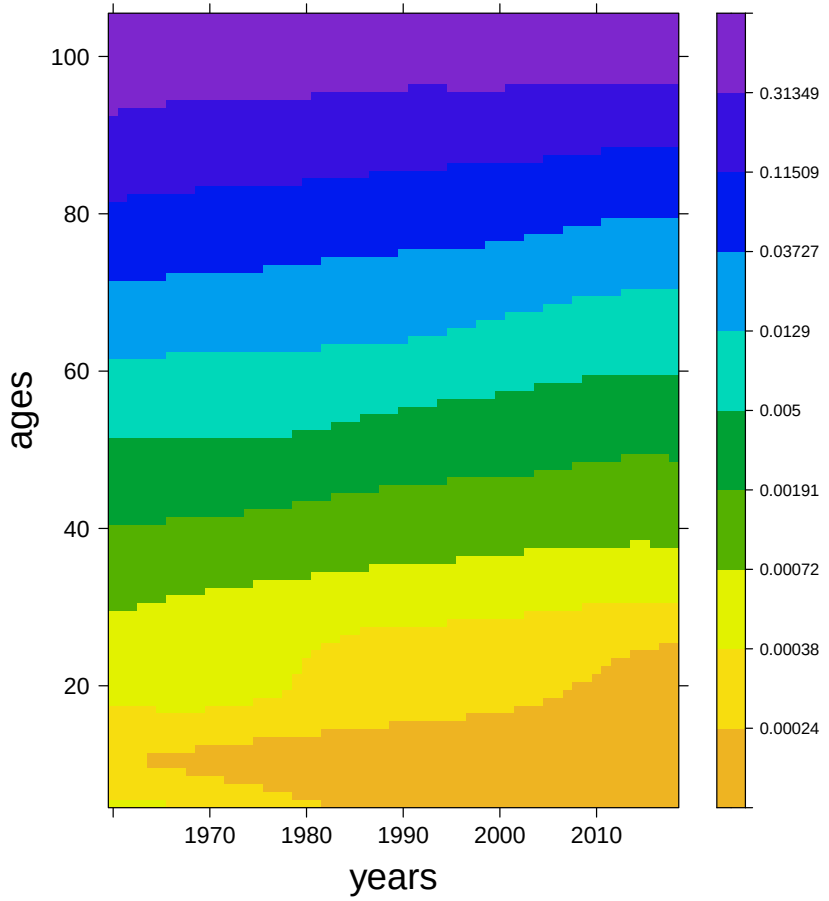
$$\begin{aligned}
\text{Roughness measure} &= \text{Penalty over the rows of } \mathcal{B} + \text{Penalty over the columns of } \mathcal{B} \\
&= \lambda_a \sum_{j=1}^{k_a} \mathcal{B}_{j\bullet}\, \boldsymbol{D}'_{d_a} \boldsymbol{D}_{d_a}\, \mathcal{B}'_{j\bullet} + \lambda_t \sum_{k=1}^{k_t} \mathcal{B}_{\bullet k}\, \boldsymbol{D}'_{d_t} \boldsymbol{D}_{d_t}\, \mathcal{B}'_{\bullet k} \\
&= \lambda_a \left|\boldsymbol{P}_a \boldsymbol{\beta}\right|^2 + \lambda_t \left|\boldsymbol{P}_t \boldsymbol{\beta}\right|^2
\end{aligned}$$

- This roughness measure in the IWLS algorithm (see slide 5) means

$$\boldsymbol{P} = \lambda_a (\boldsymbol{I}_{k_t} \otimes \boldsymbol{D}'_{d_a} \boldsymbol{D}_{d_a}) + \lambda_t (\boldsymbol{D}'_{d_t} \boldsymbol{D}_{d_t} \otimes \boldsymbol{I}_{k_a})$$

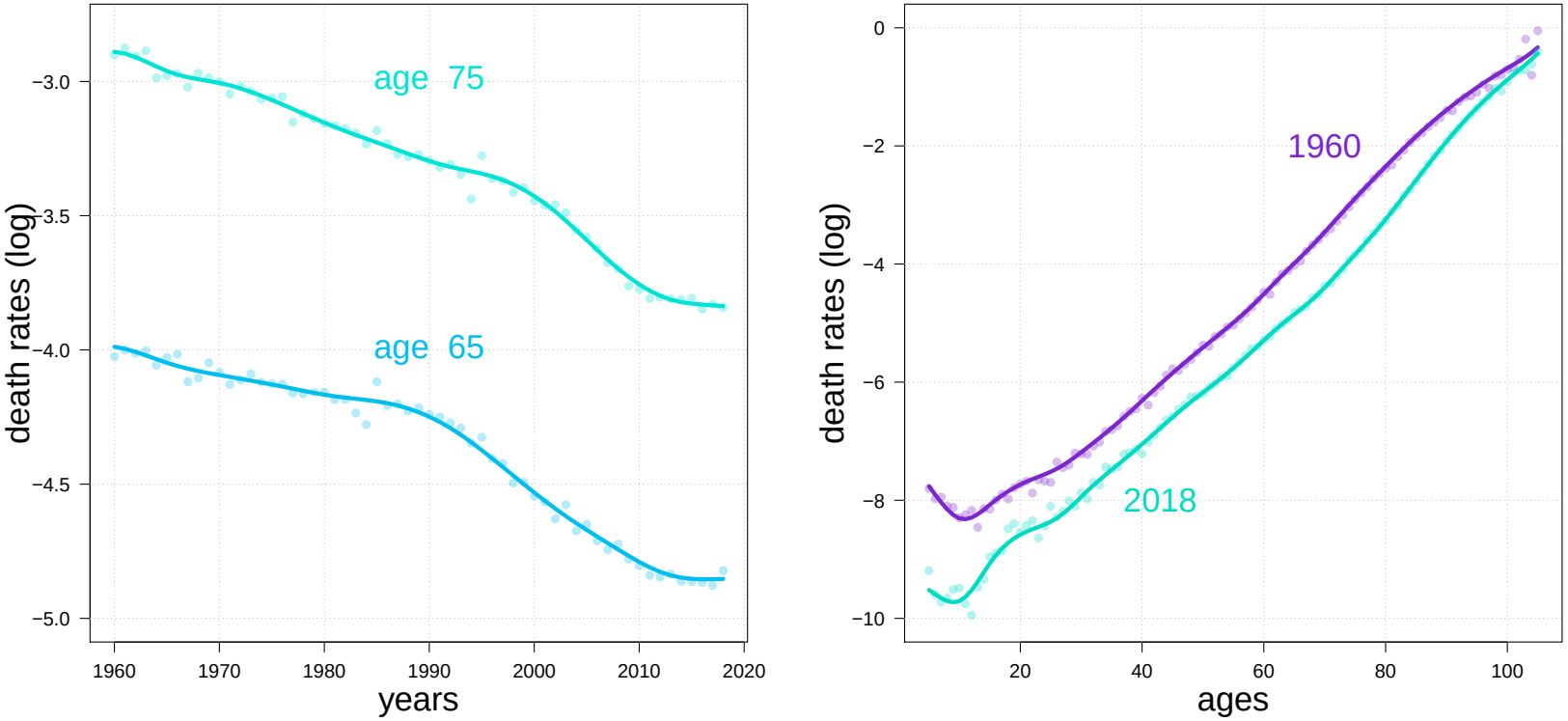
- Smoothing parameters (λ_a, λ_t) allows anisotropic smoothing
- This model could be written as a Generalized Linear Array Model ($GLAM$, Currie et al. 2006)

"Simple" mortality smoothing over age and year



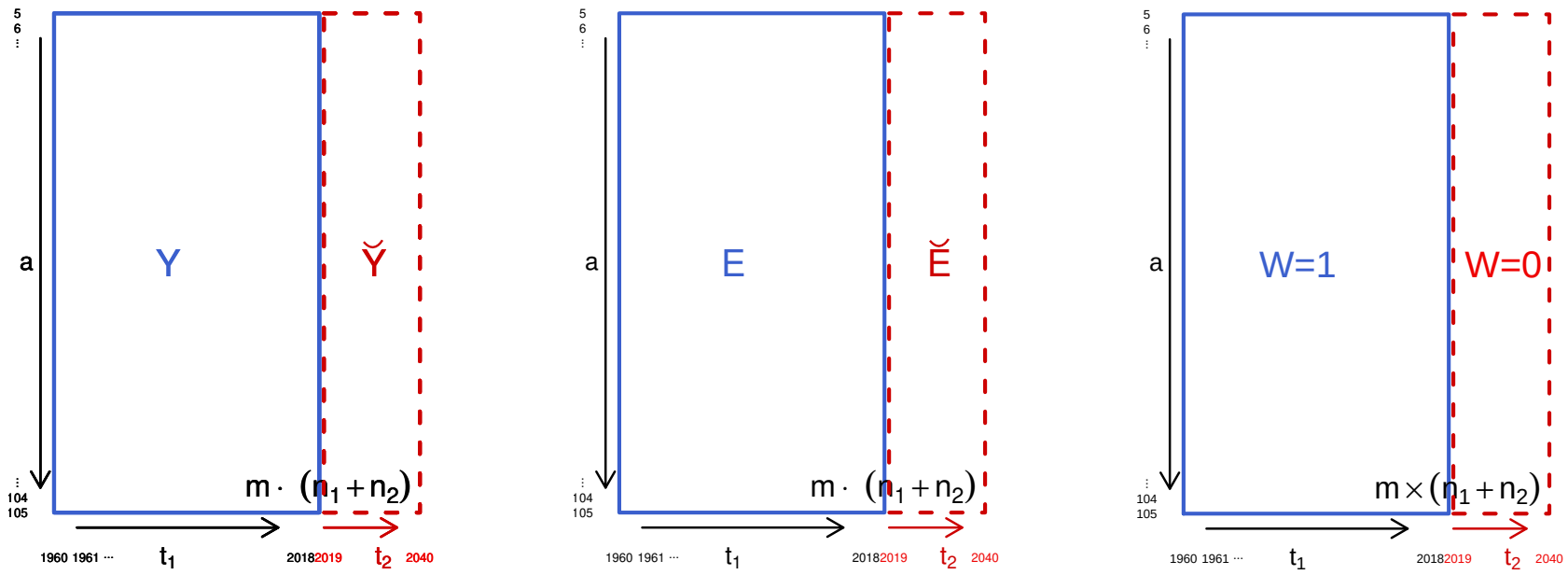
Smooth by P-splines age- and year-specific death rates over ages and years, log scale.
England & Wales, females, ages 5-105, 1960-2018.
Source: Human Mortality Database.

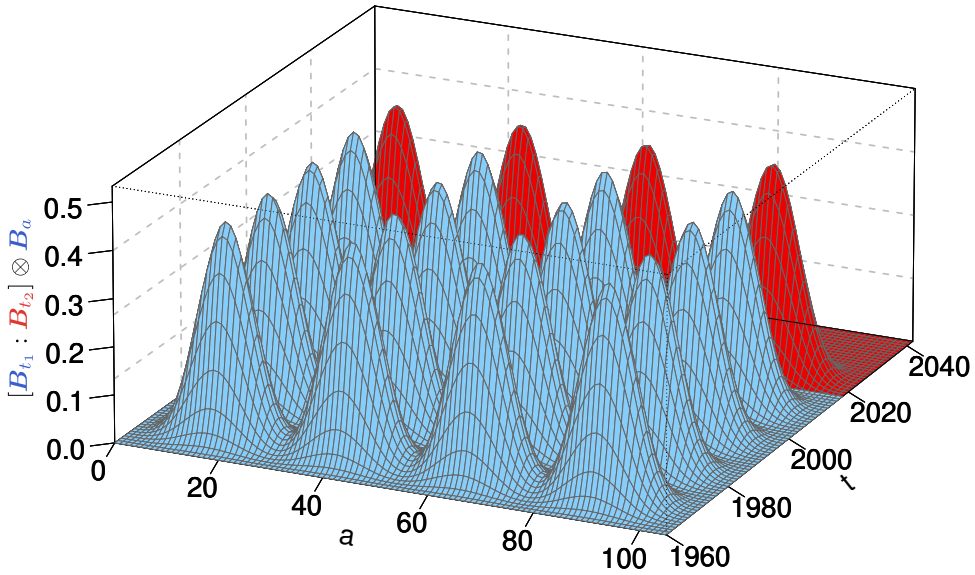
"Simple" mortality smoothing over age and year



Actual and smooth by P-splines death rates for selected ages over years and vice versa, log scale.
England & Wales, females, ages 5-105, 1960-2018.
Source: Human Mortality Database.

