LONGEVITY RISK AND ANNUITY PRICING WITH THE LEE-CARTER MODEL

By S. J. Richards and I. D. Currie

[Presented to the Faculty of Actuaries, 16 February 2009]

ABSTRACT

Several important classes of liability are sensitive to the direction of future mortality trends, and this paper presents some recent developments in fitting smooth models to historical mortality-experience data. We demonstrate the impact these models have on mortality projections, and the resulting impact which these projections have on financial products. We base our work round the Lee-Carter family of models. We find that each model fit, while using the same data and staying within the Lee-Carter family, can change the direction of the mortality projections. The main focus of the paper is to demonstrate the impact of these projections on various financial calculations, and we provide a number of ways of quantifying, both graphically and numerically, the model risk in such calculations. We conclude that the impact of our modelling assumptions is financially material. In short, there is a need for awareness of model risk when assessing longevity-related liabilities, especially for annuities and pensions.

KEYWORDS

Longevity Risk; Lee-Carter; Mortality Projections; Smoothing; Model Risk; Annuities

CONTACT ADDRESS

Stephen Richards, 4 Caledonian Place, Edinburgh EH11 2AS, U.K. Tel: +44(0)131 315 4470; E-mail: stephen@longevitas.co.uk; Web: www.longevitas.co.uk

1. Introduction

'*It's tough to make predictions, especially about the future.*'
Lawrence Peter 'Yogi' Berra, U.S. baseball player and team manager

'*I never think of the future. It comes soon enough.*' Albert Einstein, interviewed in 1930

1.1 Actuaries do not have the luxury of thinking like Albert Einstein, as almost every calculation which they make involves some sort of assumption about the future. A particularly topical assumption is that of future mortality trends, usually labelled as 'mortality improvements', due to the clear expected direction of change.

1.2 One of the most commonly used models for projecting future mortality is the Lee-Carter model (Lee & Carter, 1992). Although originally created for forecasting life expectancy, it is also used to forecast mortality rates at each age. This paper presents some recent developments in the fitting of Lee-Carter models, specifically the idea of using smoothing methods to reduce the number of effective parameters in the model. A useful by-product of this smoothing is a new approach to projecting future mortality trends within the Lee-Carter framework. We illustrate the fitting of these smoothed models, and also the financial impact of the projections produced by them. We find that model risk is a particularly important source of uncertainty for actuaries pricing and reserving for pensions and annuities, often as important as the uncertainty within the model itself.

2. Data and Data Preparation

2.1 The data used in this paper are the number of deaths aged x last birthday during each calendar year y, split by gender. Corresponding mid-year population estimates are also given. The data, therefore, lend themselves to modelling the force of mortality, $\mu_{x+\frac{1}{2},y+\frac{1}{2}}$, without further adjustment. We use two such data sets, one provided by the Office of National Statistics (ONS) and one by the CMI. The CMI data come in the form of initial exposed-to-risk, and were adjusted to an approximate mid-year central exposed-to-risk by deducting half of the deaths.

2.2 We use ONS data for England and Wales for the calendar years 1961 to 2006 inclusive. This particular data set has death counts and estimated exposures only up to age 89, and we will work here with the subset of ages 40 to 89, which is most relevant for insurance products sold around retirement ages. ONS death data are provided by the date of registration, but, between 1993 and 2005, they are also available by the date of occurrence. In this paper we have used the registration data throughout for consistency. This is the same data set as used in Richards *et al.* (2006), but with a smaller age range and some extra years of data. More detailed discussion of this data set, particularly regarding the estimated exposures, can be found in Richards (2008a).

2.3 The other data set which we will use is the experience data for assured lives collected by the CMI for calendar years 1947 to 2005 inclusive. For consistency with the ONS data set, we will use the same age range of 40 to 89. Note that the CMI data set is only useful for male lives, as the data are sparse for females.

2.4 All the data here are therefore supplied aggregated, and we will model the mortality of groups. This is in contrast to the models of individual mortality which are used for detailed life insurer data, as outlined in Richards (2008b). Note that both the models for groups in this paper and the individual models of mortality in Richards (2008b) are all models for the force of mortality (hazard function), rather than for the mortality rates often used by actuaries. Mortality rates (q_x) can, of course, be calculated precisely from knowledge of the force of mortality.

2.5 We also use three data sets for actual pensions in payment: two for annuity portfolios; and one for a defined benefit pension scheme. Details of these portfolios are provided in Section 8. The portfolios are used to simulate the lifetime of the people behind the annuities or the pensions. Prior to carrying out the simulations, we deduplicated the portfolios using the procedure outlined in Richards (2008b), i.e. pensions or annuities paid to the same person were identified and aggregated. Simulating without deduplication would be doubly misleading: first, it is common for people to have more than one annuity, so the mortality experience for annuities is not independent; and second, wealthier people have a greater tendency to have multiple annuities, so a correct picture of the financial volatility can only be obtained by adding up pension and annuity records for the same person. Failure to deduplicate prior to simulation would give a falsely comforting picture of the binomial risk, and it would likely also under-estimate the concentration risk because of large aggregated pensions.

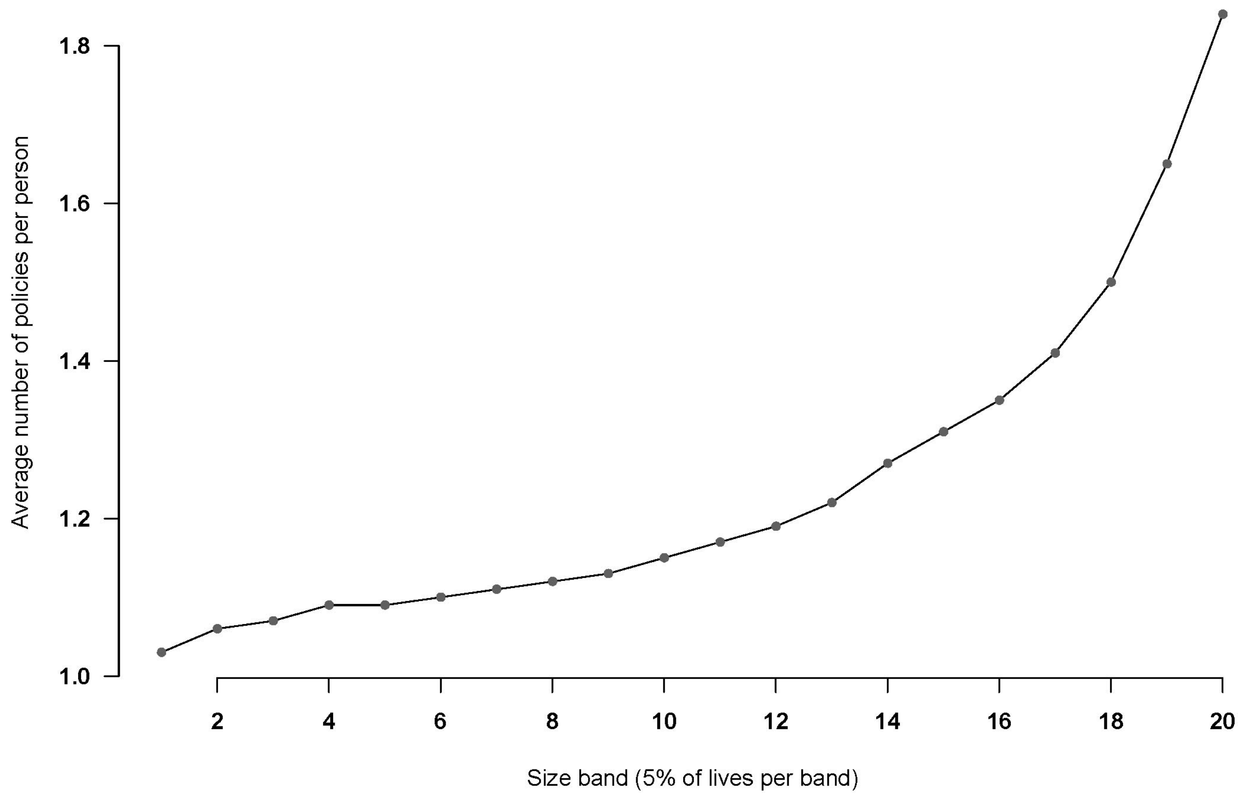

Figure 1. Average number of policies per person in each of equal-sized membership bands ordered by total annual annuity income; band 1 is the 5% of lives with the smallest annual pensions, through to band 20 which is the 5% of lives with the largest annual pensions; data taken from the large life-office annuity portfolio used in Richards (2008b)

2.6 These statements require some justification. Figure 1 shows the average number of annuities after deduplication for each life identified in a large life office annuity portfolio. Deduplication was done using the process outlined in Richards (2008b). Each of the 20 groups in Figure 1 represents the same proportion of the lives. On the left we have the 5% of lives with the smallest total pension, where there are very few duplicates and an average of 1.03 policies per person. On the right we have the 5% of lives with the largest total pension, where there are many duplicates and an average of 1.84 policies per person. This correlation of the number of policies with wealth demonstrates the importance of deduplication when performing any statistical analysis.

2.7 There is a degree of self-fulfilling prophesy in Figure 1, however. By adding together pensions across duplicates, of course larger incomes will appear to have a higher average number of policies! We can, however, prove this another way, by using geodemographic profiles according to Richards (2008b). Table 1 shows the average total annuity and average number of policies for each of 11 Mosaic groups assigned by United Kingdom postcode. These postcode profiles are not defined in relation to pension size, but there is a clear link between geodemographic group and average benefit size. This suggests that geodemographic profiles are useful in clarifying policyholder status where benefit levels are small and medium sized. As expected, there is

Table 1. Summary statistics by Mosaic group for large annuity portfolio

Mosaic group name	Average annuity (£ p.a.)	Average policies per life
Symbols of success	4,348	1.33
Rural isolation	3,405	1.30
Grey perspectives	2,708	1.29
Suburban comfort	2,203	1.24
Urban intelligence	2,489	1.22
Happy families	1,856	1.19
Ties of community	1,592	1.19
Twilight subsistence	1,394	1.17
Blue collar enterprise	1,444	1.16
Welfare borderline	1,281	1.14
Municipal dependency	1,093	1.12
Unmatched or unrecognised postcodes	2,619	1.17
Commercial addresses	4,365	1.35
All lives	2,663	1.24

Source: Postcode Mosaic groups from Experian Ltd applied to the large life-office annuity portfolio used in Richards (2008b). Statistical models require the independence assumption to hold, i.e. data sets have to be deduplicated. This is all the more critical when important subsets of policyholders have more duplicates than others, as clearly shown here.

