The role, or non-role, of constraints in the forecasting of mortality

Iain Currie

Heriot-Watt University
and
Longevitas

Longevity 14

Amsterdam, September, 2018

The Identifiability Problem

- Cairns et al, 2009 NAAJ, “identifiability problem”. These constraints “ensure the fitted cohort effect will fluctuate around 0 with no discernible linear trend” and hence can be forecast.

The Identifiability Problem

- Cairns et al, 2009 NAAJ, “identifiability problem”. These constraints “ensure the fitted cohort effect will fluctuate around 0 with no discernible linear trend” and hence can be forecast.
- Richards, Currie, Kleinow and Ritchie (2017). “constraints on the cohort parameters perform a dual purpose
  - acting as identifiability constraints

The Identifiability Problem

- Cairns et al, 2009 NAAJ, “identifiability problem”. These constraints “ensure the fitted cohort effect will fluctuate around 0 with no discernible linear trend” and hence can be forecast.
- Richards, Currie, Kleinow and Ritchie (2017). “constraints on the cohort parameters perform a dual purpose
  - acting as identifiability constraints
  - imposing behaviour on cohort effects to make forecasting assumptions valid”

The Identifiability Problem

- Cairns et al, 2009 NAAJ, "identifiability problem". These constraints "ensure the fitted cohort effect will fluctuate around 0 with no discernible linear trend" and hence can be forecast.
- Richards, Currie, Kleinow and Ritchie (2017). "constraints on the cohort parameters perform a dual purpose
  - acting as identifiability constraints
  - imposing behaviour on cohort effects to make forecasting assumptions valid"

The purpose of this paper is to show

- there is **no identifiability problem**

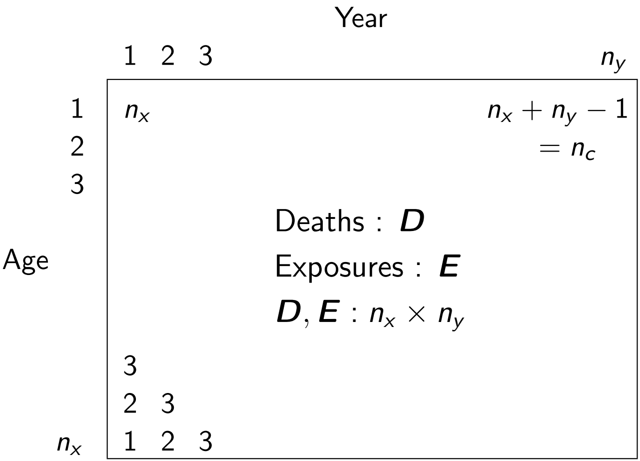
The Identifiability Problem

- Cairns et al, 2009 NAAJ, "identifiability problem". These constraints "ensure the fitted cohort effect will fluctuate around 0 with no discernible linear trend" and hence can be forecast.
- Richards, Currie, Kleinow and Ritchie (2017). "constraints on the cohort parameters perform a dual purpose
  - acting as identifiability constraints
  - imposing behaviour on cohort effects to make forecasting assumptions valid"

The purpose of this paper is to show

- there is **no identifiability problem**
- **forecasting does not depend** on the choice of constraints

Mortality data



Age of death = rows, Year of death = columns, Year of birth = diagonals

Illustrative Data

ONS: UK males: Ages: 50-104; Years: 1971-2015.

$n_x = 55$; $n_y = 45$; $n_c = 99$; $N = n_x n_y = 2475$.

Outline

Let $\mu_{i,j}$ be the **force of mortality** at age i in year j.

Outline

Let $\mu_{i,j}$ be the **force of mortality** at age i in year j.

- The **Age-Period (AP)** model is

$$\log \mu_{i,j} = \alpha_i + \kappa_j.$$

Outline

Let $\mu_{i,j}$ be the **force of mortality** at age i in year j.

- The **Age-Period (AP)** model is

$$\log \mu_{i,j} = \alpha_i + \kappa_j.$$

- The **Age-Period-Cohort (APC)** model is

$$\log \mu_{i,j} = \alpha_i + \kappa_j + \gamma_{c(i,j)}$$

  where $c(i,j) = n_x - i + j$.

Outline

Let $\mu_{i,j}$ be the **force of mortality** at age i in year j.

- The **Age-Period (AP)** model is

$$\log \mu_{i,j} = \alpha_i + \kappa_j.$$

- The **Age-Period-Cohort (APC)** model is

$$\log \mu_{i,j} = \alpha_i + \kappa_j + \gamma_{c(i,j)}$$

  where $c(i,j) = n_x - i + j$.

