Smooth models of mortality with period shocks

James Kirkby[1] and Iain Currie[1]

[1] Department of Actuarial Mathematics and Statistics, and the Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, Scotland, I.D.Currie@hw.ac.uk

Abstract: We suppose that we have mortality data arranged in two-way tables of deaths and exposures classified by age at death and year of death. It is natural to suppose that there is a smooth underlying force of mortality, the mortality surface, that varies with age and year (or period). However, observed mortality is subject to more than stochastic deviation from this smooth surface; for example, flu epidemics, hot summers or cold winters can disproportionately effect the mortality of certain age groups in particular years. We call such an effect a period shock. We describe the mortality surface with an additive model with two components: the underlying smooth surface is modelled with 2-dimensional P-splines; the period shocks are modelled with a 1-dimensional P-spline in the age direction for each year. This is a large regression model but array methods (Currie *et al.*, 2006) enable the computations to be performed. We illustrate our methods with Swedish mortality data taken from the Human Mortality Database.

Keywords: Generalized linear array model; mortality; P-splines; period shock; smoothing.

1 A smooth model of mortality with period shocks

We suppose that we have mortality data arranged in two-way tables of deaths and exposures classified by age at death and year of death. It is natural to suppose that there is a smooth underlying force of mortality, the mortality surface, that varies with age and year (or period). However, observed mortality is subject to more than stochastic deviation from this smooth surface; for example, flu epidemics, hot summers or cold winters can disproportionately effect the mortality of certain age groups in particular years. We call such an effect a period shock.

An example of a period shock is the Spanish flu epidemic of 1918 which affected the mortality of those under the age of 60. We illustrate the extent of this effect with data on Swedish males from the Human Mortality Database; the data runs from 1900 to 2003 and from age 10 to 90. The upper right panel of Figure 1 shows the differences between the logarithms of observed mortality for 1918 and 1919. The remaining panels show the

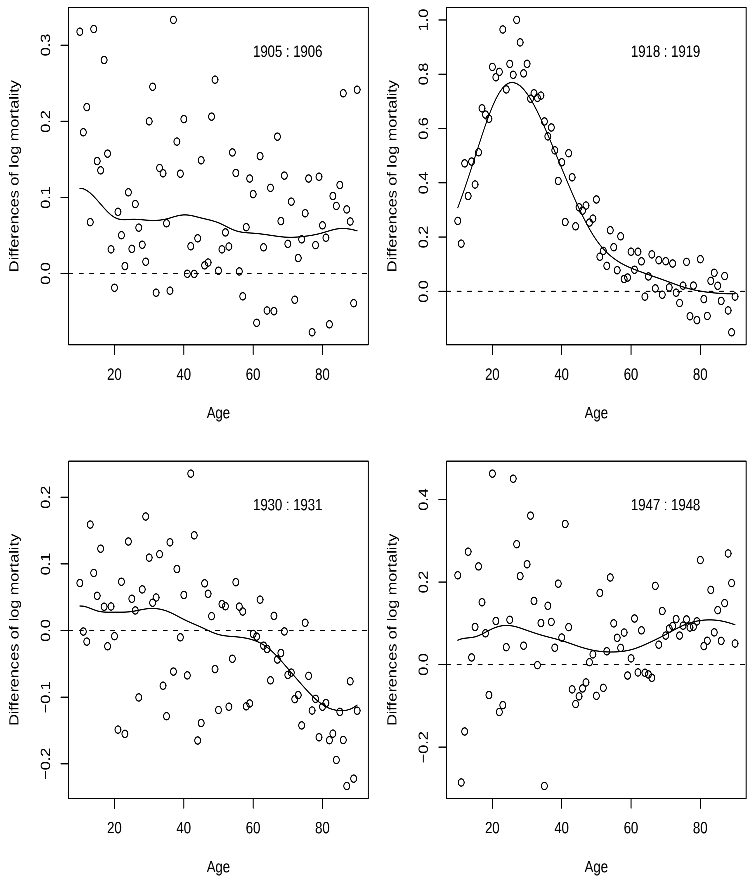

FIGURE 1. Differences of the logarithms of observed and fitted mortalities for Swedish males for selected successive years

corresponding differences for other years. The 1918/1919 experience is extreme but each panel indicates a systematic age dependent departure from a smoothly changing underlying mortality.