Table 2. Proportion of portfolio by number of policies held

Number of policies per individual	Proportion of total: (a) by lives	(b) by amounts
1	83.3%	67.6%
2	2.6%	20.3%
3	2.6%	6.4%
4	0.8%	2.6%
5	0.3%	1.3%
6	0.2%	0.6%
7	0.1%	0.4%
8	0.1%	0.3%
9	0.0%	0.1%
10	0.0%	0.1%
11	0.0%	0.1%
12	0.0%	0.1%
Total	100%	99.9%

Source: Data from large life-office annuity portfolio used in Richards (2008b). The column for amounts does not quite add up to 100%, because there are some individuals with even more than 12 policies (up to a maximum of 31 for this portfolio).

also a strong correlation with the average number of policies, thus proving that wealthier and higher-status individuals have a greater tendency to have multiple policies.

2.8 Another way of looking at the issue of duplicates is to examine the proportion of lives and amounts by each number of policies per individual. This is done in Table 2, which shows that nearly a third of pensions are in respect of individuals with multiple policies. This means that deduplication is essential in forming an accurate picture of liabilities for this annuity portfolio.

3. Models and Notation

3.1 All the models used in this paper are based on the Lee-Carter (1992) framework:

$$\log \mu_{x,y} = \alpha_x + \beta_x \kappa_y \tag{1}$$

where $\mu_{x,y}$ denotes the force of mortality (hazard rate) at age x in year y, α_x is the effect of age x, κ_y is the effect of calendar year y, and β_x is the age-specific response to the calendar year effect. Since we are dealing with two-dimensional data sets, it is convenient to rewrite equation (1) in matrix form, and we define: $\boldsymbol{M} = (\mu_{x,y})$, the matrix of the forces of mortality indexed by age and time; $\boldsymbol{\alpha}' = (\alpha_1, \ldots, \alpha_{n_a})$, the vector of age effects (n_a is the number of ages and $'$ denotes the transpose of a vector (or matrix)); $\boldsymbol{\kappa}' = (\kappa_1, \ldots, \kappa_{n_y})$, the vector of calendar year effects (n_y is the number of years); and

$\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_{n_a})$, the vector of age-specific responses. With this notation we can now rewrite equation (1) in matrix form as:

$$\log \boldsymbol{M} = \boldsymbol{\alpha}\mathbf{1}' + \boldsymbol{\beta}\boldsymbol{\kappa}' \tag{2}$$

where $\mathbf{1}$ is a vector of 1s and is of length n_y.

3.2 The Lee-Carter model is designed with forecasting in mind, since the year effect across ages is reduced to the single time-indexed parameter $\boldsymbol{\kappa}$. The Lee-Carter assumption certainly simplifies the forecasting problem, but it is as well to remember that it is not immune to the problems of any forecasting method: (a) model risk; (b) parameter uncertainty; (c) parameter stability; and (d) stochastic variation. Model risk stems from the model assumption determining the forecast, but what if the model is wrong? The main point of the present paper is to discuss the impact of model assumptions on the pricing of financial products. Parameter uncertainty exists even if we have the correct model, since parameter estimates are still subject to sampling variation. As for parameter stability, the Lee-Carter model makes the strong assumption that the estimated values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (estimated from past data) remain fixed at these values in the future. Kingdom (2008) suggests that this assumption may not hold in practice. Finally, as for stochastic variation, only if points (a), (b) and (c) are successfully negotiated will our confidence interval give a true reflection of the likely future course of mortality.

3.3 We will use three models in this paper. First, we have the Lee-Carter model, where we estimate the parameters α_x, β_x and κ_y by the method of maximum likelihood. The parameters in (1) are not identifiable, since the following transformations yield the same fitted values of $\mu_{x,y}$ for any real value of c:

$$\alpha_x \to \alpha_x^* = \alpha_x - c\beta_x \qquad \beta_x \to \beta_x^* = \beta_x \qquad \kappa_y \to \kappa_y^* = \kappa_y + c. \tag{3}$$

We therefore fix a convenient parameterisation by setting $\sum \kappa_j = 0$ and $\sum \kappa_j^2 = 1$. This has the attractive feature that $\boldsymbol{\alpha}$ is a measure of average log mortality by age (under equation (3)). Under this model, mortality projections for $\mu_{x,y}$ are obtained by projecting a time series for κ_y. This model will be denoted 'original Lee-Carter', or just 'LC', although the maximum-likelihood approach to parameter estimation was developed by Brouhns *et al.* (2002). The model in Lee & Carter (1992) was not fitted originally by maximum-likelihood estimation.

3.4 The second model is where smoothness is imposed on the β_x. The thinking behind this model is that, with data sets with small numbers of deaths (such as insurance data), the estimates of the β_x can be volatile, and this leads to inconsistent forecasts of future mortality; smoothing the β_x will fix this problem. This model was proposed by Delwarde *et al.* (2007), and they smoothed the β_x using penalised *B*-splines or *P*-splines (Eilers & Marx, 1996); this model will be denoted 'DDE'. As with the original Lee-Carter model, projections for $\mu_{x,y}$ are still obtained by projecting a time series for κ_y, although, potentially, this can differ due to the different structure for the β_x.

3.5 The third model is where smoothness is imposed on both the β_x and the κ_y by means of *P*-splines. The thinking behind this model is that, in addition to the advantages of the DDE model, we also have an estimate of the underlying trend in the κ_ys. This approach gives an alternative method of forecasting the κ_ys since the penalty function enables forecasting to take place. This model will be denoted 'CR' in this paper. Projection with penalty functions is discussed in detail in Richards *et al.* (2006).

3.6 It is not the aim of this paper to provide an exhaustive comparison of all the mortality models in existence, nor even of all the extensions to the Lee-Carter model. For example, Renshaw & Haberman (2006) proposed an extension of the Lee-Carter model to include cohort effects, which was also discussed in CMI (2007), while Kingdom (2008) raised questions about the assumption of the stability over time of the Lee-Carter parameters. Other classes of models for projection include penalised-spline models — Richards *et al.* (2006) — and a wide family of models for international data is assessed by Cairns *et al.* (2007). Rather, our aim is to show that, even within this small family of models, the financial consequences of model choice can be quite substantial.

4. Graphical Description of Penalised Splines

4.1 We provide a short description of the Eilers & Marx (1996) method of *P*-splines. Figure 2 gives a plot of the log of the observed forces of mortality, denoted by $\circ$, for those aged 70 at age of death, taken from the CMI data set. There is some variation from year to year, but an underlying trend is evident, especially the very dramatic fall in log mortality which has occurred since around 1970. Familiar polynomial regression uses polynomials as the basis for regression, but there is no reason why other functions cannot be used. The lower panel shows a basis consisting of 15 cubic *B*-splines. Each *B*-spline consists of four cubic pieces bolted smoothly together at positions known as knots to give the functions illustrated — see de Boor (2001) for more details.

4.2 We suppose that the number of deaths, d_y in year y, has a Poisson distribution with mean $e_y\mu_y$, where e_y is the central exposure and μ_y is the force of mortality. The classical Gompertz model is a generalised linear model (GLM) with $\log\mu_y = a + by$. In the same way, using the *B*-spline basis in Figure 2, we have:

$$\log\mu_y = \sum_j B_j(y)\theta_j \qquad (4)$$

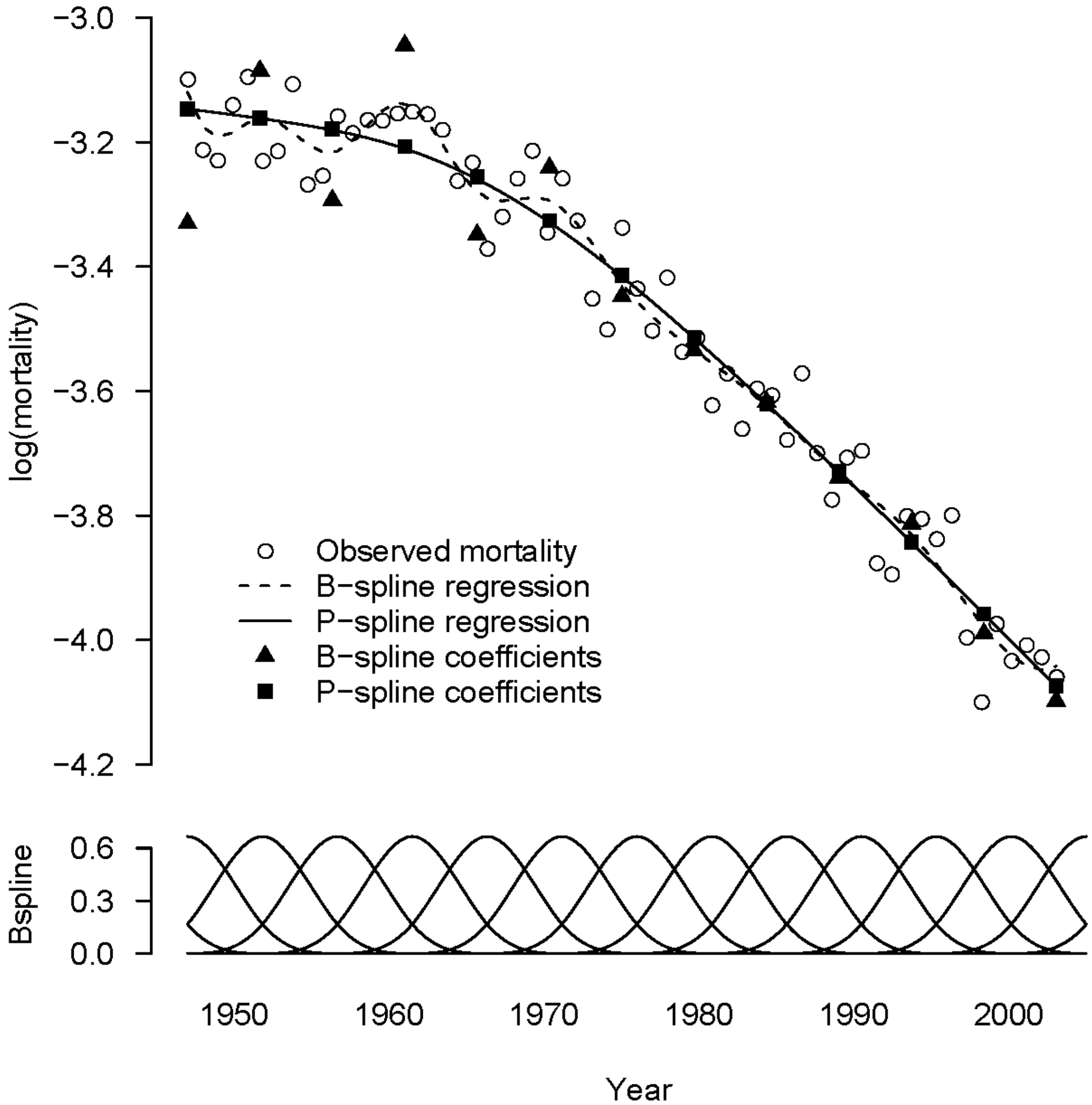

Figure 2. Log mortality for males aged 70 (CMI data) together with fitted regressions and associated coefficients; B-spline basis in lower panel

where $B_j(y)$ denotes the jth B-spline evaluated at year y; in vector/matrix notation we have:

$$\log \boldsymbol{\mu} = \boldsymbol{B\theta} \tag{5}$$

where $\boldsymbol{B} = (B_j(i))$ is the regression matrix. The resulting fitted log mortality is shown by a dashed line – – – in Figure 2. It seems that we have over-fitted or undersmoothed the data, particularly in the early years, since the dashed line oscillates rather a lot. We refer to regressing on a basis of B-splines as B-spline regression.