- The **Age-Period-Cohort-Improvement (APCI)** model is

$$\log \mu_{i,j} = \alpha_i + \kappa_j + \gamma_{c(i,j)} + \beta_i(y_j - \bar{y}).$$

  This is used by the CMI to parameterise its forecasting spreadsheet.

Generalized linear models (GLMs)

Let $\boldsymbol{D} = (d_{i,j})$ and $\boldsymbol{E} = (e_{i,j})$ and assume

$$d_{i,j} \sim \mathcal{P}(e_{i,j}\mu_{i,j})$$

where, for example, in AP model $\log \mu_{i,j} = \alpha_i + \kappa_j$.

Generalized linear models (GLMs)

Let $\boldsymbol{D} = (d_{i,j})$ and $\boldsymbol{E} = (e_{i,j})$ and assume

$$d_{i,j} \sim \mathcal{P}(e_{i,j}\mu_{i,j})$$

where, for example, in AP model $\log \mu_{i,j} = \alpha_i + \kappa_j$.

In vector form

$$\log\left(\mathbb{E}(\boldsymbol{d})\right) = \boldsymbol{e} + \boldsymbol{\mu} = \boldsymbol{e} + \boldsymbol{X}\boldsymbol{\theta}$$

where $\boldsymbol{X}$ is the model matrix.

Generalized linear models (GLMs)

Let $\boldsymbol{D} = (d_{i,j})$ and $\boldsymbol{E} = (e_{i,j})$ and assume

$$d_{i,j} \sim \mathcal{P}(e_{i,j}\mu_{i,j})$$

where, for example, in AP model $\log \mu_{i,j} = \alpha_i + \kappa_j$.

In vector form

$$\log\left(\mathbb{E}(\boldsymbol{d})\right) = \boldsymbol{e} + \boldsymbol{\mu} = \boldsymbol{e} + \boldsymbol{X\theta}$$

where $\boldsymbol{X}$ is the model matrix.

The AP, APC and APCI are all GLMs (overdispersion is ignored).

Identifiability and Rank

The Age-Period (AP) model is

$$\log \mu_{i,j} = \alpha_i + \kappa_j, \quad \boldsymbol{\theta}' = (\boldsymbol{\alpha}', \boldsymbol{\kappa}').$$

Identifiability and Rank

The Age-Period (AP) model is

$$\log \mu_{i,j} = \alpha_i + \kappa_j, \quad \boldsymbol{\theta}' = (\boldsymbol{\alpha}', \boldsymbol{\kappa}').$$

Model matrix $\boldsymbol{X}$ is $N \times (n_x + n_y)$ and has rank $n_x + n_y - 1$

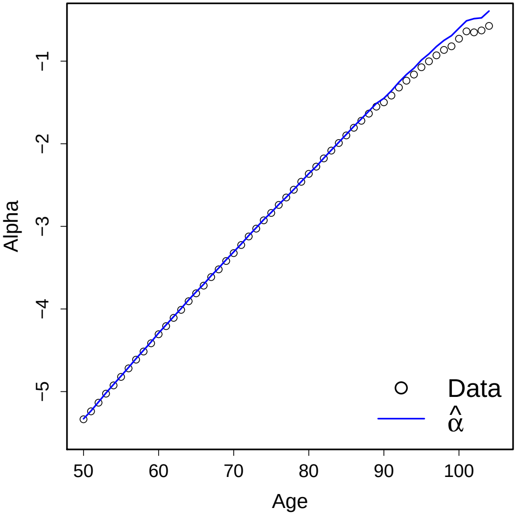
$\Rightarrow$ parameters are not uniquely estimable.

Identifiability problem

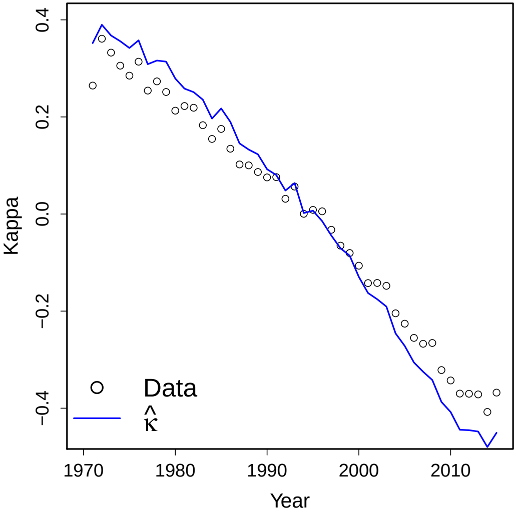
Constraints in AP model

Standard constraint: $\sum \kappa_j \;\; = \;\; 0$

$\Rightarrow$ parameters are uniquely estimable.



