



Heriot-Watt University
Research Gateway

A black box approach to fitting smooth models of mortality

Citation for published version:

Currie, ID 2023, 'A black box approach to fitting smooth models of mortality', *Statistical Modelling*, vol. 23, no. 5-6, pp. 456-464. <https://doi.org/10.1177/1471082X231181165>

Digital Object Identifier (DOI):

[10.1177/1471082X231181165](https://doi.org/10.1177/1471082X231181165)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Statistical Modelling

Publisher Rights Statement:

© 2023 The Author(s).

General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

A black box approach to fitting smooth models of mortality

Iain D Currie[†]

¹Heriot-Watt University, Edinburgh, Scotland

Abstract: Actuaries have long been interested in the forecasting of mortality for the purpose of the pricing and reserving of pensions and annuities. Most models of mortality in age and year of death, and often year of birth, are not identifiable so actuaries worried about what constraints should be used to give sensible estimates of the age and year of death parameters, and, if required, the year of birth parameters. These parameters were then forecast with an ARIMA model to give the required forecasts of mortality. A recent article showed that, while the fitted parameters were not identifiable, both the fitted and forecast mortalities were. This result holds if the age term is smoothed with P-splines. The present article deals with generalized linear models with a rank deficient regression matrix. We have two aims. First, we investigate the effect that different constraints have on the estimated regression coefficients. We show that it is possible to fit the model under different constraints in R without imposing any explicit constraints. R does all the necessary booking-keeping ‘under the bonnet’. The estimated regression coefficients under a particular set of constraints can then be recovered from the invariant fitted values. We have a black box approach to fitting the model subject to any set of constraints.

Key words: constraints, identifiability, invariance, mortality

Received March 2022; accepted April 2022

1 Introduction

Actuaries have long been interested in the forecasting of human mortality for the purpose of the pricing and reserving of pensions and annuities. Gompertz (1825), in a landmark paper, gave the first mathematical model of mortality. He only had death rates by age at ten year intervals but he observed that these rates were linear on a log scale. This enabled him to interpolate for all ages for the range of ages he had available. This was the celebrated Gompertz law of mortality.

At the present time extensive mortality data are available not only by single ages but also by year of death. The Human Mortality Database is a comprehensive resource of mortality data and gives mortality data (a) by age and year of death in a calendar year, and (b) an estimate of the mid-year number of lives exposed to the risk of death in that year. These data are available on over 40 countries for the male and female populations of these countries.

[†]Iain Currie passed away on 25 May 2022.

Address for correspondence: Paul Eilers, Erasmus University Medical Centre, 's-Gravendijkwal 230, 3015 CE Rotterdam, the Netherlands.

E-mail: p.eilers@erasmusmc.nl



These richer datasets gave rise to the possibility not only of modelling mortality by age and year but also the forecasting of mortality. Such models were in terms of age and year of death, and often year of birth. The idea was to estimate the parameters in the models and then forecast the year of death parameters and, if required, the year of birth parameters; see Cairns et al. (2009) for a comprehensive treatment of such models. However, there was a problem—most models were not identifiable so it was not obvious how to obtain parameter estimates that would give sensible forecasts. Various sets of linear constraints on the parameters were proposed that would solve this problem.

Most models were generalized linear models or GLMs. Currie (2020) showed that, while the fitted parameters were not identifiable, the fitted mortalities were identifiable (a fundamental result in regression modelling) but crucially the forecast values of mortality were also identifiable when forecasting was done with an ARIMA model. This article illustrated this result by considering two sets of constraints, one a standard set of constraints found in the literature (see Cairns et al., 2009, for example) and the other a set of random constraints. A general proof of the result was also provided.

The purpose of the present article is twofold: first, to show that we can ignore the explicit use of constraints completely and use R's `lm` and `glm` commands to fit the models, and second, to use the invariance of the fitted mortalities to explore the relationship between the estimates of the parameters under different sets of constraints. R does all the necessary book-keeping 'under the bonnet'. We have a black box approach to the fitting and forecasting of mortality.

We do not present any general theory but rather present our method through a single model, the Age-Period-Cohort model. We consider this model in a basic form and also in a smooth form where we use P-splines (Eilers and Marx, 1996) to smooth some terms in the model. It should be clear that our method applies more generally.

