Annals of Actuarial Science (2020), **14**, pp. 262–277
doi:10.1017/S174849952000007X

PAPER

Longevity trend risk over limited time horizons

Stephen J. Richards[1]*, Iain D. Currie[2], Torsten Kleinow[2] and Gavin P. Ritchie[1]

[1]Longevitas Ltd, 24a Ainslie Place, Edinburgh EH3 6AJ, UK; and [2]Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, UK
*Corresponding author. E-mail: stephen@longevitas.co.uk

(Received 26 March 2020; accepted 26 March 2020; first published online 21 May 2020)

Abstract

We consider various aspects of longevity trend risk viewed through the prism of a finite time window. We show the broad equivalence of value-at-risk (VaR) capital requirements at a p-value of 99.5% to conditional tail expectations (CTEs) at 99%. We also show how deferred annuities have higher risk, which can require double the solvency capital of equivalently aged immediate anuities. However, results vary considerably with the choice of model and so longevity trend-risk capital can only be determined through consideration of multiple models to inform actuarial judgement. This model risk is even starker when trying to value longevity derivatives. We briefly discuss the importance of using smoothed models and describe two methods to considerably shorten VaR and CTE run times.

Keywords: Longevity trend risk; Value-at-risk; Conditional tail expectation; Solvency II; ORSA; Immediate annuity; Deferred annuity; Index products

1. Introduction

Solvency II in the European Union (EU), and similar regulations elsewhere, specify a value-at-risk (VaR) approach to solvency assessment. Other territories specify regulations based around conditional tail expectations (CTE), also known as tail-VaR (tVaR). Both solvency approaches view risk over a 1-year horizon, and both calculations can be performed using the same sample data.

However, there are other horizons besides 1 year where a VaR assessment of longevity trend risk is useful. One example is the 3–5 year horizon for insurer business planning, another is for pension schemes aiming to buy out their liabilities in the medium term and a third is the pay-off of an index-based longevity hedge. This paper looks at n-year VaR and CTE assessments of longevity trend risk for each of these scenarios and contrasts the results with the familiar 1-year solvency view. We consider the longevity risk in both deferred and immediate annuities, and we discuss some methods whereby VaR and CTE run times can be considerably shortened without losing accuracy.

Section 2 provides some definitions. Section 3 outlines the data used and highights some important differences to the data used in related past work. Section 4 gives an overview of the models and how they are fitted. Section 5 considers insurer solvency calculations, while Section 6 considers insurers' business planning. Section 7 looks at pension schemes with a medium-term horizon

Data were downloaded on 21 May 2019 from the Human Mortality Database (HMD). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de.

for buyouts, while Section 8 considers the valuation of longevity hedging instruments. Section 9 concludes the paper.

2. Definitions

We denote by $V_{x,n}$ the random variable representing the expected present value of a liability contingent on a forecast of future mortality rates for a person aged x in n years' time. In this paper, $V_{x,n}$ will be either the reserve for an annuity in payment or else a deferred annuity. We assume that all basis elements for the calculation of $V_{x,n}$ are known apart from the future mortality improvement rates, i.e., that $V_{x,n}$ is random with respect to longevity trend risk only (we will discount cash flows throughout the paper at a constant net rate of 1% per annum). The best-estimate of the liability is assumed to be $\mathbb{E}[V_{x,n}]$ and the risk measures of interest are

$$\text{VaR}_{\alpha,x,n} = \left(\frac{Q_{\alpha,x,n}}{\mathbb{E}[V_{x,n}]} - 1 \right) \times 100\% \quad (1)$$

$$\text{CTE}_{\alpha,x,n} = \left(\frac{\mathbb{E}[V_{x,n} | V_{x,n} > Q_{\alpha,x,n}]}{\mathbb{E}[V_{x,n}]} - 1 \right) \times 100\% \quad (2)$$

where $Q_{\alpha,x,n}$ is the α-quantile of the distribution of $V_{x,n}$, i.e., $\Pr(V_{x,n} \le Q_{\alpha,x,n}) = \alpha$. $\text{VaR}_{\alpha,x,n}$ is the percentage extra capital on top of the best-estimate to cover a proportion α of losses that might occur over the following n years due to a change in the best-estimate assumption. $\text{CTE}_{\alpha,x,n}$ is the percentage extra capital on top of the best-estimate to cover the average expected loss if an extreme event with probability $1-\alpha$ occurs in the next n years. See Hardy (2006) for discussion and contrast of the properties of VaR- and CTE-style approaches to risk measurement. $Q_{a,x,n}$ is an order statistic, and in this paper, we will estimate it using the methodology from Harrell & Davis (1982).

In the EU, the Solvency II regime for insurer solvency calculations is based on $\text{VaR}_{99.5\%,x,1}$, while the Swiss Solvency Test is based on $\text{CTE}_{99\%,x,1}$. Both are 1-year views of risk and both typically produce values in the interval (0%, 10%) for longevity trend risk.

Solvency II further specifies an Own Risk and Solvency Assessment (ORSA) for regular use in business planning. This is also a $\text{VaR}_{\alpha,x,n}$ calculation, but with n typically 3–5 years. A specific value for α is not laid down, but we will use $\alpha = 95\%$ in this paper for ORSA-style calculations.

3. Data

We use data from the Human Mortality Database (HMD) for the United Kingdom and the Netherlands at ages 50–104 years over the period 1971–2016. The data are (i) $d_{x,y}$, the deaths classified by age last birthday, x, and year of occurrence, y, and (ii) $E^c_{x,y}$, the corresponding central exposed to risk. Data are available separately for males and females and lend themselves to modelling the central rate of mortality, $m_{x,y}$, without modification. The data also lend themselves to modelling $\mu_{x+\frac{1}{2},y+\frac{1}{2}}$, the force of mortality at age $x+\frac{1}{2}$ at time $y+\frac{1}{2}$. In this paper, we will model the force of mortality, and we will drop the $+\frac{1}{2}$ for simplicity.

There are two aspects of the data which affect comparisons of the results in this paper with those in earlier works. The first aspect concerns the UK data only, where population estimates were updated following the 2011 census; see Cairns *et al.* (2015). Figure 1 shows the ratio of population estimates for 2010 before and after the census, showing large changes (reductions) in the estimates at advanced ages. This will change the calibration of models in earlier papers that used population estimates made prior to the 2011 census.

https://doi.org/10.1017/S174849952000007X Published online by Cambridge University Press

Table 1. Models used in this paper

Label	Reference	Details
LC	Lee & Carter (1992)	$\log \mu_{x,y} = \alpha_x + \beta_x \kappa_y$
DDE	Delwarde *et al.* (2007)	As LC, but with β_x smoothed with *P*-splines
LC(S)	Currie (2013)	As LC, but with both α_x and β_x smoothed with *P*-splines
APC(S)		$\log \mu_{x,y} = \alpha_x + \kappa_y + \gamma_{y-x}$ with α_x smoothed with *P*-splines
M5(S)	Currie (2011)	$\log \mu_{x,y} = \kappa_{0,y} + \kappa_{1,y} S(x)$ where $S(x)$ is a *P*-spline-smoothed function of age
2DAP	Currie *et al.* (2004)	2D model with *P*-splines for age and period
2DAC	Richards *et al.* (2006)	2D model with *P*-splines for age and cohort

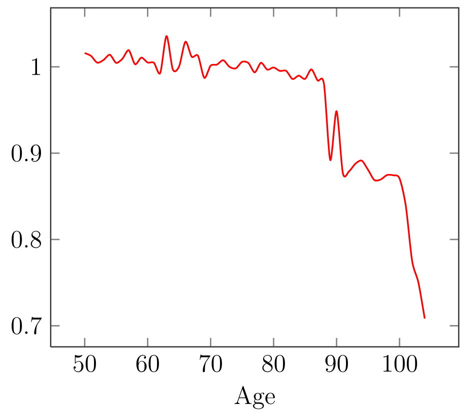

Figure 1. Population estimates for males in England & Wales in 2010: ratio of post-census estimate to pre-census estimate. After the 2011 census, the population estimates above 90 years of age were revised downwards, thus increasing stated mortality rates (the deaths data were unchanged). Source: own calculations using ONS data.