Alpha
−1
−2
−3
−4
−5
50
60
70
80
90
100
Age
Data
$\hat{\alpha}$



Kappa
Year
0.4
0.2
0.0
−0.2
−0.4
1970
1980
1990
2000
2010
Data
$\hat{\kappa}$

Comments

- Estimates have a natural interpretation.

Comments

- Estimates have a natural interpretation.
- Forecasting is simple
  - Forecast $\hat{\kappa}$
  - Use model to forecast $\mu_{i,j}$.

Comments

- Estimates have a natural interpretation.
- Forecasting is simple
  - Forecast $\hat{\kappa}$
  - Use model to forecast $\mu_{i,j}$.
- What could be simpler or more obvious?

Comments

- Estimates have a natural interpretation.
- Forecasting is simple
  - Forecast $\hat{\kappa}$
  - Use model to forecast $\mu_{i,j}$.
- What could be simpler or more obvious?

BUT

Random constraints in AP model

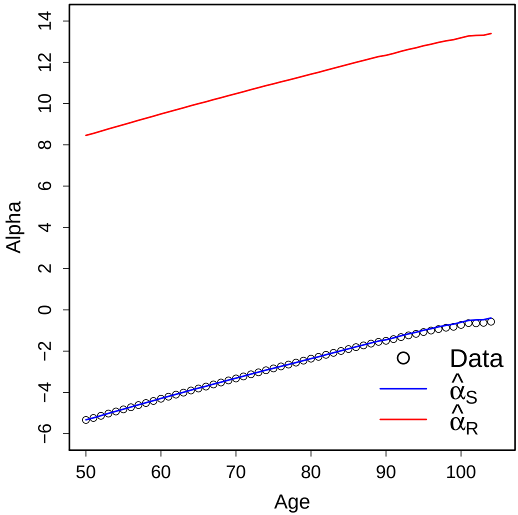
Let $\boldsymbol{\theta}' = (\boldsymbol{\alpha}', \boldsymbol{\kappa}')'$.

$$\text{Random constraint:} \quad \sum_{1}^{n_x+n_y} u_i \theta_i \;\; = \;\; 0$$

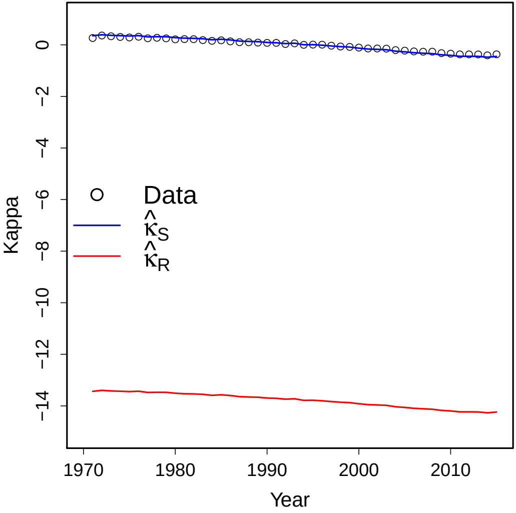
where $U_i \sim \mathcal{U}(0,1)$.

$\Rightarrow$ parameters are uniquely estimable.

Standard: $\hat{\alpha}_S$, Random: $\hat{\alpha}_R$



Standard: $\hat{\boldsymbol{\kappa}}_S$, Random: $\hat{\boldsymbol{\kappa}}_R$



Properties

Standard (centred) estimates: $\hat{\boldsymbol{\theta}}_S = (\hat{\boldsymbol{\alpha}}'_S, \hat{\boldsymbol{\kappa}}'_S)'$
Random estimates: $\hat{\boldsymbol{\theta}}_R = (\hat{\boldsymbol{\alpha}}'_R, \hat{\boldsymbol{\kappa}}'_R)'$.
Define

$$\boxed{\boldsymbol{\Delta}\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}_S - \hat{\boldsymbol{\alpha}}_R, \quad \boldsymbol{\Delta}\hat{\boldsymbol{\kappa}} = \hat{\boldsymbol{\kappa}}_S - \hat{\boldsymbol{\kappa}}_R}$$

