

# A VALUE-AT-RISK FRAMEWORK FOR LONGEVITY TREND RISK

BY S. J. RICHARDS\*, I. D. CURRIE AND G. P. RITCHIE

[Updated 9th November 2012]

## ABSTRACT

Longevity risk faced by annuity portfolios and defined-benefit pension schemes is typically long-term, i.e. the risk is of an adverse trend which unfolds over a long period of time. However, there are circumstances when it is useful to know by how much expectations of future mortality rates might change over a single year. Such an approach lies at the heart of the one-year, value-at-risk view of reserves, and also for the pending Solvency II regime for insurers in the European Union. This paper describes a framework for determining how much a longevity liability might change based on new information over the course of one year. It is a general framework and can accommodate a wide choice of stochastic projection models, thus allowing the user to explore the importance of model risk. A further benefit of the framework is that it also provides a robustness test for projection models, which is useful in selecting an internal model for management purposes.

## KEYWORDS

Longevity risk, mortality projections, value-at-risk, Solvency II, model risk, Lee-Carter, Age-Period-Cohort, Cairns-Blake-Dowd,  $P$ -spline.

## CONTACT ADDRESS

\*Corresponding author: Stephen Richards, 4 Caledonian Place, Edinburgh, EH11 2AS, United Kingdom; Telephone: +44 (0)131 315 4470; Email: stephen@longevitas.co.uk; Web: www.longevitas.co.uk

## 1. INTRODUCTION

*“Whereas a catastrophe can occur in an instant, longevity risk takes decades to unfold”*

**The Economist (2012)**

1.1 Longevity is different from many other risks an insurer faces because the risk lies in the long-term trend taken by mortality rates. However, although longevity is typically a long-term risk, it is often necessary to pose questions over a short-term horizon, such as a year. Two useful questions in risk management and reserving are “what could happen over the coming year to change the best-estimate projection?” and “by how much could a reserve change based on new information?”. The pending Solvency II regulations for insurers and reinsurers in the EU are concerned with reserves being adequate in 99.5% of situations which might arise over the coming year.

1.2 This paper describes a framework for answering such questions, and for setting reserve requirements for longevity risk based on a one-year horizon instead of the more natural long-term approach. The paper contrasts three approaches to reserving for longevity risk: the stressed-trend method, the mortality-shock method and a new value-at-risk proposal. The framework presented here is general, and can work with any stochastic projection model which can be fitted to data and is capable of generating sample paths. Like Börger (2010), Plat (2011) and Cairns (2011), we will work with all-cause mortality rates rather than rates disaggregated by cause of death.

1.3 Börger (2010) and Cairns (2011) both work with so-called forward models of mortality, i.e. models which produce multi-year survival probabilities as their output. In this paper we will work with what are sometimes described as spot models for mortality, i.e. models which produce instantaneous rates such as the force of mortality,  $\mu_x$ , or the rate of mortality,  $q_x$ . A forward model would specify directly a distribution for the survival probability,  ${}_t p_x$ , whereas a spot model would require sample-path simulation in order to empirically derive a distribution for  ${}_t p_x$ . The motivation for a forward model comes from the desire to deal with (or avoid) nested stochastic simulations.

Both Börger (2010) and Plat (2011) approach this problem by presenting a new model specifically designed for value-at-risk problems involving longevity. In writing about some specific challenges regarding dynamic hedging, Cairns (2011) evaluates some approximations to avoid nested stochastic simulations. In contrast, the framework presented in this paper avoids nested simulations by design and without approximation.

1.4 In considering the Solvency Capital Requirement (SCR) for longevity risk, Börger (2010) concluded that “the computation of the SCR for longevity risk via the VaR approach obviously requires stochastic modelling of mortality”. Similarly, Plat (2011) stated that “naturally this requires stochastic mortality rates”. This paper therefore only considers stochastic mortality as a solution to the value-at-risk question of longevity risk. The value-at-risk framework presented in this paper requires stochastic projection models.

1.5 Cairns (2011) warns of the risks in relying on a single model by posing the oft-overlooked questions “what if the parameters [...] have been miscalibrated?” and “what if the model itself is wrong?”. Cairns (2011) further writes that any solution “should be applicable to a wide range of stochastic mortality models”. The framework described in this paper works with a wide variety of models, enabling practitioners to explore the impact of model risk on capital requirements.

1.6 This paper is about the technical question of how to put longevity trend risk into a one-year, value-at-risk framework. It is not a detailed review of current industry practices, which would be a quite different study. Neither is this paper an attempt to evaluate the respective merits of the various models used — the reader is directed to Cairns et al (2009) for just such a comparison, and also to Currie (2012) for questions over the validity of model assumptions. We use a variety of all-cause mortality models to illustrate the methodology proposed in this paper, and so the question of whether to project by all-cause mortality or by cause of death is a side issue here; the interested reader could continue with Booth & Tickle (2008) or Richards (2010).

## 2. DATA

2.1 The data used in this paper are the all-cause 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 data provided by the Office for National Statistics (ONS) for England & Wales for the calendar years 1961–2010 inclusive. This particular data set has death counts and estimated exposures at individual ages up to age 104. We will work here with the subset of ages 50–104, which is most relevant for insurance products sold around retirement ages. The deaths and exposures in the age group labelled “105+” were not used. More detailed discussion of this data set, particularly regarding the estimated exposures, can be found in Richards (2008).

2.2 One consequence of only having data to age 104 is having to decide how to calculate annuity factors for comparison. One option would be to create an arbitrary extension of the projected mortality rates up to (say) age 120. Another alternative is to simply look at temporary annuities to avoid artefacts arising from the arbitrary extrapolation. We use the latter approach in this paper, and we therefore calculate continuously paid temporary annuity factors as follows:

$$\bar{a}_{x:\overline{105-x}|}^i = \int_0^{105-x} {}_t p_{x,y} v^t dt \quad (1)$$

where  $i$  is the discount rate,  $v^t = (1+i)^{-t}$  and  ${}_t p_{x,y}$  is the probability a life aged  $x$  at outset in year  $y$  survives for  $t$  years:

$$\begin{aligned} {}_t p_{x,y} &= \exp\left(-\int_0^t \mu_{x+s,y+s} ds\right) \\ &\approx \exp\left(-\sum_{s=0}^{t-1} \mu_{x+s,y+s}\right) \end{aligned} \quad (2)$$

2.3 Restricting our calculations to temporary annuities has no meaningful consequences at the main ages of interest, as shown in Table 1.

Table 1. Impact of using a different end age,  $\omega$ , for the temporary annuity factor  $\bar{a}_{70:\omega-70}^{3\%}$ . Source: own calculations using S1PA (males) and a discount rate of 3% per annum for cashflows.

End age, Percentage of S1PA:			
$\omega$	100%	80%	60%
120	11.10	12.06	13.32
105	11.10	12.06	13.30

For reference, our temporary continuous annuity factors are calculated using the trapezoidal rule for approximating an integral on a uniform grid:

$$\bar{a}_{x:\bar{n}}^i \approx \frac{1}{2} + \sum_{t=1}^{n-1} {}_t p_x v^t + \frac{1}{2} {}_n p_x v^n \quad (3)$$

2.4 In this paper we will use  $y = 2011$  as a common outset year throughout. From Equations 1 and 2 we will always need a mortality projection for at least  $(105 - x)$  years to calculate the annuity factor, even if we are only looking for the one-year change in the value of the annuity factor. While we have opted for the temporary-annuity solution, it is worth noting that the models of Cairns, Blake & Dowd (2006) and Richards, Kirkby & Currie (2006) are capable of simultaneously extrapolating mortality rates to higher (and lower) ages at the same time as projecting forward in time. These models therefore deserve a special place in the actuarial toolkit, and the subject is discussed in more detail by Richards & Currie (2011) and Currie (2011).

## 3. COMPONENTS OF LONGEVITY RISK

3.1 For high-level work it is often necessary to quote a single capital amount or percentage of reserve held in respect of longevity risk. However, it is a good discipline to itemise the various possible components of longevity risk, and a specimen list is given in Table 2.

Table 2. Sample itemization of the components of longevity risk. A diversifiable risk can be reduced by growing the size of the portfolio and benefiting from the law of large numbers.

<b>Component</b>	<b>Diversifiable?</b>	<b>Comment</b>
Model risk	No	It is impossible to know if the selected projection model is correct. Capital must be held in respect of the risk that one's chosen model is wrong.
Basis risk	No	Models often have to be calibrated to population or industry data, not the data of the portfolio in question. Capital must be held for the risk that the mortality trend followed by the lives in a portfolio is different from that of the population used to calibrate the model.
Trend risk	No	Even if the model is correct and there is no basis risk, an adverse trend may result by chance which is nevertheless fully consistent with the chosen model. Some practitioners may choose to include an allowance for basis risk in their allowance for trend risk.
Volatility	Yes?	Over a one-year time horizon, capital must be held against the case of unusually light mortality experience from seasonal or environmental variation, such as an unusually mild winter and lower-than-normal deaths due to influenza and other infectious diseases. Note that this risk may not be wholly diversifiable, as one year's light mortality experience may equally be the start of an adverse trend.
Idiosyncratic risk	Yes?	Over a one-year time horizon, capital must be held against the case of unusually light mortality experience from random individual variation. See Plat (2011) and Richards & Currie (2009) for examples. Note that this risk may not be wholly diversifiable, as the light mortality experience may be what drives a change in the expectation of the trend.
Mis-estimation risk	Yes	Uncertainty exists over the portfolio's actual underlying mortality rates, since these can only be estimated to a degree of confidence linked to the scale and richness of the data.

3.2 Table 2 is not intended to be exhaustive and, depending on the nature of the liabilities, other longevity-related elements might appear. In a defined-benefit pension scheme, or in a portfolio of bulk-purchase annuities, there would be uncertainty over the proportion of pensioners who were married, and whose death might lead to the payment of a spouse's pension. Similarly, there would

be uncertainty over the age of that spouse. Within an active pension scheme there might be risk related to early retirements, commutation options or death-in-service benefits. Such risks might be less important to a portfolio of individual annuities, but such portfolios would be exposed to additional risk in the form of anti-selection from the existence of the enhanced-annuity market.

3.3 This paper will only address the trend-risk component of Table 2, so the figures in Table 5 and elsewhere can only be minimum values for the total capital requirement for longevity risk. Other components will have to be estimated in very different ways: reserving for model risk requires a degree of subjectivity, while idiosyncratic risk can best be assessed using simulations of the actual portfolio — see Plat (2011) and also Richards & Currie (2009) for some examples for different portfolio sizes. For large portfolios the idiosyncratic risk will often be diversified away almost to zero in the presence of the other components. In contrast, trend risk and model risk will always remain, regardless of how large the portfolio is. For an overview of how capital requirements for longevity risk fit into the wider regulatory framework, see Makin (2011a, 2011b).