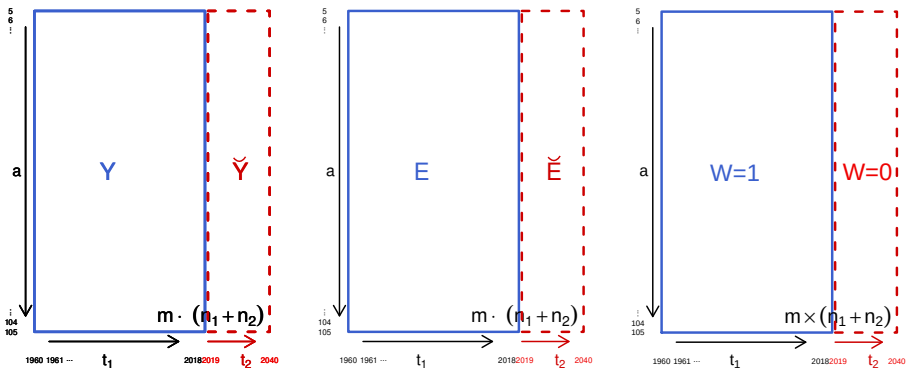
Like in 1D setting: forecasting as missing value problem

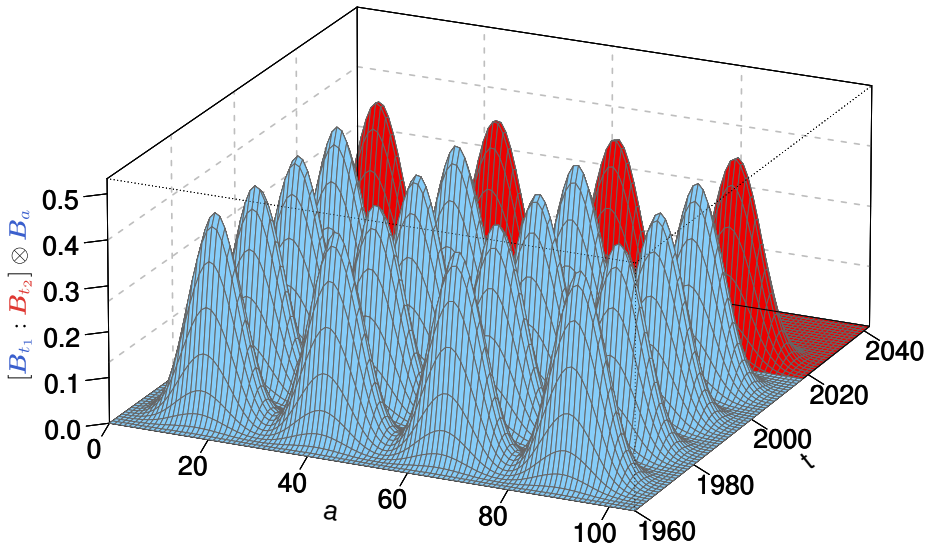




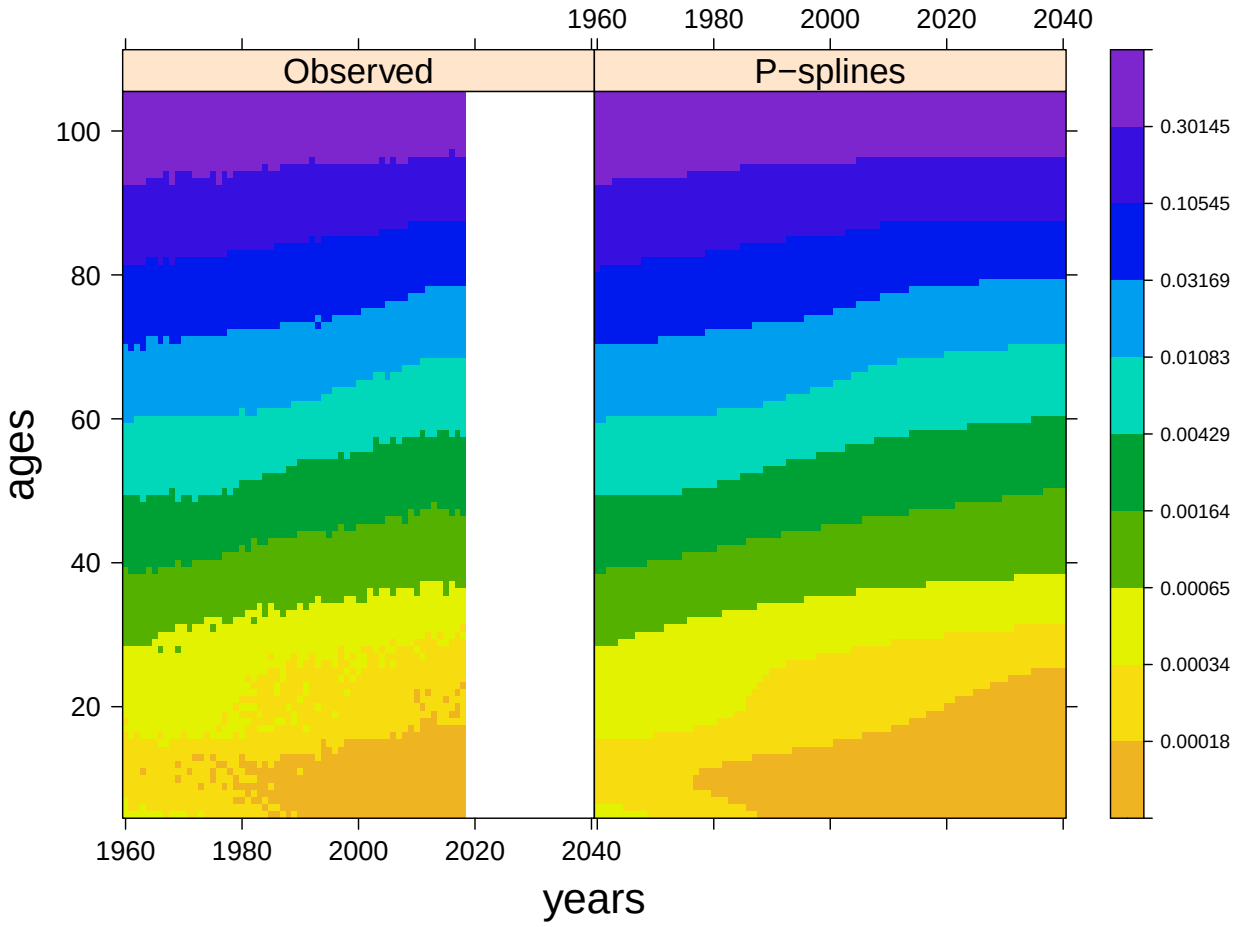
Associated equations for 2D extrapolation

- Data are augmented with arbitrary future values
- Define a 0/1 weight matrix ($\boldsymbol{W}^A$)
- We vectorize response, offset and weights ($\boldsymbol{w}^A = \mathtt{vec}(\boldsymbol{W}^A)$)
- B-spline bases are augmented too: $\boldsymbol{B} = [\boldsymbol{B}_{t_1} : \boldsymbol{B}_{t_2}] \otimes \boldsymbol{B}_a$
- Like in 1D, we add weights $\boldsymbol{w}^A$ in the IWLS algorithm



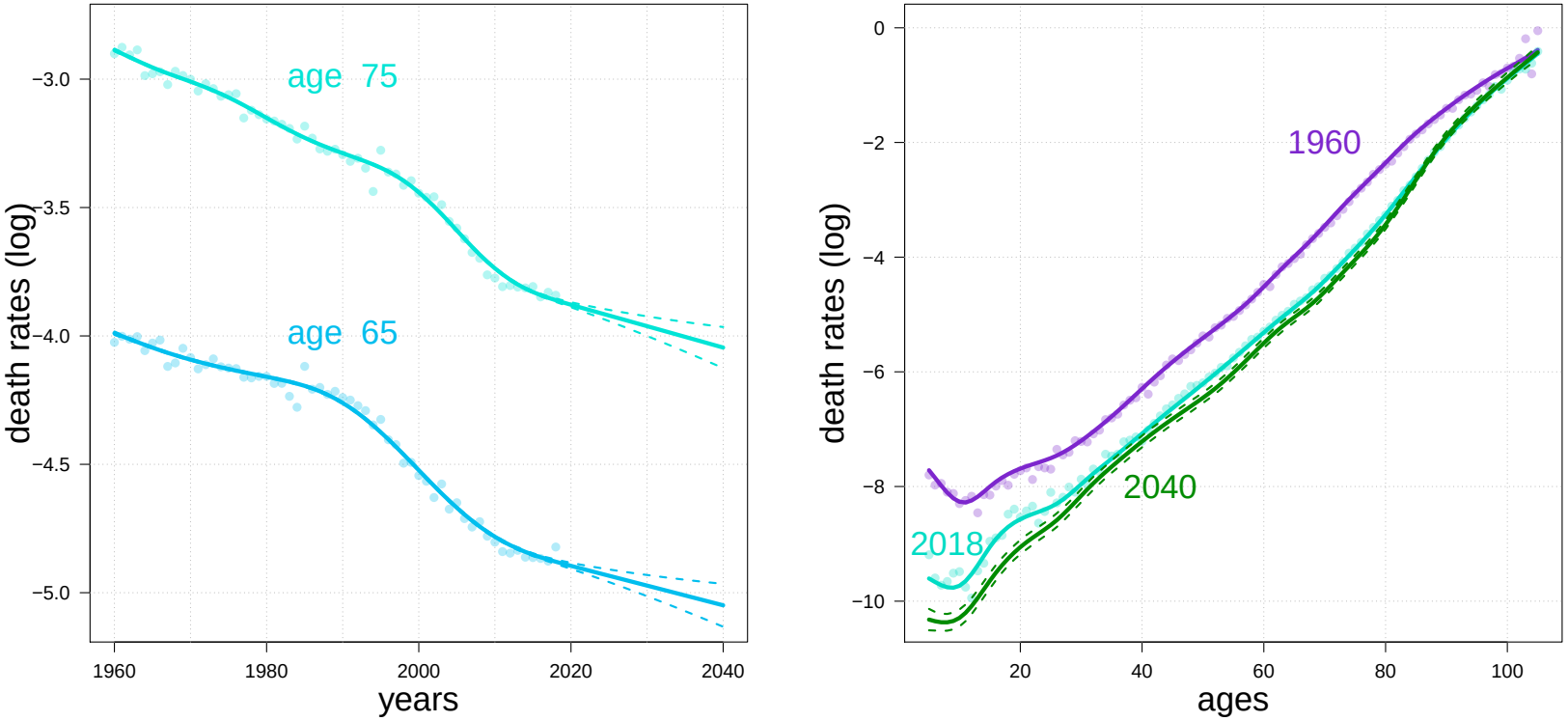


Smoothing and forecast over age and year



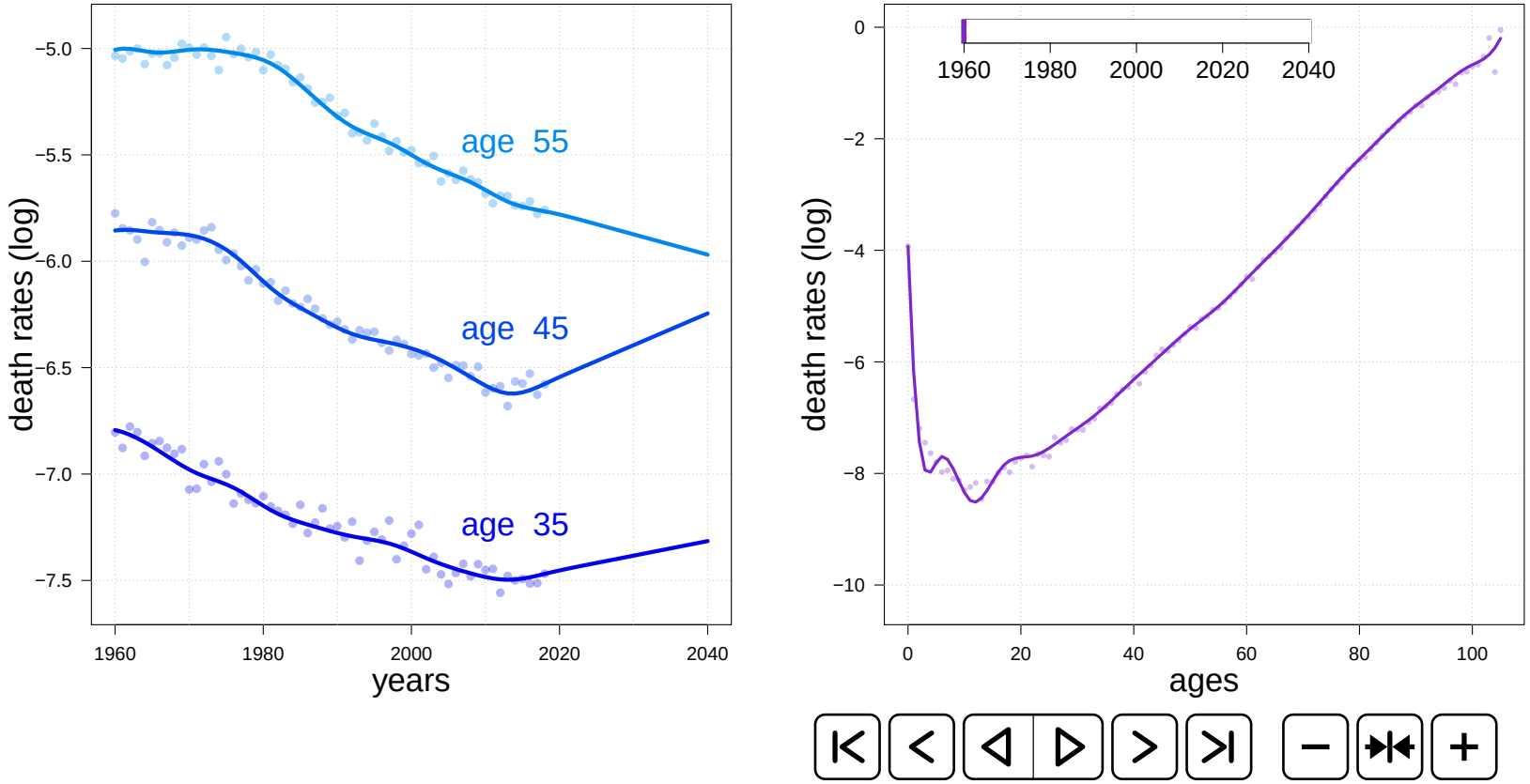
Smooth and forecast by P-splines age- and year-specific death rates over ages and years, log scale.
England & Wales, females, ages 5-105, 1960-2018.
Source: Human Mortality Database.

