MODELLING AND FORECASTING THE MORTALITY OF THE VERY OLD

BY

IAIN D. CURRIE

ABSTRACT

The forecasting of the future mortality of the very old presents additional challenges since data quality can be poor at such ages. We consider a two-factor model for stochastic mortality, proposed by Cairns, Blake and Dowd, which is particularly well suited to forecasting at very high ages. We consider an extension to their model which improves fit and also allows forecasting at these high ages. We illustrate our methods with data from the Continuous Mortality Investigation.

KEYWORDS

Forecasting, very high ages, mortality, smoothing.

1. INTRODUCTION

Mortality data at very high ages can be unreliable and this presents an additional challenge when it comes to forecasting mortality. Indeed, assured lives data from the Continuous Mortality Investigation (CMI) provide an example where observed mortality actually falls, and falls quite sharply, above age 95. We suspect that there are data-quality issues at these ages since we don’t really believe that mortality actually falls with increasing age. One solution is to discard the suspect data, and model and forecast over a restricted age range, say up to age 90; Cairns et al. (2009) and Richards & Currie (2009) both take this approach. However, actuaries still need to produce forecasts at high ages since the calculation of lifetime annuities and pensions requires mortality rates up to (say) 120, well above the ages at which good-quality data are typically available; it is this problem that we address in the present paper.

As an illustration of the challenge we consider the Continuous Mortality Investigation (CMI) assured lives data set. Thus we have policy-oriented life-assurance data where ‘deaths’ are claims and ‘exposure’ is total time at risk. We suppose that the claim and central exposure data are available in
two matrices, $D = (d_{i,j})$ and $E = (e_{i,j})$ where $i = 1, \ldots, n_a$ and $j = 1, \ldots, n_y$; the rows of $D$ and $E$ are classified by age at death, $x_a$, and the columns by year of death, $x_y$. Here we consider the data where $x_a = (40, \ldots, 100)'$ and $x_y = (1950, \ldots, 2005)'$. Let $m = (m_i)$, where

$$m_i = \log \left( \frac{\sum_j d_{i,j}}{\sum_j e_{i,j}} \right), \; i = 1, \ldots, n_a,$$

be a measure of log mortality by age averaged over year. The left panel of Figure 1 shows $m$ plotted against age. A striking feature of the plot is the fall in mortality rates for ages above 95 or so.

We consider a model proposed by Cairns et al. (2006, 2009) which is particularly well suited to forecasting the mortality of the very old in the presence of unreliable data. The plan of the paper is as follows. In section 2 we describe the Cairns et al. model and use it for forecasting at very high ages. In section 3 we propose an extension to this model which improves the fit to data and also allows forecasting at very high ages. The paper closes with a short conclusion.

2. A MODEL OF CAIRNS, BLAKE AND DOWD

Cairns et al. (2006, 2009) suggested the following two-factor model for stochastic mortality:

$$\log \mu_{i,j} = \kappa_{0,j} + \kappa_{1,j}(x_i - \bar{x})$$

where $\mu_{i,j}$ is the force of mortality in year $j$ at age $x_i$. Expression (2) together with the assumption that the number of claims follows a Poisson distribution $D_{i,j} \sim \mathcal{P}(e_{i,j} \mu_{i,j})$ specifies a generalized linear model (GLM), and the model can be fitted with standard software such as R (R Development Core Team, 2009). This model is a member of the extensive set of forecasting models described in Cairns et al. (2009); we will refer to it as CBD. We can think of CBD as a collection of Gompertz models, one model for each year. The coefficients in $\kappa_0 = (\kappa_{0,1}, \ldots, \kappa_{0,n_y})'$ then represent the intercepts of the Gompertz models while those in $\kappa_1 = (\kappa_{1,1}, \ldots, \kappa_{1,n_y})'$ represent the slopes.