#### 4. THE STRESSED-TREND APPROACH TO LONGEVITY RISK

4.1 Longevity risk lies in the long-term trend taken by mortality rates. This trend unfolds over many years as an accumulation of small changes, so a natural approach to reserving is to use a long-term stress projection applying over the potential lifetime of the annuitant or pensioner. This view of longevity trend risk is sometimes called the run-off approach, and it does not correspond with the one-year view demanded by a pure value-at-risk methodology. However, the run-off or stressed-trend approach is arguably the most appropriate way to investigate longevity trend risk, and it should therefore not be viewed as invalid merely because it is not a one-year methodology. While a great many insurance risks fit naturally into a one-year value-at-risk framework, not all risks do. It would be excessively dogmatic to insist that longevity trend risk only be measured over a one-year horizon.

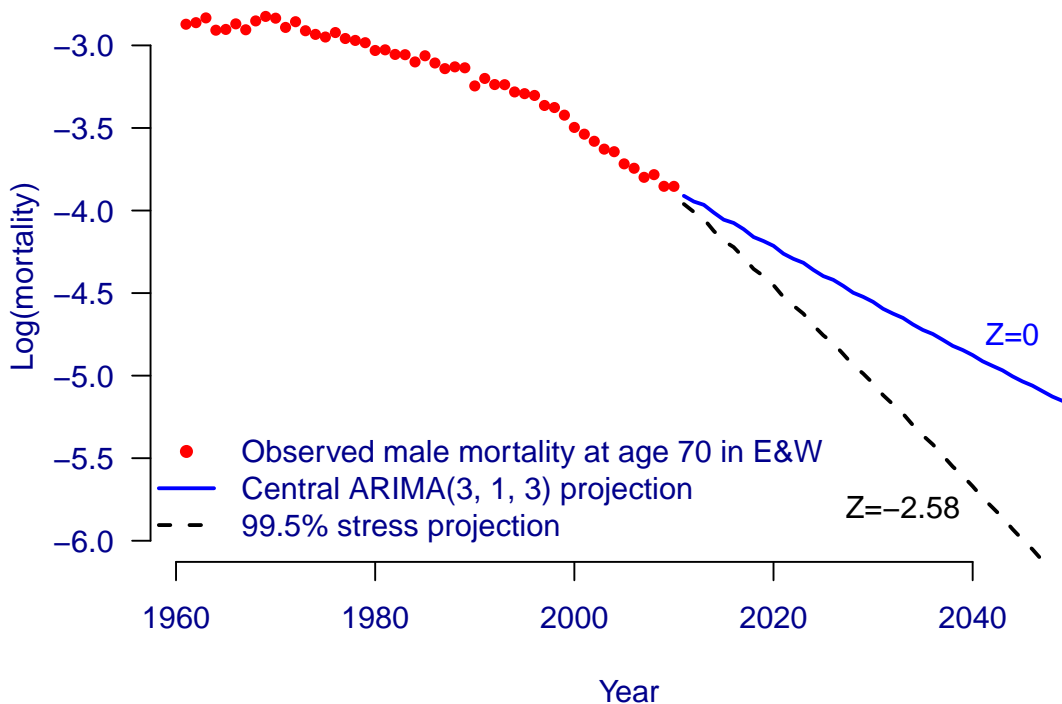


Figure 1. Central projection and stressed trend obtained by deducting 2.58 standard deviations, where  $2.58 = \Phi^{-1}(99.5\%)$ . Mortality of males aged 70 in England & Wales, modelled as per Lee & Carter (1992) with ARIMA(3,1,3) projection of  $\kappa_y$  (see Appendix 1). The ARIMA(3, 1, 3) model was selected from an array of ARIMA( $p, d, q$ ) models as having the lowest AIC for this data set

4.2 Mortality can be viewed as a stochastic process, as can the direction of mortality trends, so

it makes sense to use a stochastic projection model to calibrate a stress trend for longevity risk. This has the benefit of being able to assign a probability to the stress; Figure 1 illustrates the approach. The central projection comes from the maximum-likelihood estimate, while the stress trend is simply the relevant confidence envelope for the central projection. The confidence envelope is derived by multiplying the projection standard error by  $\Phi^{-1}(p)$ , where  $\Phi$  is the cumulative distribution function of a Normal(0, 1) random variable and  $p$  is the relevant probability level ( $\Phi^{-1}(0.005) = -2.58$  in this case). Note that the “projection standard error” here encompasses parameter risk, not process risk, and the nature of the projection standard error is discussed in more detail in Section 8. The stressed trend depicted in Figure 1 is smooth because we do not allow for annual volatility in mortality rates, a subject which is discussed in detail in Section 8.

4.3 The stressed-trend approach to longevity risk involves calculating the capital requirement as follows:

$$\left( \frac{\bar{a}_{x:105-x}^{Z=-2.58}}{\bar{a}_{x:105-x}^{Z=0}} - 1 \right) \times 100\% \quad (4)$$

where the numerator with  $Z = -2.58$  denotes the annuity factor calculated with the stressed trend as per Figure 1, and the denominator with  $Z = 0$  denotes the annuity factor calculated with the central projection. The stressed-trend approach behind Figure 1 can be used for a wide variety of regression-type models — Table 3 shows some annuity factors calculated using central estimates and 99.5% stress reserves, together with the resulting capital requirements. Table 3 also illustrates model risk, namely the tendency for different models to produce different central and stressed estimates, and hence different capital requirements.

Table 3. Central and stressed estimates of  $\bar{a}_{70:35}^{3\%,Z}$  from 2011 using models of male mortality applied to data from England & Wales, 1961–2010, ages 50–104. Temporary annuities to age 105 are used to avoid distortions caused by extrapolation to higher ages. All three variants of the Lee-Carter model produced the same values.

Model (Appendix)	Value of $\bar{a}_{70:35}^{3\%,Z}$ :		Capital requirement $\left( \frac{(b)}{(a)} - 1 \right) \times 100\%$
	(a) Z=0 central estimate	(b) Z=-2.58 99.5% stress	
LC, DDE and LC(S) (A1)	12.13	12.81	5.60%
CBD Gompertz (A2)	11.96	12.70	6.22%
CBD P-spline (A2)	11.87	12.62	6.34%
APC (A3)	12.59	13.14	4.35%
2DAP (A4)	12.76	13.38	4.92%

4.4 The stressed-trend capital requirements in Table 3 are dependent on the model, the outset age and the discount rate. Figure 2 shows how the capital requirements vary by age for four different models. Of the models shown, it is interesting to note that the Lee-Carter model yields the second-largest capital requirement up to age 70, but after age 88 it demands the lowest capital requirement. Although the Age-Period-Cohort model generates consistently low capital requirements, Currie (2012) raises concerns about the assumptions required in projecting mortality using the APC model.

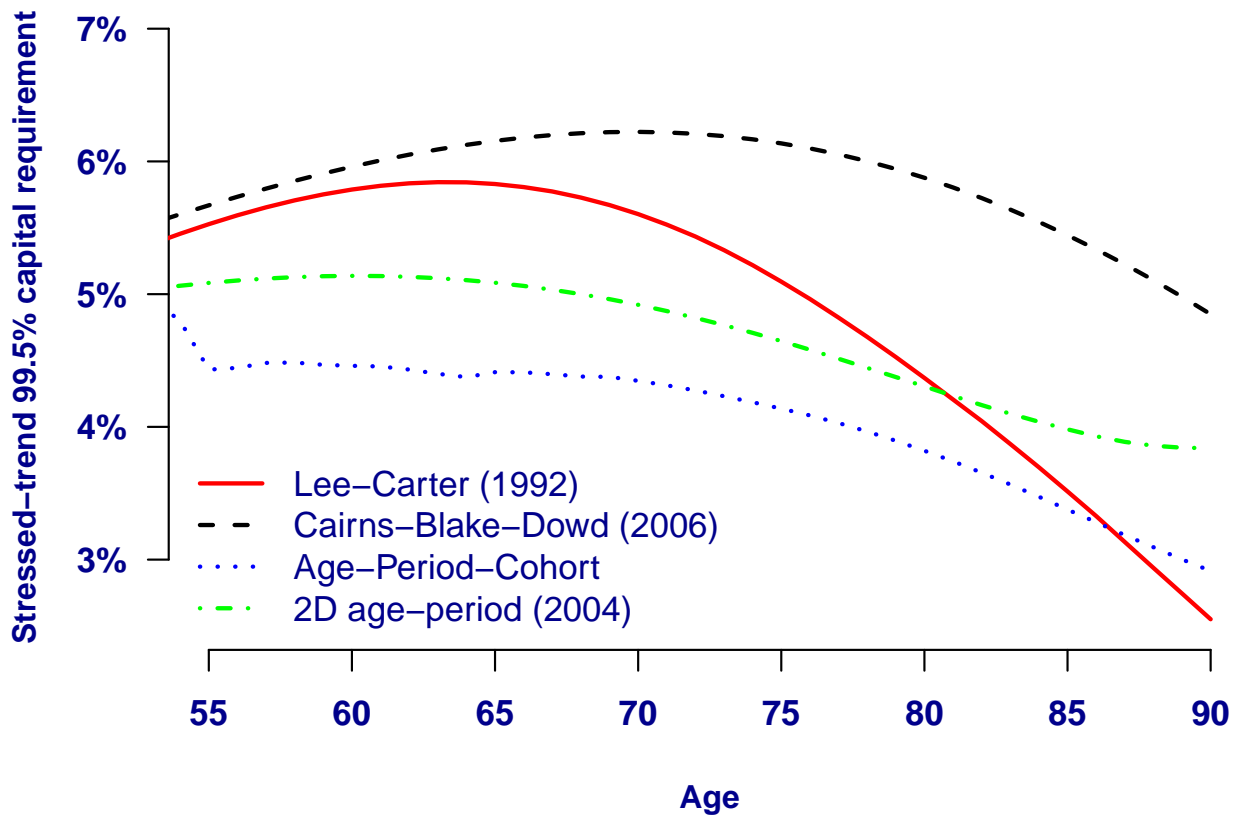


Figure 2. Capital requirement at age  $x$  on valuation date for  $\bar{a}_{x:105-x}^{3\%}$  from stressed trend at 99.5% level for four of the models in Table 3. Graph shows the capital requirement defined in Equation 4 for varying outset ages. The different patterns by annuity outset age show the importance of selecting a specimen age which represents the portfolio for which capital requirements are being calculated. Different models show different patterns by age in part because they vary in the extent to which mortality rates are correlated under each model.

4.5 There is also the question of how best to measure the impact of longevity risk. Plat (2011) carries out a whole-portfolio valuation to minimise the risks associated with representing a portfolio with model points, which makes the published results portfolio-specific. In contrast, we have often chosen a single specimen age for most calculations in this paper on the grounds of simplicity, although in practice for a given portfolio it would be better to follow the approach of Plat (2011). As to the choice of specimen age, Börger (2010) chose age 65 to illustrate the impact of different interest rates, and this age would generally be appropriate for newly written annuity business. In contrast, we have chosen age 70 as it is more likely to be representative of mature portfolios of pensions in payment (i.e. excluding deferred or active members in a pension scheme).