The second aspect is that the HMD revised its methods protocol in 2017 to improve estimates of exposures using the distribution of births for each cohort; see Wilmoth *et al.* (2017). Updated population estimates for the UK and the Netherlands became available in May 2018; see Boumezoued (2020) for further details of how exposure estimates are enhanced using fertility data, and why this is necessary. The effects of this change are twofold: first, false cohort effects are removed from the data; second, the volatility of mortality rates is reduced. Both of these effects will change model fits, projection behaviour, and thus the results of any VaR assessments. Removal of false cohort effects will obviously change the parameterisation of any model incorporating a cohort term, while the reduction in volatility will change the calibration of any projection model based on a time series. As shown in Kleinow & Richards (2016, Section 7), the volatility of the forecasting process plays a dominant role in the 1-year VaR capital requirements for longevity trend risk.

4. Models and Their Fitting

We assume that the number of deaths at each age x in year y, $d_{x,y}$, has a Poisson distribution with mean $E^c_{x,y}\mu_{x,y}$. Models are fitted by maximising the log-likelihood with the functional forms for $\mu_{x,y}$ shown in Table 1. Smoothing is achieved with the method of *P*-splines of Eilers & Marx (1996). Here, parameters for individual ages are replaced by the coefficients in a regression model where the regression matrix is computed with an appropriate basis of *B*-splines; the coefficients are chosen by optimising a penalised likelihood. Over-dispersion of the Poisson death counts is allowed for directly in the 2DAP and 2DAC models.

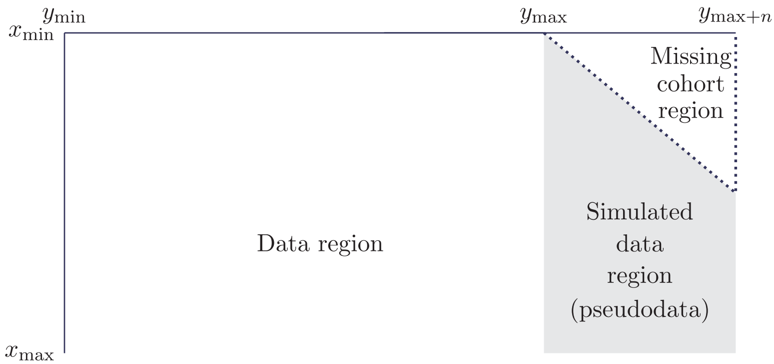

Figure 2. Simulation of *n* years of population experience data. Rolling forward the population results in an increasingly large triangle of missing pseudodata in the upper right. This region can be weighted out in the subsequent model refits, but as *n* increases, it becomes trickier to fit models containing cohort terms.

The LC, DDE, LC(S) and APC(S) models project κ_y as an ARIMA time series, with the order chosen by minimising the AIC (Akaike 1987) with a small-sample correction (Macdonald *et al.*, 2018, p 98). ARIMA models are fitted with R's `arima()` command using a limited number of attempts with different initial values and we select the fit that has the smallest AIC (see also Appendix A). Standard linear identifiability constraints were used, although the precise choice is irrelevant to both fit and forecast; see Currie (2020). The M5(S) model is a *P*-spline-smoothed variant of Cairns *et al.* (2006), although $\{\kappa_{0,y}, \kappa_{1,y}\}$ is still projected using a bivariate random walk with drift. The 2DAP and 2DAC models are projected using a second-order penalty functions, i.e., a linear extrapolation on a logarithmic scale. Further technical details of the models and their fitting can be found in Richards *et al.* (2014, Appendices 1–4) and Richards *et al.* (2019, Appendices B, C & D).

For the VaR calculations, we extend the method of Richards *et al.* (2014) as follows. Data are selected for a given age range and date range, e.g., ages $x_{\min} - x_{\max}$ over the years $y_{\min} - y_{\max}$ as depicted in Figure 2. A stochastic projection model is fitted to this data and the parameters are estimated; we call this the baseline model. The baseline model is then used to generate sample path realisations of mortality rates over the following *n* years. The details of this for the various projection methods are described in Richards *et al.* (2014, Section 8). Where projections are carried out by an ARIMA(p, d, q) model for κ, we restrict our attention to cases with $d = 1$ and where the fitted ARIMA model has a valid estimate of the parameter covariance matrix, i.e., where the leading diagonal contains only positive values. These simulated mortality rates are then used to simulate the population experience from one year to the next, i.e., generate $d_{x,y}$ and $E^c_{x,y}$ for the shaded region in Figure 2. The simulated pseudodata is then appended to the real data and the model is refitted. In most of the calculations in this paper, the refitted model has the same structure and nature as the baseline model. However, in Section 8, we consider the impact of refitting a model of a different nature to the baseline model.

The repeated model-fitting that underlies the methodology here is computationally intensive. Other authors have developed approximations for financial functions under specific models, such as Cairns *et al.* (2011) for the model of Cairns *et al.* (2006) and Liu & Li (2019) for simple drift models for the time index in the models of Lee & Carter (1992) and Renshaw & Haberman (2006). However, we want to keep our approach general and be able to use any model, and we particularly want to use ARIMA models, not drift models, for the time index of the Lee–Carter model (see Kleinow & Richards 2016 for a discussion). For these reasons, we use parallel processing to complete the calculations in a manageable time frame (see Appendix A) and we do not re-estimate the smoothing parameters between simulations (see Appendix B).

https://doi.org/10.1017/S174849952000007X Published online by Cambridge University Press

