Continuous Mortality Investigation

Mortality Committee

Working Paper 15

Projecting future mortality: Towards a proposal for a stochastic methodology

July 2005

Continuous Mortality Investigation

Mortality Committee

Working Paper 15

Projecting future mortality: Towards a proposal for a stochastic methodology

1. Introduction

1.1 *Projecting Future Mortality*

The subject of mortality improvements has been the focus of much work in recent years. The Projections Working Party of the Mortality Committeee of the CMI is tasked with producing projections of future mortality rates, and therefore projections of mortality improvements. The most recent set of projections were presented in Working Paper 1, (CMI, 2002) which reviewed historic methods, and then briefly introduced the new projections. These were distinctive in three important respects.

Firstly, they recognised the so-called cohort effect; that is, the dependence of mortality improvement rates on a person's year of birth. See Willets (1999) for more details, and also Willets (2004).

Secondly, these projections were extrapolations based on the results of a methodology new to actuaries, namely penalised spline regression (or P-splines). Details of the application of this technique to mortality data are described in Durban, Currie and Eilers (2002) and also Currie, Durban & Eilers (2003). Working Paper 3 (CMI, 2004) described this and other projection methodologies commonly used outside the actuarial profession to model future mortality, and discussed in detail the practical problems and issues surrounding mortality projection.

Thirdly, three alternative projections were offered, instead of the traditional single projection. This was an explicit recognition of the uncertainty of long-term mortality projections, although it was not a probabilistic statement.

The subject of uncertainty about mortality projections required much more work, however, not least because of recent changes in the regulatory environment and professional guidance for actuaries:

> "Where there is a considerable range of possible outcomes, the FSA expects firms to use stochastic techniques to evaluate these risks. In time, for example, longevity risk, where this constitutes a significant risk for the firm, may fall into this category."
> *PRU 7.3.18, Financial Services Authority (FSA).*

> "If it is intended to use a combined economic and mortality stochastic model to value deferred annuities, guaranteed annuity options or other liabilities, the stochastic

> variation most relevant is likely to be in the general rate of improvement of longevity rather than variation in individual longevity." *Manual of Actuarial Practice, GN46, V1.1 B46.13.*

With the change in the regulatory environment for life insurance companies, most obviously with the introduction of Individual Capital Assessments (ICAs) and the expectation of stochastic modelling, actuaries should have the option of stochastic projection of future mortality as well as the traditional deterministic approach. While ICAs are a life-office phenomenon, all actuaries examining longevity liabilities — including defined-benefit pension schemes — will be interested in the uncertainty over long-term mortality projections. Thus, both pensions actuaries and life actuaries should have the option of using a model (or models) for stochastic projection of future mortality.

The production of such a model (or models) was a major piece of new work, and the first step the Working Party took was to host a joint seminar on mortality projection with the GAD in October 2003. The purpose of this seminar was to hear speakers from outside the actuarial profession. The second step was to produce a discussion paper on the issues, which was Working Paper 3. The third step was to consult the profession on these issues, which was done by asking questions at the end of Working Paper 3 followed by a meeting in Staple Inn Hall in June 2004. The responses made both at the meeting and in subsequent private letters were summarised in Working Paper 11 (CMI, 2005a). The fourth step was that the CMI sponsored a CASE award, in conjunction with the Engineering and Physical Sciences Research Council, to support a PhD project at Heriot-Watt University, under the supervision of Dr Iain Currie, into smoothing and projection methodologies for longevity projection. The fifth step was that the CMI supported a two-day technical workshop on mortality projection in Edinburgh in September 2004, organised by Dr Currie. The next step in this process is to describe our work to date and to indicate our intended modelling framework(s) for stochastic mortality projection. These are the subjects of this working paper, which essentially adds a probabilistic approach to the projections in Working Paper 1. This is by no means the last stage in the process, as the Working Party will here raise further issues and topics for research and consultation.

Working Paper 1 introduced the three deterministic cohort projection bases, but explicitly left it up to individual actuaries to decide which one to use for a specific purpose, or whether to use a cohort projection basis at all. This working paper introduces a stochastic modelling framework, but does not provide any recommended calibrations.

This working paper has been prepared for the Mortality Committee of the CMI by a Working Party consisting of Angus Macdonald, Adrian Gallop, Keith Miller, Stephen Richards, Rajeev Shah and Richard Willets. It has been approved by the Committee.

1.2 *Summary of Working Paper 3*

Working Paper 3 was a discussion paper. It contained the deliberations of the Projections Working Party to date, including discussion of many of the issues raised following the seminar in October 2003. Some of these issues were formulated as explicit questions for the profession, and details of these questions and the responses were published in Working Paper 11 (summarised in the next section).

Working Paper 3 described the basis and background to the cohort effect, and the interim cohort-based projections which resulted in Working Paper 1. Working Paper 3 also contained a description of a joint seminar (with the GAD) held in Edinburgh on 6 October 2003 to discuss the views and approaches of demographers, statisticians and gerontologists, all of whom have a strong professional interest in the projection of future mortality and its underlying causes. Further details about the seminar can be found in Working Paper 3. A second seminar was held in Edinburgh in September 2004 to focus on particular features of the Lee-Carter method of stochastic mortality projection — see Currie *et al.* (2004).

Working Paper 3 described the background to the need for a new set of projections. This included the need to give some indication of uncertainty in projections, and to allow a more transparent approach to risk management. These are required as insurers are increasingly moving, or being moved, towards the use of such risk management tools, not least because of the IASB 'fair valuation' rules, FSA 'realistic balance sheet' requirements, and convergence of regulatory regimes in banking and insurance. Greater insight into these drivers of new mortality projections, particularly the interpretation of the FSA's requirements, can be found in Working Paper 3.

An overview of different projection methodologies was given in Section 2 of Working Paper 3, contrasting their strengths and weaknesses. Process-based projections attempt to model trends in causes of death, although this approach is not favoured because of problems in death classification and insufficient understanding of the major cause-of-death processes. Extrapolative methods are based on projecting historical trends in mortality into the future, although all such methods include some element of subjective judgement, for example in the choice of period over which the trends are to be determined. Examples of different approaches to extrapolation are discussed in Section 2.1 of Working Paper 3.

Section 2.2 of Working Paper 3 considered the various types of model in use, including the current CMI methodology, and the methodology used by the GAD for projecting mortality in the official national population projections for the U.K. and its constituent countries.

Special consideration was given to the treatment of uncertainty about future projections. In particular, three well-known sources of uncertainty associated with the use of statistical models were discussed:
(a) model uncertainty;
(b) parameter uncertainty; and
(c) stochastic uncertainty.

Section 4 of Working Paper 3 discusses uncertainty in more detail, and further insights can be found in Cairns (2000).

Working Paper 3 also discussed in detail the issues associated with fitting data and making projections. The first major question is what data set to use, and there are particular problems associated with the lack of suitable annuitant mortality experiences. Considerable care is required with projections, particularly where the experience of an annuitant population of insufficient size could potentially lead to implausibly narrow confidence intervals in projections. This problem was illustrated by an example where there were insufficient data to refute the fitting of a straight line to the mortality trend, which, by virtue of the model structure, resulted in very narrow confidence intervals around the projection. Another issue specific to regression models was that traditional polynomial methods can yield acceptable fits in the region of the data, and yet produce very poor projections outside it.

Working Paper 3 contains a lot of detailed, technical discussion and is a useful backdrop to understanding this working paper.

1.3 *Summary of Working Paper 11*

Working Paper 3 was set out as a discussion paper and asked eight specific questions of the profession. Responses to these questions were received during a seminar at Staple Inn on 4 June 2004, and also in writing from some life offices and reinsurers. These questions and a summary of the responses are set out below. For fuller details, see Working Paper 11.

The first question was what base tables and projections were currently used by life offices[1]. The few responses to this question indicated that offices chose standard tables and projections based on their own observed experiences. For the most part offices use the "92" Series standard tables, although some offices use adjusted "80" Series standard tables and a small number of offices use even older tables with adjustments. The "92" Series projections or subsequent interim cohort projections are the most widely used, particularly for male mortality, though some offices have adjusted these projections to reflect their own experiences and expectations.

The second question concerned the appropriate level of aggregation in projecting future mortality. There was general agreement that, while cause-specific projections should be carried out if they could be made to work, the currently available methods and data fell far short of being adequate. However, contributors at the seminar expressed their continuing concerns that the improvements seen in the past may have been greatly influenced by changing smoking patterns and, as smoking patterns stabilised, future improvements could follow a different pattern. The Working Party agrees that changes in smoking incidence is indeed an important factor affecting mortality improvements, and one of those factors most easily identifiable from the available data. However, the Working Party does not believe that changes in smoking incidence explain all patterns; see Willets (2004). Mortality improvements for almost all causes arise due to the interaction of a number of factors. Many of these factors are common to several causes and few of the interactions are fully understood. Therefore, it is very difficult to model cause-specific mortality rates in a robust way.

The third question was whether the CMI should continue to project cohorts. All respondents supported the projection of cohort mortality improvements, giving as their reasons the evidence of the presence of cohorts in past mortality improvements.

The fourth question was whether quantitative measures of uncertainty associated with projections were needed, and, if so, what form they should take. Feedback was unanimous on the need for a measure of uncertainty, although views differed on how this should be provided. Scenarios similar to the long, medium and short interim cohort projections were seen as very useful in presenting the mortality risks to non-actuaries, particularly boards of life offices. Some respondents indicated that they would like stress-testing scenarios to

[1] The Working Party is aware that CMI standard tables are widely used by pension scheme actuaries in their work. Nevertheless, the standard tables are based on life insurance data and are prepared primarily for use by life insurance companies, and this is reflected in the emphasis given to this paper.

be provided. In an informal show of hands, attendees at the seminar voted overwhelmingly in favour of measures of uncertainty being provided with the next set of projections. No one voted against.

The fifth question was whether distributions or percentiles of future rates of mortality, derived from statistical models of past rates of mortality, were sufficiently meaningful to be used in practice. There was general agreement on the need for quantitative measures.

The sixth question was whether projections and any measures of uncertainty should be based on the largest available appropriate populations. The feedback agreed that the largest appropriate population should be used and that this choice was greatly affected by what was practical. The Working Party believes that the only choice is between population data and male assured lives data. No other current CMI experience is old enough or large enough to be credibly used for producing projections.