4.3 Each regression coefficient θ_j can be associated with its corresponding basis function, and Figure 2 also shows each $\hat{\theta}_j$, the estimated value of θ_j, plotted ▲ at its corresponding B-spline $B_j(x)$. The erratic nature of the fitted curve is a consequence of the erratic nature of the fitted coefficients. Eilers & Marx (1996) placed a penalty on differences between nearby coefficients, as in the second order penalty:

$$(\theta_1 - 2\theta_2 + \theta_3)^2 + \cdots + (\theta_{c-2} - 2\theta_{c-1} + \theta_c)^2 \tag{6}$$

where c is the number of coefficients. We note that (6) is a measure of roughness, since it increases as the fitted function becomes less smooth. We incorporate this penalty function into the log-likelihood, creating a penalised log-likelihood function. Fitting is now a balance between the goodness of fit and the roughness of the fitted curve, i.e. a balance between maximising the log-likelihood and maximising the smoothness. For more details see Richards *et al.* (2006). Figure 2 shows the results of optimising the Bayesian Information Criterion, one method of choosing the balance between fit and roughness. The coefficients, plotted ■ in Figure 2, have been 'ironed out', and the resulting fitted curve has a pleasing smoothness to it. We refer to regressing on a basis of B-splines with penalties as P-spline regression.

4.4 We can now summarise our three models for $\log \boldsymbol{M}$:

$$\text{LC}: \qquad \alpha\mathbf{1}' + \boldsymbol{\beta\kappa}' \tag{6}$$

$$\text{DDE}: \qquad \boldsymbol{\alpha}\mathbf{1}' + \boldsymbol{B}_a\boldsymbol{b\kappa}' \tag{7}$$

$$\text{CR}: \qquad \boldsymbol{\alpha}\mathbf{1}' + \boldsymbol{B}_a\boldsymbol{b}\boldsymbol{k}'\boldsymbol{B}_y' \tag{8}$$

where $\boldsymbol{B}_a$ and $\boldsymbol{B}_y$ are regression matrices evaluated on B-spline bases for age and year respectively, and $\boldsymbol{\beta} \to \boldsymbol{B}_a\boldsymbol{b}$ and $\boldsymbol{\kappa} \to \boldsymbol{B}_y\boldsymbol{k}$. In (7) there is a penalty on the coefficients $\boldsymbol{b}$, and in (8) there are penalties on both $\boldsymbol{b}$ and $\boldsymbol{k}$.

4.5 Figure 2 also illustrates how forecasting with P-splines is achieved. Linear forecasting of the last two coefficients leaves the roughness measure unchanged, and the forecast then follows from the forecast coefficients. We note that, for the purposes of projection, the knot points for the penalised splines for κ_t are set such that one sits on the final year of the data (2006 for the ONS population data, 2005 for the CMI assured lives). In this paper we have used a five-year knot spacing in order to simplify presentation, but other spacings are possible and yield similar fits.

5. Graphical Comparison of the Model Fits

5.1 Figure 3 shows parameter plots for α_x, β_x and κ_y for the three models in the left column. The solid dots show the unsmoothed values from

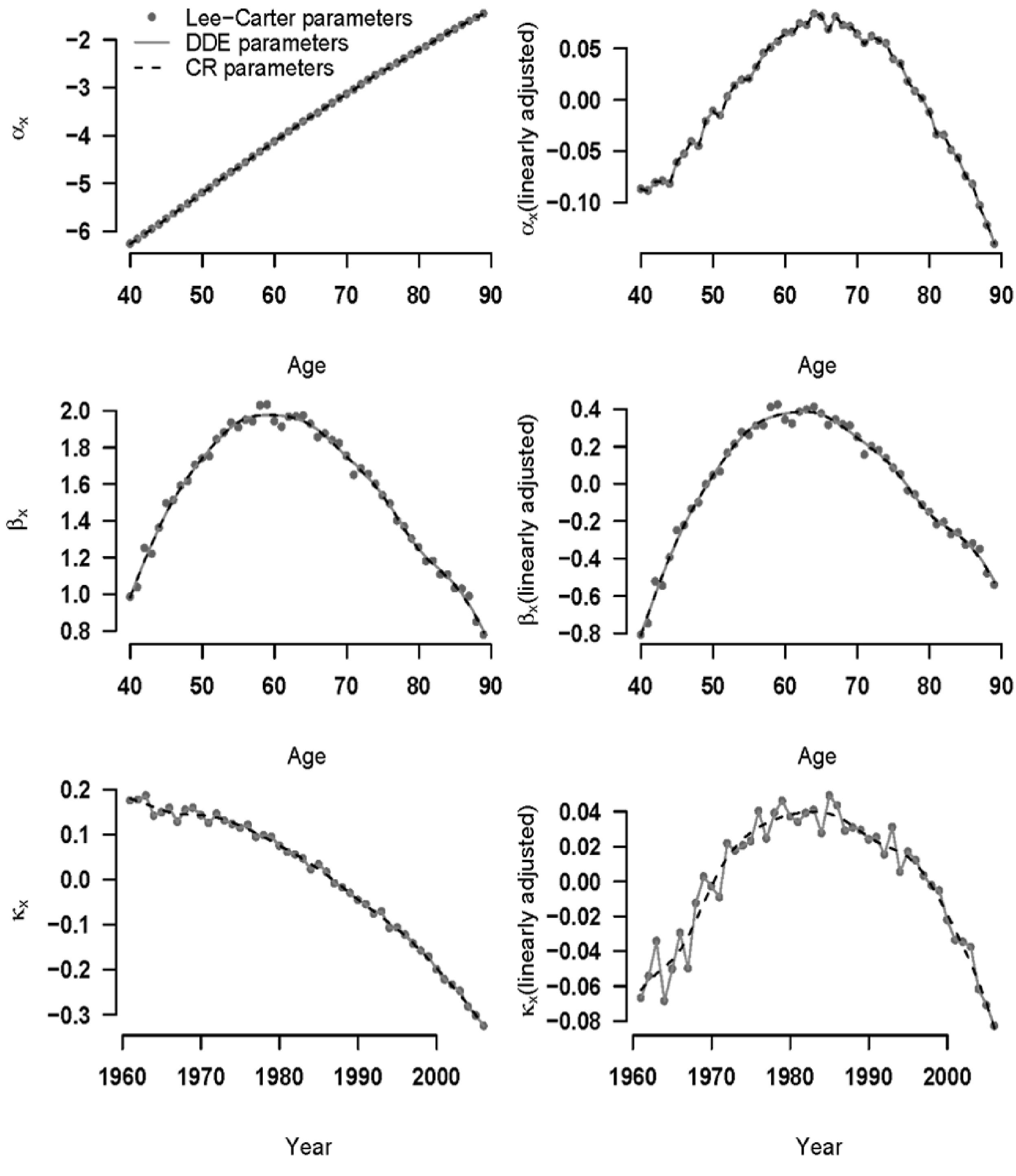

Figure 3. Parameter plots for α_x, β_x and κ_y (left column), and the same parameters after linear adjustment (right column); the original Lee-Carter parameters are shown as solid dots, while the DDE parameters are shown by a solid line and the CR parameters are shown by a dashed line; the linear-adjusted plots show the same coefficients on the left after subtracting a fitted straight line; they show, for example, that the pattern of α_x by age is not as linear as it seems (ONS data)