4.6 Another important feature is the sensitivity of the capital requirement to the rate of discount. Börger (2010) notes that capital requirements increase as the (net) discount rate decreases. Figure 3 shows the stressed-trend capital requirements under the Lee-Carter model, which show that lower discount rates lead to higher capital requirements. Interest rates are particularly low at the time of writing, so the presence of significant volumes of escalating annuities could mean an effective net discount rate close to zero for some portfolios. Indeed, interest rates are not only low but the yield curve has a very particular shape which has some consequences of its own (see Appendix A6).

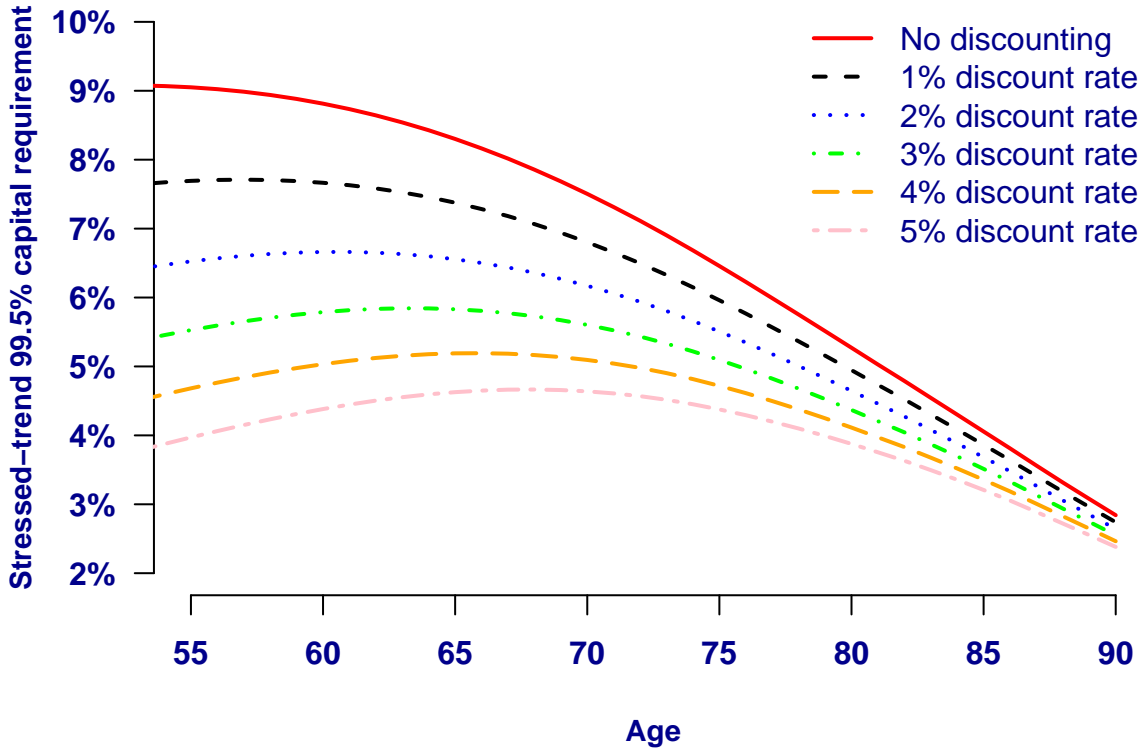


Figure 3. Capital requirement by age for  $\bar{a}_{x:\overline{105-x}|}$  from stressed trend at 99.5% level. Graph shows the capital requirement defined in Equation 4 for varying outset ages and using various discount rates. Mortality is according to the Lee-Carter model from Table 3.

4.7 One consequence of the stressed-trend approach is that, for a given p-value, it will tend to produce higher capital requirements than a one-year approach. Some practitioners use a simple approach to try and allow for this by picking a p-value for the stressed-trend which is less than the 99.5th percentile required for the one-year approach. However, this can be arbitrary, and even less subjective approaches can be inconsistent. For example, one simple approach might be to select a p-value,  $p$ , such that  $0.995^n = 1 - p$ . The parameter  $n$  could be chosen relative to the portfolio under consideration, say by setting it to the discounted mean term of the annuity payments. However, the drawback of this approach is that portfolios exposed to greater levels of longevity risk will receive less onerous values of  $p$ . For example, assume  $n = 10$  for a portfolio of level annuities, thus leading to  $p = 0.0489$ . If we then assume that the annuities are in fact escalating, the value of  $n$  might increase to 12 (say), leading to  $p = 0.0631$ . This is counter-intuitive — the portfolio of escalating annuities is arguably more exposed to longevity risk, so why should it be subjected to a weaker stress scenario? Deriving a one-year longevity stress from a multi-year calculation is tricky, and so other methodologies are required.

## 5. THE SHOCK APPROACH TO LONGEVITY RISK

5.1 A simplified standard-formula approach to longevity risk is to assume an immediate fall in current and future projected mortality rates, i.e.

$$\mu_{x,t}^{\text{shock}} = \mu_{x,t} \times (1 - f) \quad (5)$$

where  $\mu_{x,t}$  is the central mortality projection ( $Z = 0$ ) behind Table 3 and  $f$  is the shock reduction in mortality rates. Analogous to Equation 4, the shock capital requirements are:

$$\left( \frac{\bar{a}_{x:\overline{105-x}|}^f}{\bar{a}_{x:\overline{105-x}|}} - 1 \right) \times 100\% \quad (6)$$



where the numerator is the annuity factor calculated with the shocked mortality rates according to Equation 5 and the denominator is the annuity factor calculated using the central projection. The QIS5 rules for Solvency II specify  $f = 20\%$  (European Commission, 2010), and the results of this are shown in Figure 4 for the Lee-Carter model from Table 3. It is worth noting that the shock approach is designed to cover more aspects of longevity risk than just the trend risk. The subject of the various components of longevity risk is addressed in Section 2.

5.2 The standard-formula approach conflicts with the stressed-trend approach: whereas in Figure 2 the capital requirements reduce after around age 70, the standard-formula capital requirements increase steadily with age to reach excessive levels. Furthermore, the capital requirements at younger ages look to be a little too low compared with the stressed-trend approach of Figure 2. Any given portfolio contains many pensions at different ages, of course, and the appropriateness (or otherwise) of the shock approach depends on the individual portfolio. Richards & Currie (2009) suggested that it might be reasonable for a new and growing annuity portfolio.

5.3 The stressed-trend approach produces a more intuitive set of results than the shock approach — capital requirements are greater at younger ages, where there is more scope for suffering from an adverse long-term trend. Furthermore, comparing Figures 4 and 2 suggests that the stressed-trend approach produces lower capital requirements than the shock approach. Börger (2010) describes the “structural shortcomings of the shock (approach)”, specifically the too-low capital requirements at younger ages and the too-high requirements at older ages, concluding that “an age-dependent stress with smaller reductions for old ages might be more appropriate”. In theory this might not matter, as the shock approach would be intended to work at a portfolio level and excess capital at higher ages would be offset by lower capital requirements at younger ages, although whether this worked in practice would depend critically on the profile of liabilities by age. Also, the shock approach, as structured as a standard formula under Solvency II, is intended to cover more than just longevity trend risk.

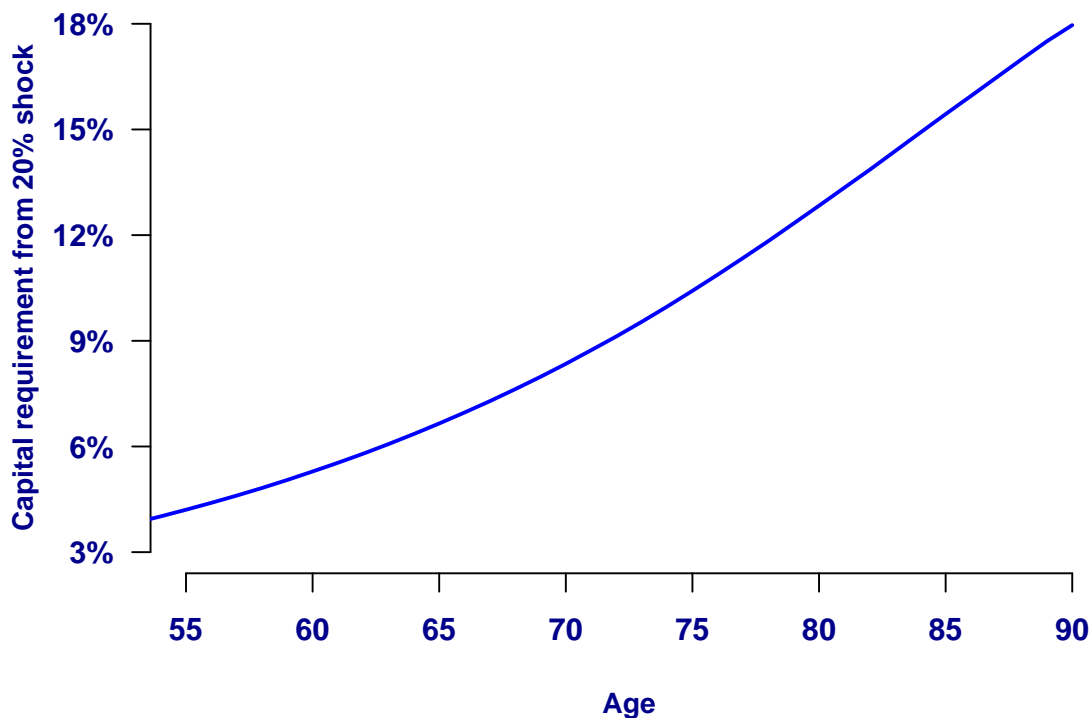


Figure 4. Capital requirement by age for  $\bar{a}_{x:105-x}^{3\%}$  from immediate fall in mortality rates of 20%. Graph shows capital requirements calculated according to Equation 6 using  $f = 0.2$  and the central Lee-Carter projection in Figure 1 ( $Z = 0$ ).

## 6. A VALUE-AT-RISK FRAMEWORK

6.1 The stressed-trend approach of Section 4 deals with changes in mortality rates over many years. This does not answer the one-year value-at-risk question, so something different is needed for the likes of Solvency II. This section describes a one-year framework for longevity risk based on the sensitivity of the central projection to new data. This approach differs from the models of Börger (2010) and Plat (2011), which seek to model the trend and its tail distribution directly. Börger (2010) and Plat (2011) also present specific models, whereas the framework described here is general and can accommodate a wide range of stochastic projection models. In contrast to Plat (2011), who modelled longevity risk and insurance risk, the framework here is intended to focus solely on longevity trend risk in pensions and annuities in payment.