We write CBD as a GLM as follows. Let $d = \text{vec}(D)$ and $e = \text{vec}(E)$ be the vectors of claims and central exposed to risk respectively; here, and below, $\text{vec}(\cdot)$ is the operator which stacks the columns of a matrix into a vector. Let $M = (\mu_{i,j})$, $n_a \times n_y$, be the matrix of forces of mortality and $\mu = \text{vec}(M)$. Let $1_a$ be the vector of 1s of length $n_a$, $x = x_a - \bar{x} 1_a$ where $\bar{x}$ is the mean age, and let $I_y$ be the identity matrix of size $n_y$. We define two regression matrices $X_0$ and $X_1$ as follows:

$$X_0 = I_y \otimes 1_a, \quad X_1 = I_y \otimes x$$
where $\otimes$ denotes the Kronecker product; see Searle (1982, pp. 265, 333), for example, for a discussion of Kronecker products. With a log link, the linear predictor of the GLM representation of CBD is

$$\log(e) + X_{0}k_{0} + X_{1}k_{1}.$$  \hspace{1cm} (4)

Figure 2 displays the fitted values of $k_{0}$ and $k_{1}$ where we have restricted the data to age $x_{a} = (40, \ldots, 89)'$ and years $x_{y} = (1950, \ldots, 2005)'$. The left panel shows how the general level of mortality has fallen over the last fifty years or so, while the right panel shows that the rate of change (slope) of mortality with respect to age has increased over time. Figure 2 suggests that it should be possible to forecast $k_{0}$ and $k_{1}$; we follow Cairns et al. (2009) and use a bivariate random walk with drift to forecast $k_{0}$ and $k_{1}$. With the forecast values of $k_{0}$ and $k_{1}$ in place, expression (2) yields a forecast of the mortality table for the age range of reliable data, here 40 to 89.

The structure of CBD now allows us to extend the forecast in the age direction. We summarize the complete process:

(i) we fit CBD with data in which we have confidence (here ages 40 to 89);
(ii) we use a bivariate random walk with drift to forecast $k_{0}$ and $k_{1}$ (here to 2050 for ages 40 to 89);
(iii) with the fitted and forecast values of $k_{0}$ and $k_{1}$ in place, expression (2) provides fitted and forecast mortalities for the suspect ages 90 to 100.

Corresponding to $m$, a measure of average observed log mortality by age defined in (1), we define $\hat{m}$, a measure of average fitted log mortality by age as follows: let $\hat{m} = (\hat{m}_{i})$, with

$$\hat{m}_{i} = \log \left( \frac{\sum_{j}d_{i,j}}{\sum_{j}e_{i,j}} \right), \ i = 1, \ldots, n_a,$$  \hspace{1cm} (5)
where we have replaced $d_{i,j}$, observed deaths, in (1) by $\hat{d}_{i,j}$, fitted deaths. The values of $\hat{m}$ have been added to Figure 1. The fitted Gompertz line (dashed) follows the data (solid) quite well up to around age 85; the model assumption then increasingly over-rides the data above age 85.

We can streamline the above process by a simple device. We define a weight matrix $V$ with $v_{i,j} = 1$ where we have confidence in the data and $v_{i,j} = 0$ where the data are unreliable; let $v = \text{vec}(V)$. We now fit model (2) over the whole data region (here ages 40 to 100 and years 1950 to 2005) but weight the observations by $v$. An important point is that using a weight matrix in this way results in exactly the same fitted values in the region of reliable data as fitting without a weight matrix over the reliable data only. We will use this device in the next section in a more general setting.

3. A SMOOTH VERSION OF CBD

The right panel of Figure 1 suggests that model (2) overestimates log mortality at low and high ages, and underestimates it at central ages. One possibility is to add a quadratic term to (2) as follows:

$$
\log \mu_{i,j} = \kappa_{0,j} + \kappa_{1,j}(x_i - \bar{x}) + \kappa_{2,j} [(x_i - \bar{x})^2 - \sigma_i^2]
$$

where $\sigma_i^2 = \sum (x_i - \bar{x})^2/n_i$. This model is a special case of another member (the seventh) in the Cairns et al. (2009) set in which the cohort effect is set to zero. A possible disadvantage of this model is the introduction of a third time series, $\kappa_3$. We can obtain a simpler version of model (6) by replacing the third time series by a constant:

$$
\log \mu_{i,j} = \kappa_{0,j} + \kappa_{1,j}(x_i - \bar{x}) + \rho [(x_i - \bar{x})^2 - \sigma_i^2].
$$
We will comment on the fit of these two extensions to model (2) below. In this paper we avoid the introduction of an additional quadratic term by setting
\[ \log \mu_{i,j} = \kappa_{0,j} + \kappa_{1,j} S(x_i - \bar{x}) \]  
(8)

where \( S(x_i - \bar{x}) \) is a smooth function of \( x_i - \bar{x} \). We will use the P-spline system of Eilers and Marx (1996) to estimate \( S(\cdot) \). For a description of P-splines from an actuarial perspective see Richards et al. (2006) and Richards and Currie (2009). Let \( \{ B_1, \ldots, B_c \} \) be a B-spline basis of dimension \( c \) defined along age with equally spaced knots and let \( B_a = (B_j(x_i - \bar{x})) \), \( n_a \times c \), be the corresponding regression matrix. It is helpful to rewrite the linear predictor (4) for CBD in matrix form as
\[ \log(E) + 1_a \kappa'_0 + x \kappa'_1. \]  
(9)

We can also express the linear predictor for (8) in matrix form. We substitute \( x = B_a a \) in (9) where \( a, c \times 1 \), is the vector of regression coefficients in age. The matrix form of the linear predictor for (8) is
\[ \log(E) + 1_a \kappa'_0 + B_a a \kappa'_1. \]  
(10)

We will refer to this model as smooth CBD. We are now faced with the same problem as we encounter in the estimation of the parameters of the Lee-Carter model (Lee and Carter, 1992, Brouhns et al., 2002): expression (10) does not define a regression equation since there is no regression variable. We fit model (8) by iterating between two GLMs. Suppose \( \hat{a} \) is an estimate of \( a \), then (10) corresponds to a GLM with linear predictor
\[ \text{LP1: } \log(e) + [I_y \otimes 1_a] \kappa_0 + [I_y \otimes B_a \hat{a}] \kappa_1; \]  
(11)

this is (4) with \( x \) replaced by its smooth estimate, \( B_a \hat{a} \). Conversely, suppose that \( \hat{\kappa}_0 \) and \( \hat{\kappa}_1 \) are estimates of \( \kappa_0 \) and \( \kappa_1 \); then (10) corresponds to a GLM with linear predictor
\[ \text{LP2: } \log(e) + [\hat{\kappa}_0 \otimes 1_a] + [\hat{\kappa}_1 \otimes B_a] a. \]  
(12)

Convenient initial estimates of \( \kappa_0 \) and \( \kappa_1 \) are provided by the estimated values from fitting CBD. We now iterate between the two GLMs with linear predictors LP1 and LP2 until convergence.

Two problems remain: first, it is not obvious what value of \( c \), the dimension of the B-spline basis, we should choose; second, unlike the original CBD model where the age structure is linear, there is no obvious way that we can extrapolate the smooth function, \( S(x - \bar{x}) \), into the unreliable data region. The P-spline system of Eilers and Marx (1996) allows us to solve both these
problems simultaneously: we choose a sufficiently large value of $c$ so that the fitted curve $S(x - \bar{x})$ is undersmoothed; we then place a second order penalty on the coefficients, $a$ in (10), and fit by penalized likelihood. We can now fit the model by iterating between a GLM with linear predictor LP1 and a penalized GLM with linear predictor LP2; see Currie et al. (2004) for details of fitting a penalized GLM. The level of smoothing of $a$ is determined by minimizing the Bayesian Information Criterion where

$$\text{BIC} = \text{Deviance} + \log(n) \times \text{Dimension},$$  \hspace{1cm} (13)

where $n = \sum v_{i,j}$ is the number of observations used to fit the model. Thus BIC balances (a) the closeness of the fit as measured by the Deviance and (b) the complexity of the fitted model as measured by the Dimension; see McCullagh and Nelder (1989, p. 25) for a discussion of deviance and Hastie and Tibshirani (1990, p. 52) for one on dimension.