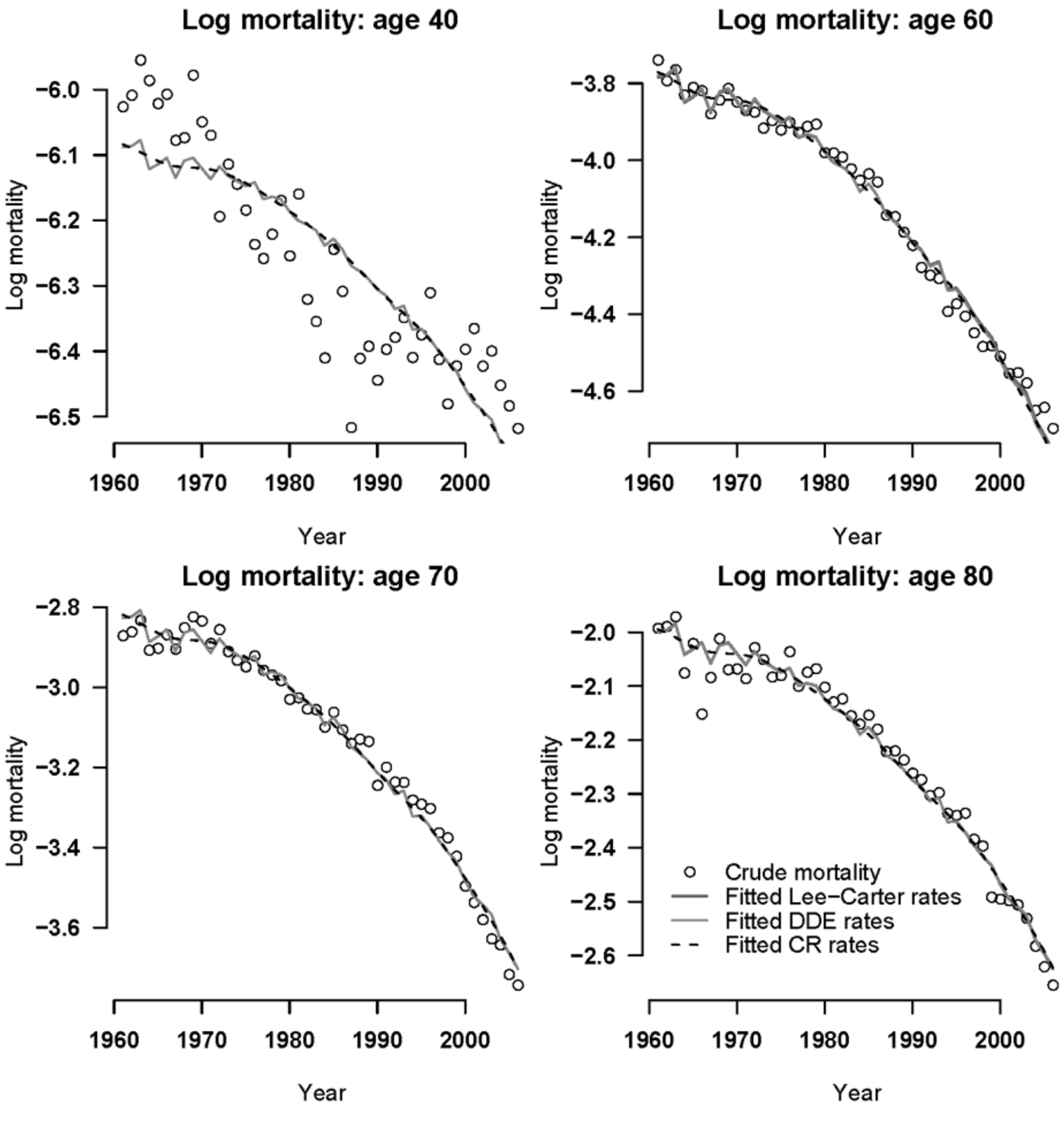

Figure 4. Log mortality at selected ages (ONS data); the dark grey line for original Lee-Carter parameters is largely obscured by the DDE line due to them being almost completely co-incident, i.e. the LC and DDE fits are almost identical

the original Lee-Carter model. In each case there is an obvious smooth pattern in the parameters, hence the extension of the DDE and the CR models to smooth β_x (DDE and CR) and κ_y (CR only).

5.2 In Figure 3 the only practical distinction between the DDE and the CR models is for the κ_y parameter, as the lines are largely coincident for the plots for α_x and β_x. The DDE model has unsmoothed κ_y values, so that the

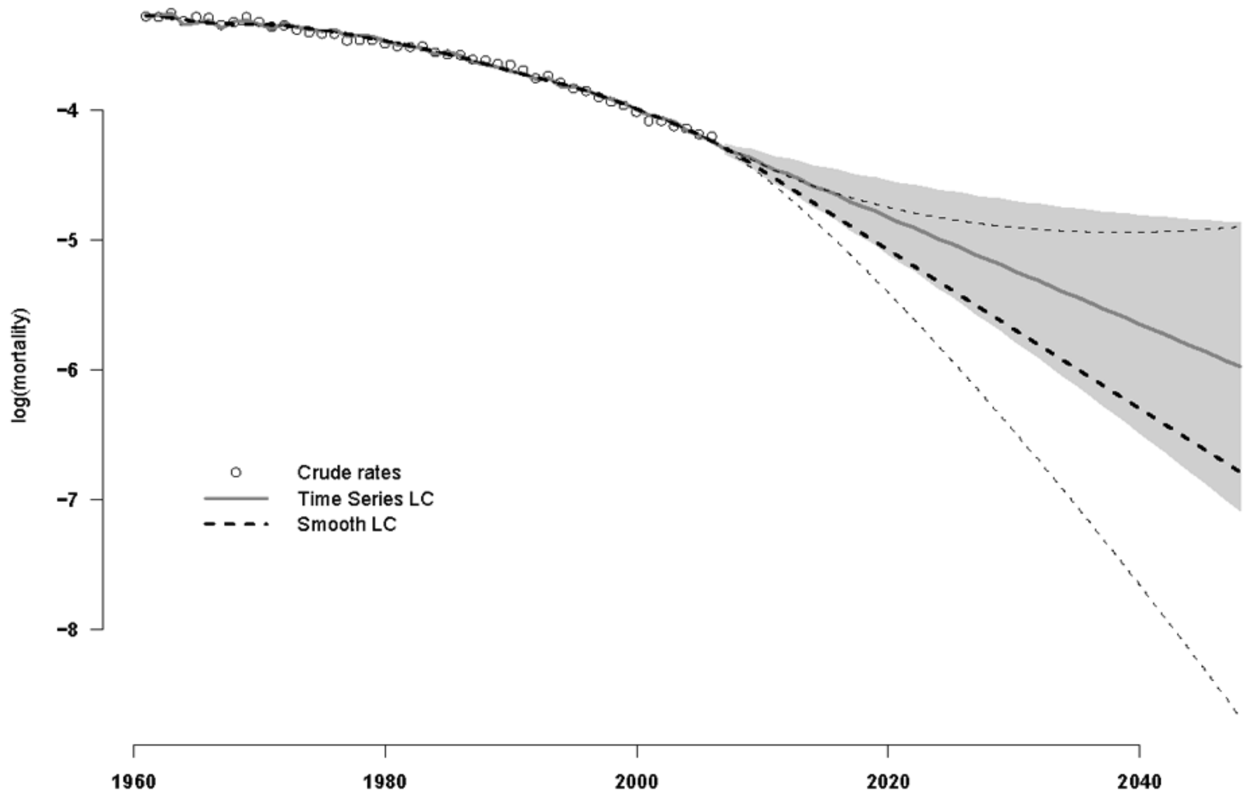

Figure 5. Mortality forecast at age 65 with 95% confidence intervals; the solid grey line is the time-series Lee-Carter forecast, together with shaded 95% confidence area; the dashed line is the smoothed-κ_y Lee-Carter forecast, together with 95% confidence bounds (ONS data for England and Wales population)

DDE parameter line, in effect, joins the dots for the original Lee-Carter model. In contrast, the CR model shows deviations from the Lee-Carter κ_y parameters, due to smoothing.

5.3 Note that the α_x parameters are unsmoothed in all of the models, hence the DDE and the CR parameter curves for α_x, in effect, join the dots for the original Lee-Carter model. The plots for α_x in Figure 3 show a high degree of regularity, and suggest that all three models are over-parameterised with respect to age; the full freedom of a separate parameter for each α_x is unnecessary when each value is so closely placed next to its immediate neighbours. Therefore, a further simplification could be achieved by spline smoothing for α_x. This is not straightforward, however, since the constraints in (3) interact with the penalty on $\boldsymbol{\alpha}$.

5.4 Figure 4 shows the crude mortality rates and the model fits under the various models. The crude mortality rates are shown as open circles, while the original Lee-Carter parameters are shown as a dark grey line. The DDE parameters are shown by a light grey line and the CR parameters are

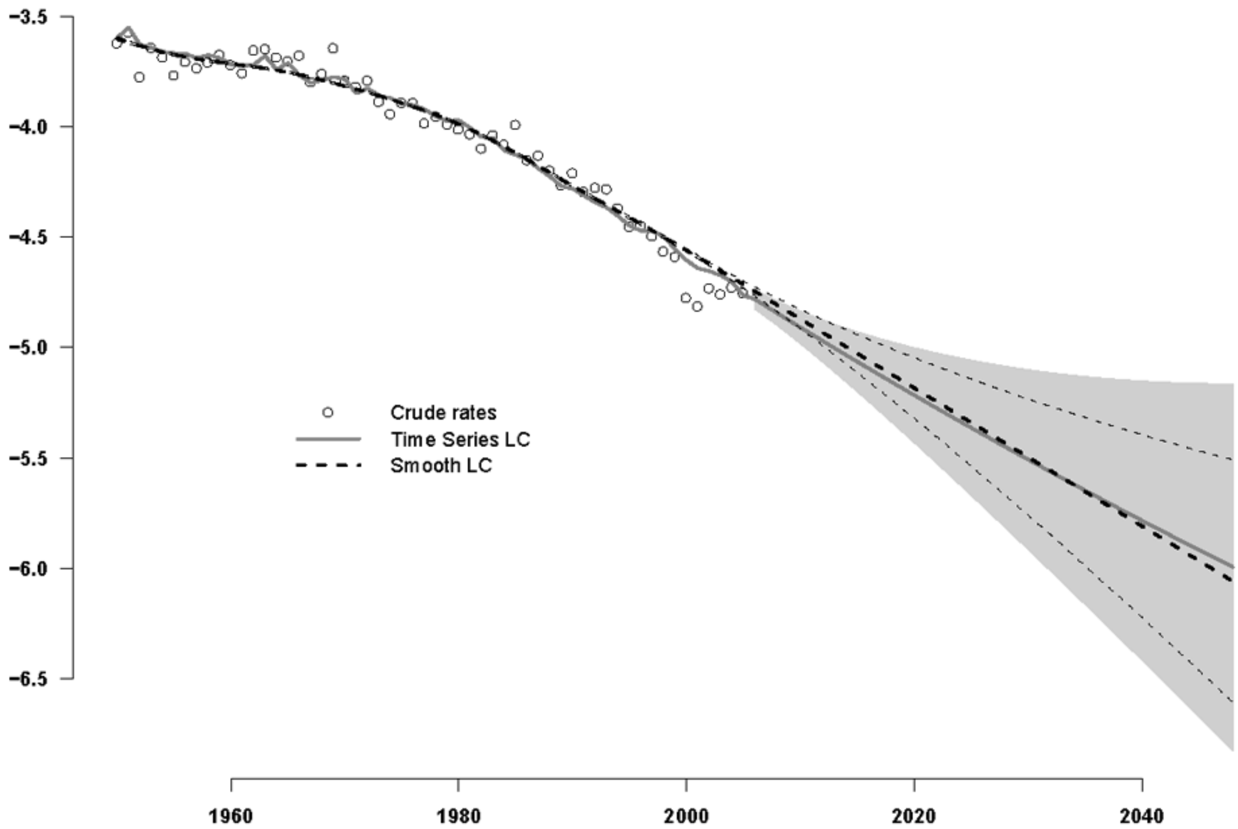

Figure 6. Mortality forecast at age 65 with 95% confidence intervals; the solid grey line is the time-series Lee-Carter forecast, together with shaded 95% confidence area; the dashed line is the smoothed-κ_y Lee-Carter forecast, together with 95% confidence bounds; in contrast to Figure 5, the projections from the two types of model are largely coincident, while this time it is the smoothed-κ_y model with the narrower confidence bounds (CMI assured lives data)

shown by a dashed line. The original Lee-Carter parameters are largely obscured by the DDE line, due to them being almost completely co-incident, i.e. the LC and DDE fits are almost identical.