2 Data and notation

We illustrate our ideas with data from the Human Mortality Database on UK males (data downloaded February 2022). We adopt the convention that matrices are denoted by upper case bold font as in \mathbf{M} and column vectors by lower case bold font as in \mathbf{m} . The data comprise two matrices: $\mathbf{D}_{obs} = (d_{x,y})$, the matrix of the number of deaths at age x in year y , and $\mathbf{E}_{obs} = (e_{x,y})$, the total time lived or central exposure at age x in year y . (We will use \mathbf{D} later to denote a difference matrix.) In general we suppose we have n_a ages and n_y years. We denote the age indices by $\mathbf{x}_a = (1, \dots, n_a)'$ and the year indices by $\mathbf{x}_y = (1, \dots, n_y)'$ where the $'$ indicates the transpose of a vector (or matrix). We index the rows and columns of \mathbf{D}_{obs} and \mathbf{E}_{obs} by \mathbf{x}_a and \mathbf{x}_y respectively. Thus \mathbf{D}_{obs} and \mathbf{E}_{obs} are both $n_a \times n_y$. Further we index the cohorts or years of birth with $\mathbf{x}_c = (1, \dots, n_c)'$ where $n_c = n_a + n_y - 1$ is the number of distinct cohorts. We adopt the convention that the oldest cohort, ie, for age n_a in year 1, is indexed 1; hence, the cohort index for cell (x, y) is $c(x, y) = n_a - x + y$. In our case we have data on ages 50 to 90 and years 1970 to 2018. Hence $n_a = 41$, $n_y = 49$ and $n_c = 89$.

In addition, the following notation is useful: $\mathbf{1}_n$ denotes a column vector of 1s of length n and \mathbf{I}_n denotes the identity matrix of size n . Further we let $\mathbf{0}_n$ stand for either a column vector of 0s of length n or an $n \times n$ zero matrix; the context should make clear which applies. We may omit the suffix n if no confusion results. We will make frequent use of the Kronecker product which we denote \otimes . The function `vec(M)` stacks the columns of a matrix \mathbf{M} on top of each other in column order.

3 The Age-Period-Cohort model

Let $D_{x,y}$ be the random variable corresponding to the observed deaths $d_{x,y}$. We suppose that $D_{x,y}$ has the Poisson distribution with mean $e_{x,y}\lambda_{x,y}$ where $\lambda_{x,y}$ is the force of mortality or instantaneous death rate at age x in year y , ie, $D_{x,y} \sim \mathcal{P}(e_{x,y}\lambda_{x,y})$. There is a slight abuse of notation here; strictly we should write $\lambda_{x+1/2,y+1/2}$ since $e_{x,y}$ is the central exposure; we will use the simpler notation throughout for clarity. Let $\mathbf{d} = \text{vec}(\mathbf{D}_{obs})$ and $\mathbf{e} = \text{vec}(\mathbf{E}_{obs})$ be the stacked vectors of deaths and exposures; let $\boldsymbol{\lambda}$ be the corresponding vector of forces of mortality.

The Age-Period-Cohort or APC model is the simplest model which involves the age and year of death, and the year of birth. We define

$$\log \lambda_{x,y} = \alpha_x + \kappa_y + \gamma_{c(x,y)}, \quad x = 1, \dots, n_a, \quad y = 1, \dots, n_y, \tag{3.1}$$

where $c(x, y) = n_a - x + y$ is the cohort index for age x in year y . Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\kappa}', \boldsymbol{\gamma}')'$. The model matrix is

$$\mathbf{X} = [\mathbf{X}_a : \mathbf{X}_y : \mathbf{X}_c], \quad n_a n_y \times (2n_a + 2n_y - 1), \tag{3.2}$$

where $\mathbf{X}_a = \mathbf{1}_{n_y} \otimes \mathbf{I}_{n_a}$, $\mathbf{X}_y = \mathbf{I}_{n_y} \otimes \mathbf{1}_{n_a}$ and the row of \mathbf{X}_c corresponding to age x and year y contains a one in column $n_a - x + y$ and zeros elsewhere. With this setup we have $\log \boldsymbol{\lambda} = \mathbf{X}\boldsymbol{\theta}$. We have defined a generalized linear model or GLM with dependent variable \mathbf{d} , model matrix \mathbf{X} , Poisson errors and log link; the exposures \mathbf{e} enter into the model as an offset. Now \mathbf{X} is $n_a n_y \times (2n_a + 2n_y - 1)$ but its rank is $2n_a + 2n_y - 4$.