Properties

Standard (centred) estimates: $\hat{\boldsymbol{\theta}}_S = (\hat{\boldsymbol{\alpha}}'_S, \hat{\boldsymbol{\kappa}}'_S)'$
Random estimates: $\hat{\boldsymbol{\theta}}_R = (\hat{\boldsymbol{\alpha}}'_R, \hat{\boldsymbol{\kappa}}'_R)'$.
Define

$$\boldsymbol{\Delta}\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}_S - \hat{\boldsymbol{\alpha}}_R, \quad \boldsymbol{\Delta}\hat{\boldsymbol{\kappa}} = \hat{\boldsymbol{\kappa}}_S - \hat{\boldsymbol{\kappa}}_R$$

- Invariance: $\boldsymbol{X}\hat{\boldsymbol{\theta}}_S = \boldsymbol{X}\hat{\boldsymbol{\theta}}_R$, i.e., fitted $\mu_{i,j}$ equal

Properties

Standard (centred) estimates: $\hat{\boldsymbol{\theta}}_S = (\hat{\boldsymbol{\alpha}}'_S, \hat{\boldsymbol{\kappa}}'_S)'$
Random estimates: $\hat{\boldsymbol{\theta}}_R = (\hat{\boldsymbol{\alpha}}'_R, \hat{\boldsymbol{\kappa}}'_R)'$.
Define

$$\boxed{\boldsymbol{\Delta}\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}_S - \hat{\boldsymbol{\alpha}}_R, \quad \boldsymbol{\Delta}\hat{\boldsymbol{\kappa}} = \hat{\boldsymbol{\kappa}}_S - \hat{\boldsymbol{\kappa}}_R}$$

- Invariance: $\boldsymbol{X}\hat{\boldsymbol{\theta}}_S = \boldsymbol{X}\hat{\boldsymbol{\theta}}_R$, i.e., fitted $\mu_{i,j}$ equal
- $\boldsymbol{\Delta}\hat{\boldsymbol{\alpha}} = k\mathbf{1}_{n_x}$, i.e., $\hat{\boldsymbol{\alpha}}_S$ and $\hat{\boldsymbol{\alpha}}_R$ are
  - parallel
  - k apart

Properties

Standard (centred) estimates: $\hat{\boldsymbol{\theta}}_S = (\hat{\boldsymbol{\alpha}}'_S, \hat{\boldsymbol{\kappa}}'_S)'$
Random estimates: $\hat{\boldsymbol{\theta}}_R = (\hat{\boldsymbol{\alpha}}'_R, \hat{\boldsymbol{\kappa}}'_R)'$.
Define

$$\boldsymbol{\Delta}\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}_S - \hat{\boldsymbol{\alpha}}_R, \quad \boldsymbol{\Delta}\hat{\boldsymbol{\kappa}} = \hat{\boldsymbol{\kappa}}_S - \hat{\boldsymbol{\kappa}}_R$$

- Invariance: $\boldsymbol{X}\hat{\boldsymbol{\theta}}_S = \boldsymbol{X}\hat{\boldsymbol{\theta}}_R$, i.e., fitted $\mu_{i,j}$ equal
- $\boldsymbol{\Delta}\hat{\boldsymbol{\alpha}} = k\mathbf{1}_{n_x}$, i.e., $\hat{\boldsymbol{\alpha}}_S$ and $\hat{\boldsymbol{\alpha}}_R$ are
  - parallel
  - k apart
- $\boldsymbol{\Delta}\hat{\boldsymbol{\kappa}} = -k\mathbf{1}_{n_y}$, i.e., $\hat{\boldsymbol{\kappa}}_S$ and $\hat{\boldsymbol{\kappa}}_R$ are
  - parallel
  - $-k$ apart