5.5 Figure 5 shows the projected log-mortality at age 65 under the DDE and CR approaches (the original LC approach is left out, as it produces near-identical results to the DDE one). While the central projections are very different, we can see that the confidence bounds substantially overlap, suggesting that the projection from one model is quite consistent with the projection from the other. It is interesting that the original Lee-Carter model has a confidence area which is essentially the more pessimistic half of the confidence area for the CR (smoothed-κ_y) model. One notable feature is that the CR model has a much wider confidence interval.

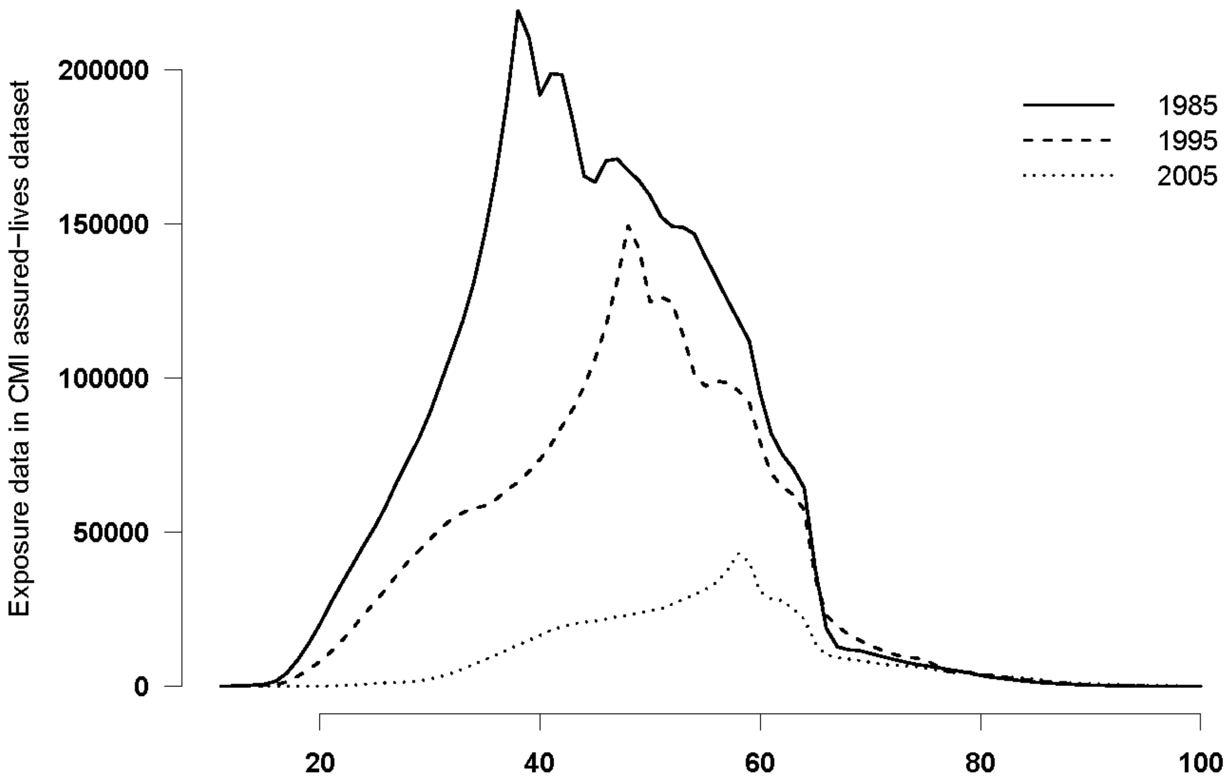

Figure 7. Exposures in CMI assured lives data set; the rapid reduction in data volumes of recent years carries the risk that the composition of the data has materially changed; the distribution by age also militates against relying on this data set for applications to post-retirement mortality

5.6 In Figure 6 the central projections are now broadly coincident, but this time it is the original Lee-Carter approach which has the wider confidence area. This is because the CMI data set is much smaller, and a much greater degree of smoothing is being applied. The smoothing is achieved by the penalty function, which is also what forms the basis of the forecast. Thus, a heavy degree of smoothing yields an apparently greater degree of certainty in the forecast. This is the reverse of Occam's Razor, where simpler models are preferable. In contrast, with smaller data sets it is hard to prove the existence of more complicated patterns, thus leading to only simplistic models being fitted with narrow confidence bands, which might give an illusion of certainty. This paradox is discussed in greater detail in CMI (2005).

5.7 One concern about the assured lives data is that the data volumes have reduced radically in recent years — see Figure 7 — which raises concerns about whether the socio-economic composition might have changed, thus affecting the projections. The other obvious comment is that the CMI data are, in any case, rather limited for post-retirement ages. It is for these reasons that most of the illustrations in the tables and figures in this paper are based on the ONS population data. There is the obvious question of whether results based on the general population are applicable to annuitants and pensioners, known as basis risk to actuaries and as bias to statisticians. In answer to this, we would say that this paper presents a methodology, not an answer, and readers can apply the methods here to the data set of their choice. For those wishing to produce projections for one population with reference to another, we refer readers to Currie *et al.* (2004).

5.8 In discussing interest rate models, Cairns (2004) wrote: "[...] probability statements derived from the use of a single model and parameter set should be treated with caution." The same can be said equally well of models of longevity risk, and is a reason for exploring trend risk with multiple data sets and multiple models.

6. Financial Comparison of the Models

6.1 One issue when presenting results from these models is what to do about the missing mortality rates above age 89. One approach would be to use mortality rates from the Human Mortality Database, and assume that each age above 89 experiences the same changes as at age 89. However, this seems somewhat arbitrary, and it risks introducing distortions for the purpose of this paper. Instead of calculating life expectancies and annuity values throughout life, we will therefore calculate temporary values up to age 90. This enables conclusions to be drawn about the various projection methodologies without worrying if they are, in part, influenced by the further assumptions above age 89.

6.2 Table 3 shows the time lived between ages 60 and 90 for a male aged 60 at outset in 2007. The first row shows the number of years lived between ages 60 and 90 at the estimated period rates in 2007. These rates are estimated using the Lee-Carter model and data up to 2006, so the small variation in the first row shows the uncertainty over projections even for just one year.

Table 3. Years lived up to age 90 for a 60-year-old male in 2007

Model	1%	5%	50%	95%	99%
Current rates	21.03	20.96	20.81	20.65	20.58
LC	24.48	24.11	23.14	22.08	21.62
DDE	24.48	24.11	23.14	22.09	21.62
CR	25.51	25.14	24.12	22.92	22.37

Source: Complete life expectancy using population data and projected current rates in 2007 (ONS data)

Table 4. Years lived up to age 90 for a 60-year-old male in 2007 according to some deterministic bases in current use

Basis	Years lived
Medium-cohort	22.97
Long-cohort	23.30
QIS4	24.65

Source: Own calculations using bases from CMI (2002) and CEIOPS (2007). Projected current rates for population in 2007, with reduction factors as listed. The QIS4 value is a 25% reduction to the central projection of mortality rates under the original Lee-Carter model, i.e. it compares with the 23.14 figure in Table 3.

6.3 The second row shows the same time lived, but this time uses the time-series projection method in the original Lee-Carter (1992) model. This projection adds 2.33 years to the time lived on current rates, but it could be as high as 3.45 years or as low as 1.04 years on the 1st and 99th percentiles, respectively.

6.4 The third row shows the same figures for the DDE model, which differs only in that the β_x values are smoothed before projection. The projection methodology is the same time-series approach as the Lee-Carter model, hence the near identical values.

6.5 The fourth row of Table 3 shows the values produced by the model with smoothed β_x and κ_y values. Unlike the LC and DDE models, the projections here are based on the penalty function used to smooth the κ_y values. As we can see, this adds anything from 0.75 to 1.04 extra years to the time lived for the original Lee-Carter model. However, even the most optimistic scenario still has an average loss of four-and-a-half years of life out of the possible 30. Table 4 shows how these life expectancies compare to those derived from some common projections in recent use.

6.6 Table 5 shows the annuity factors corresponding to the expected years lived in Table 3. To put these in perspective, Table 6 shows equivalent reserves calculated using some current deterministic bases. We can see that the medium-intensity cohort projection (‘medium cohort’) produces a lower reserve than all three central projections under the Lee-Carter models. It is

Table 5. Value of a temporary continuous annuity to age 90 for a 60-year-old male in 2007, discounting at 5% interest per annum

Model	1%	5%	50%	95%	99%
Current rates	12.50	12.48	12.41	12.35	12.32
LC	13.67	13.55	13.22	12.85	12.69
DDE	13.68	13.55	13.22	12.85	12.69
CR	14.00	13.88	13.55	13.16	12.98

Table 6. Value of a temporary continuous annuity to age 90 for a 60-year-old male in 2007, discounting at 5% interest per annum, according to some deterministic bases in current use

Basis	Reserve
Medium cohort	13.13
Long cohort	13.23
QIS4	13.78

Source: Own calculations using bases from CMI (2002) and CEIOPS (2007). The QIS4 value is a 25% reduction to the central projection of mortality rates under the original Lee-Carter model, i.e. it compares with the 13.22 figure in Table 5.

for this reason that Willets (2007) labelled the medium cohort as an extreme 'trend-reversal scenario'; mortality improvements under the medium cohort tail off too sharply to be a sensible best-estimate projection. The figure for the long cohort is almost identical to the central projections for the Lee-Carter and the DDE models. However, the long cohort is clearly weaker than the central projection for the CR model, and much weaker than the proposed Solvency II standard in QIS4 at this age.

6.7 Table 7 shows the change in the annuity factor relative to the central projection. We can see that there is a 5% chance that mortality trends will cost over $2\frac{1}{2}$% more than the central projection for each model, while, equally, there is a 5% chance that mortality trends will cost around $2\frac{1}{2}$% less. This is the normal understanding of the financial uncertainty over annuity pricing; within a given model framework, we can calculate the size of each loss along with a probability for it. The key phrase is 'within a given model framework' however, since it is not possible to know if the model is (or will be) correct. This leads us next to consider model risk, namely the consequence of a pricing actuary's model not being the correct one.