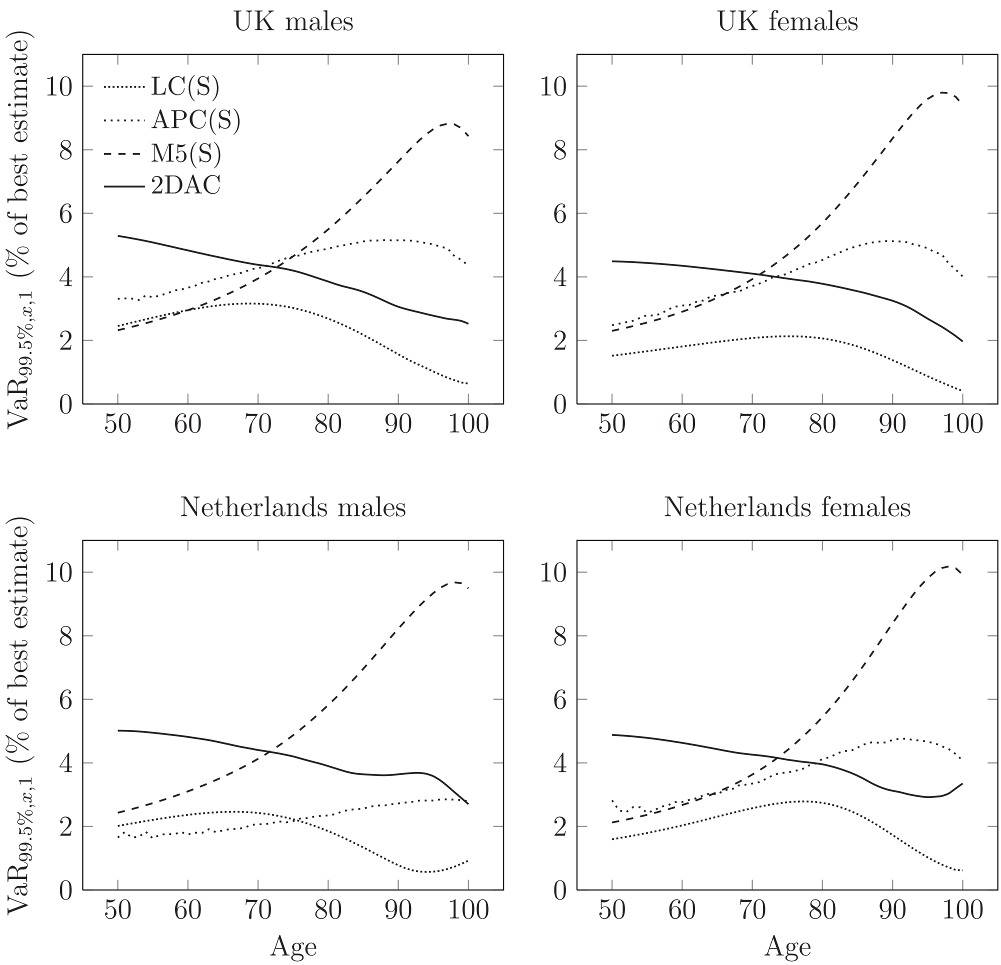

Figure 3. $\text{VaR}_{99.5\%,x,1}$ capital requirements as percentage of best-estimate reserve for a selection of models from Table 1. The age distribution of liabilities is important, but model risk is key – solvency capital requirements have to be set by actuarial judgement. Source: 1,000 refits for each model, immediate annuity liabilities with cash flows discounted at 1% p.a.

5. Insurer Solvency

Solvency II regulations in the EU specify a $\text{VaR}_{99.5\%,x,1}$ approach to longevity trend-risk capital. Figure 3 shows the capital requirements as a percentage of best-estimate reserves for males and females in the UK and Netherlands for a selection of four different models.

As has been documented elsewhere (Richards *et al.* 2014), the capital requirements in Figure 3 vary considerably by age and model choice. The variation by age means that an insurer's longevity trend-risk capital will depend on the age distribution of liabilities. The variation by model means that the final choice of trend-risk capital must be a matter for actuarial judgement – the quite different shapes in Figure 3 show that this actuarial judgement must be informed by the results of a variety of models. Bank of England Prudential Regulatory Authority (2015, p. 10) and Woods (2016, p. 8) describe using members of four model families, and the four models in Figure 3 were chosen for their very different nature.

Solvency capital for longevity trend risk is different for deferred annuities. Figure 4 shows the $\text{VaR}_{99.5\%,x,1}$ capital requirements for annuities where payment is deferred until age 67. As a rule of thumb, whatever the choice of model, the deferred annuity capital requirements at age 50 are

https://doi.org/10.1017/S174849952000007X Published online by Cambridge University Press

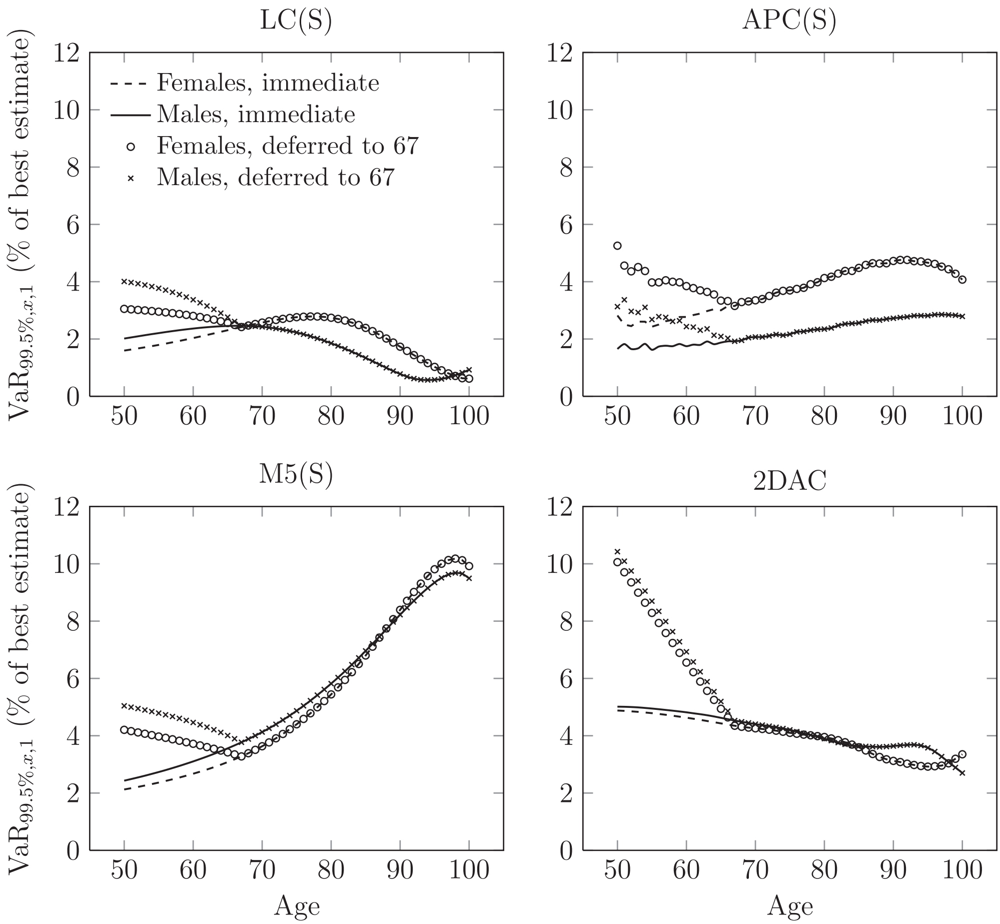

Figure 4. $\text{VaR}_{99.5\%,x,1}$ capital requirements as percentage of best-estimate reserve. Deferred annuities have up to double the capital requirements for longevity trend risk at the age of 50 years. Source: 1,000 simulations for each model, the Netherlands data, immediate and deferred annuity liabilities with cash flows discounted at 1% p.a.

around double those for immediate annuities at the same age. This is one of a number of reasons why insurers and reinsurers prefer to transact business related to annuities in payment, rather than deferred annuities or pensions. Insurers often have a fixed capital budget to support new business, and a greater volume of business can be written if that capital is applied to write immediate annuities rather than deferred annuities. This difference affects insurers' appetite for deferred annuities and thus the keenness of their pricing. As a result, pension schemes may find that it helps to let their deferred liabilities mature before approaching an insurer, i.e., allowing the proportion of deferred liabilities to shrink by waiting for more lives to reach retirement age. This is considered further in Section 7.