6.2 At a high level we use a stochastic model to simulate the mortality experience of an extra year, and then feed this into an updated model to see how the central projection is affected. This is repeated many times to generate a probability distribution of how the central projection might change over a one-year time horizon. In more detail the framework is as follows:

6.3 First, select a data set covering ages  $x_L$  to  $x_H$  and running from years  $y_L$  to  $y_H$ . This includes the deaths at each age in each year,  $d_{x,y}$ , and the corresponding population exposures. The population exposures can be either the initial exposed-to-risk,  $E_{x,y}$  or the mid-year central exposed-to-risk,  $E_{x,y}^c$ . For this process we need the exposures for the start of year  $y_{H+1}$ , so if the basic exposures are central we will approximate the initial exposures using  $E_{x,y_{H+1}} \approx E_{x-1,y_H}^c - d_{x-1,y_H}/2$ .

6.4 Next, select a statistical model and fit it to the data set in ¶6.3. This gives fitted values for  $\log \mu_{x,y}$ , where  $x$  is the age in years and  $y$  is the calendar year. We can use the projections from this model to calculate various life expectancies and annuity factors at specimen ages if desired.

6.5 Use the statistical model in ¶6.4 to generate sample paths for  $\log \mu_{x,y_{H+1}}$ , i.e. for the year immediately following the last year for which we have data. These sample paths can include trend uncertainty or volatility or both (see Section 8). In practice the dominant source of uncertainty over a one-year horizon is usually volatility, so this should always be included. We can estimate  $q_{x,y_{H+1}}$ , the binomial probability of death in year  $y_{H+1}$ , by using the approximation  $q \approx 1 - e^{-\mu}$ .

6.6 We simulate the number of deaths in year  $y_{H+1}$  at each age as a binomial random variable. The population counts are the  $E_{x,y_{H+1}}$  from ¶6.3 and the binomial probabilities are those simulated in ¶6.5. This gives us simulated death counts at each age apart from  $x_L$ , and we can calculate corresponding mid-year exposures as  $E_{x,y_{H+1}}^c \approx E_{x,y_{H+1}} - d_{x,y_{H+1}}/2$ .

6.7 We then temporarily append our simulated data from ¶6.6 to the real data in ¶6.3, creating a single simulation of the data we might have in one year's time. The missing data for age  $x_L$  in year  $y_{H+1}$  is treated by providing dummy values and assigning a weight of zero. We then refit the statistical model to this combined data set, reperform the projections and recalculate the life expectancies and annuity values at the specimen ages using the updated central projection.

6.8 Repeat steps ¶6.5 to ¶6.7  $n$  times, where  $n$  might be at least 1,000 (say) for Solvency II-style work. It is implicit in this methodology that there is no migration, or that if there is migration its net effect is zero, i.e. that immigrants have similar numbers and mortality characteristics to emigrants. The choice of  $n$  will have a number of practical considerations, but for estimating the 99.5th percentile a minimum value of  $n = 1,000$  is required.

6.9 Figure 5 shows the resulting updated central projections from a handful of instances of performing steps ¶6.3-¶6.8. Note that we do not require nested simulations, as the central projection is evaluated without needing to perform any simulations. The nature of projections and their standard errors is discussed in Section 8.

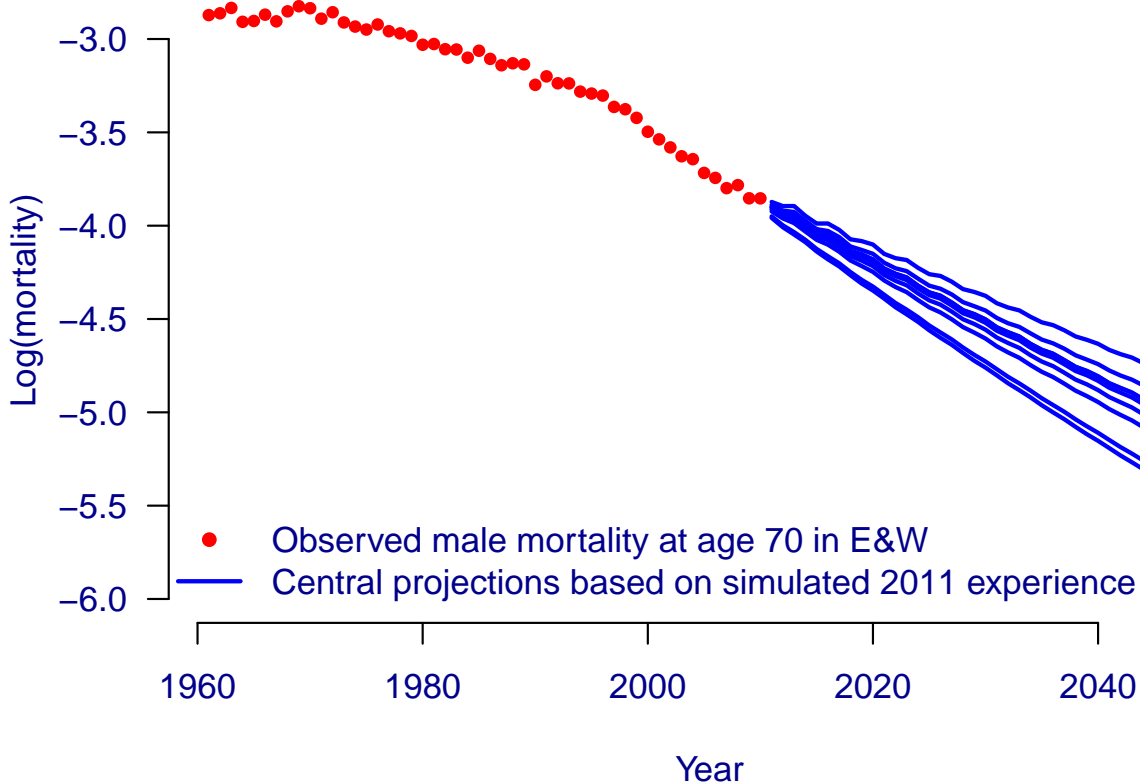


Figure 5. One-year approach to longevity risk. Experience data for 2011 are simulated using sample paths from an ARIMA(3, 1, 3) process and the Lee-Carter model is refitted each time the 2011 data are simulated. The changes in central projections give an idea of how the best estimate could change over the course of a year based on new data. For comparison with Figure 1, all of the projections here are  $Z = 0$ . Although we are interested in the one-year change in annuity factor, we have to do a full multi-year projection in order to have projected mortality rates to calculate the annuity factor in Equation 1.

6.10 After following this procedure we have a set,  $S$ , of  $n$  realised values of how life expectancies or annuity values can change based on the addition of a single year's data:

$$S = \{\bar{a}_x^j; j = 1, 2, \dots, n\} \quad (7)$$

6.11  $S$  can then be used to set a capital requirement to cover potential changes in expectation of longevity trend risk over one year. For example, a one-year value-at-risk estimate of trend-risk capital would be:

$$\left( \frac{99.5^{\text{th}} \text{ percentile of } S}{\text{mean of } S} - 1 \right) \times 100\% \quad (8)$$

6.12 Appendix 5 contains details of how percentiles are estimated from sets of data in this paper. Before we come to the results of this approach in Section 9, we must first consider which models are appropriate (Section 7) and how to generate sample paths (Section 8).

## 7. MODEL CHOICES

7.1 Suitable models for the framework in Section 6 are those which are (i) estimated from data only, i.e. regression-type models where no subjective intervention is required post-fit, and (ii) are capable of generating sample-path projections. Most statistical projection models are therefore potentially suitable, including the Lee-Carter family (Appendix 1), the Cairns-Blake-Dowd model

(Appendix 2), the Age-Period-Cohort model (Appendix 3) and 2D  $P$ -spline model (Appendix 4). This is by no means an exhaustive list, and many other models could be used. Note that a number of models will be sensitive to the choice of time period (i.e.  $y_L$  to  $y_H$ ) and some models will be sensitive to the choice of age range (i.e.  $x_L$  to  $x_H$ ).

7.2 Unsuitable models are those which either (i) require parameters which are subjectively set, or which are set without reference to the basic data, or (ii) are deterministic scenarios. For example, the CMI's 2009–2011 models cannot easily be used here because they are deterministic targeting models which do not generate sample paths — see CMI (2009, 2010). Models which project mortality disaggregated by cause of death could potentially be used, provided the problems surrounding the projection of correlated time series were dealt with — see Richards (2010) for details of other issues with cause-of-death projections. Models which contain artificial limits on the total possible reduction in mortality would not be suitable, however, as the purpose of this exercise is to estimate tail risk. When modelling tail risk, it would be self-defeating to use a model which starts by limiting the tail in question. In addition to this, Oeppen & Vaupel (2004) show that models claiming to know maximum life expectancy (and thus limiting the maximum possible improvements) have a poor record.

## 8. SAMPLE PATHS

8.1 The framework in Section 6 requires the generation of sample-path projections for the following year. There are two components to the uncertainty over projected mortality rates: (i) trend risk, and (ii) volatility. Trend risk is the uncertainty over the general direction taken by mortality rates, and is what is illustrated in Figure 1. Trend uncertainty is determined by the uncertainty over the parameters in the projection model, and is therefore a form of parameter risk. Volatility is an additional risk on top of the particular direction taken — it arises due to temporary fluctuations in mortality around the trend, such as those caused by a harsh (or mild) winter, or an outbreak of influenza. Volatility is fundamentally a short-term phenomenon and would therefore seem to be of less concern to an annuity writer than the trend risk. However, volatility plays an important role in the one-year simulations required for the framework described in Section 6.

8.2 The distinction between trend risk and volatility can be illustrated by examining a drift model, which says that the mortality index  $\kappa$  in year  $y + 1$  is related to the index in the previous year by the simple relationship:

$$\kappa_{y+1} = \kappa_y + \text{drift constant} + \epsilon_{y+1} \quad (9)$$

where  $\kappa_y$  is (say) the Lee-Carter mortality index in year  $y$  and  $\{\epsilon_y\}$  is a set of independent, identically distributed noise terms with zero mean and variance  $\sigma_\epsilon^2$ . In fitting a drift model there are therefore three parameters to be estimated: (i) the drift constant, (ii) the standard error of the drift constant,  $\sigma_{\text{drift}}$ , and (iii) the standard error of the noise process,  $\sigma_\epsilon$ . For an annuity writer, trend risk lies in the uncertainty over the drift constant, i.e. in the parameter risk expressed by  $\sigma_{\text{drift}}$ . For an assurance-type liability the situation may be reversed and the dominant risk might be the noise variance,  $\sigma_\epsilon$ , i.e. the volatility.

8.3 For a detailed risk investigation any simulation should be capable of including or excluding each type of risk independently, i.e. (i) trend risk only, (ii) volatility only, and (iii) trend risk and volatility combined. This enables the actuary to not only reproduce a realistic-looking sample-path projection when required, but also to test the relative contribution of each of the two risks to the costs for a given liability profile.