The usual actuarial approach is to choose three suitable constraints and then forecast the resulting estimates of $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}$. A common choice is

$$\sum_1^{n_y} \kappa_y = \sum_1^{n_c} \gamma_c = \sum_1^{n_c} c\gamma_c = 0 \tag{3.3}$$

see Cairns et al. (2009). We refer to (3.3) as the *standard constraints*. We can fit the APC model in R with the `glm` command. Here `Dth` and `Exp` contain the R variates holding the deaths \mathbf{d} and the exposures \mathbf{e} respectively. We have

```
Fit.APC <- glm(Dth ~ - 1 + X + offset(log(Exp)), family = poisson) \tag{3.4}
```

```
Theta <- Fit.APC$coef \tag{3.5}
```

```
Theta[is.na(Theta)] <- 0 \tag{3.6}
```

Inspection of `Fit.APC$coef` shows that the estimates of κ_{n_y} , $\gamma_{n_a+n_y-2}$ and $\gamma_{n_a+n_y-1}$ are returned as NA.

What is the connection between the approach to identifiability used by R and that often used by actuaries? We can mimic R's method with the constraints

$$\kappa_y = \gamma_{n_c-1} = \gamma_{n_c} = 0, \tag{3.7}$$

ie, we constrain the final period parameter and the final two cohort parameters to be zero. We refer to these constraints, implicit in R's approach, as the *R constraints*. We refit the APC model with the constraints (3.7). Currie (2013) gave the following generalization of the GLM scoring algorithm to fit a GLM with constraints $H\theta = \mathbf{0}$ but note that this algorithm plays no part in the treatment presented in this article

$$\begin{pmatrix} X' \tilde{W} X & H' \\ H & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\theta} \\ \hat{\omega} \end{pmatrix} = \begin{pmatrix} X' \tilde{W} \tilde{z} \\ \mathbf{0} \end{pmatrix}. \tag{3.8}$$

Here \tilde{W} is the diagonal matrix of weights, \tilde{z} is the working variable, $\tilde{\cdot}$ indicates a current estimate and $\hat{\cdot}$ denotes an updated estimate. The fitted coefficients are precisely those in `Fit.APC$coef` but with the NA values replaced by zeros. Equation (3.8) is used simply to verify this equality.

We have fitted the model without the need to specify constraints. If we wish for some reason to find a fit with specified constraints we proceed as follows. We seek a solution of the GLM scoring algorithm

$$(X' \tilde{W} X)\hat{\theta} = X' \tilde{W} \tilde{z}. \tag{3.9}$$

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two solutions of (3.9) where $\hat{\theta}_1$ satisfies the known constraint $H\hat{\theta}_1 = \mathbf{0}$ and $\hat{\theta}_2$ satisfies some unspecified constraint. We know

$$\begin{aligned} (X' \tilde{W} X)\hat{\theta}_1 &= (X' \tilde{W} X)\hat{\theta}_2 \\ \Rightarrow (X' \tilde{W} X + H' H)\hat{\theta}_1 &= (X' \tilde{W} X)\hat{\theta}_2. \end{aligned}$$

Now $X' \tilde{W} X + H' H$ is positive definite so if we can find a $\hat{\theta}_2$ we have the solution which satisfies $H\theta = \mathbf{0}$ as

$$\hat{\theta}_1 = (X' \tilde{W} X + H' H)^{-1} X' \tilde{W} X \hat{\theta}_2. \tag{3.10}$$

This expression exists in another form which is particularly illuminating. We observe that $X\hat{\theta}_2 = \log \hat{\lambda}$ and so we can write (3.10) as

$$\hat{\theta}_1 = (X' \tilde{W} X + H' H)^{-1} X' \tilde{W} \log \hat{\lambda}. \tag{3.11}$$

In other words, we can compute the estimated coefficients subject to any particular set of constraints directly from the invariant fitted values. Invariance is a fundamental and powerful idea and (11) demonstrates just one result from applying it.