We propose an additive model with two components for the mortality surface : the first component describes a smooth two-dimensional surface and the second describes the period shocks. We describe each component in turn.

We suppose that the deaths and exposures are arranged in $n_a \times n_y$ matrices $\boldsymbol{Y}$ and $\boldsymbol{E}$, that $\boldsymbol{y} = \text{vec}(\boldsymbol{Y})$ and $\boldsymbol{e} = \text{vec}(\boldsymbol{E})$ are their vector equivalents, and that the rows and columns of $\boldsymbol{Y}$ and $\boldsymbol{E}$ are classified respectively by ages $\boldsymbol{x}_a$ and years $\boldsymbol{x}_y$ each arranged in ascending order. The first component of our model uses two-dimensional P-splines (Eilers and Marx, 1996, Currie *et al.*,

2004) to model the smooth underlying mortality surface. Let $\boldsymbol{B}_a = \boldsymbol{B}(\boldsymbol{x}_a)$, $n_a \times c_a$, and $\boldsymbol{B}_y = \boldsymbol{B}(\boldsymbol{x}_y)$, $n_y \times c_y$, be one-dimensional regression matrices of B-splines evaluated at age and year respectively; $\boldsymbol{B}_a$ and $\boldsymbol{B}_y$ are known as marginal regression matrices. The Kronecker product $\boldsymbol{B}_y \otimes \boldsymbol{B}_a$ creates a two-dimensional regression basis. We suppose that the number of deaths y_{ij} at age i and year j follows a Poisson distribution with mean $\mu_{ij} = e_{ij}\theta_{ij}$ where θ_{ij} is the force of mortality. We define a generalized linear model (GLM) for $\boldsymbol{y}$ with regression matrix $\boldsymbol{B}_y \otimes \boldsymbol{B}_a$, offset $\log \boldsymbol{e}$, log link and Poisson error. If we suppose that we have a rich basis of B-splines for age and year then a smooth surface is obtained by marginal penalization. We define the penalty matrix

$$\boldsymbol{P} = \lambda_a \boldsymbol{I}_{c_y} \otimes \boldsymbol{D}_a' \boldsymbol{D}_a + \lambda_y \boldsymbol{D}_y' \boldsymbol{D}_y \otimes \boldsymbol{I}_{c_a} \qquad (1)$$

where $\boldsymbol{D}_a$, λ_a, $\boldsymbol{D}_y$ and λ_y are the difference matrices and smoothing parameters for age and year respectively. We have defined a two-dimensional P-spline model; see Currie *et al.* (2004, 2006) and elsewhere for details of fitting these models.

We require the second component to have two properties. First, it must be versatile and capable of modelling different patterns in different years and second, the underlying patterns in different years must be smooth. Figure 1 illustrates both these features: the patterns in different years are distinct and follow separate underlying smooth curves. We define a second regression matrix by $\boldsymbol{I}_{n_y} \otimes \breve{\boldsymbol{B}}_a$ where $\breve{\boldsymbol{B}}_a = \breve{\boldsymbol{B}}_a(\boldsymbol{x}_a)$, $n_a \times c$, is a marginal regression matrix of B-splines. We suppose that c is small so that the modelling by age for each year is quite crude. With c regression coefficients for each of n_y years $\boldsymbol{I}_{n_y} \otimes \breve{\boldsymbol{B}}_a$ is a large matrix, even with small c. The underlying smooth mortality surface is modelled by the first component of our model so we force smoothness on the second component by applying a ridge penalty to the age coefficients in each year. We define the penalty matrix

$$\breve{\boldsymbol{P}} = \lambda_s \boldsymbol{I}_{n_y} \otimes \boldsymbol{I}_c = \lambda_s \boldsymbol{I}_{n_y c} \qquad (2)$$