8.4 The drift model in Equation 9 is a simple and restricted subset of a full ARIMA model for  $\kappa_y$ . The situation for simulations under ARIMA models is analogous: there is a noise process as before, but now  $\kappa_{y+1}$  is related to one or more previous values as an autoregressive moving average. There are therefore more parameters to be estimated than just the drift constant, and each new parameter will also have an estimated standard error corresponding to  $\sigma_{\text{drift}}$ . As before, the variance of the noise process,  $\sigma_\epsilon^2$ , will be labelled as the volatility, while the trend risk lies in

the standard error of  $\log \hat{\mu}_{x,y}$ , which arises from the uncertainty surrounding the various ARIMA parameter estimates.

8.5 For models projecting via a penalty function the situation is different because the forecast is always smooth and there is no volatility. However, simulation of volatility can be achieved by randomly sampling columns from the matrix of unstandardised residuals and adding them column-wise to the smooth projection. This approach is very general and requires no distributional assumption for the volatility — and therefore no estimate of  $\sigma_\epsilon$  — as the empirical set of past volatility is used for sampling. This has the advantage of simulating mortality shocks in a given year while preserving the age structure. For any smooth model, therefore, a simulated sample path,  $\mu_{x,y}^{\text{simulated}}$ , can be obtained as follows:

$$\log \mu_{x,y}^{\text{simulated}} = \log \hat{\mu}_{x,y} + Z\sigma_{x,y} + Ir_{x,y} \tag{10}$$

where  $\hat{\mu}_{x,y}$  is the smooth projected force of mortality and  $\sigma_{x,y}$  the corresponding standard error of  $\log \hat{\mu}_{x,y}$ .  $Z$  is a value from a  $N(0, 1)$  distribution, while  $I$  is an indicator variable taking the value zero (if excluding volatility) or one (if including volatility).  $r_{x,y}$  is derived from a randomly chosen column  $j$  from the unstandardised residuals thus:

$$r_{x,y} = \log D_{x,j} - \log E_{x,j}^c - \log \hat{\mu}_{x,j} \tag{11}$$

where  $j$  is a year in the data region, i.e.  $j \in \{1961, 1962, \dots, 2010\}$  for the data in this paper, and  $y \geq 2011$ . With this method of simulating sample paths for smooth projection models, we again have the facility to consider the two components of risk in Equation 10 either separately or combined. To include trend risk only, we sample  $Z$  from a  $N(0, 1)$  distribution and set  $I = 0$ . To include volatility only, we set  $Z = 0$  and  $I = 1$ . To include both trend risk and volatility, we sample  $Z$  from  $N(0,1)$  and set  $I = 1$ . Note that this approach will work for any model where projections are smooth, including the fully smoothed 2D P-spline models of Richards, Kirkby & Currie (2006) and the variant of the Lee-Carter model in Richards & Currie (2009).

8.6 Although we have very different structures for our projections — drift models, ARIMA processes and penalty forecasts — we can nevertheless arrange a common framework for trend risk (represented by parameter uncertainty) and volatility (which is simulated or resampled from the past departure from the model fit). These two separate components are individually switchable to allow investigation of their respective and combined impact. This is summarised in Table 4.

Table 4. Sources of uncertainty in sample paths

Projection type	Trend risk	Volatility
Drift	$\sigma_{\text{drift}}$	$\sigma_\epsilon$
ARIMA	s.e. of $\log \hat{\mu}_{x,y}$	$\sigma_\epsilon$
Penalty	s.e. of $\log \hat{\mu}_{x,y}$	resampled residuals added to $\log \hat{\mu}_{x,y}$

8.7 There is an irony behind Table 4 and the VaR approach to longevity risk. Longevity risk is essentially about an adverse long-term trend, the uncertainty about which is predominantly described by parameter risk. However, when simulating over a time horizon as short as a year, the volatility is likely to play the major role.

## 9. RESULTS OF THE ONE-YEAR VAR APPROACH

9.1 The framework in Section 6 is applied to some of the models in Table 3. The results for some selected models are shown in Table 5. As a check, the average annuity value in column (a) in Table 5 should be very close to the central annuity value in column (a) in Table 3.

Table 5. Average and 99.5<sup>th</sup> percentile values for  $\bar{a}_{70:35}^{3\%}$  from 2011 using models of male mortality applied to data from England & Wales, 1961–2010, ages 50–104. Temporary annuities to age 105 are used to avoid distortions caused by extrapolation to higher ages. Discounting is at 3% per annum. Percentiles are estimated using the method described in Appendix 5. Results are based on 1,000 simulations according to the procedure described in steps ¶6.3–¶6.8.

Model (Appendix)	Value of $\bar{a}_{70:35}^{3\%}$		Capital requirement $\left(\frac{(b)}{(a)} - 1\right) \times 100\%$
	(a) average value	(b) 99.5 <sup>th</sup> percentile	
LC (A1)	12.14	12.72	4.80%
DDE (A1)	12.15	12.77	5.06%
LC(S) (A1)	12.15	12.76	5.04%
CBD Gompertz (A2)	11.98	12.44	3.85%
CBD <i>P</i> -spline (A2)	11.89	12.36	3.98%
APC (A3)	12.61	13.04	3.40%
2DAP (A4)	12.80	13.69	6.97%

9.2 With the exception of the 2DAP, the capital requirements shown in Table 5 are lower than the equivalent values in Table 3. This is partly because the data behind the models in Table 3 end in 2010 and project rates forward to 2011 and beyond. In contrast, the data behind the models in Table 5 end in 2011 due to the simulated extra year’s experience, so there is one year’s less uncertainty in the projections. However, the primary driver in the reduced capital requirements in Table 5 is the difference in time horizon — a single year as opposed to the run-off in Table 3. As with Table 3, model risk is illustrated by the varying capital requirements.

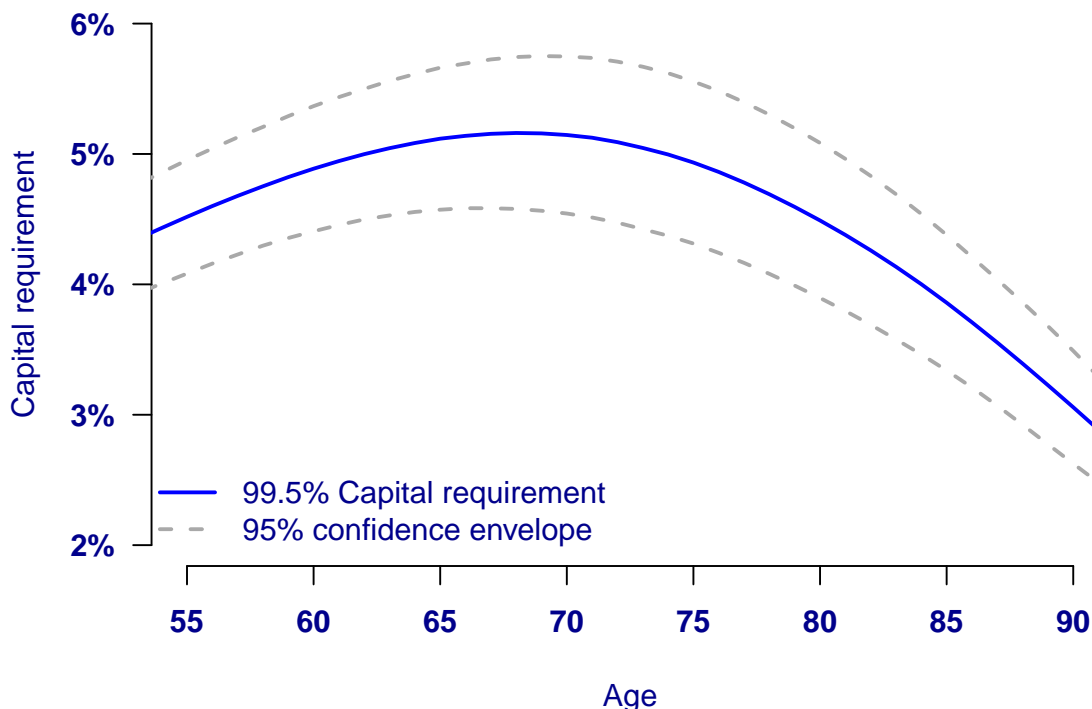


Figure 6. Harrell-Davis (1982) estimate of 99.5% value-at-risk capital requirement for LC(S) model in Table 5, together with approximate 95% confidence envelope.

9.3 One issue with the figures in Table 5 is that they are sample quantiles, i.e. they are based on the top few order statistics and are therefore themselves random variables with uncertainty

surrounding their estimated value. One approach is to use a more sophisticated estimator of the quantile, such as that from Harrell & Davis (1982). Such estimators are more efficient and produce standard errors for the estimator without any distributional assumptions. Figure 6 illustrates the Harrell-Davis estimates of the value-at-risk capital for the smoothed Lee-Carter model, LC(S), together with a confidence envelope around those estimates.

9.4 Plat (2011) felt that there was a risk of underestimation of true risk if the so-called spot model is not responsive to a single extra year’s experience. This is perhaps true of the time-series models shown in Table 5. However, this cuts both ways as overestimation of capital requirements is also possible if the model is too responsive. This is possible for models using penalty projections, such as those of Richards, Kirkby & Currie (2006) and Richards & Currie (2009), and an illustration of what can go wrong is given in ¶10.3.

## 10. A TEST FOR MODEL ROBUSTNESS

10.1 Besides answering the value-at-risk question, the framework outlined here can also be used to test a model for robustness to new data. When selecting a model for internal management use, it is important to know two things about its behaviour when new data becomes available. First, one wants the model to fit, and second, one wants the model to produce sensible results. For example, Richards et al (2007) highlight occasions when two-dimensional  $P$ -spline models can come unstuck due to under-smoothing in the time direction. A particular model must be tested in advance to see if it is vulnerable to such behaviour and either correct it or choose a different model.

10.2 Note that the question of robustness here — will the model fit and work? — is quite different from the question of model suitability, which might itself be addressed by back-testing or the kind of quantitative assessment carried out by Cairns et al (2009). Indeed, the one-year framework presented in this paper might usefully be added to the tests in Cairns et al (2009) when considering an internal model. To illustrate, consider the models listed in Table 5. In each case 1,000 simulations were carried out of the next year’s potential mortality experience, and in each case the model was successfully refitted. The resulting capital requirements are in a similar enough range to conclude that any one of these models could be used as an internal model, safe in the knowledge that once the commitment has been made to use the model the user is unlikely to be surprised by the model’s response to another year’s data.

10.3 In contrast, consider an example where we remove the over-dispersion parameter from the 2DAP model, i.e. we force  $\psi^2 = 1$  regardless of the over-dispersion apparent in the data (see Appendix 4 for details of the over-dispersion parameter). With 1,000 simulations of the framework in steps ¶6.3 to ¶6.8, there were five instances where the model could not fit. Although this failure rate sounds small (0.5%), the implied capital requirement is also of the order of 30%. When the model fits are inspected, we see that a number of the simulations have led to under-smoothing in time and thus excessively volatile projections. The 2DAP model is therefore only usable as an internal model with the England & Wales dataset with the over-dispersion parameter. Analogous situations are possible with many other models, and it is obviously important to find a model’s weak points before committing to use it.