The seventh question was whether there was any clearly preferred methodology. The majority of the respondents had no preference for any particular methodology, although a desire for simplicity was expressed. Concerns were aired about the possible complexity of stochastic mortality models and the difficulties that would arise in explaining them to non-actuaries. The Working Party did not expect the mortality model to be too sophisticated or complicated, compared to stochastic asset models for instance, and so such difficulties should not be overplayed.

The eighth and final question was what the financial consequences of allowing for uncertainty in projecting future mortality might be. There was concern about how mortality research could be misunderstood, especially outside the profession. In particular attention was drawn to the possibility of an investment analyst concluding that offices were going to strengthen reserves, with a consequent impact on statements of earnings or profits. There were also concerns that regulators may draw inappropriate conclusions from the results of research. The need for careful communication and appropriate caveats was highlighted.

In its subsequent discussions, the Working Party rapidly concluded that it could not possibly produce any definitive answers to the problems raised by mortality projections, and that any methodology that it might suggest for use in practice would inevitably be subject to criticism and to change in the light of ongoing research. We believe it is absolutely essential that users of the projections are fully aware of this. In particular, we consider the following questions to be beyond the scope of our current research: (a) model uncertainty; (b) correlation between mortality and investment risk; and (c) moving projection methodology towards cause-specific projections.

The Working Party also noted that any financial uncertainty arising from uncertainty regarding the level of aggregate future mortality rates can be swamped by the heterogeneity of the amounts of pensions within an office's portfolio, given the very large difference between the mortality of pensioners with the smallest and largest pensions. For smaller portfolios, (indicated by our modelling in Section 3 to be fewer than 5,000 lives), the heterogeneity in pension size can be one of the biggest drivers of financial uncertainty, especially over shorter time-periods — an illustration of this is given in Richards & Jones (2004).

2. Outline of Proposed Methodology

2.1 *From Deterministic Trends to Scenario Generators*

Until 2002, the CMI's projections of future mortality were based on a single deterministic projection, derived by considering past trends. Latterly, they took the form of a two-dimensional 'sheet' $r(x,t)$, being the improvement factor to be applied to the baseline table to obtain the rate of mortality $q(x,t)$ at age x in the future calendar year t.

In 2002, Working Paper 1 introduced projections allowing for cohort effects. Three scenarios were published, called 'short', 'medium' and 'long' cohorts, differing in the length of calendar time over which the cohort effect was assumed to persist. These calendar times were chosen arbitrarily, and no probabilistic interpretation was possible. The CMI described these as 'interim' projections, signalling to the profession its intention to make a more scientific attack on the problem, if possible. Working Papers 3 and 11 set out at length the background to this work, with emphasis on the extent to which it may be possible to make probabilistic statements about future mortality.

At the same time, virtually all life insurance companies and some larger pension schemes have begun to use stochastic asset-liability models, driven by economic scenarios produced by scenario generators. This is very useful for our purposes, because it means that we can draw upon most actuaries' increasing knowledge of asset models, in order to highlight ways in which our own proposals are similar or dissimilar. To fix ideas and terminology, we refer to Figure 1. This illustrates the features of stochastic projections of the rate of change of the Retail Prices Index (RPI) made with the original Wilkie model (Wilkie, 1986). This is an AR(1) autoregressive model, and values are generated by recursions of the form:

$$\log \frac{Q_t}{Q_{t-1}} = \hat{\alpha}_1 \left(\log \frac{Q_{t-1}}{Q_{t-2}} - \hat{\alpha}_2 \right) + \hat{\alpha}_3 \; z_t \tag{1}$$

where Q_t is the RPI at time t, and the z_t are a sequence of independent unit Normal random variables. The parameter is the vector $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, and the projection uses the estimated parameter $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)$.

(a) Shown as dotted lines are 10 simulated sample paths. These clearly show high volatility, which reflects the high volatility of price inflation during the period used to fit the model. Since life office finances are sensitive to sudden changes in financial conditions, it is these sample paths that an insurance company would use as the scenarios in its asset-liability model.

(b) Shown as solid lines are the 5th, 25th, 50th, 75th and 95th percentiles of the distribution of the sample paths in each future year. They are rather smooth, the slight roughness just being a result of the number of sample paths generated (here 1,000 were used). The insurance company might well express the outcomes of its asset-liability model in terms of such percentiles, but it would not use them as scenarios to drive the model, because it is clear that the smooth percentiles bear no resemblance to the volatile sample paths.

(c) Shown as a solid line with diamond markers is a projection made by setting the variances of the driving noise term z_t to zero. This is a deterministic projection consistent with the fitted structure of the model.

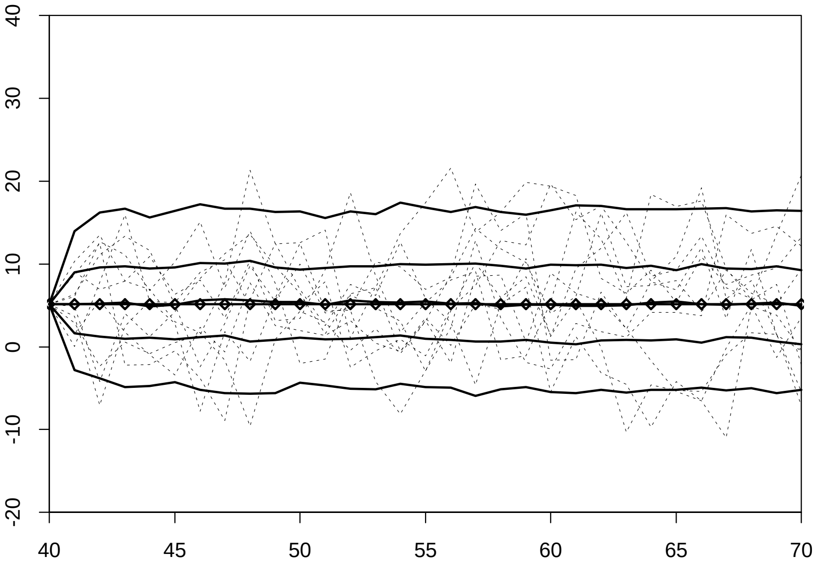

Figure 1: Projections of forces of retail price inflation based on Wilkie (1986). Ten simulated sample paths are shown (dotted) as well as the 5th, 25th, 50th, 75th and 95th percentiles. The line with diamond markers is the 'skeleton' of the projection, with the variance of the underlying random noise set to zero.

(d) Neither the sample paths nor the percentiles make any allowance for parameter uncertainty. This might appear to be surprising, since this methodology is used in most if not all asset-liability models. However, we can see that the only source of uncertainty in the generating equation (1) is the random noise series z_t. The estimated parameter $\hat{\alpha}$ is just used as if it was known to be correct. This might or might not be a significant oversight, depending on the application. If the main source of concern is the fact that the sample paths are volatile, rather than the location of the parameter, it is not likely to matter much. If, on the other hand, the sample paths were smooth, then its location might matter more, and we might miss important information by ignoring parameter uncertainty.

This is an example of a stochastic process model, in which we specify a structural part (the autoregressive form) which describes the non-random part of the time evolution of the RPI, and a source of randomness (the z_t) which perturbs it. Based on this example, the following terminology should be self-evident; we will try to use it consistently throughout this working paper.

(a) *Sample path*: A sample path is a single realization of the future course of a quantity represented by a stochastic process. It is the equivalent, in that context, of a single outcome of a probabilistic experiment.

(b) *Percentile*: At any future time t, the quantity being modelled has a probability distribution which defines its percentiles at that time. For illustrative purposes, we may join up the points of equal probability at different times, as in Figure 1. We will also call these percentiles. Different processes with very different patterns of sample paths may have identical percentiles.

(c) *Scenario*: We reserve this word to mean a projection that an actuary uses in an asset-liability model, however it is obtained. Thus the old CMI method used a single scenario; the interim projections offered three scenarios, and a stochastic process model might be used to generate many sample paths which can be used as scenarios. We emphasise that in this paper the word 'scenario' implies nothing about how it was obtained and we do not necessarily impute any probabilistic interpretation.

One attractive feature of stochastic process models is that we may be able to choose the space in which their sample paths lie so that any future that is physically plausible is a possible sample path, though not necessarily a probable one. For example, a future in which the RPI increases by 500% in one year and decreases by 400% in the next is a possible sample path. Thus, if we choose the space of sample paths sensibly, the model includes all possible futures, and if we estimate the parameters sensibly, only very small probability attaches to bizarre futures. If we assume or believe that the future will be consistent with the past, we can parameterize the model by fitting it to data; if we believe strongly that it will not be, we can adjust the parameters in line with our beliefs.

In Working Paper 3, we contrasted projections based on time-series models[2] and projections based on regression models. Time series models were there exemplified by the Lee-Carter mortality model, and regression models by the penalised spline (P-spline) mortality model. In the following sections, we describe briefly these models, and then summarise their advantages and disadvantages, in the context of an actuary wishing to generate scenarios for use in an asset-liability model.

2.2 *The Lee-Carter Model*

The Lee-Carter model is a bilinear model in the variables x (age) and t (calendar time) of the following form:

$$\log \mu(x,t) = a(x) + b(x)k(t) + z(x,t) \tag{2}$$

where $\mu(x,t)$ is the force of mortality at age x in year t and $z(x,t)$ is a random error term (or stochastic innovation). The $a(x)$ coefficients describe the average level of the $\log \mu(x,t)$ surface over time. The $b(x)$ coefficients describe the pattern of deviations from the age profile as the parameter $k(t)$ varies. If the $b(x)$ coefficient is particularly high for some ages x, then this means that mortality rates change faster at these ages than in general. If the $b(x)$ are all equal then mortality rates change at the same rate at all ages.

The $k(t)$ parameter describes the change in overall mortality. If $k(t)$ falls, then mortality rates fall, and if $k(t)$ rises, then mortality rates rise. The coefficients $b(x)$ determine how this overall change in mortality affects rates at the age in question. If $k(t)$ decreases linearly, then $\mu(x,t)$ decreases exponentially at each age, at a rate that depends on $b(x)$ (unless $b(x)$ is negative, in which case $\mu(x,t)$ increases).

The model does not specify unique choices for the parameters, because $b(x)$ and $k(t)$ appear only through their product $b(x)k(t)$. It is often assumed that $\sum_x b(x) = 1$ and $\sum_t k(t) = 0$ to enforce uniqueness.

[2]Time series models are examples of stochastic process models.