One thing that immediate and deferred annuities share is that $\text{VaR}_{99.5\%,x,1}$ and $\text{CTE}_{99\%,x,1}$ produce comparable capital requirements. Figure 5 shows how closely the $\text{CTE}_{99\%,x,1}$ capital requirements shadow the $\text{VaR}_{99.5\%,x,1}$ ones. The picture is similar for the Netherlands females and for males and females in the UK (not shown). The $\text{CTE}_{99\%,x,1}$ capital requirements are in fact slightly higher, but the difference is negligible in the context of the actuarial judgement that needs to be exercised when considering model risk. For all practical purposes, $\text{VaR}_{99.5\%,x,1}$ and $\text{CTE}_{99\%,x,1}$ can be considered equivalent for longevity trend risk.

https://doi.org/10.1017/S174849952000007X Published online by Cambridge University Press

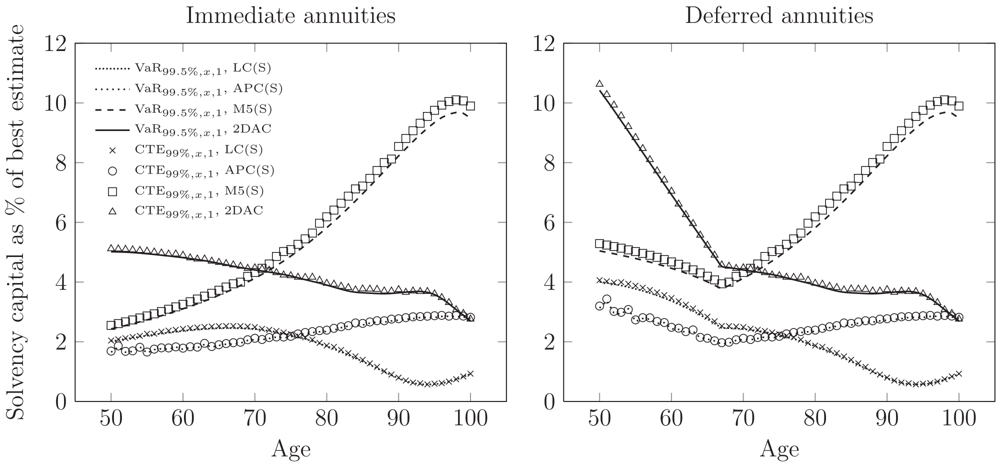

Figure 5. $\text{VaR}_{99.5\%,x,1}$ and $\text{CTE}_{99\%,x,1}$ capital requirements as percentage of best-estimate reserve. The two approaches produce comparable results, with model risk clearly dominant. Source: 1,000 simulations for each model for Netherlands males, immediate- and deferred-annuity liabilities with cash flows discounted at 1% p.a.

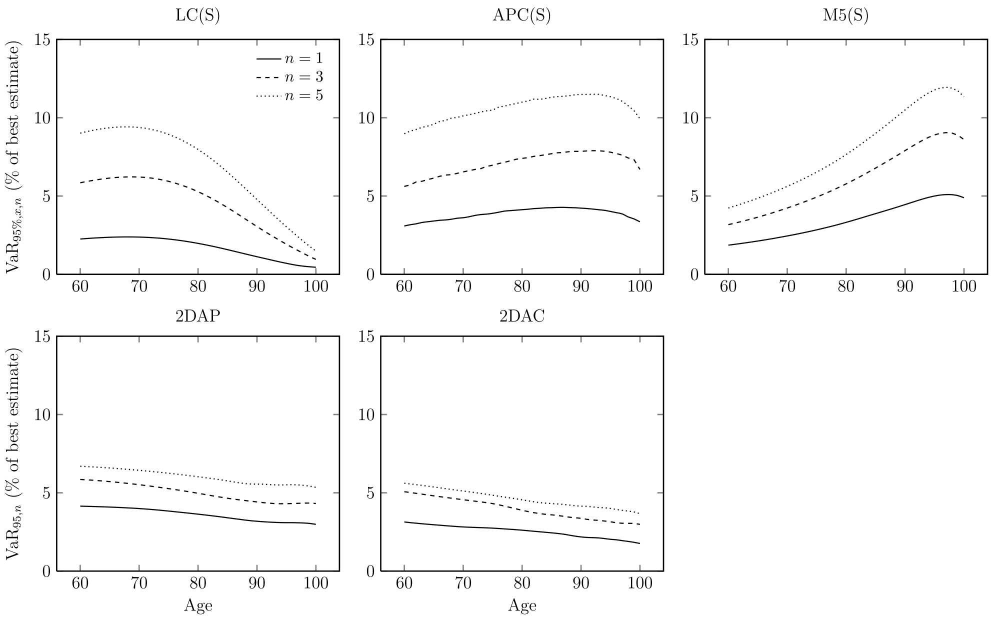

Figure 6. $\text{VaR}_{95\%,x,n}$ capital requirements for immediate annuities to single-life males as percentage of best-estimate reserve for $n \in \{1, 3, 5\}$ using various models. The capital requirements have similar shapes across time horizons, and the figures for a 3-year horizon are generally equidistant between the 1- and 5-year horizons (with the exception of the 2DAC model). As always, model risk is very material. Source: 1,000 refits for each model, data for UK males.

6. Insurer Business Planning

Beyond the 1-year solvency calculations, Solvency II also mandates the form of near-term business planning. Insurers are supposed to carry out VaR-style investigations over the short term using lower p-values under the ORSA. Fixed details are not laid down, but in this section, we consider VaR capital requirements for annuities at a p-value of 95% over 1-, 3- and 5-year horizons. Figures 6 and 7 show the results for UK males and females, respectively. As with the 1-year

https://doi.org/10.1017/S174849952000007X Published online by Cambridge University Press