10.4 The procedure described in steps ¶6.3–¶6.8 is therefore not only useful for value-at-risk calculations, it is also a practical test of the robustness of a model. The procedure can be used to check that the model selected this year will continue to perform sensibly when new data arrive next year. Furthermore, there is no reason to force the generating stochastic model to be the same as the one being tested. The reality of model risk — as shown in Tables 3 and 5— means that a thorough robustness test would involve testing a model’s ability to handle the sample paths generated by a different model. Without this sort of test, there is a risk of investing in a given model only for it to surprise (or embarrass) when new data become available. It therefore makes sense to thoroughly test a model before committing to using it.

## 11. CONCLUSIONS

11.1 There are a number of components to longevity risk, of which trend risk is just one part.

The longevity trend risk faced by insurers and defined-benefit pension schemes exists as a long-term accumulation of small changes, which could together add up to an adverse trend. Despite the long-term nature of longevity risk, there are reasons why insurers and others want to look at longevity through a one-year, value-at-risk prism. These reasons include the one-year horizon demanded by the ICA regime in the United Kingdom and the pending Solvency II regime for the European Union. This paper describes a framework for putting a long-term longevity trend risk into a one-year view for setting capital requirements. The results of using this framework tend to produce lower capital requirements than the stressed-trend approach to valuing longevity risk, but this is not uniformly the case.

11.2 Whatever the choice of method — stressed trend, mortality shock or value-at-risk — the actual capital requirements depend on the age and interest rate used in the calculations, and also the choice of model. However, the three approaches used in this paper suggest that the capital requirement in respect of longevity trend risk in level annuities should not be less than  $3\frac{1}{2}\%$  of the best-estimate reserve at the time of writing, and will often be higher. For escalating annuities, or for indexed pensions in payment, the minimum capital requirement in respect of longevity trend risk will be higher still.

#### ACKNOWLEDGEMENTS

The authors thank Stephen Makin, Matthias Börger, Rakesh Parekh, Richard Plat and David Wilkie for helpful comments. Any errors or omissions remain the sole responsibility of the authors. All models were fitted using the [Projections Toolkit](#). Graphs were done in R and typesetting was done in pdfTEX. The R source code for the graphs and the output data in CSV format are freely available at <http://www.longevity.co.uk/var.html>

#### REFERENCES

- BÖRGER, M. (2010). Deterministic shock vs. stochastic value-at-risk: An analysis of the Solvency II standard model approach to longevity risk, *Blätter DGVFM*, **31**, 225–259.
- BOOTH, H. & TICKLE, L. (2008). Mortality modelling and forecasting: a review of methods, *Annals of Actuarial Science*, Vol. 3, Parts 1 and 2, page 8.
- 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., BLAKE, D. & DOWD, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration, *Journal of Risk and Insurance*, **73**, 687–718.
- CAIRNS, A. J. G., BLAKE, D., DOWD, K., COUGHLAN, G. D., EPSTEIN, D., ONG, A. & BALEVICH, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States, *North American Actuarial Journal*, **13(1)**, 1–35.
- CAIRNS, A. J. G. (2011). Modelling and management of longevity risk: approximations to survival functions and dynamic hedging, *Insurance: Mathematics and Economics*, **49**, 438–453.
- CLAYTON, D. & SCHIFFLERS, E. (1987). Models for temporal variation in cancer rates. II: age-period-cohort models, *Statistics in Medicine*, **6**, 469–481.
- CONTINUOUS MORTALITY INVESTIGATION (2009). User Guide for The CMI Mortality Projections Model: ‘CMI 2009’, November 2009.
- CONTINUOUS MORTALITY INVESTIGATION (2010). The CMI Mortality Projections Model, ‘CMI 2010’, Working Paper 49, November 2010.
- CURRIE, I. D., DURBAN, M. & EILERS, P. H. C. (2004). Smoothing and forecasting mortality rates, *Statistical Modelling*, **4**, 279–298.
- CURRIE, I. D. (2011). Modelling and forecasting the mortality of the very old, *ASTIN Bulletin*, **41**, 419–427.
- CURRIE, I. D. (2012). Forecasting with the age-period-cohort model?, *Proceedings of 27th International Workshop on Statistical Modelling, Prague, (to appear)*.
- DE BOOR, C. (2001). A practical guide to splines, *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 penalized likelihood approach, *Statistical Modelling*, **7**, 29–48.



- DJEUNDJE, V. A. B. & CURRIE, I. D. (2011). Smoothing dispersed counts with applications to mortality data, *Annals of Actuarial Science*, **5(I)**, 33–52.
- THE ECONOMIST (2012). The ferment of finance, *Special report on financial innovation, February 25th 2012*, p8.
- EILERS, P. H. C. & MARX, B. D. (1996). Flexible smoothing with  $B$ -splines and penalties, *Statistical Science*, **11**, 89–121.
- EUROPEAN COMMISSION (2010). [QIS5 Technical Specifications](#), Brussels, p152.
- GOMPERTZ, B. (1825). The nature of the function expressive of the law of human mortality, *Philosophical Transactions of the Royal Society*, **115**, 513–585.
- HARRELL, F. E. & DAVIS, C. E. (1982). A new distribution-free quantile estimator, *Biometrika*, **69**, 635–640.
- HYNDMAN, R.J. & FAN, Y. (1996). Sample Quantiles in Statistical Packages, *American Statistician (American Statistical Association)*, November 1996, **50 (4)**, 361–365.
- LEE, R. D. & CARTER, L. (1992). Modeling and forecasting US mortality, *Journal of the American Statistical Association*, **87**, 659–671.
- LINDLEY, D. V. & SCOTT, W. F. (1984). New Cambridge Elementary Statistical Tables, *Cambridge University Press*, p35.
- LONGEVITAS DEVELOPMENT TEAM (2011). [Projections Toolkit v2.2](#), Longevitas Ltd, Edinburgh, United Kingdom.
- MAKIN, S. J. (2011a). Reserving and Regulatory Requirements, *In: Longevity Risk, E. McWilliam (ed.)*, Risk Books, 147–174.
- MAKIN, S. J. (2011b). Economic Capital, Modelling and Longevity Risk, *In: Longevity Risk, E. McWilliam (ed.)*, Risk Books, 175–199.
- OEPPEM, J. & VAUPEL, J. W. (2004). Broken Limits to Life Expectancy, *Science*, **296**, 1029–1031.
- PLAT, R. (2011). One-year Value-at-Risk for longevity and mortality, *Insurance: Mathematics and Economics*, **49(3)**, 462–470.
- R DEVELOPMENT CORE TEAM (2011). R: a language and environment for statistical computing, *R Foundation for Statistical Computing, Vienna, Austria*. ISBN 3-900051-07-0, <http://www.r-project.org>. Accessed on 11th October 2012.
- RENSHAW, A. E. & HABERMAN, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors, *Insurance: Mathematics and Economics*, **38**, 556–570.
- 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 (I)**, 5–61.
- RICHARDS, S. J., ELLAM, J. R., HUBBARD, J., LU, J. L. C, MAKIN, S. J. AND MILLER, K. A. (2007). [Two-dimensional mortality data: patterns and projections](#), *British Actuarial Journal* **13 (III)**, No. 59, 479–555.
- RICHARDS, S. J. (2008). Detecting year-of-birth mortality patterns with limited data, *Journal of the Royal Statistical Society, Series A (2008)* **171**, Part 1, 279–298.
- RICHARDS, S. J. & CURRIE, I. D. (2009). [Longevity risk and annuity pricing with the Lee-Carter model](#), *British Actuarial Journal* **15(II)** No. 65, 317–365 (with discussion).
- RICHARDS, S. J. (2010). Selected Issues in Modelling Mortality by Cause and in Small Populations, *British Actuarial Journal*, **15 (supplement)**, 267–283.
- RICHARDS, S. J. & CURRIE, I. D. (2011). Extrapolating mortality projections by age, *Life and Pensions Risk magazine*, June 2011, 34–38.
- SVENSSON, L. E. O. (1994). Estimating and interpreting forward interest rates: Sweden 1992–1994, *Working paper of the International Monetary Fund*, 94.114.

## APPENDIX 1: LEE-CARTER MODEL

A1.1 Lee & Carter (1992) proposed a single-factor model for the force of mortality defined as follows:

$$\log \mu_{x,y} = \alpha_x + \beta_x \kappa_y \quad (12)$$

where  $\mu_{x,y}$  is the force of mortality applying at age  $x$  in year  $y$ .  $\alpha_x$  is the age-related component of mortality,  $\kappa_y$  is a period effect for year  $y$  and  $\beta_x$  is an age-related modulation of  $\kappa_y$ . The Lee-Carter model is fitted by estimating the parameters  $\alpha_x$ ,  $\beta_x$  and  $\kappa_y$  by the method of maximum likelihood discussed by Brouhns, Denuit & Vermunt (2002). Note, however, that the model specified in Equation 12 cannot be fitted on its own, as there are infinitely many parameterizations which yield the same fitted values. To see this, note that the following transformations yield the same fitted values of  $\mu_{x,y}$  for any real value of  $c$ :

$$\alpha_x \rightarrow \alpha_x^* = \alpha_x - c\beta_x, \quad \beta_x \rightarrow \beta_x^* = \beta_x, \quad \kappa_y \rightarrow \kappa_y^* = \kappa_y + c. \quad (13)$$

A1.2 We therefore fix a convenient parameterisation by setting  $\sum \kappa_j = 0$  and  $\sum \kappa_j^2 = 1$ . This has the attractive feature that  $\alpha_x$  is a measure of average log mortality for age  $x$ .

A1.3 A notable variant of the Lee-Carter model is that of Delwarde, Denuit & Eilers (2007), who used penalised splines to smooth the  $\beta_x$  parameters. This is labelled the DDE model, and a useful feature is that it reduces the risk that mortality rates at adjacent ages will cross over in the projections. A further improvement can be obtained by smoothing both the  $\beta_x$  and the  $\alpha_x$  parameters, which we will denote by LC(S). All three versions of the Lee-Carter model — LC, DDE and LC(S) — could project  $\kappa$  either as a drift model or as a full ARIMA time-series process. A further variant of the Lee-Carter model was proposed by Richards & Currie (2009), which applied spline smoothing not only to  $\beta_x$  but also to  $\kappa_y$ . In this case, projections of  $\kappa_y$  are done by means of the penalty function, as in the 2D  $P$ -spline models in Appendix 4.

A1.4 The drift model is very simple and is described in Equation 9 for  $\kappa$ . In contrast, an ARIMA model is more complicated, but also more flexible and produces more realistic expanding “funnels of doubt” for the projected mortality rates. To understand an ARIMA model, we first define the lag operator,  $L$ , which operates on an element of a time series to produce the previous element. Thus, if we define a collection of time-indexed values  $\{\kappa_y\}$ , then  $L\kappa_y = \kappa_{y-1}$ . Powers of  $L$  mean the operator is repeatedly applied, i.e.  $L^i \kappa_y = \kappa_{y-i}$ . The lag operator is also known as the backshift operator.