Lee & Carter (1992) suggested using a time series model for $k(t)$, so projections based on a Lee-Carter model share many of the statistical features of time series forecasts. Having fitted the model, $k(t)$ is projected forward to give mortality rates for future years. In fact an AR(1) autoregressive model for $k(t)$ has been a popular choice, so projections often have the essential features shown in Figure 1.

The Lee-Carter model can be fitted by standard likelihood methods, assuming a Poisson model for the numbers of deaths at each age and in each calendar year in the data. The parameter can be of very high dimension, depending on whether or not $a(x)$ and $b(x)$ are estimated at each integer age x or are represented by parametric models. Assuming that the model chosen for $k(t)$ has parameter vector β, the parameter of the whole model may be as large as:

$$\alpha = (a(0), a(1), a(2), \ldots, b(0), b(1), b(2), \ldots, \beta). \quad (3)$$

It has often been observed that projections made with the Lee-Carter model, with an AR(1) model for $k(t)$, have rather narrow confidence intervals. We can now see why. Writing in full the parameters of the AR(1) model for the time trend $k(t)$ as $\beta = (\beta_1, \beta_2, \beta_3)$, along the lines of Equation (1), we see that:
(a) only $\hat{\beta}_3$, of all the many parameters, contributes at all to the uncertainty;
(b) there is no allowance for any parameter uncertainty (not even uncertainty about $\hat{\beta}_3$); and
(c) the choice of an AR(1) model for $k(t)$ imposes a stationary distribution on the projected values (this is why the percentiles in Figure 1 are parallel instead of fanning out) and hence is very influential.

We believe that the Lee-Carter model has much to recommend it; it is simple but highly-structured, the structure has some plausibility, it is familiar and much-studied, and it can be used to generate scenarios using exactly the same techniques as are already used in asset-liability models. However, we do not think it should be used in such a way that parameter risk is ignored, unless it can be shown that the impact of this is unimportant. Also, the very fact that it is highly structured introduces a degree of model uncertainty.

The following method of incorporating parameter uncertainty is essentially the same as the method used by Forfar, McCutcheon & Wilkie (1988) to generate probability distributions of graduated mortality tables; it is a widely used technique known as parametric bootstrapping. The idea is to use the fitted model (parameter $\hat{\alpha}$), and the known exposures at each age and calender year in the data, to simulate fresh data. These can then be used to re-fit the model, generating a parameter from the distribution of $\hat{\alpha}$. By doing this repeatedly, we simulate the distribution of $\hat{\alpha}$. More detail is given in the following steps, which are unavoidably technical:
(a) Suppose the data are numbers of exposures[3] $E_{x,t}$ and numbers of deaths $D_{x,t}$ at each age x and in each past calendar year t.
(b) We fit the Lee-Carter model and obtain parameter $\hat{\alpha}$, which gives us an estimate $\hat{\mu}_{x,t}$ of the force of mortality at each age and in each calendar year.

[3] We use the notation $E_{x,t}$ rather than $E^c_{x,t}$, to avoid proliferation of indices, but note that the exposures here are of central type.

(c) At each data point (x, t) we calculate the deviance residual[4] (McCullagh & Nelder, 1989):

$$r_{x,t} = \operatorname{sign}(D_{x,t} - \hat{D}_{x,t})\sqrt{2\left(D_{x,t}\log\frac{D_{x,t}}{\hat{D}_{x,t}} - (D_{x,t} - \hat{D}_{x,t})\right)} \tag{4}$$

where $\hat{D}_{x,t} = \hat{\mu}_{x,t}E_{x,t}$ is the 'expected' number of deaths.

(d) Suppose we have data for T calendar years, labelled $1, 2, \ldots, T$, and for A ages. We form the $A \times T$ table of deviance residuals, R, using (4). The ith bootstrap sample $R^{(i)}$ from R is obtained by first taking a sample of size T with replacement from the set $\{1, 2, \ldots, T\}$; suppose this sample is $\{s_1^{(i)}, s_2^{(i)}, \ldots, s_T^{(i)}\}$. Then $R^{(i)} = (r_{x,s_t^{(i)}})$. This is best thought of as taking the same bootstrap pattern across time for each age. In this way we hope to preserve any aberrant years (severe winters, flu, etc) by moving them randomly in time. Solving Equations (4), we simulate the numbers of deaths $D_{x,s}^{(i)}$.

(e) Using 'data' $D_{x,t}^{(i)}$ and $E_{x,t}$, we refit the Lee-Carter model, getting a new parameter $\hat{\alpha}^{(i)}$.

(f) After running N bootstrap simulations, we have generated parameters $\hat{\alpha}^{(1)}, \hat{\alpha}^{(2)}, \ldots, \hat{\alpha}^{(N)}$. Then the empirical distribution of these parameters can be taken as an estimate of the sampling distribution of $\hat{\alpha}$.

Generating scenarios of future mortality allowing for parameter uncertainty is then simple. If we want N scenarios, we generate the ith using the parameter $\hat{\alpha}^{(i)}$, generating a sample path from the time series model for $k(t)$ with the parameter $\hat{\beta}^{(i)}$.

Finally, given a scenario, which consists of a set of forces of mortality $\mu_{x,t}^{(i)}$ in each future year, we assume that the numbers of in-force policies (or pensioners) $E_{x,t}^{(i)}$ in each future year are available (possibly different in each scenario). The numbers of deaths in each future year can then be simulated from Poisson distributions with parameters $E_{x,t}^{(i)}\ \mu_{x,t}^{(i)}$. For annuity portfolios in which there are significant concentrations of risk, such as a small number of annuitants with very large annuities, the exposures may be subdivided by size of annuity, and the numbers of deaths simulated separately for each subgroup.

Figure 2 shows an example of this procedure. We use ages 40 (left) and 75 (right) as examples, but bear in mind that these are not separate one-dimensional projections, but cross-sections through the two-dimensional sheet. The experience is that of Male Assured Lives from 1947 to 2002.

(a) At the top we have the actual data, represented by the crude forces of mortality (the circles before 2003) and the fitted Lee-Carter model in the region of the data. After 2002, we show as a grey line a single simulation based on the fitted parameters and a time-series-like sample path of $k(t)$. The circles after 2002 represent crude rates of mortality obtained by simulating the numbers of deaths from $E_{x,2002}$ lives exposed in each future year. (Note that the surprisingly wide dispersion after 2002 at age 40 is

[4] The deviance residual is more appropriate here than the more familiar Pearson residual or standardised deviation, long used in tests of graduations, because the underlying model is Poisson.

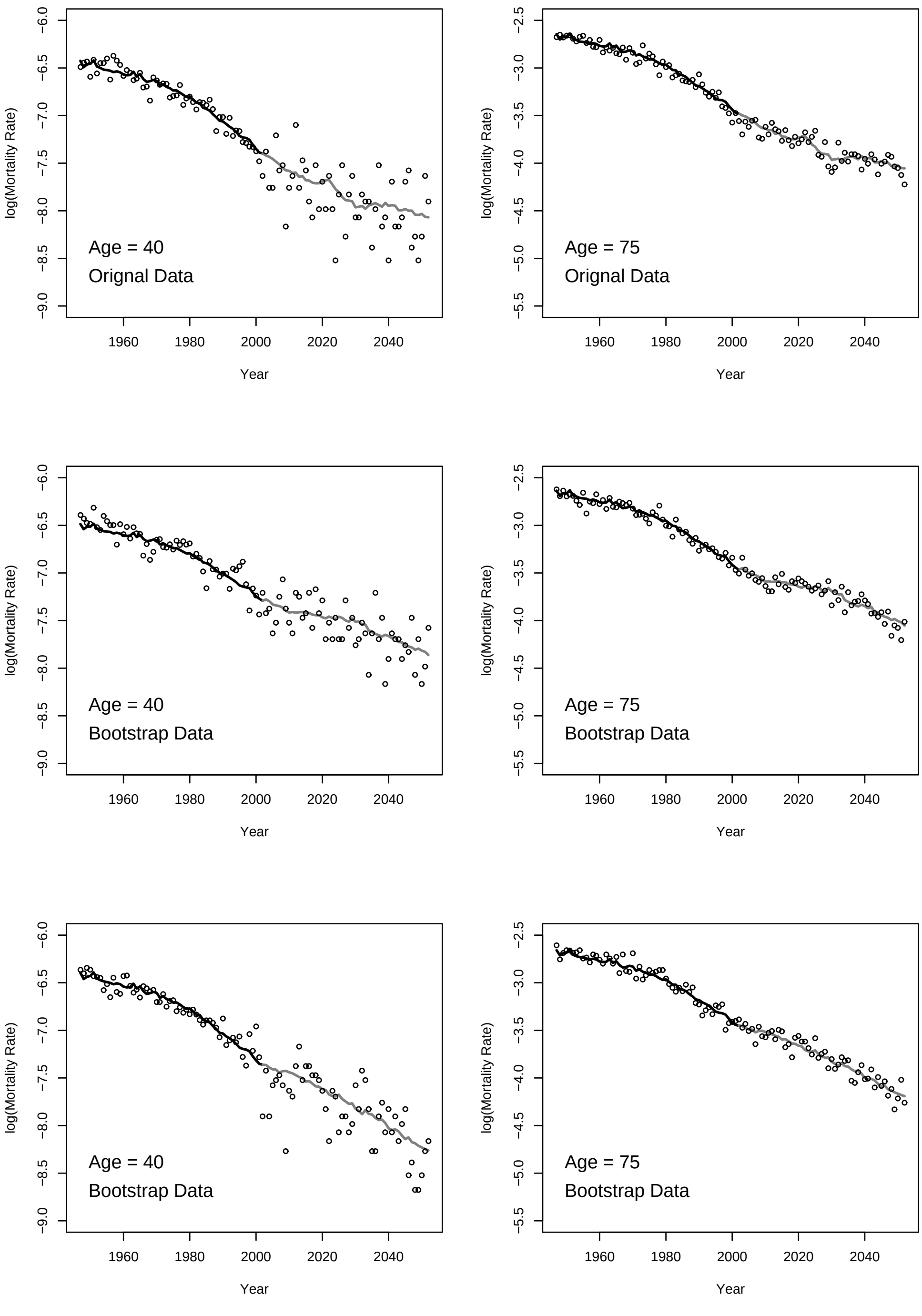

Figure 2: An example of the Lee-Carter Model. At the top are cross-sections at ages 40 and 75 of the fitted model and a single scenario based on the fitted time series model for $k(t)$. The second and third lines show cross-sections at ages 40 and 75 of two independent simulations. Circles before 2003 indicate the actual data (top) or bootstrapped data (simulations), and circles after 2002 represent added Poisson noise using the exposures $E_{x,2002}$.

because the exposure $E_{40,2002}$ had dropped sharply (by a factor of six) from its peak in 1983.)