Figure 1 emphasizes the effect that constraints can have on parameter estimates. The effect on the estimates of the time parameters is particularly striking. Under the standard constraints one may be tempted to interpret the parameter estimates as indicating improving mortality. The estimates under the R constraints indicate exactly the opposite. What can be said is that mortality does improve over time but that can only be concluded securely from the invariant fitted mortality shown in the lower right panel.

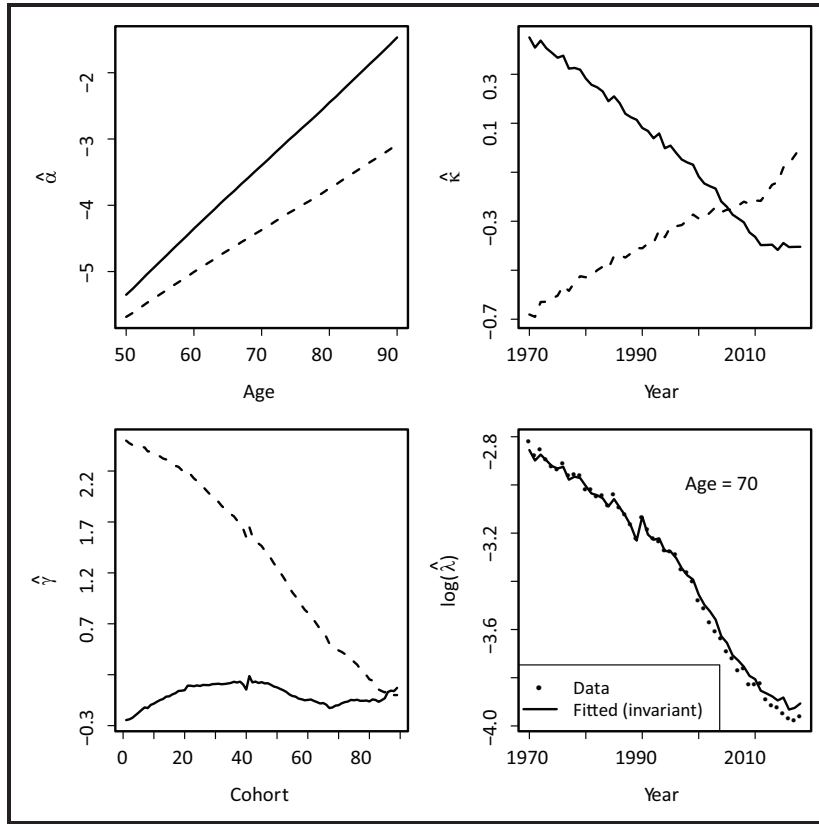


Figure 1 Estimates of α , κ and γ under standard constraints (solid line), R constraints (dashed line) and the observed and invariant fitted log mortality for age 70.

We examine the differences in the parameter estimates. We define $\Delta\alpha = \hat{\alpha}_s - \hat{\alpha}_r$ where the suffixes indicate the estimates under the standard and R constraints. We have corresponding definitions for $\Delta\kappa$ and $\Delta\gamma$. Currie (2020) showed that these differences obeyed a very simple relation. We have

$$\Delta\alpha = (A + B)\mathbf{1}_{n_a} + C\mathbf{x}_a \tag{3.12}$$

$$\Delta\kappa = -A\mathbf{1}_{n_y} - C\mathbf{x}_y \tag{3.13}$$

$$\Delta\gamma = -(B + n_a C)\mathbf{1}_{n_c} + C\mathbf{x}_c \tag{3.14}$$

for some constants A , B and C . We conclude that the three sets of differences are linearly related with slopes C , $-C$ and C . With our data and the standard and R constraints we found $A = -1.163$, $B = 1.465$ and $C = 0.03199$. Despite appearances to the contrary in Figure 1 the two sets of estimates are intimately related.

4 The smooth Age-Period-Cohort model

Gompertz smoothed the age parameters with a simple linear function. Here we use the method of P-splines to smooth the age parameters, α , in the APC model. As well as the general consideration that we expect the age effect to be a smooth monotonic function, there are good actuarial reasons for smoothing the age parameters. The resulting estimates in time by age will be more regular and hence the pricing of insurance products will also be more regular.