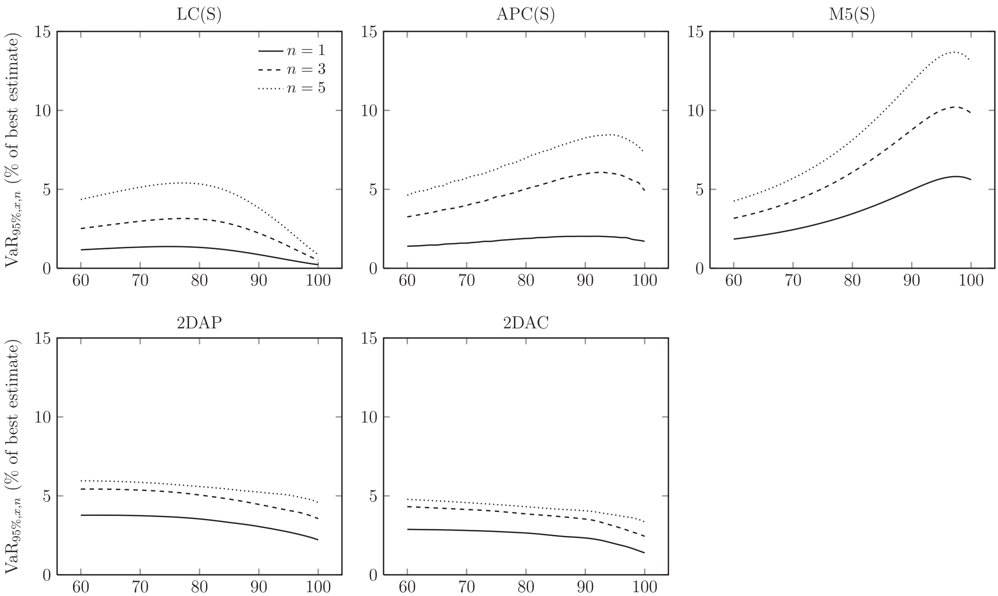

Figure 7. $\mathrm{VaR}_{95\%,x,n}$ capital requirements for immediate annuities to single-life females as percentage of best-estimate reserve for $n \in \{1, 3, 5\}$ using various models. The capital requirements have similar shapes across time horizons, and the figures for a 3-year horizon are often equidistant between the 1- and 5-year horizons, albeit not for the 2DAP and 2DAC models. As always, model risk is very material. Source: 1,000 refits for each model, data for UK females.

solvency capital calculations in Section 5, model risk is a major source of difference, as is variation by age. One new feature is that the capital requirements for a 3-year horizon are not always equidistant between the 1- and 5-year horizons. While the LC(S), M5(S) and APC(S) models have roughly equidistant gaps, the 2DAC model requires relatively little extra for the jump from a 3- to a 5-year horizon. As with solvency capital calculations, business planning capital is a question of actuarial judgement.

7. Pension Scheme Funding Targets

Many defined benefit pension schemes find themselves wanting to buy out their liabilities with a life insurer, but have insufficient funds to afford this in the near term. To achieve the buyout goal, the scheme's funding target can be defined as the buyout premium in n years' time. The scheme's funding rate and investment policy are set accordingly, and the scheme's buyout deficit is monitored over time. This situation is sometimes referred to as the "glide path" to buyout, although the actual financial journey may be rather bumpier than this expression implies.

However, the buyout premium in n years' time is not a static target. Among many other things, insurer buyout prices will be influenced by yield curves, both for nominal and index-linked securities, and forecasts of mortality improvements. To calculate an estimated buyout premium in n years, the pension scheme can use futures prices for the economic assumptions. However, no such market prices exist for mortality improvements, so how can a pension scheme judge what assumptions insurers might use for pricing in n years' time? This obviously cannot be predicted precisely, but a pension scheme can get an idea of the uncertainty and variability of insurers' future mortality forecasts by using the same VaR procedure with the time horizon extended to match the buyout goal. For funding purposes, the pension scheme is also likely to use a lower p-value, say 90% instead of 99.5%. The two panels of Figure 8 illustrate the potential levels of uncertainty caused by longevity trend risk over a 10-year term to buyout.

https://doi.org/10.1017/S174849952000007X Published online by Cambridge University Press

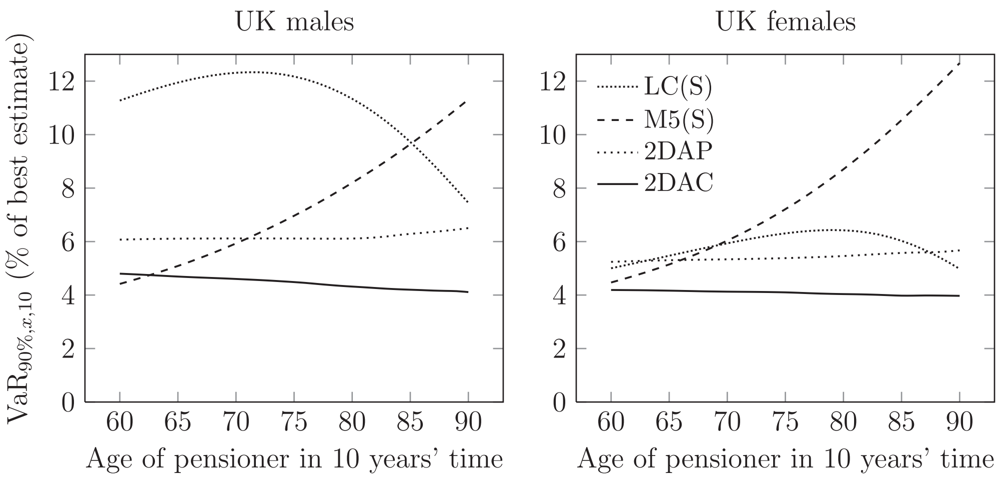

Figure 8. VaR$_{90\%,x,10}$ excess over best-estimate for pensions in payment to males. According to these four models, there is a 10% chance of buyout prices being at least 4% more expensive in 10 years' time due solely to changed expectations of longevity trend risk. Source: 1,000 refits of each model valuing immediate annuities with cash flows discounted continuously at 1% per annum.

On a practical note, the number of models in Figure 8 is rather limited. Specifically, the only model with any year of birth or cohort patterns is the 2D *P*-spline model of Richards *et al.* (2006). This is because the *P*-spline models can cope with the missing cohort triangle in Figure 2, which becomes an increasingly challenging problem for models with individual cohort terms, such as the APC(S) model. Another practical issue is that longer terms for n increase the risk of a restatement of the data of the kinds described in Section 3.

8. Valuing Index-Based Hedging Instruments

A relatively recent financial innovation for managing longevity risk is the index-based hedge. Several such hedges have been transacted in recent years by insurers in the Netherlands; see Blake *et al.* (2018) for an overview. The basic idea is that the owner of longevity risk, e.g., an insurer or pension scheme, uses a hedging instrument defined with reference to population mortality. This critically assumes that the longevity trend risk in the owner's portfolio moves in parallel with the longevity trend risk in the reference population; if it does not, the owner is exposed to basis risk. This basis risk is magnified by the tendency of portfolio liabilities to be heavily concentrated among a select subset of lives; see Richards & Currie (2009, Table 1) for an example of this. The owner of the longevity risk is also left with the volatility arising from individual lifetimes being longer or shorter than expected, known variously as idiosyncratic risk or binomial risk. An index-based hedging instrument is therefore an imperfect risk management tool compared with the more traditional practice of directly (re)insuring the entire longevity risk in the portfolio, which covers trend risk, basis risk and idiosyncratic risk. Such all-encompassing (re)insurance is sometimes known as an indemnity arrangement to contrast it with hedging instruments.