(b) In the middle, we have the data from a single bootstrap simulation $D^{(1)}_{x,t}$ say. (The circles before 2003 actually represent the simulated crude rates of mortality $\mu^{(1)}_{x,t} = D^{(1)}_{x,t}/E_{x,t}$.) The Lee-Carter model is refitted to these data, giving a parameter $\hat{\alpha}^{(1)}$ say, including new parameters for $k(t)$. Shown as a grey-green line[5] is a single sample path from this model, the circles after 2002 again representing crude rates of mortality obtained by simulating the numbers of deaths from $E_{x,2002}$ lives exposed in each future year.

(c) At the bottom, we show the outcome of a second bootstrap simulation, yielding 'data' $D^{(2)}_{x,t}$ and parameter $\hat{\alpha}^{(2)}$ say.

2.3 *The P-Spline Model*

We refer to Currie, Durban & Eilers (2004) for more details specific to mortality data, or Eilers & Marx (1996) for the original exposition. We also refer back to the discussion in Working Paper 3.

Briefly, a regression model describes the functional relationship of some data Y on some other data X. For example if the data Y consists of numbers of deaths $D_{x,t}$ at ages x in calendar years t, and the data X consists of the corresponding exposures $E_{x,t}$, basic statistical considerations suggest a Poisson model, such that $D_{x,t}$ has a Poisson distribution with parameter $E_{x,t}\ \mu_{x,t}$. The question then becomes: how can we describe the relationship between $\mu_{x,t}$ and the variables x and t? One answer (the regression approach) is to choose a number of basis functions $b_1(x,t), b_2(x,t), \ldots b_n(x,t)$, such that we can represent $\mu_{x,t}$ as a linear combination:

$$\mu_{x,t} \approx a_1\ b_1(x,t) + a_2\ b_2(x,t) + \cdots + a_n\ b_n(x,t). \qquad (5)$$

The question then becomes: how do we choose a suitable set of basis functions $b_i(x,t)$, and what criterion do we apply to choose the 'best fitting' regression coefficients a_i? There are as many different regression models as there are answers to this question.

In Working Paper 3 we gave a simple example, in one dimension (age x only) of regression using polynomials as basis functions. We used the simplest polynomials: $b_1(x) = 1, b_2(x) = x, b_3(x) = x^2, \ldots, b_n(x) = x^{n-1}$. The main point we were driving at was that polynomials can be used successfully in many cases, but are a very poor choice if we wish to extrapolate outside the region of the data. Projection of mortality rates is, of course, an example of such extrapolation.

Splines present an alternative choice of basis functions. They have appeared before in actuarial practice in the U.K., having been used to graduate the last few English Life Tables; see Benjamin & Pollard (1980) for an introduction. A spline of degree m is simply a curve made up of segments of polynomials of degree m, such that where the segments join, their derivatives of up to order $m-1$ are equal. Figure 3 shows an example of a cubic spline, and Figure 4 shows an example of a set of cubic spline basis functions. The term 'B-splines' is often used to denote a set of basis splines.

[5] A green line on viewers and printers that can show colours, a grey line otherwise.

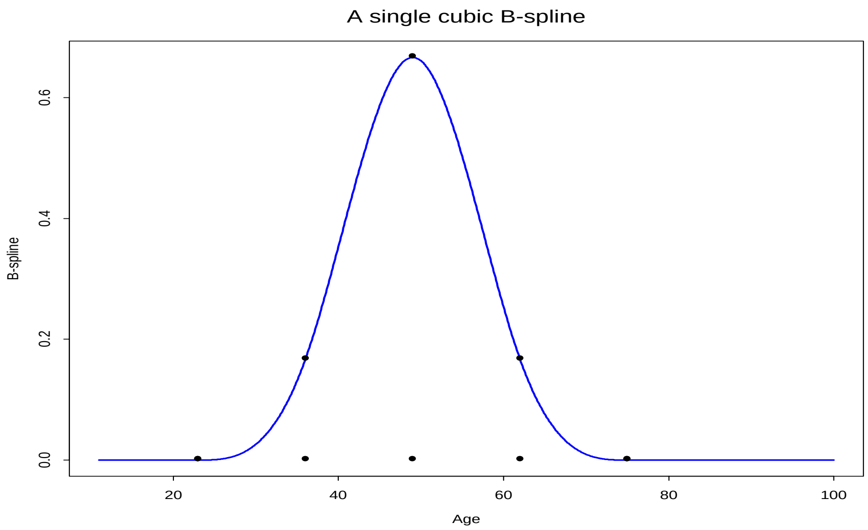

Figure 3: An example of a single cubic spline.

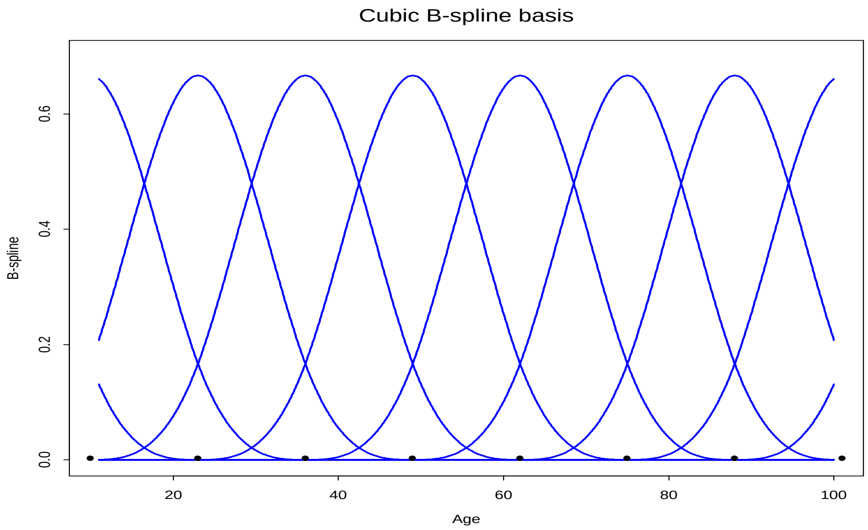

Figure 4: An example of a spline basis made up of cubic splines.

Return for a moment to polynomial regression, and consider the problem of choosing the regression coefficients a_i. We often adopt a principle of parsimony, meaning we limit n, the number of basis functions, to be as small as possible. Thus, if a quadratic regression is good enough, we do not carry out a cubic or higher-order regression. This can be viewed as a trade-off between smoothness and goodness-of-fit; by adding more higher-order polynomials to the basis, we achieve a closer fit but at the expense of less smoothness between the data points[6].

A B-spline graduation requires us to reach a similar balance between goodness-of-fit — achievable by adding more and more splines to the basis — and smoothness — achieved by limiting the number of splines in the basis. It is therefore no different to the polynomial graduation, it just presents a different (arguably better) choice of basis functions.

A genuinely different approach is the penalised spline, or P-spline, regression model. Here, we make no attempt to keep the number of basis splines small; instead we make sure that the basis is rich enough (or dense enough) to provide a good fit to almost arbitrary data. We then impose an explicit penalty on lack of smoothness, represented very conveniently by lack of smoothness in the progression of the coefficients a_i. Then the precise number of basis functions almost ceases to matter; the trade-off between smoothness and goodness-of-fit is achieved by choosing a large penalty (prefer smoothness) or a small penalty (prefer goodness-of-fit). Although its use with splines is new, the idea of penalised smoothing is not; it goes back at least as far as Whittaker-Henderson graduation (see Whittaker (1923) for example).

Splines are just as easy to define in many dimensions as in one dimension (think of hills on a plane rather than peaks on a graph). Also, the machinery of penalised spline regression carries over into more than one dimension. So, we can easily fit a surface of forces of mortality $\mu_{x,t}$ over the (x,t)-plane. Even better, there is no need to choose age x and calendar year t as the two directions; there is nothing canonical about them. We can just as easily choose age x and year of birth c (for cohort) as the dimensions, and if we do, projections will tend to carry any cohort effects into the future. We will assume in the following that we work in the (x,c)-plane.

In the region of the projection, the 'fit' is determined jointly by the data and the choice of the penalty. Moreover, because it is the penalty that regulates 'good' behaviour, the surface is guaranteed to be well-behaved in the region of the projection. This overcomes the major weakness of other (for example polynomial) regression models if the aim is extrapolation.

Thus, we can summarise the P-spline approach to mortality projection as follows. We choose a rich enough set of basis splines in two dimensions, and fit to data using a penalised likelihood, choosing the level of the penalty to enforce reasonable smoothness[7]. The fitting is actually carried out over the whole region of the (x,c)-plane covering the region of the data and the region of the projection: in the latter region, a well-behaved projection is guaranteed because of the operation of the penalty. The fitted (including projected)

[6] We refer to smoothness here in the traditional sense of the actuarial graduation; a high-order polynomial is mathematically smooth in terms of its differentiability of all orders, but that is not what we mean.

[7] Any of the common criteria for optimising smoothness *versus* goodness-of-fit can be used, such as the Bayesian Information Criterion or the Akaike Information Criterion.

surface of values $\hat{\mu}_{x,c}$ can be regarded as the 'mean sheet' of the regression. From the variance matrix of the estimated parameters, we obtain also a 'standard deviation sheet' of values $\hat{s}_{x,c}$. Precisely because it is based on the variance matrix of the parameter estimates, this 'standard deviation sheet' incorporates all the information about parameter uncertainty.

A simple method of incorporating projected longevity (and its uncertainty) into asset-liability models then suggests itself.

(a) Make the (strong) assumption that the variability of the future surface of mortality rates is adequately described by the standard deviation sheet.

(b) Make random drawings $z_1, z_2, \ldots$ from a standard Normal distribution. Then the sheet given by $\mu_{x,c}^{(i)} = \hat{\mu}_{x,c} + z_i\ \hat{s}_{x,c}$ represents the $\Phi^{-1}(z_i)$th percentile of the distribution of the sheet of future mortality rates, where $\Phi(z)$ is the cumulative Normal distribution function.