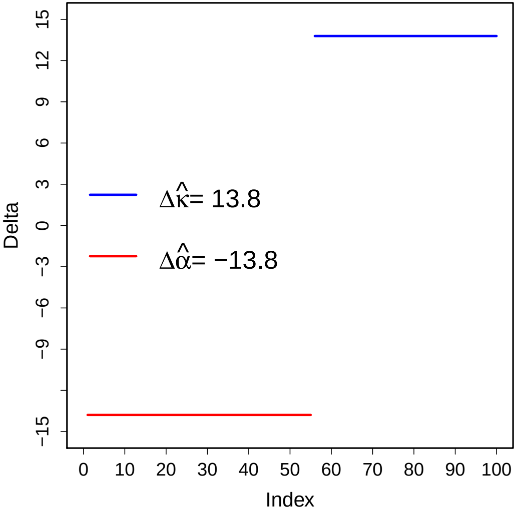
Properties

Standard (centred) estimates: $\hat{\boldsymbol{\theta}}_S = (\hat{\boldsymbol{\alpha}}'_S, \hat{\boldsymbol{\kappa}}'_S)'$
Random estimates: $\hat{\boldsymbol{\theta}}_R = (\hat{\boldsymbol{\alpha}}'_R, \hat{\boldsymbol{\kappa}}'_R)'$.
Define

$$\boldsymbol{\Delta}\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}_S - \hat{\boldsymbol{\alpha}}_R, \quad \boldsymbol{\Delta}\hat{\boldsymbol{\kappa}} = \hat{\boldsymbol{\kappa}}_S - \hat{\boldsymbol{\kappa}}_R$$

- Invariance: $\boldsymbol{X}\hat{\boldsymbol{\theta}}_S = \boldsymbol{X}\hat{\boldsymbol{\theta}}_R$, i.e., fitted $\mu_{i,j}$ equal
- $\boldsymbol{\Delta}\hat{\boldsymbol{\alpha}} = k\mathbf{1}_{n_x}$, i.e., $\hat{\boldsymbol{\alpha}}_S$ and $\hat{\boldsymbol{\alpha}}_R$ are
  - parallel
  - k apart
- $\boldsymbol{\Delta}\hat{\boldsymbol{\kappa}} = -k\mathbf{1}_{n_y}$, i.e., $\hat{\boldsymbol{\kappa}}_S$ and $\hat{\boldsymbol{\kappa}}_R$ are
  - parallel
  - $-k$ apart
- Here $k = -13.8$.

$\Delta\hat{\alpha} = \hat{\alpha}_S - \hat{\alpha}_R$, etc



Forecasting

Forecasting with ARIMA model, e.g., random walk with drift, is invariant wrt choice of constraints.

Age-Period-Cohort Model

The APC model is

$$\log \mu_{i,j} = \alpha_i + \kappa_j + \gamma_{c(i,j)}$$

where $c(i,j) = n_x - i + j$.

Constraints

Standard (Cairns et al, 2009):

$$\sum \kappa_j = \sum \gamma_c = \sum w_c \gamma_c = 0$$

where w_c is the cohort index, $w_c = 1, \ldots, n_c$.

Constraints

Standard (Cairns et al, 2009):

$$\sum \kappa_j = \sum \gamma_c = \sum w_c \gamma_c = 0$$

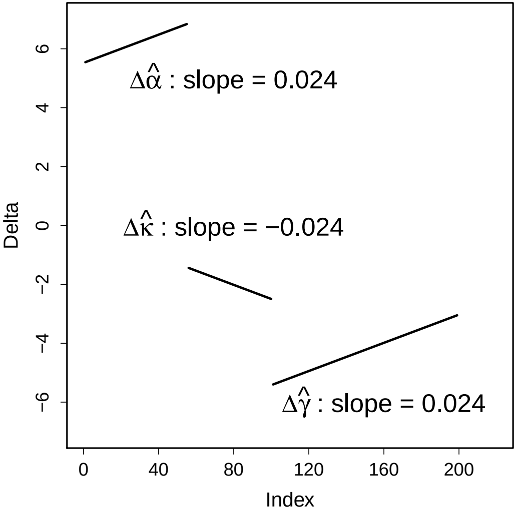
where w_c is the cohort index, $w_c = 1, \ldots, n_c$.

Random: Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\kappa}', \boldsymbol{\gamma}')'$.

$$\sum u_{1,j}\theta_j = \sum u_{2,j}\theta_j = \sum u_{3,j}\theta_j = 0$$

where the $u_{i,j}$, $i = 1, 2, 3$, $j = 1, \ldots, n_x + n_y + n_c$, are $\mathcal{U}(0,\ 1)$.

$\boldsymbol{\Delta}\hat{\alpha} = \hat{\alpha}_S - \hat{\alpha}_R$, etc



Forecasting

- Forecasting with ARIMA model, e.g., random walk with drift, is invariant wrt choice of constraints.

Age-Period-Cohort-Improvement (APCI) Model

The model is

$$\log \mu_{i,j} = \alpha_i + \kappa_j + \gamma_{c(i,j)} + \beta_i(y_j - \bar{y})$$

and forms the basis for the CMI's current forecasting spreadsheet.