An index-based hedging contract needs a risk-taker on the other side. To meet the needs of such investors, index-based hedges usually have a fixed term, n years say, as opposed to the perpetual nature of an indemnity arrangement. In order to close out the hedging contract, some financial metric needs to be calculated that bears a sufficiently strong relationship to the insurer's balance-sheet liability. An example might be that, at the end of n years, a specified mortality-forecasting model would be fitted to the data available and an annuity factor would be calculated using the resulting best-estimate projection. In such circumstances, the financial metric is the now-familiar

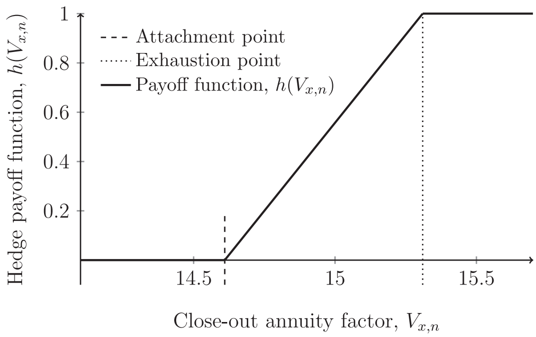

Figure 9. Example hedge pay-off function, $h(V_{x,n})$, showing the attachment and exhaustion points for the contract analysed in Table 2 for $x = 70$.

annuity factor, $V_{x,n}$. Since there is no market price for the hedge contract, we need to use a model to value it (so-called "mark-to-model"). However, while the model used for the closeout calculation will be written into the hedge contract, there is no guarantee that the mortality rates over the contract term will follow this model. In this section, we will explore the impact of model risk on the pricing and valuation of such index-based contracts.

For the purpose of illustration, we assume a term of $n = 15$ years. Since most recent index-based hedges have been written with Dutch insurers (Blake, *et al.* 2018, pp. 44, 48, 57–58), we will use mortality data for the Netherlands. For our closeout forecast model, we will use the unsmoothed Lee–Carter model for its simplicity and widespread acceptance, although Currie (2013) notes that a smoothed version of the Lee–Carter model is unambiguously superior in terms of forecasting quality. For our closeout metric, we will use an annuity factor for a life aged x in n years, with cash flows discounted continuously at an interest rate of 1% per annum. Since this annuity factor is unknown at the point of valuation of the index-based hedge, we treat it as a random variable, $V_{x,n}$, whose value will be observed after n years when the final forecast is made and the annuity factor can be calculated. We will follow Cairns & El Boukfaoui (2019) in not using the annuity factor "as is", but instead define the hedge pay-off function, $h(V_{x,n})$, as follows:

$$h(V_{x,n}) = \max\left(0, \min\left(\frac{V_{x,n} - AP}{EP - AP}, 1\right)\right) \tag{3}$$

where AP is the attachment point, i.e., the minimum value the annuity factor has to reach before any payment is made to the insurer, and EP is the exhaustion point, i.e., the annuity factor above which the payout is capped. $h(V_{x,n})$ therefore takes values in [0, 1], as depicted in Figure 9, and it can be scaled by a nominal amount to provide a suitable hedge value.

Using the same unsmoothed Lee–Carter model for both generating the sample paths and calculating the closeout value of the hedge, we further follow Cairns & El Boukfaoui (2019) in setting $AP = Q_{60\%,n}$ and $EP = Q_{95\%,n}$. If we use the same model to generate the sample paths over the following $n = 15$ years as the model used to set the attachment point, then there is by definition a probability of 0.4 of a non-zero pay-off. Using 1,000 15-year simulations of sample paths according to the unsmoothed Lee–Carter model, we agree the payout probability of 0.4; the average pay-off of $h(V_{70,10})$ is 0.338 (conditional on there being a payout, i.e., $\mathbb{E}[V_{70,10} | V_{70,10} > AP]$). The insurer and investor can then negotiate what price should be paid to the investor to enter into this contract, allowing for further items like expenses, discounting and profit margins.

But how do these figures change if mortality rates over the next n years follow a different process than the one implied by the closeout forecasting model? To illustrate this, we use the unsmoothed Lee–Carter model for both the pay-off calculation and setting the attachment and

https://doi.org/10.1017/S174849952000007X Published online by Cambridge University Press

Table 2. Impact of different sample-path models on pay-off of 15-year longevity hedge.

Model for sample paths	Payout probability	Mean pay-off
LC	0.400	0.338
M5(S)	0.548	0.499
2DAP	0.184	0.262
2DAC	0.841	0.411

Source: own calculations using population data for males in the Netherlands, ages 50–104, 1971–2016. The attachment point is 14.61, being the 60th percentile of the annuity factor under the unsmoothed Lee-Carter model, while the exhaustion point is 15.31, being the 95th percentile. The hedge pay-off function is shown in Figure 9.

Table 3. Impact of different sample-path models on pay-off of 15-year longevity hedge.

Model for sample paths	Payout probability	Mean pay-off
LC	0.397	0.413
M5(S)	0.495	0.553
2DAP	0.000	n/a
2DAC	0.004	0.062

Source: own calculations using population data for females in the Netherlands, ages 50–104, 1971–2016. The attachment point is 17.02, being the 60th percentile of the annuity factor under the unsmoothed Lee-Carter model, while the exhaustion point is 17.90, being the 95th percentile.

exhaustion points, but vary the model generating the sample paths over the n years between setting up the hedge and closing it out. Tables 2 and 3 show the resulting pay-off probabilities and mean pay-offs under some alternative models for males and females.

The diverse results in Tables 2 and 3 do not necessarily affect the hedge contract's effectiveness. If the portfolio's longevity trend truly moves in parallel with that of the reference population, then the hedge contract will be effective if the insurer's projection basis in n years' time is the Lee–Carter model. However, what Tables 2 and 3 do show is the difficulty in placing a value on the hedge contract, either for the purpose of pricing at outset or for putting on the balance sheet once written. Model risk is very significant in valuing the hedge pay-off – if mortality rates were to follow the 2DAC model, then the payout probability would be more than doubled for males but essentially zero for females. Of course, the converse could also apply: if the hedge contract were defined with reference to the 2DAC model and mortality rates actually followed the Lee–Carter model, the payout probability and mean pay-off would both be substantially lower for males, but considerably higher for females.

Placing a value on such a hedge contract is therefore not a purely mathematical exercise, but necessarily involves substantial actuarial judgement. Such judgement needs to take account of model risk by exploring the results under various different models, but it is clear that valuing an index-based hedge is even more sensitive to the choice of model than setting solvency capital requirements for either immediate or deferred annuities (see Section 5). A follow-on question for regulators is what solvency capital relief (if any) can be granted for holding such index-based contracts? A further conclusion from Tables 2 and 3 is that increasing the number of simulations would be irrelevant; there is little benefit in increasing the accuracy of estimated payout probabilities in the face of such overwhelming model risk.

As with the pension fund "glide path" to buyout in Section 7, the choice of models that can be used in Tables 2 and 3 is limited by the large number of missing cohorts in the shaded area of Figure 2. Of course, the data in this triangle will not be missing at the point the hedge contract is closed. However, the risk of a restatement of the past data – of the kinds described in Section 3 –

https://doi.org/10.1017/S174849952000007X Published online by Cambridge University Press

is even greater with longer terms like $n = 15$. As a result, hedge contracts need carefully drafted clauses to handle restated data. A particularly challenging scenario would be where a hedge contract terminated around the same time as data were restated, as the opposing parties would favour different versions of the data.

9. Conclusions

Solvency capital requirements from 1-year views of longevity trend risk are very similar between the VaR method at a p-value of 99.5% and the CTE method at 99%. Deferred annuities are more sensitive to longevity trend risk than annuities in payment; at the age of 50 years, the 1-year solvency capital requirement is doubled. This has implications not only for insurer solvency calculations but also for pension schemes aiming to buy out their liabilities.