Let $B_a, n_a \times c_a$, be a regression matrix of B-splines along age. The model matrix is as in (2) but with $X_a = \mathbf{1}_{n_y} \otimes I_{n_a}$ replaced by $\mathbf{1}_{n_y} \otimes B_a$. We define the penalty matrix $P = \text{blockdiag}\{\tau D_2', D_2, \mathbf{0}_{n_y}, \mathbf{0}_{n_c}\}$. Here D_2 is a second order difference matrix and τ is a tuning constant. See Eilers and Marx (2021) for a thorough introduction to the method of P-splines. We seek solutions of the penalized scoring algorithm

$$(X' \tilde{W} X + P) \hat{\theta} = X' \tilde{W} \tilde{z} \tag{4.1}$$

where

$$\tilde{W} = \text{diag}\{\tilde{d}\}, \quad \tilde{z} = X\tilde{\theta} + d/\tilde{d} - \mathbf{1}. \tag{4.2}$$

for a GLM with Poisson errors and a log link.

We could use a generalization of (3.8) where the upper left block becomes $X' \tilde{W} X + P + H' H$. However, our purpose here is to use R to fit the model and then use invariance to calculate the estimates of the regression coefficients under the standard and R constraints. We will see that many of the results from the unsmoothed model carry over in a natural way.

We can avoid the use of algorithms like (3.8) with data augmentation; see Eilers and Marx (2021), p35. Data augmentation enables us to fit the smooth APC model with code similar to (3.4), (3.5) and (3.6) for the original APC model discussed in section 3. We define the augmented model matrix and augmented working variable as follows:

$$X_{\text{Aug}} = \begin{bmatrix} \tilde{W}^{1/2} X_a & : & \tilde{W}^{1/2} X_y & : & \tilde{W}^{1/2} X_c \\ \sqrt{\tau} D_2 & : & \mathbf{0}_{(c_a-2) \times n_y} & : & \mathbf{0}_{(c_a-2) \times n_c} \end{bmatrix}, \quad z_{\text{Aug}} = \begin{bmatrix} \tilde{W}^{1/2} \tilde{z} \\ \mathbf{0}_{c_a-2} \end{bmatrix}. \tag{4.3}$$

It follows immediately that

$$X'_{\text{Aug}} \tilde{W} X_{\text{Aug}} = X' \tilde{W} X + P, \quad X'_{\text{Aug}} \tilde{W} z_{\text{Aug}} = X' \tilde{W} \tilde{z}. \tag{4.4}$$

Thus solutions of the penalized scoring algorithm (4.1) and

$$X'_{\text{Aug}} \tilde{W} X_{\text{Aug}} \hat{\theta} = X'_{\text{Aug}} \tilde{W} z_{\text{Aug}} \tag{4.5}$$

are equal. Equation (4.5) is weighted least squares so we can fit the model with R's `lm` function. We note that we must iterate with (4.2) until convergence. Good initial values are the observed deaths for d and the observed log mortality rates for z . Convergence is very fast since we have essentially implemented the GLM scoring algorithm via the `lm` function.