(c) As with the Lee-Carter model, given this scenario (sheet) of future mortality rates, the numbers of deaths may be simulated as Poisson random variables with parameters $E_{x,c}^{(i)}\ \mu_{x,c}^{(i)}$, if required. Also as before, the exposures may be subdivided by size of annuity and the numbers of deaths simulated separately for each subgroup.

Figure 5 shows an example of this procedure, using the same data as in Figure 2. Cohort penalties were used as described above.

(a) At the top we show the fitted mean projection $\hat{\mu}_{x,c}$, with the 2.5th and 97.5th percentiles given by $\hat{\mu}_{x,c} + 1.96\ \hat{s}_{x,c}$ and $\hat{\mu}_{x,c} - 1.96\ \hat{s}_{x,c}$ (a 95% confidence interval). Although the scale makes it hard to see, the percentiles are very close to the fitted model before 2003.

(b) In the middle, we show the first simulation supposing (for the sake of example) that $z_1 = 1$. The grey dashed line shows the percentile sheet $\mu_{x,c}^{(1)} = \hat{\mu}_{x,c} + z_1\ \hat{s}_{x,c}$, while the mean projection $\hat{\mu}_{x,c}$ is still shown for comparison. The circles after 2002 represent crude rates of mortality obtained by simulating the numbers of deaths from $E_{x,2002}$ lives exposed in each future year[8].

(c) At the bottom, we show a second simulation, assuming for the sake of example that $z_2 = -2$.

2.4 *Discussion: Comparison of the Two Approaches*

We have set out above two possible approaches to projecting longevity risk, including quantitative estimates of the associated uncertainty. We do not (cannot) single out either as preferable, but here we discuss some of their main features.

(a) *Ability to generate sample paths.* If a time series model is chosen for $k(t)$, then the Lee-Carter model is a stochastic process model capable of generating sample paths by simulation. The P-spline model is *not* a stochastic process model, and it cannot generate sample paths.

(b) *Allowance for parameter uncertainty.* The P-spline model allows for parameter uncertainty almost automatically, through the variance matrix of the regression coefficients.

[8] The circles before 2003 do not represent the data, but the simulated number of deaths based on the $\mu_{x,c}^{(1)}$ and the actual exposures $E_{x,c}$.

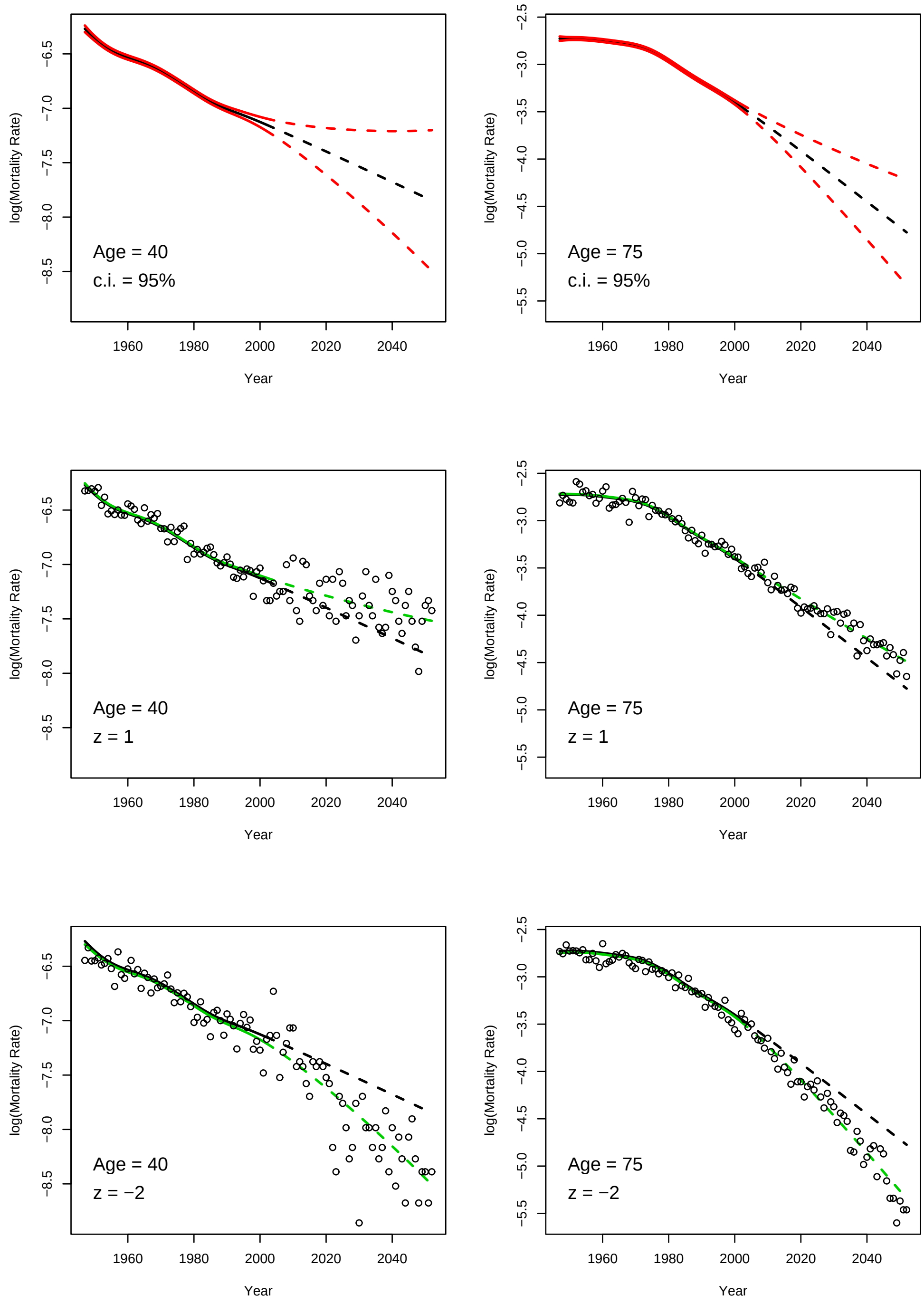

Figure 5: An example of the P-spline model. At the top are cross-sections at ages 40 and 75 of the fitted model, showing the mean projection and the 2.5th and 97.5th percentiles. The second and third graphs show cross-sections at ages 40 and 75 of two independent simulations, one with $z_i = 1$ and one with $z_i = -2$. Circles before 2003 represent added Poisson noise using the actual exposures $E_{x,t}$ before 2003 and the exposures $E_{x,2002}$ after 2002.

This governs the distribution (at least up to Normal approximation) of the surface $\mu_{x,c}$ both in the region of the data and the region of the projection. The Lee-Carter model can be made to allow for parameter uncertainty by bootstrapping[9]. There is no explicit allowance for model uncertainty, other than in being able to compare the two models.

(c) *Incorporation of cohort effects.* The P-spline model incorporates cohort effects very simply, through the choice of penalty. Specifically, there is one penalty in each dimension, which determines how strongly the main features in that dimension carry over into the projected 2-dimensional sheet. If the penalties are defined along the age and calendar year dimensions (the (x, t)-plane) cohort effects are quickly smoothed out, but if the penalties are defined along the age and year-of-birth dimensions (the (x, c)-plane) cohort effects are explicitly recognised. The Lee-Carter model, on the other hand, does not explicitly allow for cohort effects.

Apart from cohort effects, the chief difference between the two models may appear to be that one provides sample paths and the other only provides percentiles. Thus, in increasing order of sophistication and suitability for use with asset-liability models, we appear to have:

1. The single, deterministic projections used before 2002, with no cohort effects.
2. The three cohort projections published in Working Paper 1 in 2002, without any probabilistic interpretation.
3. The P-spline model that generates percentiles.
4. The Lee-Carter model that generates time-series-like sample paths.

However, we think it may be seriously misleading to regard the Lee-Carter model as a 'better' scenario generator than the P-spline model — in the sense of representing the uncertainty about future annuitants' mortality — just because it generates time-series-like sample paths. This would certainly be a strong criterion for an asset model, because the volatility within individual scenarios is critical, but it may not be such a strong criterion for mortality projections. The reason is that by far the most important source of uncertainty in a mortality projection is the trend. This is substantially embodied by the parameter uncertainty. The time-series-like sample paths of the Lee-Carter model represent only stochastic uncertainty around that trend. As we have remarked already, projections made with the Lee-Carter model, representing $k(t)$ as an AR(1) process, often have surprisingly narrow confidence intervals, precisely because all parameter uncertainty — which is, we repeat, along with model uncertainly, a major source of uncertainty — is ignored. Thus the apparent advantage of the Lee-Carter model over the P-spline model in respect of generating sample paths may be quite illusory; there is no reason

[9] We emphasise that the very narrow confidence intervals often see in graphs of Lee-Carter projections arise because of: (a) lack of allowance for parameter uncertainty; and (b) usually, the choice of an AR(1) model for $k(t)$.

to believe that these sample paths actually represent equiprobable drawings from any realistic distribution on the space of biologically plausible futures.

Therefore, we intend to continue to develop both models into proposals for methodologies, while acknowledging the limitations of each.

(a) The P-spline model allows scenarios to be generated very simply, and retains cohort effects if the penalties are chosen appropriately. But these scenarios are percentiles, lacking the properties of true sample paths.
(b) The Lee-Carter model will provide sample paths, if a time series representation is chosen for $k(t)$, but this does not really answer the question: sample paths of what? Parameter uncertainty, which we regard as very important for mortality projections, may be included *via* bootstrapping, which is computationally laborious. The model will not obviously project existing cohort patterns.

In Working Paper 3 we emphasised that we did not expect to be in a position, in 2005, to set out a methodology that was the last word on the subject. Indeed, we listed several topics that we expected to leave as the subjects of future research. We reiterate here that we do not expect projection methodology to stay unchanged for long periods of time, as it used to do.

3. Examples

This section provides examples of the risk-capital requirements for model portfolios of varying sizes using the P-Spline model and the simulation methodology set out in Section 2.3. The confidence intervals for the costs of meeting the benefits for portfolios of annuitants or pensioners ('portfolio costs') are shown for various sizes of portfolios by numbers of lives. The projected mortality improvements were derived by fitting a P-spline model with penalties on age and period (the (x, t)-plane) to male population data[10]. As the penalties largely determine the projections under P-spline models, applying penalties to age and cohort or choosing a different form of penalty could produce very different projections.