Model risk is important throughout, and solvency capital can ultimately only be determined by actuarial judgement. Such judgement should be informed by the results of using several different models. Both VaR and CTE capital requirements can be calculated from the same sample, and an adequate sample size is obviously necessary to minimise the standard error of VaR-style order statistics. However, since actuarial judgement is the final stage in determining capital requirements, too much focus on increasing the sample size risks detracting from the most important task of exploring multiple models to inform that judgement. This is particularly the case for valuing index-based hedge contracts, where model risk becomes almost overwhelming.

Acknowledgements. The authors thank Kevin Armstrong for helpful comments and an anonymous reviewer. Any errors or omissions remain the responsibility of the authors. All models were fitted using the Projections Toolkit. Graphs were done in pgfplots and tikz, while typesetting was done in LaTeX. Torsten Kleinow acknowledges financial support from the Actuarial Research Centre (ARC) of the Institute and Faculty of Actuaries through the research programme on "Modelling, Measurement and Management of Longevity and Morbidity Risk".

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrica*, **52**, 317–333. ISSN 0033–3123. doi: https://doi.org/10.1007/BF02294359.

Bank of England Prudential Regulatory Authority. Solvency II: internal model and matching adjustment update. Letter to UK-regulated insurers and reinsurers, March 2015. http://www.bankofengland.co.uk/pra/Pages/publications/ps/2016/ps3316.aspx.

Blake, D., Cairns, A.J.G., Dowd, K. & Kessler, A.R. (2018). Still living with mortality: the longevity risk transfer market after one decade. *British Actuarial Journal*, **24**, e1. doi: https://doi.org/10.1017/S1357321718000314.

Boumezoued, A. (2020) Improving HMD mortality estimates with HFD fertility data. *North American Actuarial Journal*, 1–25.

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. doi: https://doi.org/10.1111/j.1539-6975.2006.00195.x.

Cairns, A.J.G. & El Boukfaoui, G. (2019) Basis risk in index based longevity hedges: a guide for longevity hedgers. *North American Actuarial Journal*, 1–22.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D. & Khalaf-Allah, M. (2011). Mortality density forecasts: an analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, **48**, 355–367.

Cairns, A.J.G., Blake, D., Dowd, K. & Kessler, A. (2015). Phantoms never die: living with unreliable mortality data. *Journal of the Royal Statistical Society, Series A*. doi: https://doi.org/10.1111/rssa.12159.

Currie, I.D. (2011). Modelling and forecasting the mortality of the very old. *ASTIN Bulletin*, **41**, 419–427. doi: 10.2143/AST.41.2.2136983.

Currie, I.D. (2013). Smoothing constrained generalized linear models with an application to the Lee-Carter model. *Statistical Modelling*, **13**(1), 69–93. doi: https://doi.org/10.1177%2F1471082X12471373

Currie, I.D. (2020). Constraints, the identifiability problem and the forecasting of mortality. *Annals of Actuarial Science*, **14**, 537–566

Currie, I.D., Durban, M. & Eilers, P.H. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, **4**, 279–298. doi: https://doi.org/10.1191%2F1471082X04st080oa

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. doi: https://doi.org/10.1177%2F1471082X0600700103

Eilers, P.H.C. & Marx, B.D. (1996). Flexible smoothing with *B*-splines and penalties. *Statistical Science*, **11**(2), 89–121. ISSN 08834237. http://www.jstor.org/stable/2246049.

Hardy, M.R. (2006). An introduction to risk measures for actuarial applications. *Casualty Actuarial Society*.

Harrell, F.E. & Davis, C.E. (1982). A new distribution-free quantile estimator. *Biometrika*, **69**, 635–640. ISSN 00063444. doi: https://doi.org/10.1093/biomet/69.3.635. http://www.jstor.org/stable/2335999.

Hyndman, R. (2013). Bug 15396 – Initialization of regression coefficients in arima(). https://bugs.r-project.org/bugzilla/show_bug.cgi?id=15396 .

Kleinow, T. & Richards, S.J. (2016). Parameter risk in time-series mortality forecasts. *Scandinavian Actuarial Journal*, **2016**(10), 1–25. doi: https://doi.org/10.1080/03461238.2016.1255655.

Langou, J. (2010). Lapack 3.2.2 release notes. http://www.netlib.org/lapack/lapack-3.2.2.html.

Lee, R.D. & Carter, L. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, **87**, 659–671. ISSN 01621459. http://www.jstor.org/stable/2290201.

Liu, Y. & Li, J.S.-H. (in press). An efficient method for mitigating longevity value-at-risk. *North American Actuarial Journal,* Special Issue for the 13th International Longevity Risk and Capital Markets Solutions Conference.

Macdonald, A.S., Richards, S.J. & Currie, I.D. (2018). *Modelling Mortality with Actuarial Applications*. Cambridge University Press. ISBN 978-1-107-04541-5.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

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

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). doi: https://doi.org/10.1017/S1357321700005675.

Richards, S.J., Currie, I.D., Kleinow, T. & Ritchie, G.P. (2019). A stochastic implementation of the APCI model for mortality projections. *British Actuarial Journal*, **24**, e13. doi: 10.1017/S1357321718000260. https://www.longevitas.co.uk/apci.

Richards, S.J., Currie, I.D. & Ritchie, G.P. (2014). A value-at-risk framework for longevity trend risk. *British Actuarial Journal*, **19**(1), 116–167. doi: https://doi.org/10.1017/S1357321712000451. https://www.longevitas.co.uk/var.

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 (with discussion). doi: https://doi.org/10.1017/S1357321700004682.

Wilmoth, J.R., Andreev, K., Jdanov, D., Glei, D.A. & Riffe, T. (2017). Methods Protocol for the Human Mortality Database. https://www.mortality.org/Public/Docs/MethodsProtocol.pdf.

Woods, S. (2016). Reflections on the 2015 Solvency II internal model approval process. *Bank of England Prudential Regulation Authority*. http://www.bankofengland.co.uk/pra/Pages/publications/ps/2016/ps3316.aspx.

Appendices

A. Calculation platform

All models in this paper were fitted using R Core Team (2017), version 3.3.3. Note that versions of the `arima()` function in R prior to v3.0.1 could sometimes return parameters for a local maximum of the likelihood; see Hyndman (2013) for details. Despite this bug being fixed, R's `arima()` function can still return results dependent on the initial parameter values supplied. To minimise the risk of selecting a model that is a local optimum, we make several calls to `arima()` with different initial values and choose the parameters with the lowest AIC.

Our installation of R in turn relied on version 3.7.0 of the basic linear algebra subprograms (BLAS/LAPACK). Note that versions of LAPACK prior to 3.2.2 contained an unfixable bug in the QR decomposition routine; see Langou (2010). This library is used by R's `svd()` function, among other matrix functions, which means that earlier versions of LAPACK will produce slight differences in fitted parameters in R, especially where smoothing is involved.

Calculations were performed on an eight-core Debian Linux virtual server. The core count is relevant because we use process parallelism to reduce the time taken to perform the large number of model refits. A custom Java program balanced at most seven concurrent R processes at any one time. Table A.1 shows the substantial reduction in run times afforded by parallel processing, and the near-linear scaling with the number of cores used.

Table A.1. Elapsed times in seconds for $\text{VaR}_{99.5\%,x,1}$ calculations based on 1,000 simulations using an unsmoothed Lee-Carter model. The model refits are independent of each other and so can be spread over a varying number of processes. The rightmost column shows the time taken as a percentage of the single-process case.