Smoothing and forecast over age and year



Smooth and forecast by P-splines death rates for selected ages over years and vice versa, log scale.
England & Wales, females, ages 5-105, 1960-2018.
Source: Human Mortality Database.

Modelling from age 0



- Infant mortality needs a lot of flexibility to be properly described $\Rightarrow$ small selected λ_a $\Rightarrow$
  - undersmooth estimated rates in the observed years
  - unreasonable future age-patterns

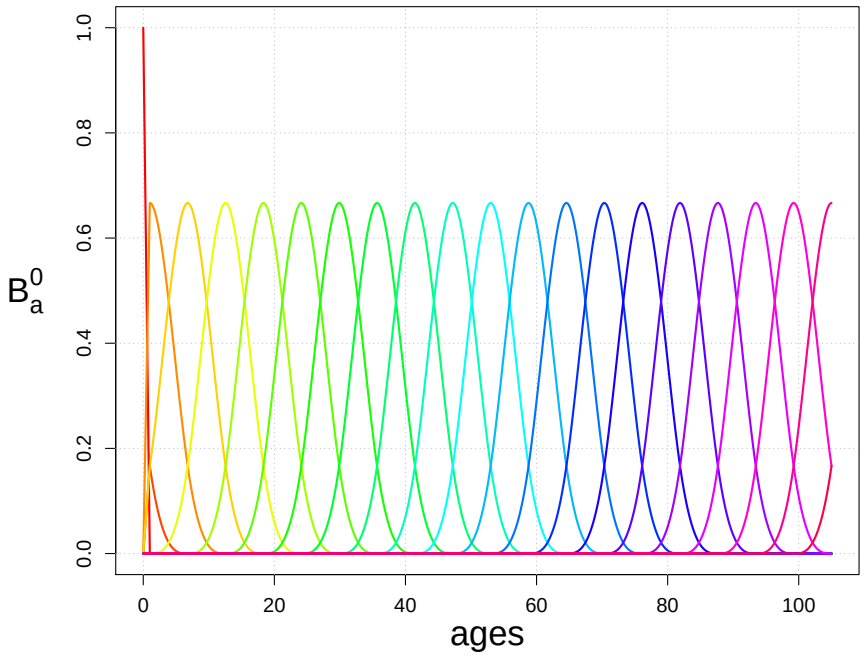
Addressing infant mortality

- The first year of life is usually treated differently in demography
- Let modify accordingly the B-spline basis over age:

$$\boldsymbol{B}_a^0 = \begin{bmatrix} 1 & \mathbf{0}_{1\times k_a} \\ \mathbf{0}_{(m-1)\times 1} & \boldsymbol{B}_a \end{bmatrix}$$

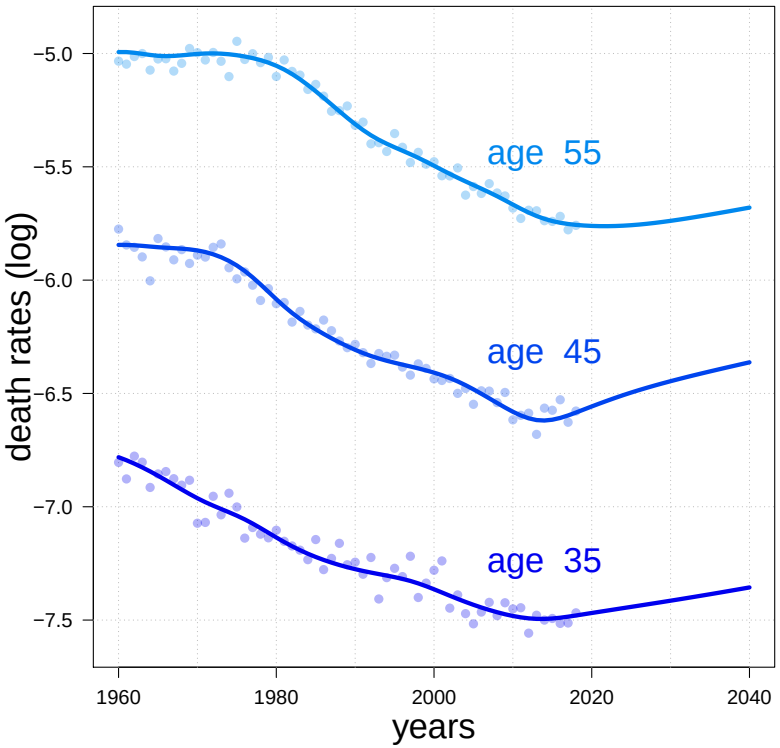
with $\boldsymbol{B}_a$: $(m-1)\times k_a$ matrix of B-splines

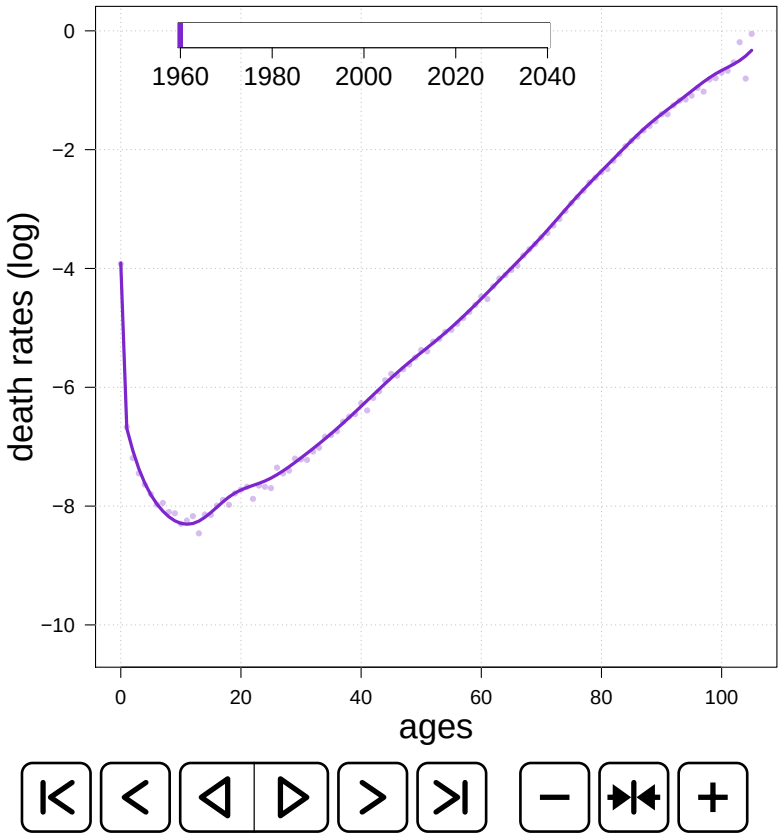
- We replace the first cell in $\boldsymbol{D}_a$ with a zero



- single and specialized coefficient for infant mortality values
- a smooth change in infant mortality over time
- overall smoothness will not be affected infant mortality behavior

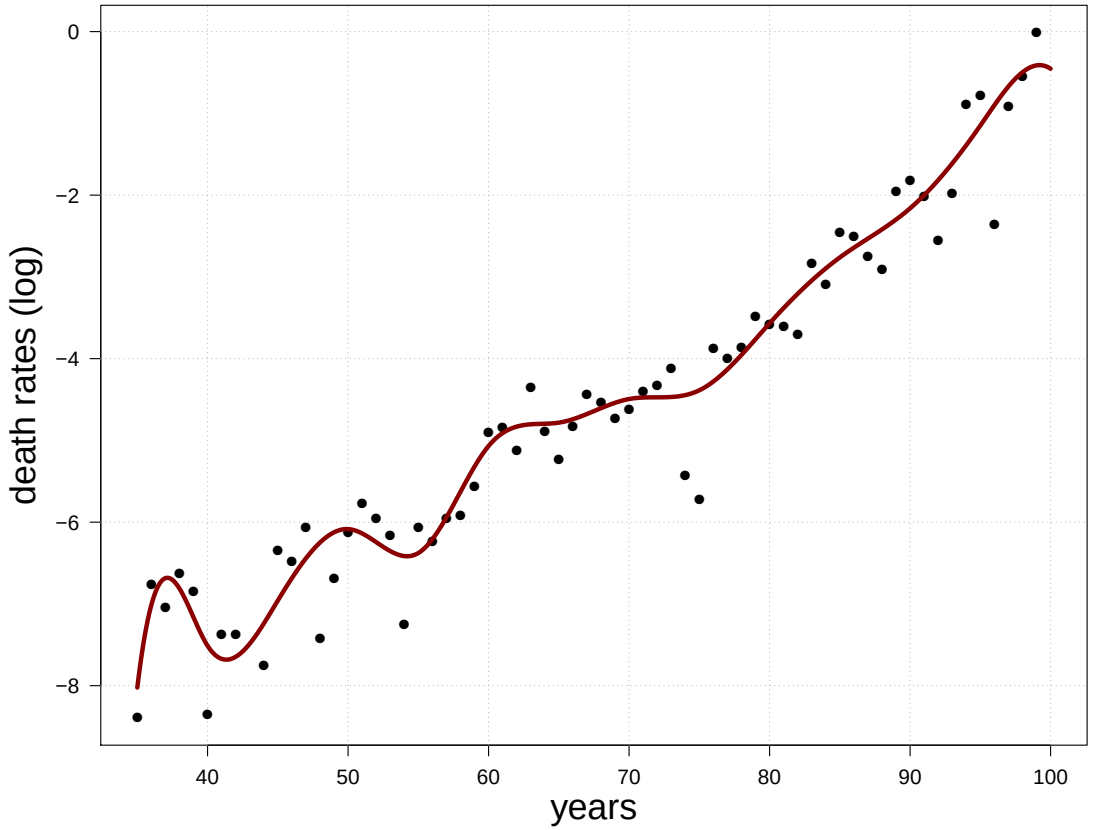
P-splines + infant specialized coefficients





- Improvement in terms of fitting, still unreasonable forecasts
  - the model purely is data-driven
  - need to incorporate demographic information

Mortality for a small population



Actual and fitted death rates over age for males with Tertiary education.
France, 2016, ages 35-100.
Source: INSEE, 2017

- We expect that mortality increases over age (after age 35)
- We aim to enforce a monotonic increase of the smooth trend
- We need to work on the derivative of the smooth trend

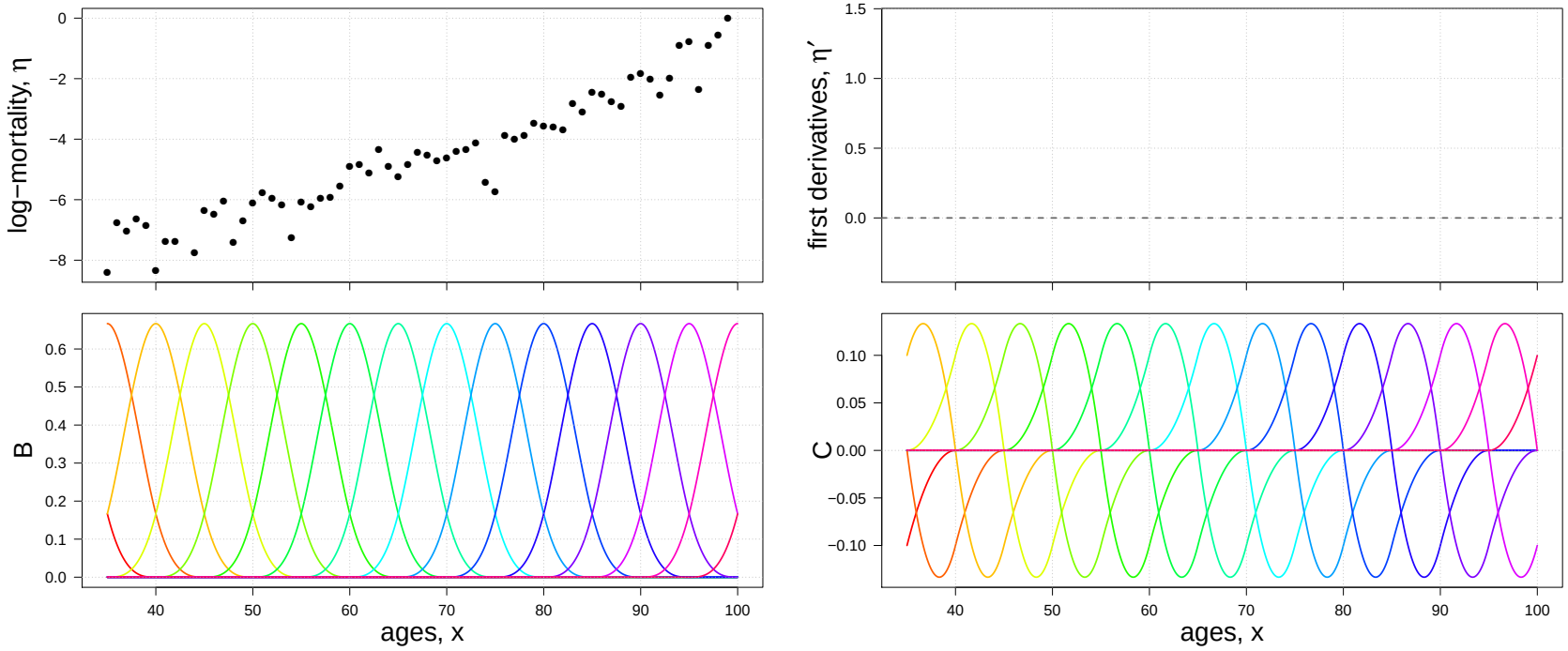
Moving to derivatives

- Linear predictor:

$$\boldsymbol{\eta} = \boldsymbol{B}\,\boldsymbol{\beta}$$

- Derivative of linear predictor:

$$\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{x}} = \frac{\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{x}}}{\boldsymbol{\mu}} = \boldsymbol{C}\,\boldsymbol{\beta}$$