where λ_s is the smoothing parameter for the shocks. Our two component model of mortality is thus a penalized generalized linear model with two additive components, linear predictor

$$\log \boldsymbol{\mu} = \log \boldsymbol{e} + \boldsymbol{B}_y \otimes \boldsymbol{B}_a \boldsymbol{a} + \boldsymbol{I}_{n_y} \otimes \breve{\boldsymbol{B}}_a \breve{\boldsymbol{a}} \qquad (3)$$

and block diagonal penalty blockdiag$[\boldsymbol{P} : \breve{\boldsymbol{P}}]$. This is a computationally demanding problem since three smoothing parameters must be chosen within the framework of a large GLM with $n_a n_y$ observations and $c_a c_y + c n_y$ regression variables. We describe how we deal with the computational problem in the next section.

2 Generalized linear array models

Generalized linear array models or GLAMs, introduced by Currie *et al.* (2006), provide a structure and a computational procedure for fitting GLMs whose model matrix can be written as a Kronecker product and whose data can be written as an array. The GLAM form of the 2-dimensional smooth model with regression matrix $\boldsymbol{B}_y \otimes \boldsymbol{B}_a$ is

$$\log \boldsymbol{M} = \log \boldsymbol{E} + \boldsymbol{B}_a \boldsymbol{A} \boldsymbol{B}_y' \tag{4}$$

where $\boldsymbol{M} = E(\boldsymbol{Y})$ and $\boldsymbol{A}$, $c_a \times c_y$, is the matrix of coefficients. In a large problem the GLAM approach gives very substantial savings in both storage and computational time over the usual GLM algorithm. The method can be extended to GLMs with additive components each of which has the GLAM form. The GLAM form of (3) is

$$\log \boldsymbol{M} = \log \boldsymbol{E} + \boldsymbol{B}_a \boldsymbol{A} \boldsymbol{B}_y' + \breve{\boldsymbol{B}}_a \breve{\boldsymbol{A}} \tag{5}$$

where $\breve{\boldsymbol{A}}$, $c \times n_y$, is a further matrix of coefficients. We use the GLAM procedure to fit model (3). Efficient computation of the linear predictor in (3) is provided by (5). The fitting of a GLM also requires the computation of a weighted inner product; for the additive model (3) we require

$$\begin{bmatrix} (\boldsymbol{B}_y \otimes \boldsymbol{B}_a)' \boldsymbol{W} (\boldsymbol{B}_y \otimes \boldsymbol{B}_a) & (\boldsymbol{B}_y \otimes \boldsymbol{B}_a)' \boldsymbol{W} (\boldsymbol{I}_{n_y} \otimes \breve{\boldsymbol{B}}_a) \\ (\boldsymbol{I}_{n_y} \otimes \breve{\boldsymbol{B}}_a)' \boldsymbol{W} (\boldsymbol{B}_y \otimes \boldsymbol{B}_a) & (\boldsymbol{I}_{n_y} \otimes \breve{\boldsymbol{B}}_a)' \boldsymbol{W} (\boldsymbol{I}_{n_y} \otimes \breve{\boldsymbol{B}}_a) \end{bmatrix} \tag{6}$$

where $\boldsymbol{W}$ is the diagonal matrix of weights. Details of the efficient computation of this matrix are given in Currie *et al.* (2006), in particular the example in section 7.1. The GLM is now fitted with the usual scoring algorithm with the linear predictor, weighted inner product and working variable all computed using array computations.
A simpler method of modelling shocks is the mean shock, by which we mean a constant shock for all ages within a year. The GLAM form of the linear predictor for this model is

$$\log \boldsymbol{M} = \log \boldsymbol{E} + \boldsymbol{B}_a \boldsymbol{A} \boldsymbol{B}_y' + \boldsymbol{1}_{n_a} \boldsymbol{h}' \tag{7}$$

where $\boldsymbol{1}_{n_a}$ is a vector of 1's of length n_a and $\boldsymbol{h}' = (h_1, \ldots, h_{n_y})$. Mean shocks are used in some well-known models of mortality, such as the Lee-Carter and the Age-Period-Cohort models; neither of these models is capable of modelling the kind of effects we see in Figure 1 but we consider model (7) for comparison.