We modelled seven portfolios containing only single male annuitants receiving level benefits. The portfolios contained 260, 500, 1000, 2,500, 5,000, 10,000 and 25,000 lives[11]. We used a deterministic interest rate of 4.5% per annum. The projected annuity rates and portfolio costs were calculated by modelling the projected mortality rates and observed deaths as set out in Section 2.3.

Stage A

We projected male population mortality for ages 40–89, from 1990, based on data from 1961–1990. Using the mean sheet of projected mortality intensities $\hat{\mu}_{x,t}$ and applying a sampled standard normal variable to the projected standard deviation sheet $\hat{s}_{x,t}$, scenarios

[10]Note that this differs from the example in Section 2.3. There, we used Assured Lives data and applied penalties to age and cohort.

[11]The smallest portfolio contained 260 lives rather than 250 in order to ensure an exposure of at least one life at each age within each sub-group when, later on, the portfolio is modelled as four homogenous sub-groups.

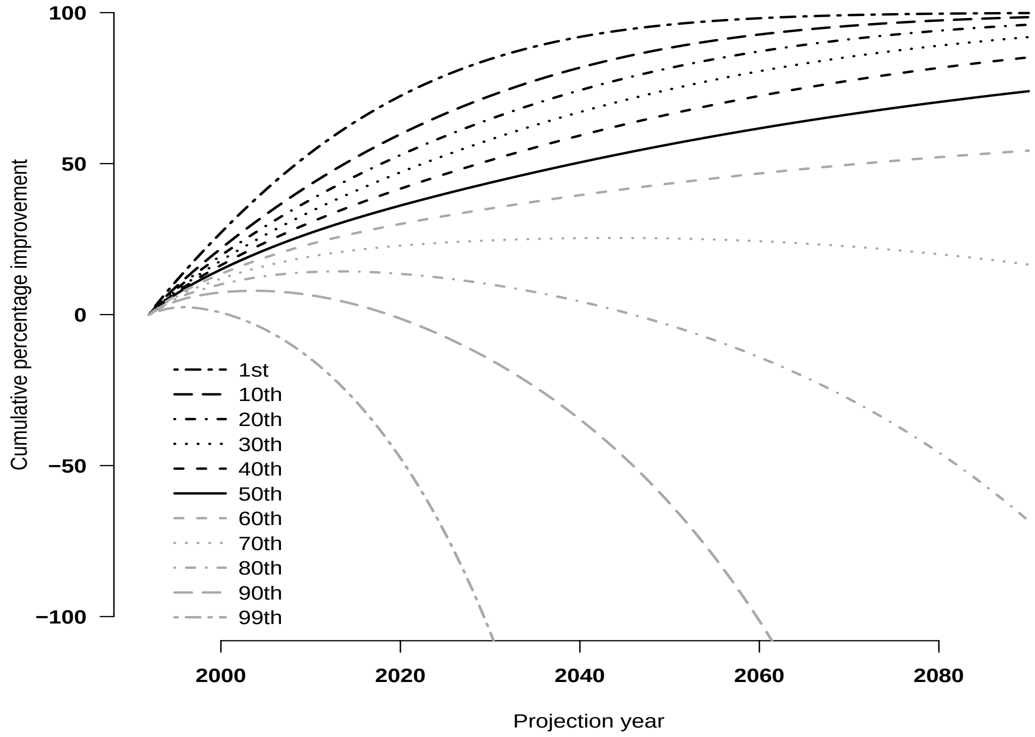

Figure 6: Percentiles of cumulative mortality improvements from calendar year 1992, at age 60.

of the future force of mortality were generated. Apart from the use of calendar year penalties rather than cohort penalties, this is just the procedure illustrated in Figure 5.

From each scenario, we derived projected improvement factors in the mortality intensities. These factors (based on population data) were then applied to the PML92C1992 table to give the projected mortality intensities suitable for use with annuitants. Improvement factors at ages above 89 were taken to be equal to those at age 89. The cumulative mortality improvements from 1992 for scenarios corresponding to the 1st to 99th percentiles at age 60 are shown in Figure 6. Projected mortality improvements cannot exceed 100% (the level at which the mortality intensity falls to zero) but can fall below −100% as the projected mortality intensities can more than double in adverse scenarios.

The number of deaths among a homogenous group of lives of a given age and in a future calendar year can then be modelled as a Poisson random variable based on the projected mortality intensity and the central exposure in the portfolio for that age and year. However, as only the systemic risks were being modelled at Stage A (that is, no stochastic uncertainty about the observed number of deaths given a particular set of mortality rates), we calculated annuity rates assuming that the numbers of deaths in each future year were equal to the means of these Poisson numbers of deaths, and portfolio costs assuming that all lives in the portfolio had the same benefit level.

Table 1: Weights by lives and amounts, and relative mortality, in an example of a heterogeneous annuity portfolio.

	Group 1	Group 2	Group 3	Group 4	Portfolio
Relative weight (lives)	10%	15%	35%	40%	100%
Relative mortality	68%	85%	110%	105%	100%
Average amount of annuity (£)	13,000	11,000	6,500	4,500	6,825
Effective weight (amounts)	19%	24%	33%	24%	100%

Stage B

In order to show the impact of heterogeneity of benefit levels within a portfolio, we allocated the lives in each portfolio to four sub-groups as shown in Table 1 below. The table shows, for each sub-group, its weight in the portfolio (by number of lives at outset), its relative mortality experience, its average amount of annuity benefit and its effective weight (at outset) by amounts. The weighted average rates of mortality and benefit levels across sub-groups equal the averages assumed in the earlier calculations in Stage A.

For each scenario from Stage A, the numbers of deaths in each sub-group in the portfolio were simulated as Poisson random variables (an alternative approach would be to model them as Binomial random variables by estimating mortality rates q_x from the mortality intensities). The annuity rates and portfolio costs were found for each simulation under the following three sets of assumptions:

(i) All sub-groups were assumed to suffer the same average underlying mortality intensities and to have the same average benefit level.

(ii) The four sub-groups faced different underlying mortality intensities at outset as set out in Table 1, but had the same average benefit level. The improvement in the mortality intensity from year to year was assumed to be the same for all sub-groups.

(iii) The four sub-groups had different average benefit levels as well as facing different underlying mortality intensities at outset as set out in Table 1.

Figure 7 below shows the 95% confidence intervals of the projected costs for the seven model portfolios, expressed as a percentage of average costs, under each of the above four sets of assumptions.

As well as illustrating how the risk capital requirements increase as the portfolio size decreases, Figure 7 also illustrates the impact of heterogeneity within the portfolio on the risk capital requirements:

(a) The risk capital calculated at Stage A reflects only uncertainty about future mortality improvements. As this is a systemic risk, the risk capital does not vary by portfolio size.

(b) The risk capital at stage B(i) reflects the capital needed to cover the risk arising from the stochastic variability related to the portfolio size as well as the systemic risk from Stage A.

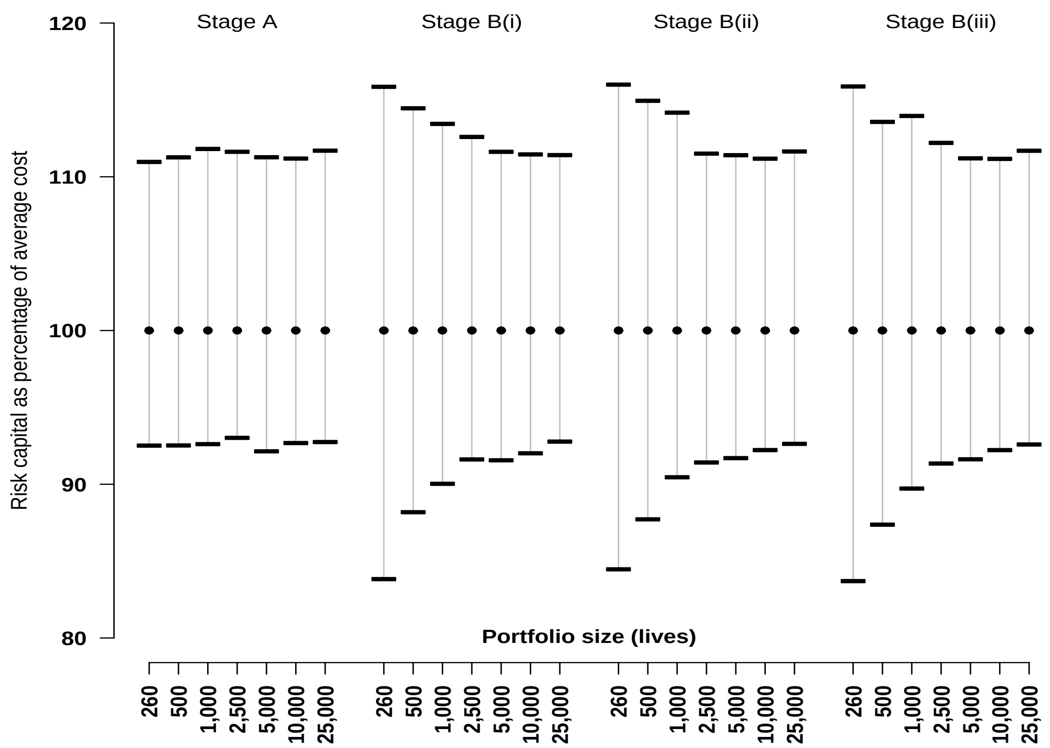

Figure 7: 95% confidence interval for portfolio costs

(c) The risk capital at stage B(ii) reflects the capital needed to cover the risk arising from heterogeneous mortality within the portfolio, as well as the risks covered at Stage B(i).

(d) The risk capital at stage B(iii) reflects the capital needed to cover the risk arising from heterogeneous mortality *and* varying benefit levels within the portfolio, as well as the risks being covered at Stage B(i).

Figure 7 shows that even large, homogenous portfolios need to hold significant capital to cover the systemic uncertainty about future mortality improvements (stage A).

The size of the risk capital to cover systemic risk is determined by the population data used to fit the model, rather than the details of the portfolio itself. If assured lives data and/or data covering different periods and ages are fitted using the P-Spline model, both the mean intensities $\hat{\mu}_{x,t}$ and the associated standard deviation sheet $\hat{s}_{x,t}$ could look very different resulting in different estimates for risk capital. While our modelling indicated risk capital to cover systemic risk of about 11% of the mean portfolio cost, this is based on population mortality up to 1990, and most life offices have significantly increased the strength of their reserving for annuity business since 1990. Further, the example does not allow for any investment risk. Therefore, the risk capital for systemic risk of about 11% indicated in the above example should not be taken as necessarily applying to annuitant portfolios today.