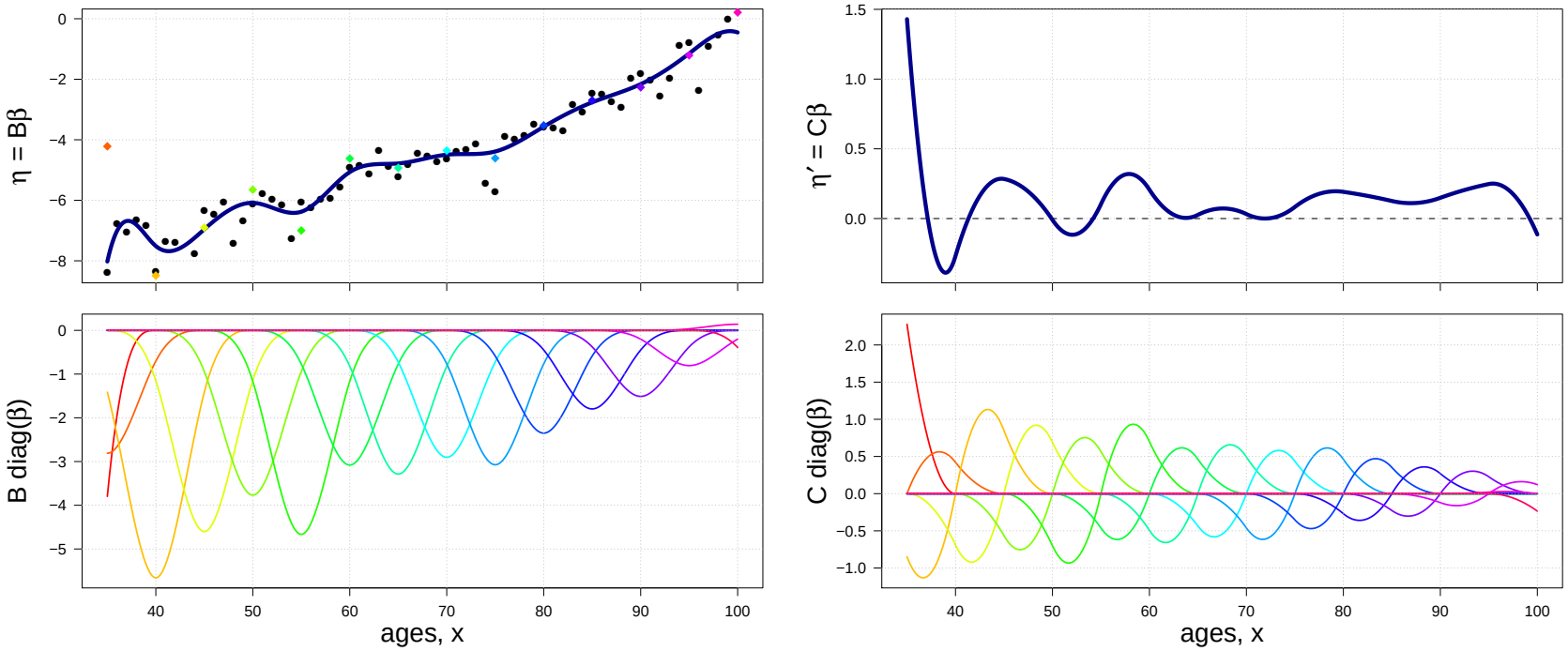
Moving to derivatives

- Linear predictor:

$$\boldsymbol{\eta} = \boldsymbol{B}\,\boldsymbol{\beta}$$

- Derivative of linear predictor:

$$\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{x}} = \frac{\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{x}}}{\boldsymbol{\mu}} = \boldsymbol{C}\,\boldsymbol{\beta}$$



- if $\boldsymbol{x}$ is age, $\boldsymbol{\eta}' = \boldsymbol{r} = (r_i)$: rate-of-aging
- if $\boldsymbol{x}$ is time, $\boldsymbol{\eta}' = \boldsymbol{\rho} = (\rho_i)$: rate-of-change/rate of mortality improvement

Asymmetric penalty

- Monotonic increase $\Rightarrow \boldsymbol{\eta}' = \boldsymbol{r} = \boldsymbol{C}\,\boldsymbol{\beta} > 0$
- We transform into a penalty this constraint:

$$\sum_i v_i\, r_i^2$$

where

$$v_i = \begin{cases} 0, & \text{if} \quad r_i \geq 0 \\ 1, & \text{otherwise} \end{cases}$$

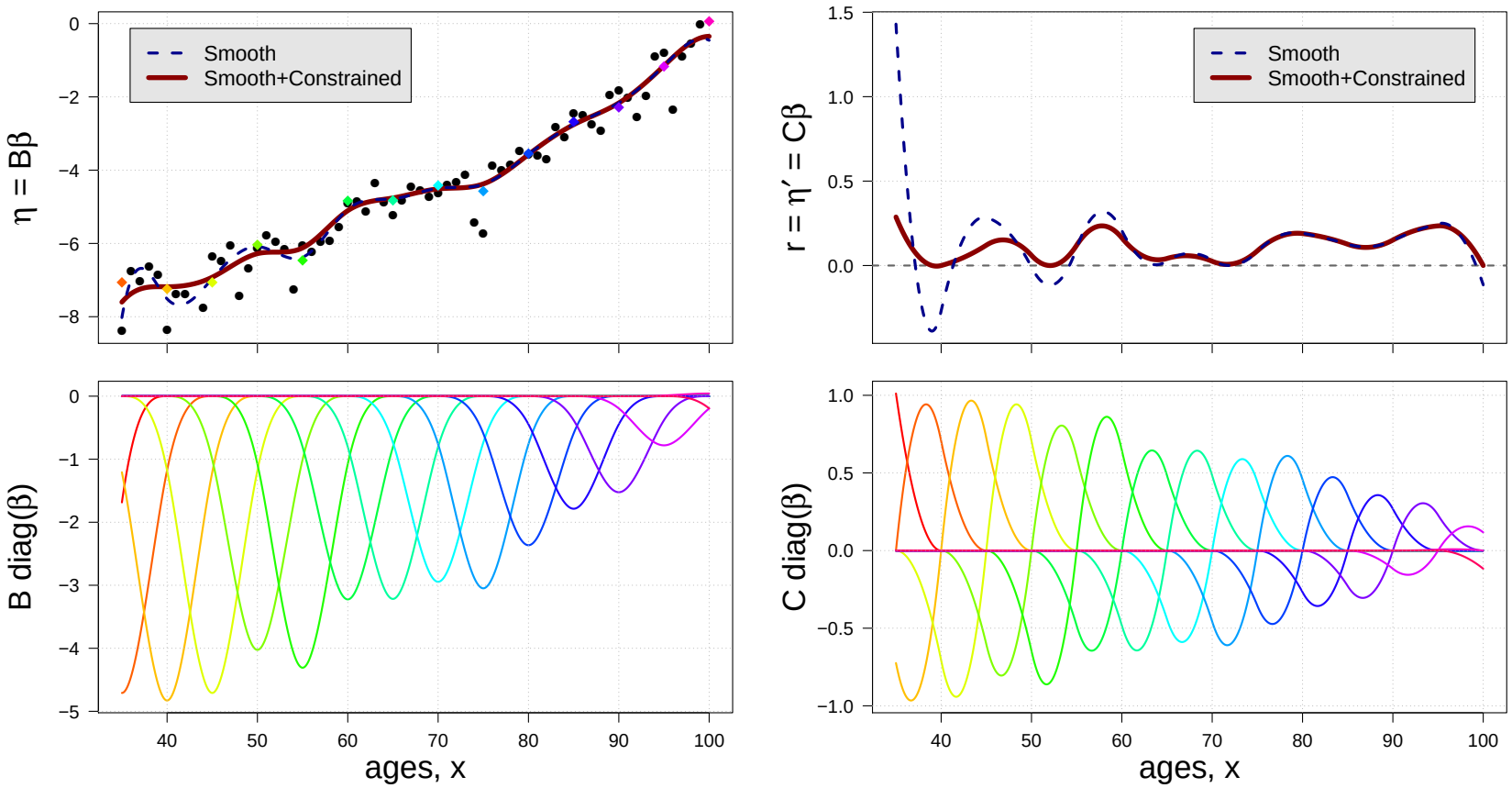
- v_i are *asymmetric* weights: negative values in r_i are penalized whereas positive are not
- We can add an *asymmetric* penalty within the penalized IWLS

$$\boldsymbol{P}_s = \kappa\, \boldsymbol{C}'\boldsymbol{V}\boldsymbol{C}$$

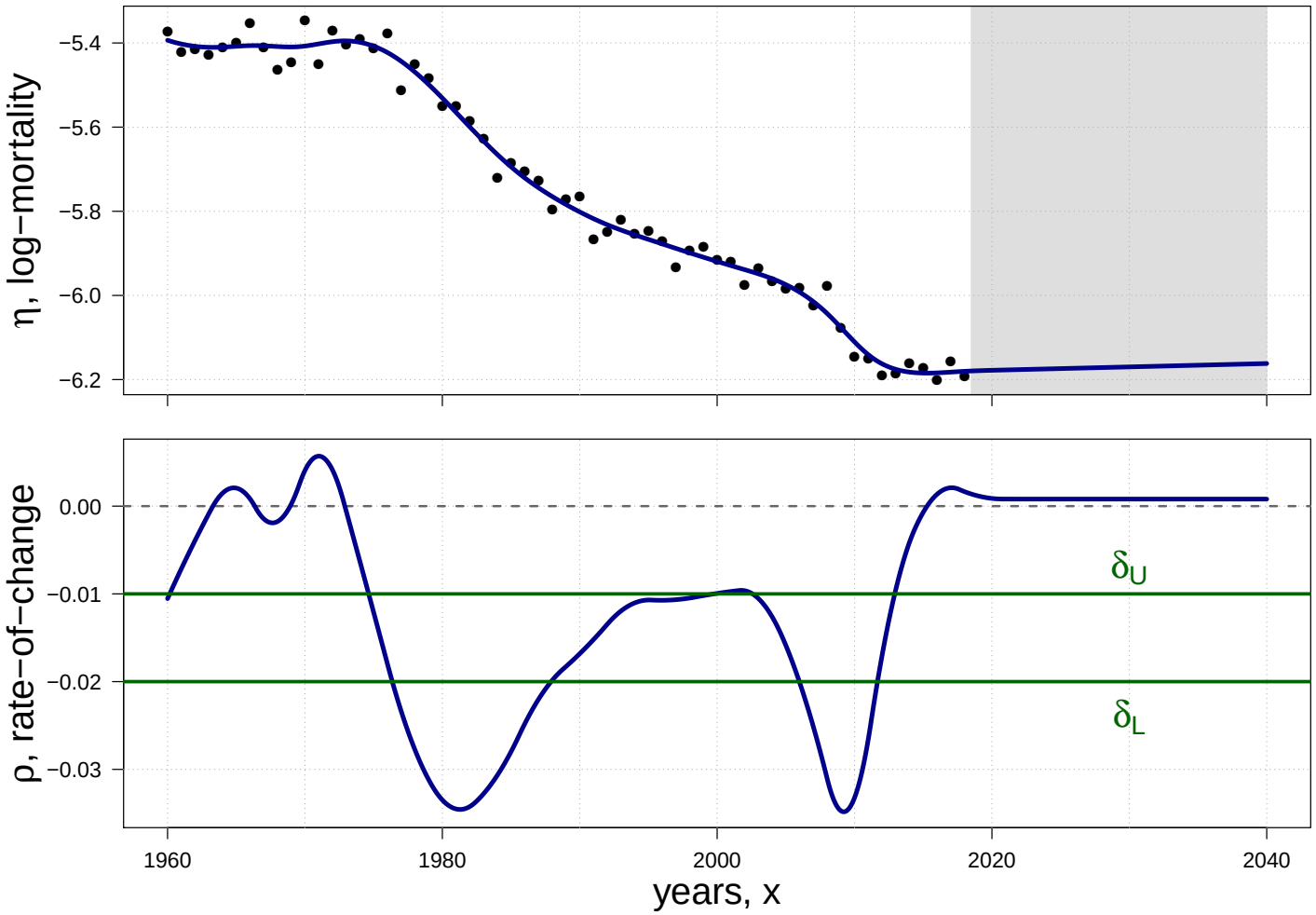
with $\kappa = 10^8$ and $\boldsymbol{V} = \texttt{diag}(v_i)$

- Values of $\boldsymbol{v}$ need to be updated every IWLS iteration
- Other types of shape can be achieved by asymmetric penalties. E.g. concavity, convexity by second derivatives

Estimating with asymmetric penalty



Derivatives over time



Log-mortality and associated rate-of-change. P-splines.
England & Wales, age 50, years 1960-2018, forecast up to 2040
Source: Human Mortality Database.