In the case of CBD, we were able to obtain fitted and forecast values in the region of unreliable data by using the weight matrix $V$. With smooth CBD, ie, model (10), the same device can be used, since the penalty function together with $V$ enables extrapolation of $B_o a$ to occur; with a second order penalty, extrapolation of $B_o a$ is the result of linear extrapolation of the last two coefficients in $a$ in the reliable age region. As in the case of CBD, fitting over the full data set with a weight matrix results in the same fitted values in the region of reliable data as fitting over the region of reliable data only; this remark assumes that the $B$-spline basis used to span all ages is an extension of that used to span the reliable ages.

Figure 3 is Figure 1 but for smooth CBD; the improvement in fit of smooth CBD over CBD is evident from the right panels of Figures 1 and 3. Table 1 gives some summary statistics not only for CBD, model (2), and smooth CBD, model (8), but also for models (6) and (7). All three extensions to CBD give very substantial improvements in fit as measured by both the deviance and BIC.

Model (6) gives the best fit but at the price of adding a large number of time series terms. The improvement in fit of smooth CBD with a deviance of

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SUMMARY TABLE FOR FOUR MODELS APPLIED TO CMI DATA.</strong></td>
</tr>
<tr>
<td><strong>MODEL (2) IS THE ORIGINAL CBD MODEL; MODEL (6) HAS AN ADDITIONAL TIME SERIES TERM; MODEL (7) HAS AN ADDITIONAL QUADRATIC TERM; MODEL (8) IS SMOOTH CBD.</strong></td>
</tr>
<tr>
<td>Model &amp; (2) &amp; (6) &amp; (7) &amp; (8)</td>
</tr>
<tr>
<td>Deviance &amp; 8856 &amp; 5315 &amp; 5917 &amp; 5822</td>
</tr>
<tr>
<td>Dimension &amp; 112 &amp; 168 &amp; 113 &amp; 116</td>
</tr>
<tr>
<td>BIC &amp; 9745 &amp; 6648 &amp; 6814 &amp; 6745</td>
</tr>
</tbody>
</table>
MODELLING AND FORECASTING THE MORTALITY OF THE VERY OLD

5822 over the quadratic model (7) with a deviance of 5917 easily outweighs
the increase in model complexity (the model dimension increases by about 3).

Finally, Figure 4 shows fitted and forecast log mortality with confidence
intervals for two ages; the width of the confidence intervals depends on param-
eter uncertainty for the fitted model in the region of the data, but is largely
determined by parameter uncertainty in the random walk in the region of the
forecast. At age 70 there is little difference in the forecasts with CBD and
smooth CBD. At age 95 both models show fitted log(mortality) higher than
observed mortality, with CBD giving larger differences from the data than
smooth CBD. Thus one agreeable feature of the extrapolation with smooth
CBD is that it gives a compromise between (a) extrapolation with CBD and
(b) the data. The width of the confidence intervals at age 95 indicates a high
level of uncertainty in the future direction of mortality, part of the price of
forecasting without data.

4. CONCLUDING REMARKS

In this short paper we have used two models whose structure allows forecasts
of future mortality for ages where data might be unreliable. Extrapolation in
age is straightforward with the original CBD model, since the age function is
a straight line. We extended this model by allowing the age function to be a
smooth function, in which case the penalty function in the P-spline system
allows extrapolation. The method described here will work for any model
which allows age extrapolation. One such class of models is the 2d P-spline
models described in Richards et al. (2006). In these models, forecasting in the
time direction is done via the time penalty, but extrapolation in the age direc-
tion can also be done simultaneously via the age penalty.
ACKNOWLEDGEMENTS

I am grateful to Andrew Cairns and Stephen Richards for useful discussions during the preparation of this paper and for their helpful comments on an earlier draft. I am also grateful for the comments of an anonymous referee which helped to clarify the presentation and to the CMI for permission to use their data.

REFERENCES


FIGURE 4: Fitted and forecast log mortality for ages 70 and 95.


IAIN D. CURRIE

*Department of Actuarial Mathematics and Statistics and the Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK.*

Tel.: +44 (0)131 451 3208,

Fax: +44 (0)131 451 3249

E-Mail: I.D.Currie@hw.ac.uk