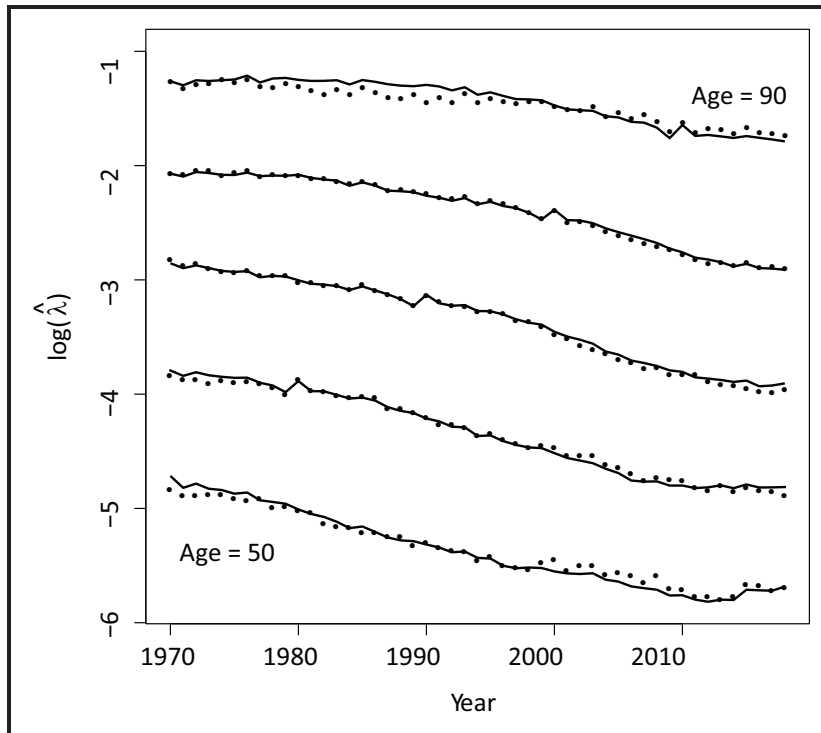


Figure 2 Observed and fitted $\log \lambda$ for ages 50 to 90 at ten year intervals for the smooth APC model.

We choose τ by minimizing AIC. The AIC curve is quite flat in the region of the optimal value and a short grid search gives $\tau = 100$ as a suitable value; this value is used in what follows.

We use the following extension of (3.11)

$$\hat{\theta}_1 = (X' \tilde{W} X + H' H + P)^{-1} X' \tilde{W} \log \hat{\lambda} \tag{4.6}$$

to investigate the relation between the two sets of estimates obtained under the standard and R constraints. We put H equal to H_s and H_r in turn where H_s and H_r are the constraints matrices for the standard and R constraints.

As in the unsmoothed case, we define Δ_a , Δ_κ and Δ_γ as the differences in the two sets of estimates but note here that Δ_a refers to the coefficients of the B-spline basis. We do not have the general theory that led to equations (3.12), (3.13) and (3.14) when $\tau > 0$; Currie (2020) gives the theory when $\tau = 0$. We proceed numerically. Plots of Δ_a , Δ_κ and Δ_γ suggest that each is linear and this is confirmed by simple regression. We find in our example that the slopes are 0.17915, -0.03583 and 0.03583 respectively. We note that $0.17915/0.03583 = 5$. We have used a knot spacing of five which suggests that the slopes are ΔC , $-C$ and C where Δ is the knot spacing. Support for this conjecture comes from Currie (2020) where it is shown that the slopes do indeed satisfy ΔC , $-C$ and C in the case $\tau = 0$.

Figure 2 shows the observed and fitted \log mortalities for age 50 to 90 at ten year intervals. Evidently, there has been a steady improvement in mortality over the observation period. This

improvement corresponds to an approximate increase in life expectancy of two years every ten years. For example, life expectancy in the UK in 1970 was 72.3 years and this has risen to 81.2 years in 2019. This has major consequences for the provision of both state and private pensions.

5 Discussion

This short article has highlighted the effect that different constraint systems can have on parameter estimates. Figure 1 is a dramatic example of such effects. We have also highlighted the role of the invariant fitted values. Equation (3.11) shows that knowledge of these invariant values yields parameter estimates under any particular constraint system. We have used the basic `lm` function in R as a black box to estimate the invariant fitted values and from these values diagrams like Figure 1 can be obtained and equation (3.11) can be applied.

Acknowledgements

I am grateful for the invitation to submit an article for this special issue. I met Brian on many occasions: it would be hard to imagine anyone with a greater zest for life in general and statistics in particular. I am also grateful for his great paper with Paul Eilers which introduced the world to P-splines. It has dominated my working life for over twenty years. I would also like to thank Paul Eilers for his encouragement to write this short article.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author received no financial support for the research, authorship and/or publication of this article.

References

- Currie ID (2013) Smoothing constrained generalized linear models with an application to the Lee-Carter model. *Statistical Modelling*, **13**, 69–93.
- Cairns AJG, Blake D, Dowd K, Coughlan GD, 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.
- Currie ID (2020) Constraints, the identifiability problem and the forecasting of mortality. *Annals of Actuarial Science*, **14**, 537–566.

- Eilers PHC and Marx BD (1996) Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–121.
- Eilers PHC and Marx BD (2021) *Practical smoothing: the joys of P-splines*. Cambridge: Cambridge University Press.
- Gomperts B (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, **115**, 513–583.
- Human Mortality Database (2023) University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org.
- R Core Team (2019) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.Rproject.org/>.