Asymmetric penalty with an interval

- Future trend (decreasing) with rate-of-change within $[\delta_L, \delta_U] \Rightarrow$

$$\boldsymbol{g}_L \leq \boldsymbol{\eta}' = \boldsymbol{\rho} = \boldsymbol{C}\,\boldsymbol{\beta} \leq \boldsymbol{g}_U$$

with $\boldsymbol{g}_L = [\delta_L, \delta_L, \ldots]$ and $\boldsymbol{g}_U = [\delta_U, \delta_U, \ldots]$ (only in future years)

- We transform this double constraint into two asymmetric penalties:

$$\boldsymbol{P}_L = \kappa\, \boldsymbol{C}'\boldsymbol{V}_L\boldsymbol{C} \qquad \boldsymbol{P}_U = \kappa\, \boldsymbol{C}'\boldsymbol{V}_U\boldsymbol{C}$$

$$\boldsymbol{V}_L = \mathtt{diag}(s_i * v_{L,i}) \qquad \boldsymbol{V}_U = \mathtt{diag}(s_i * v_{U,i})$$

$$v_{L,i} = \begin{cases} 0, & \text{if} \quad \rho_i \geq \delta_L \\ 1, & \text{otherwise} \end{cases} \qquad v_{U,i} = \begin{cases} 0, & \text{if} \quad \rho_i \leq \delta_U \\ 1, & \text{otherwise} \end{cases}$$

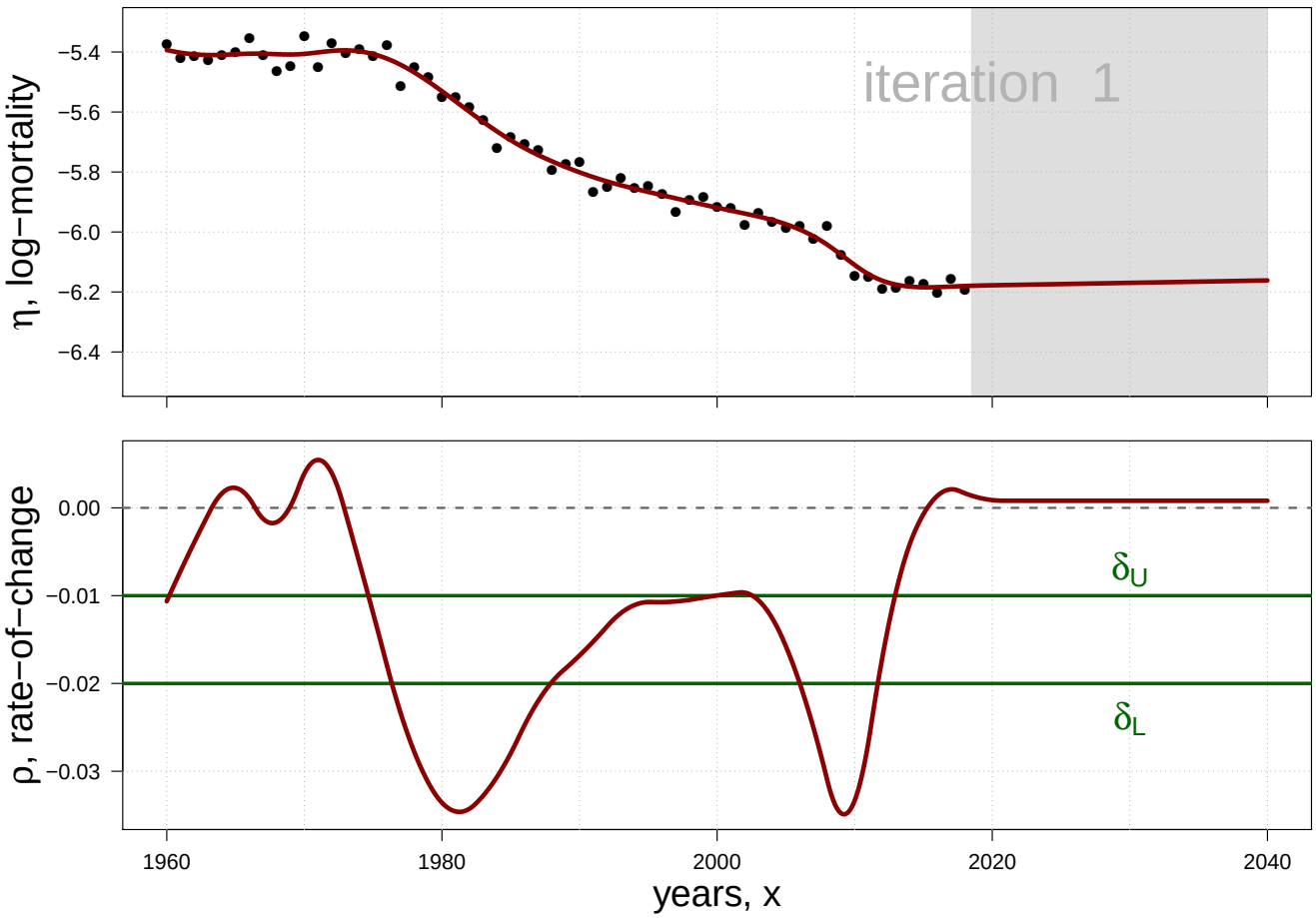
and s is a 0/1 vector equal to 1 for i in future years

- $\delta_L \neq 0$ and $\delta_U \neq 0 \quad \Rightarrow \quad$ adjust RHS of IWLS:

$$\tilde{\boldsymbol{\beta}} = (\boldsymbol{B}'\boldsymbol{W}\boldsymbol{B} + \boldsymbol{P} + \boldsymbol{P}_L + \boldsymbol{P}_U)^{-1}\left(\boldsymbol{B}'\boldsymbol{W}\boldsymbol{z} + \boldsymbol{p}_L + \boldsymbol{p}_U\right)$$

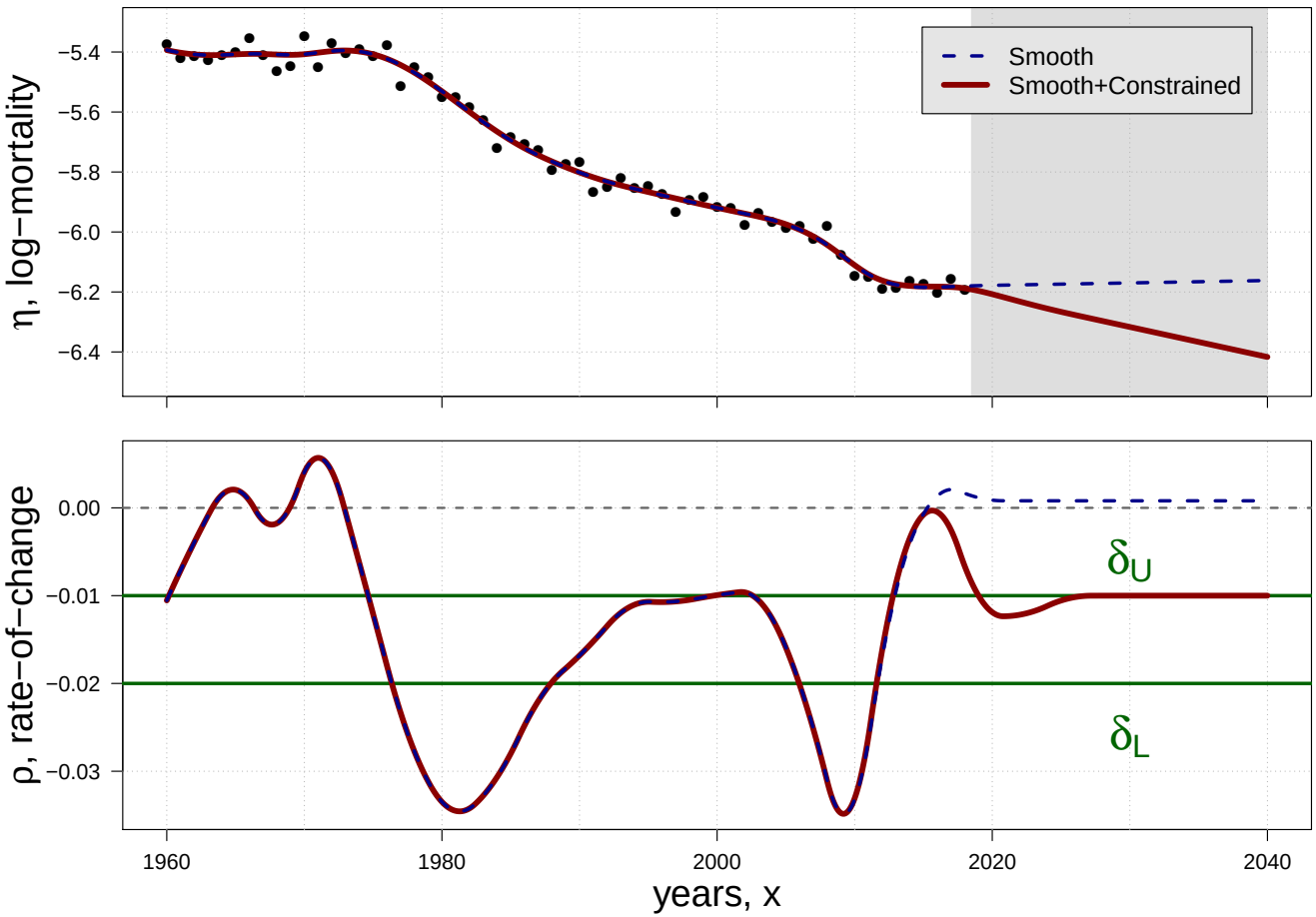
with $\boldsymbol{p}_L = \kappa\, \boldsymbol{C}'\, \boldsymbol{V}_L\, \boldsymbol{g}_L$ and $\boldsymbol{p}_U = \kappa\, \boldsymbol{C}'\, \boldsymbol{V}_U\, \boldsymbol{g}_U$

Asymmetric penalty in action



Log-mortality and associated rate-of-change. P-splines with constraints on future rate-of-change.
England & Wales, age 50, years 1960-2018, forecast up to 2040
Source: Human Mortality Database.

Asymmetric penalty in action



Log-mortality and associated rate-of-change. P-splines with constraints on future rate-of-change.
England & Wales, age 50, years 1960-2018, forecast up to 2040
Source: Human Mortality Database.

Moving to two dimensions

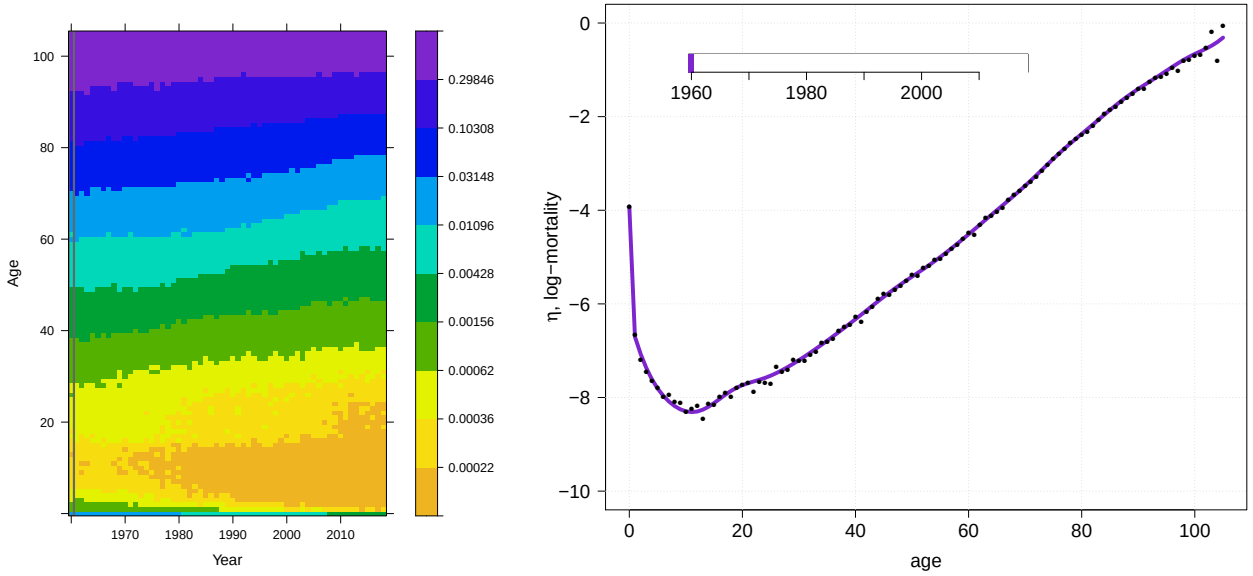
- In 2D setting, derivatives with respect to ages (years) for all years (ages):

$$\begin{aligned} \boldsymbol{r} &= (\boldsymbol{B}_t \otimes \boldsymbol{C}_a)\,\boldsymbol{\beta} \\ \boldsymbol{\rho} &= (\boldsymbol{C}_t \otimes \boldsymbol{B}_a)\,\boldsymbol{\beta} \end{aligned}$$