6.8 Table 5 shows the annuity factors corresponding to the expected years lived in Table 3, while Table 8 shows the change in the annuity factor relative to the current rates without the improvements. We can see that the median increase in reserve due to Lee-Carter improvements is an extra 6.48%

Table 7. Percentage change in the value of a temporary continuous annuity to age 90 for a 60-year-old male in 2007, discounting at 5% interest per annum; values from Table 5 expressed relative to the central projection (50th percentile)

Model	1%	5%	50%	95%	99%
Current rates	0.73	0.52	0	−0.52	−0.74
LC	3.45	2.50	0	−2.76	−3.97
DDE	3.45	2.50	0	−2.76	−3.97
CR	3.34	2.45	0	−2.88	−4.19

Table 8. Percentage increase in the value of a temporary continuous annuity to age 90 for a 60-year-old male in 2007, 5% interest per annum; values from Table 5 expressed relative to figures using current rates, i.e. without any projection; for comparison, the equivalent figure for the high-intensity cohort projection ('long cohort') is 6.56%

Model	1%	5%	50%	95%	99%
Current rates	0	0	0	0	0
LC	9.36	8.58	6.48	4.09	3.02
DDE	9.39	8.61	6.51	4.11	3.04
CR	12.00	11.26	9.16	6.58	5.37

in cost, although it could be as high as 9.36% or as low as 3.02%, according to the 1st and 99th percentiles, respectively. However, the stronger mortality improvements resulting from smoothing the κ_y values has resulted in a further increase in the reserve of 2.68% on the central scenario. This is broadly similar in size to the uncertainty cost referred to in ¶6.7 and in Table 7, suggesting that model risk is just as important as trend risk. The problem is that you can quantify uncertainty within a given model, but you cannot quantify the uncertainty over the model itself.

6.9 To put all these figures in a financial perspective, the typical pricing margin for an immediate annuity is around 4% to 5% of premium. The uncertainty over which model to use can either cut calculated annuity profits in half, or else increase them by half. This is quite separate from the uncertainty over trend, which can itself have the same effect. The model does not directly affect the profitability itself, just its measurement.

6.10 Another important risk which we have not considered is basis risk, namely whether projections based on the population data or on assured lives apply to pensioners and annuitants. People in receipt of private pensions and annuitants are a select sub-group of the population in general, and may not exhibit the same pattern of mortality trends. There may be felt to be less basis risk with the CMI assured lives data, but these mainly comprise holders of endowment policies at ages younger than 65. The population data cover the right age range, but not the select group which pensioners are. Equally, the CMI data cover private insured lives, but do not cover the right age range. Neither seems wholly satisfactory, and so some degree of basis risk must remain as long the data set for projections is not the same as the population whose benefits are being valued.

7. Reserving for Longevity Risk

'*The rates of mortality or morbidity should contain prudent margins for adverse deviation [...]. In setting those rates, a firm should take account of [...] anticipated or possible future trends in experience [...] but only where they increase the liability.*'

FSA (2008), INSPRU ¶1.2.60(5)

'*The longevity shock to be applied is a (permanent) 25% decrease in mortality rates for each age.*'

CEIOPS (2008) ¶TS.XI.C.6

'*For recovery plans based on valuations with effective dates from March 2007, mortality improvement assumptions that appear to be weaker than the long cohort assumption will attract further scrutiny and dialogue with the trustees where appropriate. Furthermore, assumptions which assume that the rate of improvement tends towards zero, and do not have some form of underpin, will also attract further scrutiny.*' Pensions Regulator (2008a) ¶2.7

'*though long cohort with some form of underpin will be used when looking at the secondary trigger, a medium cohort assumption with a stronger underpin would clearly be equivalent.*'

Pensions Regulator (2008b)

7.1 One way of looking at trend risk is to consider the probability of reserve adequacy. This is done for the DDE and the CR models in Figure 8. The curves plotted are the ogives under each model, i.e. the cumulative probability distribution function that the annuity reserve factor on the horizontal axis will be adequate. The ogive for the CR model is markedly to the right of the ogive for the DDE model, because the CR model requires higher reserves for the same probability of reserve adequacy. This is because the CR model projects faster improvements, as shown in Figure 5. The reserve factors for the medium and high-intensity cohort projections are also marked, together with the 99.5% stress scenarios under each model.

7.2 Figure 8 illustrates one of the difficulties of model risk; the 99.5% stress scenario under the DDE model might just be regarded as prudent under the CR model. Note that it is important to consider reserving margins as a whole, and it is particularly important not to overdo things by combining 99.5% stress scenarios for all the various risks. Thus, the 99.5% stress scenario in Figure 8 is a result of stressing the trend assumption, but any reserve sufficiently above 50% might be regarded as prudent with respect to trend risk.

7.3 Figure 8 also shows the reserves calculated according to two bases in recent use. The long cohort could function as a best estimate under the DDE model, but it could not be regarded as prudent. Under the CR model the long cohort would be too weak to be even a best-estimate, let alone a prudent reserving basis. The medium cohort would be regarded as inadequate under either model. The comparison with the 99.5% values is also instructive, since it suggests that the long cohort cannot be regarded as a stress scenario for ICA purposes. Figure 8 further suggests that, at this age at

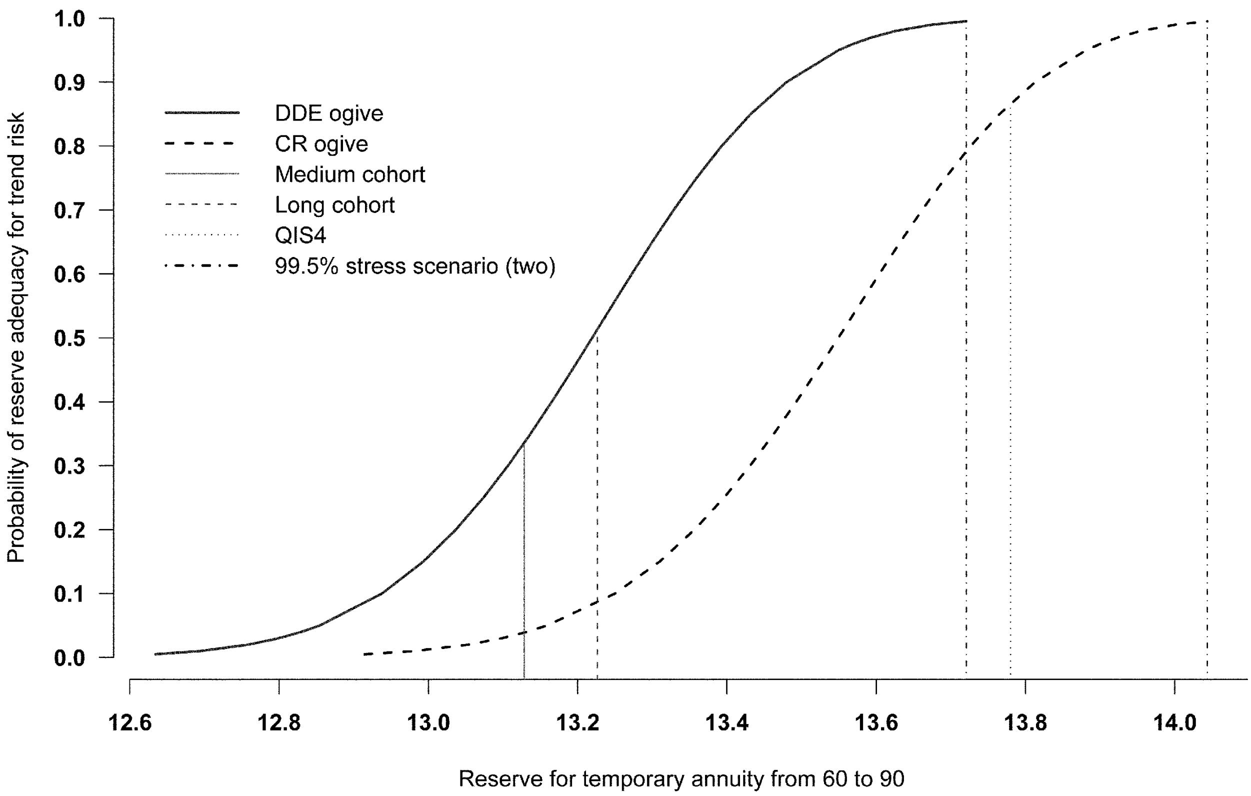

Figure 8. Probability of reserve adequacy against trend risk according to the DDE model (solid line) and the CR model (dashed line); temporary 30-year continuous annuity to male aged 60 at outset, valued at 5% interest per annum using population mortality (ONS data); for a given annuity reserving factor on the horizontal axis, you can read off the probability that it will be adequate for trend risk under the two models; the '99.5% stress scenario' is the reserve using the 99.5th worst projection under each model

least, the Pensions Regulator (2008a) was reasonable in requiring a strengthened version of the long cohort for pension scheme reserving to be considered prudent. Prudence is, of course, a matter of judgement, and therefore an opinion, rather than an absolute value. Equally, however, a claim of prudence has to be substantiated, and it is hard to do this convincingly without reference to a statistical projection model.

7.4 Figure 8 also shows the reserve calculated according to the proposed Solvency II standard known as QIS4 (CEIOPS, 2007). Under the DDE model, this would be regarded as a beyond-ICA stress scenario, with a probability of 99.8% of reserves at this level being adequate. However, under the CR model the probability of QIS4 reserves being adequate is a mere prudent-seeming 86.7%. Figure 8 suggests that the 25% shock in QIS4 is not an unreasonable reserving standard for new annuity business written around age 60.

7.5 All these calculations are for a specimen level pension at age 60, so it is important to consider the effect for some real portfolios, which we will do in the next section.