3 An application to Swedish data

We use the Baysian Information Criterion (BIC) for model selection and fit our three models to the Swedish mortality data used in section 1. We have

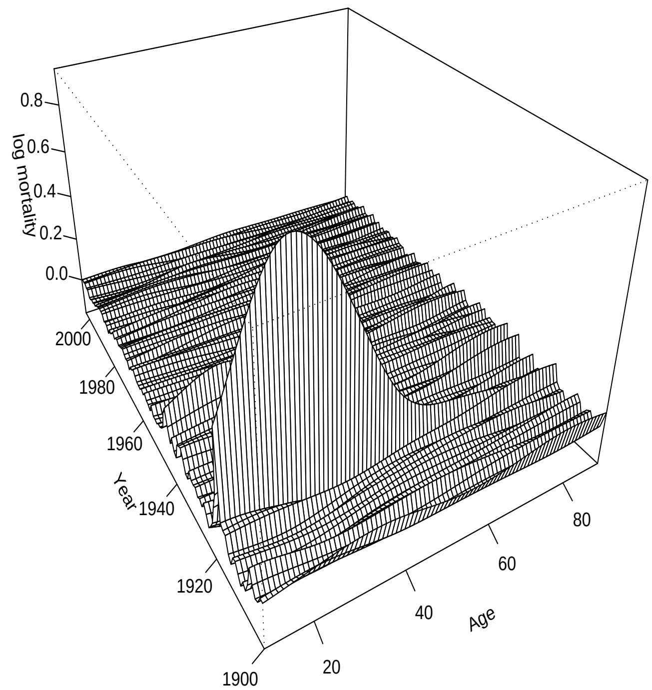

FIGURE 2. Age dependent shocks to mortality surface by year

$n_a = 81$ ages and $n_y = 104$ years and used cubic B-splines with $c_a = 19$, $c_y = 24$ and $c = 9$. The full regression matrix in (3) is 8424×1392, a large regression matrix. The fitted values have been added to Figure 1; it appears that our model has successfully modelled the systematic increases and decreases seen for our selected years. The extent of the period shocks is shown in Figure 2; the age structure of these shocks is evident.

Table 1 gives various summary statistics for all three models: the basic 2-dimensional smooth model (4), the mean shock model (7) and the age dependent shock model (5). It is clear that the period shock model is superior to the mean shock model which is, in turn, superior to the 2-dimensional model.

TABLE 1. Summary statistics

Model	$(\lambda_a, \lambda_y, \lambda_s)$	Trace	Deviance	BIC
2-d smooth	(10, 7, -)	293	21226	23871
Mean shock	(0.05, 30, 2000)	367	15538	18852
Period shock	(0.01, 1900, 850)	489	9670	14089

4 Concluding remarks

The 1918 observations are extreme. An alternative modelling strategy is to regard the 1918 data as missing. This has the effect of increasing the value of λ_s (since the remaining data are much less variable by year). The estimates of the remaining period shocks are smoother compared to the shocks from the full data set.

It is possible to express all our models in mixed model form. One consequence of this approach is that the period shocks can be regarded as random effects which is appealing given their nature. The GLAM methodology is available for mixed models (Currie *et al.*, 2006).

The models in this paper and model (3) in particular are computationally intensive. Further computational savings over and above those provided by GLAM can be made by taking advantage of the form of the regression matrices of the period components in (3) and (7).

In conclusion, the period shock model was successful in modelling age-dependent departures within years for the Swedish data considered in this paper. Experience with other data sets has shown that the model is more widely applicable. We see two main uses of the model: first, it can detect age effects within single years and this will often be of interest in its own right; second, the two component model enables the underlying smooth surface to be more successfully identified.

References

Currie, I.D., Durban, M., and Eilers, P.H.C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling* **4**, 279-98.

Currie, I.D., Durban, M., and Eilers, P.H.C. (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society, Series B* **68**, 259-80.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89-121.

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org.