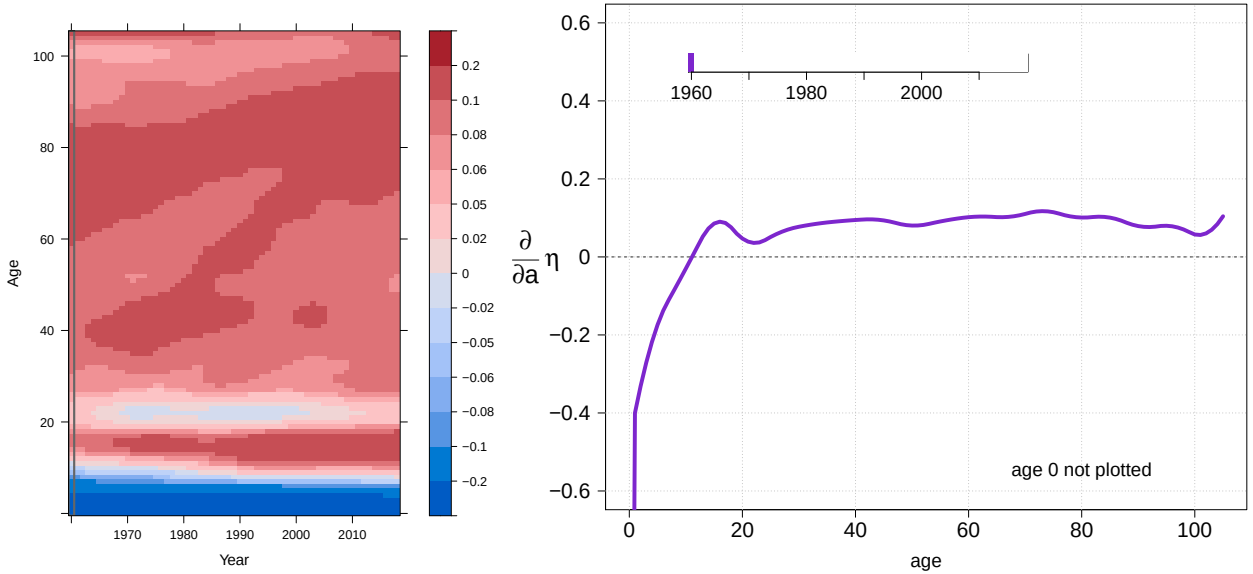
- Let's take the example of England & Wales, females, ages 0-105, years 1960-2018

Observing mortality patterns (over ages)

Actual and Smooth log-mortality

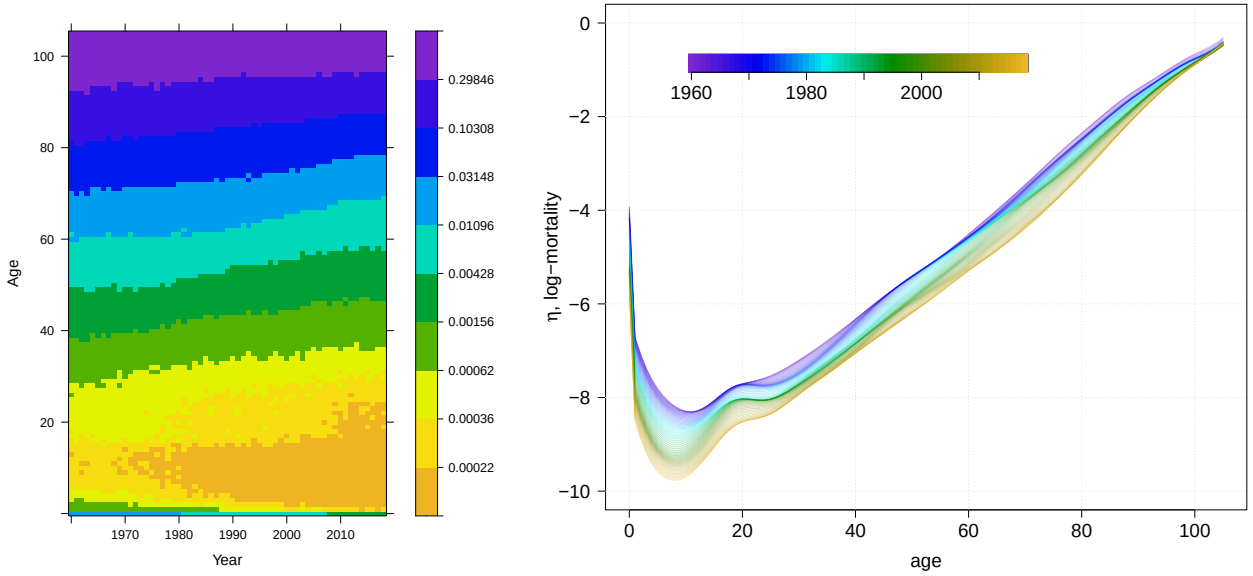


Rate-of-aging from smooth mortality

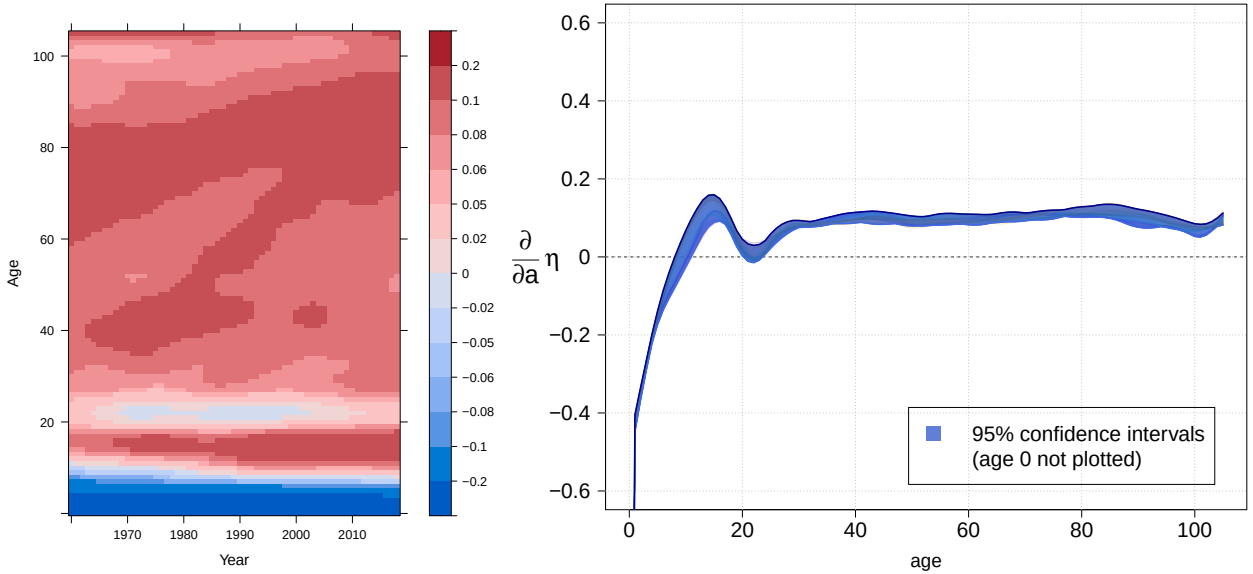


Observing mortality patterns (over ages)

Actual and Smooth log-mortality

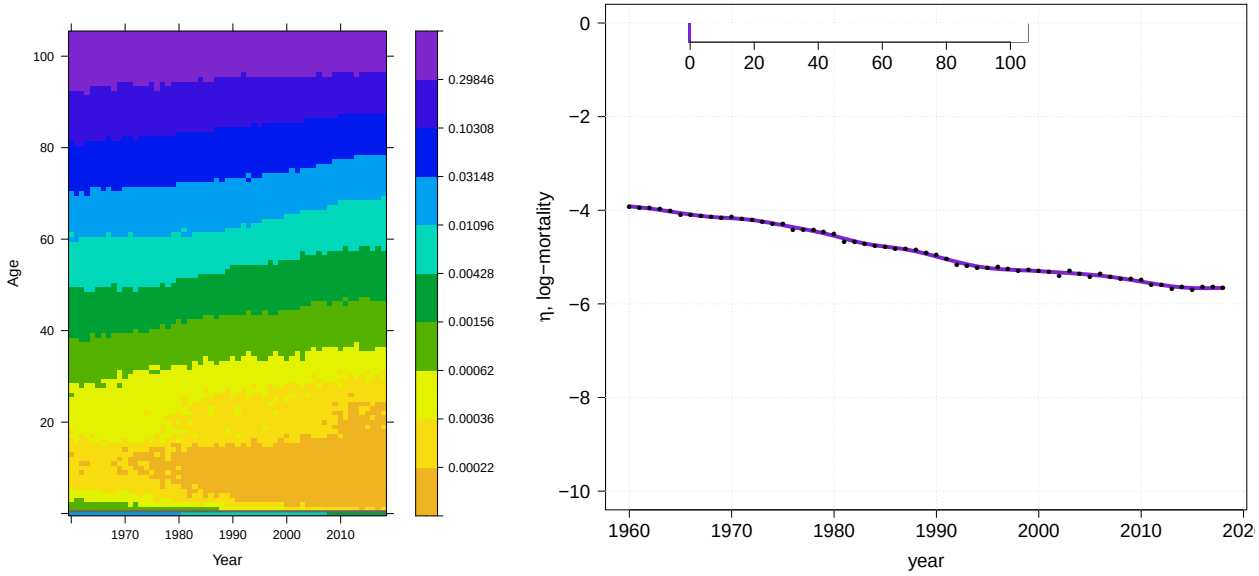


Rate-of-aging from smooth mortality

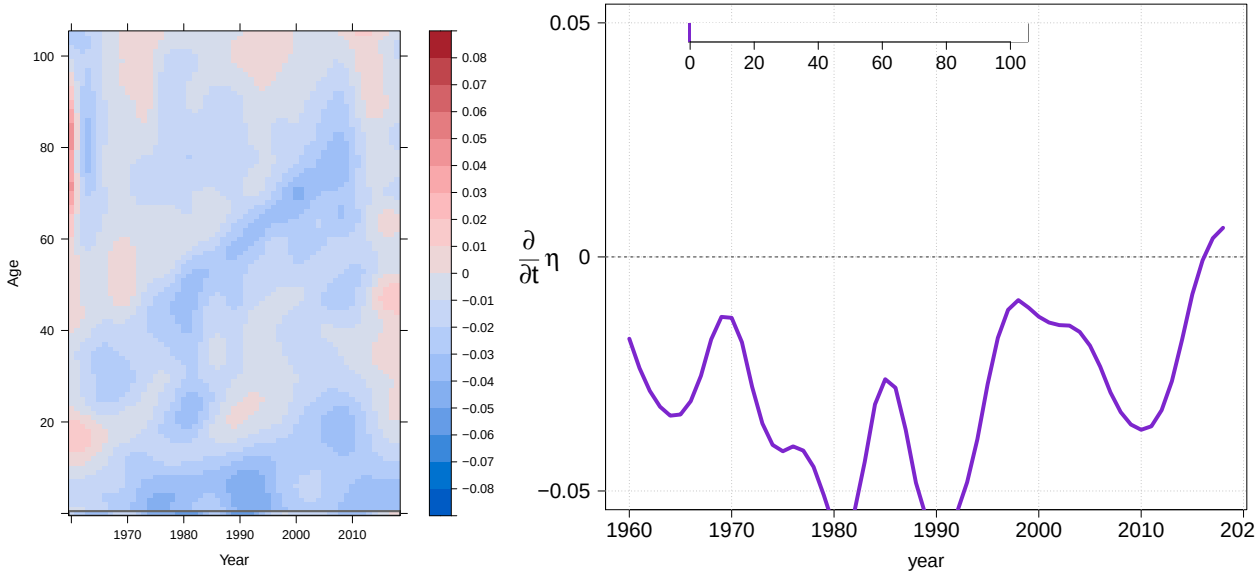


Observing mortality patterns (over years)