8. Portfolio Simulations

'*trustees [...] must take advice from their actuary [...] on best estimates and on appropriate margins for prudence. This may be by way of stochastic modelling to illustrate the variability of outcomes and their relative likelihood.*'

Pensions Regulator (2008a) ¶1.12

8.1 Leaving aside economic assumptions like interest rates, there are three components of longevity risk which we can test via simulation. We will assume that current base mortality is known, although this is not the case in practice, and is therefore also a component of longevity risk. The first component is the uncertain direction of future trends, which includes uncertainty about the model for trends as well as uncertainty within the model. The second is the binomial risk for a particular portfolio's experience, namely who happens to die when. The third is the concentration risk, where a large proportion of the financial liabilities is concentrated in a relatively small number of lives. Table 7 shows the impact of trend risk on a specimen annuity, but it is instructive to look at the impact on some actual portfolios whose age distributions are shown in Figure 9. As well as considering trend risk, we can also simulate the portfolio in run-off to examine binomial and concentration risk as well. The recent availability of inexpensive computers with multiple floating-point cores has made it possible to simulate entire portfolios quickly. This is in contrast to the error-prone approach of trying

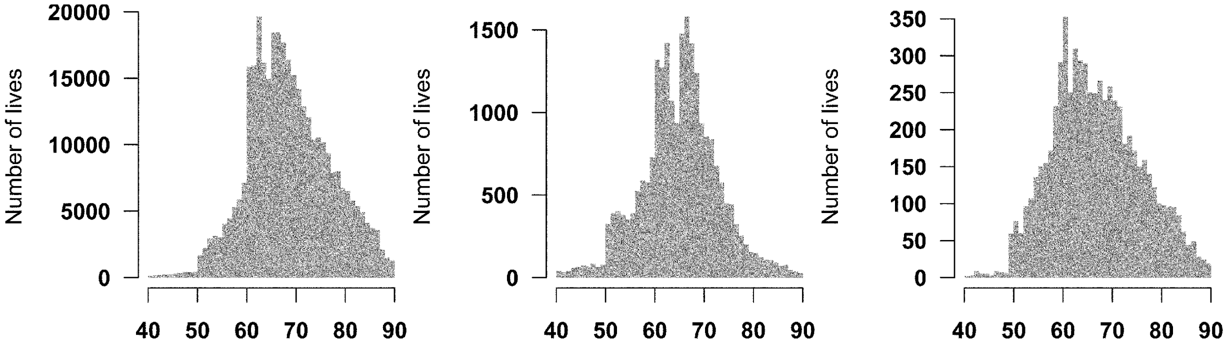

Figure 9. Age distribution of three portfolios, at 1 January 2007, for pensions or annuities in payment to lives aged between 40 and 90; the larger portfolio on the left contains both IFA introduced open-market annuities and internal-vesting business from the insurer's personal pensions (see also Table 9); the medium-sized portfolio in the middle contains internal-vesting annuities only (see also Table 10), while the small portfolio on the right is a defined benefit pension scheme (see also Table 11)

to pick a handful of policies which are supposed to be representative of the portfolio as a whole. Using purpose-written C++ programs, we can simulate the larger of the two annuity portfolios here in run-off 10,000 times in around an hour on an eight-core server.

8.2 Tables 9 and 10 show the results of 10,000 simulations of two different annuity portfolios. As expected, the smaller portfolio has to hold proportionately more extra capital for a given level of certainty. Larger portfolios benefit from the law of large numbers, as the binomial experience variation is proportionately less. Thus, there is a direct capital benefit from scale in the annuity business. However, the benefit is relatively modest, as the difference at the 99.5% level is just 0.62% of the median value, despite the larger portfolio being more than ten times the size of the smaller one. Conversely, this shows that annuity portfolios and pension schemes are exposed to a non-diversifiable amount of longevity trend risk, and that trend risk usually dominates binomial experience risk above modest portfolio sizes. For the large annuity portfolio, the extra cost at the 99.5% level is just 0.50%

Table 9. Percentage variation around the median run-off cost for a large annuity portfolio

Trend risk	Measure	Min%	0.5%	1%	5%	95%	99%	99.5%	Max%
No	Lives	−0.29	−0.20	−0.18	−0.12	0.13	0.18	0.20	0.26
	Amounts	−0.70	−0.51	−0.46	−0.32	0.31	0.45	0.50	0.74
Yes	Lives	−4.95	−3.25	−2.91	−2.01	1.85	2.56	2.88	4.89
	Amounts	−5.44	−3.61	−3.24	−2.21	2.02	2.73	3.12	5.21

Source: Own calculations using 10,000 simulations of a portfolio of 207,190 males aged between 40 and 90, with an amounts-weighted average age of 72.52. Temporary annuities to age 90 are valued continuously at 5% interest per annum. Projections are according to the CR model and population mortality. The age profile of this portfolio is given in the left panel of Figure 9.

Table 10. Percentage variation around the median run-off cost for a small annuity portfolio

Trend risk	Measure	Min%	0.5%	1%	5%	95%	99%	99.5%	Max%
No	Lives	−0.92	−0.66	−0.59	−0.42	0.41	0.57	0.63	0.87
	Amounts	−1.78	−1.12	−1.00	−0.71	0.68	0.94	1.07	1.43
Yes	Lives	−6.43	−4.11	−3.67	−2.52	2.15	3.08	3.44	4.31
	Amounts	−6.58	−3.98	−3.61	−2.52	2.18	3.10	3.50	4.74

Source: Own calculations using 10,000 simulations of a portfolio of 15,429 males aged between 40 and 90, with an amounts-weighted average age of 67.3. Temporary annuities to age 90 are valued continuously at 5% interest per annum. Projections are according to the CR model and population mortality. The age profile of this portfolio is given in the middle panel of Figure 9.

Table 11. Percentage variation around the median run-off cost for a small pension scheme

Trend risk	Measure	Min%	0.5%	1%	5%	95%	99%	99.5%	Max%
No	Lives	−3.55	−2.09	−1.85	−1.31	1.30	1.86	2.02	2.99
	Amounts	−7.64	−4.98	−4.54	−3.25	2.90	4.06	4.52	5.90
Yes	Lives	−10.11	−7.25	−6.53	−4.33	3.77	5.19	5.73	7.56
	Amounts	−12.72	−8.31	−7.37	−5.19	4.66	6.55	7.23	10.57

Source: Own calculations using 10,000 simulations of a portfolio of 2,268 males aged between 40 and 90, with an amounts-weighted average age of 67.2. Temporary annuities to age 90 are valued continuously at 5% interest per annum. Projections are according to the CR model and population mortality. The age profile of this portfolio is given in the right panel of Figure 9.

without trend risk, whereas it is multiplied six-fold to 3.12% if trend risk is included. Even for the smaller annuity portfolio, the extra cost at the 99.5% level is 1.07% without trend risk, whereas it more than triples to 3.50% if trend risk is included.

8.3 However, trend uncertainty is not always the dominant part of the mortality risk, as shown in Table 11. In this example, the combination of binomial risk and concentration risk comprises nearly two-thirds of the variation in cost at the 99.5% level; with trend risk the extra cost relative to the median is 7.23%, whereas without trend risk it is 4.52%. Indeed, the vast majority of pension schemes in the U.K. are much smaller than the one in Table 11; according to the GAD (2005) there were 7,470 private sector defined benefit pension schemes in the U.K. with fewer than 100 members. For such schemes, binomial and concentration risk often dominate all other sources of risk, leading Richards & Jones (2004) to recommend that schemes of this size should consider annuity purchase as their default investment option.

8.4 Figure 10 shows the probability of adequacy for a large annuity portfolio under the DDE and the CR models using population data. The model risk is considerable; whereas a long-cohort reserve is a prudent best estimate under the DDE model, it is wholly inadequate under the CR model. In one sense the comparison is a little harsh, as the projection basis is being required to allow for binomial risk and concentration risk, as well as trend risk. In practice, of course, a mortality basis would include a margin for adverse deviation (MAD) in the base table rates, which would have the effect of shifting the reserve lines in Figure 10 to the right. However, the comparison is not so unreasonable for this portfolio, as trend risk dominates the other two, as shown in Table 9.

8.5 An oft-unappreciated element of model risk is the choice of data set, and this is illustrated in Figure 11. Here the reserves are much higher due to the lower rates of mortality experienced by assured lives. The projections are

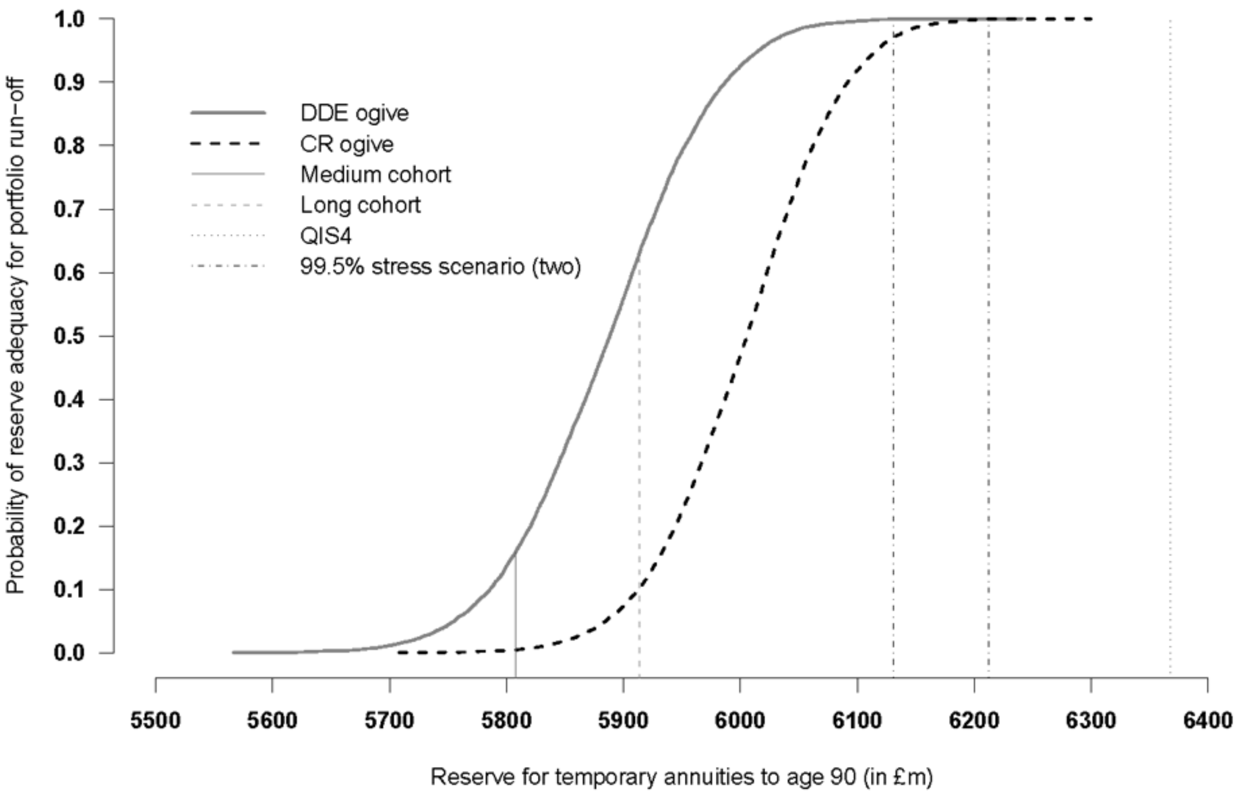

Figure 10. Probability of reserve adequacy for a large annuity portfolio against combined binomial risk, concentration risk and trend risk according to the DDE model (solid line) and the CR model (dashed line); temporary 30-year continuous annuities payable to males until age 90, valued at 5% interest per annum using population mortality (ONS data)

now quite different, as shown in Figure 6, and in Figure 11 the long-cohort projection appears beyond prudent as it produces reserves beyond what were required in any of the 10,000 simulations.