A1.5 A time series,  $\kappa_y$ , is said to be integrated if the differences of order  $d$  are stationary, i.e.  $(1 - L)^d \kappa_y$  is stationary. A time series,  $\kappa_y$ , is said to be autoregressive of order  $p$  if it involves a linear combination of the previous values, i.e.  $\left(1 - \sum_{i=1}^p \text{ar}_i L^i\right) \kappa_y$ , where  $\text{ar}_i$  denotes an autoregressive parameter to be estimated. A time series,  $\kappa_y$ , is said to be a moving average of order  $q$  if the current value can be expressed as a linear combination of the past  $q$  error terms, i.e.  $\kappa_y = \left(1 + \sum_{i=1}^q \text{ma}_i L^i\right) \epsilon_y$ , where  $\text{ma}_i$  denotes a moving-average parameter to be estimated and  $\{\epsilon_y\}$  is a sequence of independent, identically distributed error terms with zero mean and common variance. A time series,  $\kappa_y$ , can be modelled combining these three elements as an autoregressive, integrated moving average (ARIMA) model as follows:

$$\left(1 - \sum_{i=1}^p \text{ar}_i L^i\right) (1 - L)^d \kappa_y = \left(1 + \sum_{i=1}^q \text{ma}_i L^i\right) \epsilon_y \quad (14)$$

A1.6 A drift model is sometimes referred to as an ARIMA(0, 1, 0) model, emphasizing that the drift model is a narrowly restricted subset of available ARIMA models. Although an ARIMA( $p$ ,  $d$ ,  $q$ ) model is by definition more complicated than a drift model, it is also more flexible and produces more reasonable extrapolations of past trends. To illustrate this, Figure 7 shows the past mortality

rates for males aged 70 in England & Wales, together with two Lee-Carter projections: a drift model and an ARIMA model.

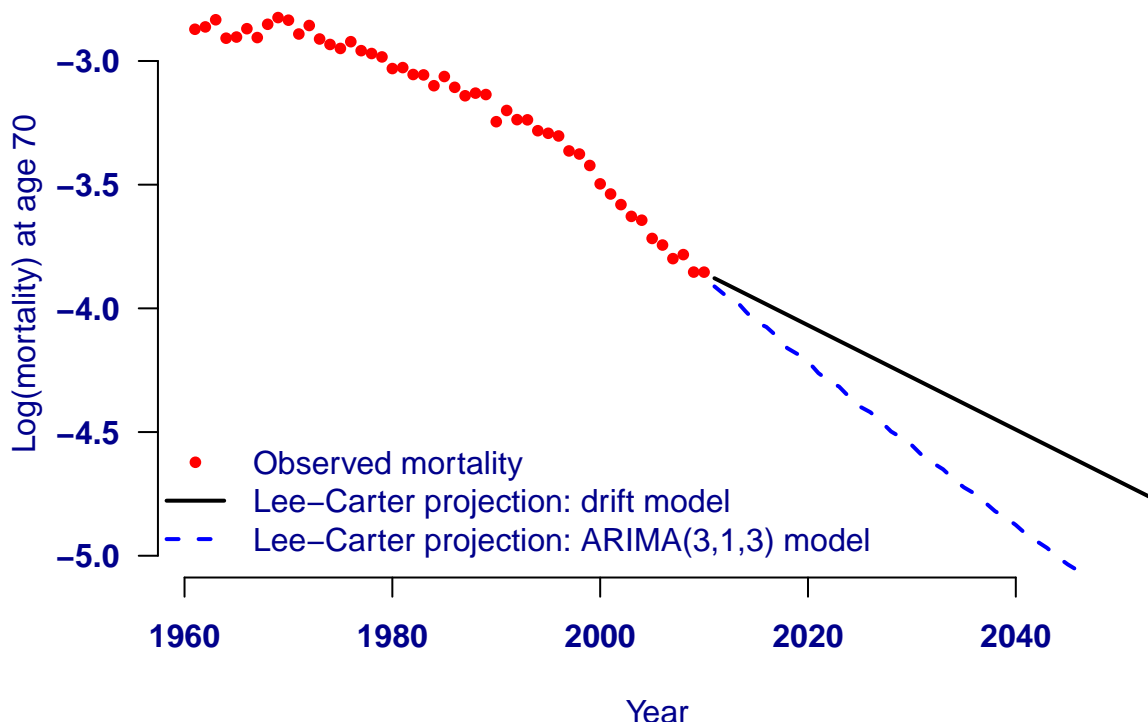


Figure 7. Observed mortality at age 70 with Lee-Carter projections using a drift model (—) and an ARIMA(3, 1, 3) model (- -). The ARIMA(3, 1, 3) model was selected from an array of ARIMA( $p, d, q$ ) models as having the lowest AIC for this data set.

A1.7 The restrictive and rigid nature of the drift-model projection is evident in Figure 7. The gradient of the trajectory of mortality rates shows an increasing rate of improvement, with the most recent twelve or so years of data showing the fastest rate of decline. This is broadly extrapolated by the ARIMA model. In contrast, the drift model projects with a shallower gradient which looks like a less credible extrapolation. Plat (2011) refers to the ARIMA(0,1, $q$ ) model as a kind of “exponential smoothing”, meaning the projection of  $\kappa_y$  depends more heavily on recent values than distant values. This is useful against a backdrop of accelerating improvements, as it yields more sensible-looking extrapolations in Figure 7.

A1.8 ARIMA processes can be fitted with or without a mean; the ARIMA processes for  $\kappa$  in the Lee-Carter models in this paper are fitted with a mean.

## APPENDIX 2: CAIRNS-BLAKE-DOWD (CBD) MODEL

A2.1 Cairns, Blake & Dowd (2006) introduced a two-factor model for the force of mortality defined as follows:

$$\log \mu_{x,y} = \kappa_{0,y} + \kappa_{1,y} (x - \bar{x}) \tag{15}$$

where  $\mu_{x,y}$  is the force of mortality applying at age  $x$  in year  $y$ . Projections of future mortality are done by means of a bivariate random walk with drift for  $\kappa_0$  and  $\kappa_1$ , where both the drift terms and their correlation are estimated from the data.

A2.2 The model in Equation 15 is essentially a Gompertz (1825) model fitted separately for each year  $y$ . Currie (2011) relaxed the Gompertz assumption to allow a general smooth function of age instead of a straight line on a logarithmic scale as follows:

$$\log \mu_{x,y} = \kappa_{0,y} + \kappa_{1,y} S(x) \quad (16)$$

where  $S()$  denotes a smooth function based on penalised splines (of which the Gompertzian straight line is a special case).

A2.3 In addition to projecting mortality rates in time, Richards & Currie (2011) showed how the models in Cairns, Blake & Dowd (2006) & Currie (2011) can also be used to extrapolate mortality rates to higher (and lower) ages. The ability to extrapolate is a useful feature shared with the 2D  $P$ -spline models in Appendix 4.

### APPENDIX 3: AGE-PERIOD-COHORT (APC) MODEL

A3.1 The Age-Period-Cohort model for the force of mortality is defined as follows:

$$\log \mu_{x,y} = \alpha_x + \kappa_y + \gamma_{y-x} \quad (17)$$

where  $\mu_{x,y}$  is the force of mortality applying at age  $x$  in year  $y$  for the birth cohort born in  $y-x$ . There are numerous issues requiring care with the APC model, including the risk of over-interpreting the parameter values and projecting parameters whose values are dictated by the choice of identifiability constraints. The three constraints used in the APC model in this paper are:

$$\begin{aligned} \text{C1} \quad & \sum_j \kappa_j = 0 \\ \text{C2} \quad & \sum_c w_c \gamma_c = 0 \\ \text{C3} \quad & \sum_c w_c c \gamma_c = 0 \end{aligned} \quad (18)$$

A3.2 In C2 and C3  $w_c$  is the cohort weight, namely, the number of times cohort  $c$  occurs in the data set. If we have  $n_x$  ages and  $n_y$  years then we have  $n_c = n_x + n_y - 1$  cohorts and  $c$  runs from 1 (oldest) to  $n_c$  (youngest). The first two constraints are used in Cairns et al (2009), but the third constraint is different (although Cairns et al (2009) do use C3 in their discussion of their models M6 and M7). We prefer constraint C3 to the constraint used in Cairns et al (2009) because we can check that C1, C2 and C3 do indeed specify a unique parameter solution. The non-linear nature of the third constraint in Cairns et al (2009) would appear to have two drawbacks: first, it is very difficult to check that a unique solution has in fact been specified, and second, the efficient numerical methods implemented in Currie (2012) are not available for non-linear constraints.

A3.3 Although popular and seemingly intuitive, Currie (2012) raises concerns about the use of the APC model. Some implementations of the APC model assume a zero correlation between the  $\alpha$ ,  $\kappa$  and  $\gamma$  terms. However, using the England & Wales population data, Currie (2012) found that these parameters were highly correlated, thus invalidating a core assumption of the model. Currie (2012) highlights similar worries about correlations amongst parameters in the Renshaw-Haberman model (Renshaw & Haberman, 2006). The limitations and challenges of the Age-Period-Cohort model are discussed in Clayton & Schifflers (1987).

A3.4 Note that in this paper  $\kappa$  in APC models is projected as a random walk with drift, while  $\gamma$  in APC models is projected as an ARIMA process without a mean.

### APPENDIX 4: 2D $P$ -SPLINE MODELS

A4.1 Currie, Durban & Eilers (2004) proposed the following model:

$$\log \mu_{x,y} = \sum_i \sum_j B_i(x) B_j(y) \theta_{ij} \quad (19)$$

where  $\mu_{x,y}$  is the force of mortality applying at age  $x$  in year  $y$ , and  $B()$  represents a basis spline — see de Boor (2001). This model smoothed the  $\theta_{ij}$  with two of the penalty functions proposed by Eilers & Marx (1996) — one penalty in age and one in calendar time.

A4.2 Richards, Kirkby & Currie (2006) presented an alternative model where the smoothing took place by year of birth instead of year of observation:

$$\log \mu_{x,y} = \sum_i \sum_k B_i(x) B_k(y-x) \theta_{ik} \quad (20)$$

A4.3 As in Richards, Kirkby & Currie (2006), we use a five-year spacing between the knots by age and by time. There is, however, an important difference between the 2D  $P$ -spline models in Richards, Kirkby & Currie (2006) and the ones used in this paper. All the models fitted in this earlier paper assumed that the number of deaths follows a Poisson distribution, a key feature of which is that the variance is equal to the mean. In practice many mortality counts exhibit greater variance than the mean, a phenomenon called over-dispersion. For time-series models this does not pose a problem, but for models using penalty projections over-dispersion can lead to under-smoothing in the time direction, and thus unstable and volatile projections. To allow for this extra variance, and to restore the stability to the penalty projections, we can include an over-dispersion parameter, as described by Djeundje & Currie (2011). This works as follows: if we let  $Y$  be a Poisson random variable with parameter  $\mu$ , we thus have  $E(Y) = \text{Var}(Y) = \mu$ . If we now suppose that the variance of  $Y$  is greater than the Poisson assumption by a factor of  $\psi^2$ , then  $\text{Var}(Y) = \psi^2 \mu$ .  $\psi^2$  is referred to as the over-dispersion parameter, and if  $\psi^2 = 1$  we have the usual Poisson distribution. In practice we expect  $\psi^2 \geq 1$ , but in our implementation we only impose  $\psi^2 > 0$ , thus theoretically allowing the phenomenon of under-dispersion. The importance of the over-dispersion parameter for projections is illustrated in ¶10.3.