Age-Period-Cohort-Improvement (APCI) Model

The model is

$$\log \mu_{i,j} = \alpha_i + \kappa_j + \gamma_{c(i,j)} + \beta_i(y_j - \bar{y})$$

and forms the basis for the CMI's current forecasting spreadsheet.

Model matrix $\boldsymbol{X}$ is $N \times (3n_x + 2n_y - 1)$ and rank $3n_x + 2n_y - 6$

and five (5) constraints are required to bring about identifiability.

Constraints

Standard:

$$\sum \kappa_j = \sum \gamma_c = \sum w_c \gamma_c = \sum w_c^2 \gamma_c = \sum j \kappa_j = 0$$

where w_c is the cohort index, $w_c = 1, \ldots, n_c$.

Constraints

Standard:

$$\sum \kappa_j = \sum \gamma_c = \sum w_c \gamma_c = \sum w_c^2 \gamma_c = \sum j \kappa_j = 0$$

where w_c is the cohort index, $w_c = 1, \ldots, n_c$.

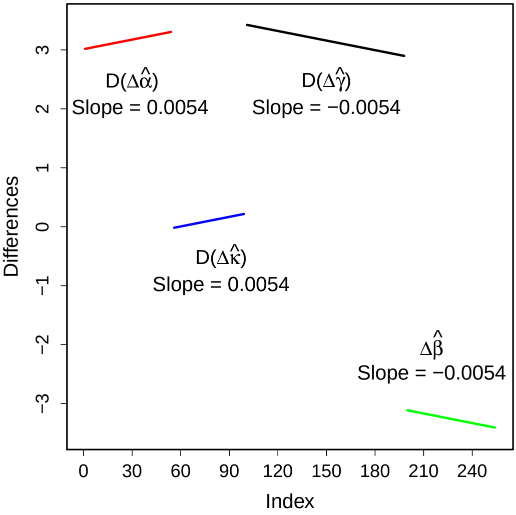
Random: Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\kappa}', \boldsymbol{\gamma}', \boldsymbol{\beta}')'$.

$$\sum u_{i,j} \theta_j = 0, \; i = 1, \ldots, 5, \; j = 1, \ldots, 2n_x + n_y + n_c,$$

where the $u_{i,j}$, are $\mathcal{U}(0,\ 1)$.

$$\boldsymbol{\Delta}\hat{\alpha} = \hat{\alpha}_S - \hat{\alpha}_R, \text{ etc}$$

D = difference operator



Forecasting

- Forecasting with ARIMA model is invariant wrt choice of constraints provided $d \geq 3$ in ARIMA model.

Smoothing

In AP and APC models smooth $\boldsymbol{\alpha}$. Set

$$\boxed{\boldsymbol{\alpha} = \boldsymbol{B}_a \boldsymbol{a}}$$

where $\boldsymbol{B}_a$ is a regression matrix of B-splines for age.

Smoothing

In AP and APC models smooth $\boldsymbol{\alpha}$. Set

$$\boxed{\boldsymbol{\alpha} = \boldsymbol{B}_a \boldsymbol{a}}$$

where $\boldsymbol{B}_a$ is a regression matrix of B-splines for age.

In APCI model additionally smooth $\boldsymbol{\beta}$. Set

$$\boxed{\boldsymbol{\beta} = \boldsymbol{B}_a \boldsymbol{b}}.$$

Use method of P-splines (Eilers & Marx, 1996).

Conclusions

- Smoothing α (and β) makes no difference.

Conclusions

- Smoothing $\boldsymbol{\alpha}$ (and $\boldsymbol{\beta}$) makes no difference.
- Fitted and forecast values are invariant wrt choice of constraints.

Conclusions

- Smoothing α (and β) makes no difference.
- Fitted and forecast values are invariant wrt choice of constraints.
- Order of
  - penalty for smoothing and
  - differencing in ARIMA model

  must be sufficiently large (see Currie (in preparation) for details).

References

1. Cairns, Blake, Dowd et al (2009) A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, **13**, 1–35.
2. Currie (2013) Smoothing constrained generalized linear models with an application to the Lee-Carter model. *Statistical Modelling*, **13**, 69-93.
3. Currie (in preparation) Constraints, the identifiability problem and the forecasting of mortality.
4. Richards, Currie, Kleinow & Ritchie (to appear). A stochastic implementation of the APCI model for mortality projections.