Number of processes	Time taken	Time factor (%)
1	5,692	100
4	1,495	26
7	867	15

Source: own calculations.

Table B.1. Elapsed times in seconds for 1,000 $\text{VaR}_{99.5\%,x,1}$ calculations using Lee–Carter models with differing degrees of smoothing. Calculations were spread over seven processes in each case and the elapsed time is shown as a percentage of the unsmoothed case.

Lee–Carter variant	Time taken	Time factor (%)
Unsmoothed, $\log \mu_{x,y} = \alpha_x + \beta_x \kappa_y$	867	100
Smoothing β_x, as per Delwarde *et al.* (2007)	1,505	174
Smoothing α_x and β_x, as per Currie (2013)	2,450	283

B. Smoothing

Many stochastic mortality models are over-parameterised. One consequence is that forecasts can have undesirable features, such as mortality rates crossing over at adjacent ages. Delwarde *et al.* (2007) showed how to improve forecast performance with the Lee–Carter model by smoothing one of the age-related parameters. Currie (2013) went further and showed how to smooth both age-related parameters in the Lee–Carter model and how this also reduced the effective dimension of the model without compromising the overall fit. Smoothing is therefore desirable for many age-related parameters, although some (non-age-related) parameters are unsuitable for smoothing; Richards *et al.* (2019) discuss when parameters should not be smoothed in mortality models.

A drawback of smoothing with penalised splines is that it is computationally intensive in the context of the repeated model refits required for the calculations in this paper. This is particularly the case where smoothing parameters are re-optimised for each model refit; see Currie *et al.* (2004). Table B.1 shows how, as the amount of smoothing increases in a model, so does the run time. Compared to Table A.1, we can see that re-optimised smoothing undoes much of the benefit of parallel processing.

However, in a VaR-style calculation, we usually add a relatively small amount of extra pseudodata to the real data, and we simulate the new pesudodata with a model which should have similar smoothing characteristics to the fitted model. To accelerate run times in this paper, we therefore applied the smoothing parameters of the baseline model without re-estimating the smoothing parameters anew each time. This is a departure from the VaR results cited in previous works like Richards *et al.* (2014), Kleinow & Richards (2016) and Richards *et al.* (2019), and a natural question is whether this "baseline-driven smoothing" materially changes the resulting capital requirements? Figure B.1 shows the $\text{VaR}_{99.5\%,x,1}$ capital requirements using baseline-driven smoothing versus reoptimised smoothing; there is no material difference in the capital requirements and so there is no reason not to accept the performance improvement from baseline-driven smoothing shown later in Table B.2.

In addition to the smoothing parameters, some models allow for over-dispersion, i.e., the tendency for mortality counts to exhibit greater variation than the Poisson assumption allows. Examples of such models are the 2D *P*-spline-smoothed models in Richards *et al.* (2006). Such models have an additional over-dispersion parameter, which is estimated from the data and which

https://doi.org/10.1017/S174849952000007X Published online by Cambridge University Press

Table B.2. Elapsed times in hours for $\text{VaR}_{99.5\%,x,1}$ calculations based on 5,000 refitted models with re-optimised smoothing parameters and baseline-smoothed parameters, together with the time factor showing the baseline case as a percentage of the re-optimised case. In each case, the simulations are spread over seven processes.

	Smoothing parameters		
Model	(i) re-optimised	(ii) baseline values	(ii)/(i) ×100%
LC(S)	1.06	1.04	98
M5(S)	3.03	1.01	33
APC(S)	4.12	2.07	50
2DAP	17.54	1.11	6
2DAC	30.19	2.08	7

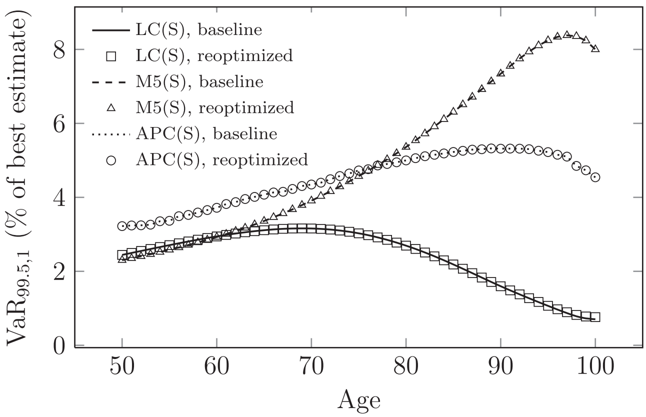

Figure B.1. $\text{VaR}_{99.5\%,x,1}$ capital requirements as percentage of best-estimate reserve, calculated using (a) full reoptimisation of smoothing parameters and (b) baseline smoothing applied to all simulations. Baseline-driven smoothing does not change the capital requirements for any of the three models shown. Source: 5,000 simulations for each model fitted to data for UK males, immediate-annuity liability with cash flows discounted at 1% p.a.

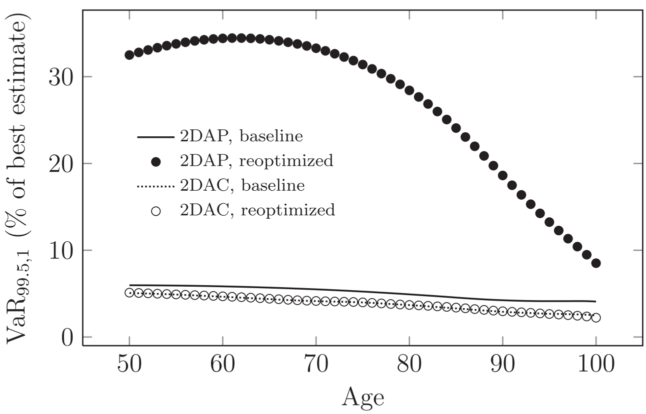

Figure B.2. $\text{VaR}_{99.5\%,x,1}$ capital requirements as percentage of best-estimate reserve, calculated using (a) full reoptimisation of smoothing parameters and (b) using baseline smoothing applied to all simulations. Baseline-driven smoothing does not affect the results from the 2DAC model, but produces more sensible results for the 2DAP model. Source: 5,000 refits of each model for UK males, immediate annuity liability with cash flows discounted at 1% p.a.

https://doi.org/10.1017/S174849952000007X Published online by Cambridge University Press

can therefore be re-estimated along with the smoothing parameters in each VaR refit (or not, if we apply the baseline value of the over-dispersion parameter along with the baseline values of the smoothing parameters). Figure B.2 shows the effect on capital requirements of re-estimating both the over-dispersion parameter and the smoothing parameters. While there is no material change for the 2DAC model, for the 2DAP model there is a wild difference between the results; the baseline-smoothed results are clearly more sensible than the re-optimised ones.

Figure B.2 leads us to prefer baseline-driven smoothing because it produces more stable and sensible results for models with over-dispersion parameters. Figure B.1 shows that there is no difference in results for models without over-dispersion parameters. However, there is an additional benefit of faster run times, sometimes dramatically so, as shown in Table B.2.

Cite this article: Richards SJ, Currie ID, Kleinow T and Ritchie GP (2020). Longevity trend risk over limited time horizons. *Annals of Actuarial Science* **14**, 262–277. https://doi.org/10.1017/S174849952000007X