A4.4 Richards & Currie (2011) showed how the models in Richards, Kirkby & Currie (2006) can be used not only to project mortality forward in time, but also to extrapolate mortality rates to higher (and lower) ages. This useful property is shared with the CBD model in Appendix 2.

## APPENDIX 5: QUANTILES AND PERCENTILES

A5.1 Quantiles are points taken at regular intervals from the cumulative distribution function of a random variable. They are generally described as  $q$ -quantiles, where  $q$  specifies the number of intervals which are separated by  $q - 1$  points. For example, the 2-quantile is the median, i.e. the point where values of a distribution are equally likely to be above or below this point.

A5.2 A percentile is the name given to a 100-quantile. In Solvency II work we most commonly look for the 99.5<sup>th</sup> percentile, i.e. the point at which the probability that a random event exceeds this value is 0.5%. This is the boundary of the top 200-quantile. There are various ways of estimating a quantile or percentile, but the one most easily accessed by actuaries is the definition used by Microsoft Excel and the R function `quantile()` with the option `type=7`.

A5.3 We will illustrate by using the following five lines of R code to (i) generate 1,000 pseudo-random numbers from a  $N(0,1)$  distribution, (ii) write these values to a temporary file called `data.csv`, and then (iii) calculate the 99.5<sup>th</sup> percentile:

```
set.seed(1)
temp = rnorm(1000)
write(temp, file="C:/data.csv", ncol=1)
sort(temp)[994:1000]
quantile(temp, 0.995, type=7)
```

A5.4 After executing the above instructions, the R console should contain something like the following:

```
> set.seed(1)
> temp = rnorm(1000)
> write(temp, file="C:/data.csv", ncol=1)
> sort(temp)[994:1000]
```

```
[1] 2.401618 2.446531 2.497662 2.649167 2.675741 3.055742 3.810277
> quantile(temp, 0.995, type=7)
99.5%
2.446787
```

where the greater-than symbol denotes a command executed by R and the other three lines are output. We can see the seven largest values, and we can also see that the R `quantile()` function has returned 2.446787. If we read the contents of the file `data.csv` into Excel we can use the equivalent `PERCENTILE(A:A,0.995)` function. It returns the value 2.446786655, i.e. R and Excel agree to at least seven significant figures.

A5.5 Note that the answer produced by R and Excel is not one of the seven largest values in the data. This is because in both cases the software is interpolating between the fifth and sixth largest values. In general, we seek a percentile level  $p \in (0, 1)$ . If  $x[i]$  denotes the  $i^{\text{th}}$  largest value in a data set, then the value sought is  $x[(n-1)p+1]$ . In the example above,  $n = 1000$  and  $p = 0.995$ , so  $(n-1)p+1 = 995.005$ . This latter value is not an integer, so we must interpolate between the 995<sup>th</sup> and 996<sup>th</sup> largest values. The final answer is then:

$$0.995 * 2.446531 + 0.005 * 2.497662 = 2.446787 \quad (21)$$

which agrees with Excel and R answers to at least seven significant figures. There are other methods of calculating quantiles and percentiles, however, and further information can be found in Hyndman & Fan (1996).

A5.6 An important point to note is that sample quantiles and percentiles are estimators, and are thus subject to uncertainty themselves. This is evidenced by the above example producing an estimate of the 99.5th percentile as 2.4468, whereas the true value for the  $N(0,1)$  distribution in question is 2.5758 (Lindley & Scott, 1984). There are many ways of calculating the confidence intervals for a quantile, including bootstrapping, but for illustration here we use the methods of Harrell & Davis (1982) by running the following two R statements (in addition to the five above):

```
library(Hmisc)
hdquantile(temp, 0.995, se=TRUE, names=FALSE)
```

A5.7 After executing the above two instructions immediately following the initial five above, the R console should contain something like the following:

```
> hdquantile(temp, 0.995, se=TRUE, names=FALSE)
[1] 2.534310
attr(,"se")
[1] 0.1360113
```

which shows three things: (i) the Harrell-Davis estimate of the 99.5th percentile (2.5343) is noticeably closer to the true known value for the standard normal distribution (2.5758), courtesy of greater efficiency, (ii) the standard error of the estimate of the percentile is substantial with 1,000 values, and (iii) the true known value is well within the 95% confidence interval implied by this estimate and standard error.

## APPENDIX 6: YIELD CURVES

A6.1 A yield curve describes the pattern of redemption yields by outstanding term for a given class of bonds. At the time of writing, a complication is that the assumption of a constant discount rate is not an accurate one — Figure 8 shows the yield curve implied by principal strips of UK government gilts on 27<sup>th</sup> April 2012. The non-constant nature of the yield curve causes some complications for the shape of capital requirements, as shown in Figure 9. Above age 80 the capital requirement for the given yield curve is best approximated by a constant discount rate of 3% p.a.,

but below age 70 a discount rate of 4% p.a. would be a closer approximation. Indeed, the capital requirements below age 65 are in fact slightly lower than those for the 4% curve, despite the fact that the yield curve in Figure 8 peaks at 3.71%. The interaction of the survival curve and the shape of the yield curve is therefore not a trivial point. Makin (2011a) considers the finer points of interaction between longevity risk and interest rates.

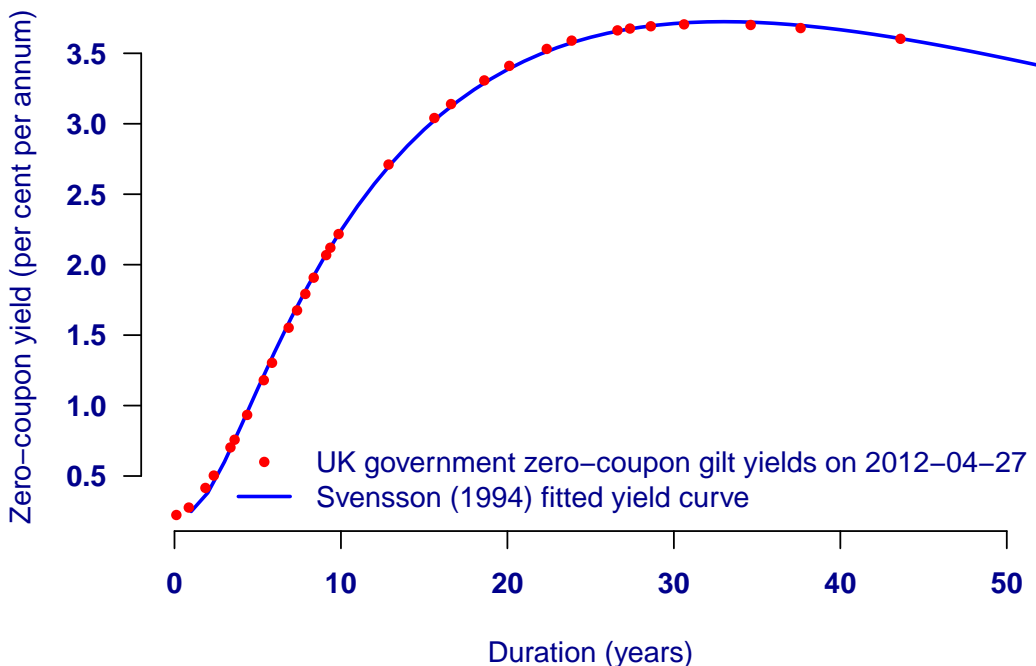


Figure 8. Zero-coupon yields implied by prices of principal strips of UK gilts on 27<sup>th</sup> April 2012 (•) and fitted Svensson (1994) model (—).

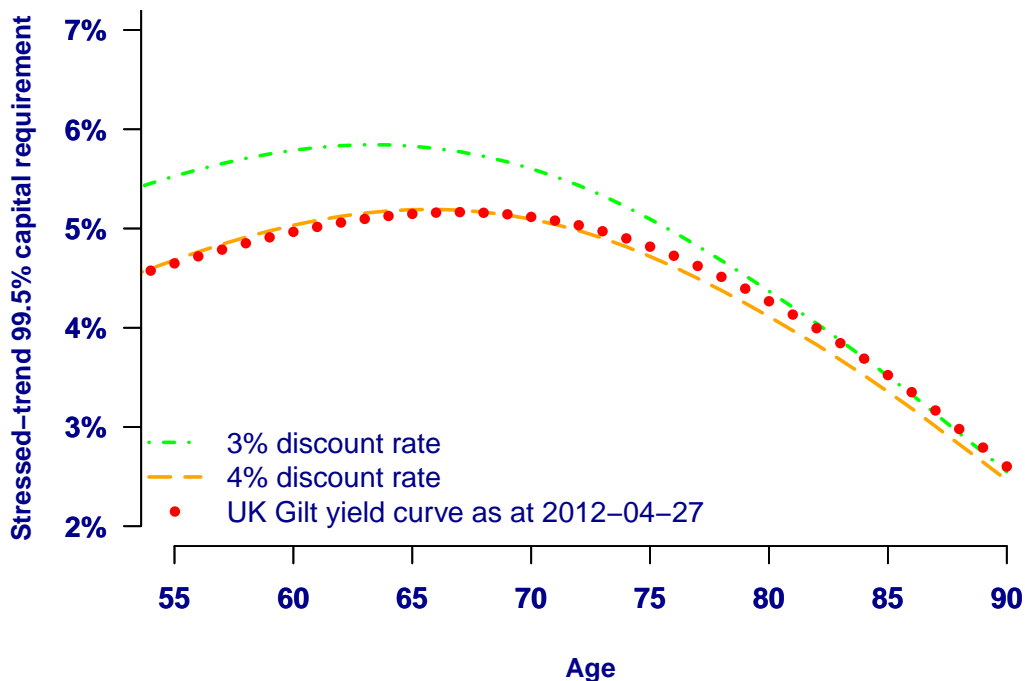


Figure 9. Capital requirement by age for  $\bar{a}_{x:\overline{105-x}|}$  from stressed longevity trend at 99.5% level for various discount rates using the Lee-Carter model from Table 3 (dashed lines) and a Svensson (1994) yield-curve model fitted to the yields implied by prices of principal strips of UK gilts on 27<sup>th</sup> April 2012 (•).