Additional risk capital is required to cover the (diversifiable) stochastic uncertainty about the number of deaths in any year. Figure 7 shows that this risk for the portfolios modelled becomes immaterial as the portfolio size exceeds 5,000 lives. For the smallest portfolios, the additional risk capital required to cover this risk can exceed 5% of the expected cost.

For the portfolios modelled as set out above, Figure 7 indicates that heterogeneous mortality, either on its own or in combination with varying benefit levels, does not lead to the need for materially higher risk capital. This may be because the systemic uncertainty about future mortality improvements at Stage A overwhelms any uncertainty arising from the diversifiable risks in Stage B for all but the smallest portfolios. Therefore, the risk capital requirements for diversifiable risk shown in Figure 7 are not conclusive:

(a) While allowing for differing mortality improvements by sub-group could be expected to lead to lower risk capital requirements at Stage A (as this risk is no longer systemic), the impact on the risk capital for the diversifiable risks seen at Stage B is not clear.

(b) If a different dataset were fitted using the P-Spline model, this may lead to a lower risk of uncertainty about future mortality improvements at Stage A and hence affect the balance in the risk capital requirements for the systemic and diversifiable risks.

Finally, note that the diverging percentiles in the region of the projection (see Figure 6) mean that uncertainty increases as time passes beyond the last calendar year in the region of the data. This means that it would not be sensible (for example) to calculate risk capital in any given calendar year using mortality projections produced some years before. In the example above, the model was fitted to data from 1961–1990, and used to calculate risk capital in 1992. Some reductions in risk capital might have been achievable by fitting the model to data for 1961–1991 or 1961–1992, if possible. Therefore, timely reporting of mortality data, and regular rebasing of projected mortality, may become necessary in future.

4. Proposed Outputs

4.1 *Introduction*

Currently, projections of future mortality are provided by the CMI for the following classes of business:

(a) Male and female immediate annuitants, (lives and amounts).

(b) Holders of retirement annuities.

(c) Male and female pensioners in insured group pension schemes (lives and amounts).

(d) Widows of members of insured group pension schemes (lives and amounts).

The same set of mortality improvement factors was used for each of the above experiences, whether males or females, lives or amounts, mainly in order to adopt a relatively simple model which avoided 'obvious' anomalies such as lower rates of mortality for males than females or for lives rather than amounts (see CMIR 17).

Table 2: Proposed annuitant tables.

Table	Investigation	Sex	Lives/Amounts	Select Period
IML00	Immediate Annuitants	Males	Lives	0
IMA00	Immediate Annuitants	Males	Amounts	0
IFL00	Immediate Annuitants	Females	Lives	1
IFA00	Immediate Annuitants	Females	Amounts	1
RMD00	Retirement Annuitants, Deferred	Males	Lives	0
RMV00	Retirement Annuitants, Vested	Males	Lives	0
RMC00	Retirement Annuitants, Combined	Males	Lives	0
RFD00	Retirement Annuitants, Deferred	Females	Lives	0
RFV00	Retirement Annuitants, Vested	Females	Lives	0
RFC00	Retirement Annuitants, Combined	Females	Lives	0
PPMD00	Personal Pensioners, Deferred	Males	Lives	0
PPMV00	Personal Pensioners, Vested	Males	Lives	0
PPMC00	Personal Pensioners, Combined	Males	Lives	0
PPFD00	Personal Pensioners, Deferred	Females	Lives	0
PPFV00	Personal Pensioners, Vested	Females	Lives	0
PPFC00	Personal Pensioners, Combined	Females	Lives	0

4.2 *Proposed Base Tables*

Proposed new tables of graduated mortality rates are being prepared by the Graduation Working Party of the Mortality Committee, for the classes of business shown in Tables 2 and 3. They will be published as one or more Working Papers later in 2005, once the work on projections is completed (Proposed tables of graduated mortality rates for permanent and temporary assurances have already been published in Working Paper 12 (CMI, 2005b).) We propose to use these graduations as the base tables for the projections. Future mortality rates will be obtained by cumulatively applying the appropriate projected rates of improvement for that age by calendar year to the mortality rates of the appropriate base table.

These new tables will be based on graduations of the experience for the relevant classes of business for 1999–2002. It is expected that the new set of tables will be denoted the "00" series, following the convention adopted for the "92" and "80" series. However this is not yet set in stone. The proposed new tables have been labelled in a way that follows the previous naming convention, although this may be subject to change when the final tables are published. It may also be that not all of the above tables are finally proposed as standard tables.

4.3 *Methods of Graduation*

As for the previous two sets of graduated tables, the methodology being used by the Graduation Working Party is to fit functions of the Gompertz-Makeham family to μ_x, by maximum likelihood (see Forfar, McCutcheon & Wilkie (1988)). Adjustments may be needed at the oldest and youngest ages to produce sensible results, and to achieve

Table 3: Proposed pensioner tables.

Table	Investigation	Sex	Lives/Amounts	Select Period
PML00	Pensioners, Normal/late retirements	Males	Lives	0
PEML00	Pensioners, Early retirements	Males	Lives	0
PCML00	Pensioners, Combined	Males	Lives	0
PMA00	Pensioners, Normal/late retirements	Males	Amounts	0
PEMA00	Pensioners, Early retirements	Males	Amounts	0
PCMA00	Pensioners, Combined	Males	Amounts	0
PFL00	Pensioners, Normal/late retirements	Females	Lives	0
PEFL00	Pensioners, Early retirements	Females	Lives	0
PCFL00	Pensioners, Combined	Females	Lives	0
PFA00	Pensioners, Normal/late retirements	Females	Amounts	0
PEFA00	Pensioners, Early retirements	Females	Amounts	0
PCFA00	Pensioners, Combined	Females	Amounts	0

consistency between (for example) males and females, or lives and amounts; see Working Paper 12.

4.4 *Data on Which to Base Projections*

As mentioned in Section 1.3, feedback on question 6 in Working Paper 3 agreed that the largest appropriate population should be used to carry out projections and to determine measures of uncertainty. The Working Party believes that the only suitable data available for and applicable to the UK are those of the national population of the UK, or of the CMI male assured lives experience. We propose that projected rates of improvement by age, sex and calendar year be produced by applying the projection methodologies which are ultimately chosen to both of these datasets (males and females separately in the case of population data).

Some points need to be considered in deciding which of these datasets might be more appropriate. The production of national demographic data is now a devolved responsibility. For the national population, the main sources of data are the mid-year population estimates and the data provided on deaths by the offices of the respective Registrars General for England and Wales, Scotland and Northern Ireland. Data for the UK are obtained by aggregation. Numbers of deaths by age last birthday and by calendar year are generally available back to 1911, although deaths by single year of age over 100 are only available for Scotland from 1973 and for Northern Ireland from 1968. Data for Scotland, Northern Ireland and for England and Wales prior to 1993 are published on a registration basis; data for England and Wales from 1993 onwards are published on an occurrence basis, although data by registration are available.

Mid-year population estimates by single year of age are generally available back to 1961, although aggregated totals only are provided for the oldest ages (this varies according to the year; most recently aggregate totals are provided for ages 90 and over). Thus, these totals need to be disaggregated by some method if mortality rates by individual years at ages 90 and over are required. It should also be remembered that population estimates are estimates; the only actual counts of the population are carried out in the decennial censuses. Population estimates for future years are then obtained by rolling forward the census counts, adding in births and immigrants and subtracting deaths and emigrants. The historical estimates are often revised following the results of a new census.

National population data are on a lives basis and hence may not be appropriate for projections of amounts. Also, the groups of lives being projected are not homogeneous subgroups of the UK population. Data by lives and amounts are available for assured lives and may be more representative of the past experience of the classes of business for which mortality rates are to be projected. The data are available for a longer period in the past than UK population data, but are not as numerous. In theory, the data for male assured lives, if recorded correctly by those offices submitting data in the past, should provide a more accurate measure of the underlying exposed to risk than do the UK population estimates. However, the past experience will include changes in the mix of business and in the offices providing the data. Factors which might influence the choice of base data are discussed further in Section 5.

4.5 *Proposed Outputs*

Full details of the methodology will be published, and it will therefore be possible, in principle, for any user to implement it as required. However, we appreciate that this may require a considerable amount of programming effort. The Mortality Committee has not decided yet what, if any, software and/or other published outputs might be produced by the CMI to assist users with the new techniques.

One possibility that the Working Party is exploring is to produce a software package which will take as its inputs the past data and choice of key parameters, and produce as its outputs scenarios of future projected rates of mortality improvement. If we do this, the software is likely to be based on an Excel front-end and a calculation engine written in R[12]. This would require users to install R, and we are aware that some companies might have to consider this requirement in the light of their IT policies. We would welcome feedback on this question. However, if it proves possible to provide such software, users may be able very easily to customise projections by changing the input data and parameters as they wished (eg age range, number of past years included).

Another, simpler possibility is that the CMI might publish a compact disc containing a suitably large number of scenarios generated by either or both of the methods described in Sections 2 and 5, once these are finalised and the Working Party has decided to propose them.

Feedback is welcome on the type of software and/or the outputs which users would prefer to have available.

[12] R is a very powerful statistical package, based on the S-Plus language. It has become a standard package in university statistics departments, and is the main resource used for all the work of the Working Party. It is available free to download over the internet at `www.r-project.org`.

5. Further Stages and Topics for Future Research

5.1 *Introduction*

So far the paper has described two main groups of models that may be used to produce future mortality scenarios. Pros and cons of the two approaches, sample results, and comments on proposed outputs have been given.

In this section of the paper we discuss some general, but fundamental, issues and outline topics for further research.

Finally, we give a list of areas where feedback is sought.

5.2 *Varying Rates of Improvement for Different Sub-groups*

In Section 4.4 we stated that modelling should be based on either population experience or the CMI experience of (male) assured lives. This raises an obvious question, namely to what extent can improvement rates based on these populations be considered appropriate for modelling the mortality of annuitants and pensioners? Willets (1999) and Willets *et al.* (2004) have provided detailed descriptions of how the pace of improvement has been consistently more rapid for higher socio-economic groups, such as annuitants and pensioners, than it has been for the general population. There is evidence that, in recent decades, the average annual rate of improvement has been around 25–50% higher for pensioners and annuitants.