Actual and Smooth log-mortality

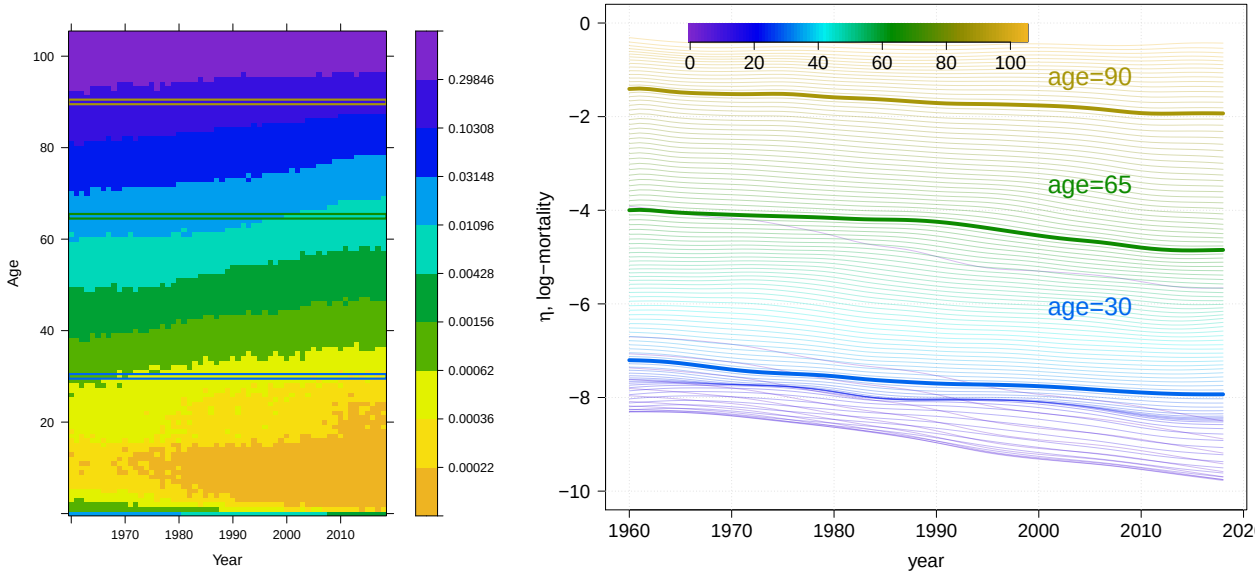


Rate-of-change from smooth mortality

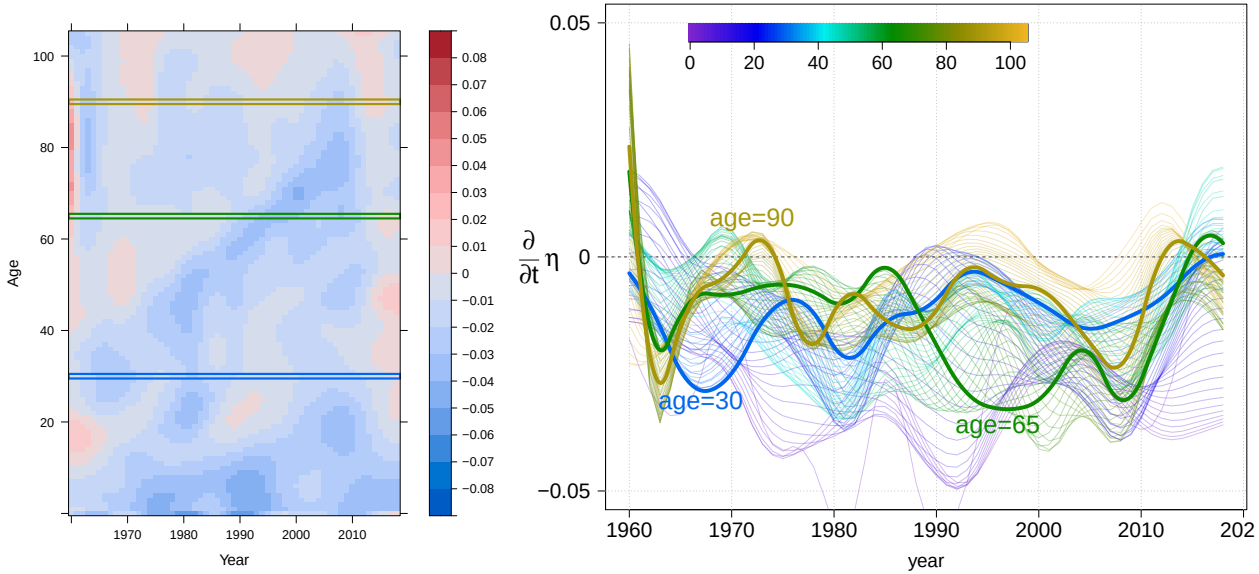


Observing mortality patterns (over years)

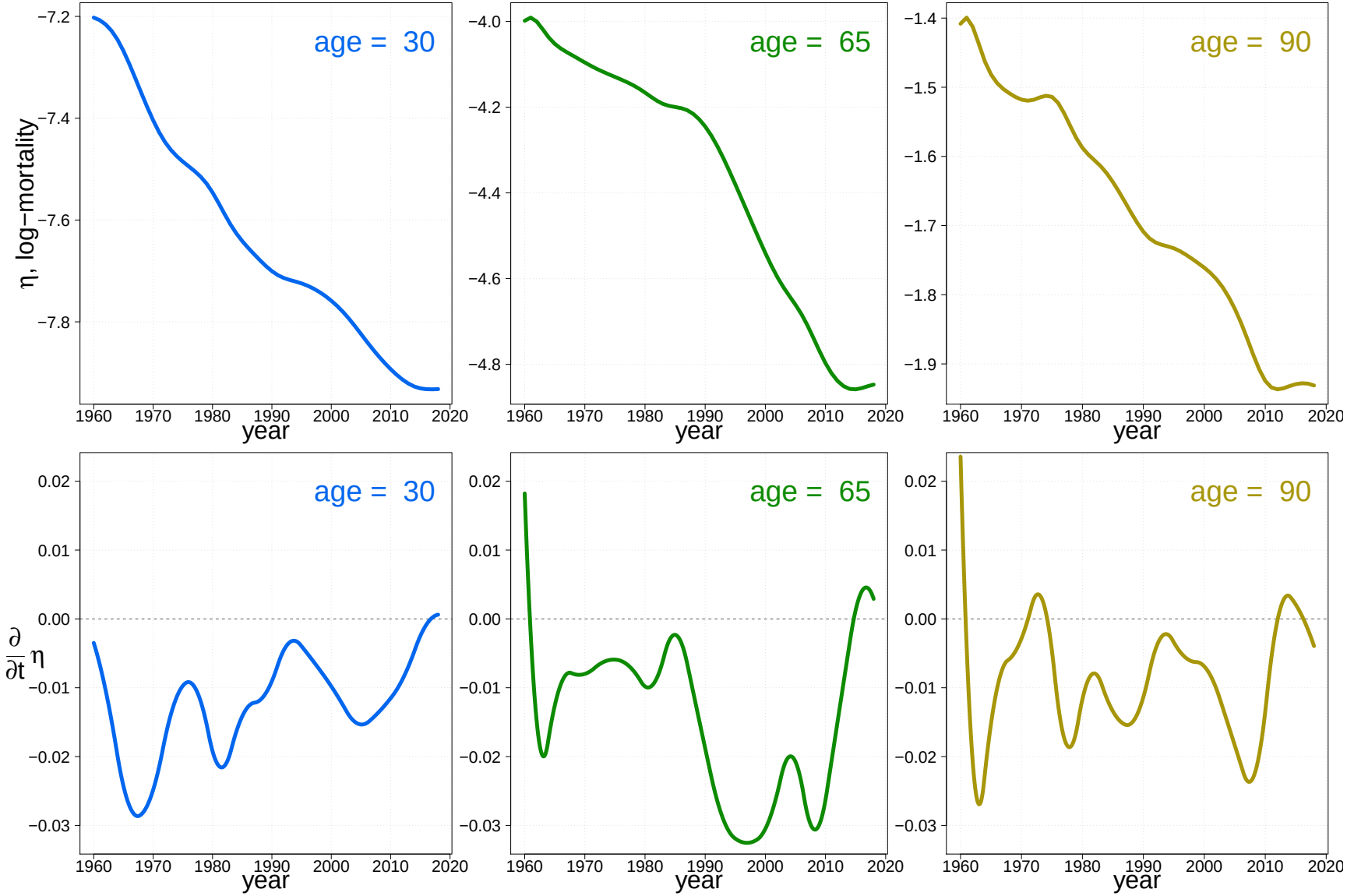
Actual and Smooth log-mortality



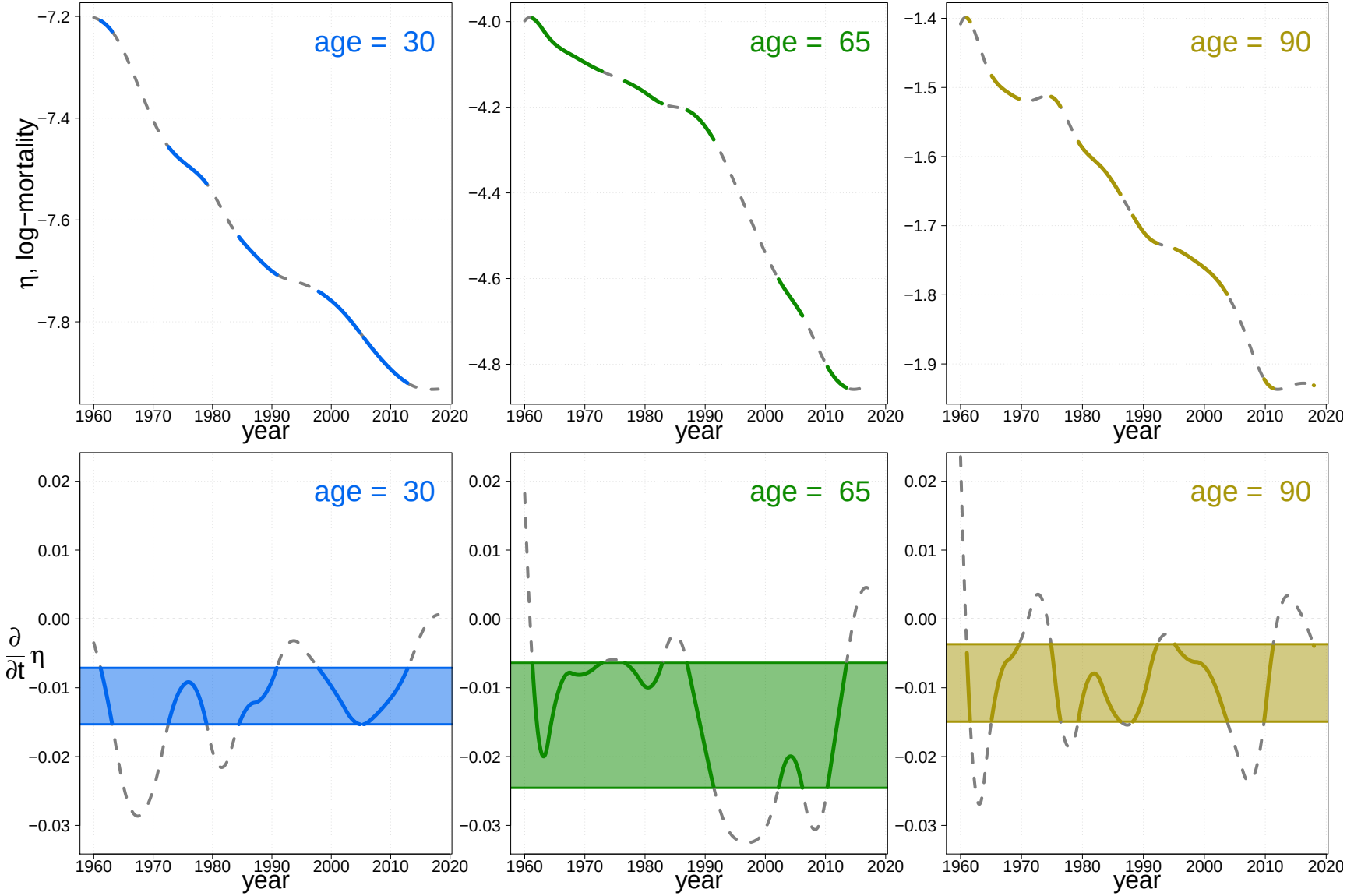
Rate-of-change from smooth mortality



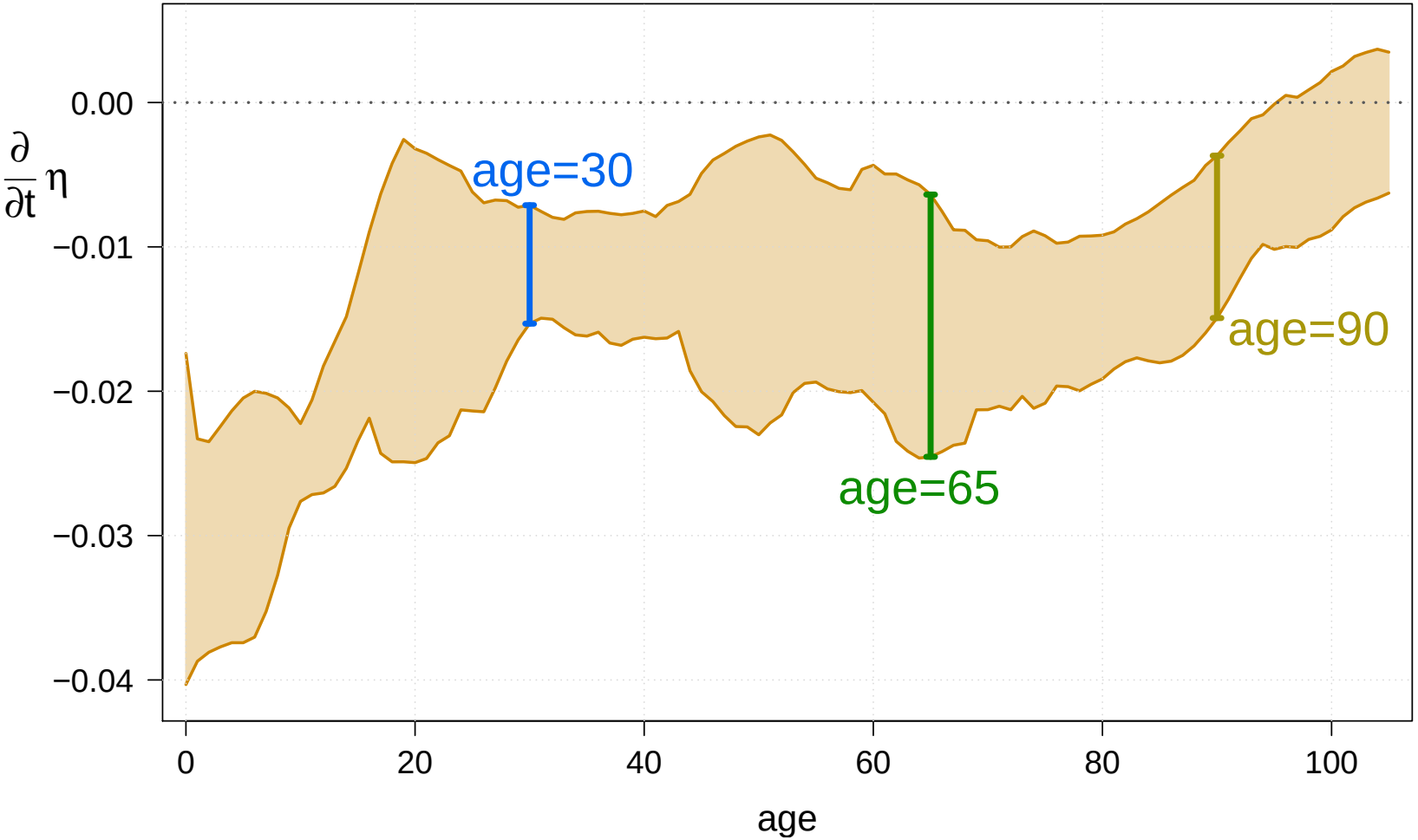
Observing mortality patterns (over years)