8.6 In both Figures 10 and 11 the reserve under the QIS4 stress scenario is higher than any of the 10,000 simulations under either model or data set. This suggests that the QIS4 shock of a 25% permanent fall in mortality rates is perhaps over-prudent for very mature portfolios, while it does not look at all unreasonable at new business ages in Figure 8.

8.7 On each occasion, we have fitted a model within the Lee-Carter framework. Whether or not we choose to smooth the κ_y values results in very different projections, with very different pictures of what is an adequate reserve for trend risk. Equally, even within the same model-fitting framework, using a different data set again gives a radically different picture. Both choices of model and of data set are aspects of model risk, which has been shown here to be very important for annuities and pensions. There are, of course, many other models available for mortality, which can only add to model risk. In light of this, it would seem sensible for insurers to have large margins in their reserving basis for the highly uncertain direction of future

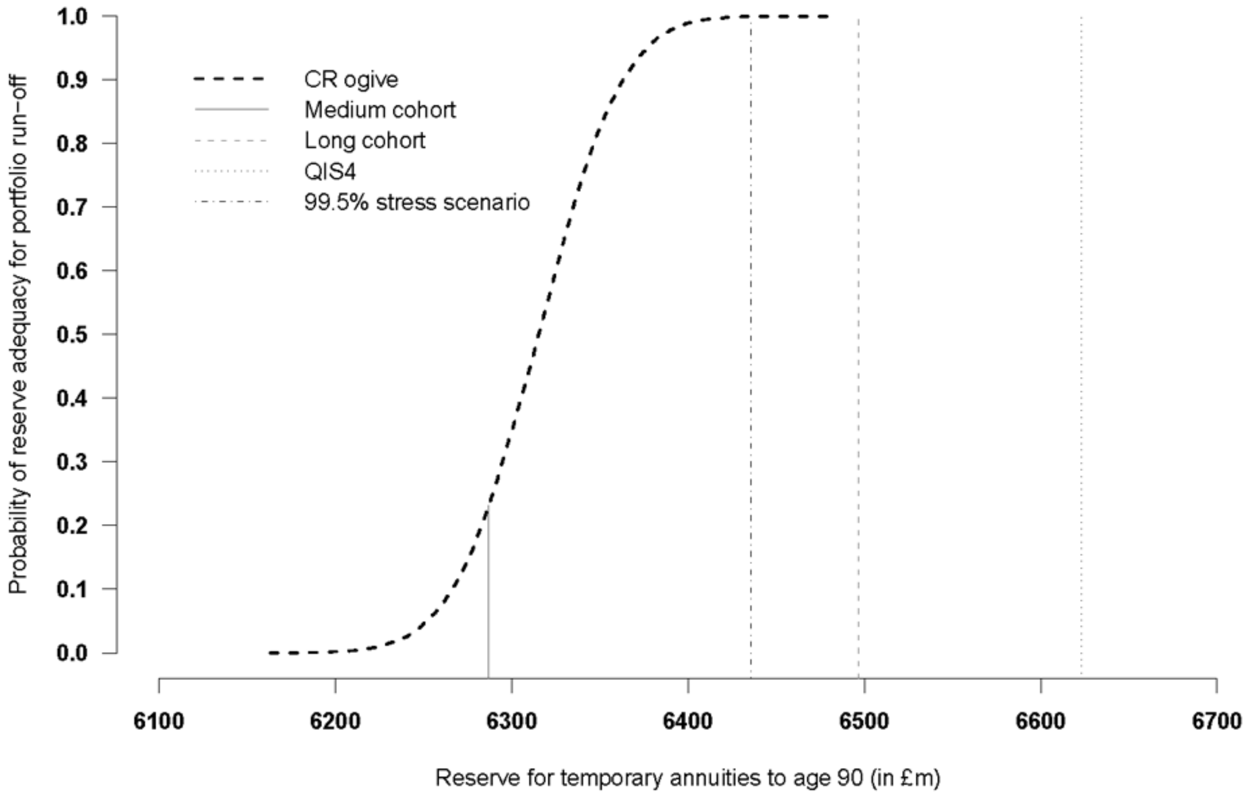

Figure 11. Probability of reserve adequacy for a large annuity portfolio against combined binomial risk, concentration risk and trend risk according to the CR model (dashed line); temporary 30-year continuous annuities payable to males until age 90, valued at 5% interest per annum using assured-lives mortality (CMI data)

mortality improvements. Equally, insurance company shareholders need to make sure that the pricing of annuity business adequately compensates them for the undiversifiable risk which they run.

9. Conclusions

9.1 This paper has presented a smoothed approach to the fitting of the parameters in a Lee-Carter model for mortality. The smoothing of the time component κ_y allows an alternative means of projection to the usual time-series approach. This approach does not materially change the fitted values of $\mu_{x,y}$, but it does change the projections. We emphasise that we have only touched on the problem of model risk, the first of the four problem areas identified in ¶3.2. Forecasting methods other than the Lee-Carter model could be used, or, indeed, other variants of the Lee-Carter model itself. Within the Lee-Carter model the principal source of variation is the variation in $\boldsymbol{\kappa}$, and this we have addressed. The variation and the stability of both $\boldsymbol{\alpha}$

and $\boldsymbol{\beta}$ have not been addressed, and these sources of variation can only add to our uncertainly about the future. Forecasting of mortality should be approached with both caution and humility.

9.2 This paper shows how the measurement of uncertainty within a given model framework is financially material to writers of immediate annuities. However, the presentation of alternative projections within the same framework shows how model uncertainty is just as important financially. This model risk limits what can be expected of long-term mortality projections, and serves as a reminder as to why explicit margins for prudence are required in pricing and in reserving for pensions and annuities.

9.3 It is not the aim of this paper to provide all the answers to the question of projecting mortality and longevity risk. However, it may prompt life office boards to ask questions about annuity pricing; if model risk is so significant, is a 5% pricing margin enough for the risk in guaranteed annuities at ages 60 to 65? Equally, trustees and employers may probe scheme funding bases; should companies with defined-benefit pension plans not buy out liabilities while there is still capacity and appetite from insurers? Perhaps the actuary's role should be a little less about calculating the value of the liability, and much more about demonstrating the depths of what we do not — and cannot — know.

Acknowledgements

The authors thank Adrian Gallop of the GAD for the population experience data, and the CMI for the assured lives data. The authors also thank two anonymous life offices for providing snapshots of their in-force annuity portfolios. The authors also thank Gavin Ritchie for invaluable help in implementing the data validation and deduplication processes, and two anonymous referees for their comments. Iain Currie thanks the Faculty and the Institute of Actuaries for a research grant supporting this paper. Any errors or omissions remain the sole responsibility of the authors.

Model fitting was done in R, which was also used for all graphs. Deduplication of annuity records was done in Longevitas. Stochastic simulation was done in purpose-written C++programs. The text was prepared in pdfTEX.

References

Brouhns, N., Denuit, M. & Vermunt, J.K. (2002). A Poisson log-bilinear approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, **31**(3), 373-393.

Cairns, A.J.G. (2004). A family of term-structure models for long-term risk management and derivative pricing. *Mathematical Finance*, **14**(3), 415-444.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A. & Balevich, I. (2007). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. Working paper, Heriot-Watt University.

CEIOPS (Committee of European Insurance and Occupational Pensions Supervisors) (2007). QIS4 technical specifications, December 2007.

CMI (Mortality Sub-Committee) (2002). An interim basis for adjusting the 92 Series mortality projections for cohort effects. Working paper No. 1.

CMI (Mortality Committee) (2005). Projecting future mortality: towards a proposal for a stochastic methodology. Working paper No. 15.

CMI (Life-Office Mortality Committee) (2007). Stochastic projection methodologies: Lee-Carter model features, example results and implications. Working paper No. 25.

Currie, I.D., Durban, M. & Eilers, P.H.C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, **4**, 279-298.

de Boor, C. (2001). A practical guide to splines. *Revised edition. Applied Mathematical Sciences*, **27**. Springer-Verlag, New York.

Delwarde, A., Denuit, M. & Eilers, P.H.C. (2007). Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: a penalised likelihood approach. *Statistical Modelling*, **7**, 29-48.

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

FSA (2008). Prudential sourcebook for Insurers (INSPRU). Financial Services Authority, United Kingdom.

Government Actuary's Department (2005). Occupational pension schemes: the thirteenth survey by the Government Actuary. GAD.

Kingdom, J. (2008). Mortality improvement. *Reinsurance News*. Society of Actuaries, August 2008, 14-17.

Lee, R.D. & Carter, L. (1992). Modelling and forecasting the time series of U.S. mortality. *Journal of the American Statistical Association*, **87**, 659-671.

Longevitas Development Team (2008). Longevitas v2.3. Longevitas Ltd, Edinburgh, United Kingdom. URL http://www.longevitas.co.uk

Pensions Regulator (2008a). Good practice when choosing assumptions for defined benefit pension schemes with a special focus on mortality. Consultation document, February 2008.

Pensions Regulator (2008b). Good practice when choosing assumptions for defined benefit pension schemes with a special focus on mortality. Response to consultation document, September 2008.

R Development Core Team (2004). R: a language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria. ISBN 3-900051-07-0, URL http://www.r-project.org

Renshaw, A.E. & Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, **38**(3), 556-570.

Richards, S.J. & Jones, G.L. (2004). Financial aspects of longevity risk. Paper presented to the Staple Inn Actuarial Society.

Richards, S.J., Kirkby, J.G. & Currie, I.D. (2006). The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal*, **12**, 5-61.

Richards, S.J., Ellam, J.R., Hubbard, J., Lu, J.L.C., Makin, S.J. & Miller, K.A. (2007). Two-dimensional mortality data: patterns and projections. *British Actuarial Journal* (to appear).

Richards, S.J. (2008a). Detecting year-of-birth mortality patterns with limited data. *Journal of the Royal Statistical Society, Series A (2008)*, **171**(1), 279-298.

Richards, S.J. (2008b). Applying survival models to pensioner mortality data. *British Actuarial Journal* (to appear).

Willets, R.C. (2007). Recent developments in mortality. Paternoster presentation.

Printed by Bell & Bain Ltd., Glasgow