Furthermore, there is evidence that, in different subgroups, the 'cohort effect' has applied to people born in different generations. In the general population, the most rapid improvements have been experienced by people born in or around 1931 (GAD, 1995) whereas in the CMI assured lives experience, the fastest improvements were experienced for male assured lives born in or around 1926 (CMI, 2002). Generally speaking, rapid improvements have occurred 'earlier' for higher socio-economic class groups, which may be the result of earlier adoption of lifestyle changes beneficial to health, such as reduction of smoking prevalence (Willets, 2004).

If population mortality is used as the basis for a stochastic projection methodology, how should results of a fitted model be adjusted to reflect the socio-economic class characteristics of the annuitants to whom the projection is applied?

It seems reasonable that the magnitude and form of any such adjustment should be left to individual users of the models. An appropriate adjustment for (say) pensioners in a scheme sponsored by a heavy engineering company would clearly be very different from one appropriate for individuals using their Open Market Option to purchase annuities through Independent Financial Advisers.

The question does remain, as to whether the CMI should facilitate making such adjustments in any software that it provides, or whether the implementation of such adjustments should be left solely to the discretion of users.

5.3 *Gender Issues*

If the models are applied independently to males and females then it is possible that seemingly 'unrealistic' results may sometimes be obtained. The pace of mortality improvement at higher ages in recent years has generally been more rapid for males than females, in the population of England and Wales. If this feature is projected forwards, male mortality rates may fall below those for females.

It may be possible for the methodology to be adjusted to ensure this does not happen, or, alternatively, it could be left to users to adjust projections appropriately if they feel that such a feature is not appropriate.

5.4 *Understanding the Forces Driving Trends*

The methods described in this paper generate scenarios by applying mathematical models to historic mortality rates. The pace of projected future improvement is therefore a function of past experience and the chosen model. As discussed in Section 2, model uncertainty may be a significant additional source of uncertainty.

Much of the actuarial debate on mortality change in recent years has focussed on understanding the forces driving mortality improvements, and shaping projections accordingly. For instance, Willets (1999) suggested the projection of trends by year of birth and making an allowance for the impact of changes in cigarette smoking prevalence over time.

Our proposed methodologies are not grounded in an understanding of mortality change, other than to the extent they may project existing cohort patterns where appropriate. However, that does not mean that users of the projections cannot, or should not, make suitable adjustments to projections when they deem them to be appropriate. This is very much in line with the most recent professional guidance for actuaries in relation to Individual Capital Assessments:

> "Mortality and morbidity risks can be divided into three broad categories: large-scale events, long-term adverse trends and year-on-year volatility of non-homogeneous blocks of business. ICAs must allow for the impact and likelihood of all types of risk . . .
>
> Significant advances in the treatment of a significant critical illness of the aged (*e.g.* cancer or heart disease) or the development of a commonly available treatment to significantly delay the normal ageing process could be considered a 'large scale event' for a portfolio of annuities or guaranteed annuity options.
>
> Long-term adverse trends are particularly important where policy terms are guaranteed. The ICA should consider firstly, with justification, how any historically observed trends (including cohort effects) might continue, or might continue to accelerate or decelerate. Extreme adverse events should then be reasonably foreseeable worsenings of the expected continuation or its rate of acceleration or deceleration." (GN46 Version 1.1 Section 8.2).

Clearly, in seeking 'justification' of assumed future trends, the professional guidance places considerable emphasis on understanding the drivers of possible future change and consideration of events, relating to the major causes of death, that may lead to substantial future reductions.

It is equally apparent that the proposed stochastic methodology does not seek to address these issues or provide justifications for the scenarios produced. The Working Party feels that such justifications are by their nature subjective and best left to individual actuaries to consider as appropriate. Instead, the proposed methodologies aim to provide a sound mathematical foundation which will aid users in developing their own views.

5.5 *The Width of Confidence Intervals*

Working Paper 3 listed three kinds of uncertainty related to the projection of future mortality trends: model, parameter and stochastic uncertainty. It is not clear which of the three is the most significant. In the case of the P-spline model the percentiles produced just reflect parameter uncertainty. However, it is not apparent whether percentiles that reflected all three kinds of uncertainty perfectly (assuming that such percentiles could be defined in some way) would be wider or narrower than those produced by the P-spline (or any other) model. There is no easy, or objective, way of judging whether the width of confidence intervals is actually appropriate for the task of projecting future mortality rates.

There are several ways in which the path of percentiles could be judged. One is to review medical evidence and consult medical experts as to possible improvements in mortality rates from different causes. However, a mortality model in which different causes of death are modelled separately makes this approach much more feasible.

The other approach is to consider past variations in actual and predicted mortality rates. This method is hampered by the lack of suitable (independent) datasets to consider. It is also implicitly assumes that past and future projection methodologies are equally likely to be successful.

Both of these methods may yield interesting results. However, both are clearly outside the scope of what the Working Party could achieve in the timescales envisaged and are arguably outside the remit of the CMI.

In much the same way as individual actuaries may want to adjust the projections to reflect their own views on the likely path of mortality improvements, some will want to adjust the width of confidence intervals (say by increasing, or reducing, the standard error terms produced by the P-spline model) to reflect their own views on the overall level of uncertainty.

5.6 *Topics for Further Research*

A number of topics could warrant further research, but appear to fall outside the possible scope of the work to be published by the Working Party this year. These include:

(a) model uncertainty;
(b) cause-specific mortality projection;
(c) the correlation between investment and mortality risk
(d) further consideration of other stochastic mortality models available; and
(e) models which project future actuarial assumptions (for future mortality) rather than the mortality rates themselves.

5.7 *Feedback*

Feedback is welcomed on any aspect of this paper, however there are a number of specific points on which the Working Party would especially welcome responses:

1. Are potential users in favour of the broad approach of developing a stochastic methodology?

2. Is there a preference for Lee-Carter or P-Spline models, or should both models be made available?
3. What are users' views on the possible production of spreadsheet models utilising Excel and R, and/or the possible production of a CD containing a suitable number of scenarios, and/or any other form of output that may be desired?
4. Should the models allow users to make appropriate adjustments to the projections (say to reflect the socio-economic class mix of their business) or should the implementation of such adjustments be left to the discretion of individual users?
5. Should the Working Party specify a preferred basis and methodology, or should this be left to the discretion of individual actuaries?

5.8 *Next Steps*

The next stage in the process is to await feedback from users. Feedback should be addressed to Rajeev Shah at `Mortality@cmib.org.uk` by 31 July 2005. The Working Party aims to complete its development of proposed projection methodologies and to publish them as a Working Paper before the end of September 2005.

We stress that the paper will not be the final say on the matter and that we anticipate that subsequent reports exploring further aspects of the work will be forthcoming in future, and indeed that independent research may also be stimulated by the proposals.

5.9 *Conclusion*

This paper has tried to outline the Working Party's progress towards the publication of a projection methodology to accompany the publication of the "00" series of standard mortality tables.

There are no magic answers when it comes to projecting future mortality. The proposed stochastic methodology should not be seen as a means of supplying definite answers to questions that have strong subjective elements. Indeed there may be aspects of actuarial work where a subjective scenario-based approach is particularly appropriate, for instance in considering what capital an annuity writer should hold to cover the possibility of a cure for cancer.

Having said that, the Working Party strongly feels that all users of mortality projections should increasingly focus on the uncertainty surrounding scenarios. Developing a stochastic methodology is an excellent way of making this uncertainty a central feature of the basis. The proposed methodologies also aim to provide mortality projections which have a transparent mathematical foundation.

Acknowledgements

The Working Party is especially grateful to Dr Iain Currie and Mr James Kirkby of Heriot-Watt University, for the major and essential part they have played in producing this Working Paper. As well as providing much advice and discussion, and comments on earlier drafts, the implementations of the models in Section 2 are theirs, and they produced all the figures in Section 2.

References

Benjamin, B. & Pollard, J.H. (1980). *The analysis of mortality and other statistics (second edition).* Heinemann, London.

Cairns, A. J. G. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics* **27**, 313–330.

CMI (2002). *Working Paper 1: An interim basis for adjusting the "92" Series mortality projections for cohort effects.* The Faculty of Actuaries and Institute of Actuaries.

CMI (2004). *Working Paper 3: Projecting future mortality: A discussion paper.* The Faculty of Actuaries and Institute of Actuaries.

CMI (2005a). *Working Paper 11: Responses to Working Paper 3 entitled Projecting future mortality: A discussion paper and further commentary thereon.* The Faculty of Actuaries and Institute of Actuaries.

CMI (2005b). *Working Paper 12: The graduation of the CMI 1999–2002 mortality experience: Feedback on Working Paper 8 and proposed assured lives graduations.* The Faculty of Actuaries and Institute of Actuaries.

Currie, I.D., Durban, M. & Eilers, P. (2003). *Using P-splines to extrapolate two-dimensional Poisson data.* Proceedings of 18th International Workshop on Statistical Modelling, Leuven, Belgium, 97–102.

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

Currie, I.D., Denuit, M., Eilers, P., Gampe, J., Renshaw, A., Oeppen, J. & Richards, S.J. (2004). *Workshop on Lee-Carter Methods* .
http://www.ma.hw.ac.uk/~iain/workshop/workshop.html.

Eilers, P.C.H. & Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–102.

Forfar, D.O., McCutcheon, J.J. & Wilkie, A.D. (1988). On graduation by mathematical formula. *Journal of the Institute of Actuaries* **115**, 1–149.

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

McCullagh, P. & Nelder, J.A. (1989). *Generalized linear models (second edition).* Chapman & Hall, London.

Richards, S.J. & Jones, G.L. (2004). *Financial aspects of longevity risk.* Staple Inn Actuarial Society, London.

Whittaker, E.T. (1923). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society* **41**, 63–75.

Willets, R.C. (1999). *Mortality in the next Millennium.* Staple Inn Actuarial Society, London.

Willets, R.C. (2004). The cohort effect: Insights and explanations. *British Actuarial Journal* **10**, 833–877.

Willets, R.C., Gallop, A.P., Leandro, P.A., Lu, J.L.C., Macdonald, A.S., Miller, K.A., Richards, S.J., Robjohns, N., Ryan, J.P. & Waters, H.R. (2004). Longevity in the 21st Century. *British Actuarial Journal* **10**, 685–832.