Observing mortality patterns (over years)



Observing mortality patterns (over years)



Asymmetric penalty over age and time

- For illustrative purposes, we show constraints for the age shape
- $\boldsymbol{\delta}_L^a$ and $\boldsymbol{\delta}_U^a$: lower and upper bounds of CI of the derivatives
- Keep same constraints for all years $\Rightarrow$ augment $\boldsymbol{\delta}$ over both dimensions:

$$\begin{aligned} \boldsymbol{g}_L^a &= \mathbf{1}_{n_1+n_2} \otimes \boldsymbol{\delta}_L^a \\ \boldsymbol{g}_U^a &= \mathbf{1}_{n_1+n_2} \otimes \boldsymbol{\delta}_U^a \end{aligned}$$

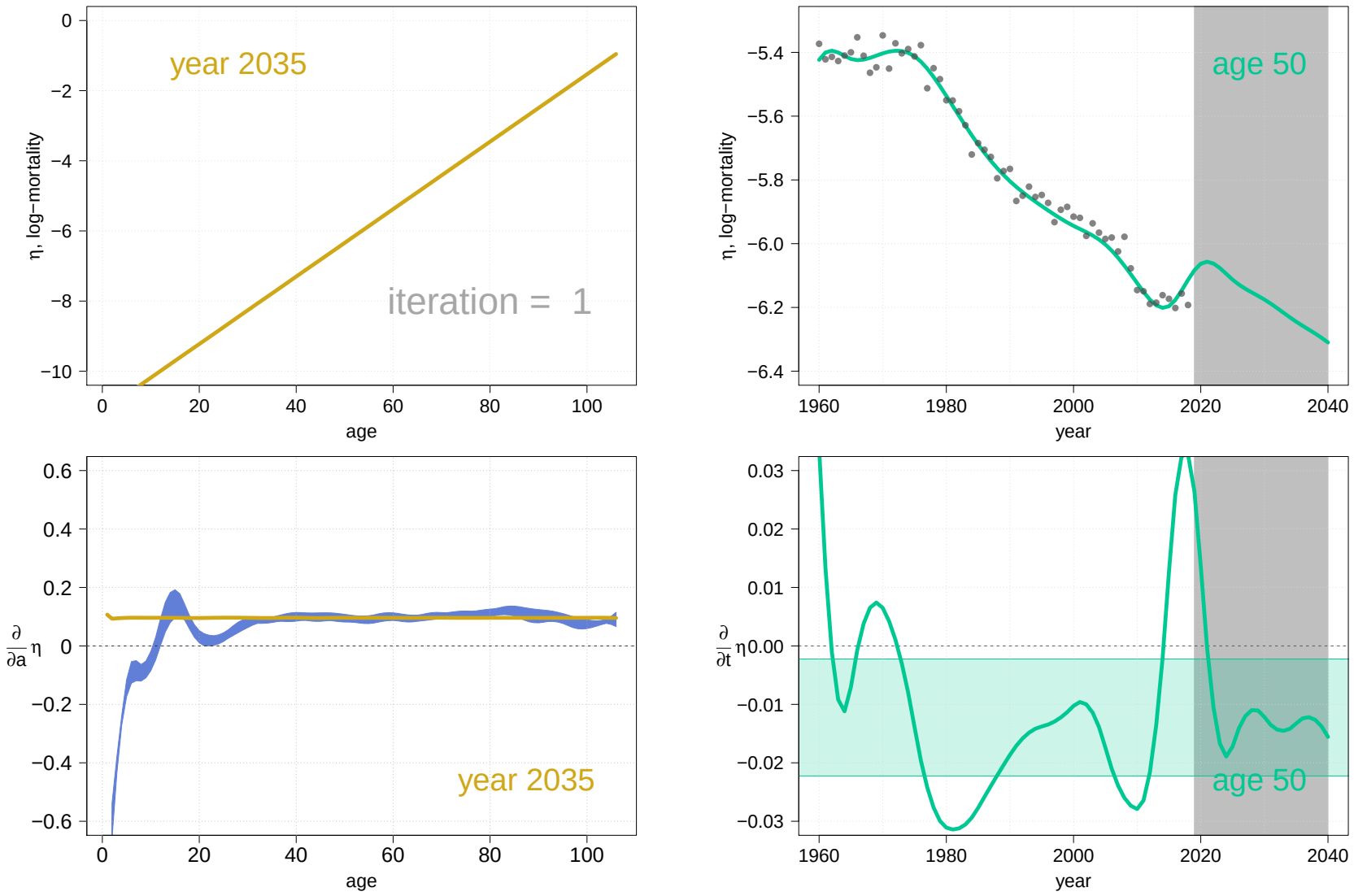
- Two asymmetric penalty terms:

$$\boldsymbol{P}^a = \boldsymbol{P}_L^a + \boldsymbol{P}_U^a \qquad \text{and} \qquad \boldsymbol{p}^a = \boldsymbol{p}_L^a + \boldsymbol{p}_U^a$$

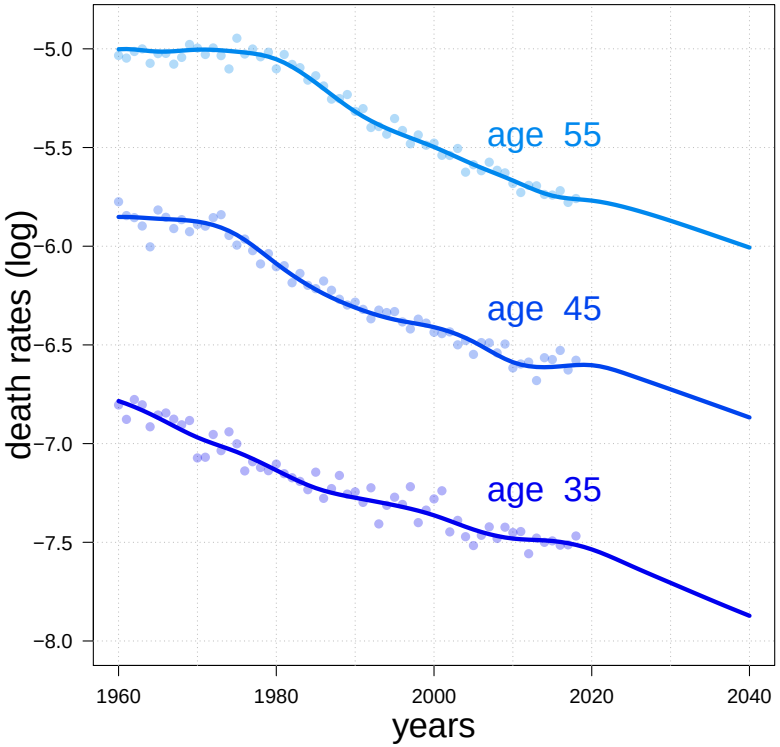
  which are a 2D generalization of previous equations (see slide 29)

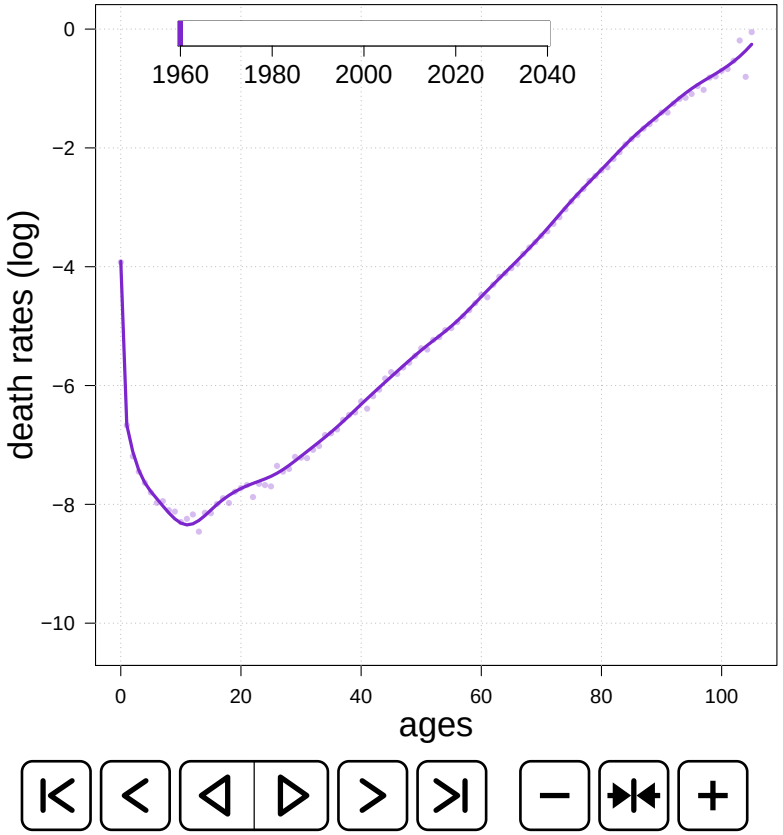
- Similar computation is performed over years

CP-splines: outcomes

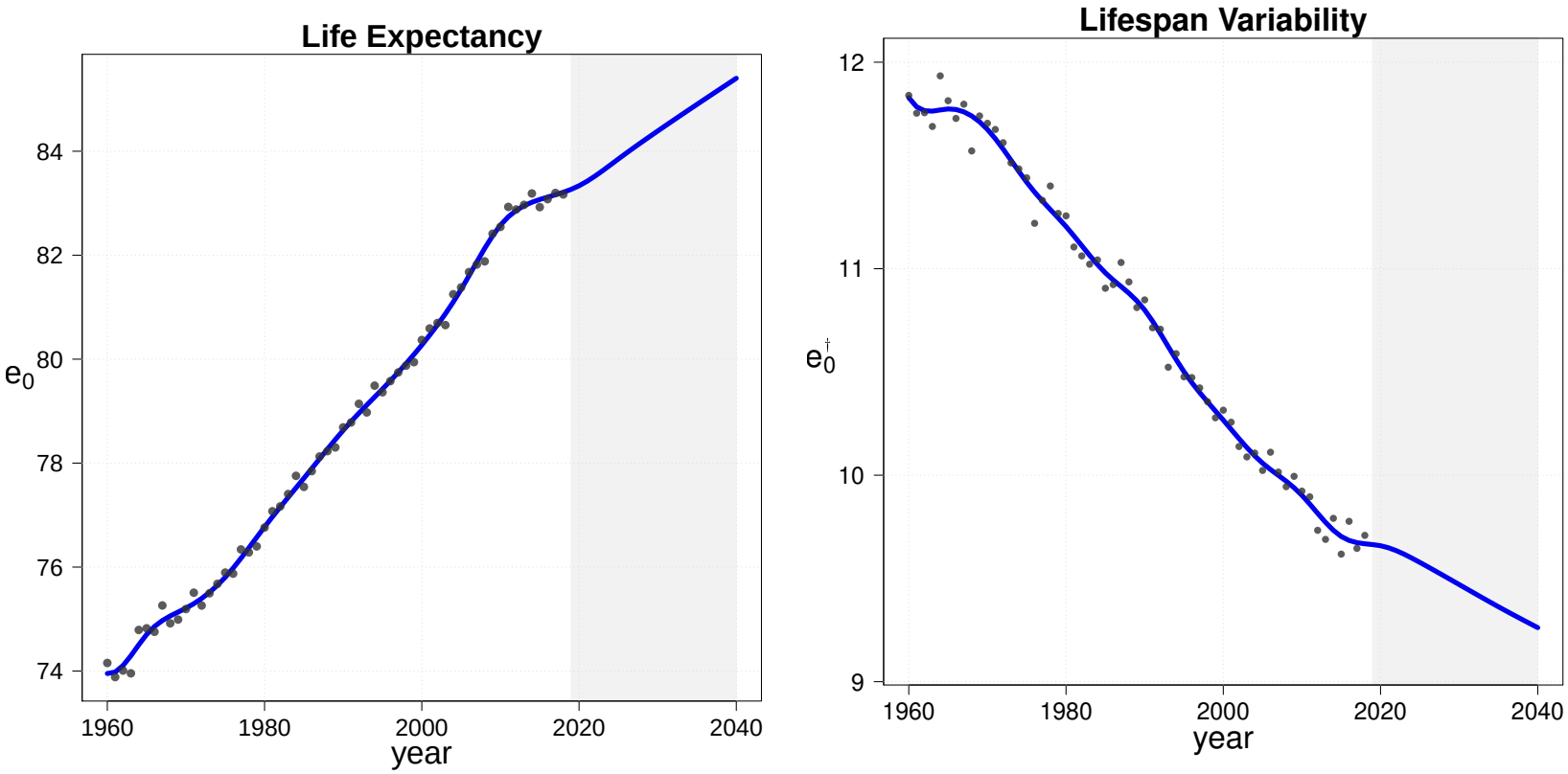


CP-splines: outcomes





CP-splines: outcomes



Life expectancy (e_0, left) and a measure of lifespan variability ($e_0^\dagger$, right). CP-splines.
England & Wales, ages 0-105, years 1960-2018, forecast up to 2040
Source: Human Mortality Database.

What we haven't shown here, . . .

- . . . but you can find in the paper below:
  - Other equations
  - Applications to more populations
  - Bootstrap procedure to obtain confidence intervals
  - Out-of-sample performance
  - Comparison with other alternative methods
  - Effect of changing time-window on the outcomes
  - Sensitivity analysis on confidence level in rate-of-change over t
  - Reproducible R-code with a GLAM framework

Camarda, C. G. (2019).
Smooth Constrained Mortality Forecasting.
Demographic Research. **41** (38), 1091-1130